



**University of
Nottingham**

UK | CHINA | MALAYSIA

Upscaling Robot-assisted Interventional Tool Manipulations based on Multimodal Endovascular Data Analysis

PhD in Computer Science & Operations Research

**Wenjing Du
20473779**

**Supervised by
Professors Boon Giin Lee and Lei Wang**

School of Computer Science, University of Nottingham Ningbo China

January 2025

Acknowledgements

The completion of this thesis would not have been possible without the support and love of many people in my life, to whom I owe my deepest gratitude and thanks.

I am deeply grateful to my supervisor, Professor Boon Giin Lee, for accepting me into the School of Computer Science at the University of Nottingham Ningbo China (UNNC) and for his support throughout my doctoral studies. The discussions we had were always enlightening and provided me with the clarity needed to guide through challenging aspects of my research. Your feedback was not only constructive but also inspiring, motivating me to push the boundaries of my knowledge and skills. The time and effort you invested in refining my manuscripts, identifying appropriate journals, and guiding me through the submission process were beyond commendable. Your support has undoubtedly played a pivotal role in my academic success. Your recommendation for my PGR Fellowship application to the UNNC-SIAT Joint Recruitment and Cultivation of PhD Students program was instrumental in my achieving this degree. I would also like to acknowledge the committee of the UNNC PGR President's Fellowship Programme for sponsoring my doctoral program. Special thanks to Professor Boon Giin Lee for the vital discussions regarding our research and his candid advice on progressing through challenging times. My profound gratitude goes to all team members, especially Renjie Wu and Haotian Bai, who helped me during my doctoral studies.

I am also grateful to my co-supervisor, Professor Lei Wang, for accepting me into the Research Centre for Medical Robotics and Minimally Invasive Surgical Devices at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (SIAT-CAS), and for his academic guidance and support throughout my doctoral studies. I want to thank you for your profound guidance in my research work. Your expertise and insights were crucial in shaping my understanding and approach to complex research problems. Your assistance in the publication of my academic papers was equally remarkable. You meticulously reviewed my drafts, providing detailed suggestions that significantly enhanced the quality of my work. Your contributions have been invaluable to both my academic and personal growth. I also acknowledge the Research Centre for Medical Robotics and Minimally Invasive Surgical Devices at SIAT-CAS for providing a conducive research environment for my degree program. My profound gratitude goes to all members of the Research Centre, especially Wenke Duan, Olatunji Mumini Omisore, Xingyu Chen, Toluwanimi Oluwadara Akinyemi, Guanlin Yi, and all the visiting students who worked with us during my doctoral studies.

My sincere appreciation goes to my husband. Words cannot fully express my gratitude for your understanding throughout this program. Thank you for your patience and support. To my

daughter, I am deeply grateful for your patience and for teaching me how to balance a PhD and motherhood. I apologize for not giving you the attention you deserved during this period and for not being able to provide you with more time to play with Dad. My family, in-laws, and loved ones made great sacrifices to move to Shenzhen. Thus, I sincerely thank my priceless parents. Without your selfless sacrifices, I would not be where I am today, and this thesis would not have been possible without your support. I also thank my siblings for their constant support and encouragement.

Finally, I would like to thank all my colleagues at SIAT, especially my friends who provided countless help, advice, and a conducive living environment over the past three years.

Wenjing Du

Ningbo, January 2025.

List of Publications

- [1] **Wenjing Du**, Olatunji Mumini Omisore, Wenke Duan, Lu Gan, Toluwanimi Oluwadra Akinyemi, Boon-Giin Lee, Jiang Liu, Lei Wang. Exploring operators' natural behaviors to predict catheterization trial outcomes in robot-assisted intravascular interventions. *IEEE Transactions on Medical Robotics and Bionics*. 2022, 4(3):682-695. (IF=3.4, JCR 2)
- [2] **Wenjing Du**, Guanlin Yi, Olatunji Mumini Omisore, Wenke Duan, Toluwanimi Akinyemi, Xingyu Chen, Jiang Liu, Boon-Giin Lee, and Lei Wang. Analyzing Surgeon-Robot Cooperative Performance in Robot-assisted Intravascular Catheterization, *IEEE Transactions on Human-Machine Systems*, 2024, 54(6):698-710. (IF=3.5, JCR 1)
- [3] **Wenjing Du**, Guanlin Yi, Olatunji Mumini Omisore, Wenke Duan, Xingyu Chen, Toluwanimi Akinyemi, Jiang Liu, Boon Giin Lee, Lei Wang. Guidewire Endpoint Detection Based on Pixel-Adjacent Relation during Robot-Assisted Intravascular Catheterization: In Vivo Mammalian Models, *Advanced Intelligent Systems*, 2024, 6(4):2300687. (IF=6.8, JCR 1)
- [4] **Wenjing Du**, Guanlin Yi, Olatunji Mumini Omisore, Wenke Duan, Toluwanimi Oluwadara Akinyemi, Xingyu Chen, Lei Wang, Boon-Giin Lee, Jiang Liu. Guidewire Endpoint Detection Based on Pixel Adjacent Relation in Robot-assisted Cardiovascular Interventions, 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 2023 Jul 24, 1-5.
- [5] Wenke Duan, Toluwanimi Akinyemi, **Wenjing Du**, Jun Ma, Xingyu Chen, Fuhao Wang, Olatunji Omisore, Jingjing Luo, Hongbo Wang, Lei Wang. Technical and clinical progress on robot-assisted endovascular interventions: A review. *Micromachines*, 2023, 14(1):197. (IF=3.0, JCR 2)
- [6] Toluwanimi Oluwadara Akinyemi, Olatunji Mumini Omisore, **Wenjing Du**, Wenke Duan, Xingyu Chen, Guanlin Yi, Lei Wang. Interventionalist hand motion recognition with convolutional neural network in robot-assisted coronary interventions, *IEEE Sensors Journal*, 2023, 23(15): 17725-17736. (IF=4.3, JCR 1)
- [7] Olatunji Mumini Omisore, Toluwanimi Oluwadara Akinyemi, **Wenjing Du**, Wenke Duan, Rita Orji, Thanh Nho Do, Lei Wang. Weighting-based deep ensemble learning for recognition of interventionalists' hand motions during robot-assisted intravascular catheterization. *IEEE Transactions on Human-Machine Systems*. 2022, 53(1):215-27. (IF=3.5, JCR 1)
- [8] Olatunji Mumini Omisore, Toluwanimi Oluwadara Akinyemi, Wenke Duan, **Wenjing Du**, Lei Wang. Multi-Lateral Branched Network for Tool Segmentation during Robot-assisted Endovascular Interventions, *IEEE Transactions on Medical Robotics and Bionics*, 2024. (IF=3.4, JCR Q2)
- [9] Wenke Duan, Zihao Li, Olatunji M Omisore, **Wenjing Du**, Toluwanimi O Akinyemi, Xingyu Chen, Xing Gao, Hongbo Wang, Lei Wang. Development of an Intuitive Interface with Haptic Enhancement for Robot-Assisted Endovascular Intervention. *IEEE Transactions on Haptics*. 2023.4. (IF=2.4, JCR 2)
- [10] Toluwanimi O Akinyemi, Olatunji M Omisore, Xingyu Chen, Wenke Duan, **Wenjing Du**, Guanlin Yi, Lei Wang. Adapting neural-based models for position error compensation in robotic catheter

- systems. *Applied Sciences*. 2022, 12(21):10936. (IF=2.5, JCR 2)
- [11] Olatunji Mumini Omisore, Guanlin Yi, Yuhong Zheng, Toluwanimi Oluwadara Akinyemi, Wenke Duan, **Wenjing Du**, Xingyu Chen, Lei Wang. Endovascular Tool Segmentation with Multi-lateral Branched Network during Robot-assisted Catheterization, 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 2023 Jul 24,1-4.
- [12] Xingyu Chen, Yinan Chen, Wenke Duan, Toluwanimi Oluwadara Akinyemi, Guanlin Yi, Jie Jiang, **Wenjing Du**, Olatunji Mumini Omisore. Design and evaluation of a learning-based vascular interventional surgery robot. *Fibers*. 2022, 10(12):106. (IF=4.0, JCR 1)

Participated Research Projects

- [1] National Natural Science Foundation of China (grant no. U1713219) Title: Research on force sensing and guidewire manipulation mechanism of wearable precision positioning interventional robot
(穿戴式精确定位介入手术机器人的力觉感知与导丝操控机理研究)
- [2] National Key Research & Development Plan (grant no. 2019YFB1311700) Title: Research on the key technology and application of multi-instrument cooperative delivery robot for vascular interventional surgery for complex lesions
(面向复杂病变的多器械协同递送血管介入手术机器人关键技术及应用研究)
- [3] National Natural Science Foundation of China (grant no. U21A20480) Title: Key robotic-assisted technologies on the accurate positioning and aspiration biopsy for the peri-pulmonary small nodules
(机器人辅助肺周小结节精准定位与穿刺活检关键技术研究)
- [4] National Key Research & Development Plan (grant no. 2024YFF1206900) Title: Research on key technologies of scene navigation in robot-assisted intelligent surgery and Hybrid operating room
(机器人辅助智慧手术与复合手术室情景导航关键技术研究)

Abstract

The integration of robotic technology has advanced endovascular intervention towards a new paradigm. Unlike traditional endovascular intervention that require operators to wear heavy protective suits and expose themselves to prolonged X-ray radiation, innovative master-slave robotic systems are being developed for endovascular intraluminal procedures. Surgeons supervise instrument positioning from outside the operating room, using a master robot to control a slave robot for operations such as guidewire delivery and stent release. This robotic approach offers enhanced safety and precision by eliminating human-induced errors such as shaking, while also reducing radiation exposure and operating time to improve surgical efficiency and reduce complications. Interventional robots, presently in development, are designed for treating endovascular diseases by manoeuvring through stenosis in endovascular paths. These designs have led to the creation of intuitive manipulation models for robot-assisted surgeries. The lack of haptic feedback significantly affects task performance in anatomical spaces. To counteract this, visual information is used to improve intuitive manipulation. This involves multi-sensor data modelling and visual perception to ensure accurate tool manipulation. Statistical analysis using multi-sensor helps identify manipulation patterns, reaching an accuracy of 93.96% in distinguishing between successful and unsuccessful robot-assisted tasks across fourteen patterns, revealing the internal relevance between tool manipulation and systems for specific robot-assisted surgical tasks.

The effectiveness and safety of tool manipulation rely heavily on the seamless collaboration between the surgeon and the robotic system. Intuitive manipulation plays an important role in improving the performance of surgical tasks using robot assistance, influencing both the force or speed of manipulation and the degree of cooperation between the operator and the robot. This process involves utilizing machine learning based on manipulation patterns to assess the operator-robot synergy, aiming to calculate the synergy ratio between the actions of the operator and the real-time response of the robot. This study used a convolution neural network considering factors such as no delay, constant delay, and variable delay to calculate the synergy ratio for precise prediction of the operator's pattern of manipulation associated with the movement of the controlled robot. Simulations with a vascular interventional robot indicate that the model performs excellently in recognizing manipulation patterns and calculating the synergy ratio. In addition, operators experienced in manual percutaneous coronary interventions show significantly improved cooperative performance with the robotic system over inexperienced operators, achieving synergy ratios of 89.66%, 90.28% and 91.12% in the three delay considerations. Experiments involving animals and simulations with multi-sensor

data-driven modelling demonstrate that intuitive manipulation significantly impacts robot-assisted surgical task performance and operator-robot synergy. Thus, improving intuitive manipulation provides significant aid in accurate and safe instrument delivery in robot-assisted interventional surgeries.

The master-slave vascular interventional robots assisted surgery primarily depends on the surgeon's use of real-time 2- or 3-dimensional medical imaging to match the patient's anatomy with the images, ensure precise tool positioning, and enhance manipulation to reduce contact with nontarget tissues. Thus, accurate visual perception of the endovascular instrument's trajectory for providing guidewire position and direction details to offset the absence of tactile feedback, is crucial for instrument navigation, reducing vessel wall injury. To achieve this, an eight-neighbourhood-based deep neural network was designed to detect the guidewire endpoint and its maximum bending region. The method operates in two phases. The first phase involves the design of an improved U-Net network, which segments the guidewire to identify regions containing endpoints, limiting interference from other anatomical elements and imaging noise. The second phase involves extracting skeletons, removing bifurcation points, and repairing breaks using pixel correlations in eight-neighbourhood zones.

Initial results demonstrate that the eight-neighbourhood strategy achieves a mean pixel error of 2.02 pixels on a rabbit dataset and 2.13 pixels on a porcine dataset, outperforming state-of-the-art approaches. This approach, reliant on pixel adjacency relationships based on segmentation quality, performs best when the segmentation is strong, showing few false negatives and false positives. However, the detection results are unsatisfactory mainly due to poor segmentation performance. To further improve visual perception of surgical instruments, a multi-branch feature fusion with a triple-pyramid network was designed to refine surgical instrument segmentation, aiding surgical decision-making, determining procedural stages, and identifying critical surgical zones. This model utilises an encoder-decoder architecture that features a sophisticated Visual Geometry Group 13 encoder for improved edge and texture detection, along with a triple pyramid decoder that improves feature maps. The method attains a mean intersection-over-union of 95.54% in a multimodal fusion dataset, delivering crucial visual input for robotic endovascular interventions.

In conclusion, this thesis focuses on developing multi-sensor-based modelling methods for surgeon intuitive manipulation behaviours and visual perception techniques to enhance tool manipulation in an underactuated master-slave vascular interventional robotic system with spatial flexibility. Key areas of emphasis include the creation of efficient models for highly accurate robot tool manipulation along spatially flexible paths, particularly for accessing complex and narrow endovascular pathways. The studies are grounded in a developed vascular interventional robotic system and aim to establish an intuitive manipulation model that clarifies the relationship between the surgeon's manipulation and the performance of robot-assisted

surgical tasks. Additionally, visual-based modelling is employed to enhance the visual perception of interventional instruments. By enhancing tool manipulation, this approach facilitates the safe and precise delivery of catheters and guidewires through single-port minimal invasion, enabling access to lesion sites along various endovascular pathways.

Keywords: Master-slave robotic system; robot-assisted endovascular interventional surgery; sensory force feedback; intuitive manipulation modelling; pattern recognition; synergy performance; visual perception; guidewire endpoint detection; interventional instrument segmentation

Contents

Acknowledgements	I
List of Publications	III
Abstract	V
Contents	VIII
List of Figures	XI
List of Tables	XIV
List of Abbreviations	XV
Chapter 1: Introduction	1
1.1 <i>Research Background</i>	1
1.2 <i>Research Motivation and Significance</i>	5
1.3 <i>Key Research Challenges, Aim and Objective</i>	8
1.3.1 <i>Efficiency improvement of surgeon-robot interaction</i>	9
1.3.2 <i>Optimization of cooperative manipulation between operator and robot</i>	10
1.3.3 <i>Enhancement of visual perception of interventional instrument in endovascular pathway</i>	10
1.4 <i>Main Contributions</i>	11
1.5 <i>Thesis Organization</i>	12
1.6 <i>Publications</i>	14
Chapter 2: Literature Review	15
2.1 <i>Concept and the Development Trends of Interventional Medicine</i>	15
2.2 <i>Vascular Interventional Robots in Interventional Surgery</i>	18
2.2.1 <i>Rationale of vascular interventional robot development</i>	18
2.2.2 <i>Evolution of VIR system</i>	19
2.2.3 <i>Clinical trials of applications of vascular interventional robots</i>	25
2.3. <i>Intuitive Manipulation of VIR</i>	29
2.3.1 <i>Underactuated intraluminal manipulation modelling</i>	32
2.3.2 <i>Force-based information fusion robot-assisted manipulation</i>	35
2.3.3 <i>Visual perception of interventional instruments for tool manipulation</i>	36
2.4. <i>Technical Challenges and Future Development</i>	39
2.4.1 <i>Cooperative mechanisms of multi-instruments and teleoperation setup</i>	40
2.4.2 <i>Automatic surgery</i>	41
2.5 <i>Endovascular Robotic System Design and Prototyping</i>	41
2.5.1 <i>Master-slave isomorphic robot-assisted system</i>	42

2.5.2 Validation of the robot-assisted interventional system.....	44
2.6 Chapter Summary.....	47
Chapter 3: Modelling and Analysis of Manipulation Pattern	49
3.1 Introduction	49
3.2 Related Work	49
3.3 Multi-Modal Sensing Data Pre-Processing	51
3.3.1 Integrated data acquisition and analysis platform for interventional surgery.....	51
3.3.2 Data acquisition and feature extraction	54
3.4 Multi-Layer Classification System.....	59
3.4.1 Manipulation technical skill framework.....	59
3.4.2 Manipulation technical modelling for control mode classification.....	61
3.4.3 Evaluative index of manipulation technical skill framework	65
3.5 Model Evaluation and Analysis.....	67
3.5.1 Inter-group behaviour analysis in the initial-decision layer.....	68
3.5.2 Inter-group natural pattern analysis in the motion-decision layer	72
3.5.3 Level-S and Level-US pattern analysis in the mixed-decision layer	77
3.6 Chapter Summary.....	78
Chapter 4: Synergy Performance Between Manipulation Behaviour and Robotic System	79
4.1 Introduction.....	79
4.2 Related Work.....	79
4.3 Methodology.....	81
4.3.1 System design.....	81
4.3.2 Signal acquisition and processing.....	83
4.4 Manipulation-Based Synergy Characteristic Modelling.....	87
4.4.1 Manipulation pattern recognition modelling	87
4.4.2 Training strategy and evaluation metric	89
4.4.3 Human-machine synergy evaluation strategy	90
4.5 Cooperative Performance Analysis in Human-Robot Systems.....	91
4.5.1 Performance evaluation of the synergy between operator and robotic system	92
4.5.2 Manipulation time analysis	95
4.5.3 Interaction force analysis	96
4.6 Chapter Summary.....	102
Chapter 5: Visual Data Modelling of Instrument Endpoints for Enhancing Tool Manipulation.....	104
5.1 Introduction	104
5.2 Related Work.....	104

5.3 Methodology.....	106
5.3.1 Dataset acquisition.....	107
5.3.2 First stage: guidewire segmentation methods	107
5.3.3 Second stage: guidewire endpoint detection	110
5.3.4 Training strategy and evaluation index.....	115
5.4 Model Detection Performances.....	116
5.4.1 Segmentation performance.....	116
5.4.2 Comparison of detection performance for heatmap methods.....	119
5.4.3 Performance comparison of heatmap regression methods	121
5.4.4 Comparison of heatmap and regression methods	123
5.4.5 Guidewire bending angle	124
5.5 Chapter Summary.....	126
Chapter 6: Visual Perception Modelling of Interventional Instruments Using Multi-Modal Images Fusion	127
6.1 Introduction.....	127
6.2 Related Work.....	127
6.3 Methodology.....	129
6.3.1 Multi-modal image dataset.....	129
6.3.2 Image pre-processing and data augmentation	131
6.3.3 Surgical instrument segmentation method	133
6.3.4 Loss function design and data training strategy.....	138
6.3.5 Model performance evaluation metrics	140
6.4 Semantic Segmentation Performance Analysis	141
6.4.1 Semantic segmentation results.....	141
6.4.2 Encoder and decoder performance in semantic segmentation.....	144
6.4.3 Loss functions for semantic segmentation	147
6.4.4 Performance of training strategy in segmentation	149
6.4.5 Performance comparison of SAM annotation and manual labelling	151
6.5 Chapter Summary.....	153
Chapter 7: Conclusions and Future Works	155
7.1 Conclusions	155
7.2 Discussions on practical applications.....	158
7.3 Limitations.....	159
7.4 Future Works.....	160
References	162
Author’s Biography.....	177

List of Figures

Figure 1.1:	Illustration of various surgical robotic systems.....	2
Figure 1.2:	Top vascular interventional robots available in international markets.	3
Figure 1.3:	Motivations towards robot-assisted interventional medicine.....	5
Figure 1.4:	The significantly difference of manipulation between robot-assisted surgery and conventional MIS.....	7
Figure 1.5:	The framework of dissertation ' s research logic.	8
Figure 2.1:	Overview of the progression trends in endovascular interventional surgery technologies.....	17
Figure 2.2:	Overview of the endovascular interventional procedure.....	19
Figure 2.3:	Development of vascular interventional robots.	21
Figure 2.4:	Overview of the robot-assisted endovascular interventional processing procedure.	27
Figure 2.5:	The manipulation analysis in endovascular interventional procedure.	33
Figure 2.6:	Illustration of DSA images in endovascular interventional procedure.....	39
Figure 2.7:	3.0 generation of VIR system.....	43
Figure 2.8:	3.5-generation of VIR system.	44
Figure 2.9:	Testing platform and processing of interventional robotic system.	45
Figure 3.1:	Experimental setup showing the original data collected procedure.	52
Figure 3.2:	Manipulation pattern of endovascular interventional surgery.	53
Figure 3.3:	Three chosen catheterization routes and path navigation analyses.....	54
Figure 3.4:	Multi-modal sensing information.....	56
Figure 3.5:	Recognition framework manipulation during robot-assisted interventional surgery.....	60
Figure 3.6:	Statistical differences of features between successful trials and unsuccessful trials.....	69
Figure 3.7:	Distribution of operating time for two groups.....	69
Figure 3.8:	Performance of S/US based on single-modal and fusion-modal features with MLP classifier.....	70
Figure 3.9:	Performance of S/US based on different feature with MLP classifier.....	71
Figure 3.10:	Distribution of statistical differences of features.	73
Figure 3.11:	The recognition accuracy of manipulation patterns.	73
Figure 3.12:	Performance of manipulation pattern based on fusion-modal features and statistical differences.	74

Figure 3.13: Performance of manipulation patterns based on thirty-one features.	74
Figure 3.14: Performance of manipulation pattern based on different fusion-modal features.	75
Figure 3.15: Performance of manipulation patterns based on different classifiers.	77
Figure 4.1: Experimental setup.	81
Figure 4.2: The nine manipulation patterns during robot-assisted interventional surgery. ...	83
Figure 4.3: Designed 32 channel flexible haptic sensor system.	84
Figure 4.4: Overview of technical skill classification.	87
Figure 4.5: Overview of the evaluation framework of surgeon-robot synergy.	88
Figure 4.6: Performance of the proposed recognition framework based on different learning rate.	90
Figure 4.7: Performance of surgeon-robot synergy in operator with three different skill level.	93
Figure 4.8: Speed manipulation of robot master and slave mechanism based on different movement patterns.	94
Figure 4.9: Time manipulation of operators with different technical skill among three groups.	96
Figure 4.10: Comparison of the average distal force among three groups, including A_A , A_B and B.	97
Figure 4.11: Proximal force among three skill group.	98
Figure 4.12: Performance of haptic force in operator with three different skill level.	100
Figure 4.13: Performance of distal forces in 4 complex vessel pathways for 3 different skill groups.	100
Figure 4.14: Performance of contact forces in 4 complex vessel pathways for 3 different skill groups.	101
Figure 5.1: Overall framework for guidewire endpoint detection.	106
Figure 5.2: Image acquisition framework during robot-assisted endovascular interventional procedure.	108
Figure 5.3: Improved U-Net model structure.	109
Figure 5.4: Guidewire endpoint detection based on pixel-adjacent-relation method.	111
Figure 5.5: Overall processing of skeletonization extraction.	112
Figure 5.6: Detection and removal of bifurcation point.	112
Figure 5.7: Overview process of the repairment of fracture points.	113
Figure 5.8: Adjacent point relationships within the eight-neighbourhood scope.	113
Figure 5.9: Overview process of determining the maximum bending region and angle values.	115

Figure 5.10: Performance analysis of proposed improved U-Net model.....	117
Figure 5.11: Guidewire segmentation results using the test set from in vivo rabbit model...118	
Figure 5.12: Mean pixel/distance error of four typical heatmap detection methods.	120
Figure 5.13: Guidewire endpoint detection results of four heatmap methods.	121
Figure 5.14: Mean pixel/distance error of six typical heatmap regression methods.	122
Figure 5.15: Comparison of guidewire endpoint detection of six heatmap regression methods.	123
Figure 5.16: Overview of the performances of guidewire endpoint detection methods.....	124
Figure 5.17: Evaluation of guidewire bending angle detection using the eight-neighbourhood method.	125
Figure 6.1: Image acquisition during robot-assisted endovascular and bronchoscope interventional procedure.....	130
Figure 6.2: Different augmentation methods.	132
Figure 6.3: Multi-branch coupling with MBTPDS-Net model.....	133
Figure 6.4: Design structure of the encoder model.....	134
Figure 6.5: Design structure of the TPD model.	136
Figure 6.6: Transposed convolution and bilinear up-sampling.....	137
Figure 6.7: Principle of depth-wise with and without dilation.....	138
Figure 6.8: Performances of different segmentation methods.	142
Figure 6.9: Performance comparison of the segmentation methods.....	144
Figure 6.10: Performance of different encoders and decoders with proposed method.....	145
Figure 6.11: Results of proposed segmentation methods with different encoders and decoders.	146
Figure 6.12: Performance of proposed MBTPDS-Net based on different loss function.....	148
Figure 6.13: Performance of MBTPDS-Net network based on different training strategy. ..	149
Figure 6.14: Performance comparison of MBTPDS-Net network using manual and SAM annotation.	152
Figure 6.15: Performance comparison of MBTPDS-Net network using manual and SAM annotations.....	153

List of Tables

Table 2.1: Technology transfer from chinese companies in vascular interventional robots....	22
Table 2.2: Summary of main clinical trial studies of VIRs.	29
Table 2.3: Summary of related VIR systems.	30
Table 2.4: The progression of previous research.	32
Table 2.5: Error measurements of axial feed.	45
Table 2.6: Error measurement of twisting angle.	45
Table 2.7: Response time of axial feed.	47
Table 3.1: Performance for S vs.US classification.	70
Table 4.1: Summary of the acquired data.	83
Table 4.2: Delivery speed among three different groups.	94
Table 5.1: Performance evaluation of different models for guidewire segmentation.	118
Table 5.2: Comparison of four typical heatmap detection methods.	119
Table 5.3: Comparison of six typical endpoint detection methods.	122
Table 6.1: Image augmentation methods and their respective parameters.	133
Table 6.2: Performance comparison of segmentation methods.	143
Table 6.3: Performance comparison of MBTPDS-Net with different encoders and decoders.	145
Table 6.4: Performance comparison with different loss functions.	147
Table 6.5: Performance comparison of different models across different datasets.	149
Table 6.6: Performance of the segmentation model across different convolutions.	151

List of Abbreviations

MIS	Minimally invasive surgery
RAS	Robot-assisted surgery
CT	Computed tomography
FDA	Food and drug administration
VIRs	Vascular interventional robots
RCS	Robotic catheter system
RNS	Remote navigation system
2D or 3D	Two or three-dimensional
DSA	Digital subtraction angiography
MRI	Magnetic resonance imaging
AI	Artificial intelligence
CVD	Cardiovascular disease
DoF	Degrees of freedom
CNN	Convolution neural network
SOTA	State-of-the-art
MBTPDS-Net	Multi-branch feature fusion coupled with a triple-pyramid deep neural network
NMPA	National medical products administration
PCI	Percutaneous coronary intervention
ADF	Active device fixation
R-PCI	Robot-assisted percutaneous coronary intervention
R-PVI	Robot-assisted peripheral vascular intervention
R-NVI	Robot-assisted neurovascular intervention
R-EPI	Robot-assisted electrophysiological intervention
EPIR	Electrophysiological interventional robot
M-PCI	Manual percutaneous coronary intervention
TEVAR	Thoracic endovascular aortic repair
TRCAS	Trans-radial carotid artery stenting
EM	Electromagnetic
sEMG	Surface electromyography
DL	Deep learning
FBG	Fiber bragg grating
MIoU	Mean intersection over union

BRA	Background reverse attention
MA	Muscle activity
FM	Finger motion
HM	Hand motion
GT	Guidewire trajectory
PL	Pull
PH	Push
CR	Clockwise rotation
CCR	Counterclockwise rotation
PHCR	Pushing combined clockwise rotation
PHCCR	Pushing combined counterclockwise rotation
PLCR	Pulling combined clockwise rotation
PLCCR	Pulling combined counterclockwise rotation
SS	Static stage
APB	Abductor pollicis brevis
FCR	Flexor carpi radialis
DI	Dorsal interossei
ECR	Extensor carpi radialis
MVC	Maximum voluntary contraction
RMS	Root mean square
ARV	Average rectified value
ZCR	Zero-crossing rate
KNN	k -nearest neighbours
SVM	Support vector machine
RBF	Radial basis function
RF	Random forest
SGD	Stochastic gradient descent
HAR	Human activity recognition
TP	True positive
TN	True negative
FN	False negative
FP	False positive
TNR	True negative rate
NPV	Negative predictive value
FDR	False discovery rate
MCC	Matthews correlation coefficient

Pre	Precision
Rec	Recall
Acc	Accuracy
ANOVA	Analysis of variance
HRNet	High-resolution network
BG	Background
GW	Guidewire
ReLU	Rectified linear units
IoU	Intersection over union
FFPS	Forward feedback processing speed
MPE	Mean pixel error
VGG13	Visual-geometry-group-13
FCN	Fully convolutional network
ASPP	Atrous spatial pyramid pooling
SI	Surgical instrument
SAM	Segmentation anything model
Sen	Sensitivity
Spe	Specificity
AROC	Area of receiver operating characteristics curve
PR-area	Precision-recall curve
Fps	Frames per second

Chapter 1: Introduction

1.1 Research Background

The development of surgical techniques can be categorized into three main stages: open surgery, minimally invasive surgery (MIS), and robot-assisted surgery (RAS). Before the 1840s, surgical progress was slow due to unresolved issues such as incision pain, wound infections, bleeding, and blood transfusions, resulting in high patient mortality rates [1]. However, as these problems were gradually addressed, surgery experienced rapid advancements. In 1881, the world's first gastrointestinal anastomosis was performed [2]; in 1882, the first cholecystectomy was conducted [3]; in 1886, the first appendectomy was carried out [4]; in 1933, the first kidney transplant was executed [5]; in 1963, the first liver transplant was achieved [6]; and in 1967, the first heart transplant was performed [7]. Despite these milestones, open surgery with large incisions and prolonged recovery periods, remains the predominant form of surgery worldwide [8, 9].

With the advancement of endoscopic and laparoscopic technologies, MIS has rapidly developed because of its advantages of smaller incisions, shorter recovery times, and improved treatment outcomes. In 1987, French surgeon Philippe Mouret performed the world's first laparoscopic cholecystectomy [10], heralding a new era in surgical practice. The 21st century marked the beginning of the "precision surgery" era, characterized by significant advancements in minimally invasive techniques. Technologies such as multi-slice computed tomography (CT), intraoperative ultrasound, and digital three-dimensional reconstruction have enabled surgeons to target lesions with greater accuracy, propelling the rapid development of surgical procedures. Laparoscopic surgery has now been extended to include 90% of general surgeries [11], 100% of urological surgeries [12], and 100% of gynaecological surgeries [13], achieving both minimal invasiveness and precision.

The quest for even less invasive procedures, which reduce patient pain and accelerate postoperative recovery, has driven the development of surgical robots. The precise positioning and movements of robots make them particularly valuable in surgeries requiring meticulous operations. Consequently, surgical robots have been applied in various fields, including neurosurgery, urology, orthopaedics, general surgery, and gynaecology. Based on the target organs, surgical robots can be divided into hard tissue robots and soft tissue robots. Hard tissue robots are primarily used in neurosurgery and orthopaedic surgeries, while soft tissue robots can be further divided into laparoscopic surgery robots, natural orifice surgery robots, vascular intervention robots, and percutaneous puncture robots, as shown in **Figure 1.1**.

The initial development of surgical robots primarily focused on transitioning from industrial robots to surgical applications. In 1985, Yik San Kwok and his team performed the first neurosurgical procedure using the PUMA 200 industrial robot, marking the advent of surgical robots [14]. In 1988, the robotic system completed the first robot-assisted prostate surgery [15]. In 1992,

the ROBODOC system, developed by Integrated Surgical Systems, successfully performed a total hip replacement and received the first “food and drug administration” (FDA) approval [16]. In neurosurgery, surgical robots represented a significant shift from traditional craniotomy to minimally invasive procedures.

Precise positioning is crucial in surgery. Traditional neurosurgery uses stereotactic tools that require patients to wear a head frame while awake, causing significant discomfort, long operation times, and limited surgical views. Surgical robots, with their precise arm positioning, are now used in the treatment of epilepsy, brain tumours, Parkinson disease, and intracerebral haemorrhage. In orthopaedic surgery, traditional procedures for joint replacements, spinal surgeries, and trauma surgeries are plagued by low precision, high osteotomy errors, and poor implant positioning. Surgical robots improve accuracy and stability [17], reduce nerve and endovascular damage and avoid severe complications such as paralysis.

The implementation of microsurgery and laparoscopic surgery marked a breakthrough in surgical robot development, addressing clinical demands for more flexible, precise movements and reducing surgeon fatigue through teleoperation [18]. Unlike early neurosurgery, microsurgery involves direct manipulation of nerves without electrode implants. The progression from open surgery to laparoscopic and natural orifice surgery reflects the trend towards less invasive procedures. In 1993,



Figure 1.1: Illustration of various surgical robotic systems, including ROSA ONE® Brain Robot for Neurosurgical procedure from American Zimmer Biomet, Inc.; Mazor X Stealth™ Edition robot for Spinal surgery from American Medtronic, Inc.; da Vinci robot for Laparoscopic surgery from American Intuitive Surgical, Inc.; The Ion robot for minimally invasive lung biopsy from American Intuitive Surgical, Inc.; CorPath GRX robot for endovascular interventional surgery from American Siemens Healthineers, Inc.; XACT ACE® Robotic System for Percutaneous puncture surgery from Israel XACT Robotics Inc.

the AESOP robot developed by Computer Motion performed the first laparoscopic surgery [19]. The da Vinci system by Intuitive Surgical was first tested on humans in 1997 and obtained FDA approval in 2000 [15], improving precision and flexibility while reducing patient trauma. The da Vinci system addresses clinical challenges in laparoscopic surgery, such as instrument movement limitations, the “fulcrum effect” at incision points, poor surgical views, and surgeon fatigue. It improves flexibility with articulated joints, allows for natural operations using master-slave controls, and provides high-resolution binocular visuals for effective hand-eye coordination. The da Vinci system, which dominates 80% of the market today [20], leads the field for Intuitive Surgical, although companies like CMR Surgical, Johnson & Johnson, and Medtronic also provide robotic surgical platforms. The commercial success of the da Vinci system has driven the diversified development of surgical robots, expanding into joint surgeries, spinal surgeries, single-port and multi-port laparoscopic surgeries, endovascular interventional surgeries, and neurosurgery.

Vascular interventional robots (VIRs) [21] is regarded as typical representation, also known as robotic catheter systems (RCS), emerged as alternatives to open heart surgery in the late 1990s. These robots navigate catheters and guidewires to the target lesion as planned, achieving precise treatment while avoiding radiation exposure and reducing surgeon fatigue. Early vascular intervention robots utilised magnetic navigation systems, with Stereotaxis designing the first-generation system, Telstar [22], in 2004, which required special guidewires and catheters and faced limitations in balloon and stent operations. In 2006, the Israeli remote navigation system (RNS) [23] used electromechanical systems to improve surgical accuracy. Based on RNS, Corindus developed the CorPath200 in 2012, later improving it to the more accurate CorPath GRX system [24]. Hansen Medical’s Sensei X1 [25] used flexible, actively steerable catheters, Sensei X2 added intervention system R-One developed by Robocath [26], a French company, assists cardiovascular force feedback at the tip of the catheter and received FDA approval in 2014. The vascular interventionists

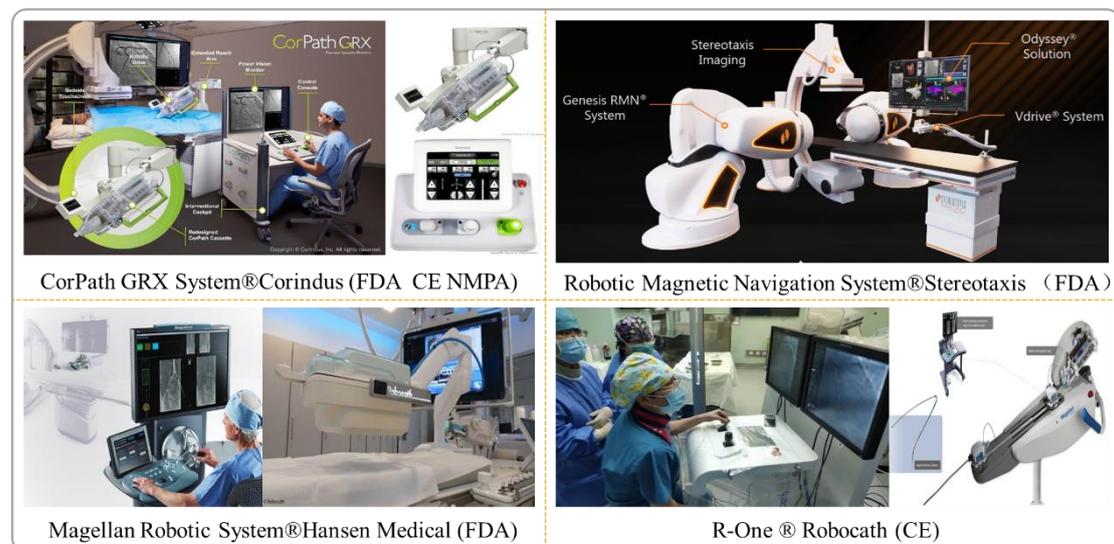


Figure 1.2: Top vascular interventional robots available in international markets, including CorPath GRX System® Corindus, Inc., USA; Robotic Magnetic Navigation System® Stereotaxis, Inc., USA; Magellan Robotic System® Hansen Medical, Inc., USA; R-One ® Robocath, Inc., France.

interventionists in the performance of stent implantation procedures. It obtained CE certification in 2019 and has been commercially available in Europe and Africa. Despite these potentials, global acceptance of robotic catheterization is limited by factors such as hardly handling complex vessel bifurcations, chronic total occlusions, severe calcified lesions, high doses of X-ray radiation, limited 3D imaging, lack of catheter-vessel contact force sensing and compatibility with online devices [27]. Consequently, vascular interventional robots are currently employed for cardiovascular interventions in only a limited number of hospitals around the world [28].

A significant advantage of vascular interventional robotic technologies lies in the design of master-slave teleoperation structures [29]. Through teleoperation, operators can manipulate the master robotic system to control slave robotic effectors, delivering catheters and guidewires to the site of the injury for interventional procedures at the distal end of the mechanism. This setup allows surgeons to sit outside the operating room in a shielded area, issuing control commands via the master robotic device to manipulate the movements of the slave robot and operate the surgical instruments (catheter or guidewire) attached to it. One of the key benefits of this master-slave teleoperation system is its ability to minimize orthopaedical injuries and radiation exposure, which are major challenges encountered by surgeons during interventional procedures. In addition to their master-slave remote control setup, essential technologies in vascular interventional robotic systems include image navigation systems for accurate tool placement using endovascular imaging, active drive catheters that can adjust to various vessels and procedural requirements, highly flexible and precise robotic arms for swift and precise catheter movement, and force feedback mechanisms that relay endovascular contact forces to the operator's interface, minimizing the possibility of endovascular rupture [30]. Together, these technological advancements guarantee the precision and safety required for successful endovascular interventions.

Endovascular imaging is essential for positioning catheters, guidewires, stents, and other surgical instruments used in endovascular interventions. Currently, digital subtraction angiography (DSA) is the main imaging technique in endovascular interventions. However, DSA is limited to providing two-dimensional images without depth information, while due to similarities between equipment and endovascular structures, interference signals can make it difficult for surgeons to discern accurately. In addition, interventional procedures are highly dependent on the experience and muscle memory of the operator, posing challenges for achieving precision and standardization in operations [31]. Medical professionals require substantial training to acquire the necessary skills for performing endovascular intervention surgeries due to the steep learning curve associated with these procedures. Moreover, there is considerable variation in the experience and skill levels of surgeons. Furthermore, a significant limitation of master-slave interventional robotic systems is the lack of force feedback or haptic perception. The lack of feedback on instrument-vessel contacts force increases the risk of complications such as thrombosis and endovascular perforation. These risks arise from potential issues with the lack of sensory force feedback in intuitive manipulation of robot-assisted surgical tasks performance [32].

1.2 Research Motivation and Significance

In the early 20th century, the advancement of radiology and imaging technology led to the rapid growth and widespread use of interventional diagnosis and treatment. In the 1950s, cardiac catheterization was extensively used for the diagnosis and treatment of heart disease [33]. By the 1970s, with the advent of angiography, interventional medicine began to address endovascular diseases [34]. Subsequently, with the continuous emergence of various new technologies and equipment, interventional diagnosis and treatment achieved significant advances in cardiovascular diseases, oncology, neurological disorders, and other fields [35]. Interventional therapy, guided by imaging devices such as X-rays, CT, magnetic resonance imaging (MRI), and ultrasound, involves the use of interventional materials such as catheters, guidewires, and puncture needles to treat diseases [36, 37]. It also obtains histological, bacteriological, biochemical, and physiological data and uses imaging data for disease diagnosis and treatment, including both endovascular and non-endovascular interventions.

Endovascular interventional treatment, which involves making a minimum incision of about 3mm on the skin surface without exposing the lesion surgically, enables precise targeting of suspected lesions directly under the guidance of devices such as DSA [38]. Due to minimal trauma, safety, efficiency, wide applicability, and fewer complications, it has become the preferred treatment method for cardiovascular diseases. However, conventional endovascular interventional procedures still have drawbacks: surgeons and patients must wear heavy lead aprons for extended periods while exposed to X-rays, which poses significant health risks over time. The procedures are also challenging, requiring high precision and extended durations, with high risks and a strong dependence on the individual experience and skill level of the surgeons. In addition, high-quality medical resources are currently concentrated in a few developed countries and major metropolitan areas. Populations in underdeveloped or remote regions and those with low incomes often lack access to these advanced medical technologies [39]. This disparity underscores the critical need for innovative solutions to bridge the gap and ensure that cutting-edge medical care becomes available to all individuals, regardless of their geographical location or economic status. Therefore, given the uneven distribution of medical resources worldwide, achieving a balanced distribution of medical

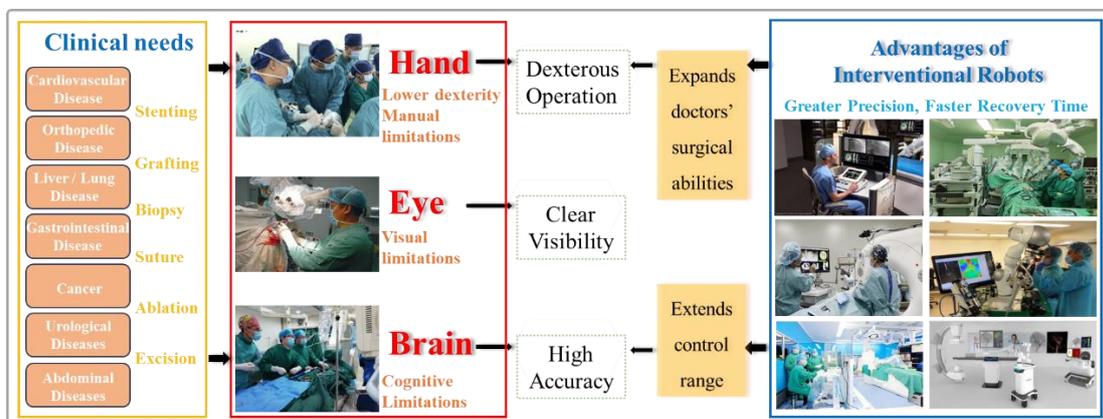


Figure 1.3: Motivations towards robot-assisted interventional medicine.

resources worldwide require urgent involvement of advanced technologies such as robotics, remote assistance, and artificial intelligence (AI).

Medical robots are increasingly being adopted for minimally invasive surgical interventions due to the varying benefits provided. For instance, patients can experience reduced perioperative pain, haemorrhage and trauma associated with traditional surgery while they are also able to recover faster after interventions. Currently, the da Vinci surgical robots, which is the most widely used minimally invasive surgical robot, have been adopted for clinical use across all continents with more than 8,887 units already installed worldwide. The da Vinci surgical robot can be used to assist surgeons in different procedures such as urological, cardiothoracic, gynaecological, and abdominal surgery [40]. Moreover, other minimally invasive surgical robots, such as the Sensei robotic catheter system [41] and the XACT ACE™ Robotic System [42], are also increasingly becoming crucial for minimally invasive surgeries in the cardiovascular and orthopaedic fields.

Cardiovascular diseases (CVD) lead to serious threats to the lives and health of patients while their family and the healthcare insurance shoulder huge financial responsibility thus, presenting significant challenges to social and economic development. With this, there is an urgent need to accelerate research, development and application of medical robotic technologies for timely CVD diagnosis and treatment. Moreover, it is crucial to promote the nationwide dissemination of advanced medical technologies and improve the uneven distribution of medical resources to alleviate the substantial public health pressure caused by cardiovascular diseases [21]. Vascular interventional robots, a category of minimally invasive surgical robots, typically operate under a master-slave mode, advancing catheters or guidewires to the lesion site in specific branch vessels, such as the Magellan system [43] of Hansen Medical (USA), the Amigo system [44] of Catheter Robotics (USA), the CorPath system [45] of Corindus Vascular Robotics (USA) and the R-one system of Robocath (France) [46]. Their performance depends on the precise control of guidewires during surgery, visual navigation, and the integration of multi-modal pathological information. Their main indications include coronary balloon dilation or stent implantation, percutaneous coronary angioplasty, and cardiac radiofrequency ablation for atrial fibrillation or tachycardia. Through master-slave design, interventional robots significantly reduce the harm of X-ray exposure to surgeons during surgery, decrease doctor fatigue, and mitigate the impact of subjective factors such as manual operation instability on surgical quality.

Global surgical statistics indicate that approximately 6 million robot-assisted procedures are performed annually [47]. Compared with the widespread adoption of surgical robots in laparoscopic and orthopaedic interventions, the clinical acceptance and utilization rate of robot-assisted vascular interventions can be said to be significantly lower, accounting for only 4.6% of total procedures [48]. A mainly reason for this disparity lies in the unresolved challenges of manipulation under under-actuated conditions. In conventional surgery, surgeons grasp, hold, and manipulate surgical instruments based on dexterous hand movements, tactile sensation, visual perception, muscle memory, clinical experience and brains' cognitive control. However, robot-assisted surgery

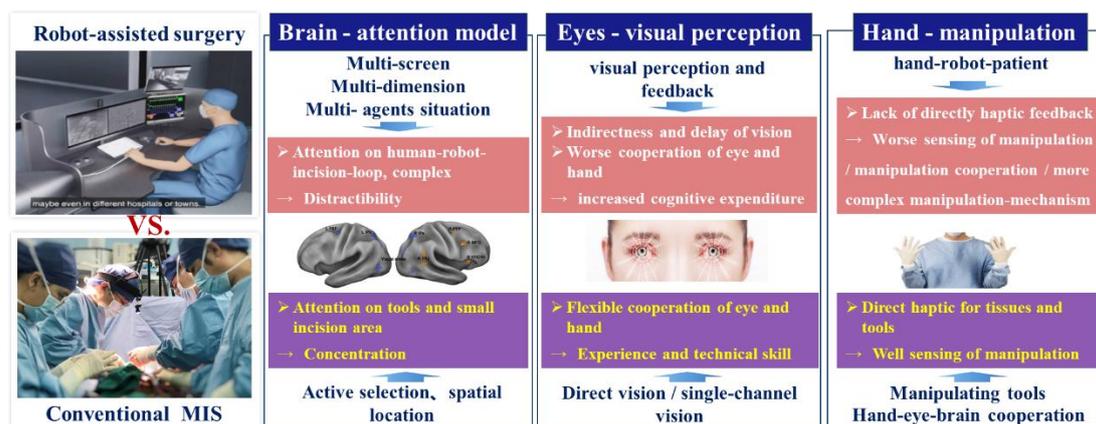


Figure 1.4: The significantly difference of manipulation between robot-assisted surgery and conventional MIS.

introduces significantly different manipulation, as show in **Figure 1.4**. These include altered visual input conventions, causes direct haptic feedback loss, and hinders multi-agent collaborative mechanism involving surgeons operating the patients with a robotic platform.

In robotic vascular interventions, surgeons must sense information across multiple screens and dimensions, integrate visual perception with indirect feedback, and simultaneously manage a more complex hand-robot-patient system. This complexity often leads to increased cognitive load, distracted attention, and reduced manipulation efficiency, ultimately requiring a much higher level of sensory adaptation and mental focus.

However, current robot-assisted systems still fail to provide an adequate sensing of manipulation. The lack of intuitive feedback and perceptual transparency limits the surgeon's ability to execute precise movements. Therefore, to improve the lives and health of indigenous peoples, our research team at the Centre for Medical Robotics and Minimally Invasive Surgical Devices, (SIAT-CAS, China) has developed a number of robotic prototypes for endovascular interventions. The master-slave structure robot with two degrees-of-freedom (DoF) is designed that can be used for the navigation of under-actuated flexible endovascular tools, such as catheter and guidewire, during endovascular interventions. The robotic systems involve spatial navigation of surgical instruments from an entry point to desired lesion sites in patient's vasculature. Utilizing dexterous, steerable endovascular tools, or highly flexible robots with high degrees of freedom can enhance interventional procedures through complex natural orifice. However, due to the lack of human (patient)-machine (robotic arm) coordination capabilities and the absence of an effective force feedback mechanism, the spatial positioning accuracy and safety of guide wires in the interventional robotic systems currently used are still inadequate. Moreover, the time required for intervascular navigation can be very long due to related constraints from communication delay, lack of experienced interventionists, intelligent level and automatic degree for operations of interventional robotic systems. Thus, this thesis is focused on learning-based modelling for improved spatially accurate operating and teleoperation control during robot-assisted surgery path navigations in human complex endovascular pathway. These are part of development projects for cardiac surgery through a single minimally incised port.

1.3 Key Research Challenges, Aim and Objective

The main goal of this research is to develop functional models using multi-sensor data-driven approaches to enhance interventional tool manipulation for the precise execution of tasks in robot-assisted endovascular interventions. However, robotic systems used for vascular interventional surgery features an underactuated mechanism. This presents significant manipulation challenges when navigating and positioning the distal tips of the flexible surgical instruments used in endovascular interventions. Moreover, the tool, robot, and surgeon collectively constitute a highly abstract and unstructured manipulation environment. This presents challenges in enhancing the surgeon's perception and control during manipulation. Thus, to improve manipulation perception, it is necessary to ensure visual clarity, operational precision and efficiency, manipulation flexibility, and effective human-robot collaboration to achieve a high degree of procedural comfort, as shown in **Figure 1.5**. Typically, accurately modelling the highly abstracted and unstructured surgeons' manipulation skills to improve surgeons' sensing of manipulation during robot-assisted surgery remains unresolved. This involves: (1) visual perception modelling to ensure reliable identification of complex states and accurately assist surgical decision-making, (2) manipulation behaviour modelling to support flexible and precise tool control, and (3) surgeon-robot synergy evaluation to assess the quality of human-robot collaboration during endovascular interventions. This research aims to address each of the above-mentioned specific challenges, as described below:

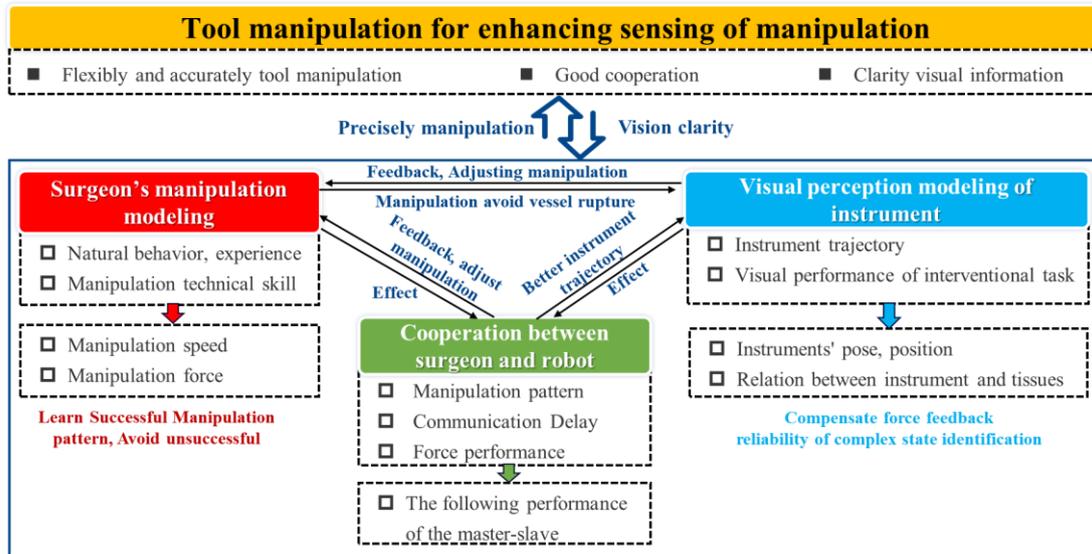


Figure 1.5: The framework of dissertation's research logic.

- Enhancing the effectiveness and accuracy of interventional instrument manipulation:** A major challenge lies in developing efficient models for intuitive manipulation through multi-sensor data analysis to improve the technical success of surgeons during robot-assisted endovascular procedures. This involves distinguishing between successful and unsuccessful manipulation patterns, understanding the internal relationships between intuitive manipulation and robotic task performance, and enabling real-time adjustment of manipulation strategies.

- **Collaborative strategies modelling for optimization synergy between operators and robot:** This challenge is to model cooperative strategies that account for varying manipulation skills among operators to enhance the intuitive delivery of tools, optimizing the synergy between surgeons and the robotic system. This requires improving the collaborative performance of human-machine manipulation to ensure precise and efficient task execution.
- **Improving the perception characteristic of interventional instrument for upscaling tool manipulation:** A critical challenge is the development of models for precise segmentation of interventional instruments and accurate endpoint localization, which are essential for improving robot-assisted surgical task performance by enhancing the visual perception of interventional instrument. This involves overcoming the difficulty of visualizing and analysing the surgical scene to minimize the risk of instrument collisions with non-target tissues, while enhancing the precision and safety of endovascular surgeries.

Remote interventional robots consist of multiple entities, such as surgeons, robots (comprising a master operational mechanism and a slave actuator mechanism), and patients. To ensure the safe, precise, and efficient performance of robot-assisted interventional surgeries, coordinated management of these participants is necessary. Consequently, this thesis intends to design multi-modal, data-driven models to enhance tool manipulation of endovascular interventional robotic systems. Furthermore, it aims to enable interventional robotic systems to accurately traverse narrow endovascular pathways, thus enhancing the ability to manipulate in human-guided robot-assisted surgeries. This is achieved through several goals that address the key research challenges, including:

1.3.1 Efficiency improvement of surgeon-robot interaction

The performance results of robot-assisted interventional procedures are largely dependent on the surgical experience of the operator, making manipulation behaviour modelling crucial for achieving successful outcomes. The vascular interventional robotic system introduces a new operational paradigm that separates the surgeon's manipulation for interventional tools, changing traditional surgical habits. This system, designed to maintain the surgeon's natural operating habits, involves complex manipulation of catheters and guidewires that mainly includes three main types of operational actions: translation, twisting, and composite actions. Surgeons perform push-and-pull manoeuvres on the guidewire along its axis. Axial translation is used to advance the guidewire through the endovascular lumen to reach the target lesion, or to pull it back if it enters an incorrect vessel. When the guidewire encounters a branch in the vessels, the surgeon uses their thumb and index finger to twist the guidewire, causing it to rotate around its circumference to navigate the bifurcation and enter the correct vessel branch. In certain narrow or calcified lesions, due to the high flexibility of the guidewire tip, surgeons often simultaneously push and twist the guidewire to adjust its movement, resulting in both axial and circumferential motion. These interventional manipulations are the result of the combined movements of the surgeon's hands, fingers, wrists, elbows, and shoulders. Multi-modal sensory data from these areas can define manipulation patterns

of interventional procedures, and data-driven machine learning models can identify various operational patterns resulting in successful or failed robot-assisted outcomes, thereby enabling interventional tools to accurately reach the desired target along the anticipated path. A multi-sensor data-driven manipulation model was proposed to uncover the inherent connection between intuitive manipulation and robot-assisted task performance, concentrating on the manipulation patterns that lead to successful outcomes in robot-assisted interventional tasks.

1.3.2 Optimization of cooperative manipulation between operator and robot

The delay encountered between the master and slave mechanisms in robot-assisted interventional surgery extends beyond mere communication factors. Variations in collaboration between the robotic system and operators of different technical skills also add to the delays. Inefficient transmissions from proximal to distal points can occur due to inaccurate motion control, communication lags, and excessive force application. While inaccurate motion control and communication lags are linked to the inherent capabilities of the robotic system, the application of excessive force is primarily influenced by the operator's level of skill. The absence of a feedback force affects operators in their real-time perception of distal and proximal forces, leading to abrupt changes in haptic force, manipulation speed, or behaviour patterns based on technical skill differences. As a result, the robotic system may have difficulty adjusting and reacting quickly to the surgeon's instructions, leading to less effective cooperation between the operator and the robot. This thesis suggests efficient evaluation methods to evaluate synergistic performance between the surgeon and the robot. These methods involve modelling the cooperative dynamics between operators and the master-slave robotic system and examining the effects of delay factors such as no delay, constant delay, and variable delay on proximal-to-distal transmissions during robot-assisted interventional procedures. In addition, it examines the interaction details, including the distal force of the tool-tissue interaction, the proximal force of the tool and mechanism, and the control force exerted on the hand-controlled ring.

1.3.3 Enhancement of visual perception of interventional instrument in endovascular pathway

In robot-assisted endovascular interventions, imaging guidance is crucial for safe manoeuvring of robots through tight endovascular routes. This involves the swift injection of a contrast agent through a catheter into the heart chambers or blood vessels to illuminate the endovascular lumen under X-ray imaging, helping to identify the lesion site. Surgeons must recognize instruments such as catheters, guidewires, balloons, and stents in X-ray fluoroscopy images to accurately direct these instruments to the lesion for efficient restoration of the endovascular pathway. In addition, the loss of force feedback coupled with potentially strong delivery forces may lead to excessive tissue damage. These reasons further increase the risk of robot-assisted endovascular interventional surgery. Studies have shown that visual feedback can partly compensate for the lack of haptic feedback in robot-assisted surgeries [49, 50], emphasizing the need to optimize visual systems to

simulate tactile sensations and improve manipulation accuracy. Thus, accurate visual feedback is critical, enabling surgeons to adjust surgical procedures effectively and ensure precise instrument manipulation, which is vital for procedural success. However, it can be challenging to visualize and track these endovascular tools through images because their appearance can resemble endovascular structures or other tissues. The guidewire often appears almost invisible, and the simple design of the instruments can easily blur with similar outlines of other objects such as bones or lungs. Therefore, the accurate tracking of endovascular tools on fluoroscopy images is essential for a fast and precise navigation through narrow endovascular pathways. Correctly identifying these tools relies on the model's capacity to derive semantic feature maps and handle contextual semantic information. In this study, models for segmenting and localizing surgical instruments were designed to ensure precise identification and effective positioning of these instruments in a variety of interventional scenarios.

1.4 Main Contributions

This dissertation focusses on intuitive manipulation of underactuated environments and improving robot-assisted interventions to enable precise and safe handling of interventional instruments for the endovascular stenosis targets within patients. These are key challenges in achieving precision manipulation control of interventional robots through intraluminal pathways. Significant research challenges in the advancement of intuitive manipulation are explored for vascular interventional robots (VIRs), with the primary research breakthrough being the main contributions of this thesis, which is as follows:

- To develop a model for manipulation based on multi-sensor data for statistical analysis, aiming to unveil the intrinsic connection between intuitive manipulation and the performance of surgical tasks with robotic assistance. This multi-layer recognition model can efficiently identify manipulation patterns in both successful and unsuccessful robot-assisted surgical tasks, emphasizing learning from successful manipulation behaviours. This process serves to alert operators to modify their manipulation techniques to prevent tissue rupture.
- To explore the utilization of a modelling technique to investigate the collaborative performance of the operator and the robot, with the aim of determining the synergy between the operator's actions and the robot's immediate responses. This ratio is calculated using a convolution neural network (CNN) trained with delay factors such as no delay, constant delay, and variable delay, to accurately predict the operator's manipulation pattern necessary to match the motion pattern of the slave robotic delivery device.
- To propose a design of an eight-neighbourhood deep neural network for detecting guidewire endpoints and the area of maximum bending achieves leading detection outcomes. This visual information modelling provides information on the position, direction, and posture of the guidewire to surgeons, thus compensating the absence of sensory force feedback to enhance

intuitive control of performance in robot-assisted surgical procedures.

- To present a design that integrates multiple-branch feature fusion with a triple-pyramid deep neural network aims at improving the visual information for segmenting interventional instruments. This approach ensures accurate segmentation to provide essential instrument visual data for surgical decision-making in robot-assisted endovascular procedures, such as identifying the surgery stage and pinpointing high-risk areas to improve intuitive manipulation.

1.5 Thesis Organization

This thesis comprises seven chapters that outline the background and advancements in general interventional robotic systems from research and industrial progress, as well as key contributions in interventional tool manipulation, cooperative performance of humans and robots, and visual perception of interventional instruments. Each chapter highlighted the main contributions of the work conducted, and the overview of each chapter is detailed below:

- **In Chapter 1** provides an overview of open surgery, minimally invasive surgery, and interventional robots, and includes a description of the progress in the development of surgical robotic systems used in cardiovascular intervention procedures for the safe and precise handling of interventional surgery. In addition, this chapter emphasizes the research motivation for the master-slave interventional robot in endovascular interventional surgery and the importance of associated interventional techniques. Finally, the chapter outlines the central research challenges, along with the objectives and main contributions of this thesis.
- **In Chapter 2** presents the literature review on interventional robots, emphasizing intuitive manipulation and precise visual perception to improve safety in robot-assisted surgeries. It outlines the transition from manual to robotic interventional procedures, specifically in endovascular surgery while detailing the successful research and clinical applications of these robots. Furthermore, intuitive manipulation modelling, force information acquisition and precise visual information perception of the interventional robots that have been proposed for optimization manipulation characteristic of the robotic system. In addition, it identifies technological challenges that must be addressed to improve the safety and acceptance of interventional robots. The design of the catheter-based interventional robotic systems upon which the research works carried out in this thesis, are presented.
- **In Chapter 3** explores the robot-assisted catheterization, highlighting its impact on cardiovascular procedures. Traditional manual surgery depends on the precise handling of the tools by surgeons, but robotic systems have changed these techniques significantly, despite their challenges. These systems face difficulties in emulating manual manipulation skills and

have a high learning curve, which poses a concern not yet addressed in research. This proposed multi-sensor data driven-based manipulation model aims to connect surgeon manipulation behaviour with the effectiveness of robot-assisted tasks, emphasizing on the learning of manipulation patterns with successful interventional procedures to enhance safety and efficiency of tool manipulation in robotic interventions. Moreover, statistics analysis was carried out to understand the redundancy of surgeon's natural behaviours or features and their relationship with the robot-assisted interventional tasks.

- **In Chapter 4** addresses the design of an isomorphic master-slave interventional system in vascular intervention robots, highlighting the inevitable system response delays. While these delays are unavoidable, communication delays significantly affect human-machine synergy performance. The chapter emphasizes how operator expertise leads to different manipulation strategies and behavioural adaptations, impacting the robot's response time and overall system cooperation. This proposed manipulation-based evaluation framework aims to improve cooperative performance by integrating data on operator muscle activity, finger movements, contact forces, and position and rotation data from both master and slave devices. In addition, three delay factors are assessed for their influence on performance and real-time system adaptability, ensuring that safety standards are maintained.
- **In Chapter 5** discusses the importance of visual perception of interventional instrument in guiding surgeons' manipulation during interventional procedures, specifically focusing on intuitive handling of surgical instruments. Therefore, the chapter highlights the use of segmented and located instruments to facilitate tracking and pose estimation, which improves the safety and effectiveness of surgeries. It emphasizes the crucial role of guidewires in endovascular surgery, noting the challenges in visualizing and tracking them due to interference and their flexible nature. To address these issues, a novel detection framework is introduced that uses an enhanced U-Net approach for extracting the entire guidewire from DSA images, followed by an eight-neighbourhood method for precise endpoint detection. This framework aims to improve manipulation strategies and reduce the risk of instrument collision while also helping to develop automated interventional procedures.
- **Chapter 6** focusses on further improving the segmentation accuracy of interventional instruments using a technique called MBTPDS-Net, which involves multi-branch feature fusion and a triple-pyramid deep neural network (MBTPDS-Net). This chapter highlights the importance of precise segmentation in multi-modal fusion images utilizing the eight-neighbourhood detection method discussed previously, aiming to further improve accuracy of the interventional instruments' segmentation performance. Even though this method works well with good segmentation quality, it fails if segmentation is poor. Challenges include low contrast, complex surgical environments, mirror reflections, and size and shape differences,

particularly with guidewires, which can be mistaken for bones or lungs in fluoroscopy images due to their simple appearance. These factors complicate accurate segmentation and can result in operational errors, longer surgery times, and higher patient risks.

- **Chapter 7** summarizes the technical achievements of previous chapters, identifies limitations in interventional robot designs explored in the research, and presents these as open research questions for future exploration.

1.6 Publications¹

In this thesis, the research works that were carried out to solve manipulation modelling and improve the visual perception of the interventional instruments for upscaling interventional tool manipulation in prototypes of the interventional robotic system developed at the Centre for Medical robotics and Minimally Invasive Surgical Devices. Some of these works have appeared in previous publications as thus; system designs for vascular interventional robot, outlined in **Chapter 2**, were first presented in Wenjing Du *et al.* [1] and Wenjing Du *et al.* [2], respectively; the multi-modal sensing data driven-based manipulation framework proposed for surgeons' manipulation technical pattern recognition of robot-assisted endovascular interventional procedure and analysis the redundancy of behaviours or features on robot-assisted performance of interventional tasks, outlined in **Chapter 3**, were first presented in Wenjing Du *et al.* [1, 6, 7]; the synergy performance assessment model proposed for exploring the cooperative capability between surgeon with different technical skill and master-slave to improve the efficient proximal-to-distal transmission based on machine learning method, outlined in **Chapter 4**, was first presented in Wenjing Du *et al.* [2]; the endpoint location method of endovascular tools proposed for robot-assisted interventional procedure, outlines in **Chapter 5**, was first presented in Wenjing Du *et al.* [3, 4]; while interventional instrument segmentation model proposed for optimization intuitive manipulation of master-slave isomorphic robot, outlined in the first part of **Chapter 5**, was first presented in Wenjing Du *et al.* [3, 4]; the interventional instrument segmentation models based on multi-modal images fusion, outlined in **Chapter 6**, were first presented in Wenjing Du *et al.* [3]. The specific details for each of these studies are described in the respective chapters of this thesis.

¹ Details please see "List of Publications" in page III.

Chapter 2: Literature Review

Despite increasing evidence that supports surgical robots as a reliable and effective platform for diagnosing and treating internal organs, advanced robotic technologies, such as the da Vinci surgical system, have inherent limitations when it comes to anatomical access through restricted pathways. These constraints involve multiple-port access, extended procedures, and challenges in docking during patient-care operations [51]. This chapter offers an overview of the background and developments in robot-assisted interventional surgeries, together with a comprehensive review of the literature on the design, master-slave controls, haptic feedback, and clinical application of interventional robots. Furthermore, it explores the levels of robotic intelligence used for minimally invasive interventions within the narrow natural channels of the human body. The chapter also provides a comprehensive review of the robotic systems designed for interventional surgery, current methods for learning and evaluating intuitive manipulation, force feedback systems, improved visual perception of the interventional instrument, as well as the limitations and technical challenges of existing robotic systems and tool manipulation.

2.1 Concept and the Development Trends of Interventional Medicine

Interventional medicine is described as a set of diagnostic and therapeutic procedures that identify and treat diseases using medical imaging equipment, such as needles, guidewires, and catheters [52]. The key characteristics of interventional medicine include being minimally invasive, precise, safe, and effective. From a clinical practice viewpoint, interventional medicine includes all diagnostic and therapeutic procedures performed with the help of medical imaging equipment, regardless of the clinical department responsible for them. However, interventional radiology, also known as interventional therapy, is understood from the diagnostic technique viewpoint. Using imaging devices such as digital subtraction angiography machines and ultrasound, it uses needles, catheters, and other materials to insert specific instruments into affected areas of the body through minimally invasive methods for various treatments, such as puncture techniques, catheter operation, angiography, radiofrequency ablation, and cryoablation. Essentially, both interventional medicine and interventional radiology refer to the same concept, but from different viewpoints. Interventional radiology is considered a developing diagnostic and therapeutic technology, whereas interventional medicine is seen as a new medical specialty.

The roots of interventional medicine can be traced back to cardiac catheterization in the early 20th century, initially used to evaluate heart function in individuals with heart disease [53]. Interventional medicine expanded rapidly with the evolution of radiology and imaging technology, having its inception in 1896 with the first diagnostic application of X-ray [54]. In 1929, Werner Forssmann achieved a significant milestone by performing the first cardiac catheterization [55]. He inserted a catheter into his own vein at the elbow and navigated it to the right atrium while X-ray

monitoring. Further advancements followed with Sven-Ivar Seldinger's method for catheter insertion in 1953 [56] and Charles Dotter's groundbreaking transluminal angioplasty in 1958 [57], which laid the foundation for contemporary endovascular interventions. The evolution of endovascular interventional surgery progressed, with Dotter and Melvin Judkins introducing catheters and guidewires in 1964 [58].

During the 1970s, interventional medicine expanded its scope to include endovascular diseases with the introduction of angiography technology. In 1973, Andreas Gruentzig carried out the first percutaneous coronary angioplasty, laying the foundation for interventional coronary heart disease treatment [59]. He also created the balloon catheter in his kitchen, enabling effective treatment of stenotic conditions and igniting a "balloon fever". In 1976, Alexander Margulis coined the term "interventional" [60], signalling the creation of a new specialized branch in radiology. This development was followed by the emergence of various interventional devices, driving the rapid growth of interventional medicine. Julio Palmaz invented the balloon-expandable stent in 1978 [61], which provided essential technical support for coronary stent implantation. The balloon catheter designed by Thomas J. Fogarty's in 1990 [62] further improved thrombectomy and angioplasty capabilities in endovascular interventions.

In the following decades, significant advances were made in interventional medicine that included cardiovascular disease, tumours, nervous system disorders, and beyond. For example, the first successful radiofrequency ablation for liver cancer was recorded in 1994 [63], inaugurating a new era in tumour interventional therapy. Subsequently, in 1996, the Cyber-Knife stereotactic radiotherapy system [64] emerged, offering new advances in interventional radiotherapy. It was not until 1998 that the US FDA approved the first cryoablation system for addressing cardiac arrhythmias [65]. From the 2000s onwards, interventional radiology experienced rapid advancement, notably with the introduction of drug-eluting stents in 2000 [66], which significantly decreased the rate of coronary artery restenosis. In 2004, the first successful application of irreversible electroporation for liver cancer marked a breakthrough in tumour ablation technology [67]. By 2005, cryoablation began to be implemented clinically for the treatment of kidney cancer in the US [68], and this technology gradually gained acceptance as a standard method of treatment for kidney cancer. In 2008, the US FDA approved the first cryoablation system for lung cancer treatment [69]. In 2010, high intensity focused ultrasound [70] technology was developed and put to clinical use as a non-invasive method for tumour treatment, while cryoablation was used to ease the pain of bone metastases, thereby effectively alleviating pain symptoms in patients. In 2011, the fenestrated endovascular aortic repair technique [71] had matured to treat aortic aneurysms, reducing the risks associated with aortic aneurysm surgery.

Furthermore, in 2012, advances in drug-eluting stents, including biodegradable variants, were made, improving the long-term success of coronary stent implantation. Concurrently, cryoablation began to be used more in the treatment of liver and pancreatic cancers, slowly becoming a complementary treatment approach [72]. In 2013, the Mitra-Clip system [73] was effectively used

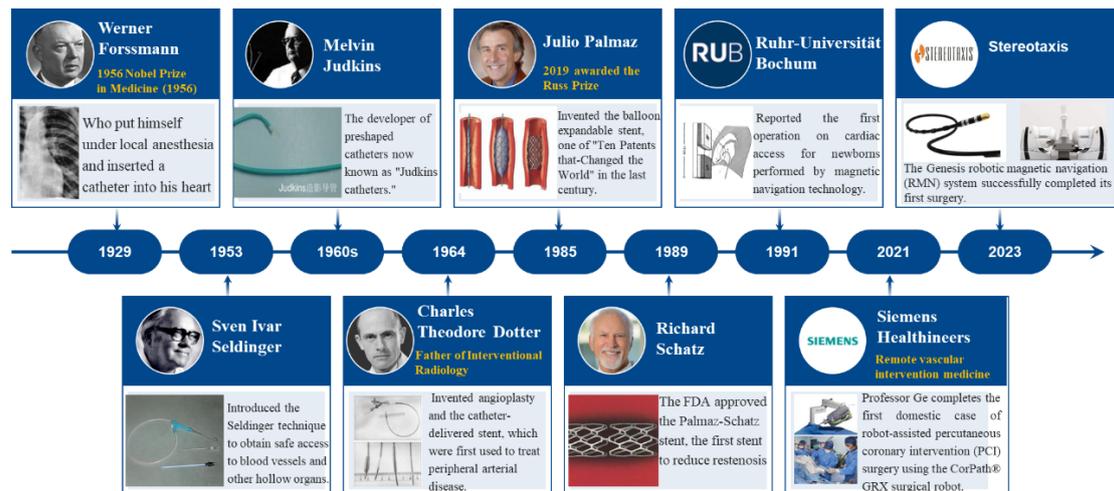


Figure 2.1: Overview of the progression trends in endovascular interventional surgery technologies.

for transcatheter mitral valve repair, decreasing the risks associated with cardiac surgery. By 2015, transcatheter heart valve replacement methods, such as transcatheter aortic valve replacement [74] and transcatheter tricuspid valve replacement [75], offered safer valve replacement alternatives for high-risk patients. In 2016, transvenous thrombectomy greatly improved outcomes in the treatment of patients with acute ischemic stroke [76]. In the same year, the US FDA approved a cryoablation system for clinical trials aimed at the treatment of breast cancer [77]. In 2017, the National medical products administration (NMPA) in China approved Hai-Jie-Ya's cutting-edge devices, the "Low-Temperature Cryoablation Surgery System" [78] and the "Disposable Sterile Cryoablation Needle" [79], through the special approval process for pioneering medical devices. By 2018, stereotactic radiotherapy technology [80] was used for accurate radiotherapy of malignant tumours, including lung and pancreatic cancers. In the same year, US researchers began investigating the joint efficacy of cryoablation and immunotherapy to improve treatment outcomes [81].

