



HOKKAIDO UNIVERSITY

Title	From Image Preprocessing to 3D Modeling : Deep Learning Approaches for Coronary Artery Analysis
Author(s)	李, 彦君
Degree Grantor	北海道大学
Degree Name	博士(保健科学)
Dissertation Number	甲第16334号
Issue Date	2025-03-25
DOI	https://doi.org/10.14943/doctoral.k16334
Doc URL	https://hdl.handle.net/2115/94971
Type	doctoral thesis
File Information	Yanjun_Li.pdf



学 位 論 文

From Image Preprocessing to 3D Modeling: Deep Learning
Approaches for Coronary Artery Analysis

(深層学習手法を用いた冠動脈解析のための画像前処理から 3D モデリン
グに関する研究)

LI YANJUN

北海道大学大学院保健科学院
保健科学専攻保健科学コース

2024年度

From Image Preprocessing to 3D Modeling: Deep Learning Approaches for Coronary Artery

Analysis

(深層学習手法を用いた冠動脈解析のための画像前処理
から 3D モデリングに関する研究)

YANJUN LI

Division of Advanced Health Sciences
Graduate School of Health Sciences
Hokkaido University
Japan

Principal Supervisor:
Prof. HIROYUKI SUGIMORI

Dec 2024

DECLARATION

I hereby declare that this thesis represents my own work which has been done after registration for the degree of PhD at Hokkaido University, and has not been previously included in a thesis or dissertation submitted to this or any other institution for a degree, diploma or other qualifications.

I have read the University's current research ethics guidelines, and accept responsibility for the conduct of the procedures in accordance with the University's Research Ethics Committee. I have attempted to identify all the risks related to this research that may arise in conducting this research, obtained the relevant ethical and/or safety approval (where applicable), and acknowledged my obligations and the rights of the participants.

Signature: YANJUNLI

Date: 2024/12/02

Acknowledgments

I would like to extend my heartfelt gratitude to my supervisor, Professor Hiroyuki Sugimori, whose expert guidance and unwavering support have been instrumental in the successful completion of my doctoral research. His insightful feedback and encouragement have helped me grow both as a researcher and as a professional. I am also deeply appreciative of Lecturer Dr. Takaaki Yoshimura for his invaluable support and the engaging, thought-provoking discussions that significantly enriched my work.

I am equally thankful to my lab members, whose camaraderie and collaboration played a crucial role in shaping my research journey. Their active participation in discussions, along with their generous assistance and friendship, has made my time in the lab an enjoyable and productive experience.

I would also like to express my sincere thanks to Hokkaido University for granting me the Ambitious Doctoral Fellowship and the EXEX Doctoral Fellowship. These scholarships provided me with the necessary resources and opportunities to pursue my research goals and advance my academic career.

Finally, I would like to acknowledge the unwavering love and support of my wife. Her constant encouragement and belief in me have been a source of strength throughout this challenging yet rewarding journey, and I am deeply grateful for all she has done.

Abstract

Coronary artery disease (CAD) remains a leading cause of mortality worldwide, necessitating advancements in diagnostic and treatment methods. This thesis explores novel deep learning approaches to improve the analysis of coronary artery imaging, addressing challenges in stenosis detection, and 3D reconstruction. Three key contributions are presented: (1) a Hessian-based image preprocessing method integrated with image fusion, enhancing visualization of coronary angiography; (2) an attention-enhanced YOLO-based framework for stenosis localization, improving detection accuracy in complex angiographic scenarios; and (3) an automated deep learning pipeline leveraging GhostNet and transformers for the extraction and 3D reconstruction of the right coronary artery (RCA) from CT images. Comprehensive experiments validate the efficacy of the proposed methods across multiple tasks in coronary artery analysis. The Hessian-based image preprocessing combined with image fusion significantly enhanced vessel visualization, leading to a 5% increase in detection accuracy ($AP_{50} = 87.1\%$) using the YOLOv10-X model compared to raw dataset. The proposed attention-enhanced YOLO-based network achieved outstanding performance in stenosis localization, with an AP_{50} of 90.5% at the zero-degree left coronary artery (LCA) imaging angle, the highest accuracy across all tested angles. Additionally, the automated deep learning pipeline for the extraction and 3D reconstruction of the right coronary artery achieved a segmentation F1 score of 0.866 and a mean IoU of 0.835. The 3D reconstruction process provided superior visual clarity, making it suitable for clinical applications. These contributions collectively advance the accuracy, efficiency, and practicality of coronary artery disease analysis, offering

transformative potential for real-time medical diagnostics and treatment planning. This study highlights the transformative potential of advanced deep learning architectures in CAD diagnosis, offering pathways for real-time, precise, and robust medical imaging solutions. Future research will focus on extending these techniques to broader datasets and refining latency for real-time clinical applications.

Keywords: Artificial Intelligence, Coronary Artery Disease, Deep Learning, 3D Reconstruction, Medical Image Processing

Abbreviations and symbols

CAD	Coronary artery disease
ICA	Invasive coronary angiography
CNNs	Convolutional neural networks
FFR	Fractional flow reserve
HFV	Hessian-based frangi vesselness
RCA	Right coronary artery
LCA	Left coronary artery
LAO	Left anterior oblique
RAO	Right coronary artery
IHS	Intensity–hue–saturation
RGB	Red–green–blue
R-FCN	Region-based fully convolutional networks
R-CNN	Region-based convolutional neural networks
FCN	Fully convolutional networks
RPN	Region proposal network
ResNet	Residual network
RoIs	Regions of interest
YOLO	You Only Look Once
NMS	Non-maximum suppression
VFL	Varifocal loss
CIOU	Complete-IoU
DFL	Deep feature loss

PGI	Programmable gradient information
GELAN	General efficient layer aggregation network
TAL	Task alignment learning
RT-DETR	Real-Time detection transformer
DETR	Detection transformer
AP	Average precision
TP	True positive
mAP	Mean average precision
IoU	Intersection over union
FLOPs	Floating point operation quantification
EMCAM	Efficient multi-scale convolutional attention module
MSCAB	Multi-scale convolution block
CAB	Channel attention block
SAB	Spatial attention block
PPA	Parallelized patch-aware attention
HCF-Net	Hierarchical context fusion network
AACN	Attention-aware compositional network
PSA	Partial self-attention
MHSA	Multi-head self-attention
RPB	relative position bias
ViT	Vision transformer
FCN	Fully convolutional network
ASPP	Atrous spatial pyramid pooling
mIoU	Mean intersection over union

Table of Contents

Declaration	i
Acknowledgments	ii
Abstract	iii
Abbreviations and symbols	v
Table of Contents	vii
List of Tables	x
List of Figures	xi
List of Algorithms	xii
General Introduction	1
Chapter 1 A Hessian-Based Deep Learning Preprocessing Method for Coronary Angiography Image Analysis	3
1.1 Introduction	3
1.2 Materials and Methods	8
1.2.1 Materials	9
1.2.2 Preprocessing Approaches	9
1.2.3 Model descriptions	14
1.2.4 Evaluation Indicators	19

1.3	Experiments and Results	22
1.3.1	Experiments	22
1.3.2	Results	24
1.4	Discussion	34
1.5	Conclusions	37

Chapter 2 YOLO-Based Attention-Enhanced Network for Coronary

Artery Stenosis Localization 39

2.1	Introduction	39
2.2	Methods	43
2.2.1	Agent Attention mechanism	43
2.2.2	Proposed method	44
2.3	Experiments and Results	47
2.4	Discussion	49
2.5	Conclusion	50

Chapter 3 Automated Extraction and 3D Reconstruction of the Right

Coronary from CT Images Using Deep Learning 51

3.1	Introduction	51
3.2	Materials and Methods	55
3.2.1	Materials	55
3.2.2	Model descriptions	56
3.2.3	3D Reconstruction	67
3.2.4	Evaluation Indicators	69
3.3	Experiment and Results	71
3.4	Discussion	81
3.5	Conclusion	84

Chapter 4 Conclusion and Future Work 85

4.1	General Conclusion	85
4.2	Future Work	87

Bibliography	89
List of Achievements	100

List of Tables

1.1	Data Distribution in CAG Dataset.	23
1.2	Performance Comparison of Different Models on Raw CAG Dataset. .	28
1.3	Performance Comparison of Different Models on Preprocessed CAG Dataset.	29
1.4	Comparison of Latency and Accuracy with and without Preprocess- ing Method.	32
1.5	Performance of YOLOv10-X Model on Different Imaging Angles. . .	34
2.1	Performance of YOLO-Based Attention-Enhanced Network on Two Datasets.	48
2.2	Performance of YOLO-Based Attention-Enhanced Network on Six Imaging Angles.	48
3.1	Performance Comparison of Different Models on CT Dataset. . . .	73
3.2	Comparison of Laplacian Pyramid and Gaussian Smoothing Interpo- lation Methods.	80