The advancement of the field progressed with breakthroughs in robotic surgery technologies, such as FDA approval of the da Vinci robot by the American Intuitive Surgical Company in 2000 [82], the introduction of AI-assisted diagnostics in 2019 [83], and the creation of new interventional materials such as drug-eluting microspheres and biodegradable stents in 2020 [84]. **Figure 2.1** highlights the major technological progress in interventional medicine.

Currently, interventional radiology is the preferred approach for treating cerebrovascular and cardiac diseases, replacing traditional surgical methods with minimally invasive endovascular procedures. Many open endovascular surgeries have been replaced by endovascular techniques. In addition, transcatheter arterial chemoembolization [85], iodine-125 seed implantation, and tumour ablation are now crucial components of cancer therapy. Furthermore, the use of interventional radiology extends to various organ systems, including the respiratory, digestive, and urinary systems, highlighting its extensive clinical influence. During the past three decades, interventional therapy has become a leading clinical discipline worldwide, known for its minimal invasiveness, rapid effectiveness, and low complication rates [86]. The formation of specialized interventional

departments in numerous hospitals emphasizes its role as the third largest clinical specialty after internal medicine and surgery.

Unlike conventional medical therapies that involve drug intake or injections, as well as open surgeries, interventional procedures use specialized catheters or minimally invasive instruments directed by sophisticated imaging technology. This approach ensures minimal trauma to the patient and provides prompt therapeutic results. Interventional procedures aim to address five major issues, including infusion, embolization, recanalization, ablation, and biopsy, using minimally invasive surgical instruments, as guided by clinical needs, with further details detailed as follows:

- **Infusion:** Administering medications directly to the site of the lesion via a catheter inserted within the natural pathways of the body.
- **Plugging up:** Utilizing catheters or other surgical tools with embolic substances to block blood vessels, stopping bleeding, or removing tumours. This technique is applied for acute arterial bleeding conditions such as haemoptysis, bleeding from a ruptured liver, spleen, kidney, postpartum haemorrhage, nosebleeds, etc.
- **Unclogging:** Using guidewires along with balloons and stents to clear narrowed or blocked blood vessels or other lumens is used for occlusive diseases such as peripheral arterial stenosis, carotid artery stenosis, renal artery stenosis, lower limb arterial stenosis, among others.
- **Ablation:** Involves the use of a catheter or surgical instrument to directly deliver radiofrequency, microwave, cryoablation, and particles into cancerous tissue to remove lesions, and is utilised in the treatment of cancers such as uterine fibroids, lung cancer, kidney cancer, pancreatic cancer, and liver cancer, among others.
- **Clamping:** Removal of tissue from deep-body lesions for pathological examination using a catheter or endoscope working channel to insert biopsy forceps or puncture needles, as seen in lung nodule biopsies.

2.2 Vascular Interventional Robots in Interventional Surgery

2.2.1 Rationale of vascular interventional robot development

Despite significant advances in endovascular interventional technology, the techniques used in performing these surgeries still face several challenges. First, X-ray imaging, which is crucial during interventional procedures, exposes surgeons to extended radiation exposure. Although lead aprons or radiation shielding panels are currently used to reduce this exposure, they do not eliminate radiation risk. Prolonged exposure to high-radiation environments increases the risk of developing conditions such as skin cancer, leukaemia, thyroid cancer, and other diseases. Furthermore, wearing heavy lead aprons for long periods of time compromises the precision of guidewire and catheter handling by surgeons, while also increasing the probability of joint injuries. Current interventional surgeries also heavily depend on surgeon expertise and muscle memory, making it difficult to ensure

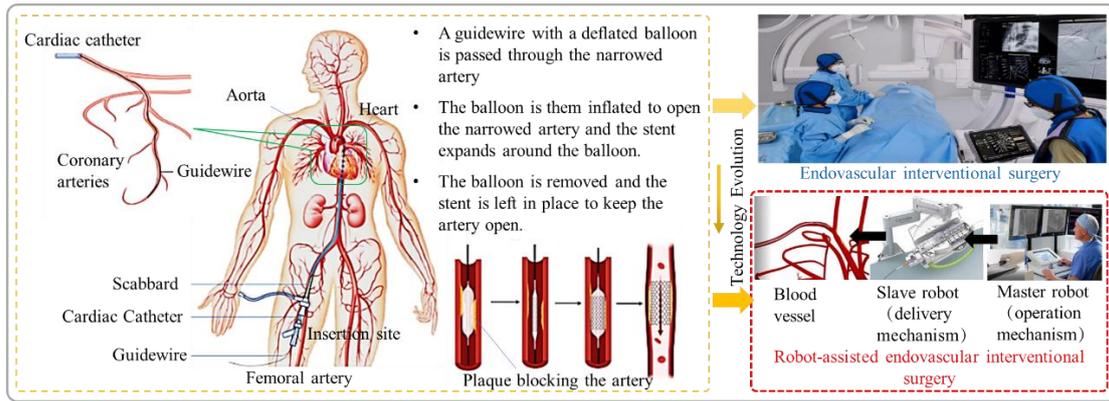


Figure 2.2: Overview of the endovascular interventional procedure.

accurate and standardized procedures. For example, hand movements while manoeuvring guide wires and catheters can lead to vibrations, misplacement, or excessive rotation. Complex scenarios, including multiple or diffuse lesions, are harder to handle and require more time to handle. Endovascular interventional surgery also involves a steep learning curve, necessitating extensive training for clinical surgeons to perform these surgeries, with considerable differences in surgical experience and proficiency among practitioners.

The introduction of VIRs is intended to address these issues. These robots transform medical tasks into mechanical processes, enhancing accuracy to millimetre level and force feedback to 0.01N. This capacity facilitates complex interventional surgeries that are difficult to perform manually, thus increasing surgical precision, safety during operations, and recovery of blood vessels after surgery. VIRs function as master-slave electromechanical tools that enable surgeons to remotely manage catheters and guide wires in coronary, neuro and peripheral endovascular procedures. By merging surgical robot technology with endovascular interventional methods, these robots assist surgeons in precisely pinpointing lesions, creating 3D models of patient vessels from preoperative and intraoperative images, and evaluating features such as endovascular intersections, curves, elasticity, and plaques. This supports the accurate tracking and placement of surgical tools, optimizes the use of balloons, stents, and catheters, and encourages the standardization of surgical procedures, as shown in **Figure 2.2**.

2.2.2 Evolution of VIR system

The introduction of VIRs has significantly minimized surgeon exposure to X-ray radiation, while improving precision and stability in surgical procedures. Over nearly two decades, various companies have developed different robotic systems. In 2006, Beyar *et al.* [23] at Haifa Hospital in Israel designed a RNS for percutaneous coronary intervention (PCI). This system uses multiple friction wheels to steer the tools (guidewire or catheter). Unlike magnetic drive catheters, RNS does not require the creation of special instruments, thus expanding its versatility.

Later, in 2007, the FDA granted approval for the Hansen Medical Sensei X1 system [87], which subsequently began its application in coronary intervention and ablation therapy. This system operates on an electromechanical principle, using an electric mechanical hand to manage a bendable,

specialized catheter and sheath. By manipulating the tail end and embedding connecting lines around the catheter or sheath, the bending direction can be modified. The Sensei X2 system [88], an upgrade of the Sensei X1, integrates visual and tactile force feedback, although it requires special head-end components. Originally from the Sensei system, the Magellan system [89] is adapted to aid in the treatment of peripheral endovascular diseases. It improves catheter flexibility and allows forward and backward motion through a series of clamping rollers. The FDA approved both the Sensei X series and the Magellan system in 2012, but they do not support standard intraluminal devices. Both systems are composed of a remote device manipulator, a surgeon workstation, and a robotic catheter [90]. The remote device manipulator includes two manipulators for controlling flexible endovascular instruments, along with a friction drive belt that ensures stable delivery and precise control of the distal tip of tools like catheters and guidewires [91]. The Magellan system, specifically designed for peripheral endovascular procedures, supports multi-specialty applications. One major benefit is its uniquely engineered flexible bending catheter, which facilitates seamless entry into complex small peripheral vessels, ensures steady placement of the guidewire, and reduces the likelihood of endovascular damage caused by contact with and friction against the vessel wall [92]. However, the system is disadvantaged by the requirement for manual placement of interventional devices (for example, balloons and stents) after surgical access is achieved, the high cost of the robotic system, and the absence of haptic feedback [93]. The necessity of specially developed catheters and sheaths also limits their usage.

Given the drawbacks of earlier design concepts in clinical settings, the creation of commercial intraluminal devices operated by robots began to more effectively meet the needs for endovascular intraluminal interventional surgery, as shown in **Figure 2.3**. Catheter Precision Co. developed the Amigo system [94], receiving FDA approval in 2012. The outer tube of the Amigo system can hold the internal catheter in place and perform movements in forward, backward, and rotational directions like a rail, enabling the catheter's multidirectional movement. The open design allows compatibility with certain commercially available catheters of varying thicknesses, but it is unable to deploy stents or carry out more complex intraluminal interventional procedures. Corindus Vascular Robotics launched a robotic system for endovascular surgeries, beginning with CorPath200 [95], which gained FDA approval in 2012. Subsequently, they advanced to the CorPath GRX system [95], obtaining FDA approval in 2016 and CE certification in 2019. This new system stands as the leading commercial solution for vascular interventional surgery robots, applicable to the coronary and peripheral arteries using an open platform. The system functions by moving the guidewire and catheter forward and backward through a clamping roller, with the roller's rotation dictating the rotation of the interventional tools (catheter or guidewire). It is compatible with 0.014" guidewires, rapid-exchange balloon catheters, and balloon-expandable stent catheters. Operation at the bedside involves a single use control box containing the roller, which can be prepped with guidewires, catheters, balloons, and related equipment after infection. The second generation CorPath® GRX robotic system [95] is equipped with a single-use disposable cassette, an extended reach arm, and a

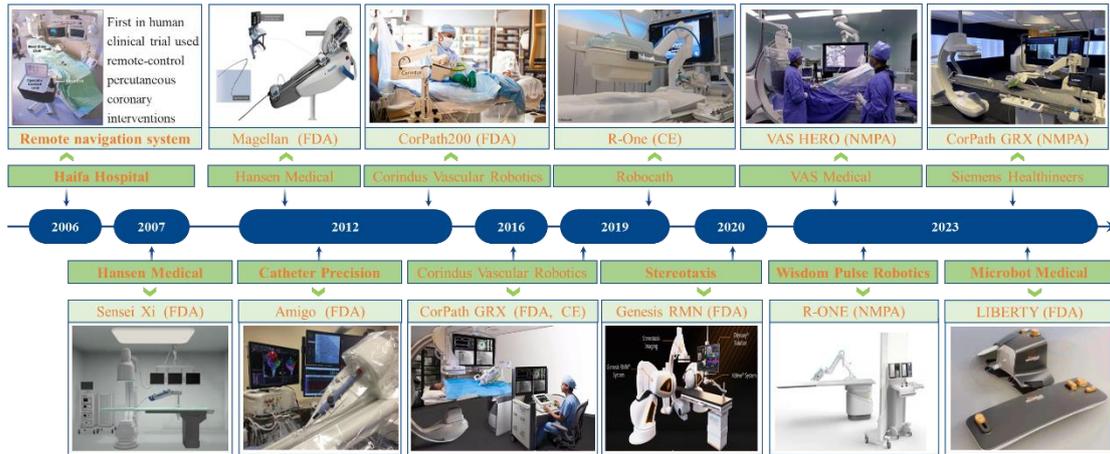


Figure 2.3: Development of vascular interventional robots.

lead-shielded robotic control workstation. This system can simultaneously guide catheter balloon, manage the guidewire, and stent catheter. In addition, the automation and accuracy of the system have been improved with features such as “dotter”, “constant speed”, “active device fixation (ADF)”, and “auto rotate-on-retract” [96-99]. In general, the compact design of the robotic system facilitates easier management of endovascular instruments through the control workstation and includes a sterilizable disposable drive cassette. Despite its advances, the system is limited by its ability to operate only one set of guidewires and catheters, which makes it less ideal for complex lesions, along with the high cost of consumables and lack of distal force feedback [100, 101].

In 2019, Siemens Medical invested \$1.1 billion in cash to acquire Corindus Vascular Robotics and develop vascular interventional surgery robots [102]. In the same year, Robocath, a French company, received CE certification for its R-One robotic system [103]. In 2020, the FDA approved the Genesis RMN system [104] from the American company Stereotaxis. In early 2020, the Israeli company Microbot Medical introduced the Liberty robotic system [105], acknowledged as the first fully disposable robotic system worldwide, targeted at surgeries involving the nervous system, cardiovascular, and peripheral endovascular regions. This system is built similarly to an old tape, featuring a specially designed catheter that is pre-rolled within the robot. The Liberty robotic system features a self-developed cost-effective catheter head known for its active rotation capability, compactness, and simplicity, which obtained FDA approval in 2023. These products are mainly available on the EU and US markets, where the development of the vascular interventional robot industries was initially developed.

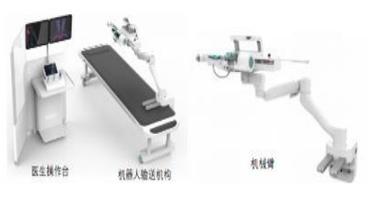
Compared to the global market, domestic development of vascular interventional robot technology began at a later stage, as detailed in **Table 2.1**. Professor Xiao Nan of Beijing Institute of Technology established wire science and pioneered the first domestic multi-site remote coronary intervention, which is currently in the animal experimentation phase. Guo Shuxiang, an international academic member of the Japanese Academy of Engineering, founded Arobo Medical and developed a pan-vascular interventional robot, which has successfully completed its first clinical trial for neurovascular intervention, receiving the fourth NMPA approval for vascular interventional

robots in China in 2024. The Shanghai Operation Robot company developed the ALLVAS™ vascular interventional robot, in partnership with Shanghai Chang-hai Hospital and Shanghai Jiao Tong University. In addition, RainMed Medical and RAYSIGHT Medical are known for their capabilities in artificial intelligence and imaging software, concentrating on comprehensive planning for functional endovascular diagnosis and treatment; they collaboratively developed the Flash Robot and an interventional vascular robot with SIAT, respectively. The CAFFR system from RainMed Medical, along with the CTFRR and coronary intelligent post-processing platforms from RAYSIGHT Medical, obtained approval from the NMPA.

WeMed's first product, the ETcath vascular interventional robot, which is a DSA machine, has passed the national innovation medical instrument examination and approval. Shanghai Huihe Healthcare technology, in collaboration with Shanghai Jiao Tong University, cofounded an

Table 2.1: Technology transfer from Chinese companies in vascular interventional robots.

Robotic systems	Company	Main features	Description
 <p>R-ONE®</p>	Wisdom Pulse Robotics (Shanghai)	The French company Robocath S.A. and Minimally Invasive Robot have formed a joint venture in China.	NMPA in 2023
 <p>VAS HERO</p>	VAS ROBOT	A surgical robotic system for minimally invasive endovascular interventions invented by the clinical research team led by Professor Li Youxiang at Beijing Tiantan Hospital and the Beijing Institute of Technology	First approved interventional NMPA
 <p>ETcath</p>	WeMed	Integration of DAS with AI for image synthesis	Has successfully completed the national innovation medical device examination and received approval
 <p>ALLVAS™</p>	Shanghai Operation Robot	Features an adaptable hand gripping technique	Achieved the world first PTA stenting
 <p>Pan-vascular Interventional Robot</p>	Abrobo	Cooperative management and supervision of several catheter guidewires, comprehensive vascular interventional robot	First clinical trial for neurovascular intervention, securing 120 million yuan in Pre-A funding, and the fourth NMPA-approved interventional

 <p>Flash Robot</p>	RainMed	Integrating the digital function diagnosis unit, merging diagnosis and treatment	Conducted the initial experiments on animals
 <p>Vascular interventional robot</p>	RAYSIGHTMED	RAYSIGHTMED & SIAT	Using 5G for remote telesurgery with force feedback, RuiXin-FFR(NMPA) is engaged in an extensive multicentre clinical trial with a robot.
 <p>Interventional robot</p>	Wire Science	First domestic multi-location remote coronary intervention with initial study conducted on animals	Secured 10 million CNY in a Series A funding round.
 <p>Robvas</p>	LANCET Robotics	Real-time master-slave regulation, accurate movement management of interventional supplies, combined force and visual feedback mechanism	Advanced to the phase of registered clinical trials
 <p>VasCure robotic system</p>	Beijing Hongtai Medical Technology co.	Cooperation with Institute of Automation CAS	Numerous successful clinical trials in human science research and remote 5G surgeries
 <p>K-Clip™</p>	Shanghai Huihe Healthcare technology co.	Global initiative for ultrasonic localisation	Received fast-tracked approval for cutting-edge medical devices
 <p>TITIAN®</p>	MEIO CARDINAV MEDICAL	Electrophysiological intervention that is compatible with the standard catheter sheath	Received fast-tracked approval for cutting-edge medical devices

Interventional robot R&D centre, resulting in the development of the ultrasound-locating interventional tricuspid valve shaping device K-Clip. Moreover, LANCET Robotics and Beijing Hongtai Medical Technology each developed the Robvas and VasCure robotic systems, which have

both entered the registered clinical trial phase and have demonstrated preliminary clinical practicability. MEIO CARDINAV MEDICAL designed the TITIAN® electrophysiology interventional robot, which is currently under review for innovative medical equipment. Hebei Yidu Robotics Technology Co., Ltd., a VAS ROBOT subsidiary, created the “VAS HERO” interventional robotic systems, certified by the NMPA in March 2023, marking it as the first approved vascular interventional robot in China. In addition, Minimally Invasive Electrophysiology has exclusive distribution rights for Stereotaxis magnetic navigation electrophysiology robots in China. To bring R-ONE to China, MicroPort Robotics collaborated with French Robocath S.A.S in 2020, forming Wisdom Pulse (Shanghai) Robotics Co., Ltd., which developed the R-ONE® robotic system, now achieving NMPA approval as the second certified vascular interventional robot in China. In 2021, Siemens medical’s CorPath GRX interventional surgery robot was approved by NMPA in 2023, making it the third certified vascular interventional robot in China. Generally, these clinical systems are remarkable, and the competitive advantage provided by local companies has played a crucial role in the rapid expansion of the vascular interventional robot sector in China.

In May 2023, Siemens chose to stop the development of the Corindus vascular interventional robot after discovering significant flaws in the CorPath product, such as its inability to support sheaths, guide catheters, or mechanisms for the deployment of self-expanding stents. The robot is limited to moving the interventional tools (guidewire or catheter) forward and backward, as well as making small rotational adjustments, but cannot perform complex intraluminal interventions. In addition, it is incapable of performing multiple activities involving the guidewire, balloon catheter, or stent at the same time and lacks a tactile feedback force feedback system.

Currently, numerous research projects are focused on force or haptic feedback, operational safety strategies, and multi-instruments collaborative delivery technology for robotic-assisted endovascular procedures. Various vascular interventional robots have been designed to meet the essential requirement in catheterization laboratories for robotic systems capable of handling multiple tools (guidewires or catheters) simultaneously during the treatment of complex coronary lesions (type B2), which is a technology that is the subject of ongoing studies [106]. The treatment of complex lesions depends on the coordinated use of multiple instruments. To address this issue, Wang *et al.* [107] developed a mechanism capable of simultaneously handling both the catheter and guidewire. The guidewire is delivered via an axial reciprocating motion, while its rotation is facilitated by three rollers that clamp and rotate the guidewire at the same speed and direction. This mechanism allows for the clamping and rotation of two guidewires, with switching between them achieved by replacing the crank rocker with a different roller set. However, despite enabling multi-instrument movement, the mechanism’s large size and weight, along with the potential for multi-segment bending in the wire clamping section, may pose risks of damaging the instrument. In contrast, Cha *et al.* [108] utilised a combination of linear reciprocating motion and rotating gear teeth for catheter movement and rotation, coupled with a friction wheel and gear arrangement to drive the guidewire’s movement and rotation [109]. They performed an in-vivo study to confirm the

suitability and efficacy of the system to manage, control, and navigate multiple instruments [110], highlighting its advantageous in straightforward disassembly and sterilization of components.

On a global scale, VIRs are considered the newcomers among major surgical robots. In comparison, orthopaedic robots have already seen the rise of leading companies, such as Stryker's Mako joint robot, approved by the FDA in 2006. Stryker's MAKO now commands a global market share of 9%, alongside Medtronic's Mazor and China's Tianzhihang. Intuitive Surgical's da Vinci leads the field of laparoscopic robots, being the first of its kind to receive FDA approval and has maintained a stronghold in the global market for many years. Meanwhile, vascular interventional robots see only a 0.5% penetration rate in the US markets. The VIRs, which require significant advances in imaging, materials science, and robotics technology, are distinct and complex to develop compared to other surgical robots, as they merge multiple disciplines, including medicine, mechanics, biomechanics, imaging, and computer science, leading to substantial technical challenges.

Unlike orthopaedic and laparoscopic robots, VIRs need to manoeuvre through more delicate human blood vessels and operate catheters, guidewires, balloons, and other interventional instruments within them. This requires high standards for product performance and stability. However, Laparoscopic robots integrate relatively few types of mechanical arms and hands. On the other hand, VIRs must handle hundreds or even thousands of consumables, which differ in thickness, length, and hardness. The robot must be able to control these pre-existing endovascular intraluminal devices to complete all surgical steps and accommodate various endovascular interventional procedures. These features present considerable technical challenges for VIRs.

2.2.3 Clinical trials of applications of vascular interventional robots

Domestic and international VIRs currently utilise a master-slave mode of operation, where the robotic propulsion mechanism accurately navigates catheters or guidewires to specific branch vessels or lesion locations. On the other hand, this master-slave setup effectively protects medical staff and patients from radiation in the surgical setting, thereby decreasing the risk of radiation-related injury, and lessens the dependence on the surgeon's expertise and muscle memory, which can often affect the accuracy and standardization of procedures. In addition, they alleviate the steep learning curve and the disparities in surgical expertise and skill levels among clinicians, which traditionally require extended training periods to achieve proficiency in such procedures. On the contrary, the use of small instruments such as catheters and guidewires allow minimally invasive procedures with lower bleeding, limited radiation exposure, and greater accessibility. However, the system's underactuated nature and numerous DoF constrain the positioning accuracy of the interventional robot.

Surgical robots, such as the da Vinci system, are mainly built for laparoscopic operations. These robots have distal robotic arms equipped with rigid tools, such as grasping forceps, cutting forceps, and ligation forceps. The tools are integrated with the robot, creating a unified system, although the range of available instruments is limited. During the development process of these robots, robotic

arms and manipulators are designed as a single unit, enabling complete control over rigid tools. In contrast, the design of vascular intervention robots involves the surgeon operating a primary robotic arm, which then remotely controls a secondary device to handle instruments like guidewires, catheters, balloons, and stents that vary in size and length (from 1 to 3 meters) for precise forward, backward, and rotational movements within blood vessels. Even though these movements are relatively simple, the extended pathways add complexity. Moreover, instruments that vary in size, length, and flexibility are inherently non-rigid, which presents challenges in managing their movement paths through complex, narrow endovascular pathways with limited trajectory control. Consequently, achieving accurate placement of the catheter or guidewire and its delivery mechanism, which together form an underactuated configuration, continues to be a significant challenge in interventional surgery.

Robot-assisted endovascular interventions in underactuated environments are typically performed in a two-stage process. The first stage involves manual procedures to establish stable arterial access using flexible surgical instruments, followed by robotic navigation, often assisted by a surgeon or aide. The second stage is the robot-assisted intervention, where the interventionalist remotely operates the robot to insert guidewires, balloons, and stent catheters into endovascular lesions, guided by real-time two-dimensional imaging. These stages are illustrated in **Figure 2.4**. VIRs are categorized into four main types: robot-assisted percutaneous coronary intervention (R-PCI), robot-assisted peripheral endovascular intervention (R-PVI), robot-assisted neurovascular intervention (R-NVI), and robot-assisted electrophysiological intervention (R-EPI). R-PCI, R-PVI, and R-NVI share similarities in their procedural workflows, differing primarily in the anatomical regions involved—such as coronary arteries, lower extremities, and cranial vasculature. Initially, commercial robotic systems were designed as specialized tools for either R-PVI or R-PCI. The Hansen Magellan system exemplifies an R-PVI system, while the CorPath and R-One systems (Robocath Inc., Rouen, France) are tailored for R-PCI. Attempts have been made to adapt these systems for broader applications in endovascular interventions with minimal modifications. For instance, the CorPath® GRX robotic system has been evaluated for R-PCI, R-PVI, and R-NVI, making it a multi-specialty vascular robotic platform [111-113]. R-PCI, R-PVI, and R-NVI all involve the use of flexible tools like catheters, guidewires, balloons, and stents, which VIR systems navigate to specific sites within blood vessels. In contrast, R-EPI focus on distinct anatomical targets and procedural flows that differ significantly from the endovascular procedures. Consequently, electrophysiological interventional robots (EPIRs) are typically specialized systems, designed with features suited for cardiac ablation procedures. These include catheter-tip flexibility, distal force measurement, enhanced catheter manoeuvrability, magnetic navigation, and comprehensive cardiac mapping. These attributes are crucial for accurately positioning the steerable catheter, particularly when treating cardiac arrhythmias and atrial flutter.

Robot-assisted endovascular interventional therapy has demonstrated promising results in early clinical trials. In 2006, Beyar *et al.* performed the first robot-assisted PCI using the RNS system in

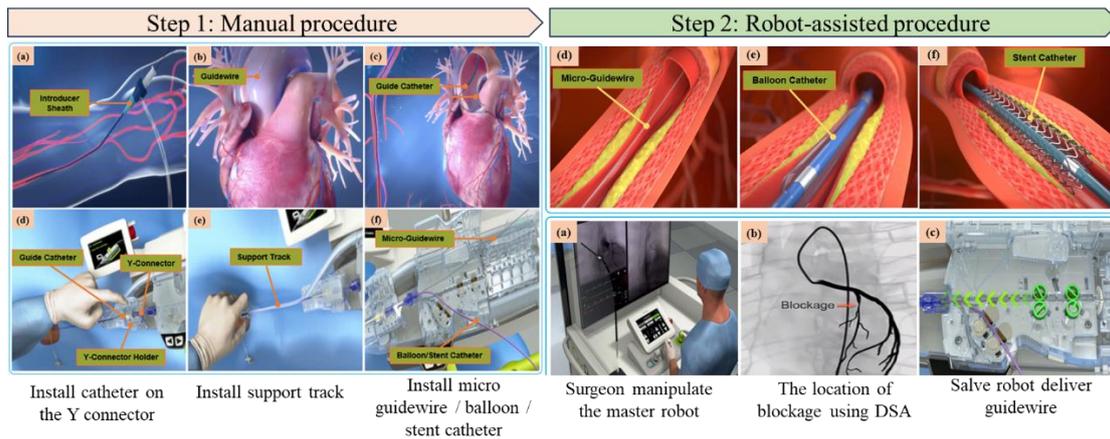


Figure 2.4: Overview of the robot-assisted endovascular interventional processing procedure.

18 patients. The trial successfully completed guidewire crossing through the lesion in 17 patients, resulting in a clinical success rate of 100%, a technical success rate of 94%, and an overall surgical success rate of 83% [23]. This trial marked a significant milestone in the development of vascular interventional surgery robots. As these robots continue to improve, they enable surgeons to perform robot-assisted endovascular interventions in various specialties, including interventional cardiology, peripheral endovascular surgery, neurovascular surgery, and cardiac electrophysiology.

Granada *et al.* reported the first human trial of the CorPath® 200 robotic system for percutaneous coronary artery intervention [114]. This clinical trial involved the robot-assisted delivery and manipulation of coronary guidewires, balloons, and stents in eight patients, assessing the system's safety and feasibility. The results showed a 97% reduction in radiation exposure to surgeons compared to manual surgery, with a technical success rate of 97.9% and no equipment-related complications. In 2013, Weisz *et al.* evaluated the safety and clinical efficacy of CorPath® 200 R-PCI in the PRECISE (Percutaneous Robotically Enhanced Coronary Intervention) clinical trial [115]. Among 164 subjects, 112 patients (68.3%) had A or B1 lesions, while the remaining cases involved type B2 (18.9%) or type C (12.8%) lesions. The maximum lesion length was 24 mm. In this trial, 160 patients (97.6%) experienced clinical success, and the operator's radiation exposure was reduced by 95.2% compared to the operating table (0.98 vs. 20.6 μ Gy). This demonstrated that R-PCI addresses some occupational hazards for interventionalists without compromising patient safety or procedural performance.

Mahmud *et al.* [116] recruited 315 patients with complex type B2 and C lesions in the CORA-PCI (Complex Robotically Assisted Percutaneous Coronary Intervention) study, dividing them into two groups: R-PCI and manual PCI (M-PCI). The results indicated that the clinical success rate of R-PCI was 99.1%, comparable to that of M-PCI, thus confirming the feasibility, safety, and high technical success of R-PCI in treating complex coronary disease. Subsequently, Smitson *et al.* [111] reported the first human clinical trial of the second-generation robotic system, CorPath® GRX, in treating complex coronary artery disease, achieving a clinical success rate of 97.5%. This further validated the safety and effectiveness of CorPath® GRX in managing complex coronary artery disease.

In a clinical report on robot-assisted PVI, Mahmud *et al.* demonstrated that feasibility and safety of CorPath® 200 in treating peripheral arterial diseases [117]. The study enrolled 20 subjects, primarily with Rutherford class 2 to 3 symptoms (90%), and treated 29 lesions. The system achieved a 100% technical success rate, safety, and clinical success rate, with no adverse events related to the robotic system. This led to FDA approval of the device for peripheral interventions.

For NVI, the CorPath® GRX system was modified, including the addition of an active device fixation feature that stabilizes the guidewire during microcatheter movement and enables guide-catheter manipulation. Clinical studies reported satisfactory outcomes with this modified robot in NVI. Weinberg *et al.* [118] confirmed the feasibility and safety of trans-radial carotid artery stenting (TRCAS) assisted by the CorPath® GRX robot. A comparison between robotic and manual procedures revealed no significant differences in baseline characteristics, contrast-agent dose, radiation exposure, catheter replacement, technical success, or transfemoral artery-conversion rates. There were also no technical or procedural complications, demonstrating that robot-assisted TRCAS is feasible, safe, and effective.

However, concerns remain regarding the lack of effective force feedback in current robotic systems compared to manual procedures and the extended learning curve required for mastering robotic manipulation skills. Perera *et al.* [119] published a clinical report indicating that arch-catheter placement reduces cerebral embolization during thoracic endovascular aortic repair (TEVAR) with the Magellan robot. A comparison of robotic and manual procedures showed improved outcomes with robotic catheter placement.

Several retrospective studies have evaluated the use of the Sensei X robotic navigation system in the ablation of atrial fibrillation and atrial flutter, showing that remote-controlled robotic systems are feasible for cardiac mapping and radiofrequency ablation, with prolonged ablation duration and a low recurrence rate of atrial fibrillation [120]. Studies have investigated the feasibility and safety of using the Amigo remote catheter system for Cavo-tricuspid isthmus ablation in patients with typical atrial flutter, particularly in right heart mapping. The results indicate that Amigo is effective and safe [121], though procedural times may be extended in resource-limited settings [122]. This finding highlights the need for optimizing the system's efficiency to facilitate broader clinical adoption.

Other studies assessed the feasibility, efficiency, and safety of Cavo-tricuspid isthmus ablation using the Amigo remote catheter system in patients with typical atrial flutter and in mapping the right side of the heart. The results demonstrated that Amigo is safe and effective for these procedures but may result in longer procedural times in resource-limited setting. Meanwhile, VIR systems have shown significant disadvantages, as they often lack tactile feedback, which is essential for manoeuvring through tight vessels. Furthermore, current robotic systems do not yet accommodate guidewire coronary intervention. Even though they facilitate balloon operations, two vital procedures, gaining arterial access and handling the guide catheter, remain manual tasks. Furthermore, in resource-limited settings, robot-assisted PCI might lead to longer procedures than

traditional manual PCI [123]. **Table 2.2** summarizes the main clinical trial studies focusing on VIRs.

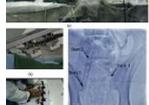
Table 2.2: Summary of main clinical trial studies of VIRs.

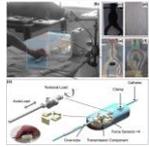
Clinical trial	Device	Intervention	Treated lesions	Technical success rate	Clinical success rate	Radiation reduction rate	Year
Granada <i>et al.</i> [114]	CorPath 200	R-PCI	8	97.9%	100%	97%	2011
Weisz <i>et al.</i> [115]	CorPath 200	R-PCI	164	98.8%	97.6%	95.2%	2013
Lopez-Gil <i>et al.</i> [121]	AMIGO	R-EPI	60	98	100%	70.73%	2014
Mahmud <i>et al.</i> [117]	CorPath 200	R-PVI	20	100%	100%	52.67%	2016
Dello Russo <i>et al.</i> [120]	Sensei X	R-PVI	40	100	90%	35%	2016
Mahmud <i>et al.</i> [116]	CorPath 200	R-PCI	315	91.7%	99.1%	5%	2017
Perera <i>et al.</i> [119]	Magellan	R-TEVAR	11	72.08%	100%	N/A	2017
Smitson <i>et al.</i> [111]	CorPath GRX	R-PCI	40	90%	97.5%	N/A	2018
Hoffmayer <i>et al.</i> [122]	AMIGO	R-EPI	25	100%	100%	1.79%	2018
Patel <i>et al.</i> [124]	CorPath GRX	R-PCI	5	100%	100%	N/A	2019
Weinberg <i>et al.</i> [118]	CorPath GRX	R-EPI	13	100%	100%	N/A	2020
Häner <i>et al.</i> [125]	CorPath GRX	R-PCI	21	81%	100%	N/A	2023
Leung <i>et al.</i> [126]	CorPath GRX	R-PCI	21	81%	100%	1.04%	2024

2.3. Intuitive Manipulation of VIR

Several modern robotic systems have been developed and commercialized to address stenosis and plaque in blood vessels without the need for large or multiple incisions in patients, as shown in **Table 2.3**. Although the effectiveness of commercial interventional robots has been clinically proven, they present difficulties for surgeons due to the steep learning curve, particularly for procedures that differ from the usual skills of a surgeon [127]. Surgeons may need extensive training to effectively operate these complex systems. Isomorphic teleoperation uses ergonomic master interfaces that allow surgeons to mimic their natural hand movements during interventions. In this configuration, the master and slave systems have similar structural and functional designs, ensuring that the commands on the master side are accurately mirrored by the slave device. This allows the slave device to demonstrate the interventionalist's dexterity with hands and fingers for precise tool handling. Isomorphic configurations are relatively new in the field of endovascular intervention. However, recent studies have shown that they can reduce the surgeon's learning curve by allowing him to directly apply his inherent catheterization skills.

Table 2.3: Summary of related VIR systems.

Group	Manipulation strategy	Control scheme	Teleoperation setup	Guidance systems	Haptic perception	Areas	Figure
CorPath® 200&GRX [111-113]	Friction roller-based or Rotating Clamped Wheel	Position and velocity	Non-Isom.	DSA	Obstacle feedback	PCI PVI NVI	 (Siemens Healthineers)
Magellan® [90, 128]	Friction roller-based or Friction wheel-based rotation	Position and velocity	Non-Isom.	DSA or CT	Haptic	PVI EPI	 (Hansen Medical)
Amigo® [129]	Linear actuator or Linear Relative Motion Drive	Position and velocity	Non-Isom.	DSA	N/A	EPI	 (Catheter Precision)
Guo <i>et al.</i> [130, 131]	Clamped-based or Rotating Clamped Wheel	Position and force	Non-Isom. or Isom.	DSA	Haptic or Proximal force	PCI	 Beijing institute of Technology
Wang <i>et al.</i> [132, 133]	Clamped-based or Rotating Clamped Wheel	Position and force	Isom.	DSA	Haptic or Proximal force	PCI	 Yanshan university
Wang <i>et al.</i> [134, 135]	Friction roller-based or Bionic finger-based rotary	Position and velocity	Non-Isom. or Isom.	DSA	N/A	NVI	 Beijing University of Aeronautics and Astronautics Robotics Institute
Wang <i>et al.</i> [136, 137]	Clamped-based or Rotating Clamped Wheel	Position and velocity	Non-Isom.	DSA	N/A	PCI PVI	 Shanghai Jiao Tong University
Thakur <i>et al.</i> [138]	Friction roller-based or Rotating Clamped Wheel	Position and velocity	Isom.	N/A	N/A	PCI	 The University of Western Ontario
Cha <i>et al.</i> [108, 139]	Friction roller-based or Rotating Clamped Wheel	Position and force	Non-Isom.	DSA	Haptic or Proximal force	PCI	 Hanyang University
Choi <i>et al.</i> [140]	Friction roller-based or Bionic finger-based rotary	Position and velocity	Non-Isom.	DSA	N/A	PCI	 University of Ulsan College of Medicine

Payne <i>et al.</i> [141]	Friction roller-based or Rotating Clamped Wheel	Position and force	Isom.	N/A	Distal and Proximal force	PCI	 Imperial College London
Tanimoto <i>et al.</i> [142]	Friction roller-based or Rotating Clamped Wheel	Position and force	Isom.	CT	Haptic or Distal and Proximal force	PCI	 Nagoya University
Wang <i>et al.</i> [29, 143, 144]	Clamped-based or Rotating Clamped Wheel	Position and force	Isom.	DSA	Haptic or Proximal force	PCI	 SIAT, CAS
Bian <i>et al.</i> [145, 146]	Friction roller-based or Bionic finger-based rotary	Position and velocity	Isom.	DSA	N/A	PCI	 Institute of Automation, CAS
Langsch <i>et al.</i> [147]	N/A	Position and velocity	N/A	US	Proximal force	PCI	 Technische University Munich
Li <i>et al.</i> [148]	Clamp-based or Rotating clamped wheel	Position and velocity	N/A	IVUSOC T	N/A	PCI	 University of Southern California
Sankaran <i>et al.</i> [149]	Friction roller-based or Rotating Clamped Wheel	Position and force	Isom.	N/A	Haptic or Proximal force	PCI	 University of Illinois at Urbana Champaign
Zhou <i>et al.</i> [150]	Clamped-based or Rotating Clamped Wheel	Position and force	Non-Isom.	N/A	Proximal force	PCI	 Xiamen University

Thakur *et al.* [138] developed an isomorphic design that uses a real input catheter as the master device, with a sensor to capture catheter motion, while the slave device replicates this movement to guide a catheter inside the vessel. Similarly, Payne *et al.* [141] designed a novel master-slave force-feedback system that aligns with the surgeon's natural operating preferences and ergonomics. The interface of this system is user-friendly and easy to learn, leading to a wider adoption in recent studies [151-153]. However, these systems still face challenges to achieve improvements in intervention visualization, clinical application efficiency, and operational convenience simultaneously.

The clinical application of VIRs is still in an early stage and its widespread acceptance is restricted. Since endovascular interventional procedures require flawless cooperation between the surgeon and the robot to achieve safe, accurate and skilful manipulation of instruments within the patient's blood

vessels, ensuring the safety of interventional procedures remains a significant challenge in the advancement of these systems. The performance of robot-assisted interventional tasks depends on the surgeons' intuitive manipulation, and researchers have extensively explored the intuitive manipulation aspects of master-slave isomorphic interventional robots for precisely and safely handling interventional instruments. The relevant progress is presented in **Table 2.4**.

Table 2.4: The progression of previous research.

Group	Surgery	Manipulation		Dataset	Instruments	Methods	Results
		Manual	Robot-assisted				
Sánchez <i>et al.</i> [154]	laparoscopic	✓	-	Accelerometer Data	Surgical forceps	Statistic difference	Novice and expert
Rolls <i>et al.</i> [155]	Vascular intervention	✓	-	X-ray video	Guidewire	Statistic difference	Novice and expert
RafiiTari <i>et al.</i> [156, 157]	Vascular intervention	✓	-	Tool forces, contact forces	Guidewire	SVM	90% Novice and expert
King <i>et al.</i> [158]	Laparoscopic surgery	✓	-	Glove	Surgical forceps	HMM	92% Novice and expert
Zhou <i>et al.</i> [159, 160]	Vascular intervention	✓	-	EM, EMG, fibre-optic sensors	Guidewire	GMM	92% Novice and expert, 90% manipulation pattern
Du <i>et al.</i> [144]	Vascular intervention	✓	✓	EM, sEMG, glove, force	Guidewire	MLP, SVM, RF, HAR-CNN	94.11% manipulation pattern
Shen <i>et al.</i> [161]	Laparoscopic surgery	-	✓	2017 and 2018 EndoVis endoscopic images, Lapavis dataset	Surgical forceps	BAANet	52.62% MIoU, 40.64% MIoU
Yang <i>et al.</i> [162, 163], Bian <i>et al.</i> [164]	Laparoscopic surgery	-	✓	2017 and 2018 EndoVis endoscopic images	Surgical forceps	MSDE-Net	94.17% MIoU (2024)
Zhang <i>et al.</i> [165, 166]	Vascular intervention	✓	-	Self-collected DSA images	Guidewire	BRA	89.9% f1-score
Omisore <i>et al.</i> [167]	Vascular intervention	-	✓	Self-collected DSA images	Guidewire	MLB-Net	84.89% MIoU
Zhou <i>et al.</i> [168, 169]	Vascular intervention	✓	-	Self-collected DSA images	Guidewire, retinal microsurgery dataset	Real-Time Endpoint Localization	2.20 pixels MPE, 5.30 pixels MPE,

2.3.1 Underactuated intraluminal manipulation modelling

Endovascular interventional surgery is a highly intricate and delicate task. Safely manoeuvring flexible endovascular instruments through blood vessels for tasks such as stent and balloon placement requires the surgeon's skill in using their forefinger and thumb [170]. These skilful hand movements entail precise manipulation of endovascular tools, allowing the surgeon to deftly steer the slender, elongated, and flexible devices back and forth within the vessels. In the case of a bifurcation, the fingers collaborate, shifting relative to one another to change the guidewire tip's

direction, thereby successfully navigating the bifurcation. **Figure 2.5** shows a schematic diagram of guidewire manipulations and the force analysis performed by the surgeon's thumb and index finger. Robot-assisted interventional surgery relies on intuitive control, which requires executing three basic actions: pressing, translating, and rotating. These actions allow for two degrees of freedom (axial and rotary) in manoeuvring endovascular instruments within blood vessels.

Surgeons perform angioplasty using 2D fluoroscopic images and tactile sensations experienced through their fingertips when handling instruments. This procedure is highly dependent on the surgeon's familiarity with 3D anatomical models of blood vessels and the operation of instruments [156]. Previous research has shown that the success rate of endovascular interventional surgery is significantly influenced by the skills and experience of the operator [171]. However, the learning curve to master endovascular catheterization is extensive and challenging. To quickly improve operator proficiency and precision in surgical methods, researchers have conducted comprehensive studies focusing on the inherent manipulation behaviour of surgeons.

Recent investigations have begun to investigate the implementation of innate behaviours, such as bodily motion signals from surgeons during surgery, within the medical field. Srimathveeravalli *et al.* [172] used EM sensors on the thumb and index finger to monitor interventionist hand movements, analyses the data of typical finger movements to establish design criteria for safe movements and forces in robotic systems. Villarruel *et al.* [173] designed a robotic surgical apparatus governed by muscle activity utilizing non-invasive electromyography (EMG) sensors to perform surgery remotely. Similarly, muscle activity was used in [174] for automatic identification of surgical manipulations and real-time detection of irregularities during simulated laparoscopic operations. Li *et al.* [175] implemented 14 bespoke bend sensors to fully record finger movement, reflecting clinical hand performance. Sánchez *et al.* [154] utilised accelerometers placed on the dominant wrists of surgeons assessed the force exerted during surgical procedures. Rolls *et al.* devised semiautomated catheter tracking software to scrutinize motion in videos from a virtual reality simulator frame by frame, calculating the path length at the end of a 2D catheter. Using it as a criterion for evaluation in carotid stenting, experienced operators were shown to generally follow a shorter path compared to novice surgeons [155]. Rafii-Tari *et al.* introduced a novel force measurement platform that measures the operator's proximal propulsion force and the catheter's interaction force with tissue. This platform also collects position sensor data linked to the catheter tip as an objective skill indicator for interventionists in a realistic simulation setting, evaluating the

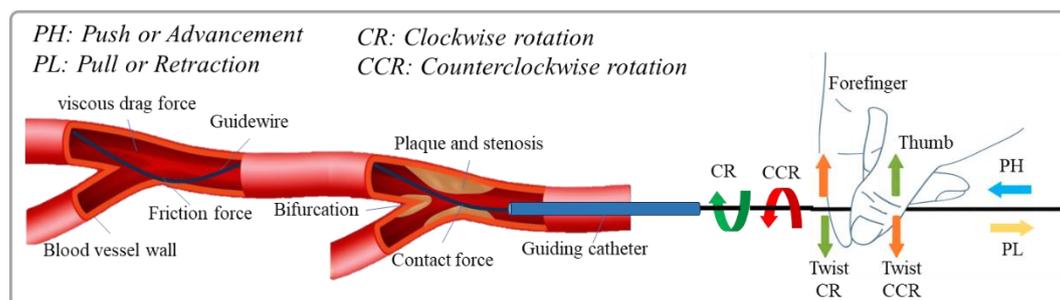


Figure 2.5: The manipulation analysis in endovascular interventional procedure [30].

natural behavioural data of operators with diverse degrees of experience [156, 157].

Efforts have been made to integrate various motion capture sensors to achieve a more precise and complete depiction of surgeons' diverse natural behaviours. For instance, Tercero *et al.* [176] from Nagoya University used photo-elastic stress analysis to display the contact stress between medical instruments and a specific endovascular model. They used an optical encoder to track the movement of the proximal catheter and implemented an EM tracker to monitor the motion of the operator's fingers and wrists. This system was created to assess and measure skill levels by obtaining parameters related to hand movements and the motion of endovascular devices. King *et al.* [158] developed a wireless glove that includes seven accelerometers and a fibre-optic bend sensor to capture force data and finger motion during laparoscopic procedures. Similarly, Perez-Duarte *et al.* [177] used EMG sensors to examine muscle activity, and a motion capture glove equipped with bend sensors to study finger motion, with a focus on evaluating surgeon ergonomics during single-site laparoendoscopic surgeries.

To date, the natural behaviours observed during surgical operations have mainly been used to evaluate skills. However, these behaviours can be used to identify endovascular tool manipulation patterns, helping to develop novel human-robot interfaces. However, current methods face obstacles such as inadequate information for real-time recognition that involves two forms of natural behaviour and a lack of redundancy analysis. For example, research conducted by [172] and [176] showed that hands movements, as recorded by EM sensors on closely positioned fingers, showed a significant level of correlation, suggesting redundancy. Furthermore, not all behaviours detected by motion sensors are relevant for the movements of the endovascular tool. Multi-DoF sensors generate multi-dimensional data that represent various sub-behaviours, some of which do not pertain to tool motions. However, the significance of these behaviours has not been thoroughly investigated in the present studies.

These limitations have driven the exploration of various motion sensors to capture the diverse natural behaviours of interventionalists during standard PCI. Previous studies [178, 179] have effectively used motion capture sensors to identify hand postures, gestures, and movements. Zhou *et al.* [159, 160] introduced a behaviour-based assessment approach, where four natural behaviours (hand movement, proximal strength, muscle activity, and finger movement) of the surgeon's arms were used as behavioural indicators to qualitatively and quantitatively evaluate the skill levels of interventionists at various stages. Du *et al.* [144] explored surface electromyography, hand movement, and tactile signals of the operator's right arm during endovascular intubation, focusing on hand and finger movement patterns of the operator to assess the precision of integrating surgeon operational patterns into robot-assisted PCI. Xie *et al.* [180] developed an innovative multimodal information fusion framework based on deep learning (DL) to recognize and analyse eight common interventionist operating behaviours using four types of sensors [159, 160, 180].

It is crucial to manipulate the robotic system accurately to guide the tool to the target site efficiently and safely in endovascular interventions. Expert operators typically exhibit superior

control skills compared to novices, leading to higher success rates. Nevertheless, both experts and novices can encounter failed attempts, and such errors elevate the risks of endovascular interventions. Therefore, understanding and distinguishing successful manipulation patterns is more critical than differentiating between expert and novice performance when transferring operator skills to robotic systems. Thus, creating an analytical framework to assess the redundancy of behaviours or characteristics and their significance to successful or unsuccessful robot-assisted interventional performance is crucial. An effective classification framework could pinpoint relevant behaviours and characteristics, enhancing the detection of endovascular manipulation patterns to improve the optimization of robot-assisted safety manipulation.

2.3.2 Force-based information fusion robot-assisted manipulation

In robotic endovascular intervention procedures, acquiring multi-sensor behaviour data is crucial for identifying manipulation patterns, while measuring proximal and distal forces is vital for maintaining safety and enhancing operational strategies. Proximal forces are gauged near the catheter's operating end using commercially available force sensors, whereas distal forces, which are more significant to the interventionalist, are assessed at the catheter's distal tip. Detecting the contact force between the catheter tip and vessel walls to measure distal force necessitates a miniaturized sensor within the coronary catheter, with important features including size, resolution, biocompatibility, measurement capacity, and precision. Force-sensing catheters are used in both endovascular and electrophysiological procedures. In these settings, they measure the contact force between the catheter electrode and the myocardium, aiding in avoiding excessive force and forming scars on abnormal heart tissue. This technology has been commercialized, featuring products like the IntelliSense® System and TactiCath® Catheter (Endosense SA, Geneva, Switzerland). The TactiCath catheter is a steerable 7-F radiofrequency ablation device equipped with a force sensor at its distal tip, which measures the contact force between the catheter tip and the cardiac tissue.

To improve stability during dynamic conditions in endovascular procedures, force-sensing catheters commonly integrate strain gauges or pressure-sensitive rubber, with the active component of the catheter encased for biocompatibility. For instance, Guo *et al.* [181] placed pressure-sensitive rubber sensing units in a frontal duct array and sealed them to sense forces at various nodes. Omisore *et al.* [143], Payne *et al.* [141], and Wei *et al.* [182] have engineered catheters with over-the-wire force sensing (e.g. using strain gauges) to assess endovascular contact forces. Even though resistive and strain sensors exhibit excellent linearity, their associated hardware, including circuits and metal substrates, is susceptible to electromagnetic (EM) interference during operations, and their dimensions are often too large for narrow endovascular passages. However, fibre bragg grating (FBG) sensors provide advantages such as compact size, resistance to electromagnetic interference, and high sensitivity. Recently, researchers have embedded FBG-based fibre optic force transducers in catheters to mitigate size limitations and electromagnetic interference problems. He *et al.* [183] devised a catheter that integrates four FBGs to measure endovascular forces and compensate for temperature. The distal force measuring catheter faces challenges in packaging, sterilization, and

size configuration, restricting its clinical use in narrow blood vessel lumens.

To address these challenges, researchers are placing more emphasis on techniques that indirectly infer distal force using proximal force readings. The proximal force consists of four components of endovascular force: the viscous force of the blood, the impact force at the front of the instrument, the friction force between the blood vessels, and the potential elastic force of the guidewire [184], as presented in **Figure 2.5**. Typically, proximal force measurement is sensor-based, involving the installation of a high-precision, high-resolution force sensor on the device delivery setup to gauge the contact force between the device and the blood vessel. However, there are differences in sensor installation and the methods of measuring force. Yang *et al.* [185] engineered a system for measuring force in guidewires that uses the lever principle, with the pressure sensor located on the propulsion finger and resistance transmitted to the force measurement setup through the guidewire. The lever magnifies the signal, making the force measurement system less prone to interruptions in the transmission framework. Bao *et al.* [186] mounted the force sensor on the clamping side of the guidewire, used a linear bearing to limit friction interference on the resistance of the guidewire during transmission, and proposed a multi-level safety control strategy based on force levels to decrease operational risks. Also, Zhou *et al.* [150] and Wei *et al.* [182] applied the technique of placing a sensor in the guidewire clamping section to assess and analyse the guidewire's resistance, thus improving robot safety control. Sankaran *et al.* [149]. applied the current from the drive motor to assess the guidewire's resistance and adopted a double-layer optimization approach for calibration. Sensor-based methods to measure the proximal force are generally the most sensitive and can accurately assess the resistance while the low translation speed, the inclination angle is minimal, and the friction of the mechanism is insignificant [187]. However, during robotic operations, the proximal resistance value is often inaccurate and prone to interference, principally due to poor resistance resulting from friction forces in the actuator, inertia forces, or jerks in the linear drive system [188, 189].

This thesis integrates high-precision sensors into the master-slave mechanisms of the surgical robot to assess the performance of interventional manipulation through dynamic variations in proximal and distal forces. The thesis investigates the correlation between these forces and interventional manipulation, leading to improved robotic system manipulation performance and ultimately increasing the accuracy and dependability of robot-assisted interventional procedures.

2.3.3 Visual perception of interventional instruments for tool manipulation

In robot-assisted interventional procedures, surgeons rely on their intuitive manipulation skills that are guided by real-time perception of the movement of their instruments. Broad visual and complex working environments, accompanied by challenges such as specular reflection, blood presence, complex backgrounds, and fogging of camera lenses, raise the risk of manipulation errors and the potential damage of tissue [190]. The use of visual data can help mitigate these risks by offering segmented surgical instruments, real-time tracking, and pose estimation [191]. Segmenting surgical instruments is a vital aspect of robot-assisted procedures, as it provides crucial

intraoperative guidance information that supports decision making [192]. Furthermore, segmenting these instruments is essential to track them, estimate their pose, use augmented reality overlays, and analyse surgical phases to ensure surgical safety and assess surgeon situational awareness to reduce manipulation errors [193].

Surgical scenes comprise 2D images or videos depicting surgical tools and patient tissues, with numerous tools used depending on surgical needs. For example, gastrointestinal endoscopy does not require an incision and only a few surgical tools like biopsy forceps, balloon dilation, snares, and submucosal injection techniques. Other surgical instruments include pro-grasp forceps, cadiere forceps, bipolar forceps, monopolar curved scissors, a large needle driver, an ultrasound probe that can be dropped into, and a vessel sealer. Segmenting surgical tools is challenging due to the variety of surgical instruments, the occlusions seen, and the anatomy of the background tissues [193].

Several studies have investigated several DL techniques for segmenting surgical instruments, which have yielded promising results. For example, Sestini *et al.* [194] proposed a fully unsupervised method for segmenting surgical instruments in endoscopic images. It uses implicit motion data and shape priors to train a segmentation model per frame, achieving performance nearly on par with fully supervised methods in minimally invasive surgery. Shen *et al.* [161] introduced a lightweight network for instrument segmentation. This network integrates a branch balance aggregate module for optimizing features and reducing noise, and a block attention fusion module for integrating global and local contexts, outperforming existing methods by up to 4.03% in mean-intersection-over-union (MIoU) scores on the Laparoscopic Vision dataset. A transformer-based network was introduced to segment surgical tools in endoscopic images [162], which uses a trapezoid ASPP block, multi-scale attention fusion, and a dual encoder unit to enhance the representation of features and the precision of segmentation. Experiments conducted on the Kvasir-Instrument and Endovis2017 datasets indicate that it outperformed the state-of-the-art (SOTA) methods. Yang *et al.* [163] developed a multi-scale dual encoding network for surgical tool segmentation. This network, named MSDE-Net, integrates CNN and transformer branches with attention and context fusion blocks to improve feature extraction across multiple scales, achieving significantly superior results in the endoscopic image datasets from Endovis2017 and the Kvasir-Instrument compared to previous methods. Bian *et al.* [164] presented a dual branch fusion network that merges CNN and transformer architectures for surgical instrument segmentation, providing MBTPDS models and offering excellent segmentation performance in endoscopically viewed sinus surgery images. Using visual information from these surgical instruments, surgeons can adapt their manipulation strategies in real time, reducing the risk of interventional instrument collisions with non-target tissues and thus improving surgical efficiency.

The above these surgical instrument segmentation methods during surgical procedures have primarily been used for rigid instruments of endoscopic images based on the popular Da Vinci robot-assisted surgery. However, these methods could also be applied to segmentation and location of endovascular tools, facilitating the development of position or posture visual information of catheter

or guidewire on upscaling master-slave endovascular interventional manipulation. Unfortunately, research on this application is limited. Endovascular intervention surgeries are based on visual feedback displayed on a screen, including guidewire and catheter paths under X-ray fluoroscopy and angiographic imaging. Precise manipulation of endovascular devices (such as catheters and guidewires) within blood vessels is crucial for MIS. In these operations, surgeons must constantly perceive the exact position of endovascular tools to navigate catheters or guidewires safely within blood vessels. At present, DSA is the leading imaging technique among specialists due to its excellent spatial and temporal resolution. Using the DSA system, surgeons can determine the best interventional route or assess the size and distribution of the lesion using 2D angiography and fluoroscopy sequences in real time, supported by their anatomical expertise. Technologies such as image recognition, decomposition, fusion, and tracking are capable of effectively using these visual inputs and combining them with surgical planning and robotic motion programming to achieve precise control, thus increasing the safety of the surgical robot with endovascular intervention.

Clinicians rely on delicate tactile feedback from handheld devices to evaluate the condition of the guidewire or catheter within vessels and to reduce risks during manual endovascular procedures. However, the current surgical robots available commercially have not developed an ideal solution for mechanical force feedback and touch perception. A major drawback of these systems is their inability to detect and offer feedback on the contact forces between the tool and the vessel during operations. Without this feedback, the risk of complications such as thrombosis and vessel perforation increase, which may be due to the reliance on unclear visual feedback, inadequate hand-eye coordination, and the mental and cognitive strain on the operator [127]. This drawback is one of the impediments to precise control and broader acceptance of VIRs. Although several strategies have been tried to address this problem, the technology still lacks maturity and stability. Researchers are investigating various methods to mitigate the deficiency caused by the absence of sensory force feedback on the handling performance of robot-assisted interventional processes. Many skilled robotic surgeons believe that the lack of haptic force feedback can be compensated for sufficiently using visual cues [195-197].

Thus, improving the perception of the position or location of catheters and guidewires is crucial to increasing the manipulation performance in robot-assisted endovascular intervention procedures. Surgeons need to perform precise and rapid manipulations based on accessible real-time data, such as the configuration and guidewires' position visible in X-ray images, tactile feedback, and patient physiological signals. To this end, real-time and precise guidewire segmentation and tracking can improve guidewire visualization and provide visual information to surgeons during the intervention, as well as help scale robot-assisted interventional manipulation. Nevertheless, this task faces challenges with elongated, deformable structures and non-rigid bodies that appear with low contrast or can be easily mistaken for similar objects like bones or lungs in the noisy fluoroscopic image sequences depicted in **Figure 2.6**, which diminishes segmentation and location precision.

Zhang *et al.* [165] introduced a jigsaw-training-based background reverse attention (BRA)

transformer network to segment single and dual guidewires in X-ray fluoroscopy sequences. The method features a BRA module to mitigate background distractions and uses robust principal component analysis to create background maps, achieving an F1 score of 89.4% and a precision of 89.9%. Zhang *et al.* [166] also proposed a deep learning framework for real-time guidewire segmentation and tracking, utilizing a Yolov5s detector refined with spatiotemporal enhancements for guidewire localization, and leveraging Hessian-based enhancement alongside dual self-attention for segmentation, resulting in an 89.9% dice score for low-quality images. Omisore *et al.* [167] presented a multilateral branched network for guidewire segmentation in angiograms during robotic catheterization, achieving robust segmentation with an MIoU of 84.89% and F1 scores of 89.01% [198], outperforming state-of-the-art methods such as DeepLabV3+, SegNet, and U-Net in both accuracy and speed.

Meanwhile, a two-stage framework using a YOLOv3 detector was proposed for real-time localization of dual-guidewire endpoints in fluoroscopic images, attaining top-tier localization results with an average pixel error of 5.30 pixels on a retinal microsurgery dataset and 2.20 pixels on the guidewire dataset [168, 169]. These visual information modelling techniques allow surgeons to observe interactions between instruments and tissues during operations. Therefore, creating a visual-based detection framework for interventional instruments aims to increase the detection precision of their position or posture, improving surgeons' precise perception of interventional instrument motion in robot-assisted tasks. A well-designed segmentation and location framework can emphasize crucial feature maps, thus improving the segmentation and location accuracy of endovascular interventional instruments for optimal robot-assisted safety operations. It also helps surgeons adapt operational strategies in response to changes in endovascular instruments, thus reducing surgical workload while improving accuracy and safety.

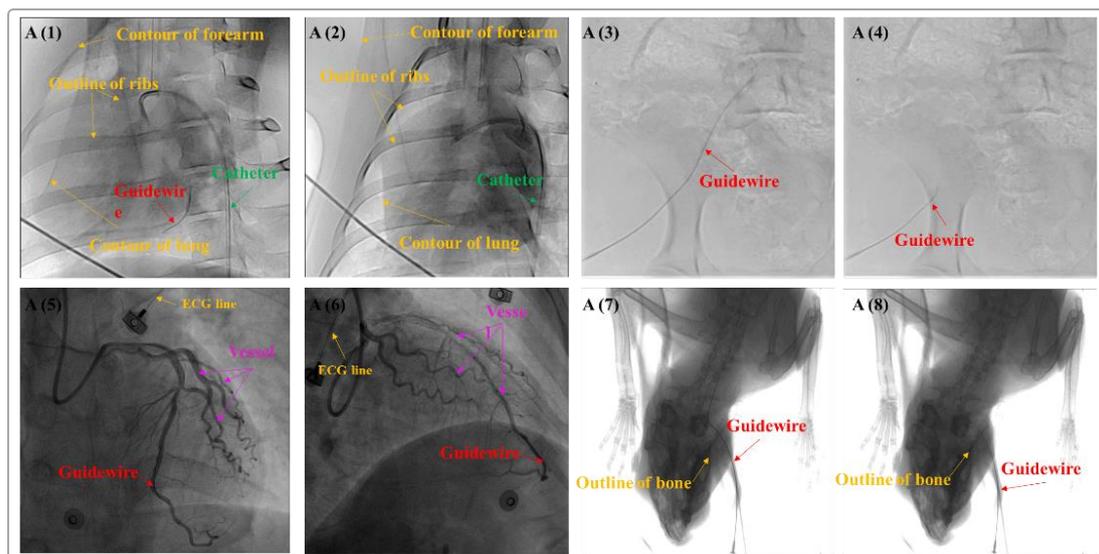


Figure 2.6: Illustration of DSA images in endovascular interventional procedure.

2.4. Technical Challenges and Future Development

The use of robotic-assisted endovascular procedures in different endovascular tissues in the human body is gaining acceptance due to the significant potential of robotic technology to improve procedural precision, improve safety, reduce health risks, and offer clear benefits to both patients and surgeons. However, despite these advancements, there are still several technical issues that need to be resolved to support the wider adoption of VIR in tertiary care facilities.

2.4.1 Cooperative mechanisms of multi-instruments and teleoperation setup

Through ongoing advancements and innovations in VIR technology, clinical constraints are progressively reducing. For example, the CorPath® GRX system, which is the second iteration of the Corindus robotic platform, has improved guide catheter handling, guidewire retraction, and other functions, thereby significantly improving the capability of robot-assisted treatment for intricate lesions and broadening the range of robot-assisted endovascular interventions. Endovascular intervention surgeries require the surgeon to work with a variety of endovascular devices such as guidewires, catheters, balloons, and stents. These procedures encompass a wide array of treatments including cardiovascular, cerebrovascular, peripheral vascular, aortic, and tumour vascular interventions. Given the multitude of guidewires, catheters, balloons, and stents available, which differ in size, length, and stiffness, creating a robot solely for proprietary instruments would be very restrictive. This involves moving forward, pulling back, and rotating guidewires, catheters, and sheaths; moving forward, pulling back, and deploying balloons; and positioning and releasing stents. If the robot is restricted to basic capabilities such as advancing and retracting guidewires or catheters with minor rotations and is unable to perform complex endovascular manoeuvres or handle more than one guidewire, balloon catheter, or stent at a time, its practical value in clinical settings would be limited. Therefore, the robot must be equipped to manipulate these commercially available tools and compatible with most commercially available endovascular instruments while providing the capability to perform most surgical manoeuvres to satisfy both clinical and market needs.

Currently, commercial VIRs are restricted to handling a single coronary guidewire and placing one balloon and stent at a time, which proves inadequate for supporting balloons and stents in intricate vasculatures and cannot handle the over-the-wire apparatus. Generally, these robotic systems are apt for simple lesion tasks, but their effectiveness in intricate procedures that require the concurrent operation of two guidewires is low. Significantly, complex lesions, such as chronic total occlusions, severe calcification, and bifurcation lesions, account for more than 50% of interventional treatments for cardiocerebrovascular conditions. Thus, there is an urgent requirement for a multi-instrument robotic delivery system that is compatible with standard endovascular intervention tools. This calls for improvements in both software and hardware control modules for the management of multiple instruments, which could ease some of the challenges surgeons face during complex lesion procedures.

Generally, teleoperation continues to be the most effective standard method for performing surgery. Interventional surgeons can perform endovascular procedures from great distances through remote teleoperation, helping to resolve the uneven regional distribution of healthcare services. The

ongoing improvement, reliability, and widespread availability of communication networks serve as a technical basis for the feasibility and safety of remote surgery. Clinical trials have shown that remote robot-assisted endovascular interventions can provide solutions and guidance to emergency centres that lack cardiovascular specialists in remote and underdeveloped regions. However, providing high-fidelity visual and tactile feedback with minimal transmission delay remains a key research challenge hindering the progress of remote telesurgery.

2.4.2 Automatic surgery

The activities involved in endovascular intervention procedures are often repetitive and follow specific patterns, such as repeated advance and release movements during the handling of the guidewire. This creates the potential for programming robots for semi-automated or automated performance after suitable programming. By automating routine and repetitive surgical duties, surgeons could focus more on the complex and critical parts of surgery. AI holds the promise to improve the automated capabilities of future surgical robots. Some robotic systems under development for surgery already integrate AI, such as those led by Ron Alterovitz [199], who has been working on developing robots that can autonomously navigate through a patient's anatomy to perform procedures with exceptional precision and safety. With advances in AI technology, the move towards surgical automation is emerging as the future trajectory in the development of surgical robots. In January 2022, for the first time, a robot independently conducted a laparoscopic small bowel anastomosis in pig soft tissue, signifying a significant milestone in the progress of surgical automation.

Surgical automation is categorized into six levels, ranging from 0, indicating no automation, to 5, representing full automation. Currently, most surgical robots are positioned in the robot-assisted phase, which poses both major challenges and significant opportunities for the ongoing advancement of surgical automation. The primary objective of medical robotics is to enable robots to independently perform certain surgical tasks, with automation degrees divided into six categories. Currently, VIRs are at automation level 1 and are moving toward achieving level 2. VIRs possess a natural advantage in pushing forward surgical automation compared to other surgical robots due to their relatively predictable and repeatable actions within the endovascular lumen. However, the adaptability of endovascular tools and the intricate, ever-changing environment inside the body pose substantial difficulties in precisely forecasting and adjusting for device movements, which is essential to reach advanced automation levels in VIRs. Moreover, ethical and legal issues related to patient safety add another layer of complexity to the increase in automation in VIRs. Consequently, progressing to higher levels of automation in future VIRs will require addressing these current challenges.

2.5 Endovascular Robotic System Design and Prototyping

This section provides a simple introduction of the vascular interventional robot our developed

that support research on safe handling and accurate guidance of interventional procedures outlined in the paper, acquiring multi-sensor data in **Chapter 3-6** based on this interventional platform for upscaling tool manipulation. It features prototypes of the 3.0 and 3.5 generation robotic endovascular interventional systems created at the Medical Robotics and Minimally Invasive Surgical Instruments Research Centre (SIAT-CAS). The key design priorities for these systems include optimizing structure, controlling manipulation, and providing force feedback, which are essential to maintain high safety standards in interventional surgeries. These issues are tackled through multi-sensor data-driven manipulation modelling and perception modelling based on interventional instruments to achieve accurate and secure manipulation of instruments in robot-assisted endovascular interventional procedures.

2.5.1 Master-slave isomorphic robot-assisted system

A. Design motivation

Endovascular interventional surgery has become a crucial standard in the treatment of cardiovascular diseases. Traditionally, these procedures require the operator to direct the wire and catheter into the blood vessels, while using imaging technologies such as DSA, CT, ultrasound, and MRI for navigation and observation. However, this approach exposes practitioners to X-ray radiation during operations, which can pose considerable health hazards over time. Furthermore, the high-risk nature of the procedure demands an elevated level of expertise, restricting these surgeries to highly specialized practitioners. This situation leads to challenges such as the scarcity of skilled professionals and the high costs and extended duration of training programmers. Furthermore, physiological tremors, manipulation mistakes, and long surgeries can cause practitioner fatigue, significantly compromising operational safety.

The intrinsic link between life and health greatly impacts the progression of medical technology. With a growing emphasis on health, the creation of robotic systems is motivated by the need to improve the safety, precision, and efficiency of endovascular interventions. Significant progress has been made in endovascular therapy through the integration of robotic systems to manipulate flexible catheters, as well as force feedback technology, cutting-edge imaging technologies, and AI. These innovations are enabling new methods to improve the safety, precision, and independence of endovascular disease management. Using minimally invasive vascular interventional surgery robots can reduce radiation exposure of surgeons, partially replace their efforts, reduce work intensity, limit physiological tremors and operational errors, and greatly improve the precision and accuracy of surgeries. Therefore, precise control devices for endovascular interventional tools are crucial for these procedures.

B. VIRs

- ***The 3.0-generation robotic system***

Various vascular interventional robotic systems have been developed to assist in minimally

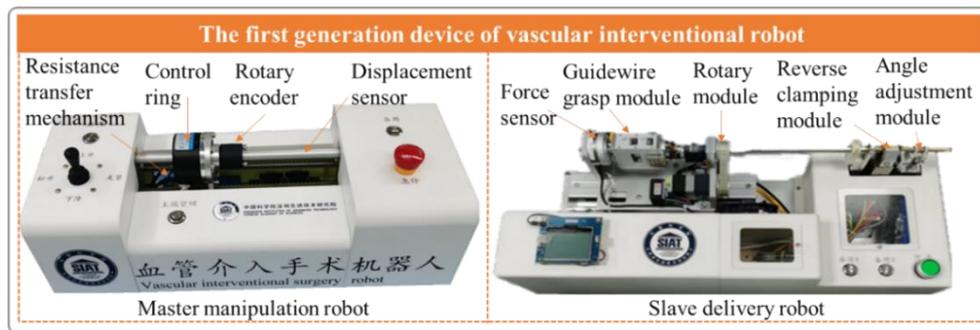


Figure 2.7: 3.0 generation of VIR system.

invasive single port surgeries, which are generally divided into two components: 1) the radiation room, where the slave mechanism manages tasks, such as holding, twisting, and advancing the catheter, and 2) the control room, where the surgeon operates the master console. Despite their technological advancements, these systems face significant limitations, such as the lack of force feedback. Force feedback, also known as haptic feedback, is essential for surgical precision and safety as it provides sensory information about the forces applied to tissues. Without this feedback, surgeons cannot accurately gauge the resistance and texture of tissues, which leads to a loss of control over surgical instruments. This can result in unintentional damage to delicate tissues or blood vessels, as surgeons can apply excessive pressure, increasing the risk of complications such as vessel perforation or bleeding. Effective manipulation of guidewires and catheters often requires precise tactile adjustments, and the absence of force feedback impairs the ability to make these fine adjustments accurately, potentially compromising procedural success. Therefore, the lack of force feedback in master-slave robotic systems impacts hand-eye coordination and operational experience, which can diminish surgical efficiency and increase the risk of complications.

To address this issue, we have designed a VIR that incorporates a high-fidelity force feedback system, as shown in **Figure 2.7**. This system features isomorphic master and slave devices, each with 3-DoF navigation capabilities, enabling the robotic catheter to perform axial translation, rotary motion, and adjustment of interventional angles. Surgeons can remotely teleoperate the master console, avoiding direct X-ray exposure, while the slave robot operates beside the patient's bed to manoeuvre interventional tools along the vasculature. The VIR utilises a two-finger operation mode to simulate wire feeding and rotation, with the master-side device transmitting motion commands to the slave-side using analogous mechanical propulsion. This design retains the clinical manual interventional surgery manipulation model, which reduces the time required for interventional surgeons to master tool manipulation techniques and improves their proficiency in performing surgeries.

- ***The 3.5-generation robotic system***

Furthermore, an update force feedback system is designed to improve interventionalists' perceptual senses for better motor control. The control architecture proposed for the VIR is presented in **Figure 2.8**. It demonstrates that the VIR uses multiple layers for intravascular catheterization. The current version is designed for a shared control paradigm in which

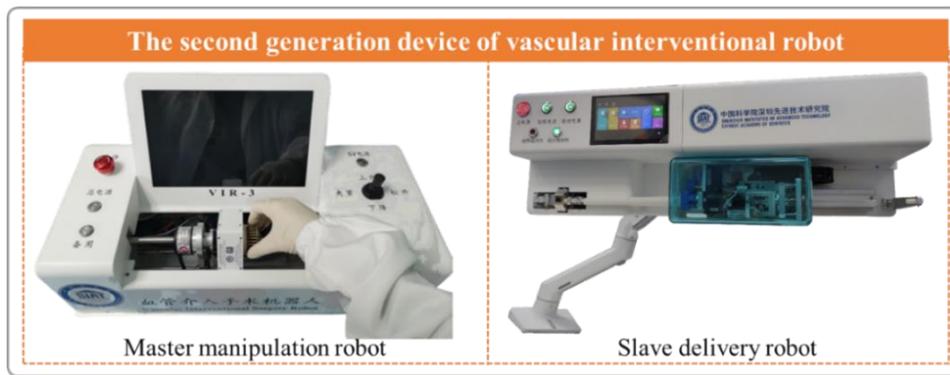


Figure 2.8: 3.5-generation of VIR system.

interventionists and the robotic system collaboratively manipulate the guidewire and catheter. The system includes a network communication based on the transmission control protocol (TCP), and a buffer of 1024 bits on each side of the TCP connection temporarily holds the incoming data. With multithreading coding, the VIR could perform the different commands and hand movements utilised for robotic-assisted catheterization independently.

The robotic system includes a force box used to measure the tactile force that interventionalists exerted with their fingers. This force is considered the interventionalist operative force on the guidewire or catheter in regular surgery. The force box includes a 32-prism, 32 tiny flexible strip sensors, PCBs, and batteries. A flexible sensor is folded and wound around the force box surface, and its data is logged as a 32-channel data multiplexed over an analogue multiplexer (Texas Instruments, Dallas, Texas, USA). Thus, interventionalists also operate the master device by manipulating the flexible sensor with their finger. Force data are processed in a STM32 microcontroller (STMicroelectronics, Geneva, Switzerland) and transmitted over a HC-04 Bluetooth module (HC Tech, Guangzhou, China) to the slave console in real time. The force box also reflects information on the rigidity of the flexible tool as it is held with a clamping mechanism on the slave manipulator. The guidewire is tightly clamped when the surgeon presses hard on the force box, and vice versa.

For measuring proximal force, an S-shaped force sensor is positioned at the rear of the clamping device to monitor the force exerted during tool handling. This sensor captures the force data proximally. The slave robot is designed to meet the requirements for precise hand motions needed in endovascular interventions, enabling actions such as pushing, pulling, and rotating the guidewire or catheter in clockwise or counterclockwise directions, allowing for tools delivery through blood vessels.

2.5.2 Validation of the robot-assisted interventional system

The performance of the master-slave vascular interventional robotic system encompasses the axial feed accuracy, the axial rotation error, the response time of the master-slave control system. To evaluate these aspects, we established a performance testing platform for vascular interventional surgery robots, as shown in **Figure 2.9(a)**, validating the developed system by measuring key

indicators such as axial feed accuracy, radial rotation error, and real-time performance of the master-slave operation. The overall verification process is illustrated in **Figure 2.9(b)**.

A. Axial feed accuracy verification

To accurately evaluate the axial feed performance of the slave controller in the vascular interventional robotic system, we systematically measured the actual axial displacement of the slave mechanism. This was achieved by entering displacement commands with various preset values into the slave propulsion mechanism and utilizing a high-precision calliper (Mitutoyo) as the measurement tool. The specific experimental steps are described subsequently.

First, the slave propulsion mechanism was positioned at a randomly selected initial location and the axial coordinates of this position were precisely recorded using a Vernier calliper as a reference.

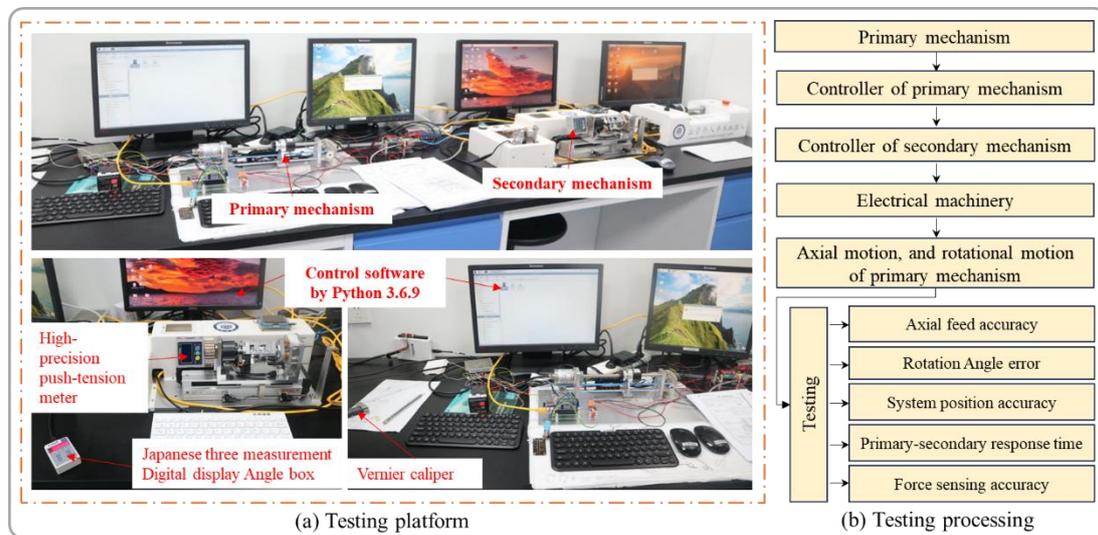


Figure 2.9: Testing platform and processing of interventional robotic system.

Table 2.5: Error measurements of axial feed.

Data	Test1	Test2	Test3	Test4	Test5	
Initial position (mm)	187.66	172.68	157.60	142.68	127.66	Average error (mm)
Set value (mm)	15	15	15	15	15	
Real position (mm)	172.69	157.65	142.65	127.62	112.66	
Real distance (mm)	14.97	15.03	14.95	14.06	15.00	
Error (mm)	-0.03	0.03	-0.05	0.06	0.00	

Table 2.6: Error measurement of twisting angle.

Data	Test1	Test2	Test3	Test4	Test5	
Initial position (°)	-40.84	-26.92	-11.56	3.34	18.3	Average error (°)
Set value (°)	15	15	15	15	15	
Real position (°)	-25.9	-11.64	3.34	18.3	33.52	
Real rotated angle (°)	14.94	15.26	14.9	14.96	15.22	
Error (°)	-0.06	0.26	-0.1	-0.04	0.22	

Next, a series of preset displacement setpoints were input into the slave controller via dedicated control software to simulate the feed commands used in surgical operations. After each command was input, the vernier calliper was employed to measure and record the actual axial displacement of the slave mechanism, ensuring real-time data accuracy. This process was repeated five times. After collecting all displacement data, we calculated the slave displacement error based on the differences between the preset and measured values. The results are presented in **Table 2.5**.

B. Twisting angle error evaluation

To accurately assess the control precision of the slave controller regarding the circumferential displacement of the rotating mechanism in the vascular interventional robotic system, we measured the actual angular displacement of the slave mechanism by inputting various preset angle values into the slave rotating mechanism, utilizing a high-precision digital angle display (Japanese three measurement digital display angle box) as the measurement tool. Prior to the experiment, the slave end rotation mechanism was calibrated to an initial horizontal position, which was recorded using the digital angle display as the starting reference point. Subsequently, the control software was used to input a series of preset angle setpoints to the slave controller, simulating the control instructions for the rotating mechanism during surgery. After each angle command was input, the digital angle display was used to measure and record the actual circumferential angular displacement of the slave end rotating mechanism, ensuring the immediacy and accuracy of the measurement data. Upon completion of the tests, the control software returned the rotating mechanism to the initial horizontal position, preparing it for subsequent tests. The circumferential displacement of each experiment was recorded, and the angular displacement error of the rotating mechanism was calculated based on the differences between the preset and measured values. The results are presented in **Table 2.6**.

C. Response time of the master-slave structure robotic system assessment

To accurately quantify the response time from the master to the slave in the vascular interventional robotic system (master-slave system), we employed a timestamp marking and comparative analysis method. First, the robotic system is initiated, and when the master is ready to send instructions, the current system time is captured as timestamp T, which is seamlessly integrated into the instruction packet. Subsequently, this instruction packet is sent to the slave for processing. When receiving the instruction containing timestamp T, the slave immediately begins to execute the required tasks. Once the task is completed, the slave captures the current system time again and calculates the time difference from the previously recorded timestamp T. This time difference, representing the interval between when the master sends instructions and when the slave completes execution, is defined as the master-slave response time of the prototype system, as shown in **Table 2.7**.

The results indicate that our self-developed master-slave homogeneous vascular interventional robotic system achieves an average axial delivery precision of approximately 0.034 mm. In the radial rotation dimension, the error is effectively maintained within approximately 0.136° , ensuring

the accuracy of the rotational movements. Additionally, the master-slave response time has been reduced to approximately 7.7956 s, providing surgeons with a more efficient and convenient surgical experience.

This section describes the specific mechanisms of robot-assisted systems designed for endovascular interventional surgery, including two prototypes of catheter-based robotic systems. Focusing on the precise manipulation of interventional systems for endovascular surgery, notable design benefits arise from the master-slave structure coupled with unique force perception devices that enhance spatial navigation of surgical tools and radiation sources. The spatial flexibility of endovascular tools depends on endovascular pathways and is further enhanced by control commands issued to the 2-DoF robotic prototypes. The following chapters propose various manipulation modelling based on multi-sensor data-driven and perception modelling of interventional instrument motion for upscaling intuitive manipulation on robot-assisted performance of interventional tasks. These models incorporate multi-sensor manipulation behaviour information and instrument perception visual information to address issues related to lack of force perception and hysteresis in surgeon-robot synergy, aiming for efficient and upscaling intuitive manipulation through anatomically flexible pathways.

Table 2.7: Response time of axial feed.

Time (s)	Test1	Test2	Test3	Test4	Test5	Average
Master	19:56:54.469	19:56:56.488	19:56:58.501	19:57:00.517	19:57:02.530	response time (s)
Slave	19:56:54.478	19:56:56.495	19:56:58.509	19:5:00.524	19:57:02.537	
Response	8.541	7.189	8.519	7.393	7.336	

2.6 Chapter Summary

In recent decades, the domain of VIR has seen significant progress, notably with the integration of advanced robotic systems that improve precision using superior manipulation and control technologies, force and haptic feedback, and the integration of AI. However, despite these advances, there are still several challenges that affect the wider clinical adoption of VIRs. It remains important to balance the improvement of capabilities of robotic systems through ensuring safe handling of endovascular interventional instruments.

One of the primary challenges in developing VIRs is the accurate and safe handling of endovascular interventional tools. This requires intuitive manipulation skills despite the absence of direct sensory force feedback to achieve the spatial precision necessary for intricate endovascular procedures. Unlike the more fixed systems used in traditional MIS, robot-assisted interventional tasks are especially vulnerable to perception errors in instrument motion, particularly for novice surgeons lacking effective haptic or force feedback. Surgeons depend on DSA imaging for guidance, which can impact precision due to overlapping tissue contours obscuring instrument paths, inadequate contrast between surgical tools and surrounding structures, and imaging artefacts like

noise and motion-induced distortions, which may affect the surgeon's ability to manipulate instruments.

These challenges are compounded by problems such as the use of undue force, the joint performance of both the surgeon and the robotic apparatus, and delays in communication between the master and slave elements of the robot during teleoperation. These factors together make achieving accurate manipulation of interventional instruments during procedures even more challenging. The modelling of technical behaviour-based manipulation and the improvement of perception for interventional instruments have emerged as effective methods for analysing and refining the safety and precise manipulation of these master-slave robotic systems. Therefore, ongoing research into intuitive manipulation modelling and scaling up intuitive manipulation with improved perception of the movement of interventional instruments is essential to advance the effectiveness of VIRs in carrying out precise and safe endovascular interventions.

Chapter 3: Modelling and Analysis of Manipulation Pattern

3.1 Introduction

This chapter focusses on modelling manipulation-based methods to improve the safety and precision of robotic-assisted endovascular interventional surgery. Controlled in vivo studies were conducted to capture four types of operator natural behaviour during 60 robotic catheterization trials, and activity signals were recorded to represent operator skills. A multilayer recognition model was developed to identify hand technical skill manipulation performed during the procedures. The model employs dense and convolutional layers to integrate features extracted from single and multiple data modalities. Beginning with an initial decision layer, the model is designed to train and classify catheterization trials recorded by nine interventionists as successful or unsuccessful. Subsequently, a motion decision layer was created to recognize the hand movements of interventionists using features from various data modalities. Finally, a mixed decision layer was integrated to identify manipulation patterns in successful and unsuccessful trials, analysing redundancy manipulation behaviour or features in robot-assisted performance of interventional tasks.

3.2 Related Work

In clinical trials, safety and efficacy evaluations of interventional robots have been reported. While it is possible to mitigate associated risks, the evaluation of the experience and skills of interventionists remains an understudied area of research [200]. While recent advancements in design have aimed to create more ergonomic master interfaces that replicate the natural motion patterns of interventionalists [201], several persistent issues remain. Robotic catheter systems operating under a master-slave configuration fail to replicate the natural manipulation skills observed in manual percutaneous interventions [202]. RCSs currently being developed in various laboratories often neglect the integration of natural skills [143, 201, 203]. These include limited flexibility in navigation and restricted tool manipulation ranges [143]. As a result, robotic systems struggle to consistently reproduce the natural motions of interventionalists, significantly hindering their practical application.

Moreover, there has been a growing interest in investigating the use of natural behaviours, specifically the motion signals generated by surgeons' manipulation behaviour during surgical procedures, within the medical field. Zhou Tao *et al.* [204] conducted an initial study analysing hand movements using surface EMG during PCI catheterization, showing that EMG signals can reveal distinct muscle activities linked to different hand movements. This study of hand and finger movement activities is reliable for intravascular catheterization. Stauder *et al.* [205] applied machine learning techniques to assess surgical workflow phases, utilizing random forests on tool usage data during laparoscopic cholecystectomy, achieving a recognition accuracy of 68.78%. Zhou *et al.* [159]

introduced a gaussian mixture model assessment model to evaluate operator skills. Rafii-Tari *et al.* [157], have designed frameworks aimed at automating skill evaluation in RCSs, employing operator motion and tactile signatures to validate assessment models. Previous research has demonstrated a correlation between a surgeon's hand movements and their level of surgical technical skill [206]. Thus far, the one or two types of surgeons' natural behaviours calculated during manual surgical procedure are mainly used for operators' natural learning and technical skills assessment.

Studies have shown that various types of surgeons' natural behaviours and features can more effectively characterize the manipulation patterns of endovascular catheterization. Zhou *et al.* [160] employed a hidden Markov model to extract a variety of features in the time and frequency domain, such as standard deviation, variance, and mean, from kinematic data that include muscle activity, hand movement, finger activity, and proximal force. They simulated the acquisition of operational skills by interventionists in traditional PCI using a endovascular simulation model, achieving an accuracy rate of 91.01%. Zhou *et al.* [170] introduced a data fusion framework that integrates hybrid motion and tactile behaviours to track guidewires during PCI, interpreting surgeons' natural behaviours through multimodal data sourced from electromagnetic sensors, sEMG, fibre optics, and acceleration data. Qualitative analyses of these multimodal signals have been carried out to understand the impact of surgeon manipulation behaviours on surgical outcomes. Wang *et al.* [180] proposed a deep learning-based multimodal information fusion architecture for recognizing eight common operating behaviours of endovascular interventionists, achieving an accuracy of 98.5%, which surpasses both existing machine learning classifiers and unimodal data. In particular, significant differences in surgeon natural hand behaviours have been observed at varying levels of technical skill in open surgery [207], laparoscopic surgery [208], and minimally invasive surgery assisted by robots [209]. Qualitative analyses of these multimodal signals obtained better results in manual interventional surgery, have been conducted to understand the relationship between surgeons' natural behaviours and manipulation technical skills. It is envisaged that modelling methods of such manipulation behaviour can be also applied to the manipulation pattern recognition of robot-assisted interventional procedure for skill learning, adaptation, and transfer application.

It is essential to consider not only the internal relationship between the natural behaviour of the operation and the surgical skill, but also the connection between operational behaviour and the outcome of the interventional procedure, particularly the natural behaviours or features that contribute to the success of the surgery. However, surgeons' manipulation behaviour is rarely involved in the master-slave robot-assisted interventional surgery, the behavioural analysis of interventionists' natural skills in robot-assisted endovascular interventions has yet to be fully integrated to ensure safe and efficient practices. Omisore *et al.* [210] proposed a novel deep integrated network model based on weighted parameters to learn the operational modes of interventional surgeons during traditional PCI and robot-assisted PCI procedures. The accuracy achieved in traditional PCI skill learning was 97.2%, while the accuracy for robot-assisted PCI skill learning was 47.8%. This disparity confirms the differences in skill transfer between traditional PCI

and robot-assisted PCI surgeries. The underlying reason is that the skills utilised by interventionists for tool delivery in robotic PCI typically differ; manual PCIs involve direct tool manipulation with the hand, whereas RCSs are employed for tool delivery in the latter. Therefore, new data-driven methods to learn manipulation skills in robot-assisted interventional surgery are essential to improve the precision of underactuated robotic systems.

3.3 Multi-Modal Sensing Data Pre-Processing

Using advanced sensors and data processing units, the system collects comprehensive and diverse control data from the intricate interaction cycle between the doctor, the vascular interventional robot, and the patient during robot-assisted endovascular procedures. These abundant data are then pre-processed, and features are extracted to facilitate further exploration of the relationships between successful and potentially failed interventions and the manipulation employed by the operators, as well as to investigate technologies to identify and categories the manipulative actions of interventional surgeons.

3.3.1 Integrated data acquisition and analysis platform for interventional surgery

A. The platform of multi-sensor data acquired interventional procedure

To thoroughly investigate the intuitive manipulation characteristic and automation processes of robot-assisted endovascular interventional surgery, a simulation platform integrated with a robotic system has been built. This platform accurately reproduces the clinical operating environment of endovascular interventional surgery using live animal models (rabbit and pig). It combines a precision manipulator system, advanced image navigation technology, and a control system to enable robot-assisted interventions to perform precise operations within small endovascular structures. By utilizing a real-time feedback and closed-loop adjustment mechanism based on DSA images, the system simulates the decision-making processes and operational behaviour of physicians in a real surgical environment. All experimental activities were conducted under a strict ethical framework. These in vivo studies were performed using the third-generation vascular interventional robotic system mentioned in **Chapter 2**. The robot-assisted catheterization trials involved navigating a 0.014" guidewire (Abbott Vascular, Diegem, Belgium) along endovascular pathways under X-ray guidance, which provides a means for CT image feedback. The experimental setup used for the mammalian subjects is presented in **Figure 3.1**.

The operators were thoroughly briefed on the catheterization trials and allowed to conduct them independently. Each trial commenced with a tutorial session in which operators were informed about the desired pathway. To ensure data consistency, the guidewire was prepared to maintain a similar initial pose at the start of each trial, allowing operators to initiate manipulation in a standardized mode. Kinematic motion data were obtained using EM trackers and an EM field generator placed in front of the master unit to ensure a stable measurement range. The sensors used for multimodal

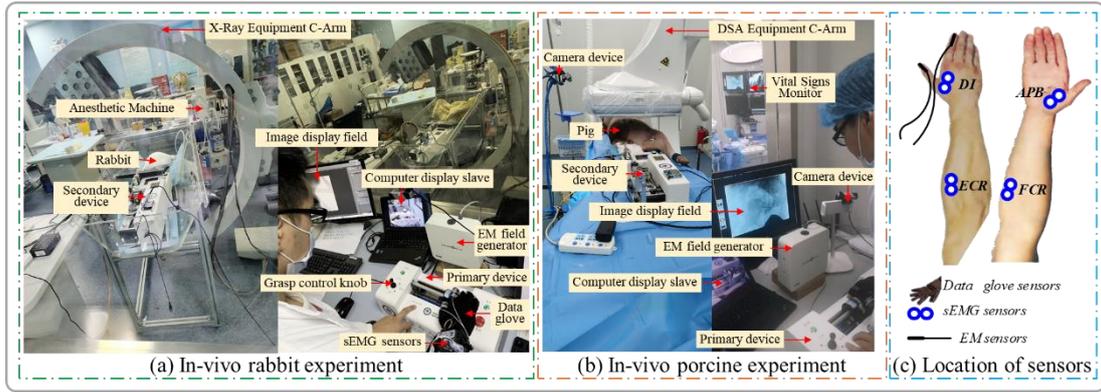


Figure 3.1: Experimental setup showing the original data collected procedure.

data acquisition is resistant to electromagnetic interference, ensuring no mutual interference among them. Each trial involved recording and deriving specific information from surface EMG data of pre-selected muscles, including EM signals related to the motion of the operator's thumb and index fingers, glove data for right-hand finger activities, and the position of the guidewire tip. The dataset recorded from the interventional procedure was labelled according to two predefined conditions: Level-S, which indicates successful path cannulation completed in less than the median time (82.5 seconds for rabbits and 92 seconds for pigs), and Level US, which includes unsuccessful trials that did not meet both conditions. The latter encompasses cases of cannulating incorrect pathways and instances of incomplete trials within the maximum allotted time. Overall, the acquired 26-channel signals include 4 channels for muscle activity (MA) EMG data, 6 channels for finger motion (FM) EM data, 14 channels for hand motion (HM) glove data, and 2 channels from CT images for guidewire trajectory (GT), all used to represent the operator's technical manipulation skills.

B. Manipulation pattern during endovascular interventional procedure

In manual PCI surgery, complex interventional manipulation involving the guidewire consists mainly of three types of manipulation: translational, twisting, and composite. Axial translational manipulation allows the interventional surgeon to push and retract the guidewire along its axis in two opposing directions. The aim of the pushing manipulation is to deliver the guidewire to the distal end of the lesion within the endovascular lumen. The surgeon pinches the guidewire between the thumb and index finger of their dominant hand and moves their entire hand to perform axial pushing manipulation (PH).

In contrast, when the guidewire reaches an incorrect vessel, the interventional surgeon must pull it (PL) in the opposite direction to return to a designated initial point, then reselect the path to deliver the guidewire to the correct endovascular branch. The coronary vascular system is intricate and has numerous branches. When the guidewire tip encounters an endovascular bifurcation, the surgeon must use two fingers to twist the guidewire, causing it to rotate circumferentially until its J-shaped tip is orientated toward the correct bifurcation. At this point, the surgeon pushes the guidewire to pass through the bifurcation and enter the correct endovascular branch. Similarly, the twisting manipulation encompasses two sub-manipulations: counterclockwise rotation (CCR) and clockwise

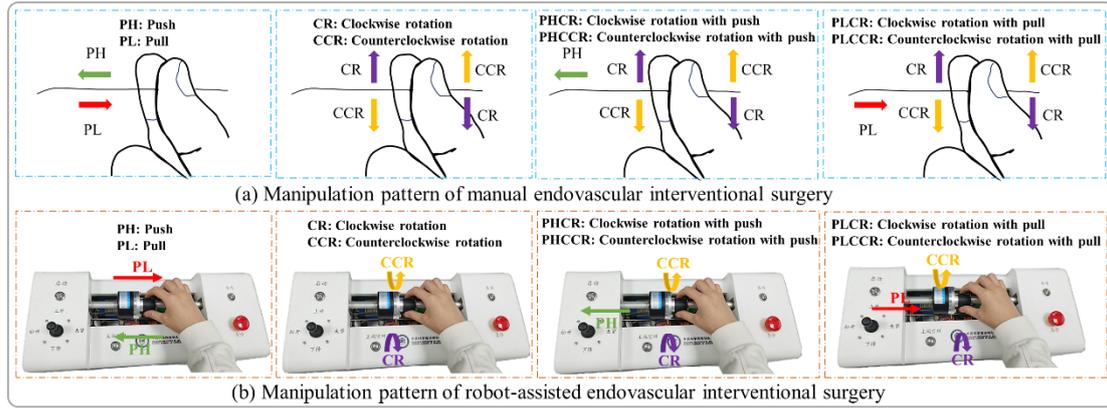


Figure 3.2: Manipulation pattern of endovascular interventional surgery.

rotation (CR). In certain narrow or calcified lesions, the flexibility of the guidewire tip can make it difficult for either of these single motion types to navigate through such lesions. Therefore, experienced surgeons often push and twist the guidewire simultaneously to dynamically adjust its movement state, generating motion both axially and circumferentially.

The composite manipulation of the guidewire includes four sub-manipulations: pushing combined counterclockwise rotation (PHCCR), pushing combined clockwise rotation (PHCR), pulling combined counterclockwise rotation (PLCCR), and pulling combined clockwise rotation (PLCR). Thus, there are eight operational modes in the endovascular interventional surgery process, which will aid future surgical robots in achieving automated standardized interventions. The self-developed master-slave vascular interventional robotic system preserves the operational paradigm of traditional interventional surgery to reduce the learning curve for interventional surgeons performing robot-assisted procedures. Thus, the robot-assisted interventional procedure also comprises eight sub-manipulations: PH, PL, CR, CCR, PHCR, PHCCR, PLCR, and PLCCR, as shown in **Figure 3.2**.

To acquire multimodal sensing information from surgeons with different technical skills, nine operators with varying PCI experience were recruited from Shenzhen Advanced Animal Study Service Centre and Shenzhen Institutes of advanced Technology. These include two veterinarians and seven novices with more than 50 and ~ 10 animal procedures completed, respectively. Subjects were introduced to the proposed study and trained on using the master-slave robotic system for endovascular tool navigation in an endovascular model. Training was done to prepare the operators. Similarly, subjects were introduced to the desired endovascular pathways in rabbits and pig animals a month before the acquisition of experimental data.

C. Endovascular cardiac pathways in mammals

The animal experiment was approved by the Shenzhen Institutes of Advanced Technology under Application No. SIAT-IACUC-200528-YGS-WL-A1289 and by the Shenzhen Advanced Animal Study Service Centre under application no. AAS191204P, which was carried out on two different mammalian animal models that are accessible. The subjects include six rabbits (2.21 ± 0.29 kg) and

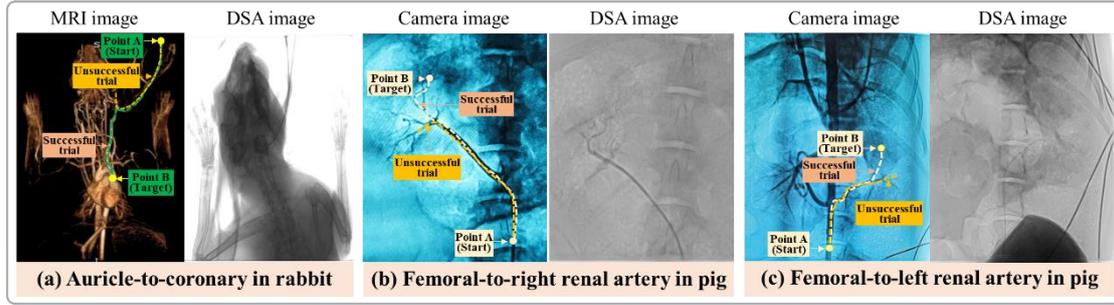


Figure 3.3: Three chosen catheterization routes and path navigation analyses.

a pig (32 kg) who were premedicated intravenously and intramuscularly, respectively.

The robotic catheterization trials started with path creation in which an indwelling needle was inserted through the auricle vessel in rabbits and the introducer sheath through the femoral vessel in pig. Sixty catheterization trials (that is, 26 in rabbits and 34 in pigs) were performed to catheterize the guidewire from the different start points A in the animals to the target points B, as shown in **Figure 3.3**. The three pathways are auricle to corona in rabbits (**Figure 3.3(a)**), femoral-to-right-renal artery and femoral-to-left-renal artery in pigs (**Figure 3.3(b)** and **Figure 3.3(c)**, respectively). Before a catheterization trial, operators were allowed to visually analyse the desired path to plan tool navigation.

3.3.2 Data acquisition and feature extraction

A. Multi-modal sensing data acquisition

The delivery of the guidewire in robot-assisted endovascular intervention is driven by the natural manipulations performed by the interventional surgeon on the primary mechanism. These natural manipulations result from coordinated movements of the hand, fingers, wrist, elbow, shoulder, and other joints. Consequently, the key factors that influence natural manipulation behaviour include muscle activity, hand movement, and finger movement. Therefore, data acquisition was achieved by recording operator manipulation behaviours through sEMG, fibre optic gloves, EM sensors, and CT imaging data. For labelling purposes, videos were recorded during 60 trials, and actual hand movements were extracted to create the labels necessary for training and validating the proposed recognition model. These offline analyses were used to label seven distinct hand movements. Each video was processed at 30 frames per second, and each frame was analysed to assess the catheterization motions performed by the operators and their progression along the desired path. The multi-modal data recording procedures are detailed in the following.

- **Acquisition of sEMG signals**

For the acquisition of muscle activity by interventional surgeons, four key skeletal muscles were selected for analysis: abductor pollicis brevis (APB), flexor carpi radialis (FCR), dorsal interossei (DI), and extensor carpi radialis (ECR). SEMG signals from these muscles were systematically collected and analysed using high-precision EMG sensors to fully characterize subtle changes and dynamic patterns of muscle activity during surgical procedures. Typical sEMG signals are captured

noninvasively to obtain physiological data that trigger limb movements corresponding to seven hand motions. A commercial configurable EMG system (BIOPAC Systems, Inc., Goleta, CA, USA) was utilised for the acquisition of the sEMG signal at a sampling rate of 1000 Hz. To ensure high signal quality, the operator's skin was thoroughly cleaned with alcohol prior to each trial. The centre-to-centre distance between each electrode in a channel was maintained at less than 20 mm. Each trial progressed from source A to target B, with the operator employing different hand motions (i.e. PL, PH, CR, CCR, PHCR, PHCCR and PLCR) for guidewire delivery. The signals related to each muscle motion were acquired and saved in separate files for further processing.

- ***Acquisition of EM signal of finger motion***

For the acquisition of finger motion, an electromagnetic tracking system was used to record the finger movements of the interventionist during the procedures. Two electromagnetic tracking sensors were affixed to the thumb and index fingertip, respectively, to capture movement behaviours throughout the intervention. EM position sensors, specifically the Aurora 6-DoF Flex Tube ($\Phi 1.3$ mm, 40 Hz) from Northern Digital Inc. (Canada), were employed to collect motion data from the operators' fingers during the trials. Each EM sensor records three-dimensional (x, y, z) position information and orientation data (pitch, yaw, roll) using a field generator with a measurement volume of $20 \times 20 \times 7$ cm. The EM field generator transmits electromagnetic signals to enable the EM sensors to function effectively. It provides position and orientation tracking accuracy of 0.9 mm and 0.8° , respectively, with approximately 20 ms latency for stable closed-loop control. The spatial pose of the surgeon's finger motion and distances are calculated based on the planar reflection of the coordinates and used for further processing.

- ***Acquisition of glove signal of hand motion***

For the acquisition of hand movement, a fibre optic glove is used to capture the complete hand motion and finger flexion data of the operators. During robot-assisted interventional surgery, the right hand is typically the dominant hand used by the surgeon. The Data Glove 14 Ultra (Fifth Dimension Technologies, Orlando, FL, USA) records the activities of the right fingers. This data glove is embedded with fiber bending sensors fixed at 14 joint points on each finger, including the proximal, middle, and distal joints. Signals are obtained from 14 channels corresponding to thumb near (tn), thumb far (tf), thumb/index (ti), index near (in), index far (if), index/middle (im), middle near (mn), middle far (mf), middle/ring (mr), ring near (rn), ring far (rf), ring/little (rl), little near (ln) and little far (lf), all at a fixed sampling rate of 60 Hz. The sensor processing unit processes this information to output the bending curvature of the hand. The glove captures kinematic data, providing flexion and abduction information between the fingers. These values are used as displacement information for hand motion and to model manipulation skills.

- ***Acquisition of CT Images of guidewire trajectory***

In robot-assisted interventional surgery, the dynamic mapping between the operator's manoeuvres and the movement trajectory of the distal endovascular device is a crucial dimension

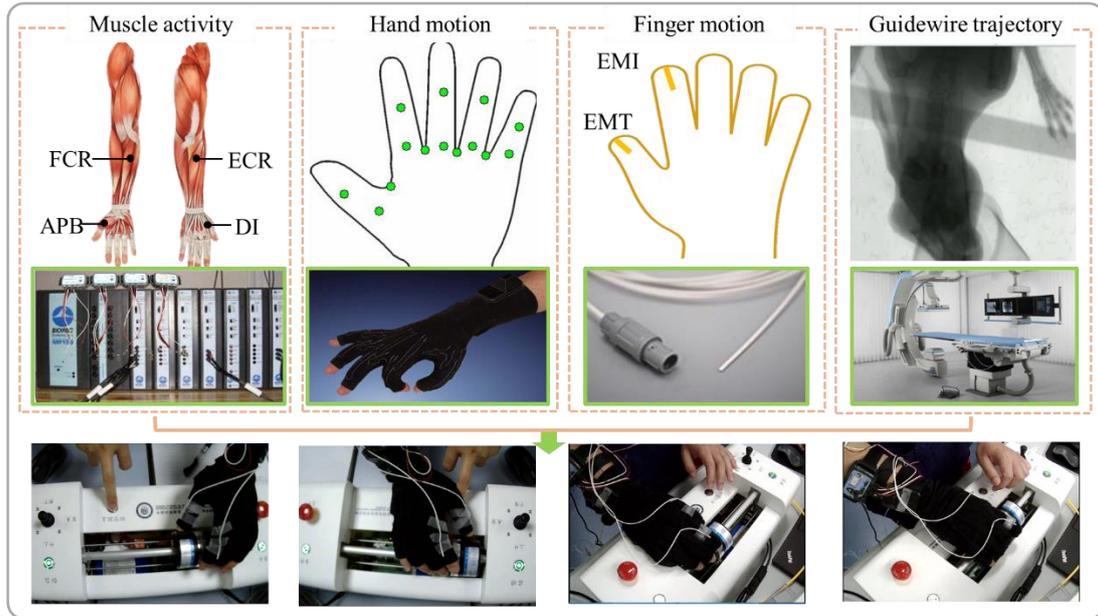


Figure 3.4: Multi-modal sensing information.

for assessing the operator's skill level. Each subtle action of the operator is directly and accurately reflected in the trajectory changes of the distal instrument within the blood vessel. This transformation process not only signifies the accuracy of the surgical operation but also provides insight into the operator's skill proficiency and decision-making abilities. Consequently, the motion trajectory data of the distal endovascular instrument serves as core information to quantitatively characterize the surgical skills of the operator. A custom X-ray machine [211] was used to produce a real-time image of the intravascular catheterization procedure. Continuous CT images tracked the motion of the guidewire in the endovascular pathways, and it was used to judge the success of the trials. Thus, guidewire pixels were segmented from CT images, and tip poses were obtained.

B. Signal processing

The study used a resolution-based dynamic time warping fusion model to align the multimodal data, given the diverse data sources used [212]. Each signal from the various sources was individually processed and the distinctive features within the signals were extracted. The sEMG signals acquired from the operators were subjected to a 10-500 Hz band-pass filter to remove unwanted components outside the specified bandwidth. Subsequently, a 50 Hz and 40 Hz notch filter was designed in the ACQ-Knowledge software to mitigate potential disturbances from power frequencies and EM sensor devices during signal acquisition. After noise reduction, a normalization procedure was implemented using the maximum voluntary contraction (MVC) method to ensure a consistent basis for comparing sEMG signals across operators, regardless of individual differences.

A CT image dataset was collected to capture the motion of the guidewire in vivo, the participants wearing sensors and performing a series of endovascular interventional surgery tasks. The data set consisted of trials recorded at 10 frames per second for rabbit experiments and 15 frames per second for pig experiments, resulting in approximately 1480 and 1545 images, respectively. To enhance

image quality, median filtering was applied to reduce nonlinear noise and diminish artefacts. Similarly, median filters were used to smooth signals from EM sensors and fibre optic gloves, eliminating spurious spikes and outliers. To synchronize the video, sEMG, EM, and fibre optic kinematics data, all four systems were connected to a computing device and time-stamped consistently throughout the experiments. The near instantaneous capture, processing, and time stamping of all incoming data streams ensured that no significant lag or temporal distortion occurred during the synchronization process.

C. Feature extraction

According to sensor specifications, the sampling rates for the muscle activity signal, EM data for finger motion, fibre optic glove data for hand motion, and CT images of guidewire motion are 1000 Hz, 40 Hz, 60 Hz and 10 frames per second, respectively. To synchronize the high-sampling-rate data with the lower-sampling-rate data, all signals were aligned in the time domain. A 0.1-second window length was used as the standard unit for each sequential signal to extract the features corresponding to the standard unit points. Data points were processed using various methods within this 0.1 second window to obtain feature values.

Following signal processing and normalization, the pre-processed EMG data were analysed using the root mean square method with a window length of 0.1 seconds, where each value was computed as the square root of 100 consecutive data points. Furthermore, the average rectified value and the zero-crossing rate, both with a window length of 0.1 seconds, were extracted to analyse the sEMG signals of each operator, facilitating the evaluation of muscle activity during robot-assisted endovascular interventions. Similarly, finger displacements were determined by averaging four consecutive data points, with each point representing the 3D coordinate distance between two consecutive finger motions recorded by the EM sensors. These values were calculated as the square root of the coordinates x , y , and z . The fibre optic glove sensors provided 14 channels of data corresponding to hand joint movements. The displacement values for each joint were obtained by calculating the absolute difference between two adjacent data points and averaging six consecutive values.

Thus, 14 types of joint displacements were derived from glove sensors. Finally, displacement data for the guidewire tip were extracted from CT images recorded during robot-assisted trials and added to the features used for outcome prediction and motion recognition. The feature extraction procedures are detailed below.

a) Manipulation feature based on muscle activity

- **Root mean square (RMS)** is used to evaluate muscle force during hand movements by measuring muscle activity. It is defined as the square root of the average sEMG signal acquired over a specific time, T . The RMS of a signal can be expressed as follows (3.1), where x_i represents the i -th data point of the signal, and N is the number of samples. In this case, N corresponds to the number of data points within a 100ms window.

$$RMS = \sqrt{\frac{\sum_{i=1}^N data[i]^2}{N}} \quad (3.1)$$

- **The average rectified value (ARV)** is used to assess the innervation input from the sEMG signals acquired from all activated muscles during a given movement. This is a parameter analysed in the time domain, and it is also a single-valued parameter that is not associated with time series of the sEMG signal. ARV is calculated as the mean amplitude of the sEMG signals over a 100ms window, as shown in equation (3.2).

$$ARV = \frac{\sum_{i=1}^N |data[i]|}{N} \quad (3.2)$$

- **Zero crossing rate (ZCR)** provides indirect information about the frequency of a signal in the time domain. It measures the number of zero crossings where the amplitude exceeds a set threshold to avoid low voltage fluctuations and background noise. In this study, the threshold (T) was set to three times the standard deviation of the background noise segment. ZCR reflects changes in the amplitude of sEMG signals recorded during muscular activities and is used for further analysis. This study investigated the speed of changes in muscle activities by analysing the ZCR of sEMG signals captured during different movements. The threshold (T) for this analysis was set at $10e-7$. A total of 12 features, including RMS, ARV, and ZCR from four muscles during muscle activity, were extracted for further analysis of the operators' technical manipulation skills based on single-modal MA data.

$$ZCR = \frac{1}{N-1} \sum_{i=0}^{N-1} sign(data[i] \times data[i+1]),$$

$$sign(x) = \begin{cases} 1 & \text{if } C_1 < 0 \wedge C_2 \geq d \\ 0 & \text{otherwise} \end{cases} \quad (3.3)$$

b) Manipulation feature based on finger motion

- **3D spatial displacements:** Operator's hand dexterity is assessed by calculating the spatial motion of the thumb and index fingers during each finger movement. This is determined as the Euclidean distance (3.4) between the two fingers using the 3D position data recorded by the EM sensors. Here, D_{jk} represents the displacement for any two consecutive finger motions; $D_{Tk}(j = th)$ corresponds to the thumb's displacement, and $D_{Fk}(j = ff)$ refers to the index finger's displacement. Similarly, $m\forall\{x, y, z\}$ denotes the displacement in different directions, which is computed for every $data_m$ obtained within a 100ms window. This results in six parameters of the two EM sensors. Consequently, the recognition of behaviour patterns in single-modal FM is based on the features extracted from the movements of the thumb and index finger.

$$D_{jk}^{hand} = \frac{\sum_{m\forall\{x,y,z\}} \sqrt{[data_m(j) - data_m(k)]^2}}{3} \quad (3.4)$$

c) Manipulation feature based on hand motion

- **Hand-tool displacement:** The interaction between the operator's hand and the lever of the master robot generates various types of displacement based on grasping and releasing motions. While some interruptions have limited effects on catheterization procedures, the

touch and frictional forces between the guidewire and the blood vessel can create significant resistance, prompting the operator to adjust their grasp or release the knob. To evaluate this in relation to manipulation skill, displacements from 14 joints are measured using a fibre optic glove sensor. Mean displacement values are calculated from data points within a 0.1-second window. The displacement values in this study are computed using equation (3.5), where $data_i$ represents the i -th data point in the sequence obtained from the fiber optic glove sensor, j is the j -th joint of the right hand, and N is the number of data points within a 100 ms window. This provides 14 displacement parameters from the fibre optic sensor, which are used to develop the recognition model.

$$D_{jk}^{knob} = \frac{\sum_{i=1}^N |data_i - data_j|}{N} \quad (3.5)$$

d) Manipulation feature based on guidewire trajectory

- **Guidewire trajectory:** The guidewire trajectory was obtained directly from CT images by calculating the coordinate values of the guidewire tip in each frame. The x and y coordinates of the CT images were used to determine the displacement of the guidewire, as shown in equation (3.6). The displacement values derived from the CT images were then used for further analysis.

$$D_g = \frac{\sum_{j \in \{x,y\}} \sqrt{(data_x(i+1) - data_x(i))^2}}{2} \quad (3.6)$$

A total of thirty-one features were extracted and used for outcome prediction and motion recognition. These features include twelve from the MA, two from the FM, fourteen from the HM, and three from the GT. For ease of reference, the features were defined as follows: AVR_{apb} , RMS_{apb} , ZCR_{apb} , AVR_{fcr} , RMS_{fcr} , ZCR_{fcr} , AVR_{di} , RMS_{di} , ZCR_{di} , AVR_{ecr} , RMS_{ecr} , ZCR_{ecr} , which were extracted from the muscle activity signals; D_{th} and D_{ff} from the finger motion; D_{tn} , D_{tf} , D_{ti} , D_{in} , D_{if} , D_{im} , D_{mn} , D_{mf} , D_{mr} , D_{rn} , D_{rf} , D_{rl} , D_{ln} , D_{lf} from the hand motion data; and D_x , D_y , D_g from the angiogram data of guidewire motion.

3.4 Multi-Layer Classification System

3.4.1 Manipulation technical skill framework

In the complex and demanding field of robot-assisted interventional surgery, the technical skill of the operator plays a pivotal role in determining the success of distal endovascular procedures. The quality of surgical outcomes is closely related to the operator's ability to master and apply various intricate operational modes during surgery. These skills are multidimensional and hierarchical, consisting not of isolated abilities but of interdependent and mutually influencing operational modes. These modes encompass fundamental hand-eye coordination, spatial perception, precise control of surgical instruments, real-time decision-making, and judgment during the surgical process, forming an intricate and sophisticated skill system. However, traditional approaches that

rely on the optimization of a single objective function are insufficient to capture and optimize all the essential elements of such a complex skill system. To address these challenges, this study introduces a novel hierarchical skill pattern learning framework driven by multimodal data, as shown in **Figure 3.5**.

The framework follows a structured four-step process. First, this framework integrates data from multiple sensors and sources, including visual feedback, mechanical sensor data, and physiological indicators, to construct a comprehensive and multi-layered skill learning system. In the second stage, advanced signal processing techniques are applied to pre-process and extract key features from multimodal data, revealing critical skill characteristics embedded within the data. Data recorded from operators during 60 trials are pre-processed through filtering and normalization to retain salient information in the multimodal signals. Feature extraction is conducted to enhance the performance of the multi-layer classifier. In the third stage, a three-layer recognition framework, consisting of an initial decision layer, a motion decision layer and a mixed decision layer, was implemented in the robot-assisted guidewire delivery system. The framework incorporates machine learning and deep learning algorithms, particularly those suited for handling complex non-linear relationships, to automatically identify and classify various operational skill patterns. This enables hierarchical modelling of skills, from lower-level concrete actions to higher-level abstractions, while also providing insights into the internal relationships and transformation rules among different skill patterns. The machine learning algorithms underpin a multi-layer classification system, which classifies skills at the skill level, motion level, and mixed level. The recognition results of these

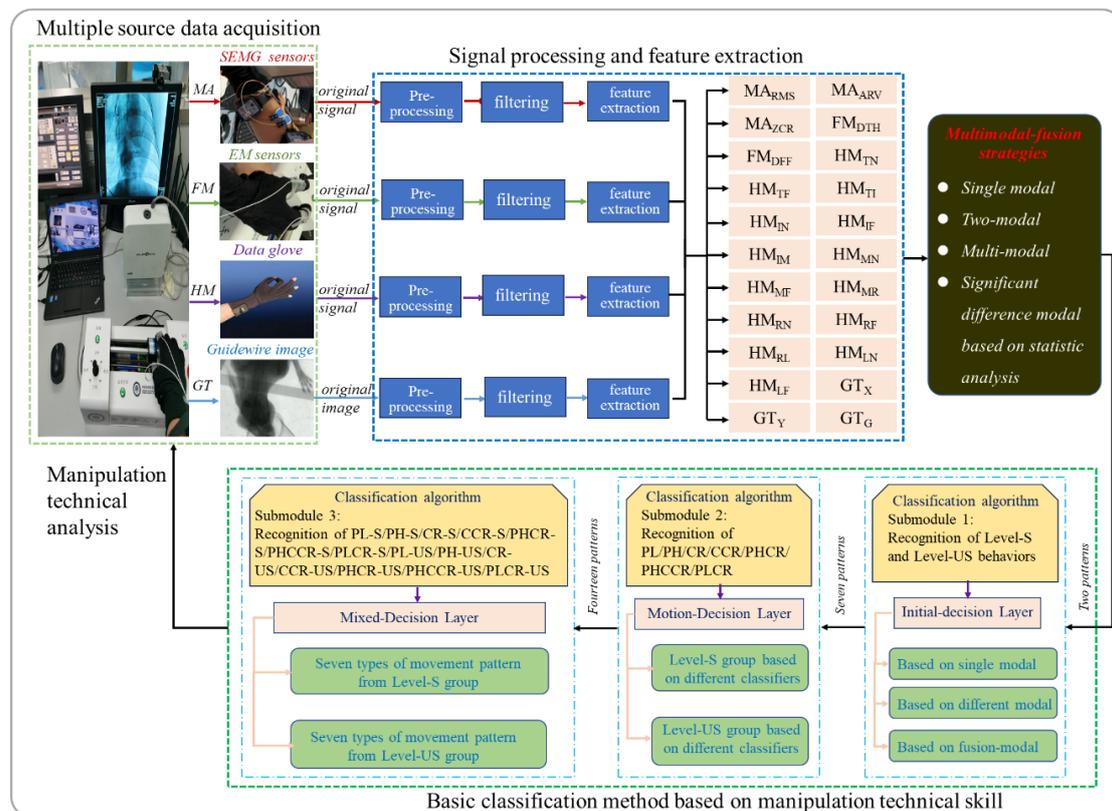


Figure 3.5: Recognition framework manipulation during robot-assisted interventional surgery.

layers, generated based on distinct objective functions, are compared. Finally, significant differences in operator manipulation skills across the 60 trials were analysed, providing deeper insights into skill variability and performance.

3.4.2 Manipulation technical modelling for control mode classification

Given the complexity and diversity of manipulation skills, four classical machine learning algorithms-KNN, SVM, RF, MLP, and HAR- are employed to investigate the accuracy and efficiency of skill learning by integrating various data types, including sensor and video data.

A. K-nearest neighbours (KNN)

Unlike training-based models, KNN classifiers keep all training examples in memory to search for the K nearest neighbours that match a sample. For classification based on imbalanced data, the KNN model seems to show good handling performance [213] Hence, it is significant to use KNN classifiers identify human activity for imbalanced data classification applications. The k-NN algorithm is a non-parametric, instance-based learning method, where the classification decision is based on the majority label of the nearest neighbours in the feature space. This model is particularly effective for rare event classification and multi-modal problems where objects may have multiple class labels.

- The k-NN algorithm classifies a test sample by finding the k -nearest training samples in the feature space. Most of the label among these k -neighbours is assigned to the test sample. The distance between samples is typically calculated using the Euclidean distance in equation 3.7. $d(x_i, x_j)$ is the distance between samples x_i and x_j , m is the number of features in the dataset.

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^m (x_{i,l} - x_{j,l})^2} \quad (3.7)$$

The main hyperparameter of the k-NN algorithm is the value of k , which represents the number of nearest neighbours considered during classification. A small k -value makes the model sensitive to noise, while a large k -value may include distant points that lead to incorrect classifications. Therefore, choosing an optimal k value is crucial to achieve high classification accuracy.

- To determine the optimal value of k , a cross-validation procedure is performed. The dataset is split into training and test sets, and cross-validation is used to evaluate different values of k in the training set. The accuracy is computed for each value of k and the value that maximizes the cross-validation accuracy is chosen as the optimal k . The cross-validation accuracy for each k is given in equation 3.8. n is the number of samples in the validation set, y_i is the true label, \hat{y}_i is the predicted label, Π is the indicator function that returns 1 if the prediction is correct and 0 otherwise.

$$Accuracy_{CV}(k) = \frac{1}{n} \sum_{i=1}^n \Pi(y_i = \hat{y}_i) \quad (3.8)$$

- Once the optimal k is determined, the model is trained on the full training set using this value

of k . The KNN model is trained using the training set with the optimal value of k . In the implementation, the optimal k was found to be 25, meaning the 25 nearest neighbours are considered for classifying a new sample. The model is then trained on the entire training dataset.

B. Support vector machine (SVM)

SVM is a powerful supervised learning algorithm widely used for classification tasks, particularly when data are not linearly separable, and is a supervised machine learning model that creates a maximum-margin hyperplane(s) to solve classification and prediction problems. SVM classifiers transform entities in the input space into different classes. SVM is a better performer in the recognition of manipulations and technical skills [214]; therefore, the SVM classifier was used to recognize operators' manipulation skills in the trials. In the context of control mode classification, SVM helps to find an optimal hyperplane that maximizes the margin between different control mode classes in the feature space. The SVM algorithm aims to find a hyperplane that best separates the data points of different classes. For multiclass classification, SVM uses a one-vs-one or one-vs-all approach, where multiple binary classifiers are trained to distinguish between pairs of classes.

- The optimization problem solved by SVM is as follows in equation 3.9 and 3.10. Where w is the weight vector defining the hyperplane, b is the bias term, C is the regularization parameter controlling the trade-off between maximizing the margin and minimizing classification errors, ξ_i are slack variables allowing for some misclassification of data points.

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (3.9)$$

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \quad (3.10)$$

- The kernel function used in the model is the Radial Basis Function (RBF) kernel, which is suitable for non-linear classification problems. The RBF kernel is defined as in equation 3.11. Where γ is a kernel hyperparameter that controls the influence of a single training example. The hyperparameters used in the model including Regularization parameter $C = 1.0$, Kernel function RBF, degree $d = 8$, Gamma $\gamma = 1/n_{feature}$, where $n_{feature}$ is the number of features.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (3.11)$$

- The SVM model is trained on the standardized training data and the corresponding labels. The kernel function of the SVM model learns the optimal hyperplane that separates the control mode classes. The trained model is then used to predict the labels for the test data, and the predicted labels are compared with the true labels to evaluate the performance of the model.

C. Random forest (RF)

The RF classifier is constructed by generating multiple decision trees during training and aggregating their predictions to enhance both the performance and robustness of the model. This

technique is particularly effective for classification tasks involving complex and high-dimensional data, such as control mode classification. The RF model operates by building a collection of decision trees, and the final prediction is obtained by combining the outputs of all the trees, typically through majority voting in classification tasks. The key advantage of RF is its ability to reduce overfitting while maintaining high performance by averaging the predictions of multiple weak learners.

The primary components of the RF model are as follows: a) decision trees, each tree is trained on a randomly selected subset of features and data points, using bootstrapping (sampling with replacement); b) aggregation, the model outputs the class most frequently predicted by the individual trees. The RF model mitigates overfitting by relying on the diversity of decision trees, each of which votes on how to classify instances of input data. Consequently, this RF model is adopted for the identification of operator hand patterns during robot-assisted guidewire delivery procedures.

The algorithm to build each decision tree in the RF model is described below.

- *Bootstrap Sampling*: For each tree, a random subset of data points is selected with replacement (bootstrapping). The size of this subset is controlled by the subsampling parameter, which is the proportion of data points used to train each tree.
- *Random Feature Selection*: A random subset of features is selected for splitting at each node of the decision tree. The size of this feature subset is controlled by the column sampling parameter, which is the number of features to consider for splitting at each node. This randomness reduces correlation between trees and improves generalisation.
- *Node Splitting*: At each node, the algorithm finds the best feature and corresponding split point that maximize the splitting criterion. In classification tasks, the Gini impurity is used as a splitting criterion. The Gini impurity is defined as in equation 3.12. Where C is the number of classes, and p_i is the proportion of samples that belong to class i at the node.

$$G_{ini} = 1 - \sum_{i=1}^C p_i^2 \quad (3.12)$$

- *Stopping Condition*: The tree is recursively grown by splitting nodes until the maximum tree depth, which increases model complexity, is reached. The number of samples at a node is smaller than the minimum samples for splitting parameter, which is the minimum number of samples required to split a node. The Gini impurity gain is less than a minimum threshold.
- *Aggregation*: When all trees are constructed, the final prediction for a sample is obtained by majority voting across all the trees in the forest.

D. Multi-layer perceptron (MLP)

The MLP is a neural network model utilized for recognizing operators' manipulation skills. As a standard feedforward artificial neural network, the MLP is valued for its non-linear mapping and generalisation capabilities. This makes it suitable for applications in signal processing domains. The control mode classification task is formulated as a supervised learning problem, with the MLP employed for classification due to its ability to effectively model nonlinear relationships in data.

The architecture of the MLP is as follows:

- The input layer processes a vector of 31 features ($m=31$).
- A fully connected dense layer with 100 neurons uses a *ReLU* activation function, defined as follow in equation 3.13. *ReLU* is chosen for its capacity to introduce non-linearity while remaining computationally efficient.

$$ReLU(x) = \max(0, x) \quad (3.13)$$

- The output layer consists of 2 neurons with a softmax activation function for multi-class classification. The softmax function, given by following in equation 3.14. Where Z_i is the input to the output neuron corresponding to class i , and C is the number of classes, ensures that the output is a probability distribution over the classes.

$$softmax(Z_i) = \frac{e^{Z_i}}{\sum_{j=1}^C e^{Z_j}} \quad (3.14)$$

- This model is trained using stochastic gradient descent (SGD) with momentum. The loss function is categorical cross-entropy, appropriate for multi-class classification problems. It is defined as follow in equation 3.15. Where, n is the number of samples, C is the number of classes, $y_{i,c}$ is a binary indicator (0 or 1) if class c is the correct label for sample i , and $\hat{y}_{i,c}$ is the predicted probability for class c for sample i

$$\mathcal{L} = -\sum_{i=1}^n \sum_{c=1}^C y_{i,c} \log(\hat{y}_{i,c}) \quad (3.15)$$

- The optimization parameters include a learning rate $lr = 0.01$, decay-rate $\lambda = 1 \times 10^{-6}$, and momentum $\mu = 0.9$. These parameters ensure a smooth and effective learning process, preventing the model from getting stuck in local minima.

E. Human activity recognition model (HAR)

This architecture leverages 1D convolutional neural networks to extract temporal features from input data, followed by fully connected layers for classification. This model aims to classify control modes by capturing complex temporal dependencies in input signals. The HAR model is made up of two main parts: the feature extractor and the classifier. The feature extractor is responsible for learning rich feature representations from the input signal data, while the classifier maps these features to specific control mode classes.

- The input data $x \in \mathbb{R}^{n \times m}$, where n is the batch size and m is the number of features, is passed through the convolutional layers to extract temporal features.
- The feature extraction part of the model is composed of a sequence of 1D convolutional layers. Each convolutional layer applies a set of filters to the input data, detecting local patterns and producing feature maps. This operation can be mathematically described as follow in equation 3.16. Where $f^{(l)}$ is the feature map output from the $l - th$ convolutional layer, $W^{(l)}$ represents the convolutional filter weights for the $l - th$ layer, $x^{(l)}$ is the input to the $l - th$ layer, $b^{(l)}$ is the bias term. σ is the activation function (*ReLU* in this case), each convolutional layer is followed by a *ReLU* activation function to introduce non-linearity. *

denotes the convolution operation. The three convolutional layers are parameterized, Layer 1 applies 64 filters of size 5 to the input, Layer 2 applies 64 filters of size 5 to the output of the first layer, Layer 3 applies 64 filters of size 5 to the output of the second layer.

$$f^{(l)} = \sigma(W^{(l)} * x^{(l)} + b^{(l)}) \quad (3.16)$$

- The model includes dropout layers to prevent overfitting and enhance the generalisation of the model. Dropout works by randomly deactivating a fraction of the neurons during training, thus improving the model's robustness.
- The classifier consists of two fully connected (Dense) layers, first fully connected layer receives the flattened feature vector and reduces its dimensionality to 128 using the following operation in equation 3.17. W_c are the weights and bias for the fully connected layer, f is the input feature vector, σ is the *ReLU* activation function.

$$h = \sigma(W_c f + b_c) \quad (3.17)$$

The second fully connected layer maps the 128-dimensional hidden representation to the final number of control mode classes. The softmax activation function is used to output a probability distribution over the classes, as defined in equation 3.18. Where \hat{y}_i is the predicted probability for class i , and Z_i is the unnormalized logit for class i .

$$\hat{y}_i = \frac{e^{Z_i}}{\sum_{j=1}^C e^{Z_j}} \quad (3.18)$$

The output of the convolutional layers is flattened into a vector of size 4032. The flattened vector is passed through the fully connected layers to produce the class probabilities.

- The model is trained using SGD with momentum, and using Cross-Entropy Loss as the loss function for classification. The update rule for each weight w is given in equation 3.19. Where η is the learning rate, $\nabla \mathcal{L}_t$ is the gradient of the loss function at time step t . μ is the momentum coefficient, which helps to smooth out the optimization process, allowing the model to converge faster and avoid getting trapped in local minima.

$$w_{t+1} = w_t - \eta \nabla \mathcal{L}_t + \mu(w_t - w_{t-1}) \quad (3.19)$$

3.4.3 Evaluative index of manipulation technical skill framework

A. Evaluation metrics

Data frames were collected from the robot-assisted interventional platform, with an average duration of 148 seconds per trial (26 trials in rabbits) and 103 seconds per trial (34 trials in pigs), using a window length of 0.1 seconds. This produced a total of 65,536 samples, which were used to develop a manipulation technique model for robot-assisted PCI. A key step in supervised machine learning is proper dataset division. Consequently, the sample datasets were randomly split into training and testing sets in a 7:3 ratio to ensure a well-generalized model with sufficient statistical power to accurately identify manipulation motions from new operators. The performance of the manipulation skill framework was evaluated using several metrics:

- **Accuracy** measures the overall proportion of correctly classified instances, including both true

positives (TP) and true negatives (TN), and is calculated as follows (equation 3.20):

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (3.20)$$

where TP represents the number of correctly predicted positive instances, TN is the number of correctly predicted negative instances, false positives (FP, Type I error or false alarm) occurs when a negative instance is incorrectly predicted as positive, and false negatives (FN, Type II error) occurs when a positive instance is incorrectly predicted as negative.

- **Precision**, positive predictive value measures the accuracy of the model's positive predictions. For a given class c , precision is defined as follows (equation 3.21). Where TP represents the number of true positives for class c , and FP is the number of false positives for class c .

$$Precision = \frac{TP}{TP+FP} \quad (3.21)$$

- **Recall**, or sensitivity, measures the model's ability to correctly identify positive instances. For class c , recall is given as follows (equation 3.22). Where FN is the number of false negatives for class c .

$$Recall = TPR = \frac{TP}{TP+FN} \quad (3.22)$$

- **Specificity**, Specificity (true negative rate, TNR) evaluates the proportion of actual negative instances correctly classified as negative:

$$TNR = \frac{TN}{TN+FP} \quad (3.23)$$

- **Negative predictive value (NPV)** measures the proportion of predicted negative instances that are true negatives:

$$NPV = 1 - \frac{FN}{FN+TN} = \frac{TN}{TN+FN} \quad (3.24)$$

- **False discovery rate (FDR)** quantifies the proportion of predicted positives that are actually false positives:

$$FDR = \frac{FP}{TP+FP} \quad (3.25)$$

- **Matthew's correlation coefficient (MCC)** offers a balanced evaluation of prediction performance, especially on imbalanced datasets, by considering true and false positives and negatives:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (3.26)$$

- **F1 Score** is the harmonic mean of precision and recall, offering a balanced measure of classifier performance for class c :

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall} \quad (3.27)$$

B. Evaluation & training strategy

To thoroughly explore the manipulation skills of interventional robots, several training strategies are employed to comprehensively evaluate and optimize their operational performance:

- **Single-modal training strategy:**

This approach uses a single type of training data, such as image data based on visual feedback or electromyography data based on muscle activity, to enhance robot manipulation ability. It aims to identify which characteristic provides the most effective representation to improve manipulation skills in specific scenarios of robot-assisted interventional surgery.

- **Strong representation mode training strategy:**

This strategy focusses on extracting highly representative features from training data by constructing complex feature extraction and representation models. It seeks to enhance the robot's understanding and adaptation of manipulation skills in the context of robot-assisted interventional surgery.

- **Significant difference mode training strategy:**

This approach uses statistical analysis to identify training modes with significant differences in robot manipulation performance under various conditions. Explore how significantly different characteristics represent interventional manipulation skills.

- **Multi-modal hybrid training strategy:**

This strategy integrates training data from multiple modalities (e.g., vision, touch, kinematics) to create a more comprehensive and diverse training environment. By leveraging the complementary nature of different modalities, it examines the representation capabilities of multi-modal features for robot-assisted interventional surgery, as well as the robustness and adaptability of the manipulation skill framework.

3.5 Model Evaluation and Analysis

Data on manipulation behaviours recorded during robot-assisted catheterization trials were analysed to assess technical proficiency and natural behaviours of the operators in relation to the results of the trials. Initially, significant differences in motion patterns between successful and unsuccessful trials were evaluated. The Mann-Whitney U-test, a nonparametric statistical method, was employed to analyse intergroup behaviour. This nonparametric approach was chosen to determine the influence of key features on outcome prediction between the two groups.

The interventionists' manipulation skills were analysed using multiple modalities, including MA features derived from sEMG signals, HM features obtained from the 15 sensors embedded in the fibre gloves, FM features from EM sensors, and the guidewire tip coordinates captured in CT images acquired via C-Arm X-ray during in vivo procedures. To identify the operators' natural behaviours and objectively classify the robot-assisted trials, we implemented a HAR model based on a CNN (HAR_CNN), which was applied to classify HM patterns.

1) Single-modal feature includes:

- Muscle activity feature: $AVR_{apb}, RMS_{apb}, ZCR_{apb}, AVR_{fcr}, RMS_{fcr}, ZCR_{fcr}, AVR_{di}, RMS_{di}, ZCR_{di}, AVR_{ecr}, RMS_{ecr}, ZCR_{ecr}$

- Hand motion feature: $D_{tn}, D_{tf}, D_{ti}, D_{in}, D_{if}, D_{im}, D_{mn}, D_{mf}, D_{mr}, D_{rn}, D_{rf}, D_{rl}, D_{ln}, D_{lf}$
- Finger motion feature: D_{th} and D_{ff}
- Guidewire trajectory feature: D_x, D_y, D_g

2) Fusion-modal feature includes:

- MA fused HM feature: $AVR_{apb}, RMS_{apb}, ZCR_{apb}, AVR_{fcr}, RMS_{fcr}, ZCR_{fcr}, AVR_{di}, RMS_{di}, ZCR_{di}, AVR_{ecr}, RMS_{ecr}, ZCR_{ecr}$ and $D_{tn}, D_{tf}, D_{ti}, D_{in}, D_{if}, D_{im}, D_{mn}, D_{mf}, D_{mr}, D_{rn}, D_{rf}, D_{rl}, D_{ln}, D_{lf}$
- MA fused FM feature: $AVR_{apb}, RMS_{apb}, ZCR_{apb}, AVR_{fcr}, RMS_{fcr}, ZCR_{fcr}, AVR_{di}, RMS_{di}, ZCR_{di}, AVR_{ecr}, RMS_{ecr}, ZCR_{ecr}, D_{th}$ and D_{ff}
- MA Fused GT feature: $AVR_{apb}, RMS_{apb}, ZCR_{apb}, AVR_{fcr}, RMS_{fcr}, ZCR_{fcr}, AVR_{di}, RMS_{di}, ZCR_{di}, AVR_{ecr}, RMS_{ecr}, ZCR_{ecr}, D_x, D_y, D_g$
- HM fused FM feature: $D_{tn}, D_{tf}, D_{ti}, D_{in}, D_{if}, D_{im}, D_{mn}, D_{mf}, D_{mr}, D_{rn}, D_{rf}, D_{rl}, D_{ln}, D_{lf}, D_{th}$ and D_{ff}
- HM fused GT feature: $D_{tn}, D_{tf}, D_{ti}, D_{in}, D_{if}, D_{im}, D_{mn}, D_{mf}, D_{mr}, D_{rn}, D_{rf}, D_{rl}, D_{ln}, D_{lf},$ and D_x, D_y, D_g
- FM fused GT feature: D_{th}, D_{ff} and D_x, D_y, D_g
- MA fused HM and GT feature: $AVR_{apb}, RMS_{apb}, ZCR_{apb}, AVR_{fcr}, RMS_{fcr}, ZCR_{fcr}, AVR_{di}, RMS_{di}, ZCR_{di}, AVR_{ecr}, RMS_{ecr}, ZCR_{ecr}$ and $D_{tn}, D_{tf}, D_{ti}, D_{in}, D_{if}, D_{im}, D_{mn}, D_{mf}, D_{mr}, D_{rn}, D_{rf}, D_{rl}, D_{ln}, D_{lf}$ and D_x, D_y, D_g
- MA fused HM and FM feature: $AVR_{apb}, RMS_{apb}, ZCR_{apb}, AVR_{fcr}, RMS_{fcr}, ZCR_{fcr}, AVR_{di}, RMS_{di}, ZCR_{di}, AVR_{ecr}, RMS_{ecr}, ZCR_{ecr}$ and $D_{tn}, D_{tf}, D_{ti}, D_{in}, D_{if}, D_{im}, D_{mn}, D_{mf}, D_{mr}, D_{rn}, D_{rf}, D_{rl}, D_{ln}, D_{lf}$ and D_{th}, D_{ff}
- HM fused GT and FM feature: $D_{tn}, D_{tf}, D_{ti}, D_{in}, D_{if}, D_{im}, D_{mn}, D_{mf}, D_{mr}, D_{rn}, D_{rf}, D_{rl}, D_{ln}, D_{lf}$ and $D_x, D_y, D_g,$ and D_{th}, D_{ff}
- MA fused HM, FM, and GT feature: $AVR_{apb}, RMS_{apb}, ZCR_{apb}, AVR_{fcr}, RMS_{fcr}, ZCR_{fcr}, AVR_{di}, RMS_{di}, ZCR_{di}, AVR_{ecr}, RMS_{ecr}, ZCR_{ecr}$ and $D_{tn}, D_{tf}, D_{ti}, D_{in}, D_{if}, D_{im}, D_{mn}, D_{mf}, D_{mr}, D_{rn}, D_{rf}, D_{rl}, D_{ln}, D_{lf},$ and $D_{th}, D_{ff},$ and D_x, D_y, D_g

3.5.1 Inter-group behaviour analysis in the initial-decision layer

A. Statistical analysis of inter-group trial outcome prediction

To investigate the differences in operator skill during robot-assisted interventional surgery, the Mann-Whitney U test, a non-parametric statistical hypothesis test, was employed to assess differences between two groups. This test is suitable for independent groups when it cannot be assumed that the population distribution follows a normal distribution. The U test was first applied to compare skill levels (Level-S or Level-US) in various motion cases in both successful and-

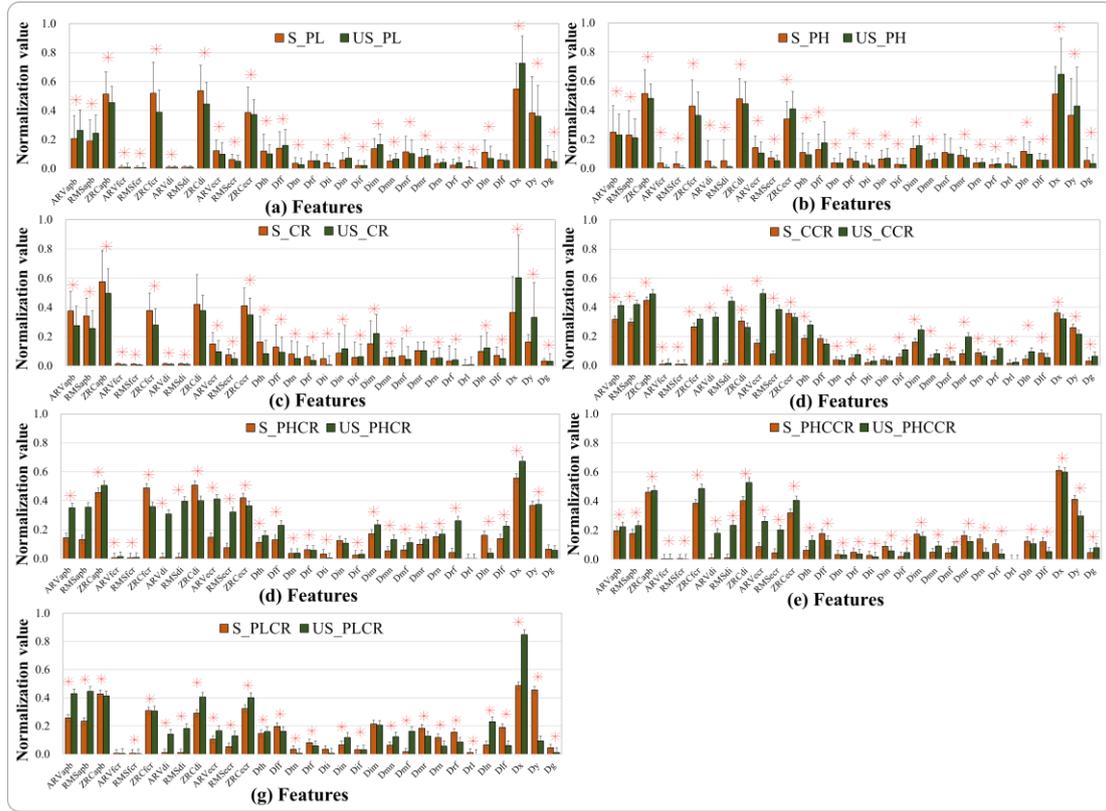


Figure 3.6: Statistical differences of features between successful trials and unsuccessful trials.

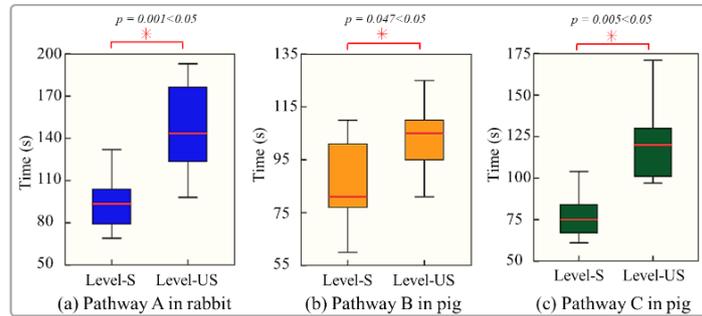


Figure 3.7: Distribution of operating time for two groups.

unsuccessful trials. The significant differences between S and US manipulations of all features are illustrated in **Figure 3.6**, with statistically significant behavioural differences indicated by an asterisk (*).

In PL operations, features have significant differences between S and US manipulations except for the RMS feature from the DI muscle and D_{tf} from hand motion. Similarly, in PH operations, aside from the D_{mf} feature, other features have significant differences between the S and US manipulation. For CR operations, the features have significant differences, except for the ZRC, D_{mr} , and D_{fl} features. In CCR operations, features also have significant differences between S and US manipulations, except for D_{th} , D_{in} , and D_{lf} features. Statistical differences in features were also evident between S and US manipulations in PHCR operations, except for the AVR and RMS features from ECR muscle activity and D_y from GT. Likewise, except for the D_{im} , D_{rl} , and D_g features, the features have significant differences between S and US manipulations in PHCCR

operations. Finally, all features have significant differences between S and US manipulations in PLCR operations.

Moreover, there is a close and intricate relationship between the operator's technical skill and the operating time in interventional surgery. Therefore, we analysed the differences in surgery time between surgeons in successful and unsuccessful trials. The results demonstrated that operators with Level-S completed guidewire delivery from point A to target B faster than those with Level-US, as shown in **Figure 3.7**. These findings suggest that the operator behaviour may vary when controlling the RCS to deliver the guidewire from the starting point to the target. This variation can be attributed to differences in experience, cognitive ability, and individual control strategies. Consequently,

Table 3.1: Performance for S vs.US classification.