List of Figures

1.1	Overview of Preprocessing Performance on Six Imaging Angles	26
1.2	Comparison of YOLOv10-X Detection Results on Original Image and Preprocessed Image.	30
2.1	Structure of Proposed YOLO-Based Attention-Enhanced Network . .	46
3.1	The Architecture of Position-Sensitive Axial-Attention	63
3.2	The Architecture of Gated Axial-Attention	64
3.3	The Architecture of Proposed Attention Mechanism	64
3.4	The Structure of the Proposed Method in Chapter 3	66
3.5	Prediction Results of Different Networks for Two Example Input Im- ages.	74
3.6	The Mesh Grid Generated from Prediction Results	76
3.7	The Result of Volume Rendering	77
3.8	Comparison Between Laplacian Pyramid Interpolation and Gaussian Smoothing Interpolation	78
3.9	An Example Position to Intuitively Observe Stenosis	79

List of Algorithms

1	Vessel Enhancement and Image Fusion	13
---	---	----

General Introduction

Coronary artery disease is a leading global health concern, characterized by the narrowing and hardening of arteries that restrict blood flow to the heart, significantly impacting mortality rates, quality of life, and healthcare systems worldwide [1, 2]. The "gold standard" for diagnosing CAD has traditionally been invasive coronary angiography (ICA), which involves visualizing coronary arteries using a contrast agent and X-ray imaging [3]. However, this method heavily relies on the subjective judgment of radiologists, leading to variability in diagnostic outcomes [4]. The anatomical complexity of coronary arteries, combined with dynamic blood flow, further complicates the accurate assessment of arterial stenosis.

Advancements in deep learning have revolutionized medical imaging, particularly in the detection and analysis of coronary artery stenosis. Convolutional neural networks (CNNs) and their variants have been successfully applied to vessel segmentation, stenosis detection, and feature extraction. However, challenges such as limited computational efficiency, high parameter counts, and difficulties in accurately processing multi-angle imaging persist [5, 6]. Furthermore, traditional approaches often overlook the complex geometry and smaller branches of the coronary network, which are crucial for comprehensive diagnostics.

Transformer-based architectures have emerged as a promising solution, offering the ability to capture global context and model dependencies across entire datasets. These models, particularly when integrated with lightweight networks like GhostNet, provide a balanced approach to achieving high accuracy with reduced computational demands [7, 8]. By combining GhostNet's efficiency with the global attention

capabilities of transformers, modern frameworks have demonstrated significant improvements in segmenting complex vascular structures, making them suitable for clinical environments requiring real-time processing.

In this thesis, we discuss three methods, presented respectively in Chapters 1, 2, and 3. These are: (1) a Hessian matrix-based image preprocessing method combined with image fusion; (2) a YOLO-based coronary artery stenosis detection network incorporating attention mechanisms and feedforward paths; and (3) a GhostNet and transformer-based approach for CT coronary artery RCA extraction and 3D reconstruction. These methods not only enhance visualization and segmentation capabilities but also enable more accurate and efficient stenosis detection, addressing critical limitations in existing methodologies. By leveraging state-of-the-art technologies, this study aims to advance CAD diagnosis and treatment planning, ultimately improving patient outcomes.

Chapter 1

A Hessian-Based Deep Learning Preprocessing Method for Coronary Angiography Image Analysis

1.1 Introduction

Coronary artery disease emerges as a critical cardiovascular medical issue characterized by progressive arterial narrowing and hardening, substantially influencing cardiac blood circulation [1]. As a mounting global health concern, CAD's prevalence continues to surge dramatically. The condition significantly impacts public health by elevating mortality and morbidity rates, dramatically diminishing patient life quality, and imposing considerable burdens on healthcare infrastructures worldwide [2]. Historically, coronary angiography has been revered as the diagnostic benchmark for identifying coronary artery disease. Invasive coronary angiography represents a pivotal diagnostic technique for detecting arterial stenosis in clinical settings. During this procedure, medical professionals introduce a radiopaque contrast agent into coronary arteries, generating a series of X-ray images to visualize

potential blockages [3]. The assessment of CAD severity predominantly relies on measuring percent stenosis, a method intrinsically dependent on physician visual interpretation. Clinical guidelines typically employ specific stenosis thresholds: a 50% narrowing often prompts consideration of coronary revascularization, while a 70% stenosis signifies clinically significant arterial obstructions [9]. Nevertheless, this diagnostic approach harbors inherent limitations that challenge its absolute precision.

The diagnosis of CAD through ICA is significantly influenced by radiologists' subjective expertise and judgment, potentially leading to inconsistent diagnostic interpretations [4]. The diagnostic process is susceptible to various human factors, including physician fatigue and disparities in professional experience, which can compromise diagnostic accuracy. Traditional angiographic imaging faces multiple technical challenges that impede precise visualization. Image quality is frequently compromised by constraints such as radiation dose limitations, non-optimal imaging perspectives, and uneven distribution of contrast agents. These factors contribute to reduced image contrast and structural blurring, thereby substantially complicating manual assessment of coronary artery stenosis. The inherent complexity of coronary arterial anatomy further compounds diagnostic difficulties. Each individual's arterial network exhibits unique three-dimensional characteristics, creating substantial interpretation challenges. The dynamic physiological environment introduces additional complexity, as blood flow velocity and pressure variations during the cardiac cycle can dramatically alter stenosis appearance. These multifaceted challenges underscore the critical need for advanced diagnostic methodologies that can transcend current limitations. Innovative imaging techniques capable of capturing real-time, precise arterial structural and functional information are essential to enhance the accuracy and reliability of CAD diagnosis.

Deep learning technologies, particularly convolutional neural networks, have emerged as transformative solutions to the persistent challenges in coronary artery disease diagnosis. Recent research has demonstrated remarkable progress in im-

age processing techniques that enable more precise and reliable stenosis detection and quantification. The existing literature reveals a diverse range of deep learning applications in cardiovascular imaging. Researchers have made significant strides in multiple domains, including blood vessel extraction [10, 11], automated stenotic lesion localization [6, 12], and non-invasive fractional flow reserve (FFR) estimation from angiographic images [13]. Innovative mathematical methodologies have also been proposed for stenosis rate calculation [5]. Despite these advancements, current research exhibits notable limitations. While groundbreaking studies often concentrate on stenotic regions within the primary vessel, they frequently neglect the critical implications of stenosis in branch vessels. This research gap stems from substantial technical challenges, including vessel overlay complexities, inconsistent contrast agent distribution, and suboptimal imaging perspectives that impede clear branch vessel visualization. Existing segmentation algorithms frequently struggle with real-time clinical implementation, and many research approaches remain confined to zero-degree imaging angles. These constraints significantly restrict the comprehensive diagnostic potential of coronary angiography. Consequently, the detection of stenosis within the intricate coronary arterial network continues to represent a complex and promising frontier of medical imaging research. The ongoing evolution of deep learning techniques offers substantial potential to address these critical diagnostic challenges, potentially revolutionizing our approach to identifying and characterizing coronary artery stenosis.

Addressing the complex challenges of the coronary arterial network, characterized by its intricate branching and geometric complexity, researchers have increasingly turned to Hessian matrix-based filtering as a promising visualization technique [14]. This innovative approach exploits the inherent relationships between Hessian matrix eigenvalues to identify and extract tubular blood vessel structures in medical imaging. The Hessian-based Frangi vesselness (HFV) filter has emerged as a crucial tool for vascular enhancement and quantification, with extensive applications demonstrated across multiple research domains [15, 16, 17]. Previous research has

highlighted the filter’s potential to improve image quality, particularly in coronary angiography, thereby enhancing deep learning networks’ ability to learn stenotic features [18]. Advanced hybrid enhancement techniques have further expanded the filter’s capabilities, combining approaches like region growth with HFV to improve background clarity and reduce noise levels [19, 20]. However, a significant research gap persists: while numerous studies have explored the HFV filter’s image enhancement potential, few have systematically investigated its impact on deep learning networks’ feature learning capabilities. This limitation may be attributed to the historical constraints in deep learning network development, which potentially hindered researchers from obtaining conclusive evidence about the filter’s comprehensive diagnostic potential. Consequently, the full capabilities of Hessian matrix-based filtering in improving coronary artery image analysis remain an exciting and promising area of ongoing research.