Method	TPR	TNR	NPV	FDR	MCC	F1-score	Accuracy
KNN	0.9889	0.9772	0.6651	0.02173	0.9665	0.9836	0.9832
SVM	0.9838	0.9765	0.5799	0.0222	0.9605	0.9808	0.9802
RF	0.9911	0.9930	0.4264	0.0066	0.9840	0.9923	0.9920
MLP	0.9921	0.9969	0.2681	0.0029	0.9888	0.9946	0.9944
HAR_CNN	0.9617	0.9615	0.4868	0.0364	0.9231	0.9626	0.9616

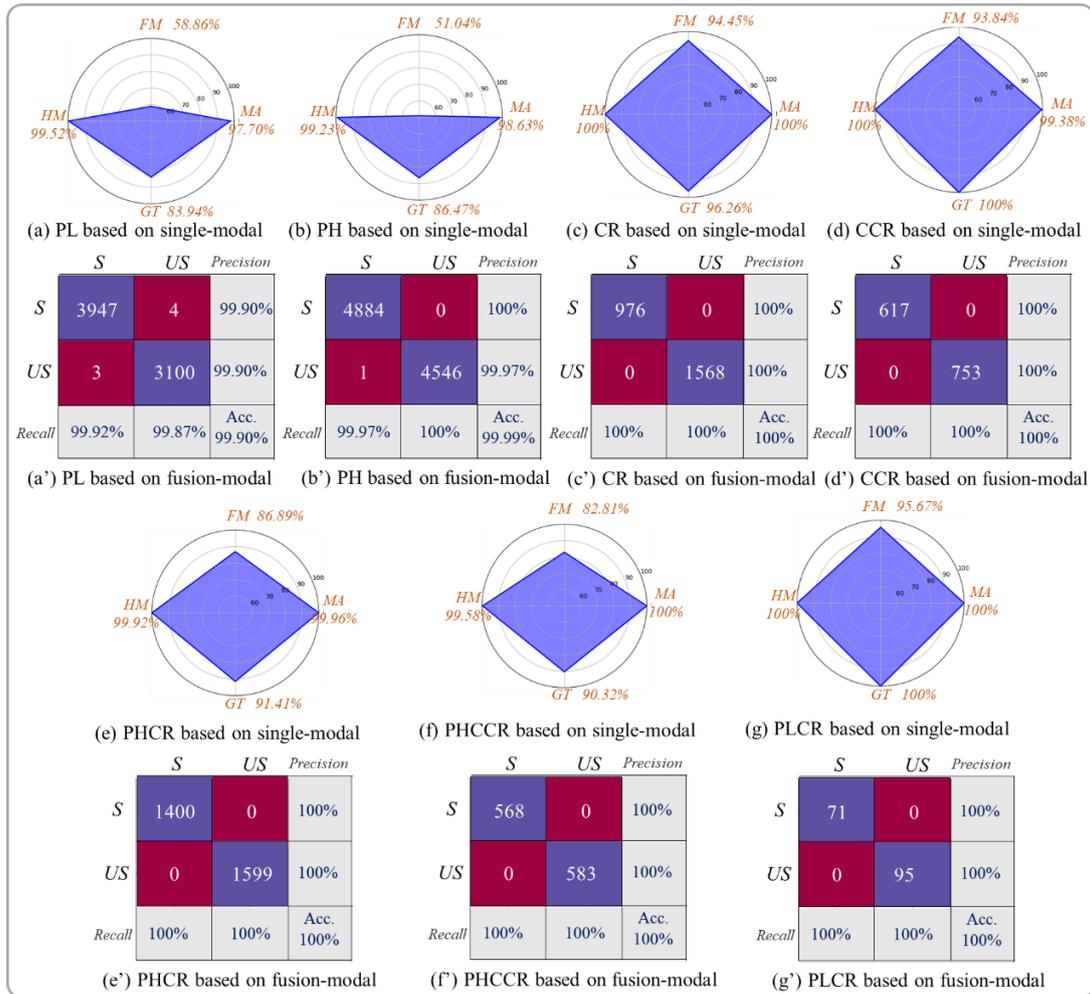


Figure 3.8: Performance of S/US based on single-modal and fusion-modal features with MLP classifier.

operators' behaviour strategies reflect varying levels of catheterization skills, which can result in successful or unsuccessful navigation of the endovascular pathway.

B. Performance analysis of Level-S vs. Level-US

The five classifiers were validated using the dataset obtained from the in vivo studies to assess their performance in predicting each trial as Level-S or Level-US based on thirty-one characteristics, as shown in **Table 3.1**. The results showed that the MLP demonstrated the best performance in recognizing the technical skills of Level-S and Level-US. The MLP method achieved the highest performance with a recognition accuracy of 99.44%, outperforming other classifiers.

Therefore, the MLP classifier was used to investigate the influence of relevant behaviours on skill level classification, distinguishing between Level-S and Level-US. The recognition accuracy based on the single-modal and fusion-modal characteristics was obtained using the MLP method, as shown in **Figure 3.8** (a)-(g) and (a')-(g'). The results show that the recognition accuracy using fusion-modal features outperformed that of single-modal features across different manipulation patterns. For the single manipulation patterns PL, PH, CR, and CCR, the single-modal feature of HM achieved the best recognition performance compared to other single modalities such as MA, GT, and FM. Similarly, for the composite manipulation patterns PHCR, PHCCR and PLCR, the single-modal feature of MA demonstrated the best performance compared to other modalities. Furthermore, the recognition accuracy of the single-modal FM feature was the lowest among the seven manipulation patterns compared to MA, HM, and GT.

Recognition performance for single-modality features revealed that MA and HM achieved better accuracy compared to FM and GT, and HM showed the best overall performance. To evaluate the impact of fusion-modal features on model accuracy, we explored the performance of bimodal and multimodal feature combinations. By combining the two best single modalities, MA and HM, as input parameters for the framework, recognition accuracy improved by 3.68% compared to using only the best single modality, HM, as shown in **Figure 3.9**. Combining HM and GT features yielded an accuracy of 97.10%, a 2.15% improvement over HM alone. Additionally, the combination of MA and GT resulted in a recognition accuracy of 96.82%, an increase of 1.87% compared to HM. However, combining FM with MA, HM, or GT led to lower accuracy compared to individual single

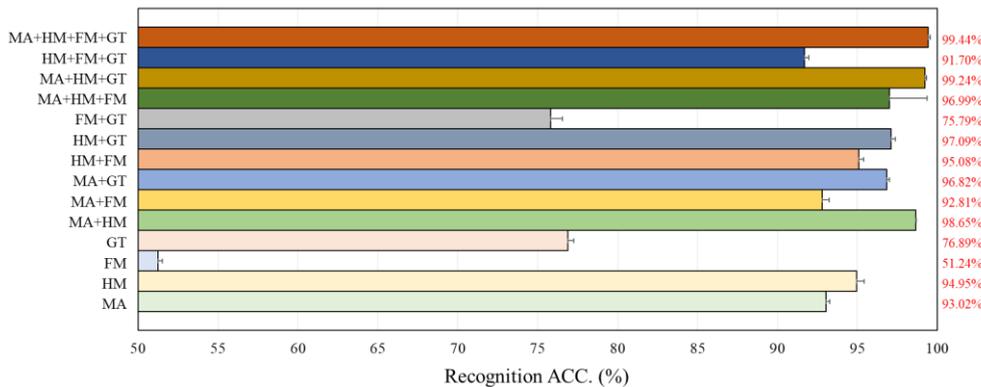


Figure 3.9: Performance of S/US based on different feature with MLP classifier.

modalities. Specifically, fusion of FM with MA resulted in a 1.65% decrease in accuracy compared to MA alone. Similarly, combining FM with HM and GT resulted in a 5.40% decrease in precision compared to HM combined with GT. These findings suggest that finger motion characteristics may not contribute significantly to the recognition of behaviour skills or may underperform in the recognition of behaviours during robot-assisted endovascular interventions.

The MLP method achieved the highest precision of 99.44% in recognizing Level-S and Level-US trials in the data set based on the MA, FM, HM, GM, and mixed-modal characteristics in the initial decision layer. Compared to single-modality schemes, multimodal behaviour showed improved performance, with 94.95% precision in distinguishing Level S from Level US trials. Similarly, natural behaviours based on a three-fusion model achieved 99.24% accuracy, which is higher than that of bimodal features. Furthermore, combining two modalities performed better than single-modal features, with an accuracy of 98.63% for the former. These results suggest that single-modality data has a lower recognition performance, whereas fusing multiple data modalities significantly improves recognition accuracy. Therefore, defining the appropriate number of combined features is essential for a better identification of operators' skill levels. Additionally, the single-modal HM features yielded the best performance, with recognition accuracy exceeding that of the MA, FM and GM modalities, indicating that hand motion information should be a primary feature in skill recognition. When merging two modalities, the best performance was achieved by combining features of muscle activity and hand motion (MA+HM), as the combined information improved the accuracy of the recognition model.

Moreover, without fusing FM features, the results show that the recognition accuracy based on bimodal features is higher than that based on single-modality features. Similarly, the accuracy based on three-modal fusion exceeds that of bimodal fusion, and the accuracy based on four-modal fusion is slightly higher than that of three-modal fusion. These findings indicate that incorporating features from multiple modalities provides richer insights into operator behaviours compared to single-modality data during robot-assisted catheterization trials and can better represent the operator's technical behaviour.

3.5.2 Inter-group natural pattern analysis in the motion-decision layer

A. Significant difference of features

To investigate whether features with significant differences impact recognition accuracy, statistical analyses were performed using Kruskal-Wallis and Mann-Whitney methods to assess the statistical differences between the two sample groups. The nonparametric test was used to compare features from seven behaviour patterns that exhibited significant differences. In *in vivo* studies, statistical analysis was conducted to explore the significant differences between PL, PH, CR, CCR, PHCR, PHCCR, and PLCR by evaluating 31 characteristics derived from MA, FM, HM and GM activities.

These results explain the influence of features with significant differences between movements

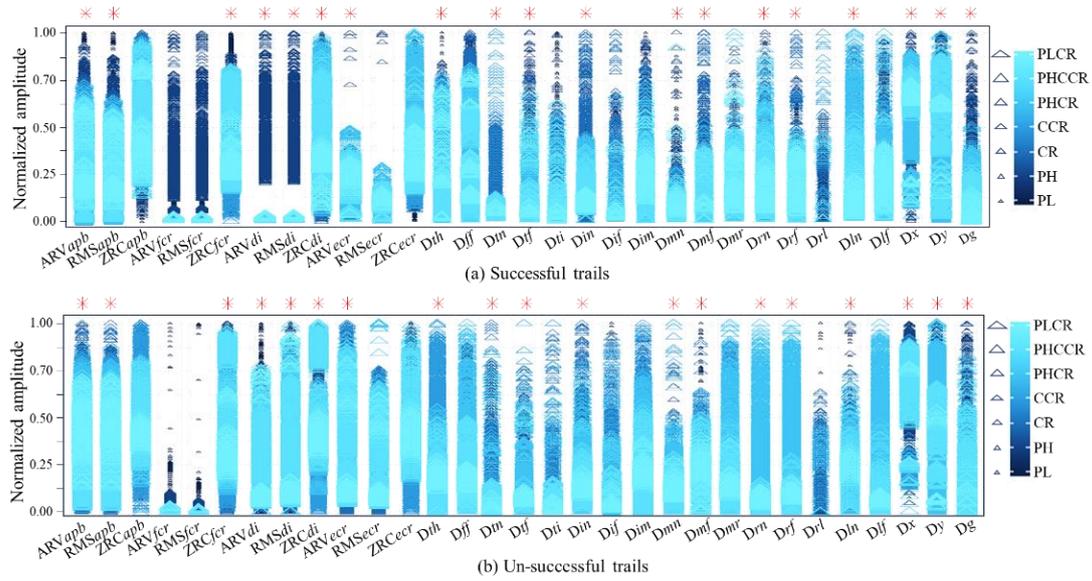


Figure 3.10: Distribution of statistical differences of features.

on the accuracy of the recognition of behaviour patterns. Pairwise comparisons among the seven movements yielded a total of twenty-one combinations. A feature was defined as efficient if it showed no significant difference in no more than two out of the 21 combinations. Thus, 19 features exhibited significant differences when comparing successful and unsuccessful trials. A significance level of $p < 0.05$ was used to identify these differences. The results are shown in **Figure 3.10**. In both successful and unsuccessful trials, the same 19 features with significant differences were identified: seven from muscle activities, one from finger motion, eight from hand motion, and three from guidewire motion.

B. Influence of different fusion-modal features

The recognition accuracy based on single-modal feature: The performance analysis of the classifier method based on single-modal features is analysed by recognizing seven manipulation patterns, as shown in **Figure 3.11**. The result show that the recognition accuracy based on HM-

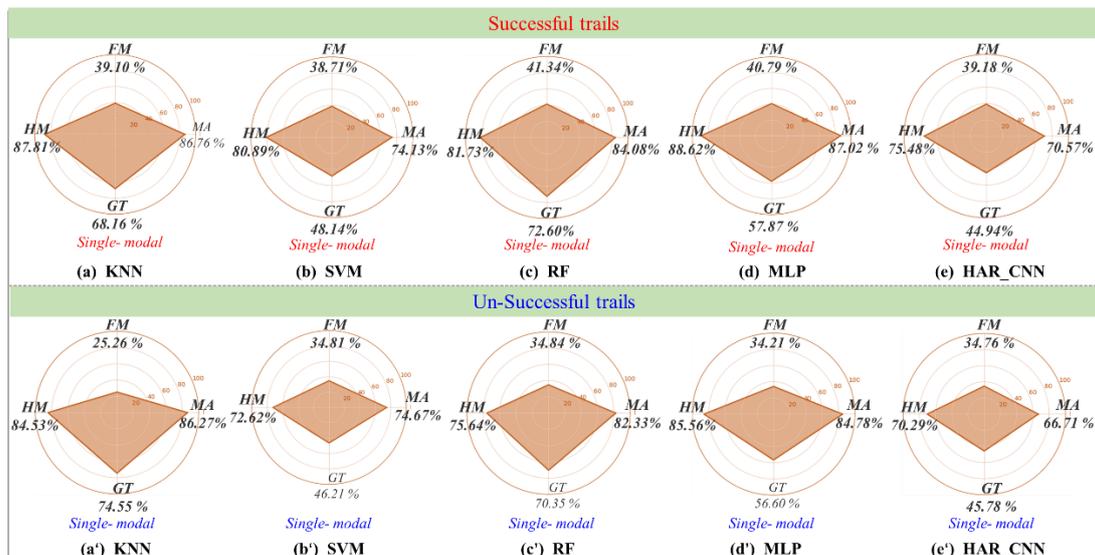


Figure 3.11: The recognition accuracy of manipulation patterns.

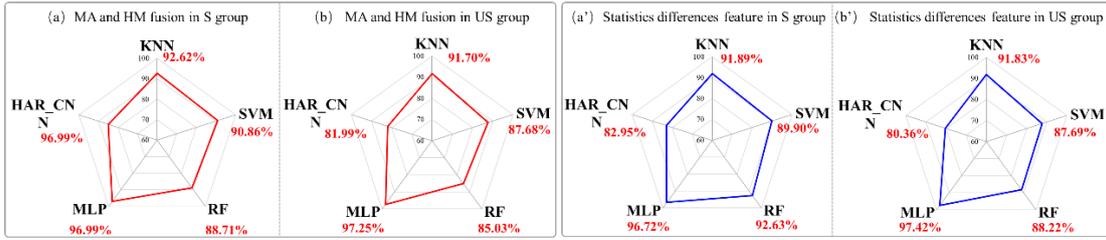


Figure 3.12: Performance of manipulation pattern based on fusion-modal features and statistical differences.

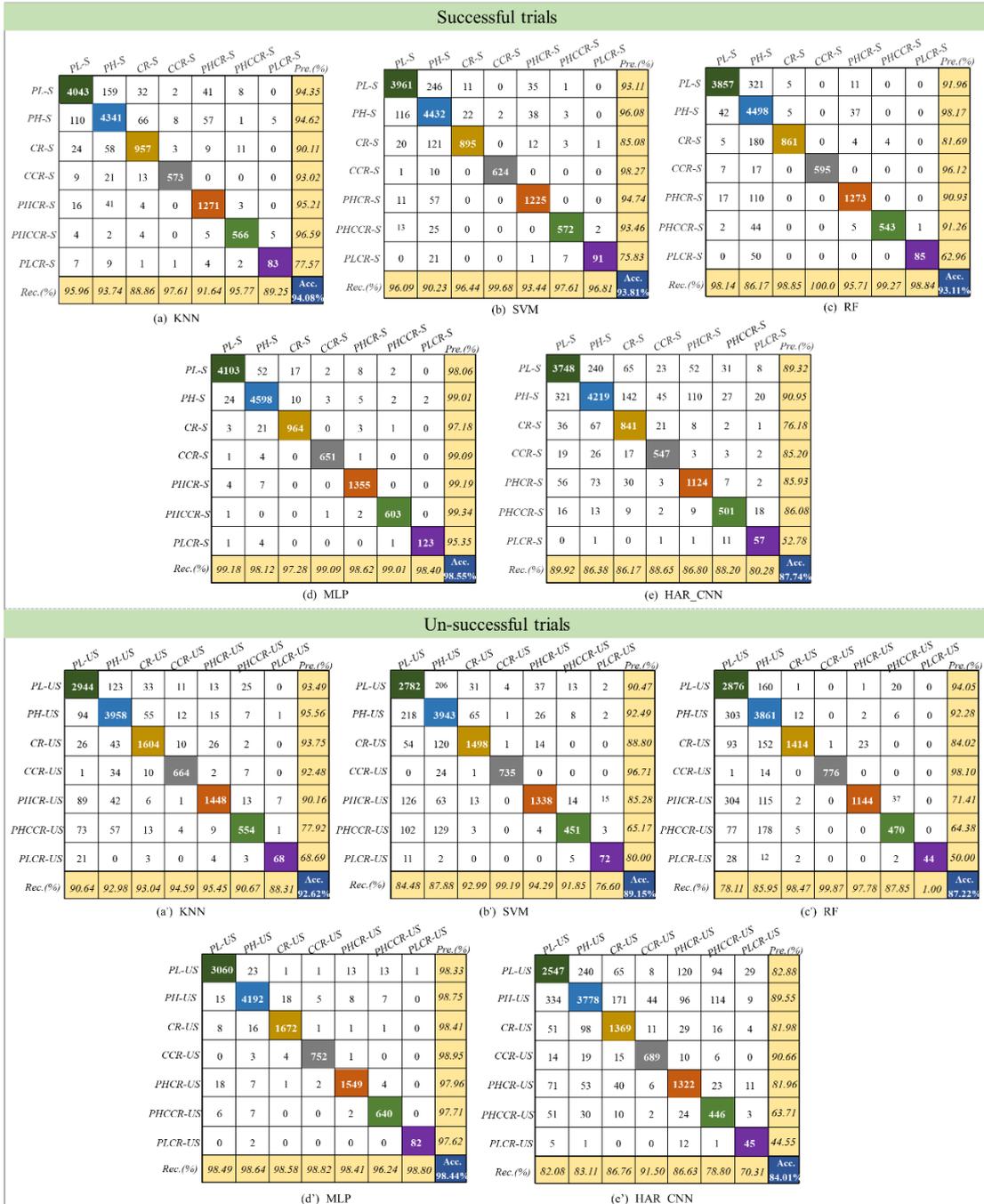


Figure 3.13: Performance of manipulation patterns based on thirty-one features.

feature has best performance than that of accuracy based on other single-modal feature in successful trails, except for RF method. Meanwhile, in unsuccessful trails, the recognition accuracy based on

MA feature has a better performance than that of accuracy based on other single-modal feature, except for MLP and HAR_CNN method. Moreover, the recognition accuracy based on the single-modal HM and MA feature is higher compared to that of single-modal FM and GT, respectively. Thus, the single-modal MA and HM feature was fused to improve the recognition accuracy performance. Fusion-modal features have 26 features, including 12 muscle activity features and 14 hand motion features.

The recognition accuracy based on fusion-modal features: The recognition accuracy based on single-modal MA features and HM features can obtain better performance compared to the single-modal FM feature and GT feature in the above discussion. Thus, using single-modal MA feature fused HM features, the accuracy achieved better performance on recognizing seven manipulation patterns, as shown in **Figure 3.12(a)** and (b). Furthermore, the recognition accuracy based on 19 features with statistical differences were obtained using different methods during robot-assisted interventional surgery, as shown in **Figure 3.12(a')** and (b').

The recognition accuracy based on all thirty-one features: The recognition accuracy of different classifier method based on all thirty-one features is shown using different classifier methods in **Figure 3.13**. The result showed that the MLP method obtained the best 98.55% performance to recognize seven manipulation patterns compared to the other four classifier methods in successful trials, meanwhile achieving 98.44% recognition accuracy in unsuccessful trials.

In surgical robotics, a precise representation of manipulation patterns is essential to improve accuracy and efficiency. Therefore, this study compares the impact of different fusion methods of characteristics on the effectiveness of manipulation patterns in both successful and unsuccessful trials, with the aim of identifying the most suitable fusion approach to optimize the performance of robotic-assisted surgical systems in **Figure 3.14**.

The results indicate that the recognition accuracy based on fused-modal features is significantly higher compared to that using single-modal features. In successful trials, a recognition accuracy of 82.95% was achieved using 19 statistically significant characteristics. When 26 characteristics derived from the two best single modalities, MA and HM, were used, the accuracy increased to

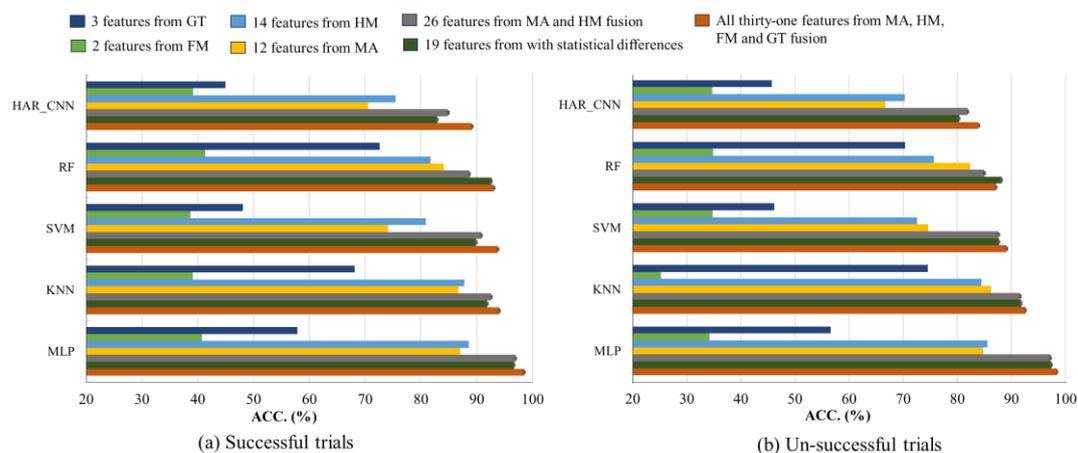


Figure 3.14: Performance of manipulation pattern based on different fusion-modal features.

84.90%. Using all thirty-one features, the HAR_CNN method demonstrated an increase of 6.24% and 4.29% in recognition accuracy compared to using 19 and 26 features, respectively. Similarly, the RF method resulted in an increase in precision of 0.48% and 4.40% with 31 features compared to 19 and 26 features, respectively. On the contrary, the SVM method showed an increase in precision of 3.91% and 2.95% when using 31 features compared to 19 and 26 features, respectively. Furthermore, the KNN method exhibited an increase of 2.19% and 1.46% in recognition accuracy with thirty-one features compared to 19 and 26 features, respectively. Finally, the MLP method showed a decrease in accuracy of 1.83% and 1.56% when using 19 and 26 features instead of thirty-one features, respectively.

In unsuccessful trials, the HAR_CNN classifier achieved a significant increase in precision with 31 features, improving by 3.65% and 2.02% compared to using 19 and 26 features, respectively. On the contrary, the RF method resulted in a 1.00% decrease in accuracy when using 19 features and a 2.19% increase when using 26 features. The SVM method saw increases of 1.46% and 1.47% in recognition accuracy with 31 features compared to 19 and 26 features, respectively. The KNN method exhibited increases of 0.79% and 0.92% with thirty-one features compared to 19 and 26 features, respectively. Lastly, the MLP method achieved higher recognition accuracy by 1.02% and 1.19% with thirty-one features compared to 19 and 26 features, respectively.

Consequently, during successful trials, the recognition accuracy based on 19 features was lower than that of 26 features, except when using the RF method (**Figure 3.14(a)**). In unsuccessful trials, the recognition accuracy based on 19 features was higher than that of 26 features, except when using the HAR_CNN method (**Figure 3.14(b)**). This suggests that statistically significant characteristics play a key role in the recognition of behaviour patterns during unsuccessful trials. However, this indicates that the features used as input to the classifiers significantly affect the recognition accuracies in successful trials and improve the recognition accuracies in unsuccessful trials.

Within the recognition framework, different feature sets were analysed to explore their influence on accuracy. In the Level-S group, feature sets comprising nineteen statistically significant features showed significantly lower recognition accuracy compared to feature sets with twenty-six features, which combine the best single-modalities MA and HM, across different classifier methods, except for the RF method. This indicates that selecting relevant muscle activity and hand motion features can potentially improve the overall recognition performance of the model. However, in the Level-US group, feature sets with 19 parameters generally exhibited higher accuracy than those with 26 parameters when using classifiers other than the HAR_CNN model. This suggests that behaviours with statistical differences contribute to improved recognition performance. Therefore, the features derived from the manipulating behaviours perform differently in the Level-S and Level-US groups. In addition, five methods were used to generate various fused feature sets from the thirty-one features.

Consequently, it indicated that the combined modality with more features from muscle activity, finger motion, hand motion, and guidewire motions have higher recognition accuracy compared to

when using a single modality with few features during robot delivering guidewire passed through endovascular path from A to target B. Thus, this suggests that more relevant behaviours are achieved from the four sources, including muscle activity, finger motion, hand motion, and guidewire motion comprehensive information on operators' behaviour activities when performing endovascular guidewire delivering for robot-assisted interventional surgery.

3.5.3 Level-S and Level-US pattern analysis in the mixed-decision layer

In the mixed-decision layer of the recognition framework, 14 behaviour patterns including PL,

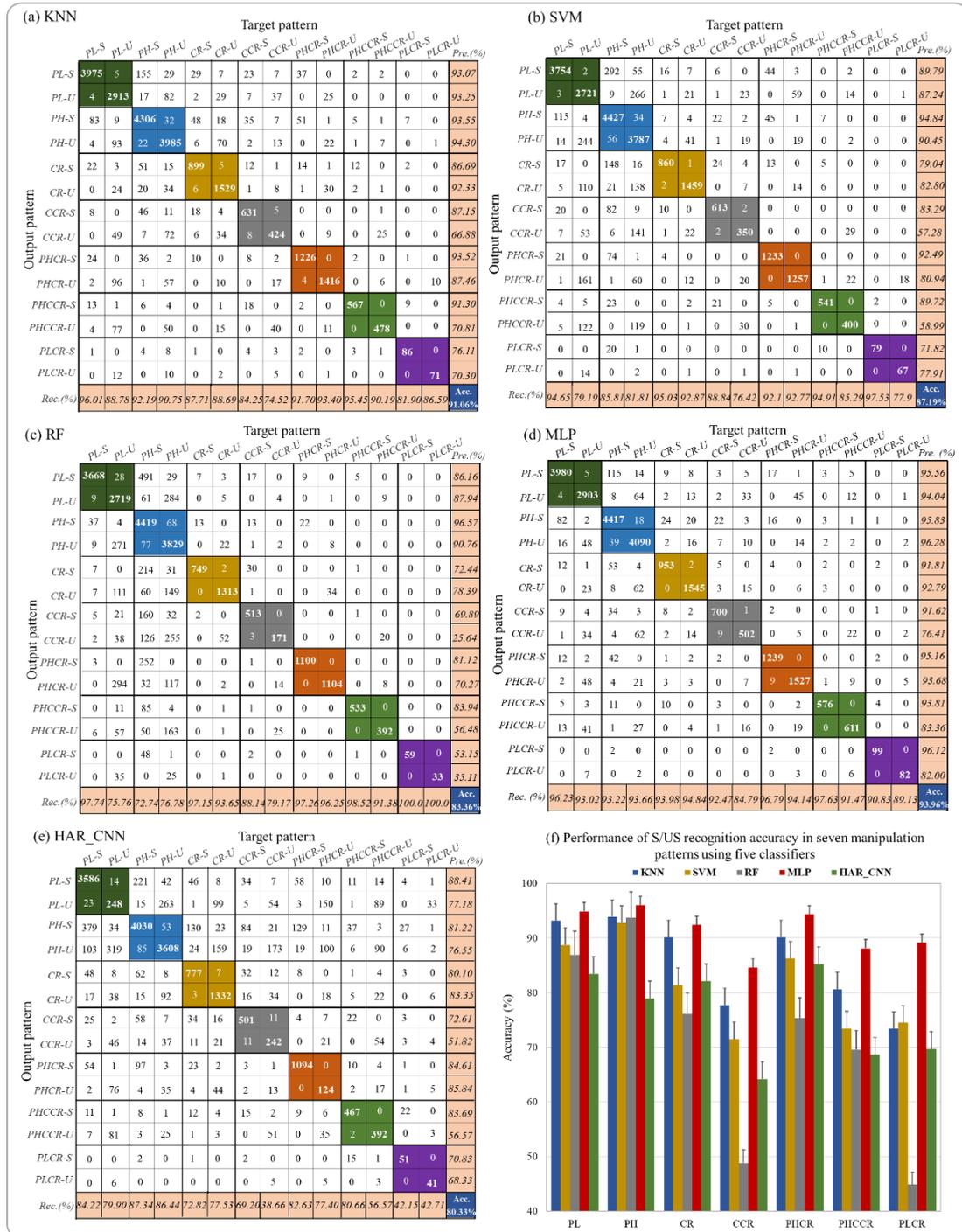


Figure 3.15: Performance of manipulation patterns based on different classifiers.

PH, CR, CCR, PHCR, PHCCR, and PLCR movements from the Level-S and Level-US groups were identified using different classifier methods. The results of each model were presented as a confusion matrix, representing the results of the recognition of manipulation behaviour, as shown in **Figure 3.15(a-e)**. The recognition accuracy of different candidate decoupling classifiers was 80.33% for HAR_CNN, 83.36% for RF, 87.19% for SVM, 91.06% for KNN, and 93.96% for MLP when using fusion-modal behaviours. When comparing the five classifiers, the MLP model achieved the highest classification performance with 31 features derived from MA, FM, HM, and GT.

The recognition performances for the motion patterns (i.e. PL, PH, CR, CCR, PHCR, PHCCR, and PLCR) are shown in **Figure 3.15(f)**. For the seven movements, the highest accuracy in recognizing Level-S vs. Level-US was achieved by the MLP method using fusion-modal features. Additionally, the Level-S vs. Level-US accuracy for PH and PL movements was higher than that for other movements, indicating that the operator's manipulation skill may have a stronger impact on these movements in robot-assisted PCI. Another reason may be that high-frequency movements are more suitable to assess operational skills. On the contrary, complex movements likely require additional information sources to accurately assess manipulation skills in robot-assisted PCI.

3.6 Chapter Summary

This chapter elaborates on a proposed operator-behaviour-based three-layer decision template designed to assess technical skills in robot-assisted PCI. Using this template, operator manipulation behaviours, including MA, FM, HM, and GM, are seamlessly integrated by analysing data from sEMG, EM, fibre optic glove sensors, and CT imaging. This integration enables the identification of the link between manipulation behaviours and guidewire delivery movements (that is, translation and rotation) within fusion-modal, dual-modal, and single-modal frameworks. Through fusion-modal analysis, the selection of suitable natural behaviours and the amalgamation of crucial information enhances the discernment of operator skill levels.

Comprehensive comparative experiments and statistical evaluations offer strong evidence that operators demonstrate distinct manipulation behaviours during successful compared to unsuccessful procedures while employing the robotic control system to navigate the wire to the target location.

Building upon the operator-behaviour-based decision template developed in this chapter, which quantitatively captures the relationship between natural manipulation behaviours and guidewire delivery performance, the next chapter further explores how these behaviours interact with the robotic system during the execution of interventional tasks. Specifically, **Chapter 4** investigates the synergy performance between the operator's manipulation strategies and the robotic system's response, aiming to evaluate the quality of human-robot collaboration under various manipulation conditions. By analysing cooperative performance indicators, such as operation delay, interaction force, and manipulation speed, the following chapter reveals how the effectiveness of intuitive manipulation influences overall procedural fluency, system responsiveness, and surgical safety in robot-assisted PCI.

Chapter 4: Synergy Performance Between Manipulation Behaviour and Robotic System

4.1 Introduction

This chapter introduces efficient interactive force analyses and a framework for synergy, designed to optimize collaboration between operators and robotic systems at different skill levels, thus enhancing collaborative catheterization in robot-assisted cardiovascular interventions. A manipulation-orientated evaluation framework is suggested to evaluate cooperative performance between operators and the robot during robotic interventional procedures using the robotic system platform described in **Chapter 2**. This framework combines various data sources, including kinaesthetic, kinematic, and haptic data from the surgeon's operating hand, along with motion and force data from the robot, gathered through multi-sensor technologies. The study also investigates interaction data, including distal force between tools and tissue and contact force from the hand control ring, to understand how operators with different technical abilities adapt their control strategies. This adaptation is intended to avoid endovascular injury from excessive force while providing sufficient tension to navigate complex pathways. Furthermore, the study examines the effect of delay factors on enhancing cooperative control strategies in master-slave isomorphic robotic systems.

4.2 Related Work

The master-slave structure of the vascular interventional robot enables precise navigation of catheters and guidewires through complex and branched pathways. Effective cooperation between the operator and the robot is essential for safe and accurate intraoperative navigation, ensuring skilled control of both the master and the slave robot components to maintain stability and prioritize patient safety. Typically, in traditional procedures, a surgeon manipulates instruments directly, such as catheters or guidewires. However, in a vascular interventional robotic system, the surgeon operates a master console that transmits control signals to the slave robot. The slave robot then grasps and directs the instrument, establishing cooperative dynamics, such as effective or inefficient master-slave synchronization, between the operator and the robot, which significantly influence the success of navigation. Consequently, designing an optimal control system for interventional endovascular robots requires understanding and incorporating manipulation techniques that create varied cooperative interactions between the operator and the robot.

Currently, there are limited models for synergistic performance analysis and interaction force measurement in vascular interventional robots. Existing studies have highlighted that imprecise motion control, communication delays, and excessive force application contribute to inefficient transmission of proximal to distal forces [215, 216]. Critical factors, including communication

delays, RCS dynamics, and operator manipulation skills, require further examination to improve performance in robot-assisted cardiovascular interventions. Although some studies explore delay factors in the context of haptic force, manipulation speed, and technical skills, research is still limited in this area. Smirnov *et al.* [217] proposed a cooperation index to coordinate human-machine intelligence in collaborative robot-assisted interventions, noting that network transmission time, packet rates, and bandwidth limitations often impact master-slave operation efficiency. Their findings suggested that teleoperation delays are unavoidable, recommending event-driven frameworks as more effective than time-driven approaches to ensure smooth master-slave operations in RCS systems [218, 219]. Xi *et al.* [220] also advocated for an event-driven approach to optimize RCS performance, while studies indicate that integrating event-based robotic intelligence with human cognition can enhance tool manipulation, potentially laying the groundwork for manipulation-driven systems in robot-assisted endovascular interventions [221].

Beyond communication delays, interaction forces, particularly excessive force application, play a significant role in cooperative performance between the operator and the robotic system. A major limitation in current master-slave systems is the lack of reliable haptic feedback, forcing surgeons to rely solely on visual cues. Without real-time tactile feedback, operators struggle to perceive distal and proximal forces accurately, leading to sudden force fluctuations, inconsistent speeds, or behavioural changes that vary by technical skill level. This lack of tactile feedback can hinder the responsiveness of the robotic system to surgeon commands, resulting in suboptimal cooperation. To address these issues, understanding the dynamics of surgical robots, endovascular tool tracking, and tool-vessel interactions, as well as the behaviours of interventionalists, is essential. Huang *et al.* [222] reviewed recent developments in modelling tool-tissue interactions, while Reiley *et al.* [223] suggested using tool-vessel interaction forces and motion data to assess catheterization skills. Zhou *et al.* [159] introduced a behaviour-based framework for skill assessment through proxies such as hand movement, proximal force, muscle activity, and finger motion. Du *et al.* [144] developed a model integrating kinematics, kinaesthetic feedback, and robot motion data to explore the transfer of surgical skills to robot-assisted PCI. Patel *et al.* [224] found that incorporating haptic and tactile feedback optimizes minimally invasive interventions. Gao *et al.* [225] and Gan *et al.* [226] advanced haptic feedback technology through magnetorheological fluid and fibre-optic sensors, respectively, allowing real-time measurement of distal forces at the catheter tip in minimally invasive procedures. Such innovations hold great potential for enhancing synergy between human operators and robots in robot-assisted PCI.

Selecting appropriate models for robotic intervention requires consideration not only of time-based communication delays and tool-tissue interaction frameworks but also of their applicability to specific surgical tasks. Modelling interactive forces and synergy performance through a manipulation-based evaluation framework remains underdeveloped. To address this gap, we propose a manipulation-based evaluation framework to improve cooperative performance between operators and master-slave robotic systems in robot-assisted interventions. This framework is

designed to assess cooperative performance during robotic-assisted cardiovascular interventions, contributing to advances in robotic-assisted surgical techniques and their clinical applications.

4.3 Methodology

4.3.1 System design

A. Experimental platform for multi-modal data acquisition

Multi-sensor technologies were employed to capture multimodal data during successful robot-assisted catheterization procedures. These data included video streams, time-series signals, robotic motion information, and interaction forces. These data were acquired using the master-slave isomorphic RCS platform developed for this study, as depicted in **Figure 4.1**. The RCS is capable of master-slave teleoperation and features multiple DoF for intravascular catheterization in emulated R-PCI. It includes a smart grasper that can change the orientation of clamped endovascular tools, such as guidewires and catheters, while sensing operational forces during R-PCI. Additionally, the RCS is equipped with potentiometer sensors for decoding control signals and transmitting them to the slave robot in real-time. This system implements all hand movements used by interventionists for tool navigation during endovascular interventions.

Compared to the third-generation RCS described in **Chapter 2**, the improved prototype logs both proximal and distal forces. A 32-channel flexible tactile sensor was designed and attached to acquire tactile forces between the interventionalist’s hand and the sleeve during procedures, as shown in **Figure 4.1** (B1). The prototype also includes a proximal sensor attached to the slave robot slider to measure frictional forces resulting from endovascular resistance, as illustrated in **Figure 4.1**(B2).

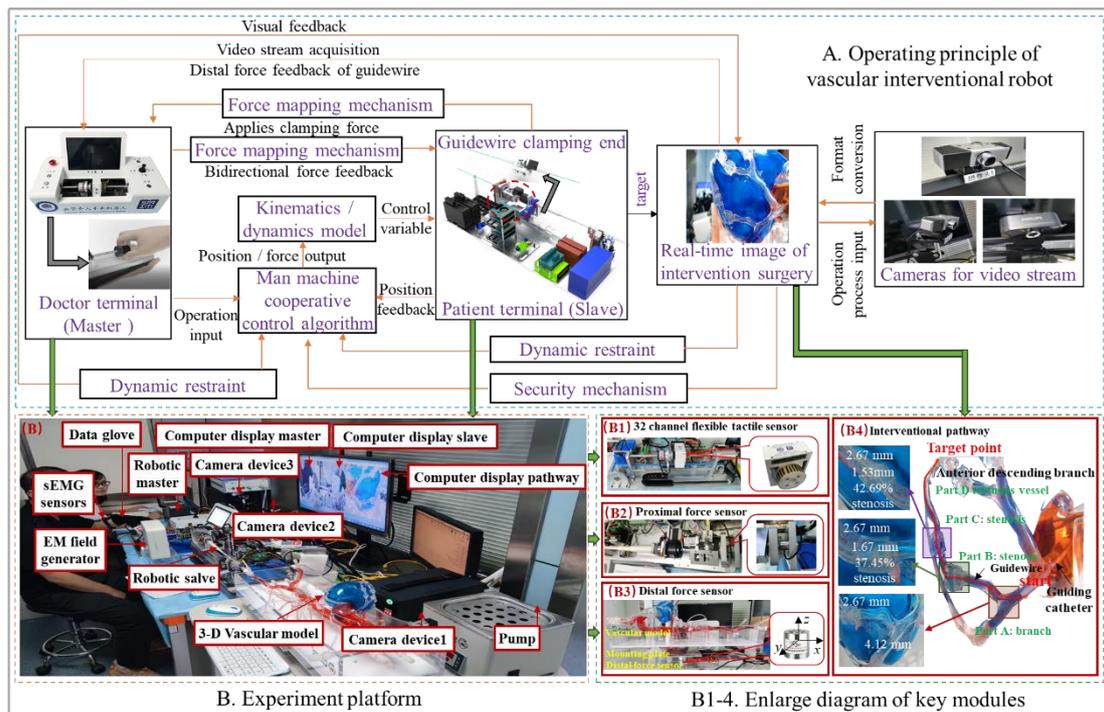


Figure 4.1: Experimental setup.

Additionally, a distal sensor was installed under the endovascular model to capture interaction forces between the instrument and the vessel or tissue, as depicted in **Figure 4.1(B3)**. This configuration facilitates the acquisition of multiple signals during R-PCIs, enabling an investigation of the synergy manipulation characteristics between the operator and the master-slave isomorphic robotic system.

To enhance the realism of the simulation of the endovascular environment, a 1:1 ratio adult endovascular simulator was used to replicate actual endovascular pathways. The steps followed in clinical practice were emulated by introducing a guidewire through the lumen of a guide catheter into the anterior descending branch of the phantom. The vessel path was filled with a blood-like fluid that flows in real time, assisted by a cycling pump connected to the endovascular phantom. Each procedure required subjects to catheterize the endovascular tool (guidewire) from a starting point to a target point.

The studies began with the creation of a path, during which a guide catheter was inserted through the femoral artery along the vessel to the coronary artery ostia in the simulator. The guidewire was then introduced via the coronary artery ostia along the anterior descending branch to the target point, as shown in **Figure 4.1(B4)**. This path consisted of one branch (Part A), two stenoses (Parts B and C, rated 37.45% and 42.69%, respectively), and tortuous routes (Part D).

B. Endovascular interventional manipulation behaviours Interpretation

A total of fourteen operators were recruited from the University of the Chinese Academy of Sciences Shenzhen Hospital and the Shenzhen Institutes of Advanced Technology in Shenzhen, China. During robot-assisted catheterization, the RCS facilitates the control of flexible endovascular tools (such as guidewires or catheters) with axial translational movement, radial rotational movement, and compound movement. Generally, the operator delivers the guidewire or catheter using nine hand motions to direct the RCS for guidewire delivery. These include pulling actions and axial retraction (that is, pull [PL] and push [PH]), radial rotation (clockwise rotation [CR] and counterclockwise rotation [CCR]), compound axial-radial motions (clockwise rotation with push [PHCR], clockwise rotation with pull [PLCR], counterclockwise rotation with push [PHCCR], counterclockwise rotation with pull [PLCCR]) and a static stage [SS], as presented in **Figure 4.2**.

When intravascular catheterization was performed in the simulator with the master-slave RCS depicted in **Figure 4.1**, the catheterization behaviours of the operators were obtained. In this study, a total of 168 robot-assisted trials were completed, which included exactly twelve trials per subject. In all cases, the operators were reformed about the chosen endovascular path through visual analysis of a pre-recorded video stream. Categorizing behaviours were characterized by hand kinaesthesia and kinematic data from operators during the completion of the robot-assisted task. To quantify the robot's contribution, low-order motion data (i.e., position and orientation) from the guidewire angiogram video stream, as well as the operational forces from the distal and proximal parts of the RCS and the haptic force received on the operator's side were also obtained. The surgeon-side and robot-side data was multiplexed to develop a manipulation-based evaluation model for the evaluation of human-machine synergy. Each session begins with a brief tutorial in

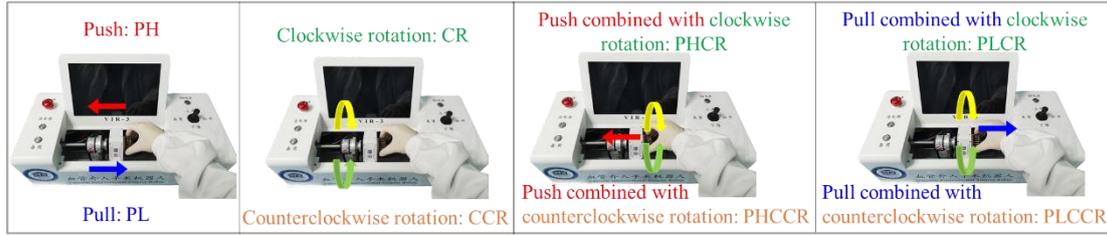


Figure 4.2: The nine manipulation patterns during robot-assisted interventional surgery.

which the operators are informed about the experimental procedure and provide a signed informed consent form. Operators were allowed to complete the task independently. For data consistency, the guidewire was prepared to start from a similar position during each experimental session. Consequently, the operators initiated the manipulation in a specific mode.

4.3.2 Signal acquisition and processing

A. Multi-sensors data acquisition

The dataset obtained from the 168 trials comprises 58 channels of signals, including 4 channels for electromyography data related to muscle activities, 14 channels of glove data for hand motion, 3 channels of distal force data for interaction information between endovascular tools and tissue, 1 channel of force data from the tool-robotic instrument, 32 channels of haptic force data for the hand-control ring, and 4 channels for position and rotation information. These various data sources are used to assess the synergistic characteristics between the operator and the robot. For further details regarding the devices used for each data source, please refer to **Table 4.1**.

- **SEMG signal acquisition**

Typically, sEMG signals are used to capture neuromuscular activity during movements of the upper and lower extremities in the human body. However, these conventional sEMG signals are designed primarily for the non-invasive acquisition of physiological signals that trigger movements of the limb. In our study, we specifically acquired sEMG signals to characterise the muscle activity of the interventionists' hand and arm. To achieve this, we used a commercial multi-channel wireless EMG system (BIOPAC Systems, Inc., Goleta, CA, USA) to capture the muscle activity signals from the operators' right hand and right arm at a sampling rate of 2000 Hz. Simultaneously acquired sEMG signals included those of the abductor pollicis brevis, flexor carpi radialis, dorsal interossei

Table 4.1: Summary of the acquired data.

Recorded Data	Source	# of Channels
Surface EMG	Pre-selected muscles	4
Glove data	Finger activities	14
Video stream	Guidewire tip motion, slave device Motion, and manipulation behaviours	3
Distal force	Tools-tissue contacting	3
Proximal force	Tool and instrument contacting	1
Contact force	Fingers with control ring contacting	32
Position and rotation data	Position and rotation value from master and slave robot	4
Multi-sensors data	Human-in the loop	61

and extensor carpi radialis muscles, as shown in **Figure 4.4** in **Chapter3**. To ensure high signal quality, the operator's skin was thoroughly cleaned with alcohol prior to each experimental session, and the centre-to-centre distance between the electrodes of each channel was maintained at less than 20 mm. In each experimental trial, the operator used the interventional robot to deliver the guidewire from point A to the target point B. Consequently, signals associated with each muscle movement were collected and saved separately for further processing.

- ***Glove signal acquisition***

During the intervention, the interventionists used axial translational movement, radial rotational movement, and compound movement to deliver the guidewire, resulting in distinct patterns of hand movement. To capture these data, a fibre optic glove, specifically the Data Glove 14 Ultra (Fifth Dimension Technologies, Orlando, FL, USA), was employed to obtain information on the bending of the operators' fingers. The hand motion data includes finger flexion and abduction. Thus, 14-channel signals were obtained from thumb near (tn), thumb far (tf), thumb/index (ti), index near (in), index far (if), index/middle (im), middle near (mn), middle far (mf), middle/ring (mr), ring near (rn), ring far (rf), ring/little (rl), little near (ln) and little far (lf) at a fixed rate of 60 Hz. The collected values provide displacement data for hand movements and are utilized to develop the evaluation framework for assessing human-machine synergy in robot-assisted endovascular interventions.

- ***Contact force data acquisition***

During manual PCI procedures, the interventionist exerts a direct contact force on the catheter and guidewire, relying on various visual cues and haptic feedback to adjust his manipulation strategy to safely reach the site of the injury. However, this dynamic shift occurs in robot-assisted PCI procedures, as robotic systems lack the capability to provide surgeons with sufficient haptic sensory data. To address this limitation, we designed a pressure measurement device featuring a multichannel flexible pressure sensor that is attached to the control ring of the master robotic device. The sensing unit incorporates a flexible sensor attached to the surface of the coaxial control ring, which the operator grasps while touching the flexible haptic sensor to complete guidewire delivery (see **Figure 4.3**); these were calibrated using an ESM303 force calibration platform (MARK-10 Inc., USA). The haptic sensor ranges from 0.01 N to 19.6 N. The 32-channel flexible pressure sensor utilized in this set-up has a sensitive area of 2 mm by 12 mm and a force range from 0.01 N to 19.6 N. This haptic force data is essential for understanding the forces applied by operators to guide the tip motion of the endovascular tool during robot-assisted catheterization and for evaluating the tool performance by both the surgeon and the robot.

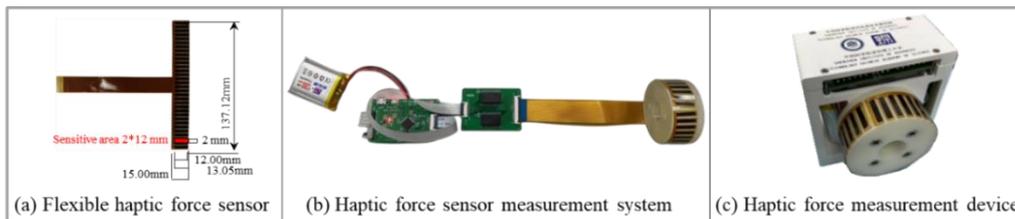


Figure 4.3: Designed 32 channel flexible haptic sensor system.

- ***Proximal robot-tool operational force data acquisition***

The proximal force comprises the interaction force between the tool and the tissue, together with the friction force between the robotic device and the tool during robot-assisted intervention procedures. This proximal force serves as an indirect measurement method for assessing the interaction force between tools and tissue within the blood vessel. To achieve this, we developed a point-contact proximal sensing platform to obtain information on the proximal force, as illustrated in **Figure 4.1(B2)**. The maximum measurement range of the proximal force sensor (SBT674-2kg, SIMBATOUCHE) is 19.6 N, with a sensitivity of $1.0 \pm 10\%$ mV/V, and the sensor provides force readings in z-axis direction. Once the flexible endovascular tool (guidewire) is secured within the slave robot and begins to move forward along the endovascular path, the force sensor captures the operational forces in real-time.

- ***Distal tool-tissue operational force data acquisition***

In typical endovascular interventional procedures, 2D real-time fluoroscopy serves as the primary method of visual guidance in endovascular interventions. However, in instances where surgeons lack 3D anatomical information regarding blood vessels, force feedback becomes crucial for interventionists. This is particularly pertinent for remote vascular intervention robots, where the interaction force between instruments and tissues remains imperceptible, thus increasing the risk of thrombosis and endovascular perforation due to excessive force during surgery [227]. To directly measure the contact force between the instrument and the tissue during the delivery of the guidewire, we designed a distal force sensing platform. This platform comprises a silicone-based anthropomorphic endovascular model mounted on a plate and rigidly coupled to a six-degree-of-freedom (6-DoF) force/torque (F/T) sensor (SBT308, SIMBATOUCHE, Guangzhou, China), with a composite error of 1.0%. The sensor is mounted close to the centre of gravity of the platform and provides force readings in each of the three directions (x, y, and z). The maximum measurement range in all three directions is 49 N, with sensitivities of $x = 1.038093$ mV/V, $y = 1.03985$ mV/V, and $z = 0.97937$ mV/V, respectively, as shown in **Figure 4.1(B3)**.

- ***Position and rotation data acquisition***

To capture the motion of the operator's hand manipulation, a magneto-strictive position sensor (SDM20T-0150-MR2P-MEP03-1, Soway Tech., Shenzhen, China) and a rotary encoder (E40H12-1000-3-N-5, Autonics, Busan, South Korea) are used to measure axial position and rotation motion, respectively. The handle operated by the operator consists of a hollow magnetic ring, a hollow rotary encoder, and a bracket connected to the slider, enabling simultaneous rotational and axial movements while preserving the operator's natural catheterization skills.

- ***Video stream data acquisition***

To document the robot-assisted intervention procedure, three distinct cameras were used for specific purposes. The first camera recorded the operator's hand motions while manipulating the master robotic device, thereby providing a comprehensive view of their catheterization manoeuvres

along the paths illustrated in **Figure 4.1(B4)**. The video data from this camera, acquired at a frame rate of 30 frame per seconds (Fps), captured both the operators' behaviour and their manipulation of the robotic master mechanism. The second camera was dedicated to capturing the motion procedures associated with the delivery of the guidewire from the slave robotic device. Illustrated how the slave device executed actions in response to commands received from the master device. The footage from this camera provided valuable insight into the coordination between the master and slave robotic components. Finally, the third camera was specifically used for offline video analysis. It facilitated the identification and annotation of pathological segments within the video frames, including Part A: branch, Part B: stenosis, Part C: stenosis, and Part D: tortuous path. This analysis was instrumental in understanding and evaluating the effectiveness of the robot-assisted intervention procedure in addressing various challenging scenarios.

B. Signal processing

Operators were tasked with catheterizing the guidewire from the coronary ostium (starting point) to the target point on the anterior descending branch. Data from multiple sources were saved separately for further processing, as described below.

- **Data Pre-processing:**

The sEMG signal is a non-stationary micro-electrical signal characterized by an amplitude range of 0 to 1.5 mV and a useful frequency range of 0 to 500 Hz. To enhance signal quality, a bandpass filter with a range of 10 to 500 Hz was applied to eliminate components outside the desired frequency bandwidth. Additionally, a 50 Hz notch filter was employed to remove power frequency disturbances. To ensure uniformity across different operators in the sEMG dataset, we utilized minimum and maximum normalization functions to standardize the sEMG signal, facilitating fair comparisons among signals from all operators. For the remaining data obtained from glove motion, distal and proximal interaction forces, haptic force, and position and rotation displacements of the master-slave system, smoothing was applied using an average filter. This step effectively removed abnormal data and spikes in motion, resulting in a more reliable and consistent dataset for further analysis.

- **Resampling:**

The sampling rates of the different signals in the setup vary; the sEMG signals have a sampling rate of 2000 Hz, which is higher than that of finger motion (60 Hz), position and rotation data in the RCS (60 Hz), tool-vessel distal force (60 Hz), haptic force signals (50 Hz), and proximal force of the tool instrument (100 Hz). These acquired sequences do not match in the time domain. To ensure alignment of the signals for developing the proposed learning-based system, the high-frequency signals are synchronized with the low-frequency signals according to their timestamps.

C. Manipulation technical skill classification

A total of fourteen operators were recruited from the University of Chinese Academy of Sciences

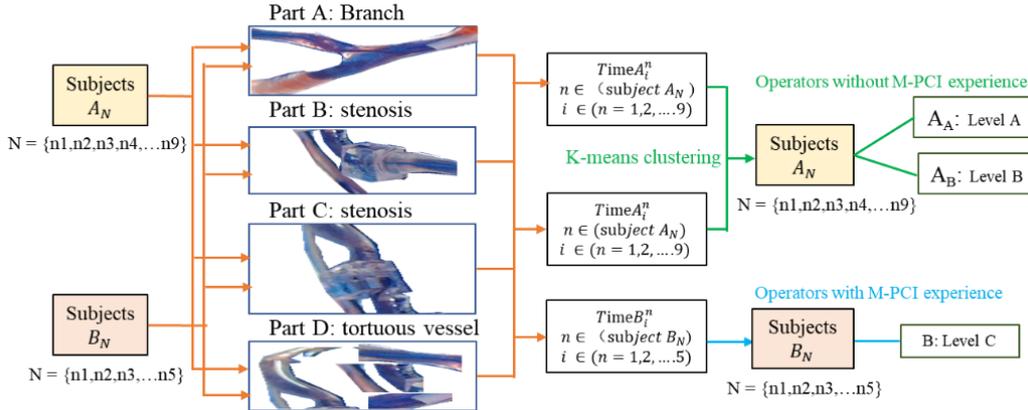


Figure 4.4: Overview of technical skill classification.

Shenzhen Hospital and the Shenzhen Institutes of Advanced Technology in Shenzhen, China. The operators participating in the study were classified into two groups: Group A, which consists of ten subjects with no prior experience in M-PCI and Group B, which consists of five subjects with experience in M-PCI.

Within Group A, further division was performed using the k-means clustering method, based on the time spent cannulating the endovascular pathway and the number of training sessions completed on the endovascular simulator for robot-assisted interventional procedures. This division resulted in two subgroups: Group A_A included operators demonstrating longer manipulation times and fewer training sessions, and Group A_B , consisting of operators exhibiting shorter manipulation times and more training sessions. Consequently, Groups A_A , A_B , and B represented different technical skill levels, defined as Level A, Level B, and Level C, respectively, as shown in **Figure 4.4**.

4.4 Manipulation-Based Synergy Characteristic Modelling

4.4.1 Manipulation pattern recognition modelling

Since the operator directly manipulates the master robotic device, with the manipulation pattern categorized as Label $A \in (\text{PH}, \text{PL}, \text{CR}, \text{CCR}, \text{PHCR}, \text{PLCR}, \text{PHCCR}, \text{PLCCR})$. Meanwhile, the slave robot follows the control model instructions sent by the master robot to deliver instruments, and its motion pattern is classified as Label $B \in (\text{PH}, \text{PL}, \text{CR}, \text{CCR}, \text{PHCR}, \text{PLCR}, \text{PHCCR}, \text{PLCCR})$. Thus, the cooperation between the human operator and the robot is reflected in the consistency and synchronization of the master-slave mechanism's operation patterns. To assess the synergy between operators with varying technical skills and the robot, a manipulation pattern recognition framework based on convolutional neural networks is proposed. This framework integrates convolutional and dense layers, augmented by an attention block to capture relevant gradients from multimodal data. The CNN model is chosen for its efficiency in learning within a simplified network structure.

The framework is composed of three stages: (1) signal acquisition and data processing, (2) feature extraction using the CNN model, and (3) manipulation pattern recognition through a fully connected

network. After aligning all data by timestamps, the pre-processed dataset is input into the CNN-attention module, which extracts deep features from the high-dimensional data consisting of 58 channels. These channels include 4 from sEMG muscle activity, 14 from finger flexion, 32 from contact force, 1 from proximal force, 3 from distal force, and 4 from the position and rotation displacement of the master-slave robotic system. The data is structured into matrices with dimensions $(6, X_i^m)$, which are then processed by the CNN model.

The CNN model for feature extraction consists of three types of layers: (1) an input layer with units L_i^0 , where the input dataset has fixed values; (2) hidden layers with units L_i^m , where values are derived from previous layers $(m-1)$; and (3) an output layer with units L_i^M , where values are derived from the last hidden layer. A set of weights, denoted as $w_{i,j}^m$, is adjusted during training, where $w_{i,j}^m$ represents the weight from an input unit L_i^m to an output unit L_j^{m+1} . The total input is denoted as $(6, X_i^m)$, and Y_i^m represents the output of unit L_i^m . Deep feature vectors (F) are extracted from data chunks, segmented using a unique sliding window approach with window sizes of $6 \times C \times S$. This segmentation reduces the impact of transient and random fluctuations in non-stationary signals. The window properties are carefully chosen to ensure that most segments encompass complete cycles of each motion type during the robot-assisted guidewire delivery procedure. Consequently, each segment represents a set of hand movements by the surgeon and is organized as a matrix, where C denotes the number of channels and S indicates the number of segments.

Figure 4.5 illustrates the CNN component, which consists of multiple convolutional blocks ($n=2,3,\dots,N$). Each block includes a convolution layer with 32 or 64 kernels of size 1, a stride of 1, and padding of 1 for feature extraction, resulting in multi-dimensional feature maps. To optimize the model by eliminating redundant features, a CNN-attention network is employed to operate on both local and global feature maps. The channel attention module extracts significant features by learning the characteristics of the multi-dimensional time series. Initial feature maps are processed through a convolutional layer with max pooling to aggregate features along the channel dimension. These feature maps are then processed using average pooling to extract vector-level local features,

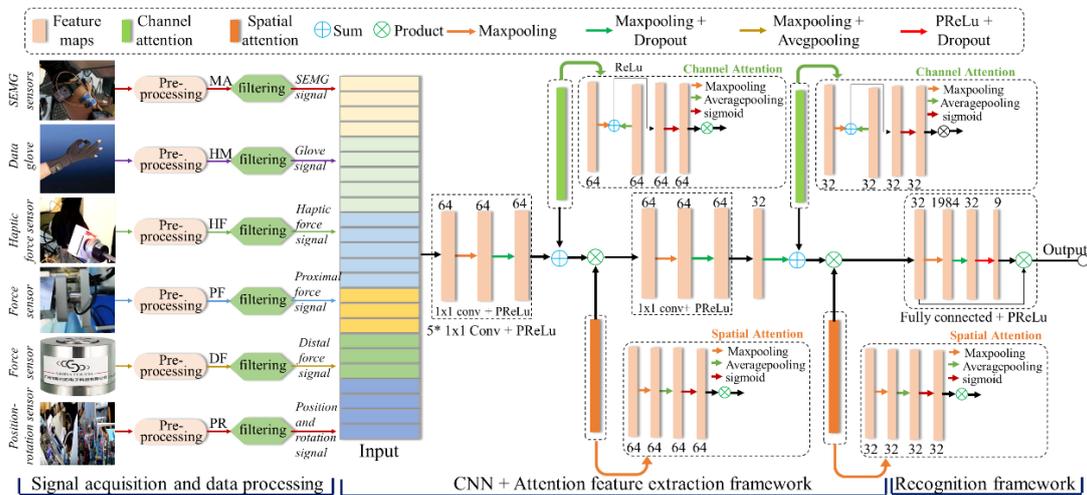


Figure 4.5: Overview of the evaluation framework of surgeon-robot synergy.

with the salient features recalculated using a weight vector. To capture global information and further enhance model performance, a spatial attention module is introduced. This module includes max-pooling, average-pooling, and sigmoid operations, along with four convolution operations. The sigmoid function constrains the output range between 0 and 1, where 0 indicates an irrelevant feature and 1 represents a highly important feature. This process transforms the feature maps into a global feature vector, and the procedure is repeated twice to obtain refined feature maps. Finally, the extracted features are passed into a SoftMax logistic module for manipulation pattern recognition, enabling the framework to effectively categorize different manipulation patterns.

4.4.2 Training strategy and evaluation metric

The network utilizes the categorical cross-entropy loss function, as defined in equation (4.1):

$$Loss = -\sum_{i=1}^n \hat{y}_{i1} \log y_{i1} + \hat{y}_{i2} \log y_{i2} + L + \hat{y}_{im} \log y_{im} \quad (4.1)$$

where n represents the number of samples, m is the number of classes $m > 2$, \hat{y}_{im} is 1 if the sample belongs to class m , and 0 otherwise. y_{im} denotes the predicted probability that the sample belongs to class m . The dataset consists of 348,760 samples, which are split into training (70% or 244,132 samples), validation (10% or 34,876 samples), and testing (20% or 69,752 samples). The Adam optimizer was selected for training, which ensures efficient and fast convergence during training, as defined in equation (4.2):

$$F_{fs} = \frac{1}{n} \sum_{i=1}^n \frac{1}{t_i} \quad (4.2)$$

The network underwent 100 epochs of training, with performance assessed using a variety of validation metrics. Throughout training, the model's parameters were dynamically adjusted based on the flow of multimodal data, ensuring that the network adapted to the dataset. Training continued until the loss value stabilized and ceased to decrease.

The manipulation pattern recognition framework can be regarded as a nine-class manipulation classification task. Model performance was evaluated using precision (Pre), recall (Rec), and accuracy (Acc), calculated as follows:

$$Pre = TP / (TP + FP) \quad (4.3)$$

$$Rec = TP / (TP + FN) \quad (4.4)$$

$$Acc = (TP + TN) / (TP + FP + TN + FN) \quad (4.5)$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively. The model, trained with a learning rate of 0.0002, achieved the highest accuracy on both the training and validation datasets, as shown in **Figure 4.6**. Furthermore, a 10-fold cross-validation was applied during training, yielding an average training accuracy of 99.50±0.16% and validation accuracy of 99.85±0.20%. These results demonstrate that the extracted features provide a solid foundation for the proposed synergy framework in evaluating human-robot coordination performance.

The evaluation framework was developed using multiple data sources, with the target function

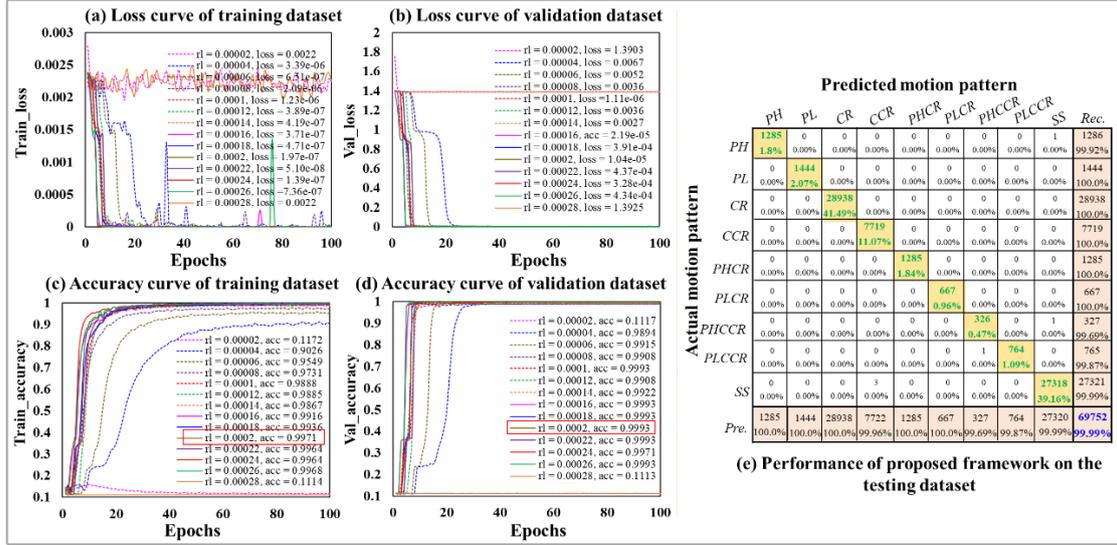


Figure 4.6: Performance of the proposed recognition framework based on different learning rate.

for model training based on the master device manipulation patterns. The effectiveness of this framework in assessing the cooperative performance between the human operator and the robot is contingent upon the accuracy of the model's prediction of the master device's motion pattern. For the framework to function as a reliable evaluation tool, it is critical that the model predicts the master device's motion pattern with high accuracy, minimizing both false positives and false negatives. Inadequate performance in predicting the motion pattern of the master device would undermine the framework's ability to assess human-robot cooperation accurately. Therefore, the reliability of the evaluation of the framework is directly related to the accuracy of the model in predicting the motion pattern of the master device.

To evaluate the performance of the model, we assessed its accuracy, recall, and precision. The results demonstrated an exceptional accuracy of 99.99% on the test data set. Additionally, the precision values for each motion pattern were notably high: PH-100%, PL-100%, CR-100%, CCR-99.96%, PHCR-100%, PLCR-100%, PHCCR-99.96%, PLCCR-99.87%, and SS-99.99% (Figure 4.6(e)). These results indicate that the proposed evaluation framework is reliable and effective, providing strong evidence for its potential to capture the synergy ratio in human-robot interactions.

4.4.3 Human-machine synergy evaluation strategy

Cooperative performance in robotic systems, particularly in human-robot interaction, is defined by the ability of the system to effectively synchronize and harmonize the actions of the operator and the robot. A key indicator of successful cooperation is the seamless interaction between the master and slave devices, which allows for the timely and accurate execution of tasks. This interaction is influenced by the flow of information between the operator and the robotic system. Key factors, such as the accuracy of motion replication by the slave robot and the responsiveness of the system to operator inputs, play a pivotal role in determining the level of synergy.

In teleoperating robotic systems, a critical factor in human-robot synergy is the ability of the

slave device to promptly and accurately replicate the motion pattern of the operator's manipulation of the master robot. Any delays in the robotic system can reduce the synergy between the operator and the robot. To evaluate this synergy, GH uses the proposed manipulation recognition framework. When the slave robot successfully mimics the operator's motions with high precision, it indicates strong cooperation between the human and the robot. Conversely, if the slave robot's motion pattern becomes erratic or lacks synchronization with the master robot, this suggests poor cooperation.

To quantify the synergy, the master manipulation pattern (Label A \in (PH, PL, CR, CCR, PHCR, PLCR, PHCCR, and PLCCR)) obtained based on the master robot's position and rotation dataset, and the recognition accuracy achieved by the proposed framework in identifying master manipulation pattern, is denoted as Acc_{master} . Similarly, the slave manipulation pattern (Label B \in (PH, PL, CR, CCR, PHCR, PLCR, PHCCR, and PLCCR)) obtained based on the slave robot's position and rotation dataset, the recognition accuracy achieved by the proposed framework in identifying slave manipulation pattern, is denoted as Acc_{slave} . When the framework achieves high accuracy (Acc_{master}) in recognizing the master device's manipulation pattern, while also achieving high accuracy (Acc_{slave}) in recognizing the slave device's manipulation pattern, which indicates that surgeon and robot existed a good cooperative performance. However, when cooperation is poor, the framework typically achieves high accuracy for the master device's manipulation pattern, but obtains low accuracy for the slave device's motion pattern. To quantify the overall synergy performance, the synergy ratio is calculated as follows:

$$\text{Synergy Ratio} = (Acc_{master} \times Acc_{slave}) \times 100\% \quad (4.6)$$

where a synergy ratio nearer to 1 signifies enhanced collaboration between the operator and the robot, whereas a ratio tending towards 0 denotes inadequate cooperation.

The synergy ratio is assessed from three perspectives. First, under the assumption of no system delays, the evaluation framework assesses operator-robot cooperation. Second, the impact of a constant communication delay (0.254 s) on the synergy ratio is considered. Finally, variable delays, which depend on the manipulation pattern, are evaluated to better capture the real-time delay characteristics of the system. This approach enables the optimization of the cooperative strategy for improved performance.

4.5 Cooperative Performance Analysis in Human-Robot Systems

In a master-slave isomorphic robotic system, delays can impact real-time interactions between operators and the master-slave mechanism, leading to a reduction in overall system control. Poor cooperation between operators and the robot can significantly compromise the stability and effectiveness of the surgical robot, posing risks and challenges to the smooth execution of PCI procedures. This section analyses the cooperative performance between human operators and the robotic system, focusing on three key delay factors: no delay, constant delay, and variable delay. The objective is to examine how these delays influence operator performance at different skill levels,

with the goal of improving the safety and effectiveness of the robotic system during robot-assisted PCI.

A critical component of this analysis is the investigation of contact forces between the operator's hand and the instrument, as well as the distal forces between the endovascular tool and vessel tissue. This approach aims to understand how operators with varying technical skills adjust their control strategies to avoid excessive force that could damage endovascular tissues while navigating complex anatomical pathways.

The proposed framework integrates multiple data sources to assess cooperative performance in three specific scenarios: 1) no delay, 2) constant communication delay, and 3) variable delay based on manipulation patterns. The Kruskal-Wallis (ANOVA) statistical method was then applied to analyse significant differences in manipulation time, haptic force, and distal force in the three levels of operator skill during delivery of robot-assisted guidewire through branches, stenotic, and tortuous vessels.

4.5.1 Performance evaluation of the synergy between operator and robotic system

This study examines cooperative performance between operators with varying technical skills and a robot, focusing on the impact of three delay factors—no delay, constant delay, and variable delay—during robot-assisted PCI procedures. The synergy ratios between the operator and the robot, as presented in **Figure 4.7**, were calculated for three operator groups with different skill levels: A_A , A_B , and B. The structural design of the master-slave isomorphic robotic system inherently introduces a delay factor. When this delay remains within an acceptable range, the system can be classified as Satisfied or Extremely Satisfied [124]. Initially, we assumed the system operated with no or minimal delay and evaluated cooperative performance accordingly. In this scenario, the synergy ratios were 49.84% for operator A_A , 51.84% for operator A_B , and 53.42% for operator B (**Figure 4.7(a)**). These results indicate a low synergy ratio across all groups, suggesting that inherent delay factors in the master-slave system may hinder the robotic device's ability to replicate operators' manipulations in real-time.

When a constant communication delay was introduced, the synergy ratios increased slightly to 51.88% for operator A_A , 56.17% for operator A_B , and 58.71% for operator B. This modest improvement suggests that a constant delay factor accounts for only part of the overall delay, with additional factors still unaddressed. Further analysis using a variable delay factor based on manipulation patterns revealed a more substantial influence on cooperative performance. When a manipulation-based variable delay was applied, the synergy ratios rose significantly to 89.66%, 90.28%, and 91.12% for operators A_A , A_B , and B, respectively. These findings demonstrate that the variable delay factor considerably enhances operator-robot cooperation compared to a constant delay.

The results show that operators with different experience perform different synergy ratios under variable delay factor. Operators B with clinical PCI experience have a higher synergy ratio, which indicated that between operators and robot has better cooperative performance, while operators A_A

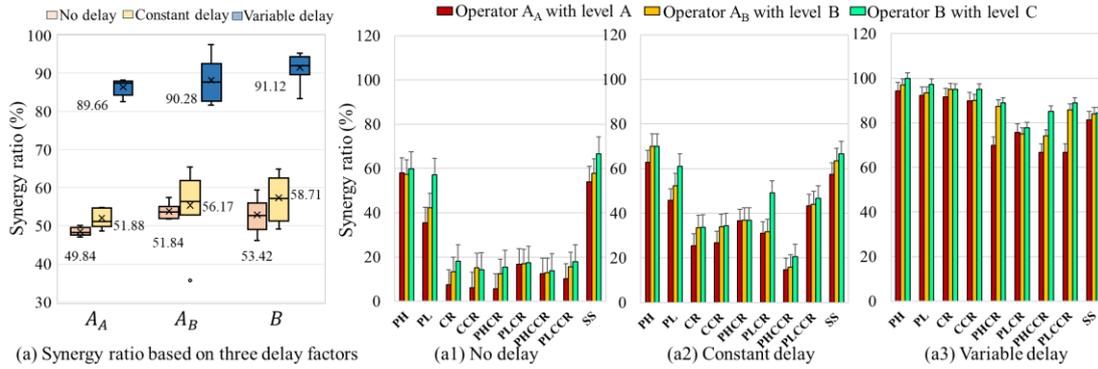


Figure 4.7: Performance of surgeon-robot synergy in operator with three different skill level.