The field of convolutional neural networks has witnessed remarkable advancements in stenosis identification and detection methodologies. In 2018, researchers introduced a modified CNN structure that demonstrated significant potential in identifying stenotic regions within angiographic images [21]. This pioneering approach involved training networks on synthetic data and subsequently refining them using clinical datasets, achieving an impressive accuracy of approximately 90%. Notably, this method outperformed traditional feedforward neural networks, marking a substantial breakthrough in diagnostic imaging. Transfer learning emerged as another critical avenue of research, with studies exploring the optimization of pre-trained CNN architectures [22]. Investigators discovered that utilizing the top three layers of pretrained networks as feature extractors could yield substantial improvements in stenosis detection, highlighting the potential of adaptive learning techniques. Building upon these foundational research efforts, the scientific community developed CathAI, an innovative fully automated method for analyzing coronary artery stenosis in conventional angiography [23]. This breakthrough approach has significantly enhanced the standardization and reproducibility of coronary artery

stenosis assessment, representing a pivotal advancement in medical imaging diagnostics. These progressive methodologies underscore the transformative potential of deep learning technologies in addressing complex cardiovascular diagnostic challenges, offering more precise, consistent, and reliable approaches to identifying arterial stenosis.

The landscape of blood vessel segmentation has undergone significant transformations, with researchers developing a diverse array of innovative models and tracking-based techniques. These methodological advances encompass a wide range of approaches, including multiscale Gabor filtering [24], sophisticated Hessian filtering integrated with flux flow measurement [25], and segmentation strategies leveraging active contour methods [26]. Moreover, deep learning technologies have played a pivotal role in advancing medical image analysis, with long short-term memory recurrent neural networks emerging as particularly promising techniques [27]. Concurrently, transformer technology has rapidly gained prominence, revolutionizing medical image processing and segmentation. Recent years have witnessed the emergence of numerous groundbreaking methods utilizing transformer architectures [28, 29, 30]. These diverse and sophisticated approaches collectively represent a comprehensive technological ecosystem dedicated to enhancing our capabilities in medical imaging. The primary focus remains on advancing the detection of coronary artery stenosis and refining blood vessel segmentation techniques, demonstrating the field’s commitment to improving diagnostic precision and understanding cardiovascular imaging.

In this chapter, we implemented a preprocessing technique, designed to enhance coronary artery imagery through the integration of Hessian matrix multiscale filtering and image feature fusion. Our methodology prioritizes feature preservation, particularly for identifying small or distal vessels. The approach involved applying Hessian-based Frangi vesselness filtering to improve blood vessel visibility, followed by image fusion to accentuate vessel characteristics. We conducted a comprehensive evaluation by training and comparing various deep learning networks under consistent conditions, assessing the preprocessing technique’s impact on their performance.

Additionally, we investigated the influence of different imaging angles on stenosis localization. The experimental outcomes demonstrated satisfactory detection results. Our study has made the following contributions:

1. We substantially improved the identification of branch vessel stenosis. Traditional object detection approaches often overlook branch vessel stenosis due to accuracy constraints, despite the considerable clinical health risks these conditions pose. Our method enhances the potential for clinical application.
2. Our dataset incorporates cases from elderly patients, many featuring artifacts such as surgical stents, which introduce additional noise. The inclusion of these diverse data samples increases dataset complexity and improves the method’s practical applicability.
3. We conducted a systematic latency assessment comparing our method to direct image processing. Recognizing latency as a critical factor in real-time clinical detection, we evaluated the method’s potential for development and application in time-sensitive diagnostic scenarios.

1.2 Materials and Methods

Our methodological approach incorporated two dataset variants: the original raw dataset and a preprocessed dataset. For the preprocessed version, we implemented our proposed preprocessing technique, which commenced with the application of the Hessian-based Frangi vesselness filter and subsequently employed image fusion techniques to further enhance image quality. Both datasets were then utilized to assess the performance of various deep learning models. The comprehensive preprocessing procedure is systematically outlined in Algorithm 1, which provides a detailed sequential breakdown of the image preprocessing steps. This approach enabled a comparative evaluation of the preprocessing method’s impact on deep learning model performance, offering insights into the potential improvements in image quality and subsequent diagnostic capabilities.

1.2.1 Materials

This study utilized a comprehensive dataset sourced from 40 patients who underwent coronary angiographic examinations at Hokkaido University Hospital. To ensure methodological precision, experienced radiologists conducted meticulous evaluations of all angiographic sequences and static images. The data collection methodology involved systematically segmenting continuous angiographic recordings into discrete frames, standardized at a frequency of 15 frames per second. This approach generated a robust collection of 400 contrast-enhanced radiographic images. These images were strategically categorized into two primary classifications: right coronary artery and left coronary artery. Each classification incorporated images captured from three distinct angular projections: left anterior oblique (LAO), right anterior oblique (RAO), and zero-degree (ZERO) orientation.

1.2.2 Preprocessing Approaches

1.2.2.1 HFV Filter

Enhancement filters serve as fundamental tools for optimizing image quality and detail across various medical imaging modalities, including X-ray, MRI, CT, and ultrasound. These enhancements are indispensable for clinicians in their diagnostic procedures and disease monitoring protocols. For vessel structure enhancement, practitioners frequently employ differential-based techniques, such as second derivative analysis and Hessian matrix operations. These differential analytical approaches significantly augment vessel visibility, rendering them instrumental in medical image segmentation and analytical processes. The methodology incorporates eigenvalue analysis of the Hessian matrix ($H(x, y)$) to identify vessel-like formations within images. The source image is denoted as $I(x, y)$. To establish a continuous Hessian matrix, convolution with a two-dimensional (2D) Gaussian filter is implemented. Equation (1.2.1) demonstrates the convolution between the Gaussian filter and the Hessian of image $I(x, y)$:

$$H(x, y) \approx G * \begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial y \partial x} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 G}{\partial x^2} & \frac{\partial^2 G}{\partial x \partial y} \\ \frac{\partial^2 G}{\partial y \partial x} & \frac{\partial^2 G}{\partial y^2} \end{bmatrix} * I(x, y) \quad (1.2.1)$$

The detection of local image structures is significantly enhanced through analysis of Hessian eigenvalue magnitudes and signs. Thus, eigenvalues and eigenvectors of $H(x, y)$ are computed to extract contrast and orientation data at each image point. This decomposition yields two eigenvalues, designated as (λ_1, λ_2) , with corresponding eigenvectors $(\bar{u}_1, \bar{u}_2)$. These elements facilitate the identification of various structural features including blobs, tubes, and plates. In the context of blood vessel detection, characteristic conditions include $|\lambda_1| \approx 0$, $|\lambda_1| \ll |\lambda_2|$, where $\bar{u}_1$ indicates vessel orientation. Furthermore, λ_2 typically exhibits negative values for bright vessels contrasted against darker backgrounds, a crucial criterion for maintaining consistency and excluding non-vessel structures. The Frangi vesselness function is consequently defined as:

$$F(v) = \begin{cases} 0 & \text{if } \lambda_2 > 0, \\ e^{-\frac{R_B^2}{2\beta^2}} \left(1 - e^{-\frac{s^2}{2c^2}}\right) & \text{otherwise,} \end{cases} \quad (1.2.2)$$

where

$$R_B = \frac{|\lambda_1|}{|\lambda_2|}, \quad S = \sqrt{\lambda_1^2 + \lambda_2^2} \quad (1.2.3)$$

The parameter R_B quantifies object eccentricity in 2D imagery, while S evaluates structural presence, effectively differentiating background regions. Background areas, characterized by minimal structural definition, exhibit low contrast and consequently low S values, whereas regions containing distinct structures generate high S values. The parameters β and c modulate the line filter's sensitivity to R_B and S . Conventionally, β is assigned a value of 0.5, while c is calibrated according to vessel gray-level intensities. The filter generates response values upon application, where elevated values correspond to vessel pixels and diminished values indicate background pixels.