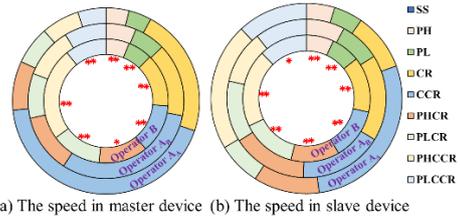
and A_B without clinical PCI experience shows slightly poor cooperation. The synergy ratio between operators and robot can help operators to know whether current manipulation is appropriate, if not, surgeon can change manipulation strategy in time to avoid damage to the blood vessel. For example, when the synergy ratio is low, operators can immediately change their manipulation patterns (push, pull, CCR, CR, PHCR, PHCCR, PLCR, PLCCR), manipulation forces, manipulation speeds to better use the robot to deliver the catheter or guidewire for reducing the risk of vascular rupture. However, when the synergy ratio is high, the operator can continue current manipulation strategy and focus more attention on more complex vascular paths, reducing the operators' attention load.

Moreover, an in-depth analysis of different manipulation patterns under each delay factor, as shown in **Figure 4.7(a1-a3)**, reveals that patterns with no delay exhibited significantly lower synergy ratios compared to those with variable delays. Among manipulation types, push, pull, and SS achieved higher synergy ratios, whereas more complex patterns, such as CR, CCR, PHCR, PLCR, PHCCR, and PLCCR, have lower ratios. These findings suggest that complex manipulation patterns exhibit reduced cooperative performance, especially when guiding the guidewire from the start to target B. Additionally, these complex patterns are associated with greater delays than simpler manipulations like push, pull, and SS. This trend persists even when considering a constant delay factor.

Figure 4.7(a3) illustrates that simple manipulation patterns, such as push, pull, CR, and CCR, generally had higher synergy ratios than compound patterns like PHCR, PLCR, PHCCR, and PLCCR. This finding indicates better cooperation between the operator and the robot during simpler manipulations. Furthermore, the synergy ratios are consistently higher for operator group B than for groups A_A and A_B, implying that operators with greater experience (group B) adopted more effective control strategies, resulting in better cooperative performance. The application of manipulation-based delay factors, where motion patterns serve as units for determining delay times, proved more effective in mitigating delays caused by packet loss than time-based units. This approach is advantageous because manipulation-based delay factors maintain the integrity of entire manipulation patterns, even if there is short-term packet loss, making them more suitable for master slave isomorphic robotic systems.

Table 4.2: Delivery speed among three different groups.

Manipulation speed (mm/s)	Master			Slave		
	A _A	A _B	B	A _A	A _B	B
PH	53.65	55.69	51.32	52.99	54.51	51.57
PL	53.33	57.31	54.03	54.18	57.97	54.45
CR	299.05	135.41	117.67	107.89	122.63	105.27
CCR	581.83	275.19	107.83	314.22	118.49	97.39
PHCR	114.58	108.25	108.22	123.19	107.48	91.77
PLCR	88.69	82.73	111.09	81.59	87.24	103.16
PHCCR	88.91	93.19	205.68	239.58	75.69	146.93
PLCCR	65.42	77.15	91.59	132.14	74.19	91.59


Figure 4.8: Speed manipulation of robot master and slave mechanism based on different movement patterns.

: significant difference (: at 0.05 level, **: at 0.01 level).

Finally, an analysis of the manipulation speed of each operator group provides information on the impact of the control strategy on cooperative performance (**Table 4.2**). **Figure 4.8** shows significant speed differences between the three skill groups. Operators B display slower speeds to push the guidewire compared to operators A_A and A_B. However, their pulling speeds are well calibrated, positioned between the faster speeds of A_B and the slower speeds of A_A. In contrast, A_A operators exhibit faster speeds during rotational manipulations to adjust the direction of the guidewire, possibly due to limited technical skills and experience in manual PCI procedures, which could increase the risk of complications.

In complex endovascular pathways, operators often deliver and retract the guidewire through stenosis or branch sites. Operators B use higher speeds for the PLCR, PHCCR, and PLCCR compound motions than operators A_A and A_B. However, they use slower speeds for PHCR, particularly when navigate challenging endovascular pathways. These results suggest that operators B adopt more appropriate control strategies based on their manual PCI experience.

Regarding the slave device, its manipulation speed is varied by operator group. For A_A operators, the slave device's speeds for push and pull motions closely match those of the master device. However, the slave speeds for CR and CCR are lower, indicating difficulty in synchronizing with the master device, which could lead to data loss. In contrast, for the PHCCR and PLCCR motions, A_A operators exhibit higher manipulation speeds on the slave device than on the master device, suggesting inconsistencies in following the master's actions. For operators A_B and B, the manipulation speeds of the master and slave devices are similar, indicating better synchronization between the operator and robot.

These results and analysis show that operators with different manipulation technical skill perform different synergy ratio, manipulation speed in different manipulation patterns (push, pull, CR, CCR, PLCR, PLCCR, PHCR, PLCCR). Composite manipulation patterns (PLCR, PLCCR, PHCR, PLCCR) show low synergy ratio, while single manipulation patterns (push, pull, CR, CCR) perform relatively high synergy ratio. The main reason is that operators need to perform radial translation with axial rotation in composite manipulation patterns, but slave delivery robot cannot switch in-time between radial translation and axial rotation due to operators' manipulation speed and force,

which cause that slave robot cannot follow the operators' composite manipulation patterns in real-time. This can help the operator to pay attention to the synergy ratio of composite manipulation patterns when performing complex vascular paths, such branch vascular pathways or stenosis site. When the synergy ratio of composite manipulation patterns is low, operators should pay attention to changing the manipulation strategies, such as manipulation force and manipulation speed, to ensure efficient and safe delivery of the catheter or guidewire.

4.5.2 Manipulation time analysis

Manipulation time in surgical procedures is a critical indicator of a surgeon's technical skill. In the context of PCI procedures, longer manipulation times lead to increased exposure to X-rays for both the patient and the surgeon, potentially compromising safety and efficiency. To assess the effect of technical skill levels on manipulation time, a non-parametric independent test was conducted to compare manipulation times across various operator groups. The results revealed a statistically significant difference ($p = 0.001 - 0.01$, $p < 0.05$) in manipulation times between the three groups when navigating the guidewire through different vessel paths, including branch path A, stenosis paths B and C, and tortuous path D (**Figure 4.9(a-d)**).

Operators in group B demonstrated shorter manipulation times when manoeuvring the guidewire through the branch vessels (path A) and stenosis (path B) compared to operators in groups A_A and A_B. This suggests that the operators in group B, with more advanced technical skills, exhibited better performance and more efficient control strategies on these pathways. However, when traversing the stenosis vessel (path C) and the tortuous vessel (path D), the operators in group B took slightly longer than the operators in group A_B but still outperformed the operators in group A_A. This indicates that operators in group B, despite their greater skill, may have hesitated due to limited familiarity with the robotic system. These results highlight the superior technical skills of the operators in group B, attributed to their extensive knowledge of surgical techniques and effective navigation strategies for complex endovascular pathways, gained from prior experience in manual PCI surgeries.

Further analysis of manipulation times suggests that operators in group B, with manual PCI experience, adapted their technical skills and control strategies more effectively than operators in groups A_A and A_B, who lacked PCI experience. For example, when navigating the branch vessel, operators in group B quickly adjusted their manipulation strategy to ensure efficient passage of the guidewire through complex pathways, demonstrating a faster learning curve. On the contrary, the operators in the A_A and A_B groups relied on their existing cognitive strategies and manual skills, resulting in longer manipulation times as they adjusted the hand movements to guide the wire to the target point.

Temporal metrics often correlate with skill level [228]. Manipulation times and speeds for each motion pattern confirmed that operators with different technical skills applied different manipulation strategies. The various technical skills they have, such as surgical knowledge, control methods, manipulation speed, applied contact force, and behavioural patterns, all contributed to improving the collaborative interaction between the operators and the robot. The positive results of

this study suggest that the proposed evaluation framework has significant potential to assess the cooperative interaction between operators and robot.

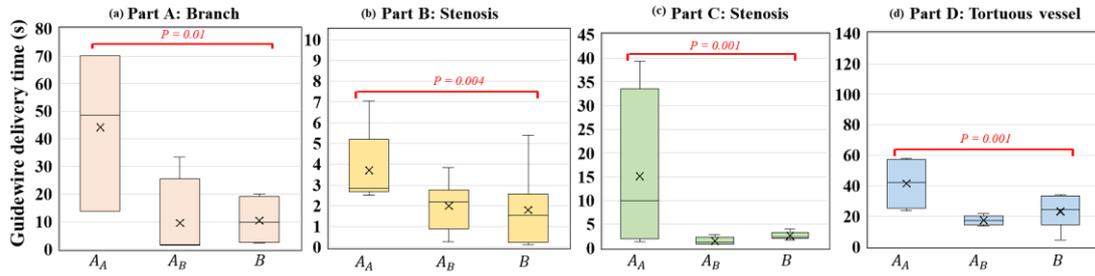


Figure 4.9: Time manipulation of operators with different technical skill among three groups.

4.5.3 Interaction force analysis

In R-PCI procedures, a key factor is controlling the robotic system to deliver the catheter or guidewire to the target site safely and efficiently. Although experts typically have more precise control skills than novices in cardiac interventional surgery, it remains unclear whether this advantage extends to R-PCI, where robotic manipulation skills are required. To address this, we statistically analysed robot-assisted procedures in terms of manipulation time, distal force between tools and tissues, and contact force between the fingers and the control ring. Data from these analyses characterize technical skill levels when navigating branched, stenotic, and tortuous vessels, providing a benchmark for skill evaluation in R-PCI procedures.

The analysis involved assessing both the distal force between tools and tissues and the contact force between fingers and the control ring, categorized by the operators' expertise levels. A non-parametric Kruskal-Wallis ANOVA test was employed for statistical analysis, with significant differences indicated by asterisks (* for $p < 0.05$ and ** for $p < 0.01$). This approach provided valuable insights into the relationship between operator skill and the delicate balance of applying sufficient force to ensure safe and effective navigation during the procedure.

The findings offer critical insights for designing haptic feedback systems tailored to operators with varying levels of expertise. Specifically, this information is essential for selecting sensor materials, precision, range, and stability to enhance force feedback accuracy. Proper selection of components in these areas also supports the evaluation of interaction forces during robot-assisted PCI, providing an efficient mechanism for alerting operators to adjust their applied force to safely guide the guidewire through complex endovascular paths.

A. Distal force between instrument and tissue

The generated endovascular force, measured as both the average distal force and the maximum distal force, was found to be higher for the operators in group A_A than for those in group B when navigating the guidewire through a branch vessel, as shown in **Figure 4.10(a1)** and (a2). This finding suggests that operators in group A_A may not have the ability to adjust their manipulation strategy effectively to navigate the branch vessel safely and efficiently during robot-assisted catheterization. Elevated force between tool and tissue increases the risk of endovascular rupture,

highlighting the importance of controlled force application.

Similarly, operators in group B exhibited greater average and maximum distal forces between the

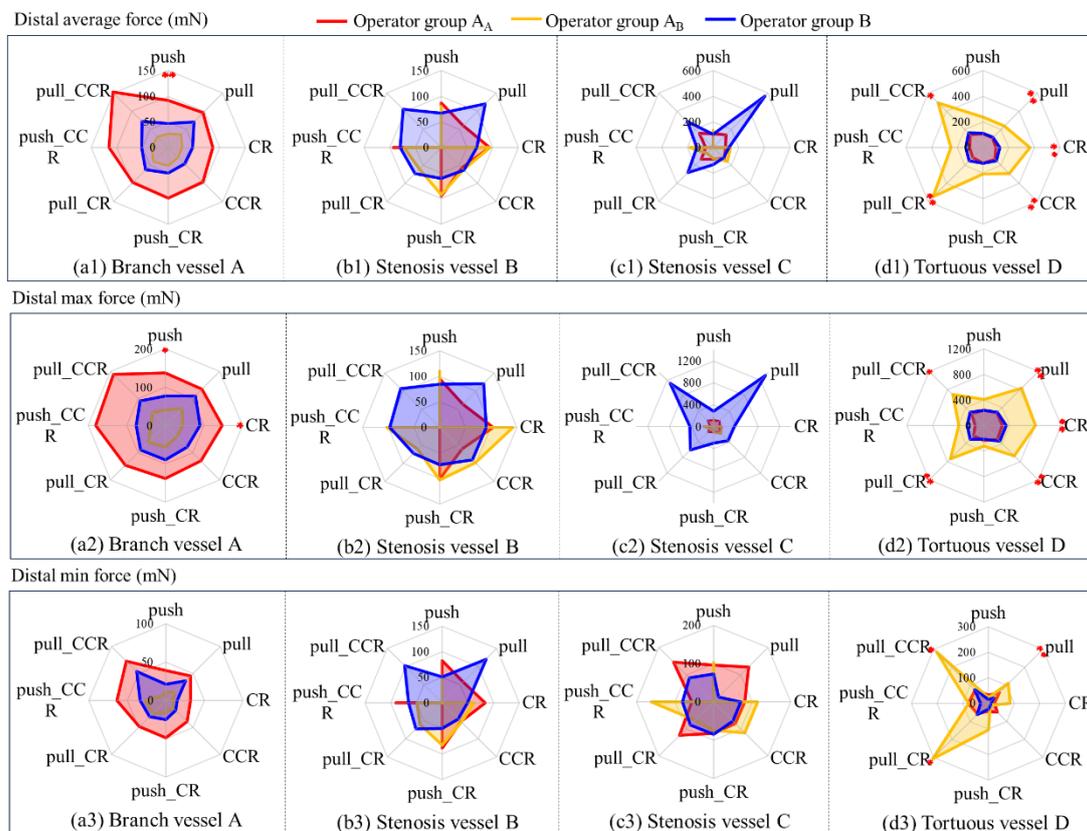


Figure 4.10: Comparison of the average distal force among three groups, including A_A , A_B and B.

tool and tissue compared to operators in groups A_A and A_B when providing the guidewire through two stenotic sites. This observation suggests that their manipulation strategy includes applying enough tension to navigate stenosis pathways effectively. Furthermore, during guidewire delivery through a tortuous vessel (**Figure 4.10(d1)**), operators in group A_B applied a manipulation strategy that resulted in higher average and maximum distal forces between the guidewire and tissue compared to those applied by operators in groups A_A and B. This result may be due to the absence of complex anatomical features, such as stenosis or branches, in the tortuous vessel, allowing operators to guide the wire more freely along the blood flow direction. Furthermore, the operators in group A_B have more experience in robot-assisted interventional procedures than those in groups A_A and B, which may contribute to their distinct manipulation strategy.

Furthermore, the operators in group B applied a manipulation strategy that generated a higher average and maximum distal force when delivering the guidewire through a more severe stenosis pathway (42.69%) compared to a less severe stenosis pathway (37.45%). This finding suggests that navigating higher degrees of stenosis requires greater applied tension to safely guide the wire through the constricted areas.

B. Proximal force

The proximal force is defined as the resulting force generated by interactions between tools and robot, including resistance at the tip of the tool against the wall of the vessel, as well as friction between the tool, endovascular tissue, and the wall of the vessel. As shown in **Figure 4.11**, we

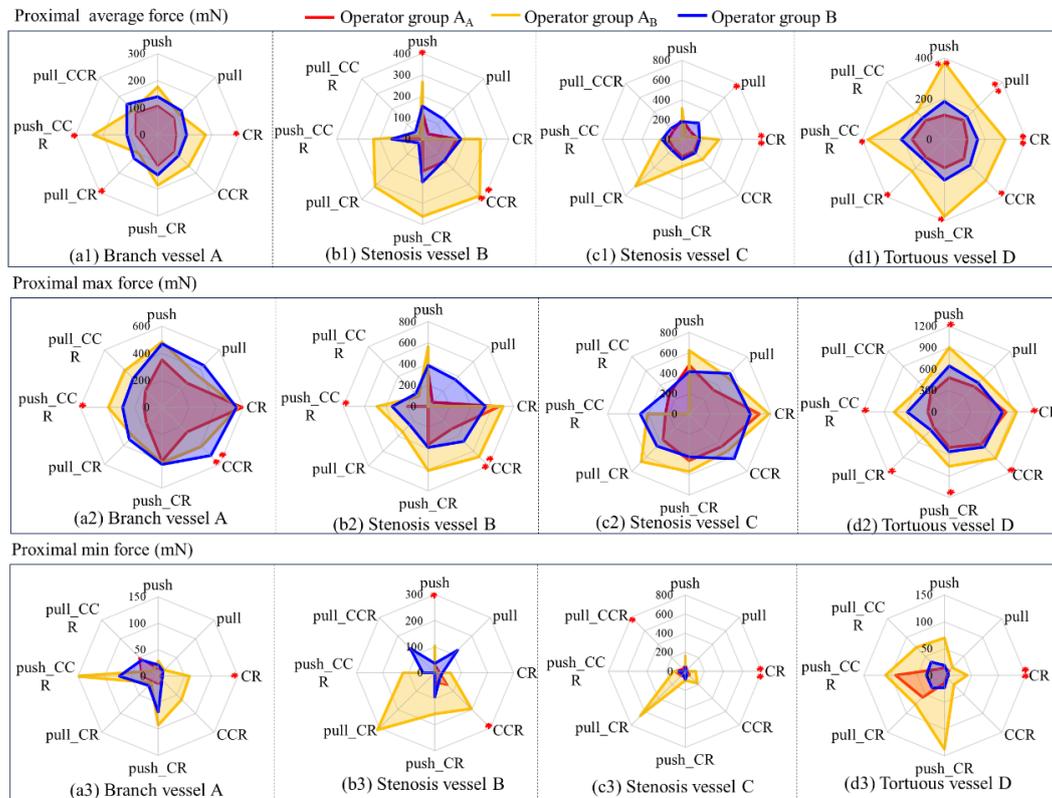


Figure 4.11: Proximal force among three skill group.

analysed the differences in proximal force among the three operator groups while navigating through branch, stenotic, and tortuous vessel paths. The results indicate that the operators of group A_B applied higher proximal force values when cannulating through stenotic and tortuous vessels compared to groups A_A and B, except in the branch vessel path.

This finding suggests that operators in the A_B group, who have experience with robotic systems but lack manual PCI experience, tend to control the RCS with greater force. Their limited experience in direct manual manipulation may mean that they lack the cognitive and haptic perception needed to accurately sense endovascular forces. Consequently, their focus may be on skilfully operating the robotic system to guide the guidewire to the target site, potentially overlooking the risks associated with higher manipulation forces or frequent operational adjustments, which could damage the endovascular wall.

C. Contact force between finger and manipulation-ring

A significant difference in the average contact force between the finger and the manipulation ring was observed among the three operators' groups, as shown in **Figure 4.12**. Specifically, operators in group B, who have manual PCI experience, exhibited substantially higher average contact forces when performing PL (81.10 mN), CR (83.87 mN), CCR (66.07 mN), PHCR (96.58 mN), and

PHCCR (79.87 mN) movements compared to operators in groups A_A and A_B , who lack prior PCI experience, during guidewire delivery through the branch vessel pathway (**Figure 4.12(a1)**). This increased contact force among operators in group B can be attributed to their familiarity with manually handling the guidewire in traditional PCI surgery, where firm grip is essential to prevent slippage. Consequently, when gloves are used to manipulate the catheter and guidewire in robot-assisted PCI, the grip strength developed from the manual PCI experience inadvertently affects its contact force during robotic procedures.

Furthermore, operators in group A_A applied the lowest average contact force compared to groups A_B and B on various endovascular pathways, including branch, stenotic, and tortuous paths (**Figure 4.12(a1, b1, c1, d1)**). This finding suggests that operators in the A_A group, lacking both surgical experience and robotic manipulation skills, approached catheter and guidewire handling with caution to avoid vessel rupture or potential damage to the robotic system. Additionally, **Figure 4.12(a2, a3; b2, b3; c2, c3; d2, d3)** shows significant differences in maximum and minimum contact forces between the three groups, highlighting the influence of experience on the application of contact force.

The results for distal, proximal, and contact forces showed significant differences when the operators navigated the guidewire through branched, stenotic, and tortuous vessel pathways. For example, when operators in the A_A group delivered the guidewire through the branched pathway, a greater distal force was observed between tools and tissues (average 100mN) compared to the A_B (25mN) and B (50mN) groups. Although the contact force between the finger and the control ring for the operators of group A_A was relatively low (average 20mN), their higher manipulation speeds, especially during rotation, could contribute to a larger distal interaction force. This elevated force

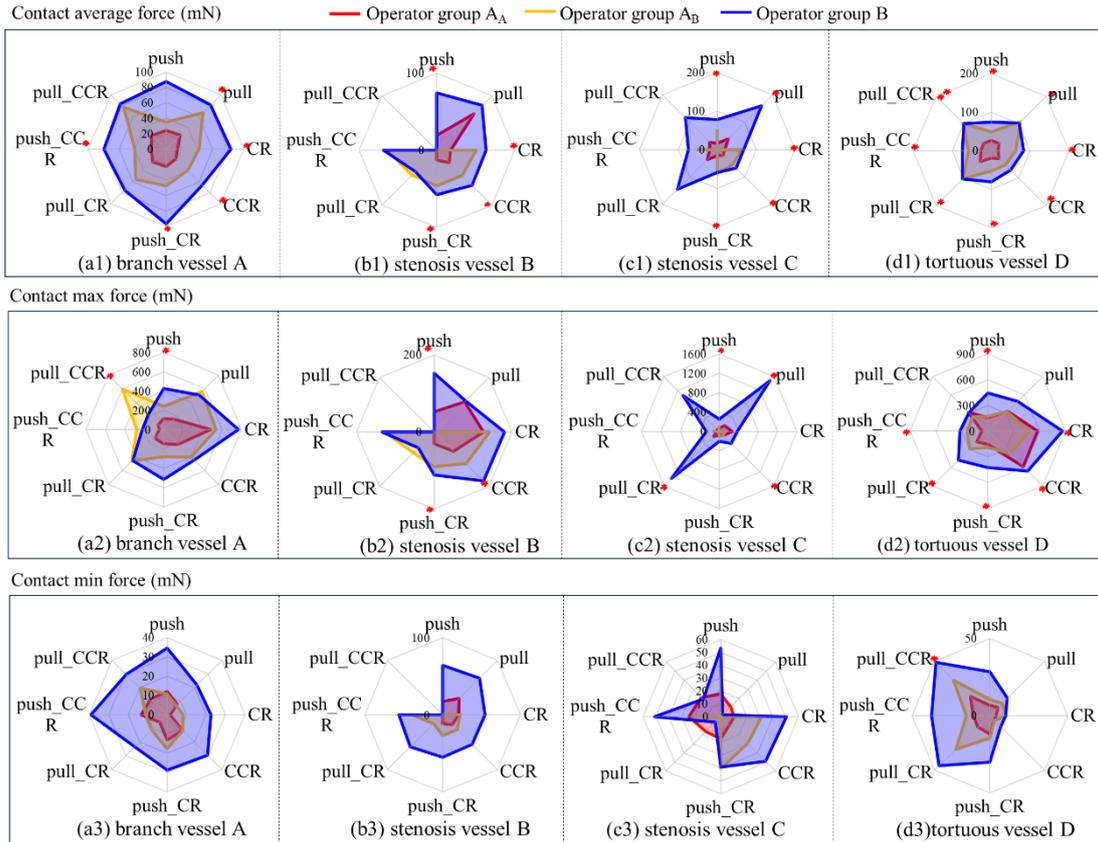


Figure 4.12: Performance of haptic force in operator with three different skill level.

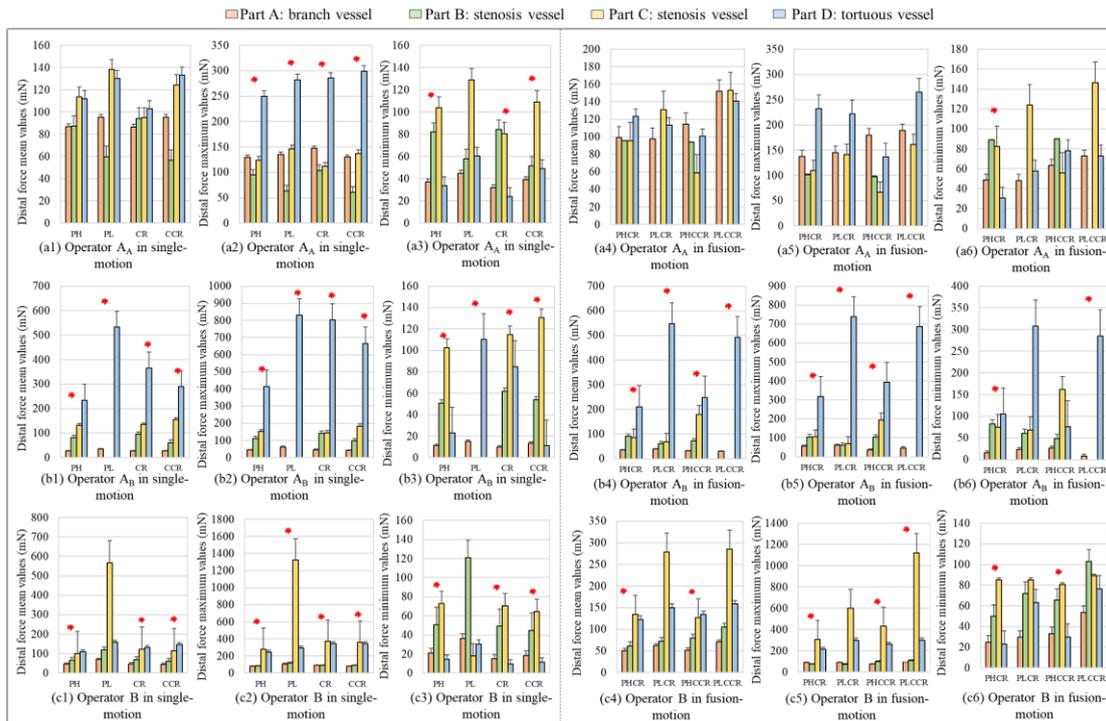


Figure 4.13: Performance of distal forces in 4 complex vessel pathways for 3 different skill groups.

suggests that operators in group A_A , who lack the control strategies and cognitive experience required to navigate complex vessel pathways, may face an increased risk of vessel rupture.

Operators in group B, in contrast, applied higher contact forces to manipulate the control ring, but exhibited lower distal interaction forces when passing through branch and tortuous pathways. This finding suggests that the operators in group B were able to transfer their manipulation strategies and cognitive skills from manual PCI to robotic PCI, effectively adapting their guidewire control. In manual PCI, operators rely on greater contact force to detect forces of endovascular interaction and prevent accidental slippage of the catheter or guidewire. Consistent with these practices, in robot-assisted PCI, operators applied greater contact force to ensure control.

Furthermore, when navigating stenotic pathways, the application of larger contact forces was correlated with an increase in distal force (average 100mN). This higher force was necessary to maintain sufficient tension to manoeuvre through narrow sections of the vessel and to overcome potential obstacles that could hinder passage. These findings are consistent with previous research [229], underscoring the importance of appropriate force application in the safe navigation of complex endovascular pathways.

D. Comparison of interaction force

Distal forces and contact force recorded during guidewire catheterization along vessels with different lesion types were analysed to quantify tool-vessel interactions and haptic force. Quantitative data on force metrics, including maximum and minimum forces, were collected to establish endovascular force measurement standards and inform the design of haptic feedback systems. Specifically, the maximum force serves as an alert for operators to adjust their approach to

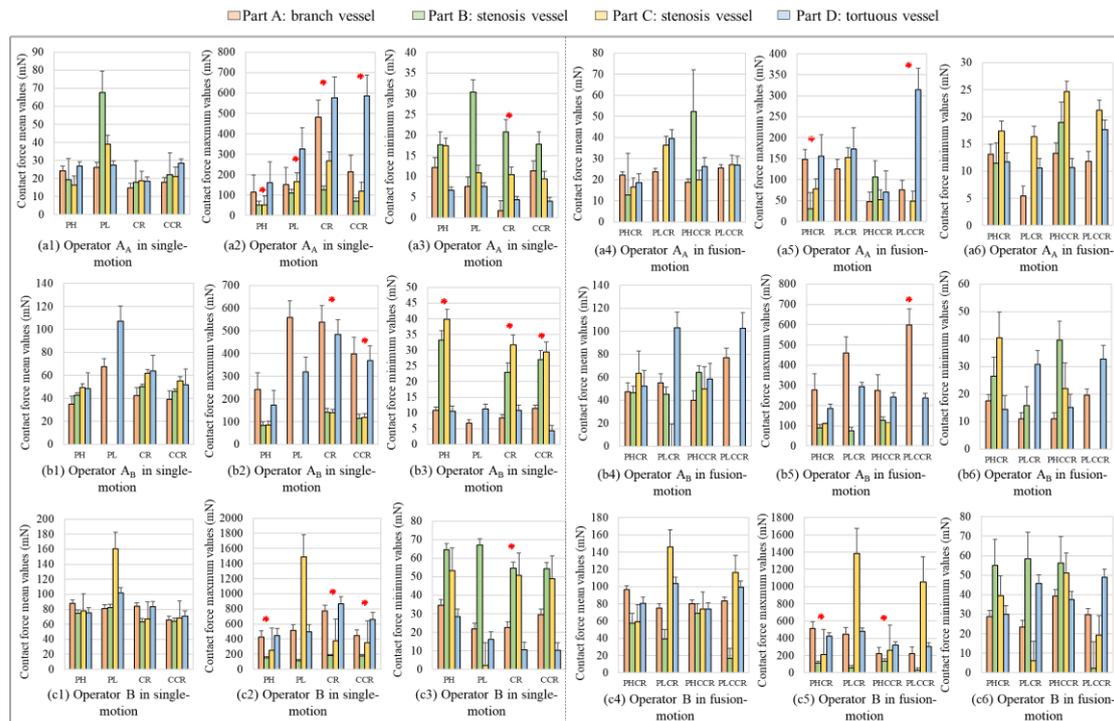


Figure 4.14: Performance of contact forces in 4 complex vessel pathways for 3 different skill groups.

prevent vessel damage when the catheter or guidewire contacts the vessel wall or surrounding tissues.

The significant differences in the interaction forces between branched, stenotic and tortuous vessels are presented in **Figure 4.13** and **Figure 4.14**, with statistically significant results indicated by an asterisk (*). These findings suggest that the operators used different control strategies to navigate the guidewire through branched, stenotic, and tortuous pathways, with the aim of preventing endovascular rupture among the three groups of operators. Such results could also be instrumental in evaluating the characteristics of endovascular lesions (e.g. branches, degrees of stenosis, calcification, and other complex lesions) in populations with coronary heart disease.

In **Figure 4.13**(a2, a5, b2, b5), operators in the A_A and A_B groups exhibited higher distal maximum forces when navigating tortuous vessels compared to branch and stenotic vessels, except during the PHCCR motion pattern. On the contrary, the operators of group B, who have PCI experience, generated greater maximum forces in stenotic vessels than in branching, slight stenosis, or tortuous pathways, as shown in **Figure 4.13**(c2, c5). This observation suggests that the operators in group B applied greater tension to maintain control through more complex paths.

The characteristics of contact force between the fingers and control ring are displayed in **Figure 4.14**. For operators in group A_A , the maximum contact force when cannulating tortuous vessels was higher than when navigating branched and stenotic vessels, except for the PHCCR pattern. Meanwhile, the operators in group A_B exhibited a higher maximum contact force in branch vessels than in stenotic and tortuous paths. Operators in group B generated the highest contact force in tortuous vessels when using single-motion strategies and in stenotic vessels when employing fusion-motion strategies. These results indicate that each group of operators applied different control strategies in response to the specific challenges posed by different endovascular lesions.

Furthermore, the study revealed significant differences in both the distal force between the tool and the tissue and the contact force between the operator's finger and the control ring across different pathways (branch A, stenosis B, stenosis C, and tortuous D) between the three groups of operators. These variations highlight that different manipulation strategies are required to navigate complex vessels safely in robot-assisted interventional procedures. Given these insights, it may be beneficial to develop a numerical model that allows operators to forecast the type of endovascular pathway (branch, stenosis, or tortuous path) in advance. Such a tool would serve as a valuable guide, allowing operators to adapt their manipulation strategies, accordingly, thereby reducing the risk of endovascular rupture in future procedures.

4.6 Chapter Summary

This chapter presents a manipulation-based framework designed to assess cooperative performance between operators of varying technical skills and a robot within a 3-D endovascular simulator for robot-assisted interventional procedures. The framework integrates multiple data sources, including muscle activity, finger motion, contact force between fingers and the control ring, distal force between tools and tissues, proximal force between tool instruments and tissues, as well as position and rotation data from both the master and slave devices in the RCS. Using training,

validation, and testing datasets, the framework demonstrates reliability and effectiveness in calculating the synergy ratio, a key measure of operator-robot cooperation.

The investigation explores three delay factors: no delay, constant delay, and variable delay. Among these, the manipulation-based variable delay factor shows significant potential to enhance operator-robot cooperation, allowing for more seamless and efficient interactions during procedures. Additionally, this chapter examines the distal forces between the catheter or guidewire and tissue, as well as the contact forces between fingers and the control ring, to understand how operators with differing technical skills adjust their control strategies. This analysis aims to prevent endovascular damage due to excessive force while ensuring sufficient tension to navigate complex pathways. These insights are instrumental in informing sensor selection and the design of haptic feedback systems with appropriate perceptual resolution, contributing to safer and more effective robot-assisted procedures.

This chapter focuses on evaluating the cooperative performance between operator and robot by analysing multi-sensor manipulation data and quantifying cooperative performance through the synergy ratio value, which suggests that operators with different manipulation technical skill (A_A , A_B , B) use different manipulation strategies. However, operators change their manipulation strategies based on the visual perception for vascular instruments' motion. Thus, changing manipulation patterns is closely relationship with the accuracy of tool perception. Moreover, the lack of direct haptic feedback increases the risk of vascular rupture, because the operator cannot intuitively sense the interaction force between instruments and vascular tissues.

Therefore, the next **Chapter 5** proposes a guidewire endpoint detection model based on visual information, enhancing the perception of endovascular instruments (guidewire). By providing operators the relationship between tools and surrounding anatomical structures (branch vessel or stenosis) based on guidewire's position and bending angle, the proposed models compensate the lack of directly haptic force. This visual perception enhancement of vascular instruments is essential for improving manipulation accuracy, enabling safer navigation through complex vascular pathways.

Chapter 5: Visual Data Modelling of Instrument Endpoints for Enhancing Tool Manipulation

5.1 Introduction

Minimally invasive surgery commonly involves multiple tools guided by medical imaging. Expert operators significantly influence outcomes, relying on precise control over instruments. Accurate visual perception and real-time response systems are vital for managing robotic systems and executing tasks securely. Real-time processing of 2D or 3D images is essential for merging patient anatomy with images to accurately locate the surgical site, reducing the risk of tool collisions [230, 231]. Guidewires are critical in endovascular procedures, directed by live fluoroscopy images, but tracking them is challenging. Issues include limited visibility (only the last 3 cm is visible), simplistic shape, flexibility, and low signal-to-noise ratio with X-ray images causing navigation errors. Model-based learning systems are needed to accurately track guidewire endpoints. This chapter suggests an eight-neighbourhoods-based approach to enhance the precision of the guidewire endpoint localization. This framework addresses the challenges of tiny target detection in two stages: first, extracting the entire guidewire from X-ray images using an improved U-Net method and second, implementing guidewire endpoints detection based on the proposed eight-neighbourhood method after guidewire segmentation. An eight-neighbourhood-based method was designed to detect the endpoint of all guidewires, including the introduction of skeletonization extraction, removal of bifurcation points, repair of fracture points, and the design of method details.

5.2 Related Work

Numerous studies have focused on tool segmentation to facilitate and visualize tool navigation in X-ray angiograms. The medical guidewire, a key surgical instrument in interventional procedures, requires accurate end point location for effective navigation. Current computer vision approaches address challenges in the segmentation, detection, and localization of surgical instruments in medical images. These algorithms typically extract features hand-made from the images to identify and track surgical tools based on their appearance characteristics. However, such traditional methods are limited in their ability to obtain high-level semantic information from images, reducing their effectiveness in complex tasks.

In recent years, deep-learning methods have been increasingly applied to surgical tool segmentation [232, 233], detection [234], and key-point localization [235] with validation on endovascular datasets. The high-resolution network (HRNet), a heatmap estimation framework, is used in object detection and semantic segmentation to achieve high spatial precision [236, 237]. The HRNet model, a novel bottom-up approach to human pose estimation, enables precise location of key points for small objects [238]. Similarly, the Hourglass network, known for human pose

estimation, uses dense residual blocks within an encoder-decoder structure to enhance precision [239]. For guidewire detection tasks, which involve small, flexible structures within X-ray images, high detection precision is essential, particularly for the guidewire endpoint.

The rapid growth of surgical instrument research has led to the development of heatmap regression methods that offer new approaches to surgical tool detection. To improve the precision of small object detection, ResNet-50 is frequently used to locate key points of surgical tools [240]. ResNet-101 and ResNet-152, extensions of ResNet-50, generate high-quality object predictions [241-243]. The Hourglass network model, widely used in instance segmentation, performs detection and segmentation for individual instances in a unified model [244]. Zhou *et al.* [245] introduced an enhanced RetinaNet model, incorporating a ResNet-18-based encoder and a specialized subnetwork for needle detection. These deep-learning approaches significantly improve detection accuracy over traditional methods by extracting high-level semantic information from images.

Despite their advantages, direct application of these methods has limitations. First, these models were primarily designed for rigid tools in laparoscopic and retinal surgeries, where points on the tool maintain relatively fixed spatial relationships. This makes them unsuitable for detecting the guidewire tip, which has a flexible, radiopaque-coated endpoint. Second, the small and soft appearance of the guidewire differs from that of larger instruments such as catheters, forceps, or endoscopes. Consequently, current key-point detection methods for large targets are inadequate for the precise detection of guidewire endpoints because of the need for higher accuracy with tiny targets.

Existing literature reveals a limited focus on the location of the guidewire endpoint in X-ray images. For example, Cronin *et al.* [246] applied impedance-based electroanatomic mapping for real-time guidewire localization. Bedel *et al.* [247] used transthoracic echocardiography to optimize central venous catheter positioning, also employing point-of-care ultrasound to refine catheter tip localization [248]. Rapid advances in deep learning [249] have furthered the development of computer vision applications for guidewire localization [250-252]. Given the strong performance of CNN in medical image segmentation [253], commonly used guidewire localization models are often CNN variants [254]. Zhou *et al.* [255] developed a real-time multifunctional framework for morphological analysis of the guidewire using a fast attention recurrent network. Li *et al.* [168] improved guidewire localization accuracy by proposing a two-stage framework that first uses the YOLOv3 detector to detect the guidewire, followed by post-processing with a segmentation attention hourglass network for endpoint prediction. This approach yielded high performance; however, the original segmentation was modified to focus on large targets rather than small ones, creating a new dataset for localization by labelling the entire guidewire within a bounding box.

Li *et al.* [169] further proposed a CNN model based on the key-point localization region to detect and localize the guidewire endpoint. This approach relied on a bounding-box dataset of guidewires as the model input. Although these methods achieved efficient localization accuracy, they altered the scale of the guidewire within the overall image, overlooking the complexity and variability of

surgical procedures and shifting the focus from detecting small objects to large target detection tasks.

5.3 Methodology

In this chapter, we propose a novel detection framework for tiny targets, as illustrated in **Figure 5.1**. The framework addresses the challenges of detecting tiny targets in two stages: first, by extracting the entire guidewire from X-ray images using an improved U-Net method, and second, by detecting the guidewire endpoints with an eight-neighbourhood-based method after guidewire segmentation. When the guidewire is extracted from the complete X-ray image, both background (BG) removal and isolation of the guidewire (GW) are achieved, which diminishes visual interference and simplifies the issue created by the guidewire's straightforward appearance before detecting its endpoints.

With recent advancements in deep CNNs (DCNNs) in medical image processing [256, 257], we designed an improved U-Net network based on a CNN for the first stage of our framework. This network segments the guidewire from the full X-ray image, minimizing interference from background elements, and can also be applied to other instruments for key-point detection. In the second stage, an algorithm based on eight neighbourhoods is implemented for precise detection of tiny endpoints of the guidewire. This algorithm relies on spatial relationships between pixels in eight neighbouring regions, facilitating processing for guidewire key-point tasks, including 1) medial axis skeletonization, 2) bifurcation point removal, 3) breakage-band repair, and 4) endpoint detection.

A surgical guidewire tip is soft and flexible and forms complex shapes (twist and circle), which easily deforms the guidewire body such that it can no longer pass through the branch or stenosis path. In such a case, the operator can increase the risk of endovascular rupture if they attempt to pass the guidewire through this complex path. To mitigate this risk, we compute angle information based on the distance between each pixel and the line connecting the start and endpoint pixels after endpoint detection. This information helps inform operators or surgical robots of the current bending angle of the guidewire, guiding them to adjust their manipulation strategy and reduce surgical risk.

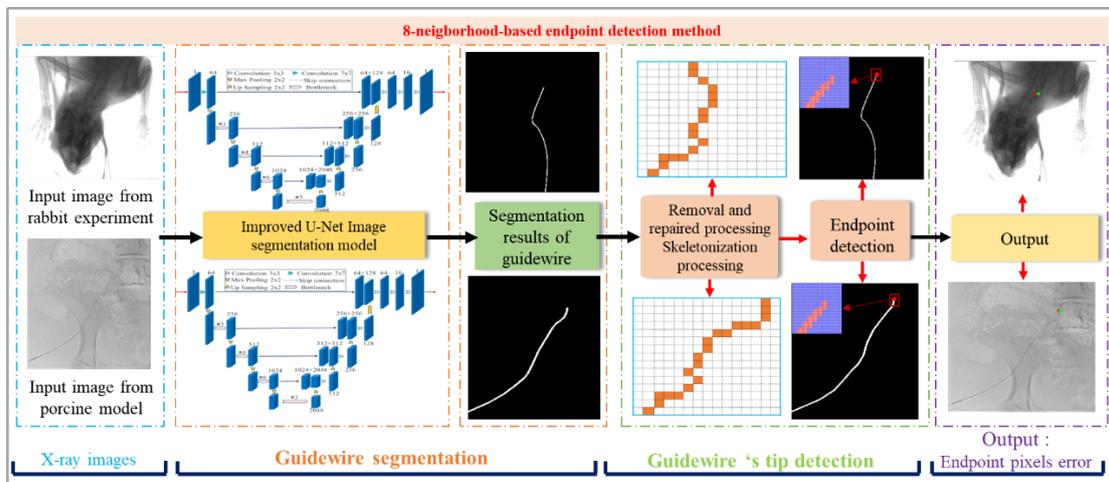


Figure 5.1: Overall framework for guidewire endpoint detection.

5.3.1 Dataset acquisition

The research datasets utilized in this chapter were gathered through the robot-assisted endovascular interventional platform outlined in **Chapter 2**. **Figure 5.2** shows the comprehensive framework for acquiring datasets. Two datasets were compiled: Dataset A, derived from in vivo experiments with a rabbit endovascular model, and dataset B, sourced from in vivo studies involving a porcine model.

Dataset A: The master-slave robotic system performed several in vivo interventional procedures by cannulating the endovascular pathway in six rabbits (average weight: 2.21 ± 0.29 kg), navigating a 0.014-inch guidewire along the auricle-to-coronary arterial path. All experimental protocols were approved by the Shenzhen Institutes of Advanced Technology (Application No. SIAT-IACUC-200528-YGS-WL-A1289). A proprietary X-ray machine [211] (a custom five DoF C-arm machine equipped with a 70 kV, 5 mA generator, and a flat detector) was used to capture real-time images during endovascular catheterization. This X-ray system produced sequences of fluoroscopic images at an average rate of 10 fps, resulting in a total of 3300 images from the six in vivo rabbit experiments, capturing the guidewire's motion trajectory. Each image had a resolution of 1560×1440 pixels at 96 dpi, yielding a spatial resolution of 0.26×0.26 mm². The guidewire trajectories in all images were annotated using the LabelMe interface and saved as JSON files for further processing. Furthermore, continuous CT images were used to verify the successful navigation of the guidewire to the target site within the endovascular pathways.

Dataset B: A porcine model (weight: 35.5 kg) was used to simulate the human renal artery system, closely replicating clinical practice. All other experimental setups mirrored those typically employed in clinical settings. The robotic slave system was operated under a commercial CT arm (digital subtraction angiography), providing real-time angiograms for visualization of the procedures. The commercial C-arm was configured to capture image sequences at an average rate of 15 fps, with a window level and width of 128 and 256, respectively. A total of 206 angiogram sequences were processed to create X-ray images, each transformed into a “.jpg” format with dimensions of 515×512 pixels at 96 dpi, resulting in a resolution of 0.26×0.26 mm². After filtering for images containing the guidewire trajectory, a total of 1880 effective X-ray images were selected for Dataset B. Guidewire trajectories were annotated using LabelMe and saved as JSON files. During catheterization with the vascular interventional robot, a contrast dye was injected to enhance angiographic views of both endovascular tools and endovascular anatomy. The 2D or 3D X-ray images, with or without angiographic subtraction, enabled the visualization of operator navigation. The experimental protocol was approved by the Shenzhen Advanced Animal Study Service Centre (No. AAS 191204P).

5.3.2 First stage: guidewire segmentation methods

The proposed guidewire endpoint detection framework consists of two stages: (i) guidewire segmentation and (ii) guidewire endpoint detection. In the first stage, an improved U-Net network is used to segment the entire guidewire from each X-ray image, producing a feature image.

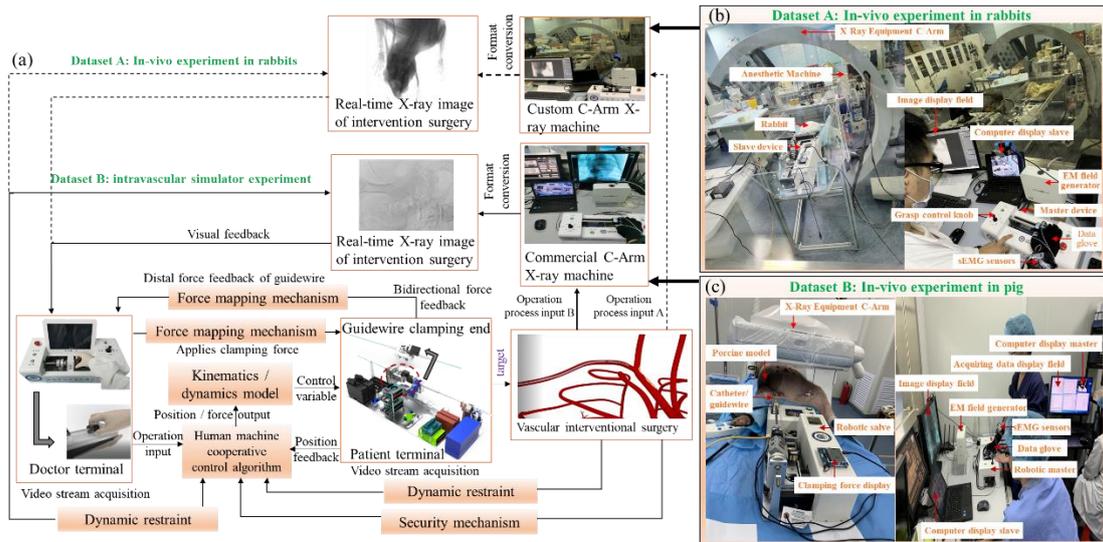


Figure 5.2: Image acquisition framework during robot-assisted endovascular interventional procedure.

A. Improving the U-Net network

Guidewire segmentation in X-ray images is particularly challenging due to the small size and flexibility of the target. In the first stage of our proposed approach, the entire guidewire is segmented as an object with a pixel value of 255, while the background is assigned a pixel value of zero. Although existing semantic segmentation networks developed for real-world and medical imaging scenarios address general segmentation needs, they often lack the accuracy needed for detecting tiny targets such as guidewires. The U-Net model, with its well-structured encoder-decoder architecture and skip connections [258] (illustrated in **Figure 5.3(a)**), provides a promising solution. U-Net++, an extension of U-Net, incorporates nested skip pathways at each network level [259], improving its applicability for small medical image datasets [260, 261].

In this study, the U-Net model was adapted for guidewire segmentation to capture high-level semantic information from X-ray images. To enhance performance, batch normalization was implemented throughout the U-Net network to standardize input values across layers, achieving a normalized distribution with a mean of 0 and a variance of 1. This adjustment accelerates convergence, reduces initialization requirements, and improves overall network performance.

Within the encoder stage, an improved convolution strategy was implemented to boost local feature extraction and strengthen global feature representation. Specifically, a convolution operation with padding was used to replace traditional convolution, preventing feature map size reduction and retaining more spatial information. Additionally, in the feature extraction stages (f2, f3, f4, and f5), we replaced the standard 3x3 convolution kernel with the Xception module, as shown in **Figure 5.3(b)**. Leveraging depthwise separable convolution, the Xception module not only reduces computational load but also enhances the fusion of local features with global context information by combining depthwise and pointwise convolutions. This integration makes the network more adaptable to features of varying scales, enhancing its ability to handle complex scenes while maintaining computational efficiency.

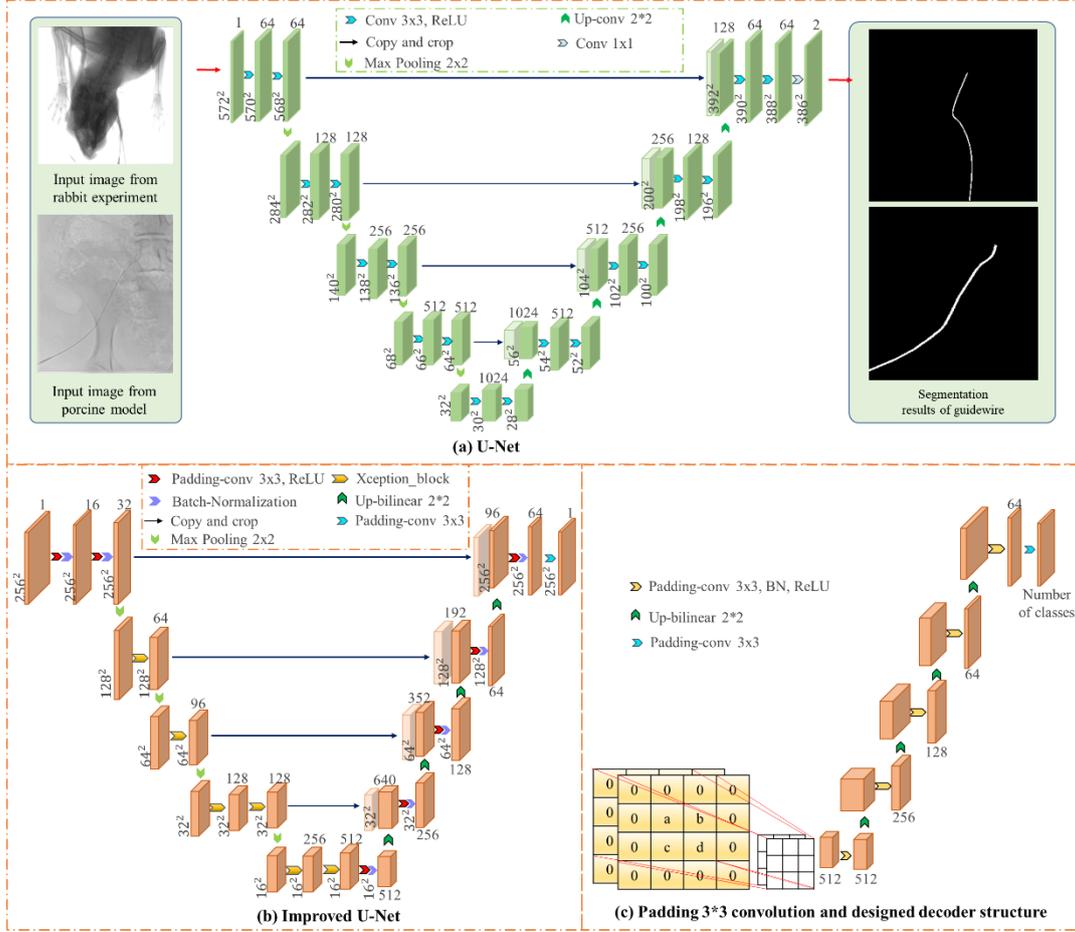


Figure 5.3: Improved U-Net model structure.

Further modifications were made to the decoding structure. The original U-Net uses cropping to match the encoding and decoding network sizes for concatenation, which can lead to information loss. In our enhanced U-Net, we replaced the original 3x3 convolution with a 3x3 padding convolution in both the encoding and decoding stages to maintain consistent feature map sizes without cropping. Additionally, 3×3 padding convolutions were introduced to reduce the number of channels to one-fourth of those in the original U-Net, and the middle convolution layer in the decoding structure was removed to reduce network redundancy, as shown in **Figure 5.3(c)**.

Finally, a 3x3 convolution with zero padding was applied after each bilinear up-sampling step, and once more in the final layer, to serve as input to the SoftMax layer. This approach optimizes the limited receptive field of the preceding 1x1 convolution, enhancing segmentation precision.

B. Design of decoding network

The activation function is a crucial component in connections within convolution layers. Compared to the sigmoid function, the rectified linear units (ReLU) activation function [262] accelerates the convergence speed and prevents the gradient from vanishing when it is greater than zero. Additionally, ReLU has low computational complexity, making it an effective choice for improving the decoding block of a network. The ReLU activation function is defined as follows:

$$ReLU = \max(0, x) \quad (5.1)$$

The output size of the convolution layer is determined by:

$$Output_{size} = \frac{(I_{size} - K_{size} + 2P_{size})}{S_{size}}, \quad (5.2)$$

where W , K , P , S and O represent the input feature size, kernel size, padding size, stride, and output feature size, respectively. The convolution process is illustrated in **Figure 5.3(c)**, where $p = 1$, $k = 3$, and $s = 1$. If the output is set to the input, applying a 3×3 convolution with padding can generate equally sized input and output feature maps. This convolution operation in the decoding process is used to resize the feature map to its original dimensions and prevent information loss. The decoding network structure is shown in **Figure 5.3(b)**.

During the convolution process, the number of convolution kernels is significantly reduced, and convolution layers with the same number of channels are not stacked in the up-sampling process. This design minimizes the number of network parameters to be trained, enhancing efficiency. Furthermore, the encoder part of the improved U-Net integrates elements of ResNet50 [256] to enhance the segmentation performance. ResNet50 consists of identity and convolutional blocks. By removing the fully connected layer of ResNet50 and integrating the remaining layers into the improved U-Net, we obtain an optimized U-Net model.

5.3.3 Second stage: guidewire endpoint detection

In the second stage, the segmented guidewire is analysed using our proposed eight-neighbourhood algorithm to accurately detect the guidewire endpoint. Guidewire endpoints represent the connections between different pixels [263]; therefore, we propose an eight-neighbourhood endpoint detection method that takes advantage of these connections. This method tracks the guidewire endpoint by examining pixel points within eight neighbouring relationships. Each pixel in the binarized image is assigned a value of either 0 or 255, where 0 represents the black background and 255 represents the target guidewire. Each pixel has up to eight neighbouring pixels, forming the basis for the eight-neighbour algorithm.

Following binarization, the entire guidewire skeleton is extracted based on these pixel points. However, the segmented guidewire often exhibits non-smooth characteristics, resulting in discontinuities or fractures between the start and endpoints. Despite skeleton extraction, the guidewire may remain fragmented. To address this, we implemented post-skeleton processing, which involves removing branch pixels and repairing fractured pixels to preserve the continuity of the guidewire skeleton across consecutive pixels. Finally, using the eight-neighbourhood algorithm, we completed the guidewire endpoint detection. The overall framework is illustrated in **Figure 5.4**.

A. Guidewire skeletonization extraction

Skeletonization captures the topological structure of an object and is used in image recognition to reduce redundancy and eliminate extraneous information. Skeletonization extraction is defined as the removal of boundary pixels while preserving the connectivity of the image. The objective is to reduce a connected region to a single-pixel width, where the skeleton is mathematically defined as the set of central points of the maximum tangent hypersphere along the boundary. This process

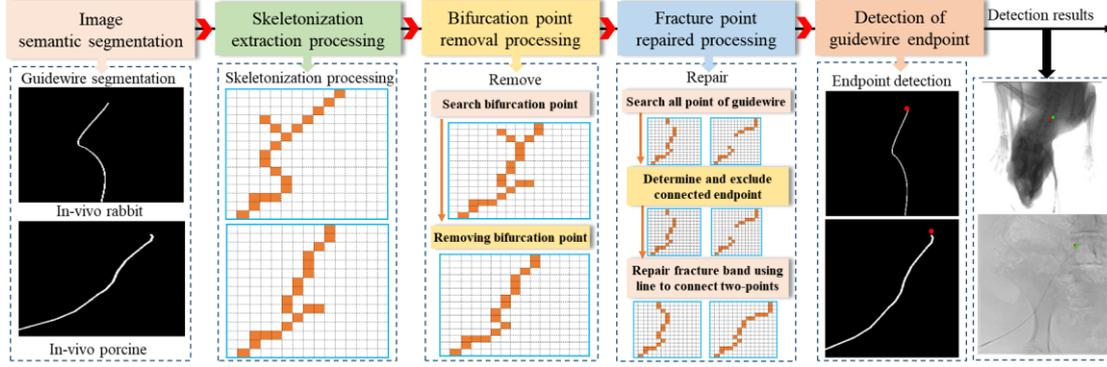


Figure 5.4: Guidewire endpoint detection based on pixel-adjacent-relation method.

can be envisioned as if the target’s edge lines are uniformly illuminated and a fire front propagates evenly towards the interior. As the fronts intersect, the “flame” extinguishes, and the union of these extinguished points forms the skeleton.

The skeletonization procedure begins by assuming that a white pixel (value 255) represents the object, and a black pixel (value 0) represents the background. For each pixel in the image, we check its surrounding pixels in all directions. If the current point of the pixel (x, y) is a white pixel, we evaluate four neighbouring pixels: $[(x - 1, y + 1), (x - 1, y), (x - 1, y - 1), \text{ and } (x, y - 1)]$. The current pixel point is assigned a new layer value of 0 if it is a black pixel or an incremented layer value based on the minimum value of its neighbours if it is white. This procedure is defined as follows:

$$F(x, y) = \begin{cases} \min \begin{pmatrix} F(x - 1, y + 1), F(x - 1, y), \\ F(x - 1, y - 1), F(x, y - 1) \end{pmatrix} + 1, & \text{if } P_1(x, y) = 255 \\ 0, & \text{if } P_1(x, y) = 0 \end{cases} \quad (5.3)$$

The second step of this algorithm mirrors the first. We scan each pixel from the bottom up and from right to left, evaluating the same neighbouring pixels around each point (x, y) . The process for this layer is defined as follows:

$$G(x, y) = \begin{cases} \min \begin{pmatrix} G(x - 1, y + 1), G(x - 1, y), \\ G(x - 1, y - 1), G(x, y - 1) \end{pmatrix} + 1, & \text{if } P_2(x, y) = 255 \\ 0, & \text{if } P_2(x, y) = 0 \end{cases} \quad (5.4)$$

The first step calculates the value of the upper enclosing layer, while the second step calculates the value of the lower enclosing layer. The final layer value for each pixel is the minimum of the values from these two enclosing layers:

$$M(x, y) = \min(F(x, y), G(x, y)) \quad (5.5)$$

In the last step, the layer values of the eight neighbouring pixels around each pixel are compared to the current pixel value. If the current pixel layer value is the maximum, it is retained; otherwise, it is removed. The result of skeletonization extraction is illustrated in **Figure 5.5**. After semantic segmentation, the guidewire is processed using this skeletonization method to obtain a topological structure that captures essential shape features, such as intersections, inflection points, and fracture points.

B. Removing of pixel bifurcation points

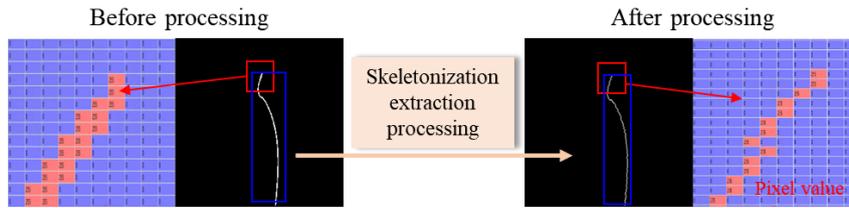


Figure 5.5: Overall processing of skeletonization extraction.

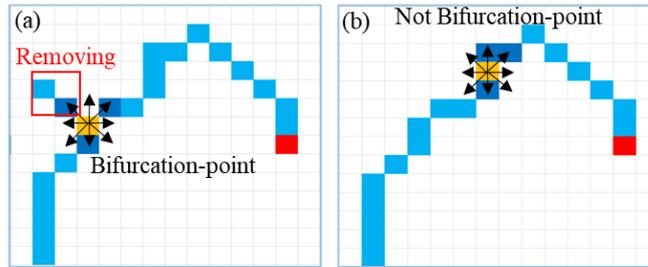


Figure 5.6: Detection and removal of bifurcation point.

The extracted guidewire often appears irregular after semantic segmentation, resulting in bifurcation points after skeletonization. To accurately detect the endpoint of the guidewire, these bifurcation points must be removed from the skeletonized feature map. **Figure 5.6** shows that bifurcation points are defined by three pixels within the eight-neighbourhood area, with each pixel linking to another pixel outside this area, creating branch paths through the joining of neighbouring pixels across different layers. In comparison, non-bifurcation points are linked by just one or two pixels.

This structural distinction allows us to identify and eliminate bifurcation points efficiently. Upon locating a bifurcation point, we computed the concatenated paths in three directions within the eight-neighbourhood region. **Figure 5.6(a)** illustrates these paths depicted with red, blue, and green dotted lines. The abnormal branch at the bifurcation point is recognized as the shortest path, which is subsequently eliminated to uncover the actual guidewire endpoint.

C. Repairment of pixel fracture points

The precision of the semantic segmentation method directly affects the quality of the extracted guidewire feature map; higher accuracy results in fewer fracture points in the segmented guidewire. Although our proposed segmentation method outperforms other semantic segmentation techniques, some fracture points still appear in the guidewire after segmentation. These fractures persist even after skeletonization. Therefore, a repair process was implemented to address all fracture points, from the initial point to the final pixel, with the aim of restoring guidewire connectivity, a critical factor for accurate endpoint detection.

To begin with, each point on the pixel was evaluated to determine if it was a fracture point. If identified as a fracture, the two adjacent pixel points around it were connected to ensure continuity in the guidewire. The detailed repair process is illustrated in **Figure 5.7**. Each pixel points on the

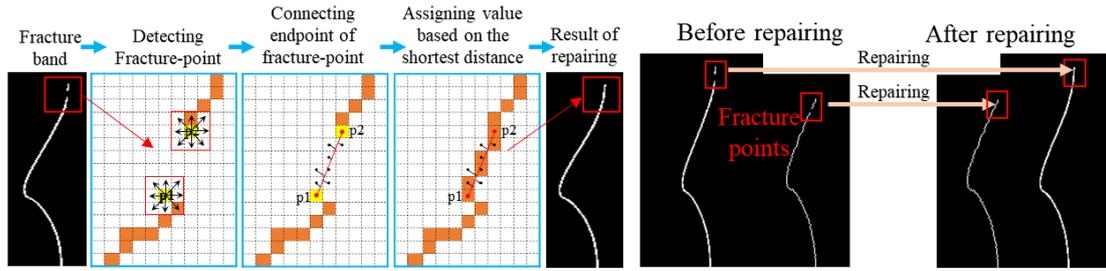


Figure 5.7: Overview process of the repairment of fracture points.

guidewire skeletonization feature map was examined sequentially from bottom to top and from left to right to assess the number of connected pixels in its eight-neighbourhood. If the current pixel had more than one connected neighbour, it was not a fracture point; if only one connected neighbour was found, it indicated a fracture.

In this case, pixel P_1 represented one endpoint of the fracture, and pixel P_2 represented the other. To repair the fracture, P_1 and P_2 were connected with a straight line that traversed background pixels. Background pixels closest to the line, based on the shortest center-to-line distance, were reassigned a pixel value of 255 to mark them as guidewire points, thereby restoring connectivity.

D. Guidewire feature map detection

The result of the feature map of the guidewire after semantic segmentation indicated that the detection stage was responsible for predicting the coordinates of the endpoints of each guidewire. First, skeletonization processing was applied in the feature map of the guidewire to obtain the skeletonization structure of the guidewire. The bifurcation point in this structure was removed and the fracture points were repaired. The given guidewire feature map was used as input to complete further detection processing. The guidewire feature map used skeletonization and defect interpolation processing, and therefore there is only one pixel connected to the endpoint in the eight-neighbourhood region of the endpoint and two-pixel points connected to the non-endpoint in the

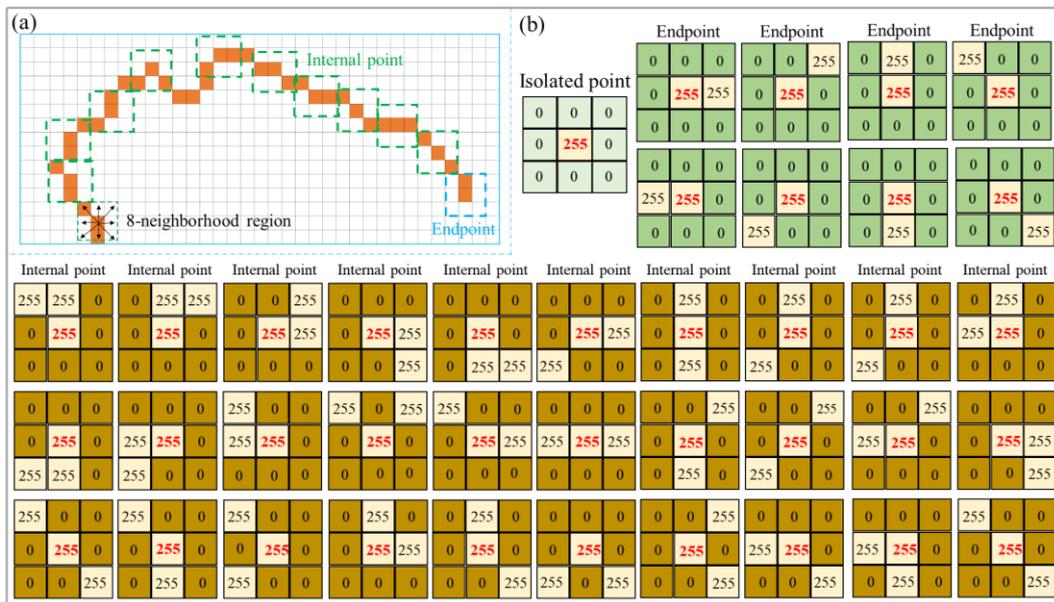


Figure 5.8: Adjacent point relationships within the eight-neighbourhood scope.

eight-neighbourhood region of the non-endpoint pixel (**Figure 5.8**). The eight-neighbourhood detection method based on this idea used the output feature maps of the removal and repair processing as input to scan each pixel to detect the guidewire endpoint.

E. Guidewire bending detection

In endovascular interventional surgery, the guidewire, though tiny and flexible, can cause mechanical damage to the arterial lining when it bends excessively. This damage can lead to platelet aggregation and stimulate intimal hyperplasia, increasing the risk of endovascular rupture. To mitigate this risk, we calculate the bending angle of the guidewire based on its current pose. This angle provides the operator with valuable information to adjust the control strategy and reduce the potential damage to the vessel.

To compute the bending shape of the guidewire, we calculate the bending angle using the connecting line between the starting point $A(x1_i, y1_j)$ and the endpoint $B(x2_i, y2_j)$. The process begins by performing a reverse-to-forward search to identify the starting pixel. Once identified, the line is drawn connecting the starting point to the endpoint. Next, we calculate the perpendicular distance from each pixel along the guidewire to this line. The pixel with the greatest distance from the line is identified as the highest point, denoted as $C(x3_i, y3_j)$. The bending angle is then determined by the angle formed at point C between the line segments from A to B and from C to B . This angle represents the degree of bending of the guidewire, as shown in Eq. (5.6-5.9).

$$a = \sqrt{(x2_i - x3_i)^2 + (y3_j - y2_j)^2} \quad (5.6)$$

$$b = \sqrt{(x3_i - x1_i)^2 + (y3_j - y1_j)^2} \quad (5.7)$$

$$c = \sqrt{(x2_i - x1_i)^2 + (y2_j - y1_j)^2} \quad (5.8)$$

$$angle = \cos^{-1}\left(\frac{a^2 + b^2 - c^2}{2ab}\right) \quad (5.9)$$

In the calculation of the bending angle, let a represent the length of the line connecting the highest point $C(x3_i, y3_j)$ to the endpoint $B(x2_i, y2_j)$, b represent the length of the line connecting the highest point $C(x3_i, y3_j)$ to the start point $A(x1_i, y1_j)$, and c represent the length of the line connecting the start point $A(x1_i, y1_j)$ to the endpoint $B(x2_i, y2_j)$. The points $(x1_i, y1_j)$, $(x2_i, y2_j)$, and $(x3_i, y3_j)$ are pixel coordinates, where $i, j \in (1, 2, \dots, m)$, with m being the total number of pixels in the image.

The bending pose of the guidewire can be evaluated using both the longest distance and the bending angle, which serve as crucial reference points for the operator to adjust the manipulation strategy as shown in **Figure 5.9**. Using the methods described above, we perform a sliding window scan across the entire guidewire segment, where the length of the scanned guidewire and the scanning interval are key parameters. During each scan, the following steps are performed:

- 1) Calculate the included angle for all segments of the guidewire.
- 2) Measure the distance from each pixel of the guidewire to the corresponding straight line, determining the point with the maximum distance, known as the “highest point”.

Compute the bending angle based on the positions of the start point, endpoint, and the highest pixel.

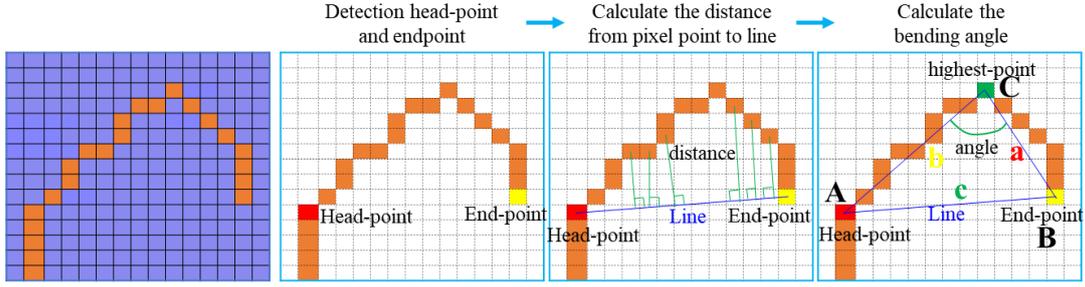


Figure 5.9: Overview process of determining the maximum bending region and angle values.

5.3.4 Training strategy and evaluation index

A. Training strategy

The improved U-Net model, referred to as JSUNet due to its J-shaped decoder network design, utilizes the categorical cross-entropy loss function [242], which is defined as:

$$Loss = -\sum_{i=1}^n \hat{y}_{i1} \log y_{i1} + \hat{y}_{i2} \log y_{i2} + L + \hat{y}_{im} \log y_{im}, \quad (5.10)$$

where n denotes the number of samples, m represents the number of classes (with $m \geq 2$), \hat{y}_{im} is 1 if sample i belongs to category m and 0 otherwise, and y_{im} denotes the probability that sample i is predicted to belong to class m .

The experimental setup to train the networks included a Windows 10 operating system, an NVIDIA GeForce® GTX 1080 graphics card, the NVIDIA CUDA 10.1 acceleration toolkit, and a Pytorch-based Keras interface as a training framework. The experiments utilised two self-constructed datasets: Dataset A, comprising 3300 X-ray images, and Dataset B, consisting of 1800 fluoroscopy images. Each dataset was divided into training, validation, and test sets in an 8: 1: 1 ratio. Specifically, Dataset A (in vivo rabbit model) included 2640 images for training (173,015,040 pixels), 330 images for validation (21,626,882 pixels), and 330 images for testing (21,626,882 pixels). Similarly, Dataset B (in vivo porcine model) included 1440 images for training (94,371,840 pixels), 180 images for validation (11,796,482 pixels) and 180 images for testing (11,796,482 pixels). All images were in “.png” format with a resolution of 256×256 pixels.

Parameter initialization employed the Xavier normal distribution to generate random tensors. The Adam optimizer [243] was selected due to its efficiency in deep learning applications. During training, if the validation loss did not decrease after every 3 epochs, the learning rate was halved to facilitate convergence. Each dataset was trained for a total of 50 epochs, with the model achieving the lowest validation loss selected for segmentation and prediction comparisons.

B. Performance evaluation index

The intersection-over-union (IoU) is commonly used as a performance evaluation metric to measure the overlap between the segmentation results for each category and the original labelled image. MIoU is also employed to assess the network model’s effectiveness in semantic

segmentation, calculated as:

$$MIoU = \frac{1}{n+1} \sum_{i=0}^n \frac{q_{ii}}{\sum_{j=0}^n (q_{ij} + q_{ji} - q_{ii})} \quad (5.11)$$

where $n + 1$ represents the number of pixel categories in the image, q_{ii} is the total number of correctly classified guidewire pixels, q_{ij} represents the total number of guidewire pixels classified as background, and q_{ji} represents the total number of background pixels misclassified as guidewire pixels.