1.2.2.2 Image Fusion

Medical image fusion serves as a critical methodology for combining multiple medical images, whether derived from single or multiple imaging modalities. This integration technique aims to maximize image quality while preserving and highlighting clinically relevant features, ultimately enhancing the diagnostic and analytical value of the resulting images. The field draws from multiple scientific disciplines, including image processing, computer vision, pattern recognition, machine learning, and artificial intelligence. Its widespread clinical applications enable healthcare practitioners to gain comprehensive understanding of pathological conditions through the integration of multi-modal imaging data, facilitating more precise diagnostic processes and treatment decisions. The intensity-hue-saturation (IHS) transformation represents a key element in IHS-based medical image fusion approaches. The process begins with converting source medical images from their original red-green-blue (RGB) format through RGB-to-IHS transformation. This conversion translates the initial image data into the IHS color space, which comprises intensity, hue, and saturation components, creating a matrix representation within this alternative color framework. The fusion process concludes with an inverse IHS transformation. In this methodology, medical images from different modalities first undergo conversion to the IHS color space through RGB-to-IHS transformation. The final fused image is then generated through the reverse process using IHS-to-RGB conversion:

$$I = \frac{R + B + G}{3} \quad (1.2.4)$$

$$\begin{cases} H = \frac{G-B}{3I-3B}, S = \frac{I-B}{I} & \text{if } B < R, G, \\ H = \frac{B-R}{3I-3R}, S = \frac{I-R}{I} & \text{if } R < B, G, \\ H = \frac{R-G}{3I-3G}, S = \frac{I-G}{I} & \text{if } G < R, B \end{cases} \quad (1.2.5)$$

$$\left\{ \begin{array}{ll} R = I(1 + 2S - 3S \times H), \\ G = I(1 - S + 3S \times H), \\ B = I(1 - S) & \text{if } B < R, G, \\ R = I(1 - S), \\ G = I(1 + 5S - 3S \times H), \\ B = I(1 - 4S + 3S \times H) & \text{if } R < B, G, \\ R = I(1 - 7S + 3S \times H), \\ G = I(1 - S), \\ B = I(1 + 8S - 3S \times H) & \text{if } G < R, B \end{array} \right. \quad (1.2.6)$$

This methodology capitalizes on the distinctive advantages of the IHS color space framework, enabling seamless integration of diverse medical imaging modalities into a unified visual format. Through the IHS transformation process, critical diagnostic features are preserved and enhanced while effectively merging complementary information from various imaging sources. Moreover, this sophisticated integration technique maintains the fundamental anatomical and pathological characteristics present in the source images while simultaneously improving their visibility and interpretability. The resulting fused images offer healthcare professionals enhanced visual information, significantly advancing their capacity for accurate clinical interpretation and thorough analysis. Meanwhile, the improved clarity and comprehensive nature of these fused visualizations facilitate more precise detection of subtle pathological changes and anatomical relationships. This enhanced diagnostic capability leads to more informed clinical decisions and treatment strategies. Additionally, this advanced fusion approach substantially streamlines clinical workflows by consolidating complex multi-modal imaging data into a single, easily interpretable format.

Algorithm 1 Vessel Enhancement and Image Fusion

Input: Medical image $I(x, y)$, Gaussian filter G , parameters β, c

Output: Enhanced and fused image I_{fused}

1: Compute Hessian matrix $H(x, y)$:

$$H(x, y) \leftarrow G * \begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial y \partial x} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix}$$

2: Perform eigenvalue decomposition:

$$(\lambda_1, \lambda_2) \leftarrow \text{Eig}(H(x, y))$$

3: Identify vessel-like structures:

$$\text{Condition: } |\lambda_1| \ll |\lambda_2|, \lambda_2 < 0$$

4: Compute vesselness function:

$$F(v) \leftarrow \begin{cases} 0 & \text{if } \lambda_2 > 0, \\ e^{-\frac{R_B^2}{2\beta^2}} \left(1 - e^{-\frac{S^2}{2c^2}} \right) & \text{otherwise} \end{cases}$$

5: Enhance image: $I_{\text{enhanced}} \leftarrow F(v)$

6: Set fusion inputs: $R \leftarrow I_{\text{enhanced}}, B, G \leftarrow \text{Original channels}$

7: Transform RGB to IHS: $I, H, S \leftarrow \text{RGB-to-IHS}(R, G, B)$

8: Enhance intensity channel I and adjust H, S

9: Transform IHS to RGB: $R, G, B \leftarrow \text{IHS-to-RGB}(I, H, S)$

10: Combine vessel-enhanced and fused images:

$$I_{\text{fused}} \leftarrow \text{Combine}(I_{\text{enhanced}}, \text{Fused GB})$$

11: **return** I_{fused}

1.2.3 Model descriptions

1.2.3.1 Region-Based Fully Convolutional Networks

Region-Based Fully Convolutional Networks (R-FCN) represent advanced deep learning architectures optimized for targeted object detection. These networks combine the strengths of Region-based Convolutional Neural Networks (R-CNN) and Fully Convolutional Networks (FCN) to achieve both efficient and accurate object detection. As detailed in [31], the R-FCN architecture incorporates four fundamental components: the Region Proposal Network (RPN), the residual network (ResNet), and modules for classification and regression, each contributing distinctly to the network’s effectiveness.

Following a methodology similar to Faster R-CNN, the Region Proposal Network (RPN) within R-FCN functions primarily to generate region proposals, identifying potential Regions of Interest (RoIs) in input images. This process enables convolutional operations across the entire image, facilitating weight layer computations. The R-FCN workflow begins with image resizing to maintain consistent short-side dimensions, followed by feature extraction using ResNet-101’s five convolutional blocks. The RPN receives its input specifically from ResNet-101’s fourth convolutional layer.

To address the high dimensionality in ResNet-101’s fifth convolutional layer output, additional convolutional layers are integrated to reduce channel count, generating a 1024-dimensional feature map. The network then implements parallel convolutional layers for classification and regression tasks. Position-sensitive score maps, dimensioned $k^2(C + 1)$ for classification and $4k^2$ for regression, incorporate crucial positional information. In the final phase, these score maps undergo pooling with the RPN-extracted Regions of Interest (RoIs), producing classification and regression results that demonstrate the network’s capability in object identification and localization within input images.

1.2.3.2 YOLO Series

1.2.3.2.1 Core Concept

The fundamental concept of the YOLO (You Only Look Once) series is to convert object detection into a regression task. This methodology involves using the whole image as input to the neural network, which subsequently forecasts the locations of the bounding boxes along with their associated classes.

1.2.3.2.2 Steps of YOLO Series Algorithms

1. Image Division: YOLO divides the input image into a fixed-size grid;
2. Predicting Bounding Boxes and Classes: In each grid cell, YOLO estimates a set number of bounding boxes (usually 5 or 3). Every bounding box is defined by five key attributes: the center coordinates, width, height, and the confidence score indicating the likelihood of an object being present. Furthermore, each bounding box also predicts the class of the detected object.
3. Single Forward Pass: YOLO uses a single forward pass through a CNN to predict all bounding box positions and classes simultaneously. This approach provides faster speed compared to other object detection algorithms, such as those based on sliding windows or region proposals, because it only requires one forward pass to complete the prediction;
4. Loss Function: YOLO utilizes a composite loss function to optimize its network training, which comprises location loss, confidence loss, and class loss. The location loss quantifies the discrepancy between the predicted bounding boxes and the actual ground truth bounding boxes. The confidence loss assesses the correctness of the object predictions and penalizes incorrect predictions for background areas. Finally, the class loss evaluates the accuracy of the object class predictions made by the network.
5. Non-Maximum Suppression (NMS): Among the predicted bounding boxes, multiple overlapping boxes may represent the same object. To eliminate re-

dundant boxes, YOLO uses the non-maximum suppression algorithm, which selects the best bounding box based on confidence and overlap.

1.2.3.2.3 Structure

The object detector architectures within the YOLO family typically comprise three primary components: Backbone, Neck, and Head. The Backbone component, commonly implemented as a pre-trained convolutional neural network, performs hierarchical feature extraction from input images. This process progressively identifies features from basic elements like edges and textures in early layers to more complex semantic content and object components in deeper network layers.

The Neck component functions as an intermediary structure between the Backbone and Head, performing feature refinement on the Backbone’s extracted features. Through additional convolutional layers, Feature Pyramid Networks (FPN), or similar mechanisms, it enhances both spatial and semantic information quality. The Head component, as the final architectural element, generates predictions using the refined features from preceding components. This component manages critical tasks including classification, localization, and instance segmentation, employing Non-Maximum Suppression in post-processing to eliminate redundant predictions and preserve only the most confident detections [32].

1.2.3.2.4 YOLOv8

YOLOv8 [33] represents an evolution of the YOLO architecture, incorporating innovative features and improvements to enhance both performance and versatility. This iteration is engineered for speed, accuracy, and user-friendliness, making it particularly effective for diverse tasks including object detection, image segmentation, and image classification.

Key architectural innovations include a redesigned backbone network, an Anchor-Free detection head, and a sophisticated loss function that integrates three distinct components: Varifocal Loss (VFL) for classification, and a combination of Complete-IoU (CIOU) and Deep Feature Loss (DFL) for regression. The archi-

ture maintains backward compatibility with previous YOLO versions, enabling seamless version transitions and comparative performance analysis. Beyond its primary object detection capabilities, YOLOv8’s versatility extends to advanced tasks such as instance segmentation and keypoint detection. This flexibility is supported by a robust community and ecosystem that ensures continuous development and support, enhancing its practical utility across diverse applications.

1.2.3.2.5 YOLOv9

YOLOv9 [34] introduces an innovative methodology called Programmable Gradient Information (PGI), which enhances deep networks’ capacity to address multiple objectives simultaneously. PGI facilitates the incorporation of comprehensive task-specific input data for objective function computation, ensuring precise gradient information for network weight optimization. The model also introduces the General Efficient Layer Aggregation Network (GELAN), a lightweight architecture developed through gradient path planning, demonstrating PGI’s effectiveness in streamlined model designs.

The model presents several significant contributions: First, it provides a theoretical analysis of existing deep neural network architectures through the lens of reversible functions, offering clarity to previously complex phenomena. This analysis culminated in the development of PGI and auxiliary reversible branches, yielding exceptional results. Second, PGI extends the applicability of deep supervision beyond very deep neural networks, enabling the practical implementation of new lightweight architectures in everyday applications. Third, GELAN achieves superior parameter utilization compared to contemporary deep convolutional designs while maintaining lightweight characteristics, speed, and accuracy, using only traditional convolutions.

1.2.3.2.6 YOLOv10

YOLOv10 [35], developed within the Ultralytics Python framework, presents significant improvements over its predecessors. Traditionally, YOLO methodologies employ Task Alignment Learning (TAL) during training, assigning multiple posi-

tive samples to each instance. While this one-to-many assignment approach generates rich supervisory signals and enhances optimization, it necessitates NMS post-processing, potentially compromising inference efficiency in deployment scenarios. Although previous research has explored one-to-one matching to reduce redundant predictions, such approaches often result in either increased inference overhead or suboptimal performance.

The researchers introduce a novel YOLO training strategy that eliminates the need for NMS by implementing dual label assignment and consistent matching metrics, achieving both high efficiency and competitive performance. In contrast to one-to-many assignment, one-to-one matching assigns a single prediction to each ground truth, eliminating the need for NMS post-processing. However, this approach typically results in weaker supervision, potentially compromising accuracy and convergence speed. To address these limitations, the authors developed a dual label assignment approach for YOLO, combining the advantages of both strategies. The model’s components undergo comprehensive optimization for both efficiency and accuracy, substantially reducing computational demands while enhancing capabilities. Extensive experimental validation demonstrates superior accuracy-latency trade-offs across various model scales.

1.2.3.3 Real-Time Detection Transformer

The Real-Time Detection Transformer (RT-DETR) [36] represents an innovative object detection model optimized for real-time applications. This architecture combines two significant approaches in object detection: Transformer and Detection Transformer (DETR). While Transformers have demonstrated exceptional capabilities in sequence modeling for computer vision tasks, DETR transforms object detection into an object query problem utilizing the Transformer architecture. RT-DETR enhances DETR’s foundational structure to enable efficient real-time object detection.

DETR [37], developed by Facebook AI researchers, integrates Transformer-based

encoding and decoding in an end-to-end object detection framework. Unlike traditional object detection methods, DETR conceptualizes each image pixel position as a query vector, which the Transformer encoder converts into feature vectors. The decoder then processes these vectors to predict object categories, bounding box coordinates, and feature vectors for detected objects through query vector interactions.

While maintaining DETR’s basic encoder-decoder architecture, RT-DETR introduces several crucial optimizations. These include the use of reduced feature maps to minimize computational costs, fewer attention heads to decrease model parameters, and an innovative grouped attention mechanism to enhance performance. The RT-DETR architecture specifically employs a modified ResNet50 network in its encoder, utilizing only the initial four residual blocks to reduce feature map dimensions. Its decoder combines a Transformer decoder with an object embedding network. The model implements just two attention heads, compared to DETR’s eight, and introduces a grouped attention mechanism that partitions attention calculations to improve computational efficiency. The object embedding network integrates object feature vectors for subsequent processing tasks.

1.2.4 Evaluation Indicators

1.2.4.1 $AP_{50}(\%)$

$AP_{50}(\%)$ represents a distinct form of the Average Precision (AP) metric, defined by setting the Intersection over Union (IoU) threshold to 0.50. This metric encapsulates various sub-metrics to provide a comprehensive evaluation:

- Intersection over Union: This metric evaluates the extent of overlap between the predicted bounding boxes and the actual bounding boxes. This metric offers a measure of the model’s accuracy in localization. IoU is computed by dividing the area of intersection between the predicted and ground truth boxes by the area of their union.
- IoU Threshold: In order for a prediction to be classified as a True Positive

(TP), the Intersection over Union between the predicted bounding box and the actual ground truth box must surpass a predefined threshold. For the $AP_{50}(\%)$ metric, this threshold is established at 0.50.

- Precision: This metric indicates the precision of the model’s positive predictions. It is determined by dividing the number of true positive outcomes by the total number of true positive and false positive outcomes.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1.2.7)$$

- Recall: This metric evaluates the model’s effectiveness in recognizing all pertinent instances. It is calculated by dividing the number of true positives by the combined total of true positives and false negatives.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (1.2.8)$$

- Average Precision: The Average Precision metric encapsulates the precision-recall relationship by calculating the area under the precision-recall curve. This metric signifies the model’s overall precision when considering all levels of recall.

$$AP = \int_0^1 \text{Precision}(\text{Recall}) d(\text{Recall}) \quad (1.2.9)$$

1.2.4.2 $AP_{\text{val}}(\%)$

The metric $AP_{\text{val}}(\%)$ offers a more thorough assessment of a model’s performance by evaluating both precision and recall across various IoU thresholds, rather than focusing on just one. This measure generally represents the Average Precision computed over a range of IoU thresholds, typically from 0.50 to 0.95, in increments (e.g., every 0.05). In some benchmarks, this is referred to as mAP (mean Average Precision).

$$AP_{\text{val}}(\%) = \frac{1}{n} \sum_{k=1}^{k=n} AP_k \quad (1.2.10)$$

Here, n denotes the quantity of IoU thresholds evaluated, while AP_k signifies the mean precision for class k .

1.2.4.3 Parameters (M)

The measurement of learnable weights in neural network architectures, expressed in millions (M), represents a key metric for evaluating model complexity. This parameter count serves as a quantitative measure of the network’s capacity to encode information and recognize patterns. While higher parameter counts generally enable more sophisticated pattern modeling capabilities, they also demand greater computational and memory resources. Implementing too many parameters can lead to overfitting issues and reduced generalization effectiveness, highlighting the need for careful balance between architectural complexity and computational resource usage.

1.2.4.4 FLOPs(G)

Neural architecture computational requirements are primarily measured through floating point operations (FLOPs), quantified in gigaflops (G), with each gigaflop representing one billion individual operations. This metric captures the arithmetic computational load required for forward propagation during inference. FLOPs provide a meaningful measure of operational costs in model deployment, particularly crucial in environments with limited resources where minimizing computational demands is beneficial. However, efforts to reduce computational complexity must be carefully balanced, as overly aggressive optimization can adversely affect predictive accuracy and overall model performance.

1.2.4.5 Latency(ms)

Latency represents the time duration between input reception and response generation, measured in milliseconds (ms). This performance metric is particularly crucial in applications requiring real-time processing capabilities, such as medical diagnostic systems, autonomous operations, and interactive technologies. The speed of response is influenced by multiple factors, including architectural design choices, hardware specifications, and software optimization techniques. Minimizing latency is essential for enhancing user experience and enabling effective decision-making in

time-critical scenarios where delays can significantly impact operational outcomes.

1.3 Experiments and Results

1.3.1 Experiments

1.3.1.1 Data Acquisition

To address the variability in contrast agent distribution during different angiography phases, which affects vascular image clarity, we carefully selected 400 representative images from our dataset for model development. These images were specifically chosen from the mid-phase of coronary angiography to ensure consistent contrast agent distribution within the vessels. The annotation process for stenotic regions was conducted using MATLAB R2022a following a systematic approach. Expert radiologists from Hokkaido University Hospital provided patient diagnoses, which formed the basis for image annotations. We prioritized regions showing stenosis rates exceeding 50%, considering their higher clinical significance and improved visibility. The dataset was organized according to coronary angiography imaging angles, comprising 280 images of the left coronary artery and 120 images of the right coronary artery. These images were further subdivided based on zero, left anterior oblique, and right anterior oblique angles. A comprehensive breakdown of the data distribution is provided in Table 1.1.

1.3.1.2 Data Preparation

Our methodology began with applying the proposed preprocessing technique to enhance input image quality across the dataset. The enhanced images were then randomly partitioned into three distinct sets: 70% allocated for training, 15% for validation, and the remaining 15% for testing purposes. To address the constraints of limited data availability and enhance model robustness, we implemented several augmentation strategies. The primary augmentation technique involved rotating

Table 1.1: Data Distribution in CAG Dataset.