Additionally, *precision*, *recall*, *F1 – score*, and *macro – F1* are used as performance metrics for segmentation accuracy, defined as follows:

$$recall_i = \frac{TP_i}{TP_i + FN_i} \quad (5.12)$$

$$precision_i = \frac{TP_i}{TP_i + FP_i} \quad (5.13)$$

$$F1 - score_i = 2 \frac{recall_i \times precision_i}{recall_i + precision_i} \quad (5.14)$$

$$macro - F1 = \frac{\sum_{i=1}^n F1 - score_i}{N} \quad (5.15)$$

In these equations, TP denote correctly detected guidewire pixels, FP represent background pixels misclassified as guidewire pixels, and FN are guidewire pixels incorrectly classified as background. The *F1 – score* balances model accuracy and recall, which is critical in medical image segmentation where missing or extra foreground regions can impact clinical outcomes.

To evaluate the model's processing speed, the forward feedback processing speed (FFPS) is calculated as:

$$FFPS = \frac{1}{N} \sum_{m=1}^N \frac{1}{time_m} \quad (5.16)$$

where N is the total number of test samples, m denotes the $m - th$ sample index, and $time_m$ is the forward feedback processing time for each sample. FFPS, measured in frames per second, indicates the efficiency of forward feedback propagation within the network model.

The mean pixel error (MPE) is used to assess the accuracy of endpoint detection, calculated as the average Euclidean distance between the predicted and ground-truth pixel points:

$$MPE = \frac{1}{N} * \sum_{i=1}^N (\|P_i - G_i\|) \quad (5.17)$$

where N is the total number of test samples, $\| \cdot \|$ is the total number of test samples, P_i is the predicted endpoint pixel for the $i - th$ sample, and G_i is the ground-truth endpoint pixel.

5.4 Model Detection Performances

5.4.1 Segmentation performance

Semantic segmentation was applied to guidewire extraction to enhance its suitability for guidewire endpoint detection in robot-assisted interventions. The improved U-Net model, known as JSUNet, was used to evaluate segmentation performance by processing images of tool trajectories from these interventions. After segmentation, the results of the proposed method were validated

using training data from Datasets A and B, with frame intervals of 10 fps for Dataset A and 15 fps for Dataset B.

The performance of the improved U-Net model, measured in terms of training accuracy, training loss, and validation accuracy and loss, is depicted in **Figure 5.10(a)** and (b). The model was trained for 50 iterations. For Dataset A, training accuracy started at 60% in the first iteration and reached a maximum accuracy of 83.45% by the 40th iteration. Dataset B showed similar performance, with the accuracy rate remaining stable from the 35th iteration onward. Training loss began at 0.4 in the initial iteration and decreased to 0.2 by the 50th iteration, indicating good model characteristics. The improved U-Net model achieved validation accuracies of 89.46% and 92.43% on Datasets A and B, respectively. Validation loss decreased rapidly within the first 10 epochs for Dataset A and the first 15 epochs for Dataset B, reflecting effective learning within those epochs. Throughout training and validation, the model demonstrated stability, indicating that it did not overfit, as evidenced by the alignment between training and validation losses.

The performance of the improved U-Net model was further validated in the test Datasets A and B in **Figure 5.10(a')** and (b'). Confusion matrix analysis showed that the model accurately segmented guidewire pixel data, achieving Macro-F1 scores of 94.57% on the rabbit data set and 95.48% on the porcine dataset. Detailed analysis of the confusion matrix revealed high recall and precision for both background and guidewire pixels: for the rabbit dataset, recall and precision reached 99.94% and 99.95% for background pixels, and 90.53% and 87.87% for guidewire pixels, respectively. For the porcine dataset, the background pixel prediction achieved 99.89% recall and 99.91% precision, while the guidewire pixel prediction achieved 92.07% recall and 90.07% precision.

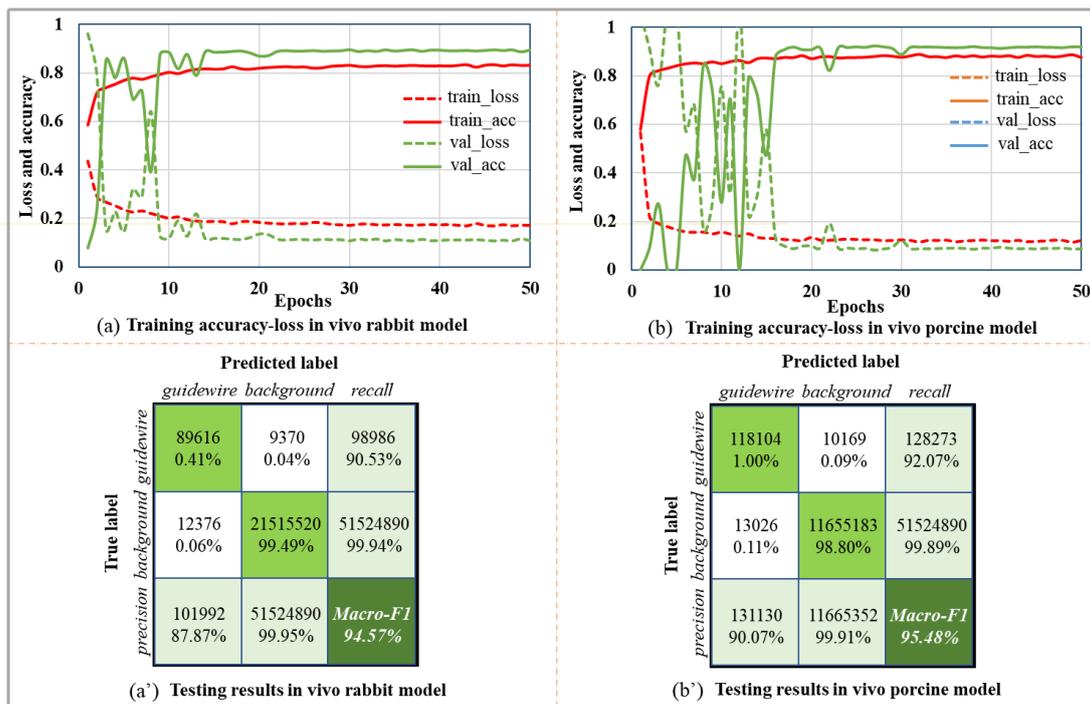


Figure 5.10: Performance analysis of proposed improved U-Net model.

Table 5.1: Performance evaluation of different models for guidewire segmentation.

Network model	IoU [%]			Macro-F1 [%]	FFPS [fps]	Dataset	Pars. No [M]
	BG	GW	MIoU				
U-Net	99.54	52.93	51.35	85.47	51.35	A	7.76
	99.73	60.13	39.32	87.34	39.32	B	
U-Net++	99.86	74.92	87.39	92.8	38.32	A	9.04
	99.7	74.78	87.24	92.71	28.14	B	
DeepLabV3+	99.89	78.86	89.38	94.07	31.39	A	6.44
	99.69	73.51	86.60	92.29	30.57	B	
Improved U-Net	99.90	80.47	90.19	94.57	45.79	A	8.58
	99.80	83.58	91.69	95.48	31.40	B	

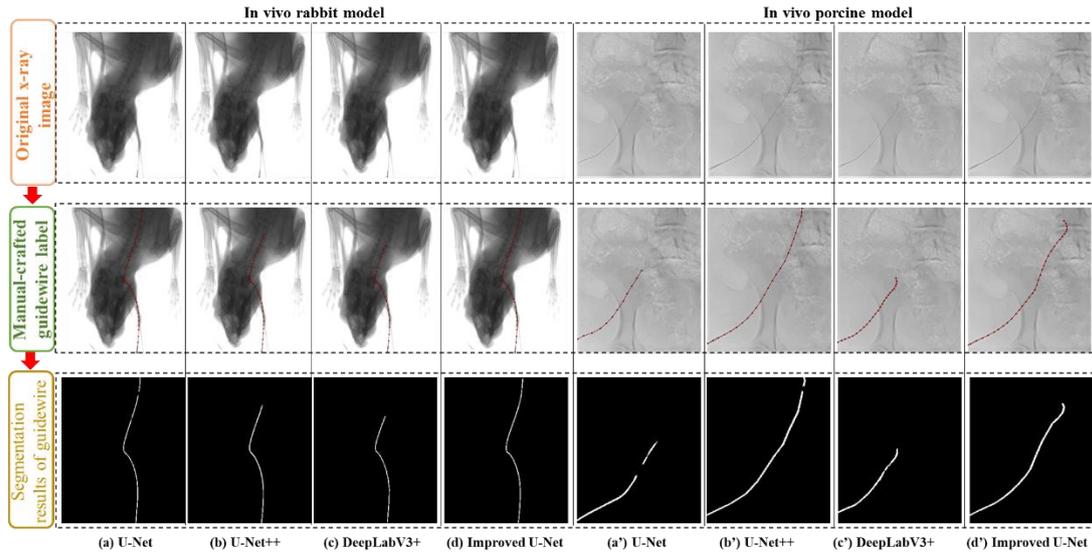


Figure 5.11: Guidewire segmentation results using the test set from in vivo rabbit model.

To comprehensively assess the improved efficiency of U-Net, a comparison was made with three other models: U-Net, U-Net++, and DeepLabV3+ in the datasets. Evaluation metrics that include MIoU, FFPS and model parameters were recorded during training on both datasets, with the results summarized in **Table 5.1**. The improved U-Net achieved an MIoU of 99.90% for background pixels and 80.47% for guidewire pixels, resulting in an overall MIoU of 90.19% for Dataset A. For Dataset B, it achieved 83.58% for guidewire pixels and 99.80% for background pixels, with an overall segmentation accuracy of 91.69%. Compared to the original U-Net, the improved U-Net demonstrated a significant performance increase, with MIoU increments of 13.95% for Dataset A and 11.76% for Dataset B. U-Net++ performed better than the base U-Net, but the improved U-Net still showed an MIoU increase of 2.8% for Dataset A and 4.45% for Dataset B compared to U-Net++.

Moreover, the improved U-Net model also required fewer parameters than the classic U-Net and U-Net++ models, although DeepLabV3+ required the fewest parameters at approximately 6.44 million. The improved U-Net used 8.85 million parameters, striking a balance between memory efficiency and performance. It outperformed the other models in MIoU and Macro-F1 metrics, achieving feedforward image processing speeds of 45.79 fps on Dataset A and 31.40 fps on Dataset B. **Figure 5.11** displays the segmentation results for the improved U-Net and the three comparison

models. Given the small target size of the guidewire relative to the background, accurately segmenting guidewire pixels remains challenging. The high accuracy achieved by the improved U-Net model demonstrates its effectiveness in using batch normalization and an enhanced decoding structure to accurately segment the guidewire in X-ray images, ultimately improving the performance of the model.

5.4.2 Comparison of detection performance for heatmap methods

After completion of guidewire segmentation, the guidewire feature maps were processed in the second stage to detect the endpoint. Using the eight-neighbourhood characteristics of the pixel distribution, skeletonization extraction was applied to the feature maps to derive the topological structure of the guidewire in pixel form. Subsequently, abnormal branches were removed, and fracture bands were repaired within the guidewire topology to accurately identify the ground truth endpoints. The performance of the eight-neighbourhood-based detection method was then assessed and compared with existing methods using the MPE metric, evaluated in data sets A and B. For comparison purposes, four standard approaches, including Hourglass, PoseResNet, HRNet, and HigherHRNet, were implemented and evaluated for their performance in addition to the newly introduced method, as shown in in **Table 5.2**.

The results indicate that our method achieved the best performance, with an MPE of 2.02 (± 0.14) pixels in Dataset A and 2.13 (± 0.37) pixels in data set B. For the rabbit dataset, the MPEs of the four other methods were as follows: Hourglass achieved 4.08 (± 0.36) pixels, PoseResNet showed the lowest endpoint detection accuracy with an MPE of 6.46 (± 0.63) pixels, HRNet achieved 5.38 (± 0.54) pixels, and HigherHRNet reached 5.74 (± 0.63) pixels. Similar comparisons were made for Dataset B, with our method achieving an MPE of 2.13 pixels in guidewire endpoint detection, outperforming the other methods. Specifically, in the porcine dataset, the MPEs for the other methods were as follows: Hourglass at 2.86 (± 0.09) pixels, PoseResNet at 3.46 (± 0.13) pixels, HRNet at 3.06 (± 0.08) pixels, and HigherHRNet at 3.13 (± 0.18) pixels, as shown in **Figure 5.12(a)**.

To provide a more intuitive representation of the model’s performance in guidewire endpoint detection, pixel-level detection errors were converted into millimetre-based errors, considering the physical dimensions of the images. This conversion is illustrated in **Figure 5.12(b)**. Our proposed method achieved the lowest average error, with a 0.53 mm endpoint detection error for the rabbit

Table 5.2: Comparison of four typical heatmap detection methods.

Methods	Dataset A: In vivo rabbit model		Dataset B: In vivo porcine model	
	MPE [pixel] / [mm]	p-value (vs. our)	MPE [pixel] / [mm]	p-value (vs. our)
Hourglass	4.08 \pm 0.36 / (1.27 \pm 0.09)	1.0332E-9	2.86 \pm 0.09 / (0.76 \pm 0.02)	0.000106
PoseResNet	6.46 \pm 0.63 / (1.71 \pm 0.17)	2.9648E-9	3.46 \pm 0.13 / (0.92 \pm 0.03)	2.7605E-7
HRNet	5.38 \pm 0.54 / (1.42 \pm 0.14)	8.9866E-9	3.06 \pm 0.08 / (0.81 \pm 0.02)	0.000015
HigherHRNet	5.74 \pm 0.63 / (1.52 \pm 0.17)	1.488E-8	3.13 \pm 0.18 / (0.83 \pm 0.05)	0.000003
Our	2.02 \pm 0.04 / (0.53 \pm 0.01)	-	2.13 \pm 0.37 / (0.56 \pm 0.09)	-

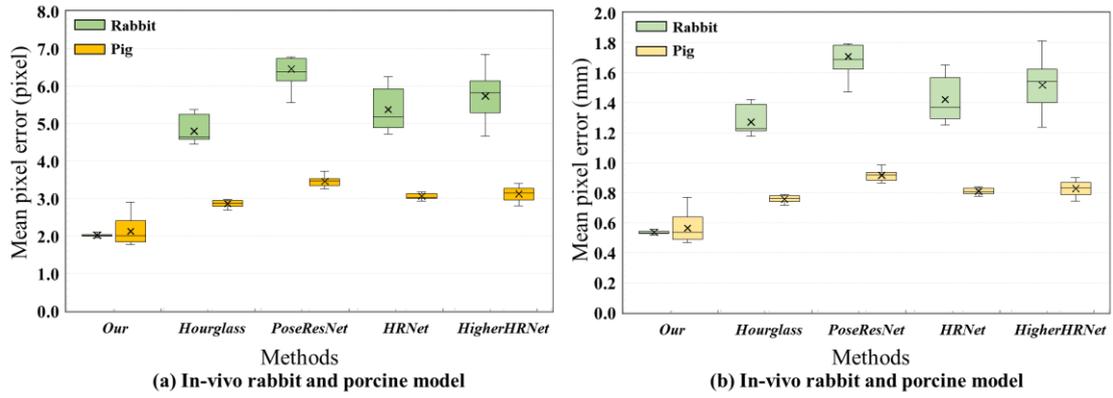


Figure 5.12: Mean pixel/distance error of four typical heatmap detection methods.

experiment and a 0.56 mm error for the porcine experiment, outperforming the other four heatmap-based methods. **Figure 5.12(b)** presents the visual results of the detection of guidewire end points for all five models.

A T-test was conducted to determine whether the MPE differences between Hourglass, PoseResNet, HRNet, HigherHRNet, and our method were statistically significant for Datasets A and B. As shown in **Table 5.2**, all p-values were less than 0.05, indicating that our method is significantly better than the four other typical heatmap-based methods. These findings suggest that the eight-neighbourhood method provides significantly improved detection accuracy compared to other learning-based methods. In contrast, other guidewire endpoint detection methods tend to have higher mean pixel errors, as they do not establish a close relationship between endpoint detection and the detection needs for endovascular interventional instruments.

Compared to the Hourglass network, our proposed method demonstrated superior performance in guidewire endpoint detection. While the Hourglass network, based on heatmap methods, is effective for human pose estimation and outperforms other models such as PoseResNet, HRNet, and HigherHRNet in endpoint detection, it is not optimal for detecting the tiny guidewire endpoint. This limitation arises from repeated bottom-up and top-down convolution operations in DCNNs, which significantly reduce the initial image resolution and hinder the accurate localisation of small objects such as the guidewire endpoint. The PoseResNet method, commonly used for heatmap extraction in human pose estimation tasks, addresses network model degradation issues, but is less effective for detecting small targets. Our results show that PoseResNet exhibited the largest performance gap compared to our method, even more so than the differences with Hourglass, HRNet, and HigherHRNet. This finding suggests that while PoseResNet is highly sensitive to human pose estimation, it lacks the precision needed to detect small objects such as guidewire endpoints.

HRNet, however, performed better than other ResNet models in guidewire endpoint detection, highlighting the importance of high-resolution representations in tasks requiring precise localization. Unlike other models such as PoseResNet, ResNet-50, ResNet-101, ResNet-152, and VGGNet, which encode input images into low-resolution representations before recovering higher resolution through high-to-low-resolution convolution connections, HRNet preserves high-resolution-

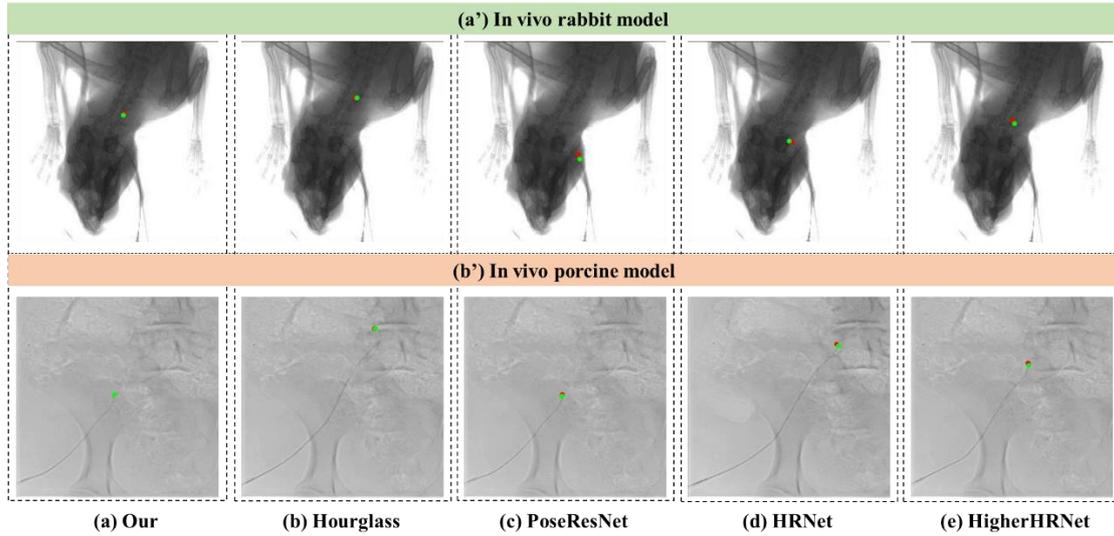


Figure 5.13: Guidewire endpoint detection results of four heatmap methods.

representations throughout the network. This continuity in high-resolution processing appears crucial for accurately localizing the guidewire endpoint. HigherHRNet, which combines high-resolution feature pyramids to learn scale-aware representations, was also tested for endpoint detection by estimating feature maps from the combined high-resolution representations. However, its performance was like or slightly worse than HRNet in this task. These findings suggest that while heatmap and heatmap regression methods, such as Hourglass, PoseResNet, HRNet, and HigherHRNet, are well suited for human pose estimation, they are less effective for detecting the tiny guidewire endpoint, where precise, high-resolution representation is essential.

5.4.3 Performance comparison of heatmap regression methods

To evaluate the performance of the proposed method, we also applied six existing heatmap regression methods from previous studies on Datasets A and B. Details of the implementation and validation for these existing methods are provided in **Table 5.3**. The results indicate that the MPEs of the six regression methods are significantly higher than those achieved by our method. Although the lightweight model MobileNetv2 reduces the number of parameters and computations, it shows the worst guidewire endpoint detection performance on both datasets, with MPEs of 13.07 ± 1.43 and 11.09 ± 0.74 for Datasets A and B, respectively. DenseNet121, which also has a low parameter count, performed somewhat better, achieving MPEs of 7.63 ± 1.94 and 8.24 ± 1.42 on Datasets A and B, respectively. Hourglass and ResNet50 produced comparable detection results. In particular, the deeper 101-layer and 152-layer ResNet architectures did not outperform 50-layer ResNet on either dataset. The results for ResNet-101 were slightly better than those for ResNet-152, suggesting that deeper layers do not improve performance in guidewire endpoint detection tasks. Furthermore, ResNet50 outperformed ResNet-101, indicating that the guidewire endpoint detection task is relatively simple and does not require a deeper network to extract additional information for successful detection. In general, the accuracy of the endpoint detection of the guidewire is significantly improved by using this two-stage detection framework.

Table 5.3: Comparison of six typical endpoint detection methods.

Methods	In vivo rabbit model		In vivo porcine model	
	MPE [pixel]/[mm]	p-value (vs. our)	MPE [pixel]/[mm]	p-value (vs. our)
ResNet50	5.90 ± 0.43 / (1.56 ± 0.11)	3.19E-10	7.71 ± 0.98 / (2.04 ± 0.26)	2.05E-9
ResNet101	6.73 ± 1.10 / (1.78 ± 0.29)	2.75E-7	8.69 ± 1.60 / (2.30 ± 0.42)	1.86E-7
ResNet152	8.24 ± 3.51 / (2.18 ± 0.93)	3.35 E-4	10.54 ± 1.28 / (2.79 ± 0.34)	1.08E-9
Hourglass2	5.76 ± 0.32 / (1.52 ± 0.09)	2.43E-11	7.35 ± 1.79 / (1.94 ± 0.47)	0.50E-5
DenseNet121	7.63 ± 1.94 / (2.02 ± 0.51)	0.70E-5	8.24 ± 1.42 / (2.18 ± 0.38)	1.02E-7
MobileNetv2	13.07 ± 1.43 / (3.46 ± 0.38)	1.52E-9	11.09 ± 0.74 / (2.93 ± 0.20)	2.92E-14
Our	2.02 ± 0.04 / (0.53 ± 0.01)	-	2.13 ± 0.37 / (0.56 ± 0.09)	-

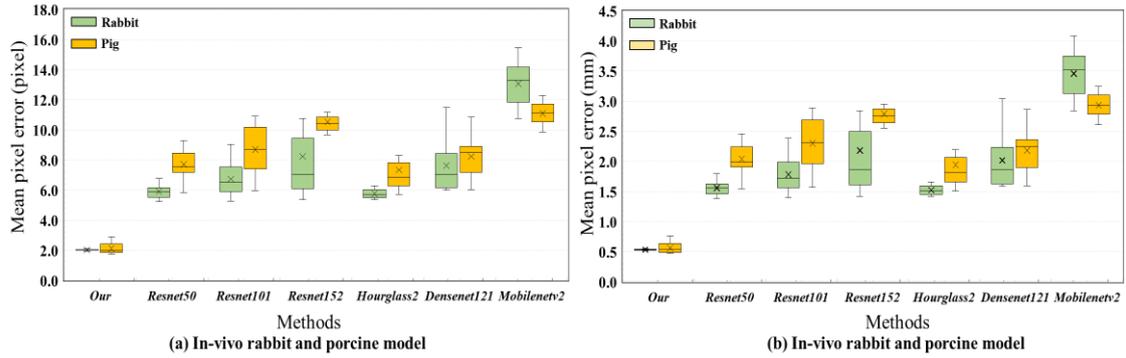


Figure 5.14: Mean pixel/distance error of six typical heatmap regression methods.

To interpret the model’s detection errors in a clinically relevant context, we converted pixel-level errors to millimetre-based measurements, reflecting the actual physical dimensions of the images, as illustrated in **Figure 5.14**. Our proposed method demonstrated superior performance, achieving the lowest millimetre error compared to the other six heatmap regression methods. **Figure 5.14** visualizes the guidewire endpoint detection results across different models. To further confirm the effectiveness of our approach, we performed a T-test to compare the eight-neighbourhood module with other learning methods. **Table 5.3** shows the differences between the seven guidewire endpoint detection methods are statistically significant.

We present the improvement to the skeletonization-repair approach implemented in the eight-neighbourhood area of each pixel. The eight-neighbourhood detection technique is specifically designed to generate regional information maps for each pixel, which can effectively remove bifurcation zones and repair breaks in the guidewire structure within the segmented feature maps. Using the relationships between adjacent pixels in the guidewire, this method achieves more accurate endpoint detection.

The proposed method also demonstrated superior performance compared to six commonly used heatmap regression methods. Among the ResNet models, ResNet50 outperformed ResNet101 and ResNet152 in the detection of guidewire endpoints, suggesting that a smaller network architecture is better suited for low resolution images where the average resolution of the guidewire is significantly limited. The ResNet-50 architecture, featuring residual blocks with identity connections, allows it to preserve and combine existing information with newly learnt features,

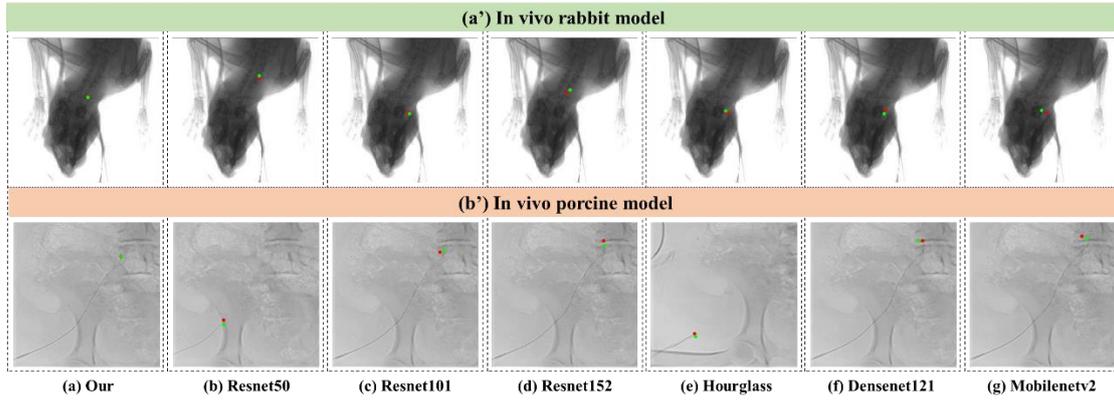


Figure 5.15: Comparison of guidewire endpoint detection of six heatmap regression methods.

resulting in improved detection accuracy for small targets. This combination makes ResNet-50 more effective than DenseNet121 and MobileNetv2 for guidewire endpoint detection in X-ray images, indicating that residual blocks provide a distinct advantage for detecting tiny targets.

In contrast, MobileNetv2 performed worse than our proposed method, as well as ResNet50, ResNet101, ResNet152, Hourglass, and DenseNet121. Although MobileNetv2's inverse residuals and linear bottleneck structures optimize memory use and reduce computational load, the model lacks sensitivity to small target features. The deeper network structure of MobileNetv2 may affect the detection of tiny objects such as guidewire endpoints, as this task does not benefit from additional network depth and instead requires efficient feature extraction from shallow layers. This finding suggests that, for guidewire endpoint detection, deeper layers may reduce the model's ability to learn the essential features needed for accurate localization.

5.4.4 Comparison of heatmap and regression methods

In the guidewire endpoint detection task, methods based on the heatmap model outperformed those based on the heatmap regression model on both Datasets A and B, as shown in **Figure 5.16**. This suggests that heatmap-based methods are more effective in handling noise and image errors. With increased down-sampling, image resolution decreases, simplifying classification tasks while making regression tasks more challenging. Heatmap-based methods exhibit better fault tolerance to noise and errors, whereas heatmap regression models are more sensitive to input noise, leading to inaccuracies in key-point localization. In contrast, our proposed method, based on the pixel-adjacent relationship approach, differs from both heatmap and regression models, achieving the highest performance.

Although many models perform well in training and standardized public datasets, they may struggle with real-world data. To evaluate unbiased performance, we tested our proposed method and other standard detection methods on self-acquired X-ray images from robot-assisted endovascular interventions. Our approach, employing various deep neural networks, demonstrated competitive performance in X-ray image analysis without requiring highly specialized deep learning hardware or extensive datasets. The proposed method maintained high detection accuracy for guidewire endpoints, highlighting the effectiveness of the two-stage pixel-adjacent-relation-based

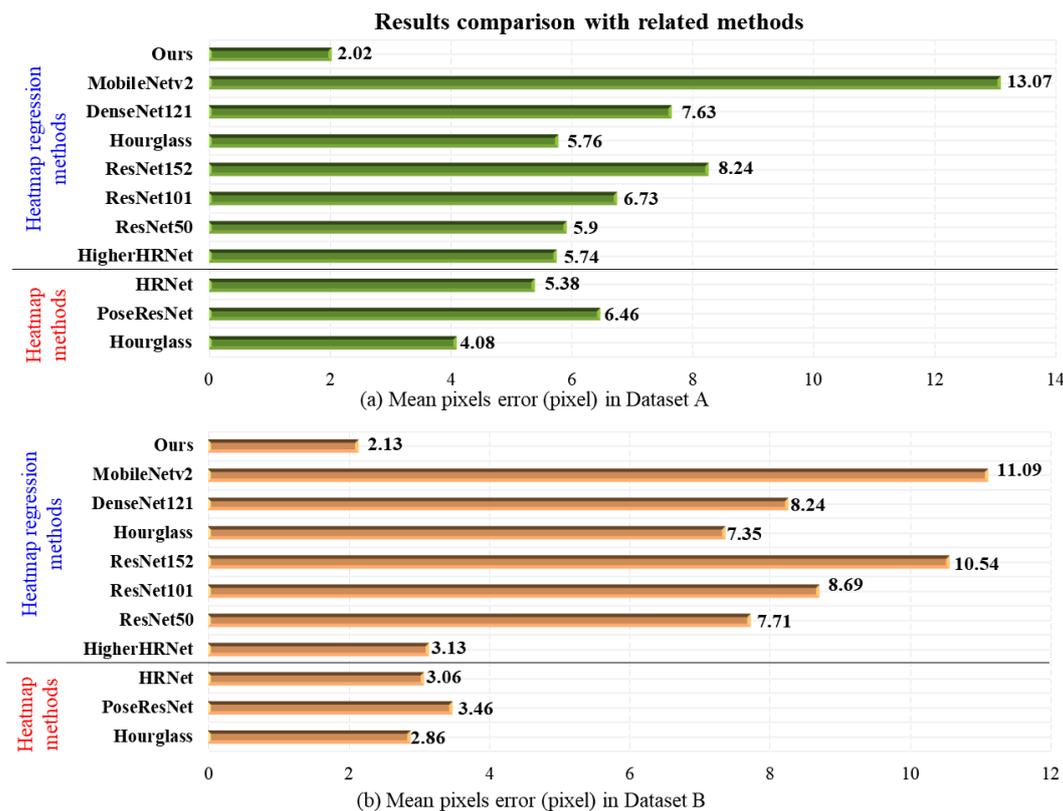


Figure 5.16: Overview of the performances of guidewire endpoint detection methods.

method in accurately identifying guidewire endpoints in 2D angiograms, even with a limited training dataset.

An analysis of 10-fold cross-validation results for detecting guidewire endpoints in Dataset A (derived from an in vivo rabbit model) and Dataset B (derived from an in vivo porcine model) shows analogous performance across both datasets. Although acquired with different equipment, both datasets contain X-ray images with similar pixel points that require localization. This uniformity indicates that our method demonstrates robust generalisation capabilities for pinpointing small object endpoints and can efficiently handle new datasets not encountered during training. In addition, this model can be applied to key-point detection tasks on other surgical instruments.

5.4.5 Guidewire bending angle

The angle of bending of the guidewire is a critical parameter that reflects its posture within the endovascular pathway. Accurate detection and measurement of the guidewire's bending angle can provide valuable feedback to the surgeon, helping them adjust their manipulation strategy to minimize the risk of endovascular rupture during surgical procedures. To this end, the proposed method identifies regions of maximum bending and calculates the bending angle after completing the guidewire endpoint detection. This approach offers a quantitative assessment of the degree of bending at specific points along the guidewire.

The performance of the guidewire bending detection method was validated using Datasets A and B. Examples of detected bending regions and corresponding angle values using the pixel-adjacent-

relation-based method are presented in **Figure 5.17**. The proposed method achieved an average accuracy of $91.13 \pm 1.12\%$ in Dataset A and 93.18% in Dataset B, demonstrating its sensitivity and effectiveness in detecting maximum bending regions.

Finally, bending angles are extracted as a reference for operators to adjust their control strategy and minimize the risk of endovascular rupture. The strong performance of the maximum bending region detection method, validated in multiple datasets, underscores its potential clinical utility [264]. This bending detection approach provides a reliable solution for the identification of guidewire bend poses. In the future, bending angle and maximum distance parameters could be integrated into a human–computer interaction interface, allowing operators to adopt optimal manipulation strategies that enhance the safety and stability of robot-assisted interventions [265].

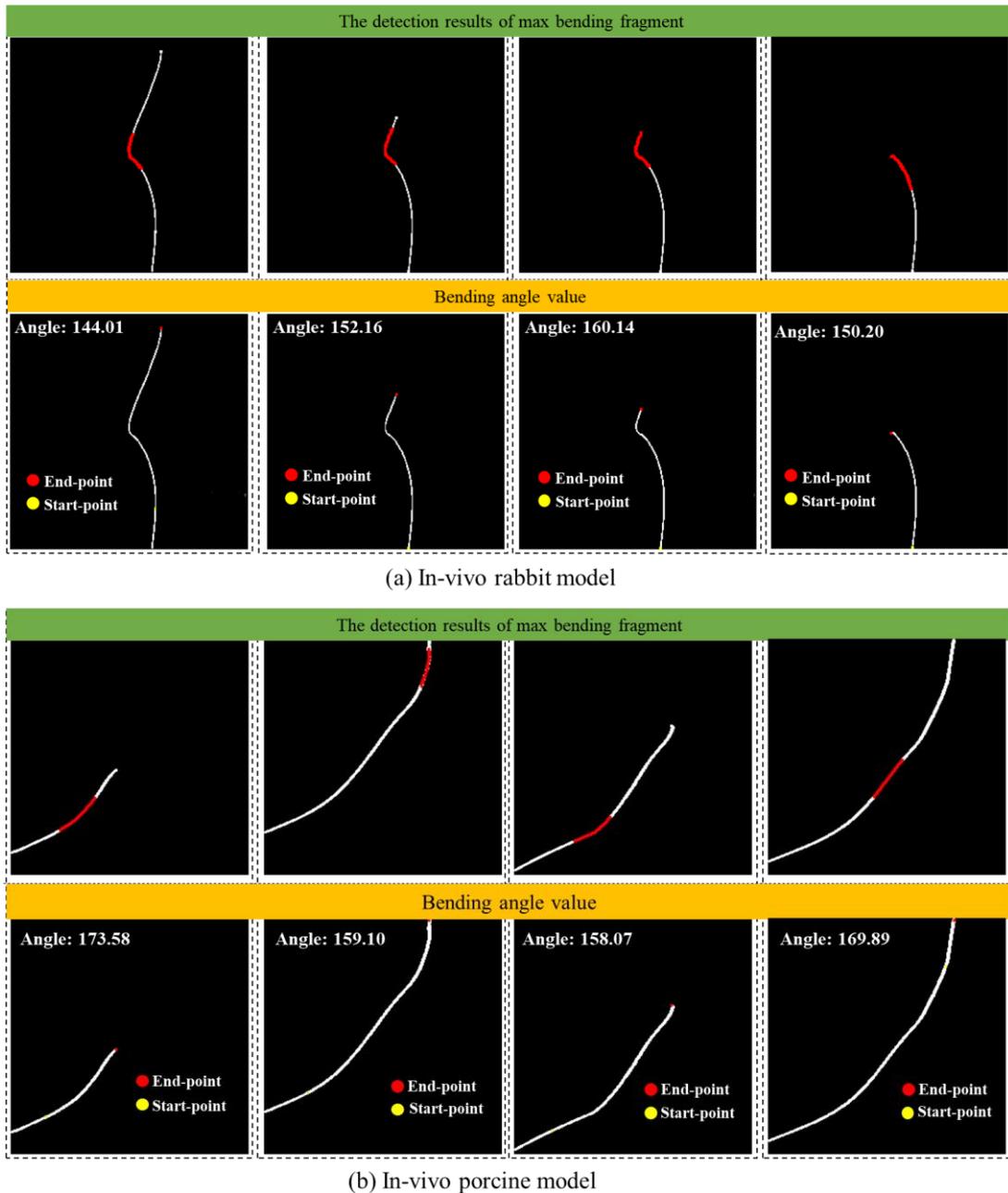


Figure 5.17: Evaluation of guidewire bending angle detection using the eight-neighbourhood method.

5.5 Chapter Summary

This chapter introduces a two-stage, eight-neighbour approach for guidewire endpoint detection. The first stage uses an improved U-Net network to segment guidewire data, improving ROI extraction and reducing interference from anatomical structures and imaging artefacts. The second stage detects guidewire endpoints using segmented data through pixel adjacency within the eight-neighbourhood framework. It involves pixel-level skeletonization, removal of the bi-furcation point, and repair of the fracture point to preserve structural integrity. Validation in a benchmark dataset showed mean pixel errors of 2.02 (0.534 mm) in in-vivo rabbit X-ray and 2.13 (0.563 mm) in in-vivo porcine X-ray, surpassing ten conventional heatmap methods. This method can also detect endpoints of other surgical tools, such as needle tips, stents, and balloons, while maintaining guidewire flexibility.

The improved U-Net model excels at segmenting tiny guidewire instances from X-ray images. Although the eight-neighbourhood detection method requires high-quality segmentation to minimize false positives and negatives, superior results are achievable with accurate segmentation. Our modified U-Net outperformed standard U-Net, U-Net++, and DeepLabV3+ models in guidewire segmentation tests. Effective endpoint detection relies on two main aspects: (1) targeted segmentation of small objects, reducing interference, and (2) skeletonization-to-repair processing. This approach improves pixel proximity, limiting non-essential pixels and removing anomalies during endpoint detection, thus improving precision based on pixel-adjacent relationships. Our method yielded the best detection results on Datasets A and B, outperforming standard heatmap and regression models. In addition, it offers precise angle measurements and valuable feedback to surgeons, which aids in human-robot interaction for safer catheterization. Challenges remain as the application of the methods to multi-instrument scenarios is untested, and performance on Datasets A (rabbit model) and B (porcine model) may not fully reflect human surgery variability, highlighting the need for future research to use diverse datasets for broader applicability. Detecting maximum bending regions is also a limitation; although effective for complex bends, the method struggles with simpler bends, affecting the accuracy of identifying the highest bending point.

To further advance visual perception capabilities in more complex and clinically realistic settings, the next **Chapter 6** proposes a multi-modal image fusion framework for the comprehensive modelling of interventional instruments. Unlike the single-modality detection approach discussed here, the upcoming work uses visual information from multiple imaging modalities to simultaneously enhance instrument segmentation, feature extraction, and perceptual robustness across various tools and conditions (DSA images, video, simple and complex vascular pathways, different size guidewires). This enables the operators to better sense the surgical environment, perception the instruments with greater accuracy, and adapt to procedural complexity. By integrating multi-source image data, the visual perception model aims to help surgeons improve intraoperative guidance, reduce ambiguity in instrument localization, and support more intuitive manipulation.

Chapter 6: Visual Perception Modelling of Interventional Instruments Using Multi-Modal Images Fusion

6.1 Introduction

Accurate instrument location and intuitive representation of anatomical structures are essential to improve tool manipulation in robot-assisted interventions. By providing surgeons with a comprehensive perception of the surgical environment, optimized visual perception systems not only improve procedural precision and safety, but also increase the overall success rate and reduce surgical risks. This chapter proposes a MBTPDS-Net network to achieve precise segmentation of surgical instruments in multi-modal fusion images. This method uses an encoder-decoder architecture with an improved visual-geometry-group-13 (VGG13) encoder to extract detailed edge and texture information. The encoder structure integrates batch normalization after each convolution layer, stabilizing and accelerating the training process while extracting multiscale features comprising both low-level and high-level semantic information. This design also reduces computational complexity and the number of parameters.

The TPD complements the encoder by combining feature maps from different scales, improving the network's ability to capture global contextual semantics and compensating for the limitations of the encoder branch. The decoder structure includes three pyramid modules, each focusing on distinct feature fusion methods and enabling parallel processing of information at various scales. These advanced segmentation modules form the foundation of the proposed vision-based semantic segmentation method, which enables precise and efficient visualization of surgical instrument information for safety and precisely intuitive manipulation during interventional navigations with robotic systems.

6.2 Related Work

Precision surgical instrument segmentation presents significant challenges due to various complex factors, including low contrast of surgical instruments, intricate surgical environments, mirror reflections, and variations in instrument scale and shape. Furthermore, there is a pronounced imbalance between the surgical instrument pixels and the background pixels, often resulting in blurred boundaries and misclassification in the results of segmentation. These issues compromise segmentation accuracy, potentially leading to errors in surgical navigation, extended operation times, and increased risks of complications for patients.

To address these challenges and improve segmentation accuracy, many scholars have proposed and refined various methods to achieve automatic segmentation of surgical instruments. Before the advent of DL, surgical instrument segmentation was predominantly performed using model-based semiautomatic traditional image segmentation methods. These approaches typically relied on basic

image features such as colour, edges, and textures. Common methods included edge detection [266], clustering [267], graph theory [268], and thresholding [269]. Although traditional methods found extensive applications in medical image segmentation and often succeeded in accurately identifying boundaries, they faced limitations such as high computational complexity and low efficiency. These limitations restricted their applicability to surgical instrument segmentation tasks, particularly in dynamic and complex surgical environments.

With the advancement of artificial intelligence and big data, DL algorithms have been increasingly used in medical image segmentation. These methods excel at processing raw data directly and automatically learning complex, abstract and high-dimensional features, enabling autonomous image segmentation. The most prominent DL-based image segmentation models include fully convolutional networks (FCN) [270], DeepLab [271], SegNet [272], and Unet [273]. Shuvo *et al.* [274] proposed a lightweight segmentation network by integrating a novel CNL module with the traditional Unet architecture, improving the focus of the network on the pixels critical to segmentation accuracy. While this approach reduced computational costs and energy consumption, it suffered from loss of local characteristics due to Unet's limited capacity to capture localized and contextual information effectively. Chen *et al.* [271, 275] introduced a series of DeepLab networks that utilized multi-scale encoders, atrous spatial pyramid pooling (ASPP), and depth-wise separable convolutions to improve multi-scale feature extraction. Despite these enhancements, the results exhibited detail loss and insufficient contextual understanding. Similarly, Cao *et al.* [276] combined conditional random fields with SegNet to achieve good segmentation accuracy and speed on public datasets, but the approach was hampered by complex training processes and misclassification of small or local targets due to inadequate contextual information extraction.

Unet has emerged as a benchmark for medical image segmentation owing to its robust performance. Numerous Unet variants have been developed, focusing on two primary improvement strategies: incorporating new modules and refining encoder-decoder structures. Feng *et al.* [277] proposed the CPFNet (context pyramid fusion network) for skin lesion segmentation, introducing a scale-aware pyramid fusion module in the bottleneck layer to extract advanced multi-scale contextual information. Additionally, a module guided by the global pyramid in the skip connection provided detailed and contextual information at different resolutions. Similarly, Li *et al.* [278] developed VCE-Net, where a dense atrous convolution block and residual multi-kernel pooling blocks extracted high-level semantic and contextual information, improving segmentation performance. Leveraging pre-trained classical network structures, such as ResNet [279] and VGG [280], as encoders, researchers have enhanced feature extraction for complex tasks. For instance, Shen *et al.* [161] introduced BAA-Net, which incorporated lightweight encoders and novel modules: branch balance aggregate (BBA) and block attention fusion (BAF) to optimise global and local surgical instrument location. Yang *et al.* [281] proposed an attention-guided network with a residual path module at the encoder for robust propagation of low-level features and a dual-attention module at the decoder to emphasise relevant features while suppressing distractions.

Despite advancements, segmenting non-rigid, small, soft, and low-contrast guidewires remain a challenging task. Unlike forceps, scalpels, and needles, rigid surgical instruments that exhibit good contrast and clear, just as the methods mentioned above, continuous relationships in endoscopic images during da Vinci-assisted procedures, guidewires face unique segmentation difficulties in endovascular interventional surgeries. During interventional endovascular surgery, surgeons rely solely on 2D X-ray images generated by DSA equipment, which often provides low spatial resolution, so tissue occlusion often occurs during the guidewire delivery process in DSA images, further complicating its visualisation. This limitation hampers effective visual analysis, increasing the likelihood of misjudgement. Moreover, guidewires, defined by their basic forms, often have contours that are hard to distinguish from similar objects. Changes in movement and posture during surgery also lead to notable changes in shape and size, making segmentation more challenging.

Complex surgical environments exacerbate these challenges. Factors such as blood, non-target organs and tissues, mirror reflections, and motion artefacts interfere with accurate guidewire segmentation. Therefore, it is critical to develop robust automated segmentation methods that are tailored to guidewires. Such methods are essential for improving surgical navigation accuracy, reducing procedural risks, and improving the success rate of robot-assisted endovascular interventions.

6.3 Methodology

A method for semantic segmentation is introduced to achieve accurate segmentation of surgical instruments (SI). From the segmentation outcomes, the instrument's pose can be estimated, and its usage status can be deduced, providing essential visual data to aid operators in tool manipulation. By quantitatively describing the tools' position and orientation within a complex lumen, this visual perception can greatly improve operators' intuitive control performance in computer-assisted procedures. This enhancement is especially crucial to overcome the challenges posed by the lack of haptic feedback in master-slave robotic systems used in interventional surgery.

6.3.1 Multi-modal image dataset

A semantic segmentation method was designed based on multi-modal image datasets to achieve precise segmentation of surgical instruments. Accurate and representative data sets were essential for effective segmentation. This section presents images obtained from two self-designed interventional robotic systems: a vascular interventional system and a teleoperated robotic bronchoscopy system for pulmonary lesion biopsy. In these systems, surgeons operated a robotic master control terminal (doctor terminal), while the slave mechanism (patient terminal) executed instructions to deliver surgical instruments through natural orifices to the target site.

Figure 6.1 illustrates the dataset acquisition framework. The vascular interventional robotic system was developed to assist surgeons in performing endovascular interventions, aiming to reduce exposure to X-ray radiation and alleviate the fatigue associated with the wearing of heavy protective

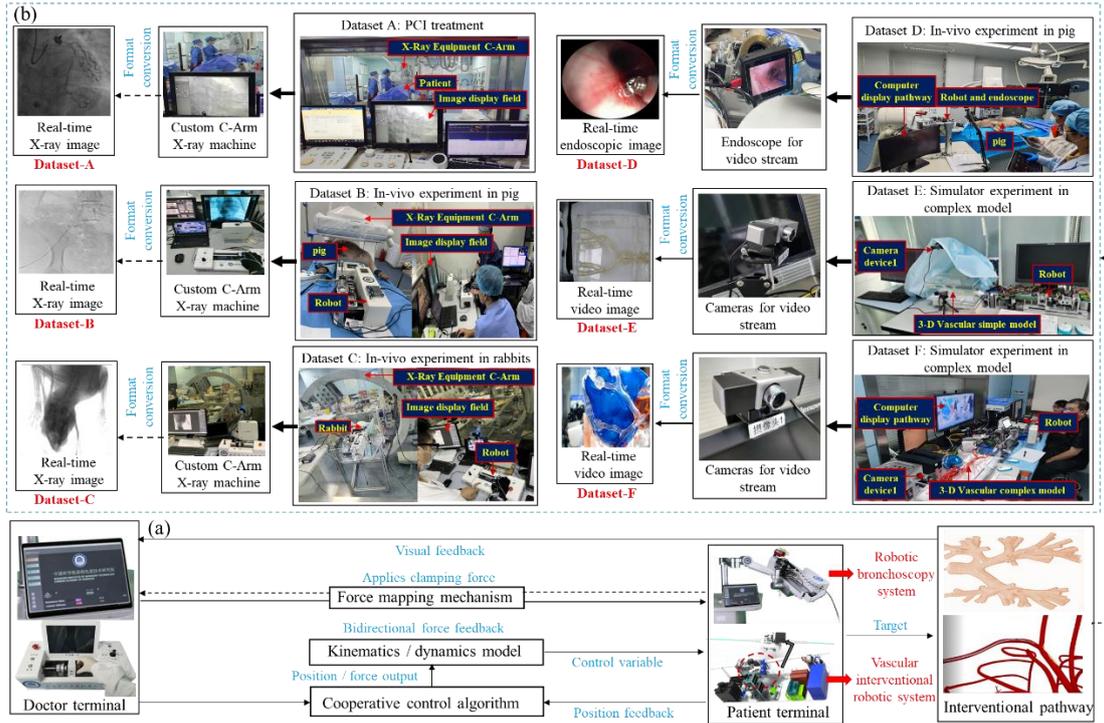


Figure 6.1: Image acquisition during robot-assisted endovascular and bronchoscope interventional procedure.

lead aprons. Similarly, the robot-assisted bronchoscopy system provided a minimally invasive and effective solution for early lung cancer intervention. Guided by a bronchoscope, the system navigated through the oral or nasal cavity to reach specific bronchial locations, allowing biopsy forceps or needles to obtain tissue samples from the sites of the injury.

Six datasets were used, including three sets of X-ray angiograms, one set of endoscopic images, and two sets of video data to segment surgical instruments. Five of these datasets were derived from the two robotic interventional platforms. All experimental procedures adhered to the ethical guidelines approved by the respective committees. Details of each dataset are summarised below:

Dataset-A: This dataset comprised coronary angiographic images of cardiovascular patients obtained from angiographic examinations or percutaneous coronary interventions at Shenzhen GuangMing District People’s Hospital. Using a 0.025" guidewire, surgeons navigated the radial artery to identify endovascular blockage sites. During PCI, a 0.014" guidewire was operated from the catheter ostia to the stenotic site. The images were captured by a Philips X-ray fluoroscopy CT arm at 16 frames per second. All patient information was anonymised to ensure privacy.

Dataset-B: Derived from an in vivo porcine model simulating the human renal artery vascular system, this dataset involved a robotic slave operating a 0.014" guidewire under Philips X-ray fluoroscopy guidance at 15 frames per second. The images provided 2D or 3D visualisations with optional angiography subtraction. The study adhered to ethical guidelines approved by the Shenzhen Advanced Animal Study Service Centre (No. AAS 191204P).

Dataset-C: This dataset involved a master-slave robotic system navigating a 0.014" guidewire along the auricle-to-coronary arterial pathway of rabbits. Images were captured using a self-

developed X-ray machine equipped with a 70 kV [282], 5 mA generator, and a flat detector, producing 10 frames per second. Ethical approval was obtained from the Shenzhen Institutes of Advanced Technology (Application No. SIAT-IACUC-200528-YGS-WL-A1289).

Dataset-D: Derived from an in vivo porcine bronchial model, this dataset involved a robotic bronchoscopy system that delivers an endoscope through the oral cavity to specific bronchial positions. Biopsy tools, such as 1.0-mm diameter forceps, were guided through the endoscope's working channel to obtain tissue samples from nodules. The bronchoscopy robot included components such as electric sliders, rotary motors, and gear sets, with a bronchoscope featuring a 5.2 mm external diameter and a 2.6 mm working channel capable of bending 160° upward and 130° downward.

Dataset-E and Dataset-F: These datasets were obtained from simulations of interventional endovascular surgeries using self-developed master-slave robotic systems. Dataset E involved a simple abdominal aorta model with an empty lumen, where a robot navigated a 0.025" guidewire from the proximal cavity to the distal end, guided by a camera at 30 frames per second. Dataset F featured a complex endovascular simulator infused with blood-like fluid circulated by a pump, simulating real-time flow. A robot operated a 0.025" blue guidewire from the coronary artery orifice to the anterior descending branch, with motion trajectories recorded at 30 frames per second.

6.3.2 Image pre-processing and data augmentation

Dataset-A consisted of 856 X-ray images with dimensions of 512×512 pixels, while Dataset-B contained 904 angiogram images sized at 665×880 pixels. Dataset-C included 1650 X-ray images, each originally sized at 1560×1440 pixels. Dataset-D was made up of 1978 images from endoscopic videos, sized at 890×970 pixels. Dataset-E consisted of 1870 video images with dimensions of 1280×720 pixels, and Dataset-F contained 880 video images sized at 700×700 pixels. All images were standardised to a resolution of 96 dpi and converted to ".jpg" format for uniform processing, resulting in a total of 8138 images.

To annotate the trajectory information of surgical instruments, LabelMe was used for manual marking and the annotations were saved as JSON files. To improve efficiency, the Segmentation Anything Model (SAM) developed by Meta AI was evaluated for its ability to automate the marking process. SAM successfully annotated instrument trajectories across the dataset, generating JSON files for subsequent processing. To enhance the generalisation capability of the proposed model, data augmentation techniques were applied to create a diverse training set, thus improving model performance. This augmentation process involved generating synthetic data through random transformations to simulate various clinical conditions. Transformations were categorised into four types: conventional augmentation, deformation, blurring, and affine transformations. Each type of transformation had a 20% probability of being applied, with a randomly selected transformation performed per image.

Conventional augmentation: Included brightness and contrast adjustments and Gaussian noise addition to simulate environmental effects like lighting variations or equipment noise. Brightness

and contrast adjustments reflected changes due to illumination, while Gaussian noise simulated electronic or electromagnetic interference.

Elastic deformation: Applied to replicate non-rigid changes in human anatomy or soft tissues. Elastic deformation generated displacement fields using smoothed random noise and warped images based on scaling (α) and smoothing (σ) factors. This enhanced robustness to shape variability by mimicking real-world anatomical distortions while maintaining structural integrity.

Gaussian blur: Used to simulate optical effects or motion-induced blurring. By convolving the image with a Gaussian kernel, the process smoothed pixel intensity variations, reducing high-frequency noise while preserving essential structural features. The degree of blur was controlled by the standard deviation (σ) of the Gaussian kernel to ensure a realistic balance.

Affine transformations: Simulated motion effects of the endoscope or surgical instruments. These included translations, rotations, and scaling that reflected realistic procedural dynamics.

Parameter details for each method are summarised in **Table 6.1**, with the image augmentation processing results illustrated in **Figure 6.2**. Training network models for surgical instrument segmentation required an understanding of various spatial information to accurately identify similar

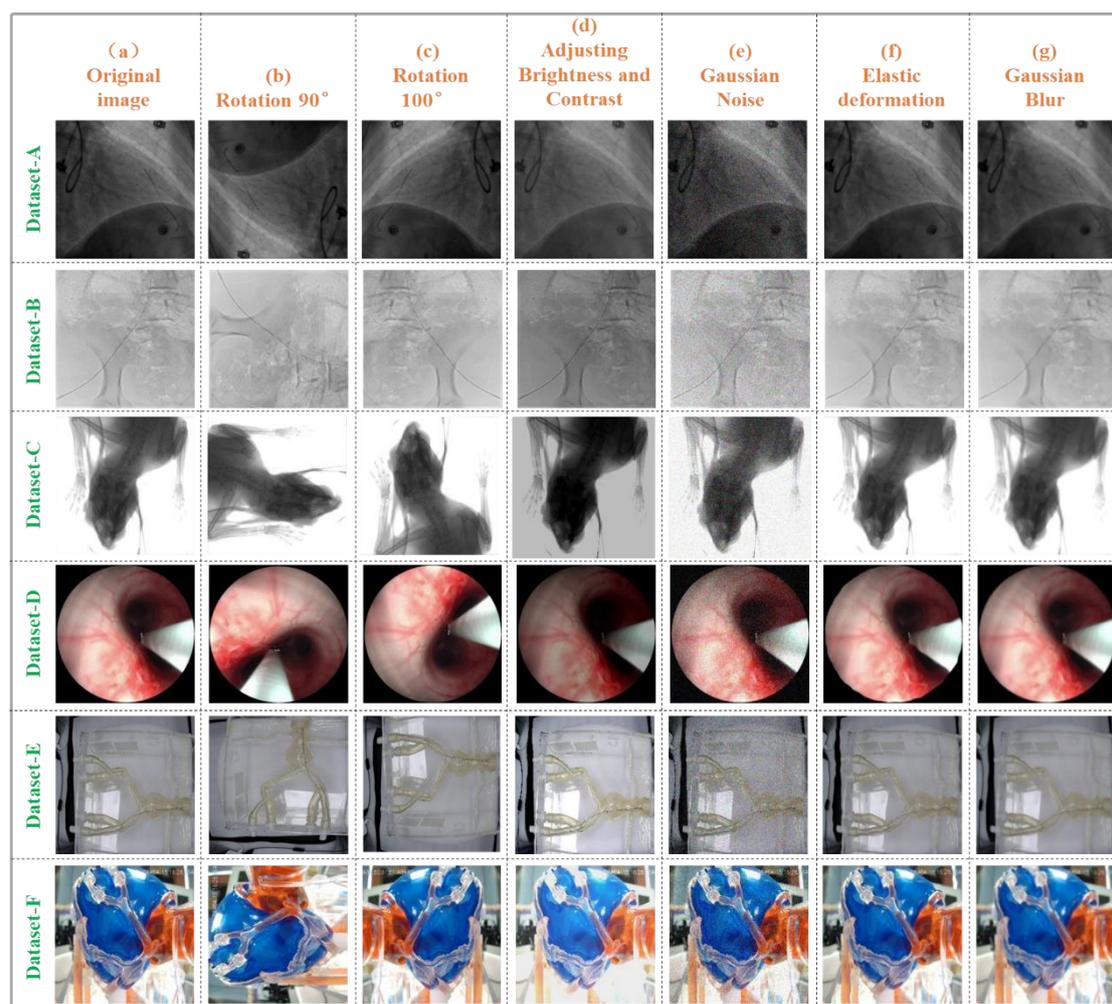


Figure 6.2: Different augmentation methods.

Table 6.1: Image augmentation methods and their respective parameters.

Type	Augmentation methods	Parameters
Traditional augmentation	Adjusting Brightness and Contrast	[0.5, 1.5]
	Adding Gaussian Noise	scale = [0, 25]
Deformation	Elastic deformation	alpha= [10, 20], sigma=5
Blur	Gaussian Blur	sigma = [0, 5]
Affine transformation	Rotation	[-90, 90]
	Translation	[-180, 180]

pixel intensities across various contexts. Instead of relying solely on acquiring and annotating additional data, the augmentation procedure strategically improved model performance. Thus, the augmentation strategies aimed to replicate the complexity of the interventional procedure in the real world by addressing both environmental and procedural variability. These enhancements allowed the model to better simulate the actual operating environment, improving the accuracy and robustness of segmentation in various clinical scenarios. During validation and evaluation, pure white masked pixels representing the surgical instrument trajectory were used to facilitate efficient matrix matching.

6.3.3 Surgical instrument segmentation method

A novel deep segmentation network, termed the MBTPDS-Net, was developed to achieve precise surgical instrument localisation at the pixel level, as shown in **Figure 6.3**. The network employs an encoder–decoder architecture, integrating multiple innovative modules to improve segmentation performance. The overall architecture of MBTPDS-Net, along with the detailed design of its key submodules, including the multibranch encoder unit, feature fusion block, and triple decoder unit, is comprehensively described. These components work in tandem to extract and process multi-scale features, facilitating accurate segmentation even in complex surgical environments.

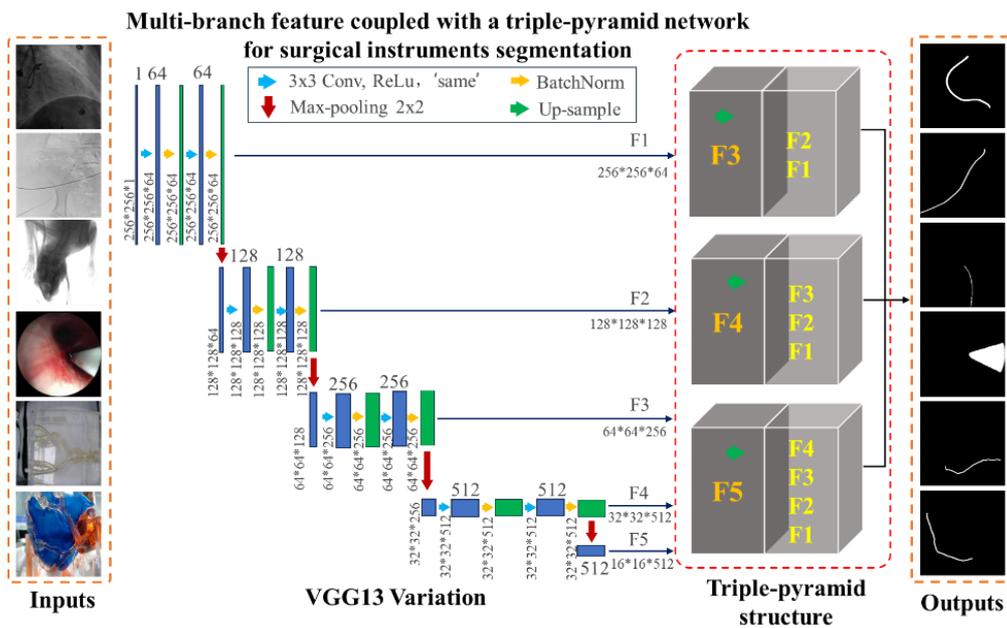


Figure 6.3: Multi-branch coupling with MBTPDS-Net model.

A. Design of encoder network

The typical UNet architecture and its variants commonly employ stacked 3×3 convolution and down-sampling layers in the encoder unit. While this approach enables feature extraction, the relatively small kernel size limits the receptive field, restricting the network's ability to capture global contextual information. Although the receptive field gradually expands with increasing layer depth, it often remains insufficient for extracting global features, particularly in complex tasks. Moreover, simply stacking additional convolutional layers does not consistently enhance recognition performance. In some cases, network performance may degrade as depth increases. To address this, ReLU activation functions are used to mitigate issues such as gradient vanishing. However, at deeper network levels, improperly added convolutional layers can exacerbate loss divergence, making gradient issues unavoidable. Therefore, a feature extractor capable of effectively enlarging the receptive field and capturing global contextual information is critical to achieve optimal semantic segmentation results.

Despite the advancements in simpler architectures like ResNet and Inception networks, the VGG network remains a robust choice for feature extraction due to its high accuracy and larger receptive field [283]. VGG excels at filtering out unrelated background information while capturing detailed spatial features. Its stacked multi-convolution layers enhance the network's feature learning capabilities, enabling the extraction of sparse and meaningful features. Furthermore, the consistent use of 3×3 convolution layers and 2×2 pooling layers in VGG contributes to its simplicity and strong generalisability, making it adaptable to various datasets [284].

To enhance performance, this study proposed a multi-branch encoder unit based on an improved VGG model. The encoder is designed to preserve the ability to learn both local and global context features, as illustrated in **Figure 6.4**. A branch-based deep feature fusion architecture was introduced, enabling efficient encoding of multi-modal image features. Feature maps at various scales were extracted and subsequently processed in the decoder.

The encoder network employed a five-branch structure, where each feature fusion block was optimised to learn scale-specific information from a multi-modal training dataset. A modified version of the VGG structure was adopted for feature extraction [285], incorporating an alternate

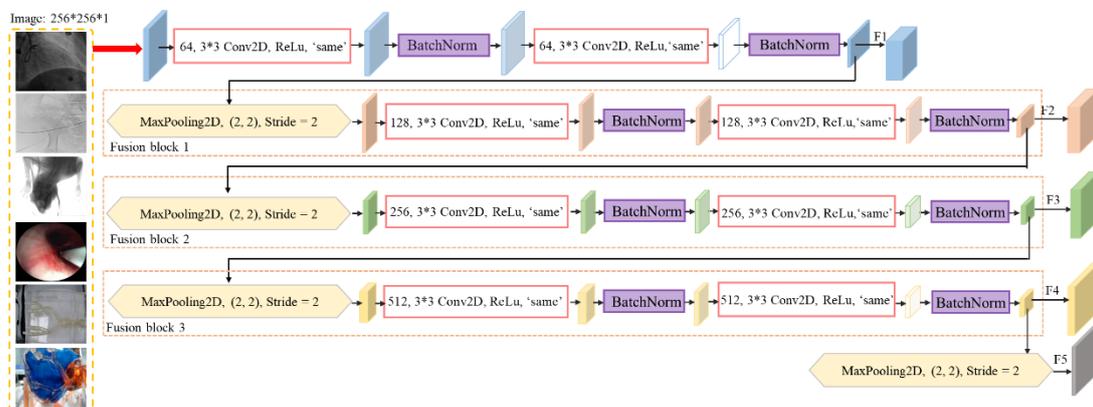


Figure 6.4: Design structure of the encoder model.

combination of convolution, batch normalisation, and maximum-pooling layers in each block. Batch normalisation was applied after each convolution layer to stabilise activations and accelerate training. To preserve spatial dimensions, the same padding was consistently used, replacing the mixed use of the same and valid padding found in the original VGG model. Although deeper layers are beneficial for capturing complex characteristics, they also increase computational complexity, training difficulty, and the risk of parameter explosion.

For applications requiring precise segmentation, such as detecting small and thin surgical instruments such as guidewires and catheters, low-level features are crucial to capture edges and textures. To ensure a focus on lower-level features, the last two convolutional layers (conv-512) of the VGG13 model were removed. This adjustment allowed the encoder to retain more spatial detail, allowing accurate segmentation of small instruments. Additionally, the last three fully connected layers of VGG were replaced in the decoder, reducing computational complexity and increasing processing speed for real-time navigation systems.

To improve computational efficiency, convolution operations were optimised to learn features within a relatively fixed receptive field, while small convolution kernels were used to minimise computational costs and hardware demands. However, relying solely on a limited receptive field could impede the extraction of global contextual features. Although depth expansion progressively enlarges the receptive field, it may still fail to encompass the entire image, limiting the encoder's ability to capture comprehensive global information. Different layers of convolution capture features at varying levels of abstraction: low-level features provide high-resolution spatial detail, while high-level features encapsulate rich semantic information [286]. Integrating these features into the decoder is critical to producing accurate predictions, especially for segmentation tasks that require precision.

In this section, feature maps were progressively integrated into the decoder after every two convolutional and normalisation processes. This strategy enabled the encoder to extract and utilise multiscale features, combining lower-level boundary details with high-level semantic information. By combining features at different scales, the network maximised the utilisation of layer-specific information, achieving comprehensive integration of local and global context during the up-sampling process. Compared to directly using output feature maps from lower layers, this method reduced parameter requirements while enhancing the flow of useful information, resulting in improved segmentation performance.

B. Design of decoding network

The decoding network was designed to enhance segmentation performance through the integration of a triple-pyramid multi-scale feature fusion decoder, as depicted in **Figure 6.5**. This decoder incorporated three pyramid modules, each tasked with up-sampling and fusing feature maps from multiple encoder levels to generate refined segmentation maps. When the original image is input into the encoding network for feature extraction, it undergoes resizing and subsequent down-sampling at each layer of the network. This progressive reduction in the size of the output feature

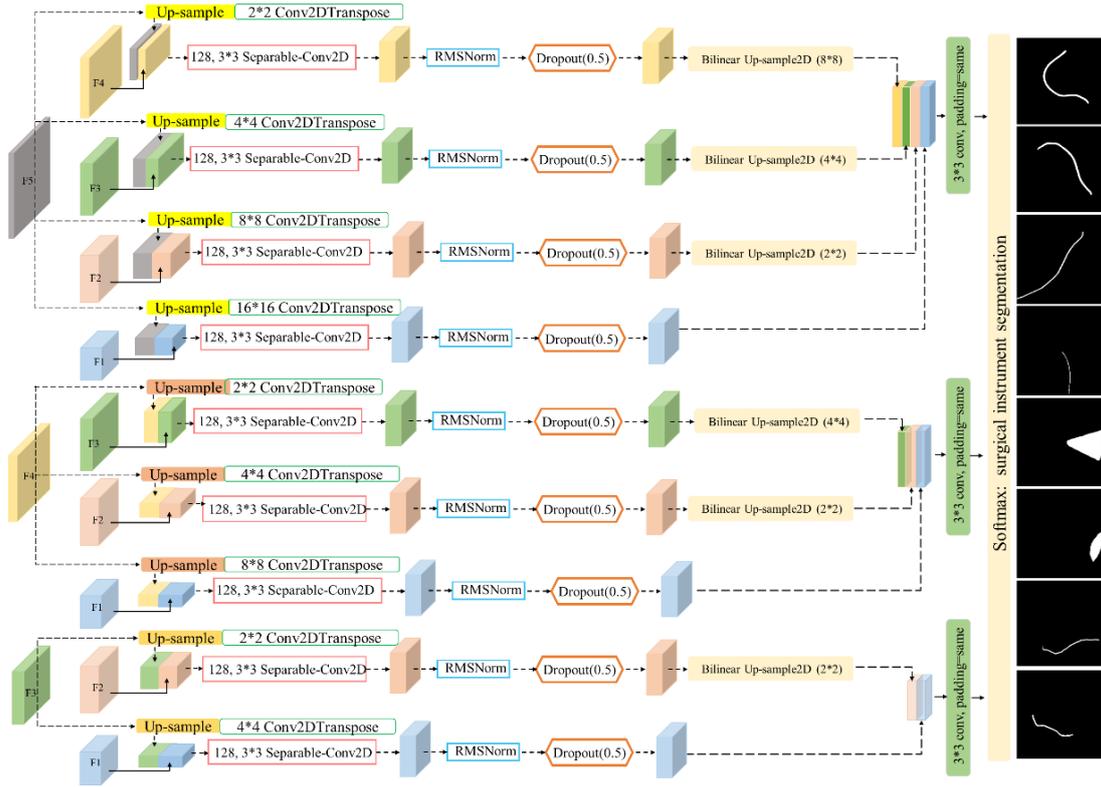


Figure 6.5: Design structure of the TPD model.

map is a natural result of encoding operations. However, to generate the final segmentation output, the image must be restored to its original resolution. This restoration process is performed during the decoding phase, where up-sampling operations are employed to expand the feature map size, mapping it from a lower resolution back to a higher resolution.

Transposed convolutions, often referred to as deconvolutions, operate by transposing matrices before performing convolution, enabling the up-sampling of feature maps to higher resolutions. Unlike predefined interpolation methods, transposed convolutions require training, allowing the network to dynamically learn optimal reconstruction patterns for feature maps. This adaptability makes them highly effective for maintaining pixel-level accuracy and improving feature reconstruction in semantic segmentation tasks.

Bilinear interpolation, on the other hand, is a computationally efficient method that estimates the pixel values in the target image using the four nearest pixel values of the original image. By performing linear interpolation in two directions, it determines the target pixel's value without requiring training. This makes bilinear interpolation particularly suitable for applications requiring fast processing and minimal computational resources, albeit at the cost of reduced adaptability to complex feature representations.

In our decoder design, transposed convolution was utilised at the initial stages of up-sampling within each branch, replacing bilinear interpolation. This choice enabled the network to learn a more optimal method for increasing the spatial resolution of feature maps, leveraging the flexibility of trainable parameters to refine the reconstruction process. At the final stage of up-sampling, bilinear

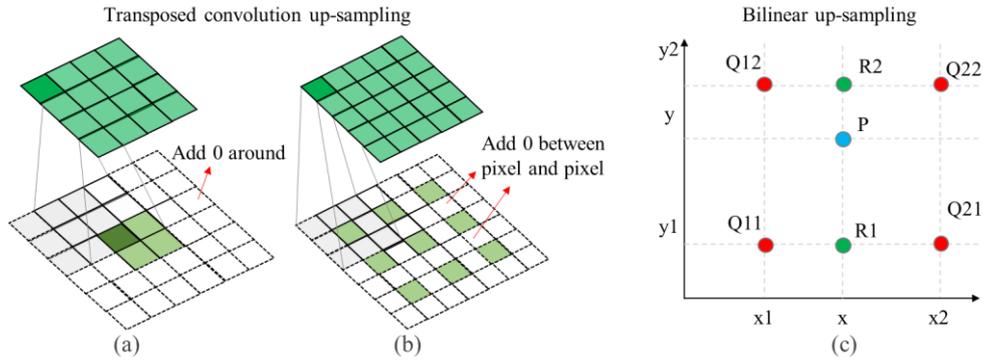


Figure 6.6: Transposed convolution and bilinear up-sampling.

interpolation was employed to further enhance the feature maps. This complementary use of transposed convolution and bilinear interpolation ensured an effective balance between adaptive learning and computational efficiency, as shown in **Figure 6.6**. This hybrid approach improved spatial resolution while maintaining the precision necessary for high-quality segmentation outputs. Additionally, RMS-Norm normalisation and dropout layers were introduced before the final up-sampling stage in each branch to enhance stability and mitigate overfitting.

Convolution operations were used during the decoding phase to decode feature maps, restore their original size, and minimise information loss. However, standard convolution operations demand significant computational resources and memory. To address this limitation, each feature fusion branch utilised depth-wise separable dilated convolutions, which effectively reduced parameter requirements and computational load. Depth-wise separable convolution significantly reduces the parameter count of the original convolution kernel while preserving network performance, making it particularly well-suited for feature recovery tasks. In the multi-scale feature recovery module, feature map dimensions are first expanded, followed by the application of dilated convolution based on depth-wise separable operations with a dilation rate of 2, as illustrated in **Figure 6.7**. This approach increases the receptive field of the convolution kernel in the preceding feature layer without adding additional parameters. Consequently, it enhances the network's ability to interpret and recover image details. By utilising this technique, the proposed network demonstrates an improved capacity for reconstructing and interpreting the intricate details of complex images, thereby achieving higher segmentation accuracy. This adjustment significantly improved the efficiency and processing speed of the model.