Imaging Angles	Numbers	Training	Validation	Testing
LCA zero	109	76	16	17
LCA LAO	96	67	14	15
LCA RAO	75	53	11	11
RCA zero	30	21	5	4
RCA LAO	51	36	8	7
RCA RAO	39	27	6	6
SUM	400	280	60	60

images in 5-degree increments within a -45° to 45° range. We supplemented this with additional augmentation methods, including horizontal and vertical flips, to further expand dataset variety. These augmentation techniques were designed to simulate diverse clinical scenarios, enhancing the model’s capacity to generalize effectively to novel data. Through these augmentation procedures, we successfully expanded our training dataset to 7320 images, substantially increasing its diversity and generalization potential. Additionally, all images were standardized to uniform dimensions to ensure consistent model input during the training phase. The entire preprocessing and augmentation workflow was executed using Matlab2022a.

1.3.1.3 Model Training and Evaluation

We employed multiple deep learning frameworks, R-FCN, YOLOv8, YOLOv9, YOLOv10, and RT-DETR to develop models using our enhanced dataset, evaluating their effectiveness in identifying stenotic features within vascular structures. To ensure result integrity, all models were trained from scratch without pre-trained weights. The data structure was adapted according to each repository’s requirements, and model performance was evaluated under consistent conditions for both preprocessed and unprocessed images.

Our quantitative performance evaluation focused on the models’ ability to ac-

curately detect and delineate stenotic regions in coronary angiography images. The evaluation utilized metrics including $AP_{val}(\%)$ and $AP_{50}(\%)$ to measure the models' proficiency in recognizing and representing stenotic areas. We also assessed each model's real-time performance capabilities by measuring latency using original coronary stenosis detection videos.

To establish robust and unbiased evaluation results, we implemented a comprehensive validation methodology. This involved training each model multiple times with different random initializations to account for training outcome variability. The final performance metrics were calculated by averaging results across these multiple training iterations, ensuring reliable and consistent performance assessment.

1.3.2 Results

We performed a comparative investigation to evaluate how HFV filter preprocessing and image fusion affected deep learning models' stenosis detection capabilities compared to using unprocessed images. The study also examined the optimal model's performance across different imaging angles. The comparative methodology involved applying the HFV filter for image preprocessing to enhance fine vascular structure visibility through high-frequency component emphasis. We then combined these filtered images with the original images through fusion, creating composite inputs that leveraged the advantages of both preprocessing approaches. This evaluation aimed to determine whether the fusion methodology could enhance stenotic region detection compared to using unprocessed source images.

Following the preprocessing phase, we carefully annotated stenosis locations in both main and branch vessels, particularly focusing on regions where stenosis rates exceeded 50%, guided by expert medical diagnostics. For comprehensive visualization of our findings, we included Figure 1.1, which systematically presents different aspects of our investigation. Figure 1.1(a1), (b1), (c1), (d1), (e1), and (f1) show the original, unprocessed images as baseline references. Figure 1.1(a2), (b2), (c2), (d2), (e2), and (f2) demonstrate the results after HFV filter preprocessing, while Fig-

ure 1.1(a3), (b3), (c3), (d3), (e3), and (f3) display the outcomes of the image fusion process, effectively illustrating the progression and success of our image processing techniques.

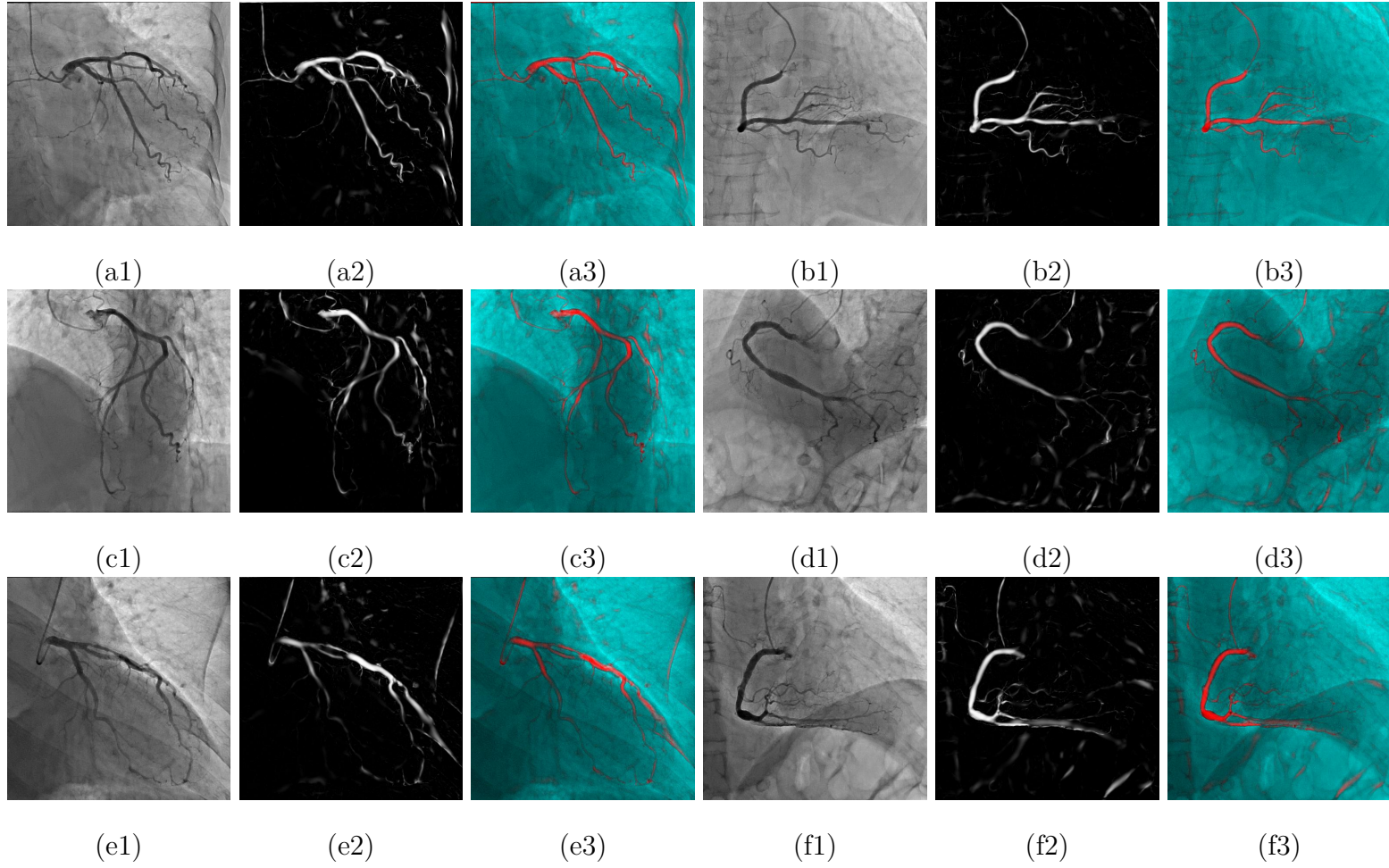


Figure 1.1: The performance of the proposed preprocessing method applied to six imaging angles. Specifically, (a1--f1) portray the original images from the dataset for the conditions of Left Coronary Artery (LCA) zero, LCA Left Anterior Oblique (LAO), LCA Right Anterior Oblique (RAO), Right Coronary Artery (RCA) zero, RCA LAO, and RCA RAO, respectively. The images (a2--f2) show the results after applying the HFV filter to these original images. Finally, (a3--f3) illustrate the output images following the application of the proposed preprocessing method. Moreover, (a1--f1) belong to the initial image dataset, while (a3--f3) will be classified as the preprocessed image dataset for training.

The HFV filter application notably enhances vascular feature visibility, particularly in primary vessels, while reducing background noise. As illustrated in Figure 1.1(a2), the filter produces clearer tubular structure visualization. However, it also generates pseudo-vessels that closely resemble branch vessel structures throughout the image. These artificial structures appear in areas where the original image shows either no vessels or poorly defined vessels.

Figure 1.1(b2) further demonstrates this effect, showing enhanced prominent features but diminished small vessels in low-contrast regions. To optimize vascular feature enhancement while preserving fine vessel details, we implemented an image fusion approach. Figure 1.1(a3), (b3), (c3), (d3), (e3), and (f3) demonstrate the results of applying this fusion technique to processed images from six different imaging angles. This methodology successfully preserved essential features from the original images while substantially reducing noise and random morphological features that might appear as vessel-like structures due to HFV filter application. This enhancement provided our deep learning models with higher quality input images, improving their capacity to accurately detect and outline stenotic regions, thereby enabling more reliable coronary artery analysis. Our testing revealed that the pre-processing module alone requires an average execution time of 37.1 ms, a factor that must be considered when evaluating overall system performance in real-time applications.

To establish a performance baseline for subsequent experiments, we conducted a comprehensive evaluation of various neural network architectures using unprocessed images. Our assessment included advanced models: R-FCN-InceptionResNetV2, YOLOv8-X, YOLOv9-E, YOLOv10-L, YOLOv10-X, RT-DETR R50, and RT-DETR R101. We evaluated each model’s effectiveness by measuring the Average Precision at an Intersection over Union threshold of 0.50, maintaining consistent training and testing conditions across all models.

The experimental results revealed that R-FCN-InceptionResNetV2, which was the top-performing network in our previous research [18], achieved an $AP_{50}(\%)$ of

only 47.6 on the raw dataset. In comparison, the newer architectures demonstrated significantly better performance: YOLOv8-X, YOLOv9-E, YOLOv10-L, YOLOv10-X, RT-DETR R50, and RT-DETR R101 achieved $AP_{50}(\%)$ values of 77.3, 82.1, 78.2, 82.9, 74.6, and 82.7, respectively. These results are comprehensively presented in Table 1.2.

This baseline evaluation serves as a crucial reference point for our subsequent experiments involving image preprocessing and augmentation techniques. The findings indicate that while older architectures like R-FCN-InceptionResNetV2 may have limitations with complex data, contemporary models exhibit robust performance that could potentially be further enhanced through appropriate preprocessing methods.

Table 1.2: Performance Comparison of Different Models on Raw CAG Dataset.

Model	$AP_{val}(\%)$	$AP_{50}(\%)$
R-FCN-InceptionResNetV2	29.3	47.6
RT-DETR R50	45.7	74.6
YOLOv8-X	46.2	77.3
YOLOv10-L	47.3	78.2
YOLOv9-E	50.3	82.1
RT-DETR R101	50.5	82.7
YOLOv10-X	51.6	82.9

Following our experiments with the raw dataset, we applied the same neural networks (excluding the underperforming R-FCN-InceptionResNetV2) to the pre-processed dataset to evaluate their detection performance under identical conditions. We conducted additional latency testing using preprocessed videos and documented the networks' training process data. All results from these experiments are compiled in Table 1.3.

Table 1.3: Performance Comparison of Different Models on Preprocessed CAG Dataset.

Model	Parameters (M)	FLOPs(G)	AP _{val} (%)	AP ₅₀ (%)	Latency (ms)
RT-DETR R50	13.8	40.0	45.9	79.2	3.02
YOLOv10-L	10.6	50.3	49.4	82.4	3.16
YOLOv8-X	28.1	112.4	50.5	84.2	5.54
YOLOv9-E	23.6	89.4	52.5	86.0	5.11
RT-DETR R101	25.1	98.6	53.8	86.7	4.51
YOLOv10-X	12.2	65.3	53.9	87.1	4.43

Among the six networks evaluated, RT-DETR R50 shows the lowest performance, despite demonstrating improvement over previous results, achieving an AP of 0.792 at an IoU threshold of 0.50. YOLOv10-X maintains its position as the superior model, demonstrating an AP of 0.871 at an IoU of 0.50, and an $AP_{val}(\%)$ of 53.9 when averaging AP values across different IoU thresholds.

A horizontal comparison across networks reveals progressive improvements in detection accuracy with successive YOLO versions. The transformer-based RT-DETR-R101 achieves detection accuracy comparable to YOLOv10-X. In terms of processing speed, YOLOv10-X demonstrates superior performance with a latency of 4.43ms, compared to RT-DETR-R101’s 4.51ms. These results underscore the enhanced detection performance achieved through our preprocessing methods. YOLOv10-X’s consistent leadership in both accuracy and efficiency emphasizes its suitability for real-time coronary artery stenosis detection. Its slight advantage in latency over RT-DETR-R101 further establishes its superiority in time-sensitive applications.

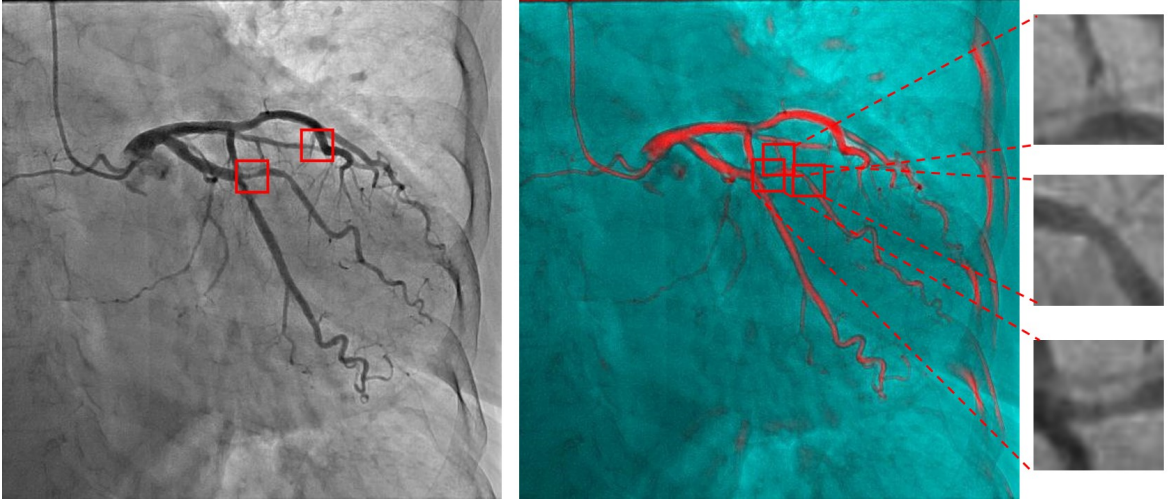


Figure 1.2: Comparison of YOLOv10-X detection results on original image and preprocessed image.

Figure 1.2 presents a comparative analysis of LCA zero angle images, contrasting YOLOv10-X’s detection performance on both original and preprocessed images. The left image shows detection results using the original image, while the right image

demonstrates results using the preprocessed image. The detection boxes have been enlarged to facilitate clearer observation of the detection outcomes. The comparison reveals that the preprocessed image enables more precise and distinct detection of stenotic regions, with the enlarged detection boxes emphasizing YOLOv10-X’s enhanced stenosis identification capabilities when applied to preprocessed images.

To provide a more concrete demonstration of our preprocessing approach’s advantages in terms of processing performance, we have compiled both latency and accuracy metrics for preprocessed and raw images in Table 1.4, including calculations of improvement margins.

Table 1.4: Comparison of Latency and Accuracy with and without Preprocessing Method.

Model	$AP_{50}(\%)$			Latency(ms)		
	Original	Preprocessed	Improvement Margin(%)	Original	Preprocessed	Improvement Margin(%)
RT-DETR R50	74.6	79.2	6.17	2.73	3.02	-9.60
YOLOv8-X	77.3	84.2	8.95	5.05	5.54	-8.84
YOLOv10-L	78.2	82.4	5.37	2.77	3.16	-12.21
YOLOv9-E	82.1	86.0	4.79	4.58	5.11	-10.33
RT-DETR R101	82.7	86.7	4.84	4.04	4.51	-10.47
YOLOv10-X	82.9	87.1	5.07	3.90	4.43	-11.91

The results in Table 1.4 demonstrate that while our preprocessing method enhanced stenotic region detection accuracy, it introduced additional processing time. The accuracy improvements varied across models, with YOLOv8-X showing the highest improvement at 8.95%, while YOLOv9-E demonstrated the smallest gain at 4.79%. In terms of processing speed, YOLOv10 models experienced the most significant impact, with both YOLOv10-L and YOLOv10-X showing approximately 12% increases in processing time.

The increased latency can be attributed to two main factors. First, the transformation from original to preprocessed images requires additional computational time, particularly due to the iterative Hessian eigenvalue calculations needed by the HFV filter across various parameter settings and the image fusion process. Second, the enhanced complexity of the preprocessed images, resulting from the fusion of original and enhanced images, demands additional processing time from the deep learning networks.

Despite these processing delays, the method’s significant improvement in detection accuracy suggests its potential to effectively balance the trade-off between real-time processing capabilities and diagnostic precision, a persistent challenge in clinical applications. The application of this preprocessing technique to optimized networks like RT-DETR R101 and YOLOv10-X, which have already undergone substantial computational optimization, shows promise for real-time diagnostic applications. These findings suggest that with further optimization, our method could be successfully integrated into clinical workflows, contributing to more accurate and timely coronary artery disease diagnosis.

Additionally, to thoroughly assess our the efficacy of our approach from various imaging perspectives and to determine the optimal angle for analyzing coronary artery stenosis, we performed an in-depth analysis for each angle within the test dataset. This comprehensive review included segmenting the dataset and evaluating the performance of the top-performing YOLOv10-X model. Specifically, we measured three performance metrics: $AP_{50}(\%)$ for the raw images, and both $AP_{50}(\%)$

and $AP_{val}(\%)$ for the preprocessed images. The detailed results of these assessments are presented in Table 1.5:

Table 1.5: Performance of YOLOv10-X Model on Different Imaging Angles.