Within each pyramid decoding module, feature maps from encoder levels were first up-sampled in individual branches and subsequently fused. These up-sampled feature maps were then combined with those produced by other branches, generating multi-scale outputs with rich semantic information. Each branch focused on capturing semantic details at varying scales, enabling the triple-pyramid decoder to build comprehensive feature representations. This hierarchical fusion process enhanced the model's ability to accurately detect and localize objects, supporting precise pixel-level classification. In the final stage, features extracted from the three parallel pyramid modules, along with depth supervision information, were merged using a dense layer with a sigmoid

activation function to complete the semantic segmentation task.

The activation function played a critical role in the connectivity between convolution layers. Compared to the sigmoid function, the ReLU activation function [287] offered faster convergence and alleviated gradient vanishing issues for values of $x > 0$. Its low computational complexity made it a preferred choice for optimising the decoding block. The ReLU function is mathematically expressed as:

$$\text{ReLU} = \max(0, x) \quad (6.1)$$

$$\text{Output}_{\text{size}} = \frac{(I_{\text{size}} - K_{\text{size}} + 2P_{\text{size}})}{S_{\text{size}}} \quad (6.2)$$

where W , K , P , S , and O represent the input feature size, kernel size, padding size, stride, and output feature size, respectively. **Figure 6.7** illustrates the convolution process involved parameters set as follows: $P = 1$, $K = 3$, and $S = 1$. Setting O equal to W using a 3×3 convolution with padding, it was possible to produce input and output feature maps of equal size.

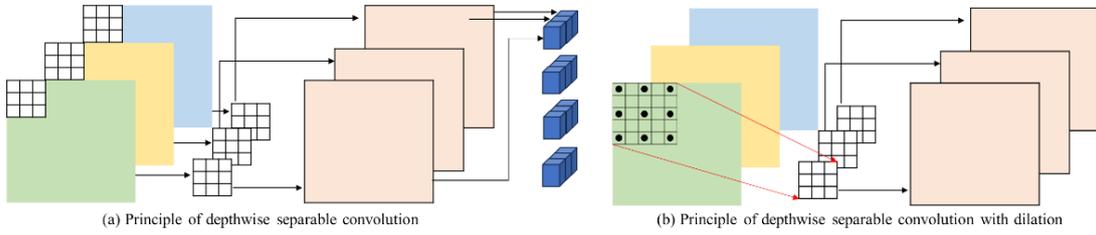


Figure 6.7: Principle of depth-wise with and without dilation.

6.3.4 Loss function design and data training strategy

A. Loss function

Micro-target semantic segmentation often faces a significant imbalance between the number of surgical instrument pixels and background pixels in the image dataset. Furthermore, small surgical instruments in the background contributed minimally to the update of the image loss function parameter. This imbalance in the pixel distribution led to suboptimal results in instrument segmentation. To address this issue, a MixLoss function [167] based on multi-task learning was proposed to adjust the allocation weights of the losses. This approach transformed the pixel discrimination problem into minimising the dissimilarity between two samples, thereby resolving the disparity between the instrument and background pixels. The MixLoss formula was as follows:

$$\text{Mixloss} = \text{Jaccard} + \text{Loss}_{xy} * \eta 1 \quad (6.3)$$

where $\eta 1 = 0.1$ adjusted the balance between the contributions of the two loss functions. The formulas were detailed follow:

$$\text{Jaccard} = 1 - \frac{\Sigma(y_t \times y_p) + \lambda_1}{\Sigma(y_t + y_p - y_t \times y_p) + \lambda_1} \quad (6.4)$$

where y_t represented the number of true position pixels, y_p denoted the number of predicted pixels, and $\lambda_1 = 1$ served as a smoothing parameter to prevent division by zero. The Loss formula was detailed as follows:

$$Loss_{xy} = loss_{xy}^1 + loss_{xy}^2 * \eta_2 \quad (6.5)$$

$$loss_{xy}^1 = f(y_t) / \left\{ \frac{[\{f(y_t) - y_t\} * \{f(y_t) - y_p\} + \{f(y_t) - y_t\} * \alpha_2]}{+(y_t * y_p + \alpha_2 * y_t)} + \lambda_2 \right\} \quad (6.6)$$

$$loss_{xy}^2 = f(y_t) / \left\{ \frac{\{((f(y_t) - y_t) * y_p) + \{f(y_t) - y_t\} * \alpha_2)\}}{+\{y_t * (f(y_t) - y_p) + \alpha_2 * y_t\}} + \lambda_2 \right\} \quad (6.7)$$

where $f(y_t)$ generated a tensor of all ones with the same shape as y_t , $\alpha_2 = 0.1$ adjusted the balance between positive and negative samples, $\lambda_2 = 10^{-7}$ prevented zero division by smoothing, and $\eta_2 = 0.2$ controlled the contribution of the loss function. The *Jaccard* and $Loss_{xy}$ functions used different strategies to measure the precision of the model in pixel-level target segmentation, demonstrating robustness to imbalanced sample issues.

B. Training strategy

The proposed model was implemented using TensorFlow, and all experiments were conducted on a Linux server equipped with an NVIDIA RTX A6000 GPU with 48GB of memory. To accelerate both training and testing, the NVIDIA CUDA 11.1 acceleration toolkit was utilised.

For model initialisation, random tensors were generated according to the Xavier normal distribution. The Adam optimiser was selected with an initial learning rate of 0.0001. This setup enabled rapid model iteration. To avoid overfitting and ensure stable training, the learning rate was halved if the validation loss did not decrease over ten consecutive epochs. Each dataset was trained for 200 epochs and training continued until the loss ceased to decrease. The model with the lowest validation loss was selected for segmentation and used to compare the prediction results.

The self-constructed dataset included five categories of images: Dataset-A, Dataset-B, and Dataset-C consisting of fluoroscopy images; Dataset-D containing endoscopic images; and Dataset-E comprising video streams. These datasets were pre-processed and randomly combined for training, with images resized to 256×256 pixels and stored in PNG format. The datasets were divided into training, validation and test sets following an 8:1:1 ratio, ensuring a balanced distribution throughout the experimental pipeline.

Existing segmentation methods often rely on homogeneous datasets, typically sourced from public datasets such as EndoVis2018, EndoVis2017, and Kvasir-Instrument datasets. While these methods achieve commendable performance on their respective training datasets, their ability to generalise to diverse real-world scenarios is significantly limited, particularly when tested on independently collected datasets with heterogeneous characteristics. To address this limitation, two training strategies were employed for the proposed model. The first involved training on a multi-modal fusion dataset integrating six different types of datasets, which aimed to enhance the model's ability to generalise across varying environments. The second strategy used a single dataset to train the model, facilitating a focused comparison of performance metrics. The trajectories of surgical instruments in robot-assisted interventions were analysed to assess the precision and robustness of

the segmentation method under these two training regimes.

6.3.5 Model performance evaluation metrics

Commonly used performance evaluation metrics for segmentation tasks include accuracy, sensitivity (Sen), specificity (Spe), and the Matthew correlation coefficient, as defined in equations 6.8 to 6.11. In these metrics, TP represents the number of instrument pixels correctly identified, false positive (FP) denotes background pixels incorrectly classified as instrument pixels, TN refers to background pixels correctly identified, and FN represents instrument pixels incorrectly classified as background. In medical image segmentation, sensitivity to both missing and superfluous predictions in the foreground is critical, making it an essential metric. Balances precision and recall, providing a comprehensive measure of the model's segmentation performance.

$$Acc = \frac{TP+TN}{TP+FP+TN+FN} \quad (6.8)$$

$$Sen = \frac{TP}{TP+FN} \quad (6.9)$$

$$Spe = \frac{TN}{TN+FP} \quad (6.10)$$

$$Mcc = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}} \quad (6.11)$$

To thoroughly evaluate the proposed method, four additional metrics commonly used in computer vision are applied: the area under the curve of receiver operating characteristics (AROC), the area under the precision-recall curve (PR-area), the Dice score, and the mean intersection over union. The Dice score quantifies the similarity between the predictions and ground truth, providing insight into the alignment between the actual and segmented regions. Meanwhile, MIoU measures the degree of overlap between predicted and ground truth segmentation, defined as follows:

$$recall_i = \frac{TP_i}{TP_i+FN_i} \quad (6.12)$$

$$precision_i = \frac{TP_i}{TP_i+FP_i} \quad (6.13)$$

$$F1 - score_i = 2 \frac{recall_i \times precision_i}{recall_i + precision_i} \quad (6.14)$$

$$macro - F1 = \frac{\sum_{i=1}^n F1 - score_i}{N} \quad (6.15)$$

$$Dice = \frac{2|T \cap P|}{|T \cup P|} = \frac{2TP}{2TP+FP+FN} \quad (6.16)$$

$$MIoU = \frac{1}{n+1} \sum_{i=0}^n \frac{q_{ii}}{\sum_{j=0}^n q_{ij} + \sum_{j=0}^n q_{ji} - q_{ii}} \quad (6.17)$$

where $n + 1$ represents the total number of pixel categories in the image; q_{ii} indicates the total number of real and predicted instrument pixels for category i , q_{ij} represents the count of real instrument pixels in category i and their incorrect prediction as background pixels in category j ; and q_{ji} refers to the number of real background pixels in category j that are incorrectly predicted as instrument pixels in category i . These metrics collectively provide a robust framework for assessing segmentation accuracy, overlap, and precision, ensuring a thorough evaluation of the model's performance in detecting and delineating surgical instruments.

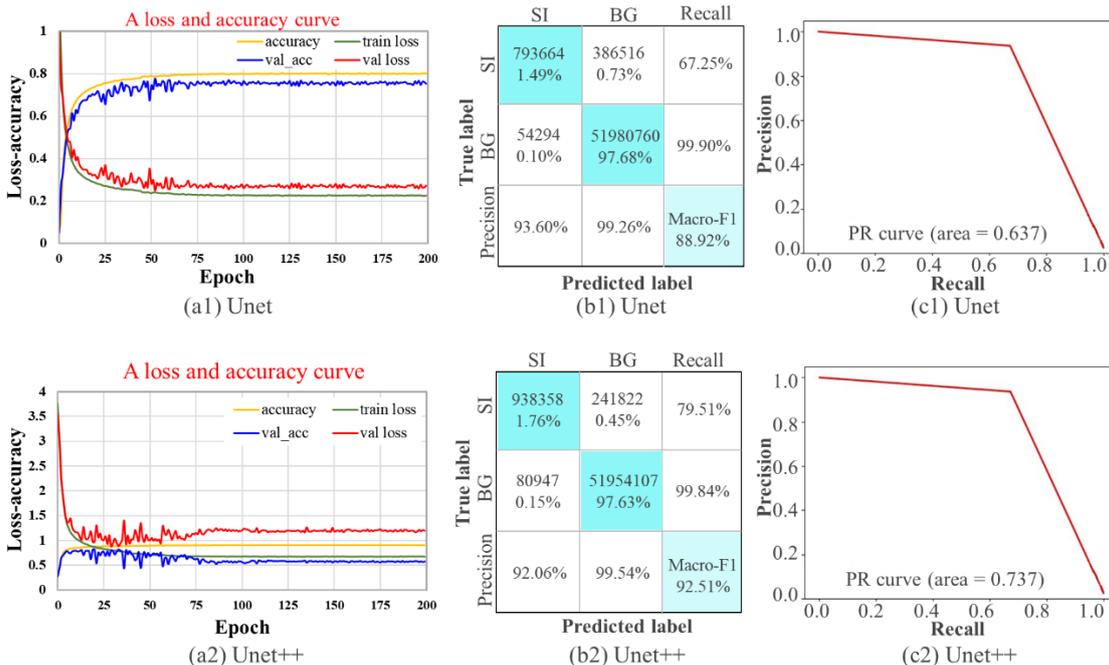
6.4 Semantic Segmentation Performance Analysis

6.4.1 Semantic segmentation results

Deep learning has advanced the development of data-driven solutions for surgical instrument segmentation, yet challenges related to precision, parameter optimisation, and generalisation persist, particularly in the dynamic and complex environments of clinical surgery. Surgical instruments vary in shape and interact with ever-changing surroundings, complicating segmentation tasks. Traditional methods have often relied on deeper or more complex networks to achieve higher precision, but this comes at the cost of increased parameter demands and memory usage. The proposed MBTPDS-Net model achieved the highest segmentation accuracy with the lowest parameter count among the tested methods. The detail description as follow:

The effectiveness of MBTPDS-Net was evaluated using a multi-modal fusion dataset, and its performance was compared to several SOAT segmentation networks. These included CNN-based methods such as UNet [273], UNet++[259], FCN_VGG13, FCN_VGG16 [288], SegNet [289], DeepLabv3+ [290], and JSUnet [282, 291]. Details of the implementation and validation processes for these methods are illustrated in **Figure 6.8**.

Table 6.2 shows the MIoU and Dice scores of the seven competing methods were significantly lower than those of MBTPDS-Net. The proposed network achieved segmentation results closer to ground truth, outperforming other models in terms of accuracy. Specifically, MBTPDS-Net recorded the highest segmentation performance with Dice and MIoU scores of 95.44% and 95.54%, respectively. Compared to the best performing alternative, JSUnet, MBTPDS-Net improved the IoU-SI value for surgical instrument segmentation by 4.09%, while requiring approximately 1.5 times fewer parameters. Moreover, JSUnet exhibited segmentation discontinuities in Dataset-A and deletions in Dataset-D, which undermined its robustness.



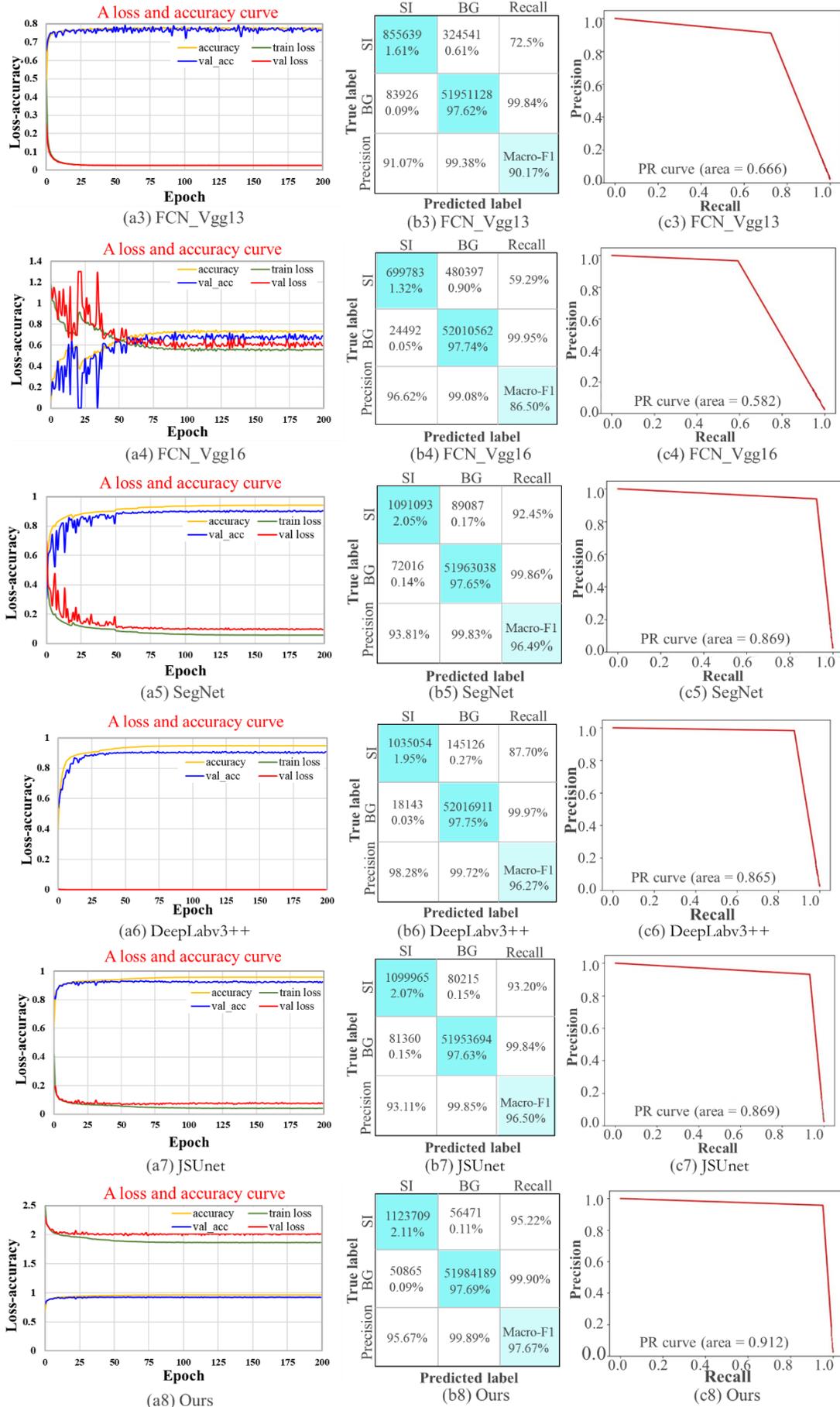


Figure 6.8: Performances of different segmentation methods.

Table 6.2: Performance comparison of segmentation methods.

Methods	Params (M)	Acc (%)	Sen (%)	Spe (%)	Mcc (%)	Macro-F1 (%)	Dice (%)	IoU (%)		MIoU (%)	AROC	PR- area	Fps
								BG	SI				
Unet	7.759	99.17	67.25	99.89	78.96	88.92	78.26	99.16	64.29	81.73	0.8357	0.6366	<u>273.6</u>
Unet++	9.041	99.39	79.51	99.84	85.26	92.51	85.33	99.38	74.41	86.89	0.8967	0.7365	241.16
FCN_Vgg13	128.953	99.23	72.50	99.84	80.89	90.17	80.73	99.19	66.86	83.03	0.8617	0.6663	226.45
FCN_Vgg16	134.313	99.05	59.29	99.95	75.30	86.50	73.49	99.04	58.09	78.56	0.7962	0.5819	215.40
SegNet	15.551	99.69	92.45	99.86	92.97	96.49	93.12	<u>99.69</u>	87.13	93.41	0.9615	0.8689	211.79
DeepLabv3+	<u>6.435</u>	99.69	87.70	<u>99.97</u>	92.69	96.27	92.68	<u>99.69</u>	86.38	93.03	0.9383	0.8646	202.97
JSUnet	8.612	<u>99.70</u>	<u>93.20</u>	99.84	<u>93.00</u>	<u>96.50</u>	<u>93.16</u>	<u>99.69</u>	<u>87.19</u>	<u>93.44</u>	<u>0.9652</u>	<u>0.8693</u>	221.89
Proposed	5.688	99.80	95.22	99.90	95.34	97.67	95.44	99.79	91.28	95.54	0.9756	0.9120	233.83

JSUnet, despite being competitive, achieved an IoU value of only 87.19% for surgical instrument segmentation. Similarly, UNet and UNet++ demonstrated poor performance, with IoU values of 64.29% and 74.41%, respectively. MBTPDS-Net outperformed UNet by improving the Dice score by 17.18% and the MIoU value by 13.81%. Compared to UNet++, MBTPDS-Net demonstrated a 10.11% improvement in Dice score and an 8.65% improvement in MIoU value.

DeepLabv3+, while achieving a respectable IoU value of 86.38% with the fewest parameters among the six competing methods, suffered from segmentation fault zones in Dataset-A and Dataset-F, as shown in **Figure 6.9** (i). Similarly, FCN_VGG16, which had the highest number of parameters (approximately 134 million), performed the worst, achieving an IoU value of only 58.09%. FCN_VGG13, with fewer parameters, slightly improved upon this performance, achieving an IoU value of 66.86%—an 8.77% improvement over FCN_VGG16. These findings suggest that deeper networks do not necessarily lead to better segmentation performance for surgical instruments.

The encoder-decoder MBTPDS-Net framework, built on an improved FCN_VGG13 encoder and a triple-pyramid decoder, demonstrated significantly superior performance compared to conventional FCN_VGG13 and FCN_VGG16 networks. Both FCN_VGG13 and FCN_VGG16 failed to segment surgical instruments in Datasets B, C, E, and F, with marginal success observed only in Datasets A and D, as depicted in **Figure 6.9**. Furthermore, SegNet performed moderately well, achieving a Dice score of 93.12% and an IoU-SI value of 87.13%. However, these metrics were 2.32% and 4.15% lower, respectively, than those achieved by MBTPDS-Net.

Furthermore, the above experimental results highlighted the superiority of two-stage networks over single-stage networks in surgical instrument segmentation tasks, such as UNet++, SegNet, DeepLabv3+, and JSUnet, consistently outperformed single-stage networks such as FCN_VGG13 and FCN_VGG16, except for UNet. Although UNet and UNet++ achieved higher Fps, their segmentation accuracy was lower, and their parameter requirements were higher compared to MBTPDS-Net. Similarly, SegNet, DeepLabv3+, and JSUnet delivered slightly lower segmentation performance than MBTPDS-Net while requiring more parameters and exhibiting slower Fps.

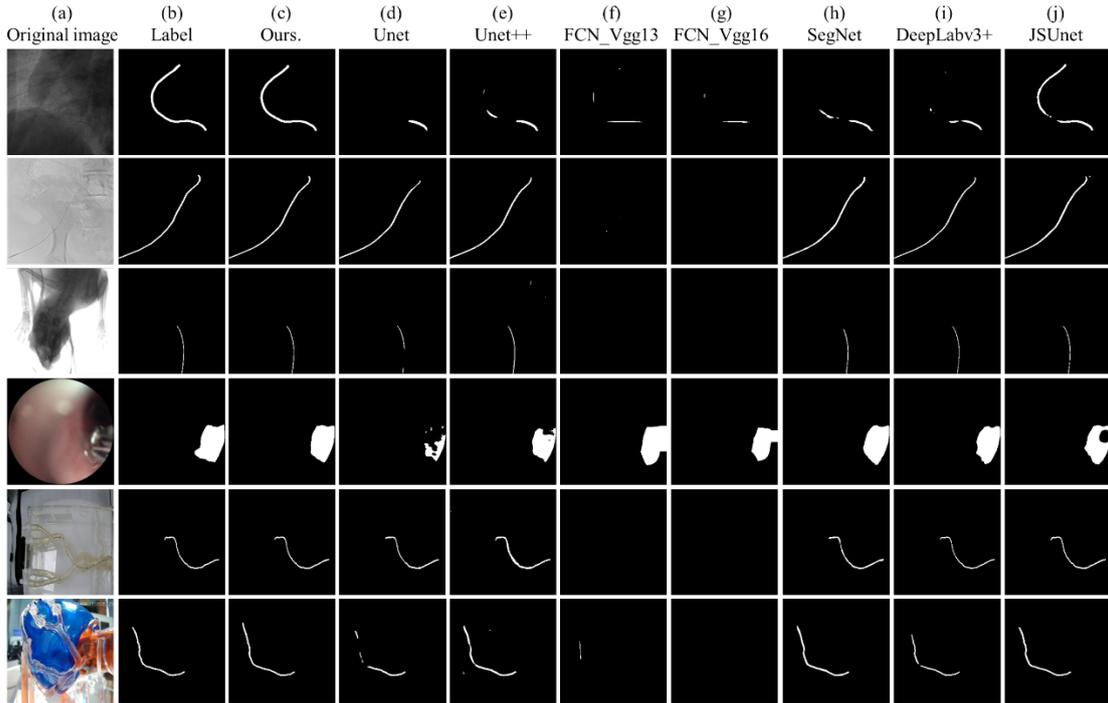


Figure 6.9: Performance comparison of the segmentation methods.

In summary, the suggested multibranch feature fusion network combined with a triple-pyramid structure improved accuracy and generalisation while alleviating the parameter load in surgical instrument segmentation. Its efficiency and effectiveness established a new benchmark in surgical instrument segmentation, tackling key challenges in precision, generalisation, and optimisations of computational resources.

6.4.2 Encoder and decoder performance in semantic segmentation

To understand the impact of various optimisations in the improved FCN_VGG13 encoder and the triple-pyramid decoder on performance, segmentation performance assessments were conducted. These assessments compared typical encoders coupled with the triple-pyramid decoder and the improved FCN_VGG13 encoder paired with typical decoders, four existing encoder models were implemented with the triple-pyramid decoder, and five existing decoder models were paired with the multibranch encoder. These were applied to the datasets, as detailed in **Table 6.3**.

The results showed that the ResNet50 encoder, when combined with the triple-pyramid decoder, required higher memory, with approximately 26.34 million parameters. It showed poorer segmentation performance, with a Dice score of 3.21% and an MIoU of 42.45%. The Xception model, known for its fast computation and memory efficiency in related computer vision tasks such as image object detection [292], showed a slight performance decrease when paired with the triple-pyramid decoder. It had a 1.18% reduction in the Dice score and a 1.09% reduction in MIoU compared to the proposed MBTPDS-Net, despite having higher parameter requirements at 6.114 million. Although the RepVgg model combined with the decoder had higher parameter requirements at 8.893 million, it showed a slight improvement with a Dice score of 95.47% and an MIoU of 95.57%

Table 6.3: Performance comparison of MBTPDS-Net with different encoders and decoders.

Type	Method	Dice (%)	IoU (%)		MIoU (%)	Params (M)
			BG	SI		
Decoder	ResNet50	3.21	83.26	1.63	42.45	26.340
	Xception	94.26	99.75	89.15	94.45	6.114
	RepVgg	95.47	99.8	91.34	95.57	8.893
	MobileNetV2	4.34	95.11	3.79	49.45	0.735
Encoder	Unet	91.07	99.6	83.6	91.6	18.014
	ResNet	76.33	99.14	61.73	80.44	10.227
	SegNet	78.21	99.19	64.21	81.70	9.047
	JSUnet	7.78	60.9	4.05	32.48	10.227
	Pyramid	10.44	73.72	5.51	39.61	5.084

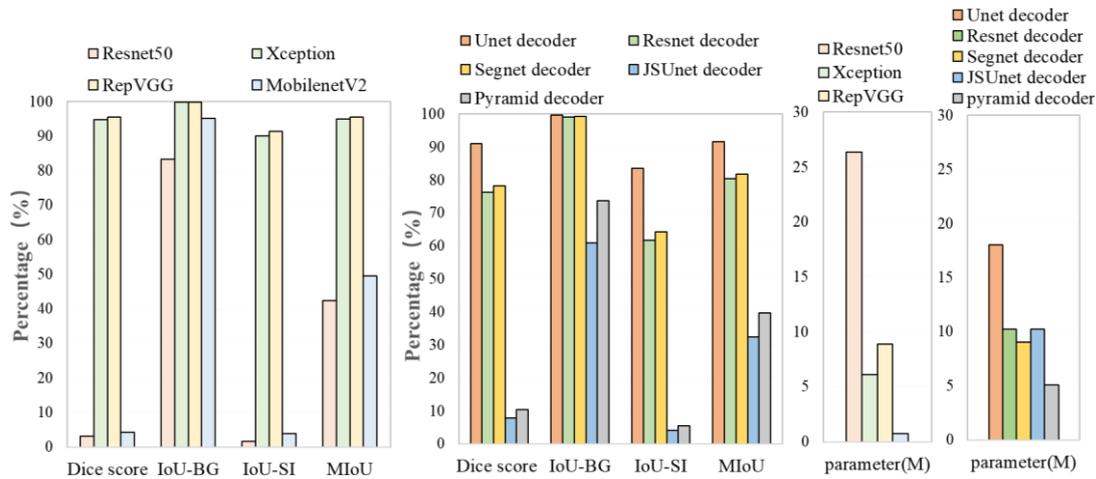


Figure 6.10: Performance of different encoders and decoders with proposed method.

compared to the proposed method. On the contrary, MobileNetV2, enabled with the triple-pyramid decoder, had the lowest parameter requirements at 0.735 million, but was unsuitable for surgical instrument segmentation. It showed poor segmentation performance, with a Dice score of 4.34% and an MIoU value of 49.45%.

These results indicated that the Xception and RepVgg encoders, when paired with the triple-pyramid decoder, achieved better segmentation performance than ResNet-50 and MobileNetV2, as illustrated in **Figure 6.10**. This indicated that the proposed triple-pyramid decoder was more sensitive to the feature representations extracted by the Xception and RepVgg encoders. The use of deep separable convolutions by Xception enhanced its ability to capture fine-grained features by increasing network depth and width. RepVgg improved inference performance by structural reparameterization, resulting in high complexity and expressiveness of the model. Thus, Xception and RepVgg were better suited for extracting features for instrument segmentation in multimodal fusion data, effectively capturing details and semantic information.

Moreover, the Unet decoder, when integrating with the multi-branch encoder, achieved a 91.07% Dice score and a 91.60% MIoU value. This represented a reduction of 4.37% in the Dice score and 3.94% in the MIoU value compared to the proposed MBTPDS-Net network. The ResNet and SegNet decoders, each integrating the multi-branch encoder module, achieved lower segmentation

performance, with Dice scores of 76.33% and 78.21% in Dice score and MIoU values of 80.44% and 81.70%, respectively. Additionally, they had higher parameter requirements at 10.227 and 9.047 million. These results indicated that the Unet decoder coupled with the improved VGG13 encoder achieved the highest segmentation accuracy, so that the Unet decoder could better reconstruct image details and boundary information when paired with the improved VGG13 encoder. On the contrary, the ResNet and SegNet decoders achieved relatively low segmentation accuracy, suggesting that they performed worse in feature reconstruction and detail retention. JSUnet, an improved segmentation method derived from our previous studies [291], demonstrated suboptimal performance when paired with the multi-branch encoder module. Specifically, it achieved a Dice score of 7.78% and an MIoU value of 32.48%, highlighting its inability to effectively process multi-modal fusion feature maps with this combination. This indicates that the JSUnet decoder, when integrated with the multi-branch encoder, is unsuitable for handling complex multi-modal surgical instrument segmentation tasks.

Similarly, despite the reduced parameter requirement (approximately 5.084 million), the combination of the multi-branch encoder module and the pyramid decoder module performed poorly. This pairing resulted in a Dice score of 10.44% and an MIoU value of 39.61%, reflecting an 85% decrease in the Dice score and a 55.93% decrease in MIoU compared to the triple-pyramid decoder module. These results suggest that the pyramid decoder module failed to preserve critical information during the feature fusion and up-sampling processes. It also lacked the capacity to effectively couple with the improved VGG13 encoder, which is instructed the traditional single pyramid decoder module lacks the sophistication needed to exploit global and local contextual features effectively, leading to its inability to improve segmentation outcomes.

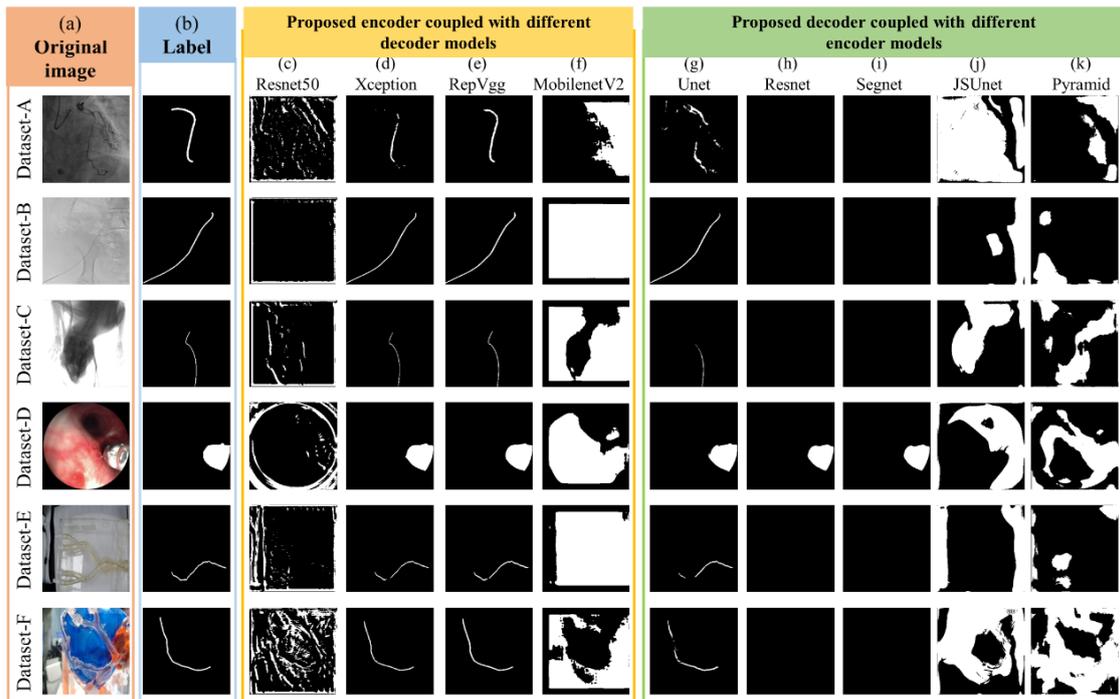


Figure 6.11: Results of proposed segmentation methods with different encoders and decoders.

The lower accuracy of both the JSUnet decoder and the pyramid decoder highlights their limitations in leveraging global contextual information and generating multi-scale image-level features, as depicted in Figure 6.11. These deficiencies resulted in poor recovery of edge and texture details, which are crucial for enhancing segmentation accuracy.

6.4.3 Loss functions for semantic segmentation

To achieve optimal performance in DL architectures for semantic segmentation, it is essential to select an appropriate loss function. For tasks with class imbalance, an effective objective function not only facilitates model convergence, but also improves segmentation accuracy. This study evaluated the performance of the proposed MBTPDS-Net using several well-known loss functions commonly applied in medical semantic segmentation. The assessment included six categories of loss functions: Dice loss, binary cross-entropy, Jaccard loss, Tversky loss, focal loss, and the proposed Lossxy function. The results for each function, listed in **Table 6.4** and **Figure 6.12**, provided critical insights into their suitability for surgical instrument segmentation.

Dice loss, binary cross-entropy, and Jaccard loss prioritise minimising loss, but do not address class imbalance. Among these, the Jaccard loss emerged as the most effective, achieving a Dice score of 94.59% and an MIoU value of 89.75%. However, these functions often fail in imbalanced datasets by disproportionately favouring dominant classes. On the contrary, the Tversky loss, the Focal loss, and the Lossxy function incorporate modulating and balancing factors, allowing the model to focus more on hard-to-classify examples. Among these, Lossxy demonstrated superior performance, achieving a Dice score of 95.13% and an IoU-SI value of 90.71%. This highlights the importance of addressing class imbalance through adaptive weighting.

The proposed MBTPDS-Net leveraged a hybrid MixLoss function, which combined Jaccard loss and Lossxy (with $\alpha = 0.1$, $\gamma = 0.2$) to address sample imbalance and improve segmentation performance. The MixLoss function enabled MBTPDS-Net to achieve the highest results, with a Dice score of 95.44%, an MIoU of 95.54%, and a PR area of 91.20%. Compared to single-loss functions, MixLoss showed a significant improvement, outperforming Jaccard loss by 0.85% in the Dice score and 1.53% in IoU-SI, and Lossxy by 0.31% in the Dice score and 0.57% in IoU-SI.

Despite the class imbalance in datasets A through F, the hybrid MixLoss function consistently

Table 6.4: Performance comparison with different loss functions.

Category	Loss Function	PR-Area	Dice (%)	IoU (%)		MIoU (%)
				BG	SI	
A: minimize loss without sample imbalance	Dice loss	0.8811	93.79	99.71	88.30	94.00
	Binary Cross-entropy	0.8884	94.08	99.74	88.83	94.28
	Jaccard loss	0.8961	94.59	99.76	89.75	94.75
B: minimize loss with sample balancing factor	Tversky loss	0.8966	94.63	99.75	89.81	94.78
	Focal loss	0.8989	94.68	99.77	89.89	94.83
	Loss _{xy}	0.9061	95.13	99.78	90.71	95.24
C: mix loss	Our proposed	0.9120	95.44	99.79	91.28	95.54

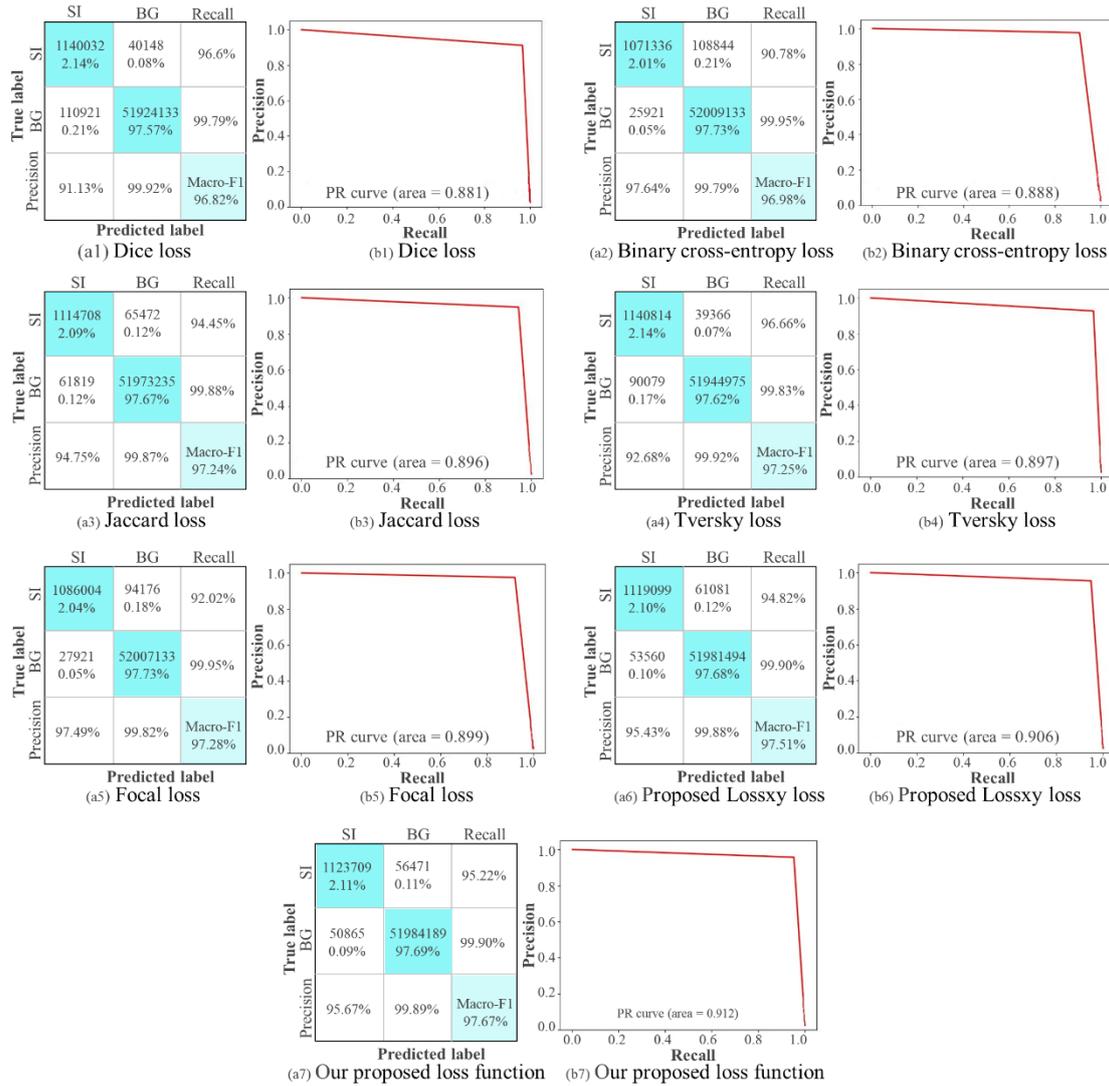


Figure 6.12: Performance of proposed MBTPDS-Net based on different loss function. (Notes: SI: surgical instrument pixel, BG: background pixel)

delivered superior performance. Single loss functions such as Dice loss, binary cross-entropy, and Jaccard loss tended to overfit dominant classes, neglecting minority class features, and reducing generalisation ability. Although the Tversky loss, Focal loss, and Lossxy partially mitigated the class imbalance through parameter adjustment, they occasionally overcompensated for minority classes, leading to a trade-off in overall performance. The MixLoss function, by combining Jaccard loss and Lossxy, effectively balanced the representation of different classes, enhancing the model's ability to handle imbalanced datasets while ensuring better generalisation.

The findings underscore that no single loss function is universally optimal for all tasks. While the MixLoss function demonstrated superior performance, further improvements could be achieved by integrating additional loss functions tailored to specific issues such as curve smoothing and extreme class imbalance. Parameter tuning also plays a critical role in optimising segmentation performance, particularly for surgical instruments in robot-assisted interventions. Future work could explore the combination of complementary loss functions to further enhance the robustness and accuracy of the model in complex clinical scenarios.

6.4.4 Performance of training strategy in segmentation

The performance of the proposed MBTPDS-Net model was evaluated using both multi-modal fusion datasets and single-modal datasets. **Table 6.5** and **Figure 6.13** presented that the model trained on multi-modal fusion datasets demonstrated superior segmentation accuracy compared to training on single-modal datasets, achieving an ultrahigh MIoU value of 95.54%. In particular, the segmentation IoU value for surgical instruments using the MBTPDS-Net network trained in multimodal fusion datasets exceeded that of single-modal training by 23.45%, 7.45%, 11.53%, 9.33%, and 17.23% in Datasets-A, -B, -C, -E, and -F, respectively, while showing marginally lower performance on Dataset-D.

Table 6.5: Performance comparison of different models across different datasets.

Multi-Fusion	Single-Model	Dataset	Dice (%)	IoU (%)		MIoU (%)	Fps
				BG	SI		
√	-	Multi-modal	95.44	99.79	91.28	95.54	233.83
√	-	A	90.82	99.60	83.19	91.39	5.46
-	√		80.83	99.60	67.83	83.71	23.98
√	-	B	93.65	99.86	88.06	93.96	29.01
-	√		91.20	99.81	83.83	91.82	50.07
√	-	C	90.69	99.91	82.96	91.44	56.59
-	√		88.74	99.90	79.75	89.83	60.20
√	-	D	99.12	99.87	98.26	99.06	67.58
-	√		97.42	99.62	94.96	97.29	111.55
√	-	E	94.75	99.94	90.03	94.99	35.77
-	√		90.08	99.89	81.95	90.92	28.51
√	-	F	95.22	99.92	90.87	95.40	13.82
-	√		85.09	99.75	74.05	86.90	25.40

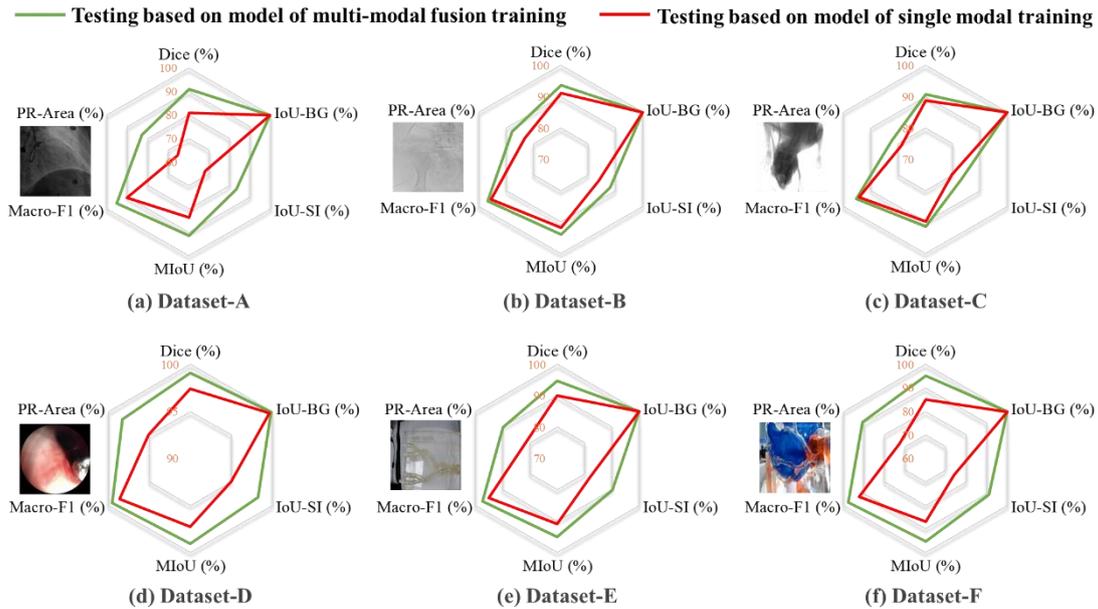


Figure 6.13: Performance of MBTPDS-Net network based on different training strategy. (Notes: the green line indicates: the MBTPDS-Net model, was trained based on multi-modal fusion datasets, then to test each single dataset; red line indicated: the MBTPDS-Net model, was trained based on single-modal datasets, then to test corresponding to single dataset.)

Furthermore, the MBTPDS-Net network, trained on multi-modal fusion datasets, was evaluated across various single-modal datasets, as depicted in **Table 6.5**. The results demonstrated that employing multi-modal fusion improved the MBTPDS-Net network's ability to segment surgical instruments in each single modality compared to training solely on the corresponding single-modal datasets. For instance, the MBTPDS-Net network from multi-modal fusion training enhanced the Dice score by 9.99% and the IoU-SI value by 15.36% in Dataset-A, by 2.45% and 4.23% in Dataset-B, by 1.95% and 3.21% in Dataset-C, by 1.70% and 3.30% in Dataset-D, by 4.67% and 8.08% in Dataset-E, and by 10.13% and 16.82% in Dataset-F, respectively, as shown in **Figure 6.13**.

These results indicated that the proposed MBTPDS-Net model demonstrated better generalisability when trained on multi-modal fusion datasets compared to single-modal training. The diverse nature of multi-modal datasets allowed the model to adapt to various noise levels and variations, improving its performance in complex tasks. Furthermore, applying the multi-modal fusion-trained model to single-modal datasets consistently resulted in higher segmentation accuracy compared to models trained on the respective single-modal datasets. This highlights the role of multi-modal datasets in stabilising model performance across different environments and enhancing generalisation.

Moreover, among the datasets, the MBTPDS-Net achieved its best segmentation results on the single-modal endoscopic dataset (Dataset-D). The superiority of the results can be attributed to the inherent characteristics of the images. Unlike angiographic datasets (Datasets A, B, and C), which track slender and flexible guidewires with minimal contrast, endoscopic images capture biopsy forceps with a pronounced contrast against surrounding tissue. The rigid nature of the forceps, coupled with their high visibility, facilitated accurate segmentation. These factors enabled the model to effectively discern the relevant pixel attributes, resulting in higher segmentation accuracy for well-connected and rigid instruments.

In addition, to optimise the efficiency of the model, separable convolutions were used to replace standard convolutions in the designed encoder network. The impact of this substitution on segmentation performance was analysed, as shown in **Table 6.6**. Although separable convolutions significantly reduced the number of parameters, 1.534 million compared to 5.688 million for standard convolutions, standard convolutions outperformed their separable counterparts in segmentation accuracy. Specifically, the IoU-SI values for surgical instrument segmentation using the MBTPDS-Net network with standard convolutions were 7.6%, 3.07%, 2.5%, 1.47%, 5.38% and 10.61% higher than those achieved with separable convolutions in datasets A, B, C, D, E, and F, respectively. Similarly, the Dice score improvements using standard convolutions were 4.72%, 1.76%, 1.52%, 0.75%, 3.06%, and 6.17% higher on the same datasets. Moreover, the MIoU values for segmentation using the MBTPDS-Net network with standard convolutions reached 95.54% on the multi-modal fusion dataset, 91.39% on Dataset-A, 93.96% on Dataset-B, 91.44% on Dataset-C, 99.06% on Dataset-D, 94.99% on Dataset-E, and 95.40% on Dataset-F. These MIoU values consistently surpassed those obtained with separable convolutions in all datasets.

These results underscore a trade-off between computational efficiency and segmentation performance. While separable convolutions offer a considerable reduction in parameter count, standard convolutions deliver superior segmentation accuracy, particularly for tasks requiring high precision, albeit with a greater computational cost.

Table 6.6: Performance of the segmentation model across different convolutions.

Method	Conv	Sepcov.	Dataset	Dice (%)	IoU (%)		MIoU (%)	
					BG	SI		
Proposed	√	-	MMF	95.44	99.79	91.28	95.54	
	-	√		95.23	99.78	90.89	95.34	
	√	-	A	90.82	99.60	83.19	91.39	
	-	√		86.10	99.40	75.59	87.49	
	√	-	B	93.65	99.86	88.06	93.96	
	-	√		91.89	99.82	84.99	92.41	
	√	-	C	90.69	99.91	82.96	91.44	
	-	√		89.17	99.90	80.46	90.18	
	√	-	D	99.12	99.87	98.26	99.06	
	-	√		98.37	99.76	96.79	98.27	
	√	-	E	94.75	99.94	90.03	94.99	
	-	√		91.69	99.91	84.65	92.28	
	√	-	F	95.22	99.92	90.87	95.40	
	-	√		89.05	99.81	80.26	90.04	
	Params	√	-			5.688 M		
		-	√			1.534 M		

6.4.5 Performance comparison of SAM annotation and manual labelling

With advances in AI technology, generative AI applications have been increasingly applied in diverse industries. ChatGPT, developed by OpenAI, is a chatbot built on a large language model, renowned for its conversational interactivity and human-like performance in various cognitive tasks, including those in medicine [293]. In computer vision, the Segment Anything Model has garnered significant attention as an innovative foundational model for promotable segmentation [294]. The exceptional performance and interactivity of SAM in image segmentation have established it as a highly regarded model in the field. However, improving surgical instrument segmentation performance remains constrained by the insufficient availability of labelled surgical instrument datasets during the model’s pre-training phase.

To determine whether SAM annotation could replace manual annotation to reduce time and costs, this study evaluated the influence of SAM annotations on segmentation performance and assessed the gap between SAM and manual annotations. SAM was used to mark surgical instruments and the impact of its annotations compared to manual annotation standards on segmentation accuracy was investigated. The performance of the proposed MBTPDS-Net network was assessed using manual annotation standards as a benchmark against SAM annotations for masking surgical instruments. Segmentation results are visualised in **Figure 6.14** and **Figure 6.15**.

The results showed that the MBTPDS-Net network, using SAM annotations, achieved a Dice

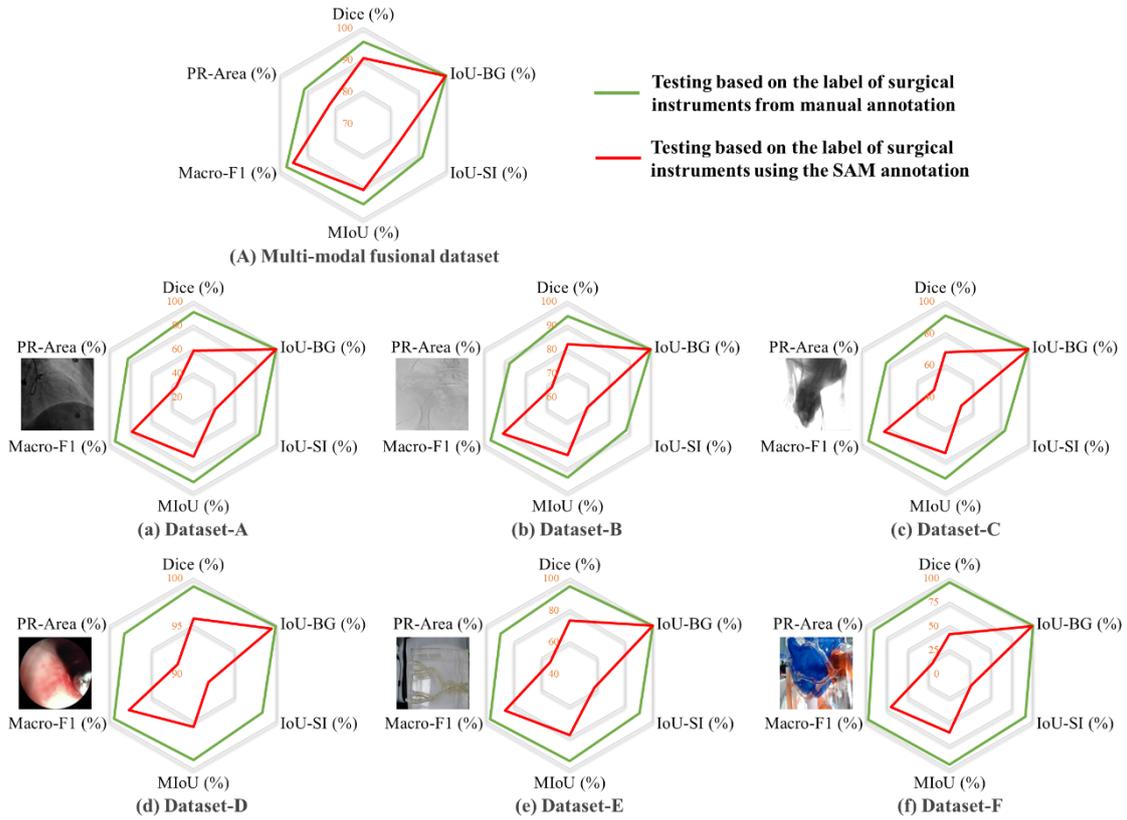


Figure 6.14: Performance comparison of MBTPDS-Net network using manual and SAM annotation. (Notes: the green line indicates: the MBTPDS-Net model, was trained based on datasets from manual annotation, then to test each datasets including multi-modal fusion datasets and each single modal dataset; red line indicated: the MBTPDS-Net model, was trained based on dataset from SAM large model annotation, then to test corresponding to dataset including multi-modal fusion datasets and each single modal dataset.)

score of 90.46% and an IoU value of 82.58%. Compared to manual annotation standards, these figures represent a decrease of 4.98% in the Dice score and 8.7% in the IoU. The segmentation efficacy of the MBTPDS-Net network based on SAM annotations also showed reductions in Dice score and IoU values across six single-modal datasets compared to manual annotation standards. Specifically, reductions of 32.61% in Dice score and 42.13% in the IoU value of surgical instruments were observed in Dataset-A; reductions of 11.72% in Dice score and 18.67% in IoU value were observed in Dataset-B; reductions of 22.72% in Dice score and 31.48% in IoU value were observed in Dataset-C; reductions of 3.37% in dice score and 6.42% in IoU value were observed in Dataset-D; reductions of 21.48% in Dice score and 32.21% in IoU value were noted in Dataset-E; and reductions of 54.39% in Dice score and 65.22% in IoU value were noted in Dataset-F, respectively.

The findings suggest that manual annotations yield superior segmentation results compared to SAM annotations. SAM struggled to accurately segment entire surgical instruments, particularly in complex structural environments, often producing fragmented outcomes. This limitation can be attributed to the highly specialised appearance of surgical instruments, which differ significantly from natural objects typically encountered in SAM’s training data. SAM also encountered difficulties in distinguishing adjacent tissues and edges with similar greyscale intensities, as seen in Dataset-F. Another challenge for SAM was its dependence on explicit frame-by-frame prompt boxes,

which is impractical in surgical contexts. Instruments with extreme aspect ratio differences, such as guidewires, further complicated SAM's performance as it struggled to distinguish target instruments from the background. Consequently, SAM exhibited challenges in segmenting elongated and curved targets, failing to capture complete objects or misidentifying them as nonsurgical instruments in datasets like Dataset-A, Dataset-C, and Dataset-E. The segmented instrument contours generated by SAM were irregular and lacked refinement, resulting in lower accuracy and efficiency compared to manual annotation.

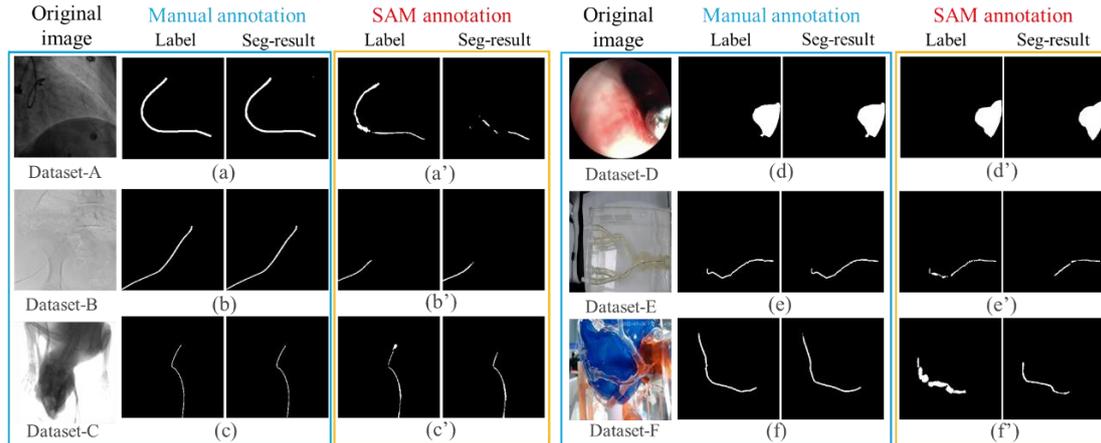


Figure 6.15: Performance comparison of MBTPDS-Net network using manual and SAM annotations.

6.5 Chapter Summary

This chapter presents an innovative network for feature fusion with multiple branches, combined with a deep neural structure using a triple-pyramid design, known as MBTPDS-Net, intended for precise segmentation of surgical tools. This approach aims to equip surgeons with detailed visual data, thus enhancing their natural ability to manipulate instruments during operations. The model leverages an enhanced VGG13 as a multi-branch encoder for feature extraction, proficiently capturing multi-scale feature representations, including fine edge and texture intricacies. To manage these features, the triple-pyramid decoder applies cross-stacked techniques to widen the receptive field, facilitating the merging of interconnected multi-scale features and boosting the capture of global contextual semantic data. The research validated this proposed approach using self-constructed multi-modal fusion datasets featuring various surgical instruments. MBTPDS-Net showcased exceptional performance, attaining a Dice score of 95.44% and an MIoU of 95.54% in these datasets. In addition, it achieved fast segmentation rates of 233.83 frames per second with fewer parameter needs. These results surpassed those of other semantic segmentation techniques, such as Unet, Unet++, FCN_VGG13, FCN_VGG16, SegNet, DeepLabv3+, and JSUnet, achieving top-tier segmentation results. Comparative trials on the multi-modal fusion dataset highlighted MBTPDS-Net's unique benefits. By adeptly extracting global and local contextual features, the model delivered accurate segmentation output with a reduced parameter load, as evidenced by thorough quantitative and qualitative evaluations. These findings highlight the potential of

MBTPDS-Net to establish a new standard in surgical tool segmentation tasks. Furthermore, its versatility extends to segmenting additional endovascular and surgical tools, including forceps, scalpels, needle tips, stents, and balloons, illustrating its extensive applicability in various medical domains.

Chapter 7: Conclusions and Future Works

7.1 Conclusions

Interventional robotics has recently been implemented to perform procedures such as diagnosis and surgery, either through natural openings or by making small incisions into internal cavities. This advancement holds significant promise in expanding the boundaries of surgical procedures. The limitations posed by traditional interventional surgery, which involve prolonged X-ray exposure for both the surgeon and the patient, have driven the development of robotic technology for endovascular interventional surgery. This technology involves manoeuvring surgical instruments and end effectors through complex anatomical pathways, indicating possible transformative changes in endovascular surgical practices. However, this new surgical approach also presents certain technological and technical issues, such as the lack of direct tactile feedback, which need to be resolved to improve safety and facilitate greater acceptance in clinical settings around the world.

This research concentrated on developing multi-sensor methods for surgeons' intuitive manipulation behaviours and visual-based feedback approaches for improving intuitive manipulation in an underactuated master-slave vascular interventional robotic system with spatial flexibilities. The focus was on designing efficient models for precise intuitive manipulation of robots along spatially flexible paths, based on the design of vascular interventional robotic systems for accessing complex and narrow pathways. The studies were based on a developed vascular interventional robotic system, modelling an intuitive manipulation model to demonstrate the internal link between intuitive manipulation and the performance of robot-assisted surgical tasks, and using visual-based modelling to address the absence of direct force feedback that results in inadequate robot-assisted intuitive manipulation characteristics. Improved intuitive manipulation enables this robotic system to accurately deliver a catheter or guidewire through minimally invasive single port procedures along various blood pathways to reach the lesion site. Hence, in **Chapter 3**, the clinical association between intuitive surgeon manipulation and the performance of robot-assisted interventional tasks is explored in detail, highlighting the significant influence of manipulation performance on surgical outcomes (success or failure performance). Using multi-sensor quantitative analysis of the operator's manipulation behaviour, the manipulation patterns that lead to successful or unsuccessful performance in robot-assisted surgical tasks are identified, offering insights into methods for improving intuitive manipulation performance. **Chapter 4** further investigates the inherent relationship between collaborative performance of humans and robots and intuitive manipulation in the context of the absence of direct force feedback in robot-assisted interventional procedures. Generally, **Chapters 3** and **4** highlight the vital role of manipulation performance in improving surgical efficiency and safety.

Chapter 5 focusses on improving intuitive manipulation in robot-assisted interventional surgery without tactile force feedback. This improvement is achieved through visual modelling, specifically

tracking instruments in distal blood vessels during procedures. Numerous accomplished robotic surgeons believe that visual signals, such as local tissue deformation due to tension, retraction, or needle insertion, can effectively compensate for the absence of sensory force feedback. This technology plays a crucial role in increasing manipulation performance, especially in master-slave robot-assisted interventional tasks where haptic feedback is lacking. Instrument tracking offers not only real-time positional data, but also helps predict the instrument's movement path, thereby allowing operators to exercise better control throughout the surgical procedure.

Chapter 6 emphasises the improvement of instrument recognition accuracy and reliability through visual-based modelling techniques. The outcomes not only cover technological advancements but also offer a more profound understanding of surgeon manipulation patterns and performance improvements. This chapter aims to present a more comprehensive and efficient robotic-assisted system for interventional surgery, ultimately striving for optimal surgical results and ensuring patient safety. The major research breakthroughs of this dissertation, which also serve as its key contributions, include:

- Modelling a multi-level manipulation recognition model including initial-decision, and motion-decision, and mixed-decision layers for indicating the internal relation between operators' manipulation patterns based on multi-sensors different trained strategies and outcome of robot-assisted surgical tasks for deeper understanding of surgeons' manipulation patterns during robot-assisted interventional surgeries. This model provides real-time feedback, prompting interventionist' technical skill to adjust control strategies to prevent tissue damage and optimize procedural outcomes. When surgeons understand which manipulation strategies are critical success factors and learn how to avoid ineffective or potentially harmful manipulation patterns, their trust in and acceptance of robot-assisted surgery can be significantly improved. This understanding not only enhances their confidence during intraoperative decision-making but also promotes more effective human-robot collaboration by aligning the robotic system's behaviour with the surgeon's intent. Furthermore, by identifying and modelling successful manipulation strategies, the system can provide targeted feedback and training support, contributing to skill acquisition and shortening the learning curve for new users. Such contributions are essential for improving procedural safety, operational efficiency, and overall user satisfaction in robot-assisted vascular interventions.
- Development of a manipulation-based machine-learning framework designed to analyse the synergy performance between operators and the robot. Investigation of the impact of delay factors on the synergy ratio between human and robot from three perspectives, namely, no delay, constant delay, and variable delay. Analysis of interaction forces, such as distal force and haptic force, and manipulation speed concerning complex endovascular paths. This analysis offers insights into how operators with different manipulation technical skills to adjust their control strategies for achieving good outcome of robot-assisted surgical tasks as

well avoiding causing damage to endovascular vessels due to excessive force while still providing sufficient tension to navigate complex paths. By evaluating the cooperative performance between the surgeon and the robot, the surgeon can more intuitively perceive the accuracy and quality of the operative process. Through the analysis of manipulation speed and interaction force, the surgeon can derive the most efficient and precise manipulation strategies, thereby facilitating more effective use of the robotic system and improving both the acceptance and proficiency of robot-assisted surgery.

- Visual perception modelling method compensates the effects of the lack of direct force feedback on the performance of robot-assisted cardiothoracic surgery. An improved U-Net model with semantic segmentation is proposed to extract guidewire feature maps from X-ray images during a robot-assisted endovascular interventional procedure. The proposed model obtains a better segmentation performance for a small object. Two experiments are carried out to prove the effectiveness of the proposed method for tiny objects detection.
- A two-stage guidewire endpoint detection method is proposed to track the guidewire endpoint position including skeletonization processing, removing the bifurcation pixel point, repairing the breakage-band of the guidewire pixel feature maps, and endpoint detection based on a pixel-adjacent-relationship. The good performance of our proposed method is verified compare four common heatmap methods and six typical heatmap regression methods proposed. This contribution enables the surgeon to perceive the position of the instrument and its spatial relationship with surrounding tissues more accurately. This enhanced perception allows for more efficient control and fine-tuning of the instrument's motion trajectory, thereby improving the surgeon's manipulation awareness and operational safety. Moreover, it facilitates dynamic adjustment of manipulation strategies based on anatomical context, helping the surgeon to avoid vessel injury, maintain procedural efficiency, and increase confidence in complex navigation tasks.
- The proposed pixel-adjacent-relationship-based method also demonstrates its effectiveness in detecting the maximum bending regions and computing the angle value. By utilizing the bending angle information of the instrument, the surgeon can anticipate and respond more effectively to unexpected intraoperative situations, such as sudden vascular resistance changes or potential deviations in navigation paths. This predictive capability enhances the surgeon's situational awareness and control over the operating environment, particularly in complex lesions. Furthermore, bending feedback allows the surgeon to infer tool-tissue interaction states, assess anatomical constraints, and make timely adjustments to manipulation strategies. This significantly improves the surgeon's sensing of tool manipulation, supports safer and more precise navigation.
- A multi-branch feature coupled with a triple-pyramid deep architecture segmentation method was proposed based on multimodal fusion dataset, achieved SOAT segmentation performance during robot-assisted interventional procedures. The improved VGG13

encoder effectively generated feature maps of different scales and reduced parameter requirements. The triple-pyramid model fused different scale features to generate better output. This lightweighted network, trained on multi-modal fusion dataset, learned target characteristics from each modality, and demonstrated better generalisation capability and computational efficiency compared to single-modal data. The introduction of a Mixloss function into the multimodal fusion dataset segmentation task enhanced the model's adaptability to diverse image data types, effectively addressing class imbalance while improving the model's ability to recognize different classes.

In summary, these proposed high-precise manipulation models and perception methods improve the safety of robot-assisted interventional surgery by enhancing the surgeons' intuitive manipulation and the surgeon's ability to perceive the position of the instrument and its spatial relationship with surrounding tissues. They help surgeons identify manipulation strategies that are critical to surgical success and avoid techniques that may lead to failure. In addition, the models facilitate intuitive assessment of human-robot collaboration during the procedure, enabling more efficient and precise manipulation of the robotic system. These advancements enhance the surgeon's perception of manipulation, increase trust in the robotic system, and improve overall acceptance and proficiency in robot-assisted surgery. These contributions, discussed in **Chapters 3 to 6**, address several of the unmet technical challenges in robot-assisted endovascular intervention, specifically those related to manipulation improvement, the lack of direct force feedback, human-robot cooperation, and precise visual perception of interventional instruments. The foundations for these solutions were laid in **Chapters 1 and 2**, which outline the SOAT developments and the key issues in vascular interventional robotic systems.

7.2 Discussions on practical applications

To translate the proposed models and methods into practical clinical value, it is essential to consider their integration within real-world robot-assisted endovascular surgical systems. First, the multi-level manipulation recognition model can be embedded into the control software of the robotic system as a real-time monitoring and feedback module. This module can continuously evaluate the surgeon's operational inputs using sensor data and classify them into successful or unsuccessful manipulation patterns. Real-time feedback can then be displayed on an interface to help surgeons adjust their strategies accordingly, especially during complex navigation or resistance changes in tortuous vessels. This integration can support intraoperative decision-making and help prevent procedural errors, thus improving safety and precision.

Second, the instrument segmentation framework (MBTPDS-Net) and guidewire endpoint localization modules can be integrated into the vision perception subsystem of the robot. By embedding these methods into the intraoperative imaging pipeline (e.g., real-time X-ray or fluoroscopy feed), the robot system can automatically recognize, segment, and track the surgical instruments with high precision. These outputs can be used not only for visualization but also for

closed-loop control, enabling adaptive motion planning and haptic substitution strategies when force feedback is absent.

Furthermore, the bending angle analysis module can serve as an advanced alert system. By continuously monitoring instrument shape deformation and calculating angle changes in high-risk regions, the system can generate real-time warnings or recommend strategy adjustments to avoid vessel injury. This predictive feedback can be particularly useful in minimally invasive catheterization procedures under restricted visual and tactile conditions.

To realize these functions in practice, a unified architecture that supports data fusion from multi-modal sources (force sensors, imaging, manipulation kinematics) is necessary. The proposed modules should be encapsulated as lightweight, plug-and-play components—compatible with surgical robots' existing hardware and real-time software constraints. Hardware-level synchronization and GPU-accelerated computation will be important for ensuring low-latency responses, which are critical in dynamic surgical environments. Their successful deployment will require interdisciplinary efforts in system design, real-time computing, and regulatory adaptation.

7.3 Limitations

Conducting endovascular interventional surgeries in complex and narrow endovascular pathways is naturally labour intensive, often resulting in prolonged radiation exposure to surgeons and decreased precision and stability due to physiological tremors. Using master-slave interventional robots can alleviate these problems. In the treatment phase, these systems allow surgeons to perform minimally invasive, safe, efficient, and radiation-free procedures, allowing for accurate identification of critical endovascular stenoses and steering the interventional process to completion, especially in intricate cases, thus improving patient outcomes. However, the limitations of the skill learning model built on surgical operational methods, the human-robot synergy evaluation framework, the endovascular instrument tip localization model, and the instrument semantic segmentation model examined in this thesis are summarized as follows.

A. Surgeon's skill diversity

The skill modelling of surgeons using data-driven methods discussed in **Chapter 3** relies on the procedures performed by a small group of both interventional and nonspecialist surgeons. However, studies show that the effectiveness of robotic-assisted interventional surgery is greatly dependent on surgeon expertise and muscle memory. A comprehensive dataset that includes surgeons of different skill levels, beginning, intermediate and advanced, would enable more precise skill acquisition and objective skill evaluation. In addition, the skill modelling in this thesis is centred on the use of a single instrument, while clinical situations often require the coordination of several instruments, especially in the case of complex lesions.

B. Difference between simulated and clinical settings

Effective remote manipulation requires accurate haptic feedback, but current systems struggle to

achieve this across interfaces, affecting their support for surgeons. Capturing and using haptic data from master, slave, and remote interfaces is important for modelling surgeon skills and creating a human-robot collaboration framework for surgery. Our framework uses EMG data, data gloves, and force information; however, while sensor-based validation is useful for measuring catheter movements, it is not suitable for clinical settings. In-vivo studies are recommended to gain a clearer understanding of differences in catheter insertion and tissue interaction.

C. Limitations of sample datasets

Developing path planning and navigation is important to improve the autonomy and safety of interventional surgeries by providing timely guidance through navigation images, real-time instrument position detection, and deformation modelling. The datasets A (rabbit model) and B (pig model) used in **Chapters 5** and **6** for guidewire tip localization do not reflect human endovascular complexity and tissue interference. Integrating imaging data from human vessels is crucial for successful training and improved catheterization, and comprehensive datasets covering a range of surgical scenarios and endovascular types are needed to assess the generalizability of the methods. The study also highlights challenges such as sample distribution mismatches and labour-intensive data annotation, limiting model generalisation, and calling for dataset expansion to improve model performance.

7.4 Future Works

These initiatives are part of development projects focused on minimally invasive cardiac surgery at the Centre for Medical Robotics and Minimally Invasive Surgical Devices (SIAT-CAS, China). Future work will involve developing efficient technical learning models for operators based on data-driven approaches, surgical tool endpoint location models, and semantic segmentation methods, in response to the above limitations, aimed at increasing the manipulative safety of robot-assisted interventional procedures.

Moreover, building upon the current framework for improving intuitive manipulation and perception in robot-assisted vascular interventional surgery, future work can explore two promising and complementary research directions: the integration of large-scale pre-trained models (large models) and the advancement toward autonomous surgical capabilities. **1) Integration of Large Multimodal Foundation Models for Surgical Perception and Decision Support.** Recent developments in vision-language large models (e.g., ChatGPT, CLIP, LLaVA) and surgical foundation models have demonstrated strong generalization across diverse modalities and medical contexts. Future work may explore the fusion of pre-trained large models with surgical robotic systems to enable richer semantic understanding of surgical environments. For instance, a large model could be used to interpret intraoperative imaging, annotate anatomical landmarks, and generate natural language guidance during complex procedures. Additionally, by integrating surgical manipulation data (force, motion, endoscopic imagery), fine-tuning or prompting these

models could yield context-aware reasoning and interactive assistance, especially in ambiguous or high-risk scenarios. This would empower the system to transition from low-level reactive feedback to high-level semantic support for surgical planning and real-time decision-making. **2) Towards Context-Aware and Semi-Autonomous Robotic Surgery.** Based on the multi-sensor manipulation recognition and perception modules developed in this study, future research can advance toward building semi-autonomous or context-adaptive robotic systems. Rather than passively executing surgeon commands, the robotic system can actively learn and infer the surgeon's intent, autonomously adjust manipulation strategies, or even pre-emptively suggest tool trajectories based on anatomical and manipulation context. This requires the integration of: 1) Dynamic scene understanding based on real-time visual and force feedback; 2) Reinforcement learning or imitation learning to capture expert manipulation strategies; 3) Hierarchical control architectures, where low-level trajectory generation is guided by high-level semantic or task-specific goals.

Combining large-scale model reasoning capabilities with multi-modal manipulation sensing and visual modelling can lay the foundation for next-generation intelligent surgical robotics. This will enable not only safer and more precise operations, but also pave the way toward more autonomous, personalized, and intelligent interventional procedures.

References

- [1] D. Cranston, "Surgery in oxford: A history," *Journal of the Nuffield Department of Surgical Sciences*, vol. 4(2), 2023.
- [2] C. Crenner, "Operative innovation and surgical conservatism in twentieth-century ulcer surgery," *Journal of the History of Medicine and Allied Sciences*, vol. Early Access, no. jrad065, 2023.
- [3] K. Suresh, "Evolution of cholecystectomy-gall bladder must be removed not because it contains stones but forms them," *Journal of Quality in Health Care & Economics*, vol. 6(1), pp. 1-9, 2023.
- [4] P. J. Herrod, A. T. Kwok, and D. N. Lobo, "Three centuries of appendectomy," *World journal of surgery*, vol. 47(4), pp. 928-936, 2023.
- [5] O. N. Reznik, D. O. Kuzmin, A. E. Skvortsov, *et al.*, "Biobanks are an essential tool for transplantation. history, current state, perspectives," *Vestnik Transplantologii i Iskusstvennyh Organov*, vol. 18(4), pp. 123-132, 2016.
- [6] A. Zarrinpar and R. W. Busuttil, "Timeline liver transplantation: past, present and future," *Nature reviews Gastroenterology & hepatology*, vol. 10(7), pp. 434-440, 2013.
- [7] T. K. Rajab, A. D. Vogel, V. S. Alexander, *et al.*, "The future of partial heart transplantation," *Journal of Heart and Lung Transplantation*, vol. 43(6), pp. 863-865, 2024.
- [8] C. D. Blackstock and B. M. Jackson, "Open surgical repair of abdominal aortic aneurysms maintains a pivotal role in the endovascular era," *Semin Intervent Radiol*, vol. 37(4), pp. 346-355, 2020.
- [9] I. S. company, "INC. united states securities and exchange commission," ed, 2024.
- [10] H. S. Himal, "Minimally invasive (laparoscopic) surgery-The future of general surgery," *Surgical Endoscopy and Other Interventional Techniques*, vol. 16(2), pp. 1647-1652, 2002.
- [11] G. Sermonesi, B. W. C. A. Tian, C. Vallicelli, *et al.*, "Cesena guidelines: WSES consensus statement on laparoscopic-first approach to general surgery emergencies and abdominal trauma," *World Journal of Emergency Surgery*, vol. 18(1), Art no. 57, 2023.
- [12] M. S. Menezes, I. K. Tanaka, V. F. Vicini, *et al.*, "Patologia benigna e beneficios da laparoscopia na cirurgia urológica: Benign pathology and benefits of laparoscopy in urological surgery," *Brazilian Journal of Development*, vol. 9(1), pp. 1746-1756, 2023.
- [13] R. Akhtar, R. Karim, and Z. Inayat, "Role of laparoscopic surgery in gynecology," *The Professional Medical Journal*, vol. 30(07), pp. 912-916, 2023.
- [14] P. Probst, "A review of the role of robotics in surgery: To davinci and beyond," *Missouri medicine*, vol. 120(5), p. 389, 2023.
- [15] A. Brassetti, A. Ragusa, F. Tedesco, *et al.*, "Robotic surgery in urology: history from PROBOT® to HUGOTM," *Sensors*, vol. 23(16), p. 7104, 2023.
- [16] K. Kim, S. Kwon, J. Kwon, *et al.*, "A review of robotic-assisted total hip arthroplasty," *Biomedical Engineering Letters*, vol. 13(4), pp. 523-535, 2023.
- [17] J. Wang, J. Miao, Y. Zhan, *et al.*, "Spine surgical robotics: current status and recent clinical applications," *Neurospine*, vol. 20(4), pp. 1256-1271, 2023.
- [18] R. M. Yeisson, S. Echevarria, C. Vidal-Valderrama, *et al.*, "Robotic surgery: a comprehensive review of the literature and current trends," *Cureus J Med Science*, vol. 15(7), Art no. e42370, 2023.
- [19] I. F. Zidane, Y. Khatta, S. Rezek, *et al.*, "Robotics in laparoscopic surgery-A review," *Robotica*, vol. 41(1), pp. 126-173, 2023.
- [20] R. Cao and H. Zhu, "Industry status and future development trend of Da vinci surgical robot," *Journal of Contemporary Medical Practice*, vol. 6(6), pp. 102-106, 2024.
- [21] Y. Zhao, Z. Mei, X. Luo, *et al.*, "Remote vascular interventional surgery robotics: a literature review," *Quantitative Imaging in Medicine and Surgery*, vol. 12(4), pp. 2552-2574, 2023..

- [22] G. Bassil, S. M. Markowitz, C. F. Liu, *et al.*, "Robotics for catheter ablation of cardiac arrhythmias: Current technologies and practical approaches," *Journal of Cardiovascular Electrophysiology*, vol. 31(3), pp. 739-752, 2020.
- [23] R. Beyar, L. Gruberg, D. Deleanu, *et al.*, "Remote-control percutaneous coronary interventions: Concept, validation, and first-in-humans pilot clinical trial," *Journal of American College Cardiology*, vol. 47(2), pp. 296-300, 2006.
- [24] P. Chen, Y. Wang, D. Tian, *et al.*, "The catheter and guidewire operating systems of vascular interventional surgical robots: a systematic review," *IEEE Transactions on Medical Robotics and Bionics*, vol. 5(2), pp. 180-195, 2023.
- [25] T. Datino, A. Arenal, M. Pelliza, *et al.*, "Comparison of the safety and feasibility of arrhythmia ablation using the Amigo robotic remote catheter system versus manual ablation," *Amer. J. Cardiol*, vol. 113(5), pp. 827-831, 2014.
- [26] E. Durand, R. Sabatier, P. C. Smits, *et al.*, "Evaluation of the R-One robotic system for percutaneous coronary intervention: the R-evolution study," *EuroIntervention*, vol. 18(16), pp. e1339-1347, 2023.
- [27] J. Bonatti, G. Vetrovec, C. Riga, *et al.*, "Robotic technology in cardiovascular medicine," *Nature Reviews Cardiology*, vol. 11(5), pp. 266-275, 2014.
- [28] L. Cruddas, G. Martin, and C. Riga, "Robotic endovascular surgery: current and future practice," *InSeminars in Vascular Surgery*, vol. 34(4), pp. 233-240, 2021.
- [29] W. K. Duan, Z. Li, O. M. Omisore, *et al.*, "Development of an Intuitive Interface with Haptic Enhancement for Robot-Assisted Endovascular Intervention," *IEEE Transactions on Haptics*, pp. 1-13, 2023.
- [30] W. K. Duan, T. Akinyemi, W. Du, *et al.*, "Technical and clinical progress on robot-assisted endovascular interventions: A review," *Micromachines*, vol. 14(1), pp. 1-25, 2023.
- [31] D. Anastasiou, Y. Jin, D. Stoyanov, *et al.*, "Keep your eye on the best: contrastive regression transformer for skill assessment in robotic surgery," *IEEE Robotics and Automation Letters*, vol. 8(3), pp. 1755-1762, 2023.
- [32] F. Hasan and J. Bonatti, "Robotically assisted percutaneous coronary intervention: Benefits to the patient and the cardiologist," *Expert Review of Cardiovascular Therapy*, vol. 13(11), pp. 1165-1168, 2015.
- [33] A. S. Lea, "Pyrite standards: medical uncertainty, ground truth, and AI model evaluation in historical perspective," *Journal of General Internal Medicine*, vol. Early Access, pp. 1-2, 2024.
- [34] P. S. Rao, "Introductory Chapter: Aortic Valve Disease-Recent Advances" *Aortic Valve Disease-Recent Advances*, 2024.
- [35] M. Teraa and C. E. Hazenberg, "Endovascular aortic interventions and aneurysm repair: recent Advances and future prospects," *MDPI*, pp. 1-172, 2024
- [36] T. Fang, J. Xiao, Y. Zhang, *et al.*, "Combined with interventional therapy, immunotherapy can create a new outlook for tumor treatment," *Quantitative Imaging in Medicine and Surgery*, vol. 11(6), pp. 2837-2860, 2021.
- [37] H. Li, L. Chen, G. Y. Zhu, *et al.*, "Interventional treatment for cholangiocarcinoma," *Front Oncology*, vol. 11, pp. 1-12, 2021.
- [38] K. A. Mashayekhi, S. A. Pyxaras, G. S. Werner, *et al.*, "Contemporary issues of percutaneous coronary intervention in heavily calcified chronic total occlusions: an expert review from the European CTO Club," *EuroIntervention*, vol. 19(2), pp. e113-122, 2023.
- [39] WHO, "Department of Data and Analytics (DNA) Division of Data, Analytics and Delivery for Impact (DDI) WHO, Geneva," *WHO methods and data sources for country-level causes of death 2000-2019*, 2020.
- [40] C. Sunnysvale. "A global technology leader in minimally invasive care and the pioneer of robotic-assisted surgery"
- [41] P. Kanagaratnam, M. Koa-Wing, D. T. Wallace, *et al.*, "Experience of robotic catheter ablation in humans using a novel remotely steerable catheter sheath," *Journal Intervention Cardiovascular Electrophysiology*, vol. 21(1), pp. 19-26, 2008.

- [42] H. Hughes and M. N. Patlas, "Robotics and Radiology," *Canadian Association of Radiologists Journal*, vol. 75(4), pp. 695-696, 2024.
- [43] T. Datino, A. Arenal, M. Pelliza, *et al.*, "Comparison of the safety and feasibility of arrhythmia ablation using the Amigo robotic remote catheter system versus manual ablation," *American Journal of Cardiology*, vol. 113(5), pp. 827-831, 2014.
- [44] A. A. Khokhar, A. Zelias, A. Zlahoda-Huzior, *et al.*, "Complication during robotic-PCI: Iatrogenic guiding catheter dissection," *Catheterization Cardiovascular Intervent.*, vol. 99(5), pp. 1526-1528, 2022.
- [45] B. Jones, C. Riga, C. Bicknell, *et al.*, "Robot-assisted carotid artery stenting: A safety and feasibility study," *CardioVascular and Interventional Radiology*, vol. 44(5), pp. 795-800, 2021.
- [46] M. Nicholls, "Mark Nicholls speaks to Professor Eric Durand and Professor Remi Sabatier about Europe's first remote robotic-assisted angioplasty procedure," *European Heart Journal*, vol. 42(32), pp. 3033-3035, 2021.
- [47] T. Deb, "Robotic Surgery Statistics 2025 By – Procedures, Safety, Complications," in "Statistics," Market.us Media, 2025
- [48] S. Zoting and A. Shivarkar, "Robotic Surgery Market Driving the Future of Minimally Invasive Procedures," in "Robotic Surgery Market Sizing," Towards Healthcare, 2025.
- [49] G. Lajkó, R. N. Elek, and T. Haidegger, "Endoscopic Image-Based Skill Assessment in Robot-Assisted Minimally Invasive Surgery," *Sensors*, vol. 21(16), 2021.
- [50] C. E. Reiley, T. Akinbiyi, D. Burschka, *et al.*, "Effects of visual force feedback on robot-assisted surgical task performance," *J Thorac Cardiovasc Sur*, vol. 135(1), pp. 196-202, 2008.
- [51] C. Freschi, V. Ferrari, F. Melfi, *et al.*, "Technical review of the da Vinci surgical telemanipulator," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 9(4), pp. 396-406, 2013.
- [52] J. Escaned, F. A. Jaffer, J. Mehilli, *et al.*, "The year in cardiovascular medicine 2021: interventional cardiology," *European heart journal*, vol. 43(5), pp. 377-386, 2022.
- [53] B. K. Spencer, "The Development of Interventional Cardiology," *J Am Coll Cardiol*, vol. 31(4), pp. 64B-88B, 1998.
- [54] R. F. Mould, "The early history of x-ray diagnosis with emphasis on the contributions of physics 1895-1915," *Physics in Medicine & Biology*, vol. 40(11), pp. 1741, 1995.
- [55] A. John and M. D. Meyer, "Werner Forssmann and catheterization of the heart, 1929," *The Annals of Thoracic Surgery*, vol. 49(3), pp. 497-499, 1990.
- [56] H. Hafiani, R. C. Saibari, N. Morsad, *et al.*, "Sven ivar seldinger (1921-1998): the founding father of interventional radiology," *Cureus J Med Science*, vol. 16(5), pp. e60397, 2024.
- [57] M. Barton, J. Grüntzig, M. Husmann, *et al.*, "Balloon angioplasty—the legacy of andreas grüntzig, MD (1939–1985)," *Front Cardiovasc Med*, vol. 1, pp.1-15, 2014.
- [58] M. M. Payne, "Charles theodore dotter: the father of intervention," *Texas Heart Institute Journal*, vol. 28(1), pp. 28, 2001.
- [59] H. V. Anderson, "Andreas R. gruentzig, MD (1939–1985)," *Cardiology*, vol. 147(1), pp. 107-112, 2022.
- [60] J. Rösch, F. S. Keller, and J. A. Kaufman, "The birth, early years, and future of interventional radiology," *Journal of Vascular and Interventional Radiology*, vol. 14(7), pp. 841-853, 2003.
- [61] J. C. Palmaz, "The balloon-expandable stent," *EuroIntervention*, vol. 2(4), pp. 416-421, 2007.
- [62] T. J. Fogarty, A. Biswas, and C. E. Newman, "Catheter thromboembolectomy," *Peripheral Endovascular Interventions*, pp. 289-308, 1996.
- [63] S. A. Curley, "Radiofrequency ablation of malignant liver tumors," *The oncologist*, vol. 6(1), pp. 14-23, 2001.
- [64] S. D. Chang, W. Main, D. P. Martin, *et al.*, "An analysis of the accuracy of the cyber knife: a robotic frameless stereotactic radiosurgical system," *Neurosurgery*, vol.52(1), pp.140-147, 2003.
- [65] D. B. Rukstalis, J. L. Goldknopf, E. M. Crowley, *et al.*, "Prostate cryoablation: a scientific rationale for future modifications," *Urology*, vol. 60(2), pp. 19-25, 2002.

- [66] W. Khan, S. Farah, and A. J. Domb, "Drug eluting stents: developments and current status," *Journal of controlled release*, vol. 161(2), pp. 703-712, 2012.
- [67] E. I. Cohen, D. Field, G. E. Lynskey, *et al.*, "Technology of irreversible electroporation and review of its clinical data on liver cancers," *Expert review of medical devices*, vol. 15(2), pp. 99-106, 2018.
- [68] C. A. Bisbee, J. Zhang, J. Owens, *et al.*, "Cryoablation for the treatment of kidney cancer: comparison with other treatment modalities and review of current treatment," *Cureus*, vol. 14(11), pp. 1-9, 2022.
- [69] S. Yilmaz, M. Özdoğan, M. Cevener, *et al.*, "Use of cryoablation beyond the prostate," *Insights into imaging*, vol. 7, pp. 223-232, 2016.
- [70] O. Al-Bataineh, J. Jenne, and P. Huber, "Clinical and future applications of high intensity focused ultrasound in cancer," *Cancer treatment reviews*, vol. 38(5), pp. 346-353, 2012.
- [71] E. L. Verhoeven, A. Katsargyris, K. Oikonomou, *et al.*, "Fenestrated endovascular aortic aneurysm repair as a first line treatment option to treat short necked, juxtarenal, and suprarenal aneurysms," *Eur J Vasc Endovasc*, vol. 51(6), pp. 775-781, 2016.
- [72] H. Bastani, "Cryoablation of cardiac arrhythmias," *Karolinska Institutet (Sweden)*, 2011.
- [73] D. Sürder, G. Pedrazzini, O. Gaemperli, *et al.*, "Predictors for efficacy of percutaneous mitral valve repair using the MitraClip system: the results of the MitraSwiss registry," *Heart*, vol. 99(14), pp. 1034-1040, 2013.
- [74] J. G. Webb and D. A. Wood, "Current status of transcatheter aortic valve replacement," *J Am Coll Cardiol*, vol. 60(6), pp. 483-492, 2012.
- [75] J. Schofer, K. Bijuklic, C. Tiburtius, *et al.*, "First-in-human transcatheter tricuspid valve repair in a patient with severely regurgitant tricuspid valve," *J Am Coll Cardiol*, vol. 65(12), pp. 1190-1195, 2015.
- [76] C. Castaño, S. Remollo, M. R. García, *et al.*, "Mechanical thrombectomy with 'ADAPT' technique by transcervical access in acute ischemic stroke," *The Neuroradiology Journal*, vol. 28(6), pp. 617-622, 2015.
- [77] R. M. Simmons, K. V. Ballman, C. Cox, *et al.*, "A phase II trial exploring the success of cryoablation therapy in the treatment of invasive breast carcinoma: results from ACOSOG (Alliance) Z1072," *Annals of surgical oncology*, vol. 23, pp. 2438-2445, 2016.
- [78] T. D. Potter, M. Klaver, A. Babkin, *et al.*, "Ultra-low temperature cryoablation for atrial fibrillation: primary outcomes for efficacy and safety: the cryocure-2 study," *Clinical Electrophysiology*, vol. 8(8), pp. 1034-1039, 2022.
- [79] B. Surtees, S. Young, Y. Hu, *et al.*, "Validation of a low-cost, carbon dioxide-based cryoablation system for percutaneous tumor ablation," *PLoS One*, vol. 14(7), p. e0207107, 2019.
- [80] Y. Nagata and T. Kimura, "Stereotactic body radiotherapy (SBRT) for stage I lung cancer," *Japanese Journal of Clinical Oncology*, vol. 48(5), pp. 405-409, 2018.
- [81] J. Abdo, D. I. Cornell, S. K. Mittal, *et al.*, "Immunotherapy plus cryotherapy: potential augmented abscopal effect for advanced cancers," *Frontiers in oncology*, vol. 8, pp. 85, 2018.
- [82] T. J. McGinnis, "Significant FDA approvals in 2000," *American Family Physician*, vol. 63(10), pp. 2061-2064, 2001.
- [83] J. Shen, C. J. P. Zhang, B. Jiang, *et al.*, "Artificial intelligence versus clinicians in disease diagnosis: systematic review," *JMIR medical informatics*, vol. 7(3), pp. e10010, 2019.
- [84] I. Rykowska, I. Nowak, and R. Nowak, "Drug-eluting stents and balloons—materials, structure designs, and coating techniques: a review," *Molecules*, vol. 25(20), pp. 4624, 2020.
- [85] R. Minici, R. Serra, M. Giurdanella, *et al.*, "Efficacy and safety of distal radial access for transcatheter arterial chemoembolization (TACE) of the liver," *Journal of Personalized Medicine*, vol. 13(4), p. 640, 2023.
- [86] V. Ascenti, A. M. Ierardi, M. Alfa-Wali, *et al.*, "Damage control interventional radiology: the bridge between non-operative management and damage control surgery," *CVIR endovascular*, vol. 7(1), p. 71, 2024.

- [87] J. Zhou, X. Ma, Z. Xu, *et al.*, "Overview of medical robot technology development," presented at the 37th Chinese Control Conference (CCC), Wuhan, China, 07 October 2018, 2018.
- [88] J. Rassweiler, M. Fiedler, N. Charalampogiannis, *et al.*, "Robot-assisted flexible ureteroscopy: an update," *Urolithiasis*, vol. 46, pp. 69-77, 2018.
- [89] C. V. Riga, C. D. Bicknell, A. Rolls, *et al.*, "Robot-assisted fenestrated endovascular aneurysm repair (FEVAR) using the magellan system," *Journal of Vascular and Interventional Radiology*, vol. 24(2), pp. 191-196, 2013.
- [90] J. Bismuth, E. Kashef, N. Cheshire, *et al.*, "Feasibility and safety of remote endovascular catheter navigation in a porcine model," *Journal of Endovascular Therapy*, vol. 18(2), pp. 243-249, 2011.
- [91] G. A. Antoniou, C. V. Riga, E. K. Mayer, *et al.*, "Clinical applications of robotic technology in vascular and endovascular surgery," *Journal of Vascular Surgery*, vol. 53(2), pp. 493-549, 2011.
- [92] A. Rolls and C. Riga, "Endovascular robotics the current state and the future direction," *Annals of the Royal College Surgeons of England*, vol. 100, pp. 14, 2018.
- [93] C. V. Riga, C. D. Bicknell, A. Rolls, *et al.*, "Robot-assisted fenestrated endovascular aneurysm repair (FEVAR) using the magellan system," *Journal of Vascular and Interventional Radiology*, vol. 24, pp. 191-196, 2013.
- [94] Z. A. Shaikh, M. F. Eilenberg, and T. J. Cohen, "The Amigo™ remote catheter system: from concept to bedside," *The Journal of innovations in cardiac rhythm management*, vol. 8(8), pp. 2795, 2017.
- [95] K. C. Sajja, A. Sweid, F. Al Saiegh, *et al.*, "Endovascular robotic: feasibility and proof of principle for diagnostic cerebral angiography and carotid artery stenting," *Journal of neurointerventional surgery*, vol. 12(4), pp. 345-349, 2020.
- [96] C.B. Beaman, N. Kaneko, P. M. Meyers, *et al.*, "A review of robotic interventional neuroradiology," *American Journal of Neuroradiology*, vol. 42(5), pp. 808-814, 2021.
- [97] D. R. Mangels, J. Giri, J. Hirshfeld, *et al.*, "Robotic-assisted percutaneous coronary intervention," *Catheterization and Cardiovascular Interventions*, vol. 90(6), pp. 948-955, 2017.
- [98] K. C. Sajja, A. Sweid, F. Al Saiegh, *et al.*, "Endovascular robotic: feasibility and proof of principle for diagnostic cerebral angiography and carotid artery stenting," *Journal of Neurointerventional Surgery*, vol. 12(4), pp. 345-349, 2020.
- [99] S. S. Panesar and G. W. Britz, "Endovascular robotics: the future of cerebrovascular surgery," *World Neurosurg*, vol. 129, pp. 327-332, 2019.
- [100] E. Mahmud, A. Pourdjabbar, L. Ang, *et al.*, "Robotic technology in interventional cardiology: current status and future perspectives," *Catheterization Cardiovascular Interventions*, vol. 90(6), pp. 956-962, 2017.
- [101] A. Pourdjabbar, L. Ang, R. R. Reeves, *et al.*, "The development of robotic technology in Cardiac and vascular interventions," *Rambam Maimonides Medical Journal*, vol. 8(3), pp. e30, 2017.
- [102] M. Riordan and A. Micklus, "Medtech dealmaking and financing trends in 2019.," *Nature*, 2019.
- [103] E. Durand, R. Sabatier, P. C. Smits, *et al.*, "Evaluation of the R-One robotic system for percutaneous coronary intervention: the R-evolution study," *EuroIntervention*, vol. 18(16), pp. e1339, 2023.
- [104] G. Koulaouzidis, D. Charisopoulou, P. Bomba, *et al.*, "Robotic-assisted solutions for invasive cardiology, cardiac surgery and routine on-ward tasks: a narrative review," *Journal of Cardiovascular Development and Disease*, vol. 10(9), pp. 399, 2023.
- [105] I. Altun and N. Nezami, "Role of robotics in image-guided trans-arterial interventions," *Techniques in Vascular and Interventional Radiology*, pp. 101005, 2024.
- [106] H. Y. Yu, H. B. Wang, W. Zhang, *et al.*, "Master-slave system research of a vascular interventional surgical robot," the IEEE International Conference on Real-time Computing and Robotics (RCAR), Kandima, Maldives, 2018.
- [107] H. Wang *et al.*, "Design of propulsion mechanism for vascular interventional operation robot," presented at the International Conference on Advanced Mechatronic Systems (ICAMechS), Zhengzhou, China, 2018.

- [108] H. J. Cha, B. J. Yi, and J. Y. Won, "An assembly-type master-slave catheter and guidewire driving system for vascular intervention," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 231(1), pp. 69–79, 2017.
- [109] J. Woo, H. Song, H. Cha, *et al.*, "Advantage of steerable catheter and haptic feedback for a 5-DOF vascular intervention robot system," *Applied Sciences-Based*, vol. 9(20), 2019.
- [110] H. Song, J. Woo, J. Won, *et al.*, "In vivo usability test of vascular intervention robotic system controlled by two types of master devices," *Applied. Sciences-Based*, vol. 11(12), 2021.
- [111] C. C. Smitson, L. Ang, A. Pourdjabbar, *et al.*, "Safety and feasibility of a novel, second-generation robotic-assisted system for percutaneous coronary intervention: first-in-human report," *Journal Invasive Cardiology*, vol. 30, pp. 152–156, 2018.
- [112] P. Legeza, G. W. Britz, T. Loh, *et al.*, "Current utilization and future directions of robotic-assisted endovascular surgery," *Expert Review of Medical Devices*, vol. 17, pp. 919–927, 2020.
- [113] R. D. Madder, S. VanOosterhout, J. Parker, *et al.*, "Robotic telestenting performance in transcontinental and regional pre-clinical models," *Catheterization Cardiovascular Interventions*, vol. 97, pp. E327–E332, 2021.
- [114] J. F. Granada, J. A. Delgado, M. P. Uribe, *et al.*, "First-in-human evaluation of a novel robotic-assisted coronary angioplasty system," *Jacc-Cardiovasc Inte*, vol. 4(4), pp. 460-465, Apr 2011.
- [115] G. Weisz, D. C. Metzger, R. P. Caputo, *et al.*, "Safety and feasibility of robotic percutaneous coronary intervention precise (percutaneous robotically-enhanced coronary intervention) study," *J Am Coll Cardiol*, vol. 61(15), pp. 1596-1600, 2013.
- [116] E. Mahmud, J. Naghi, L. Ang, *et al.*, "Demonstration of the safety and feasibility of robotically assisted percutaneous coronary intervention in complex coronary lesions results of the CORA-PCI study (complex robotically assisted percutaneous coronary intervention)," *Jacc-Cardiovasc Inte*, vol. 10(13), pp. 1320-1327, 2017.
- [117] E. Mahmud, F. Schmid, P. Kalmar, *et al.*, "Feasibility and safety of robotic peripheral vascular interventions results of the RAPID trial," *Jacc-Cardiovasc Inte*, vol. 9(19), pp. 2058-2064, 2016.
- [118] J. H. Weinberg, A. Sweid, K. Sajja, *et al.*, "Comparison of robotic-assisted carotid stenting and manual carotid stenting through the transradial approach," *J Neurosurg*, vol. 135(1), pp. 21-28, 2021.
- [119] A. H. Perera, C. V. Riga, L. Monzon, *et al.*, "Response to commentary on "robotic arch catheter placement reduces cerebral embolisation during thoracic endovascular aortic repair (TEVAR)," *Eur J Vasc Endovasc*, vol. 54(1), pp. 126-127, 2017.
- [120] A. D. Russo, G. Fassini, S. Conti, *et al.*, "Analysis of catheter contact force during atrial fibrillation ablation using the robotic navigation system: results from a randomized study," *J Interv Card Electr*, vol. 46(2), pp. 97-103, 2016.
- [121] M. López-Gil, R. Salgado, J. L. Merino, *et al.*, "Cavo-tricuspid isthmus radiofrequency ablation using a novel remote navigation catheter system in patients with typical atrial flutter," *Europace*, vol. 16(4), pp. 558-562, 2014.
- [122] K. S. Hoffmayer, F. Krainski, S. Shah, *et al.*, "Randomized controlled trial of AmigoA® robotically controlled versus manually controlled ablation of the cavo-tricuspid isthmus using a contact force ablation catheter," *J Interv Card Electr*, vol. 51(2), pp. 125-132, 2018.
- [123] E. Maor, M. F. Eleid, R. Gulati, *et al.*, "Current and future use of robotic devices to perform percutaneous coronary interventions: a review," *J Am Heart Assoc*, vol. 6(7), 2017.
- [124] T. M. Patel, S. C. Shah, and S. B. Panchoy, "Long distance tele-robotic-assisted percutaneous coronary intervention: a report of first-in-human experience," *Eclinicalmedicine*, vol. 14, pp. 53-58, 2019.
- [125] J. D. Häner, L. Räber, C. Moro, *et al.*, "Robotic-assisted percutaneous coronary intervention: experience in Switzerland," *Front Cardiovasc Med*, vol. 10, 2023.
- [126] J. Leung, J. French, J. Xu, *et al.*, "Robotic assisted percutaneous coronary intervention: initial australian experience," *Heart Lung Circ*, vol. 33(4), pp. 493-499, 2024.

- [127] F. Hasan and J. Bonatti, "Robotically assisted percutaneous coronary intervention: Benefits to the patient and the cardiologist," *Expert Review of Cardiovascular Therapy*, vol. 13, pp. 1165–1168, 2015.
- [128] Y. Thakur, J. S. Bax, D. W. Holdsworth, *et al.*, "Design and performance evaluation of a remote catheter navigation system," *Ieee T Bio-Med Eng*, vol. 56, pp. 1901–1908, 2009.
- [129] C. J. Payne, H. Rafii-Tari, and G. Yang, "A force feedback system for endovascular catheterization," presented at the IEEE/RSJ International Conference on Intelligent Robots and Systems, Algarve, Portugal, 2012.
- [130] H. Rafii-Tari, C. J. Payne, J. Liu, *et al.*, "Towards automated surgical skill evaluation of endovascular catheterization tasks based on force and motion signatures," the IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 2015.
- [131] H. Yu, H. Wang, J. Chang, *et al.*, "A novel vascular intervention surgical robot based on force feedback and flexible clamping," *Appl Sci-Basel*, vol. 11, pp. 611, 2021.
- [132] O. M. Omisore, T. O. Akinyemi, W. Du, *et al.*, "Weighting-based deep ensemble learning for recognition of interventionalists' hand motions during robot-assisted intravascular catheterization," *Ieee T Hum-Mach Syst*, vol. 53, pp. 1–13, 2022.
- [133] A. K. Ghamraoui and J. J. Ricotta, "Current and future perspectives in robotic endovascular surgery," *Current Surgery Reports*, vol. 6, pp. 21, 2018.
- [134] E. M. Khan, W. Frumkin, G. A. Ng, *et al.*, "First experience with a novel robotic remote catheter system: Amigo™ mapping trial," *J Interv Card Electr*, vol. 37, pp. 121–129, 2013.
- [135] S. Guo, Y. Song, X. Yin, *et al.*, "A novel robot-assisted endovascular catheterization system with haptic force feedback," *Ieee T Robot Autom*, vol. 35, pp. 685–696, 2019.
- [136] X. Bao, S. Guo, N. Xiao, *et al.*, "Compensatory force measurement and multimodal force feedback for remote-controlled vascular interventional robot," *Biomedical Microdevices*, vol. 20(74), 2018.
- [137] H. Wang, J. Chang, H. Yu, *et al.*, "Research on a novel vascular interventional surgery robot and control method based on precise delivery," *IEEE Access*, vol. 9, pp. 26568–26582, 2021.
- [138] X. Yang, H. Wang, L. Sun, *et al.*, "Operation and force analysis of the guidewire in a minimally invasive vascular interventional surgery robot system," *Chinese Journal of Mechanical Engineering*, vol. 28, pp. 249–257, 2015.
- [139] T. Wang, D. Zhang, and D. Liu, "Remote-controlled vascular interventional surgery robot," *International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 6, pp. 194–201, 2010.
- [140] W. S. Lu, W. Y. Xu, F. Pan, *et al.*, "Clinical application of a vascular interventional robot in cerebral angiography," *International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 12, pp. 132–136, 2016.
- [141] K. Wang, Q. Lu, B. Chen, *et al.*, "Endovascular intervention robot with multi-manipulators for surgical procedures: Dexterity, adaptability, and practicability," *Robotics and Computer-Integrated Manufacturing*, vol. 56, pp. 75–84, 2019.
- [142] K. Wang, B. Chen, and X. Xu, "Design and control method of surgical robot for vascular intervention operation," the IEEE International Conference on Robotics and Biomimetics (ROBIO), Qingdao, China, 2016.
- [143] J. Woo, H. Song, H. Cha, *et al.*, "Advantage of steerable catheter and haptic feedback for a 5-DOF vascular intervention robot system," *Appl Sci-Basel*, vol. 9(4305), 2019.
- [144] J. Choi, S. Park, Y. H. Kim, *et al.*, "A vascular intervention assist device using bi-motional roller cartridge structure and clinical evaluation," *Biosensors*, vol. 11, pp. 329, 2021.
- [145] M. Tanimoto, F. Arai, T. Fukuda, *et al.*, "Telesurgery system for intravascular neurosurgery," the International Conference on Medical Image Computing and Computer-Assisted Intervention, Pittsburgh, PA, USA, 2000.
- [146] O. M. Omisore, S. P. Han, L. X. Ren, *et al.*, "Towards characterization and adaptive compensation of backlash in a novel robotic catheter system for cardiovascular interventions," *Ieee T Biomed Circ S*, vol. 12, pp. 824–838, 2018.

- [147] W. J. Du, O. M. Omisore, W. Duan, *et al.*, "Exploration of interventionists' technical manipulation skills for robot-assisted intravascular PCI catheterization," *IEEE Access*, vol. 8, pp. 53750-53765, 2020.
- [148] Z. Q. Feng, G. B. Bian, X. L. Xie, *et al.*, "Design and evaluation of a bio-inspired robotic hand for percutaneous coronary intervention," the IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 2015.
- [149] G. B. Bian, X. L. Xie, Z. Q. Feng, *et al.*, "An enhanced dual-finger robotic Hand for Catheter manipulating in vascular intervention: A preliminary study " the IEEE International Conference on Information and Automation (ICIA), Yinchuan, China, 2013.
- [150] F. Langsch, S. Virga, J. Esteban, *et al.*, "Robotic ultrasound for catheter navigation in endovascular procedures," the IEEE/RSJ International Conference on Intelligent Robots And Systems (IROS), Macau, China, 2019.
- [151] X. Li, J. Li, J. Jing, *et al.*, "Integrated IVUS-OCT imaging for atherosclerotic plaque characterization," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 20, pp. 196–203, 2014.
- [152] N. K. Sankaran, P. Chembrammal, A. Siddiqui, *et al.*, "Design and development of surgeon augmented endovascular robotic system," *Ieee T Bio-Med Eng*, vol. 65, pp. 2483–2493, 2018.
- [153] J. Zhou, Z. Mei, J. Miao, *et al.*, "A remote-controlled robotic system with safety protection strategy based on force-sensing and bending feedback for transcatheter arterial chemoembolization," *Micromachines*, vol. 11(805), 2020.
- [154] X. H. Zhou, G. B. Bian, X. L. Xie, *et al.*, "An interventionalist-behavior-based data fusion framework for guidewire tracking in percutaneous coronary intervention," *Ieee T Syst Man Cy-S*, vol. 50, pp. 4836–4849, 2020.
- [155] H. Raffii-Taril, C. J. Payne, C. Riga, *et al.*, "Assessment of navigation cues with proximal force sensing during endovascular catheterization," *Med Image Comput Comput Assist Interv*, vol. 15, pp. 560-567, 2012.
- [156] F. Verzini, P. D. Rango, G. Parlani, *et al.*, "Carotid artery stenting: technical issues and role of operators' experience," *Perspectives in Vascular Surgery and Endovascular Therapy*, vol. 20, pp. 247-257, 2008.
- [157] G. Srimathveeravalli, T. Kesavadas, and X. Li, "Design and fabrication of a robotic mechanism for remote steering and positioning of interventional devices," *International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 6, pp. 160–170, 2010.
- [158] J. E. G. Villarruel and B. T. Corona, "Proposal for a remote surgery system based on wireless communications, electromyography and robotics," the 5th Electronics, Robotics and Automotive Mechanics Conference, Cuernavaca, Mexico, 2008.
- [159] C. Y. Wang and Y. Y. Huang, "System design of electrophysiological monitoring in pelvic surgery," the 9th International Conference on Intelligent Computing and Signal Processing (ICSP), Xian, China, 2024.
- [160] X. D. Li, R. W. Wen, Z. Shen, *et al.*, "A wearable detector for simultaneous finger joint motion measurement," *Ieee T Biomed Circ S*, vol. 12(3), pp. 644-654, 2018.
- [161] A. Sánchez, O. Rodríguez, R. Sánchez, *et al.*, "Laparoscopic surgery skills evaluation: analysis based on accelerometers," *Jsls-J Soc Laparoend*, vol. 18(4), 2014.
- [162] A. E. Rolls, C. V. Riga, C. D. Bicknell, *et al.*, "A pilot study of video-motion analysis in endovascular surgery: development of real-time discriminatory skill metrics," *Eur J Vasc Endovasc*, vol. 45(5), pp. 509-515, 2013.
- [163] H. Raffii-Tari, C. J. Payne, C. Bicknell, *et al.*, "Objective assessment of endovascular navigation skills with force sensing," *Ann Biomed Eng*, vol. 45(5), pp. 1315-1327, 2017.
- [164] C. Tercero, H. Kodama, C. Shi, *et al.*, "Technical skills measurement based on a cyber-physical system for endovascular surgery simulation," *International Journal of Meddical Robotics and Computer Assisted Surgery*, vol. 9(3), pp. e25-33, 2013.
- [165] R. C. King, L. Atallah, B. P. L. Lo, *et al.*, "Development of a wireless sensor glove for surgical skills assessment," *Ieee T Inf Technol B*, vol. 13(5), pp. 673-679, 2009.

- [166] F. J. Pérez-Duarte, M. Lucas-Hernández, A. Matos-Azevedo, *et al.*, "Objective analysis of surgeons' ergonomics during laparoendoscopic single-site surgery through the use of surface electromyography and a motion capture data glove," *Surg Endosc*, vol. 28(4), pp. 1314-1320, 2014.
- [167] L. Atallah, B. Lo, R. King, *et al.*, "Sensor positioning for activity recognition using wearable accelerometers," *Ieee T Biomed Circ S*, vol. 5(4), pp. 320-329, 2011.
- [168] K. V. Volkinburg and G. Washington, "Development of a wearable controller for gesture-recognition-based applications using polyvinylidene fluoride," *Ieee T Biomed Circ S*, vol. 11(4), pp. 900-909, 2017.
- [169] X. H. Zhou, G. B. Bian, X. L. Xie, *et al.*, "Qualitative and quantitative assessment of technical skills in percutaneous coronary intervention: porcine studies," *Ieee T Bio-Med Eng*, vol. 67(2), pp. 353-364, Feb 2020.
- [170] X. H. Zhou, G. B. Bian, X. L. Xie, *et al.*, "Analysis of interventionalists' natural behaviors for recognizing motion patterns of endovascular tools during percutaneous coronary Interventions," *Ieee T Biomed Circ S*, vol. 13(2), pp. 330-342, 2019.
- [171] S. Wang, Z. Liu, W. T. Yang, *et al.*, "Learning-based multimodal information fusion and behavior recognition of vascular interventionists' operating skills," *Ieee J Biomed Health*, vol. 27(9), pp. 4536-4547, 2023.
- [172] J. Guo, S. Guo, and Y. Yu, "Design and characteristics evaluation of a novel teleoperated robotic catheterization system with force feedback for vascular interventional surgery," *Biomedical Microdevices* vol. 18(5), pp. 76, 2016.
- [173] W. Wei, D. Yang, L. Li, *et al.*, "An intravascular catheter bending recognition method for interventional surgical robots," *Machines*, vol. 10(42), 2022.
- [174] C. He, S. Wang, and S. Zuo, "A linear stepping endovascular intervention robot with variable stiffness and force sensing," *Int J Comput Ass Rad*, vol. 13, pp. 671-682, 2018.
- [175] H. Rafii-Tari, C. J. Payne, and G. Yang, "Current and emerging robot-assisted endovascular catheterization technologies: a review," *Ann. Biomed. Eng*, vol. 42, pp. 697-715, 2014.
- [176] Y. Yan, S. Guo, Y. Guo, *et al.*, "Machine learning-based surgical state perception and collaborative control for a vascular interventional robot," *IEEE Sensors Journal*, vol. 22, pp. 7106-7118, 2022.
- [177] X. Bao, S. Guo, Y. Guo, *et al.*, "Multilevel operation strategy of a vascular interventional robot system for surgical safety in teleoperation," *IEEE Transaction on Robot*, vol. 38, pp. 2238-2250, 2022.
- [178] E. P. W.-V. D. PuttenP, R. H. M. Goossens, J. J. Jakimowicz, *et al.*, "Haptics in minimally invasive surgery—A review," *Minimally Invasive Therapy Allied Technologies*, vol. 17(1), pp. 3-16, 2008.
- [179] A. Hooshiar, S. Najarian, and J. Dargahi, "Haptic telerobotic cardiovascular intervention: a review of approaches, methods, and future perspectives," *IEEE Reviews in Biomedical Engineering*, vol. 13, pp. 32-50, 2020.
- [180] Z. Hu, J. Zhang, L. Xie, *et al.*, "A generalized predictive control for remote cardiovascular surgical systems," *ISA Trans*, vol. 104, pp. 336-344, 2020.
- [181] D. Bouget, M. Allan, D. Stoyanov, *et al.*, "Vision-based and marker-less surgical tool detection and tracking: a review of the literature," *Med Image Anal*, vol. 35, pp. 633-654, 2017.
- [182] T. Mahmood, S. W. Cho, and K. R. Park, "DSRD-Net: dual-stream residual dense network for semantic segmentation of instruments in robot-assisted surgery," *Expert Syst Appl*, vol. 202, 2022.
- [183] A. A. Shvets, A. Rakhlin, A. A. Kalinin, *et al.*, "Automatic instrument segmentation in robot-assisted surgery using deep learning," presented at the 17th IEEE International Conference on Machine Learning and Applications (IEEE ICMLA), Orlando, FL, 2018.
- [184] N. Kalavakonda, B. Hannaford, Z. Qazi, *et al.*, "Autonomous neurosurgical instrument segmentation using end-to-end learning," the 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, 2019.

- [185] L. Sestini, B. Rosa, E. D. Momi, *et al.*, "FUN-SIS: A fully unsupervised approach for surgical instrument segmentation," *Med Image Anal*, vol. 85, 2023.
- [186] W. T. Shen, Y. Wang, M. Liu, *et al.*, "Branch aggregation attention network for robotic surgical instrument segmentation," *Ieee T Med Imaging*, vol. 42(11), pp. 3408-3419, 2023.
- [187] L. Yang, Y. G. Gu, G. B. Bian, *et al.*, "TMF-Net: a transformer-based multiscale fusion network for surgical instrument segmentation from endoscopic images," *Ieee T Instrum Meas*, vol. 72, 2023.
- [188] L. Yang, Y. G. Gu, G. B. Bian, *et al.*, "MSDE-Net: a multi-scale dual-encoding network for surgical instrument segmentation," *Ieee J Biomed Health*, vol. 28(7), pp. 4072-4083, 2024.
- [189] L. Yang, C. X. Zhai, H. Y. Wang, *et al.*, "A dual-branch fusion network for surgical instrument segmentation," *Ieee Transactions on Medical Robotics and Bionics*, vol. 6(4), pp. 1542-1554, 2024.
- [190] M. D. Samad, "Effect of haptic and visual information on asymmetric hand motion in a robot-assisted task," the 2013 International Conference on Communication and Computer Vision (Icccv), Coimbatore, INDIA, 2013.
- [191] J. Huang, X. Li, T. Kesavadas, *et al.*, "Feature extraction of video data for automatic visual tool tracking in robot assisted surgery," the 4th International Conference on Robotics, Control and Automation (ICRCA), Guangzhou, PEOPLES R CHINA, 2019.
- [192] Z. Y. Chen, L. Cruciani, K. Fan, *et al.*, "Towards safer robot-assisted surgery: A markerless augmented reality framework," *Neural Networks*, vol. 178, 2024.
- [193] G. F. Zhang, H. C. Wong, Z. J. Hu, *et al.*, "Jigsaw training-based background reverse attention transformer network for guidewire segmentation," *Int J Comput Ass Rad*, vol. 18(4), pp. 653-661, 2023.
- [194] B. C. Zhang, M. Bui, C. Wang, *et al.*, "Real-time guidewire tracking and segmentation in intraoperative X-ray," *Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 12034, 2022.
- [195] O. M. Omisore, T. O. Akinyemi, W. K. Duan, *et al.*, "Multi-lateral branched network for tool segmentation during robot-assisted endovascular interventions," *Ieee Transactions on Medical Robotics and Bionics*, vol. 6(2), pp. 433-447, 2024.
- [196] O. M. Omisore, G. Yi, Y. Zheng, *et al.*, "Endovascular tool segmentation with multi-lateral branched network during robot-assisted catheterization," the 45th Annual International Conference of the Ieee Engineering in Medicine & Biology Society, Sydney, AUSTRALIA, 2023.
- [197] R. Q. Li, X. L. Xie, X. H. Zhou, *et al.*, "Real-time multi-guidewire endpoint localization in fluoroscopy images," *Ieee T Med Imaging*, vol. 40(8), pp. 2002-2014, 2021.
- [198] R. Q. Li, X. L. Xie, X. H. Zhou, *et al.*, "A unified framework for multi-guidewire endpoint localization in fluoroscopy images," *Ieee T Bio-Med Eng*, vol. 69(4), pp. 1406-1416, 2022.
- [199] A. Kuntz *et al.*, "Autonomous medical needle steering in vivo," *Science Robotics*, vol. 8(82), p. eadf7614, 2023.
- [200] C. Y. He, A. Ebrahimi, M. Roizenblatt, *et al.*, "User behavior evaluation in robot-assisted retinal surgery," the 27th IEEE International Symposium on Robot and Human Interactive Communication (IEEE RO-MAN), Nanjing, PEOPLES R CHINA, AUG 27-31, 2018.
- [201] L. Cercenelli, B. Bortolani, and E. Marcelli, "CathROB: a highly compact and versatile remote catheter navigation system," *Appl Bionics Biomech*, vol. 2017, 2017.
- [202] X. Bao, S. Guo, N. Xiao, *et al.*, "Operation evaluation in-human of a novel remotecontrolled vascular interventional robot," *Biomed. Microdevices*, vol. 20, pp. 1-13, 2018.
- [203] P. M. Loschak, L. J. Brattain, and R. D. Howe, "Algorithms for automatically pointing ultrasound imaging catheters," *Ieee T Robot*, vol. 33(1), pp. 81-91, 2017.
- [204] T. Zhou, O. M. Omisore, W. J. Du, *et al.*, "A preliminary study on surface electromyography signal analysis for motion characterization during catheterization," the 12th International Conference on Intelligent Robotics and Applications (ICIRA), Shenyang, China, 2019.

- [205] K. C. Demir, H. Schieber, T. Weise, *et al.*, "Deep learning in surgical workflow analysis: a review of phase and step recognition," *IEEE Journal of Biomedical and Health Informatics*, vol. 27(11), pp. 5405 - 5417, 2023.
- [206] N. X. Anh, R. M. Nataraja, and S. Chauhan, "Towards near real-time assessment of surgical skills: A comparison of feature extraction techniques," *Comput Meth Prog Bio*, vol. 187, 2020.
- [207] K. Lam, J. Chen, Z. Wang, *et al.*, "Machine learning for technical skill assessment in surgery: a systematic review," *Npj Digit Med*, vol. 5(1), 2022.
- [208] X. A. Nguyen, D. Ljuhar, M. Pacilli, *et al.*, "Surgical skill levels: classification and analysis using deep neural network model and motion signals," *Comput Meth Prog Bio*, vol. 177, pp. 1-8, 2019.
- [209] A. Soleymani, A. A. S. Asl, M. Yeganejou, *et al.*, "Surgical skill evaluation from robot-assisted surgery recordings," presented at the International Symposium on Medical Robotics (ISMR), Atlanta, GA, 2021.
- [210] O. M. Omisore, T. O Akinyemi, W.J. Du, *et al.*, "Weighting-based deep ensemble learning for recognition of interventionalists' hand motions during robot-assisted intravascular catheterization," *Ieee T Hum-Mach Syst*, vol. 53(1), pp. 215-227, 2023.
- [211] O. M. Omisore, W.K. Duan, W.J. Du, *et al.*, "Automatic tool segmentation and tracking during robotic intravascular catheterization for cardiac interventions," *Quantitative Imaging in Medicine and Surgery*, vol. 11(6), pp. 2688-2710, 2021.
- [212] O. M. Omisore, W. K. Duan, W. J. Du, *et al.*, "On task-specific autonomy in robotic interventions: A multimodal learning-based approach for multi-level skill assessment during cyborg catheterization," *TechRxiv*, 2021.
- [213] N. Li, H. P. Kong, Y. F. Ma, *et al.*, "Human performance modeling for manufacturing based on an improved KNN algorithm," *Int J Adv Manuf Tech*, vol. 84(1-4), pp. 473-483, 2016.
- [214] H. Su, W. J. Shang, G. Li, *et al.*, "An MRI-guided telesurgery system using a fabry-perot interferometry force sensor and a pneumatic haptic device," *Ann Biomed Eng*, vol. 45(8), pp. 1917-1928, 2017.
- [215] A. Takács, L. Kovács, I. J. Rudas, *et al.*, "Models for force control in tele-surgical robot systems," *Acta Polytech Hung*, vol. 12(8), pp. 95-114, 2015.
- [216] J. Inga, M. Ruess, J. H. Robens, *et al.*, "Human-machine symbiosis: s multivariate perspective for physically coupled human-machine systems," *Int J Hum-Comput St*, vol. 170, 2023.
- [217] A. Smirnov and A. Ponomarev, "Human-machine collective intelligence environment for decision support: conceptual and technological design," the 27th IEEE Conference of Open-Innovations-Association (FRUCT), Electr Network, 2020.
- [218] O. Kostyukova, F. P. Vista, and K. T. Chong, "Design of feedforward and feedback position control for passive bilateral teleoperation with delays," *Isa T*, vol. 85, pp. 200-213, 2019.
- [219] S. Zhai and J. Senders, "Investigating coordination in multidegree of freedom control II: Correlation analysis in 6 DOF tracking," the the 41st Annual Meeting of the Human Factors and Ergonomics Society, Santa Monica, CA, 1998.
- [220] N. Xi, T. J. Tarn, and A. K. Bejczy, "Intelligent planning and control for multirobot coordination: An event-based approach," *Ieee T Robotom Autom*, vol. 12(3), pp. 439-452, 1996.
- [221] A. E. Sklar and N. B. Sarter, "Good vibrations: tactile feedback in support of attention allocation and human-automation coordination in event-driven domains," *Hum Factors*, vol. 41(4), pp. 543-552, 1999.
- [222] K. Huang, D. Chitrakar, R. Mitra, *et al.*, "Characterizing limits of vision-based force feedback in simulated surgical tool-tissue interaction," the 42nd Annual International Conference of the IEEE-Engineering-in-Medicine-and-Biology-Society (EMBC), Montreal, CANADA, 2020.
- [223] C. E. Reiley and G. D. Hager, "Task versus subtask surgical skill evaluation of robotic minimally invasive surgery," the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention, Imperial Coll, London, ENGLAND, 2009.
- [224] R. V. Patel, S. F. Atashzar, and M. Tavakoli, "Haptic feedback and force-based teleoperation in surgical robotics," *P Ieee*, vol. 110(7), pp. 1012-1027, 2022.

- [225] Q. Gao, J. B. Wu, Y. Song, *et al.*, "A novel contact force measurement scheme for slave catheter robot in robotic endovascular surgery," *IEEE International Conference on Mechatronics and Automation (IEEE ICMA)*, pp. 164-169, 2021.
- [226] L. Gan, W. K. Duan, T. O. Akinyemi, *et al.*, "Development of a fiber bragg grating-based force sensor for minimally invasive surgery-case study of ex-vivo tissue palpation," *Ieee T Instrum Meas*, vol. 72, 2023.
- [227] N. Werner, G. Nickenig, and J. M. Sinning, "Complex PCI procedures: challenges for the interventional cardiologist," *Clin Res Cardiol*, vol. 107, pp. S64-S73, 2018.
- [228] T. D. Nagy and T. Haidegger, "Performance and capability assessment in surgical subtask automation," *Sensors*, vol. 22(7), 2022.
- [229] W. J. Du, O. M. Omisore, W. Duan, *et al.*, "Exploration of interventionists' technical manipulation skills for robot-assisted intravascular PCI catheterization," *Ieee Access*, vol. 8, pp. 53750-53765, 2020.
- [230] Z. B. Song, W. J. Zhang, W. H. Zhang, *et al.*, "A novel biopsy capsule robot based on high-speed cutting tissue," *Cyborg Bionic Syst*, vol. 2022, 2022.
- [231] L. Li, X. J. Li, B. Ouyang, *et al.*, "Three-dimensional collision avoidance method for robot-assisted minimally invasive surgery," *Cyborg Bionic Syst*, vol. 4, 2023.
- [232] J. Z. Flores and J. P. Radoux, "Catheter tracking using a convolutional neural network for decreasing interventional radiology X-ray exposure," *IRBM*, vol. 44(2), pp. 100737, 2023.
- [233] W. J. Wang, Y. Luo, J. Wang, *et al.*, "ToolNet-X: surgical instrument detection combined with high-order spatial interaction," *IEEE Transactions on Medical Robotics and Bionics*, vol. 5(4), pp. 857-866, 2023.
- [234] D. Sarikaya, J. J. Corso, and K. A. Guru, "Detection and localization of robotic tools in robot-assisted surgery videos using deep neural networks for region proposal and detection," (in English), *Ieee T Med Imaging*, vol. 36(7), pp. 1542-1549, 2017.
- [235] M. Allan, S. Ourselin, S. Thompson, *et al.*, "Toward detection and localization of instruments in minimally invasive surgery," *Ieee T Bio-Med Eng*, vol. 60(4), pp. 1050-1058, 2013.
- [236] J. D. Wang, K. Sun, T. Cheng, *et al.*, "Deep high-resolution representation learning for visual recognition," (in English), *Ieee T Pattern Anal*, vol. 43(10), pp. 3349-3364, 2021.
- [237] R. Y. Xia, G. Q. Li, Z. W. Huang, *et al.*, "Classify and localize threat items in X-Ray imagery with multiple attention mechanism and high-resolution and high-semantic features," *Ieee T Instrum Meas*, vol. 70, 2021.
- [238] B. W. Cheng, B. Xiao, J. D. Wang, *et al.*, "HigherHRNet: scale-aware representation learning for bottom-up human pose estimation," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5385-5394, 2020.
- [239] E. Benson, M. P. Pound, A. P. French, *et al.*, "Deep hourglass for brain tumor segmentation," *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Brainles 2018, Pt II*, vol. 11384, pp. 419-428, 2019.
- [240] M. Shafiq and Z. Q. Gu, "Deep residual learning for image recognition: a survey," *Appl Sci-Basel*, vol. 12(18), Sep 2022.
- [241] G. Quellec, M. Lamard, B. Cochener, *et al.*, "Real-time segmentation and recognition of surgical tasks in cataract surgery videos," *Ieee T Med Imaging*, vol. 33(12), pp. 2352-2360, 2014.
- [242] C. Loukas, "Surgical phase recognition of short video shots based on temporal modeling of deep features," presented at the 12th International Joint Conference on Biomedical Engineering Systems and Technologies, Prague, CZECH REPUBLIC, 2019.
- [243] Y. Q. Hou, W. Zhang, Q. Liu, *et al.*, "Adaptive kernel selection network with attention constraint for surgical instrument classification," *Neural Comput Appl*, vol. 34(2), pp. 1577-1591, 2022.
- [244] R. Q. Li, G. B. Bian, X. H. Zhou, *et al.*, "A two-stage framework for real-time guidewire endpoint localization," *Medical Image Computing and Computer Assisted Intervention - Miccai 2019*, vol. 11768, pp. 357-365, 2019.
- [245] M. C. Zhou, X. Y. Guo, M. Grimm, *et al.*, "Needle detection and localisation for robot-assisted subretinal injection using deep learning," *Caai T Intell Techno*, vol. Early Access, 2023.

- [246] E. M. Cronin, B. J. Wisnoskey, R. A. Rizzo, *et al.*, "Real-time guidewire localization using impedance-based electroanatomic mapping: experimental results and clinical validation during cryoballoon ablation of atrial fibrillation," *Europace*, vol. 15(11), pp. 1669-1676, 2013.
- [247] J. Bedel F. Vallée, A. Mari, *et al.*, "Guidewire localization by transthoracic echocardiography during central venous catheter insertion: a periprocedural method to evaluate catheter placement," *Intens Care Med*, vol. 39(11), pp. 1932-1937, 2013.
- [248] O. Galante, R. Sadeh, A. Smoliakov, *et al.*, "Real time wire localization by ultrasound during central line insertion improves accurate positioning in all upper torso sites," *J Crit Care*, vol. 67, pp. 79-84, 2022.
- [249] W. Farsal, S. Anter, and M. Ramdani, "Deep learning: An overview," the 12th International Conference on Intelligent Systems - Theories and Applications (SITA), Mohammed V Univ Rabat, Ecole Mohammadia Ingenieurs, Rabat, MOROCCO, 2018.
- [250] S. G. Jiang, S. T. Teng, J. Lu, *et al.*, "PixelTopoS: a pixel-topology-coupled guidewire tip segmentation framework for robot-assisted intervention," *Int J Comput Ass Rad*, vol. 17(2), pp. 329-341, 2022.
- [251] Y. J. Zhou, S.Q. Liu, X.L. Xie, *et al.*, "A real-time multi-task framework for guidewire segmentation and endpoint localization in endovascular interventions," the IEEE International Conference on Robotics and Automation ICRA, Xian, China, 2021.
- [252] D. H. Li and A. Barbu, "Training a steerable CNN for guidewire detection," the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), ELECTR NETWORK, 2020.
- [253] R. T. Liu, C. H. Ren, M. N. Fu, *et al.*, "Platelet detection based on improved YOLO_v3," *Cyborg Bionic Syst*, vol. 2022, 2022.
- [254] I. Ullah, P. Chikontwe, and S. H. Park, "Guidewire tip tracking using U-Net with shape and motion constraints," presented at the 1st International Conference on Artificial Intelligence in Information and Communication (Icaic 2019), Okinawa, Japan, 2019.
- [255] Y. J. Zhou, X. L. Xie, X. H. Zhou, *et al.*, "A real-time multifunctional framework for guidewire morphological and positional analysis in interventional X-Ray fluoroscopy," *Ieee T Cogn Dev Syst*, vol. 13, no. 3, pp. 657-667, 2021.
- [256] M. Soleimani, A. Vahidi, and B. Vaseghi, "Two-dimensional stockwell transform and deep convolutional neural network for multi-class diagnosis of pathological brain," *Ieee T Neur Sys Reh*, vol. 29, pp. 163-172, 2021.
- [257] X. Li, Y. C. Jiang, Y. L. Liu, J *et al.*, "RAGCN: region aggregation graph convolutional network for bone age assessment from X-Ray images," *Ieee T Instrum Meas*, vol. 71, 2022.
- [258] J. R. Cheng, W. Xiong, W. Y. Chen, *et al.*, "Pixel-level crack detection using U-Net," the IEEE-Region-10 Conference (IEEE TENCON), IEEE Reg 10, SOUTH KOREA, 2018.
- [259] Z. W. Zhou, N. Tajbakhsh, J. M. Liang, *et al.*, "UNet plus plus: a nested U-Net architecture for medical image segmentation," *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Dlmia 2018*, vol. 11045, pp. 3-11, 2018.
- [260] Yadavendra and S. Chand, "Semantic segmentation of human cell nucleus using deep U-Net and other versions of U-Net models," *Network-Comp Neural*, vol. 33(3-4), pp. 167-186, 2022.
- [261] C. Beeche, J. P. Singh, J.K. Leader, *et al.*, "Super U-Net: a modularized generalizable architecture," *Pattern Recogn*, vol. 128, 2022.
- [262] J. X. You, D. H. Jiang, Y. Ma, *et al.*, "SpindleU-Net: an adaptive U-Net framework for sleep spindle detection in single-channel EEG," *Ieee T Neur Sys Reh*, vol. 29, pp. 1614-1623, 2021.
- [263] X. J. Wu, L. T. Qiu, X. D. Gu, *et al.*, "Deep learning-based generic automatic surface defect inspection (ASDI) with pixelwise segmentation," *Ieee T Instrum Meas*, vol. 70, 2021.
- [264] H. Zhou, C. Tawk, and G. Alici, "A 3D printed soft robotic hand with embedded soft sensors for direct transition between hand gestures and improved grasping quality and diversity," *Ieee T Neur Sys Reh*, vol. 30, pp. 550-558, 2022.
- [265] G. Canal, Y. Diaz-Mercado, M. Egerstedt, *et al.*, "A low-complexity brain-computer interface for high-complexity robot swarm control," *Ieee T Neur Sys Reh*, vol. 31, pp. 1816-1825, 2023.

- [266] D. Gupta and R. S. Anand, "A hybrid edge-based segmentation approach for ultrasound medical images," (in English), *Biomed Signal Proces*, vol. 31, pp. 116-126, 2017.
- [267] Z. C. Jing, J. T. Ye, and G. L. Xu, "A geometric flow approach for region-based image segmentation-theoretical analysis," *Acta Math Appl Sin-E*, vol. 34(1), pp. 65-76, 2018.
- [268] F. LaRocca, S. J. Chiu, R. P. McNabb, *et al.*, "Robust automatic segmentation of corneal layer boundaries in SDOCT images using graph theory and dynamic programming," *Biomed Opt Express*, vol. 2(6), pp. 1524-1538, 2011.
- [269] Y. B. Wang, Y. X. Xiao, and Z. J. Wang, "Transition algorithm based on jump operator applied to image threshold segmentation," the 7th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 2022.
- [270] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *Ieee T Pattern Anal*, vol. 39(4), pp. 640-651, 2017.
- [271] L. C. Chen, G. Papandreou, I. Kokkinos, *et al.*, "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *Ieee T Pattern Anal*, vol. 40(4), pp. 834-848, 2018.
- [272] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *Ieee T Pattern Anal*, vol. 39(12), pp. 2481-2495, 2017.
- [273] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention*, vol. 9351, pp. 234-241, 2015.
- [274] M. B. Shuvo, R. Ahommed, S. Reza, *et al.*, "CNL-UNet: a novel lightweight deep learning architecture for multimodal biomedical image segmentation with false output suppression," *Biomed Signal Proces*, vol. 70, 2021.
- [275] L. C. E. Chen, Y. K. Zhu, G. Papandreou, *et al.*, "Encoder-decoder with atrous separable convolution for semantic image segmentation," the 15th European Conference on Computer Vision (ECCV), Munich, GERMANY, 2018.
- [276] P. Cao, S. M. Noworolski, O. Starobinets, *et al.*, "Development of conditional random field insert for unet-based zonal prostate segmentation on t2-weighted mri," *arXiv preprint arXiv:2002.06330*, 2020.
- [277] S. L. Feng, H. M. Zhao, F. Shi, *et al.*, "CPFNet: context pyramid fusion network for medical image segmentation," *Ieee T Med Imaging*, vol. 39(10), pp. 3008-3018, 2020.
- [278] C. Li and C. Z. Xu, "VCE-Net: a new type of real-time polyp segmentation network," presented at the IEEE 8th International Conference on Signal and Image Processing (ICSIP), Wuxi, China, 2023.
- [279] F. X. He, T. L. Liu, and D. C. Tao, "Why resnet works? residuals generalize," *Ieee T Neur Net Lear*, vol. 31(12), pp. 5349-5362, 2020.
- [280] K. Simonyan and A. Zisserman, "Very deep convolutional networks for largescale image recognition," *arXiv preprint arXiv:1409.1556*, 2023.
- [281] L. Yang, Y. G. Gu, G. B. Bian, *et al.*, "An attention-guided network for surgical instrument segmentation from endoscopic images," *Comput Biol Med*, vol. 151, 2022.
- [282] Y. H. Zheng, M.J. Er, S.W. Shen, *et al.*, "An improved image segmentation model based on U-Net for interventional intravascular robots," the 2021 4th International Conference on Intelligent Autonomous Systems (Icoias 2021), Wuhan, China, 2021.
- [283] X. Y. Xu, S. Tu, Y. F. Xue, *et al.*, "Visualizing CNN: An axperimental comparative study," Lecture Notes in Computer Science, pp. 199-212, 2023.
- [284] Q. Guari, Y.J. Wang, B. Ping, *et al.*, "Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study," *J Cancer*, vol. 10(20), pp. 4876-4882, 2019.
- [285] L. Geng, S. Q. Zhang, J. Tong, *et al.*, "Lung segmentation method with dilated convolution based on VGG-16 network," *Comput Assist Surg*, vol. 24, pp. 27-33, 2019.
- [286] A. Haider, M. Arsalan, S. H. Nam, *et al.*, "Multi-scale feature retention and aggregation for colorectal cancer diagnosis using gastrointestinal images," *Eng Appl Artif Intel*, vol. 125, 2023.

-
- [287] B. Hanin, "Universal function approximation by deep neural nets with bounded width and ReLU activations," *Mathematics-Basel*, vol. 7(10), 2019.
- [288] T. Y. Hu, Y. C. Deng, Y. W. Deng, *et al.*, "Fully convolutional network variations and method on Small Dataset," the IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, Peoples R China, 2021.
- [289] S. Alqazzaz, X. F. Sun, X. Yang, *et al.*, "Automated brain tumor segmentation on multi-modal MR image using SegNet," *Comput Vis Media*, vol. 5(2), pp. 209-219, 2019.
- [290] Y. Y. Liu, X. T. Bai, J. F. Wang, *et al.*, "Image semantic segmentation approach based on DeepLabV3 plus network with an attention mechanism," *Eng Appl Artif Intel*, vol. 127, 2024.
- [291] W. J. Du, G.L. Yi, O.M. Omisore, *et al.*, "Guidewire endpoint detection based on pixel-adjacent relation during robot-assisted intravascular Catheterization: In Vivo Mammalian Models," *Adv Intell Syst-Ger*, vol. 6(4), 2024.
- [292] W. Yang, J. L. Zhang, Z. B. Chen, *et al.*, "An efficient semantic segmentation method based on transfer learning from object detection," (in English), *Iet Image Process*, vol. 15(1), pp. 57-64, 2021.
- [293] S. Kalidindi and J. Baradwaj, "Advancing radiology with GPT-4: Innovations in clinical applications, patient engagement, research, and learning," *European Journal of Radiology Open*, vol. 13, 2024.
- [294] Y. H. Huang, X. Yang, L. Liu, *et al.*, "Segment anything model for medical images?," *Med Image Anal*, vol. 92, 2024.

Author's Biography

Name:	Wenjing Du
Sex:	Female
Place of Origin:	Shanxi Province, China
Date of Birth:	1989.03.11
2022.02 — 2024.12	Ph.D. in Computer Science & Operations Research, University of Nottingham, Ningbo, China. Shenzhen Institutes of Advanced Technology, Chinese Academic Science, Shenzhen, China.
2019.01 — 2022.01	Assistant Professor, Shenzhen Institutes of Advanced Technology, Chinese Academic Science, Shenzhen, China.
2016.07 — 2018.12	Research Assistant, Shenzhen Institutes of Advanced Technology, Chinese Academic Science, Shenzhen, China.
2013.09 — 2016.06	Master of Instrument and Meter Engineering, Guilin University of Electronic Technology, Guilin, China.
2009.09 — 2013.07	Bachelors of Communication Engineering, Chang'an College of Xidian University, Xian, China.

【Publication During the Doctoral Degree】

Publication Related to Thesis:

- [1] **Wenjing Du**, Guanlin Yi, Olatunji Mumini Omisore, Wenke Duan, Xingyu Chen, Toluwanimi Akinyemi, Jiang Liu, Boon Giin Lee, Lei Wang. Guidewire Endpoint Detection Based on Pixel-Adjacent Relation during Robot-Assisted Intravascular Catheterization: In Vivo Mammalian Models, *Advanced Intelligent Systems*, 2024, 6(4):2300687. (IF=6.8, JCR 1)
- [2] **Wenjing Du**, Guanlin Yi, Olatunji Mumini Omisore, Wenke Duan, Toluwanimi Akinyemi, Xingyu Chen, Jiang Liu, Boon-Giin Lee, and Lei Wang. Analyzing Surgeon-Robot Cooperative Performance in Robot-assisted Intravascular Catheterization, *IEEE Transactions on Human-Machine Systems*, 2024, 54(6):698-710 (IF=3.5, JCR 1)
- [3] **Wenjing Du**, Olatunji Mumini Omisore, Wenke Duan, Lu Gan, Toluwanimi Oluwadra Akinyemi, Boon-Giin Lee, Jiang Liu, Lei Wang. Exploring operators' natural behaviors to predict catheterization trial outcomes in robot-assisted intravascular interventions. *IEEE Transactions on Medical Robotics and Bionics*. 2022, 4(3):682-695. (IF=3.4, JCR 2)
- [4] **Wenjing Du**, Guanlin Yi, Olatunji Mumini Omisore, Wenke Duan, Toluwanimi Oluwadara

- Akinyemi, Xingyu Chen, Lei Wang, Boon-Giin Lee, Jiang Liu. Guidewire Endpoint Detection Based on Pixel Adjacent Relation in Robot-assisted Cardiovascular Interventions, 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 2023 Jul 24, 1-5
- [5] Wenke Duan, Toluwanimi Akinyemi, **Wenjing Du**, Jun Ma, Xingyu Chen, Fuhao Wang, Olatunji Omisore, Jingjing Luo, Hongbo Wang, Lei Wang. Technical and clinical progress on robot-assisted endovascular interventions: A review. *Micromachines*, 2023, 14(1):197. (IF=3.0, JCR 2)
- [6] Toluwanimi Oluwadara Akinyemi, Olatunji Mumini Omisore, **Wenjing Du**, Wenke Duan, Xingyu Chen, Guanlin Yi, Lei Wang. Interventionalist hand motion recognition with convolutional neural network in robot-assisted coronary interventions, *IEEE Sensors Journal*, 2023, 23(15): 17725-17736. (IF=4.3, JCR 1)
- [7] Olatunji Mumini Omisore, Toluwanimi Oluwadara Akinyemi, **Wenjing Du**, Wenke Duan, Rita Orji, Thanh Nho Do, Lei Wang. Weighting-based deep ensemble learning for recognition of interventionalists' hand motions during robot-assisted intravascular catheterization. *IEEE Transactions on Human-Machine Systems*. 2022, 53(1):215-27. (IF=3.5, JCR 1)
- [8] Olatunji Mumini Omisore, Toluwanimi Oluwadara Akinyemi, Wenke Duan, **Wenjing Du**, Lei Wang. Multi-Lateral Branched Network for Tool Segmentation during Robot-assisted Endovascular Interventions, *IEEE Transactions on Medical Robotics and Bionics*, 2024. (IF=3.4, JCR 2)
- [9] Wenke Duan, Zihao Li, Olatunji M Omisore, **Wenjing Du**, Toluwanimi O Akinyemi, Xingyu Chen, Xing Gao, Hongbo Wang, Lei Wang. Development of an Intuitive Interface with Haptic Enhancement for Robot-Assisted Endovascular Intervention. *IEEE Transactions on Haptics*. 2023.4 (IF=2.4, JCR 2)
- [10] Toluwanimi O Akinyemi, Olatunji M Omisore, Xingyu Chen, Wenke Duan, **Wenjing Du**, Guanlin Yi, Lei Wang. Adapting neural-based models for position error compensation in robotic catheter systems. *Applied Sciences*. 2022, 12(21):10936. (IF=2.5, JCR 2)
- [11] Olatunji Mumini Omisore, Guanlin Yi, Yuhong Zheng, Toluwanimi Oluwadara Akinyemi, Wenke Duan, **Wenjing Du**, Xingyu Chen, Lei Wang. Endovascular Tool Segmentation with Multi-lateral Branched Network during Robot-assisted Catheterization, 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 2023 Jul 24, 1-4.
- [12] Xingyu Chen, Yinan Chen, Wenke Duan, Toluwanimi Oluwadara Akinyemi, Guanlin Yi, Jie Jiang, **Wenjing Du**, Olatunji Mumini Omisore. Design and evaluation of a learning-based vascular interventional surgery robot. *Fibers*. 2022, 10(12):106. (IF=4.0, JCR 1)
- [13] Yinan Chen, Heyun Feng, Xingyu Chen, Wenke Duan, **Wenjing Du**, Lei Wang, “一种手术机器人的主端力感知反馈操纵装置及方法”, CN202210529183.0

Other Publications:

- [14] Olatunji Omisore, Wenke Duan, **Wenjing Du**, Shipeng Han, Toluwanimi Akinyemi, Lei Wang, On Task-specific Autonomy in Robotic Interventions: A Multimodal Learning-based Approach for Multi-level Skill Assessment during Cyborg Catheterization, *Authorea Preprints*, 2023

- [15] Xingyu Chen, Xiaohui Xiong, Xuemiao Wang, Peng Li, Shimei Wang, Toluwanimi Oluwadara Akinyemi, Wenke Duan, **Wenjing Du**, Olatunji Omisore, Lei Wang, Design of a teleoperated robotic bronchoscopy system for peripheral pulmonary lesion biopsy. arXiv preprint arXiv:2306.09598, 2023
- [16] Xiaohui Xiong, Peng Li, Xingyu Chen, Jie Jiang, Lei Wang, Olatunji Mumini Omisore, **Wenjing Du**, “一种刚度连续可调鞘管及其刚度调节方法和手术设备”, CN202310240820.7

【Research Projects During Degree Program】

- [1] National Natural Science Foundation of China (grant no. U1713219) Title: Research on force sensing and guidewire manipulation mechanism of wearable precision positioning interventional robot
(穿戴式精确定位介入手术机器人的力觉感知与导丝操控机理研究)
- [2] National Key Research & Development Plan (grant no. 2019YFB1311700) Title: Research on the key technology and application of multi-instrument cooperative delivery robot for vascular interventional surgery for complex lesions
(面向复杂病变的多器械协同递送血管介入手术机器人关键技术及应用研究)
- [3] National Natural Science Foundation of China (grant no. U21A20480) Title: Key robotic-assisted technologies on the accurate positioning and aspiration biopsy for the peri-pulmonary small nodules
(机器人辅助肺周小结节精准定位与穿刺活检关键技术研究)
- [3] National Key Research & Development Plan (grant no. 2024YFF1206900) Title: Research on key technologies of scene navigation in robot-assisted intelligent surgery and Hybrid operating room
(机器人辅助智慧手术与复合手术室情景导航关键技术研究)