Imaging Angle	$AP_{val}(\%)$		$AP_{50}(\%)$
	Preprocessed	Original	Preprocessed
LCA Zero	55.8	85.0	90.5
LCA LAO	55.8	83.3	87.8
LCA RAO	54.0	83.9	88.3
RCA Zero	51.3	81.1	85.2
RCA LAO	49.2	78.8	80.1
RCA RAO	51.7	80.7	84.6

The analysis demonstrates YOLOv10-X’s superior performance on preprocessed images compared to raw images across all imaging angles. The LCA zero imaging angle exhibited particularly noteworthy results with preprocessed data, achieving the highest metrics with an $AP_{50}(\%)$ of 90.5 and an $AP_{val}(\%)$ of 55.8. These findings suggest that the LCA zero angle may provide optimal conditions for analyzing coronary artery stenosis.

1.4 Discussion

Our research explored the benefits of combining a general HFV filter for vascular contrast enhancement with image fusion to improve the accuracy of deep learning-based stenosis detection in coronary angiography images. The study investigated the impact of image preprocessing on processing latency and evaluated how different imaging angles affect coronary artery stenosis detection efficiency. The investigation involved training and evaluating seven deep learning models using both preprocessed and raw images. Advanced models like YOLOv10 and RT-DETR demonstrated

superior detection performance. The method’s robustness was enhanced through increased annotations and more complex image processing. Our approach extended beyond static image analysis to include real-time detection testing, while the detailed examination of various imaging angles provided valuable clinical insights. The results indicate that preprocessing methods incorporating HFV filtering and image fusion can enhance deep learning models’ precision in analyzing complex coronary angiography images while maintaining reasonable training complexity. However, it’s important to note that HFV-filtered images may generate pseudo-vascular structures or inadequately enhance actual vessels, potentially leading to misclassifications and false positives during both training and detection phases. These findings underscore the importance of carefully evaluating the optimal scenarios for HFV filter preprocessing application. Critical assessment is needed to determine whether enhanced visual quality comes at the cost of image accuracy, establishing a foundation for future image processing research.

The training performance of the model was significantly influenced by the quantity of relevant features present in the images. We employed image fusion with HFV filtering to enhance stenosis feature identification in coronary angiography images, particularly focusing on improving the visibility of stenosis in smaller vessels. This approach specifically addressed feature loss following vascular enhancement and minimized edge enhancement issues in small features caused by filtering. Experimental results demonstrate that this methodology substantially improves branch vessel stenosis detection in complex coronary angiography images. The fusion technique successfully preserves essential vessel characteristics while reducing HFV filtering-induced noise and artifacts.

Our investigation also examined the technique’s impact on feature learning capabilities. We conducted thorough assessments of the $AP_{50}(\%)$ improvements achieved through our method, while addressing critical dataset considerations including class imbalance and dataset size effects. Results confirmed superior neural network performance on preprocessed images compared to raw images. Furthermore, our research

included a comprehensive analysis of how different imaging angles affect coronary artery stenosis detection accuracy. The experimental findings revealed a significant correlation between imaging angle and detection precision. Notably, Left Coronary Artery imaging angles demonstrated higher detection accuracy compared to Right Coronary Artery angles, with peak $AP_{50}(\%)$ values observed at the LCA Zero angle. This marked performance disparity between arterial segments suggests a promising direction for future research. It indicates that future efforts to enhance coronary artery stenosis detection efficiency should consider the advantages offered by specific imaging angles, particularly given the superior accuracy demonstrated in LCA imaging.

From a clinical perspective, the enhanced detection of branch vessel stenosis carries substantial implications for patient care and procedural planning. The improved visualization achieved through our preprocessing method offers interventional cardiologists better insight into complex vascular anatomy, potentially reducing procedure times and contrast agent usage during catheterizations. This advancement is particularly valuable when working with elderly patients who have existing surgical stents, as this growing demographic often presents complex cases requiring careful intervention planning. The clearer visualization of stenotic regions could lead to more precise interventional strategies and better-informed decisions about stent placement and sizing. Furthermore, the method’s ability to handle cases with surgical artifacts makes it particularly relevant for real-world clinical scenarios where previous interventions often complicate image interpretation.

Our research in this chapter encountered several significant limitations that warrant discussion, each providing important context for understanding the constraints of our analysis.

Firstly, while our dataset incorporates typical coronary angiography angles to mirror clinical scenarios, it presents specific challenges. The complex backgrounds and subtle stenosis features in these images limit deep learning model accuracy, highlighting the need for further research to enhance model robustness in real-world

conditions.

Secondly, the process of extracting key frames from coronary angiography video streams faces inherent challenges, particularly regarding branch vessels. The predominant elderly patient demographic often presents surgical artifacts that obscure vessels, while coronary blood flow dynamics complicate comprehensive capture of the aorta and its branches, resulting in dataset size limitations.

Thirdly, despite careful attention to branch vessel stenosis annotation, many instances remain unidentified and unmarked, limiting the model’s detection capabilities in these areas. This emphasizes the need for future research focused on improving branch vessel stenosis identification.

Fourthly, dataset distribution imbalance presents a significant challenge, primarily due to overlapping blood vessels in RCA imaging angles complicating stenosis location marking. Future work aims to address this by expanding the dataset with additional RCA images.

Fifthly, the preprocessing stage did not specifically account for image attribute variations such as brightness, potentially affecting experimental outcomes. This factor will be considered in future experiments.

These limitations underscore the need for continued research to overcome these challenges and enhance methodology effectiveness. Future developments should focus on advanced image processing techniques and larger, more diverse datasets to establish a more reliable clinical system for coronary artery stenosis detection.

1.5 Conclusions

Overall, the implementation of the HFV filter in coronary angiography has notably enhanced the clarity of images, emphasizing vascular details while minimizing noise in complicated imaging contexts. The image fusion technique has compensated for the feature loss encountered during the vascular enhancement phase. Experimental outcomes revealed that compared with the unprocessed images, training deep learning models on fusion images can generate about 5% accuracy improvement under

the premise of less speed loss. Looking forward, it is crucial to increase both the volume and diversity of the dataset, particularly addressing any potential imbalances. Additionally, further development and refinement of existing networks are necessary to improve their ability to recognize stenotic features in complex coronary angiography images.

Chapter 2

YOLO-Based Attention-Enhanced Network for Coronary Artery Stenosis Localization

2.1 Introduction

Coronary artery disease remains a leading cause of mortality worldwide, making accurate and efficient diagnosis of arterial stenosis a critical area of research in medical imaging. Advances in deep learning have significantly improved the capabilities of medical image analysis, with convolutional neural networks and transformer models being widely adopted for tasks like segmentation and object detection. However, challenges such as computational efficiency, redundancy in image processing, and the balance between model accuracy and processing speed persist. These limitations necessitate the development of novel methods that address both performance and resource constraints while maintaining high accuracy in medical imaging applications.

Among the state-of-the-art models, the YOLO series has demonstrated exceptional performance in object detection and image segmentation tasks. YOLO models are known for their speed and efficiency, achieved by processing the entire image in

a single pass through the network. The recent evolution in the YOLO series, culminating in Yolov10, has introduced further enhancements, including the displacement of Non-Maximum Suppression and an advanced matching mechanism for efficient region identification. These features make Yolov10 a suitable candidate for modifications targeting specific use cases such as coronary artery segmentation, where precise delineation of stenosis location is critical.

Attention mechanisms have also become a cornerstone of modern deep learning architectures, enabling models to focus on relevant features while minimizing irrelevant noise. Traditional attention mechanisms, such as Softmax attention[38], offer powerful global context modeling but are computationally expensive. The recent development of Agent Attention[39] provides a promising alternative, blending the strengths of Softmax attention and linear attention for a more efficient yet expressive approach. This mechanism is particularly relevant for tasks requiring both detailed local feature extraction and global context modeling, such as the segmentation of coronary arteries in angiographic images. The Efficient Multi-scale Convolutional Attention Module (EMCAM)[40] is a component of the EMCAD framework, designed to enhance medical image segmentation by improving both performance and computational efficiency. EMCAM comprises three key sub-modules: the Multi-scale Convolution Block (MSCB), which captures features across various resolutions using depth-wise convolutions; the Channel Attention Block (CAB), which assigns varying levels of importance to each channel to emphasize relevant features; and the Spatial Attention Block (SAB), which focuses on specific spatial regions within the input to enhance pertinent features. By integrating these sub-modules, EMCAM effectively captures intricate spatial relationships and concentrates on salient regions within medical images, leading to improved segmentation outcomes. Notably, EMCAD achieves state-of-the-art performance with significant reductions in parameters and floating-point operations, making it suitable for scenarios with limited computational resources. Parallelized Patch-Aware Attention (PPA)[41] is a lightweight attention mechanism designed to enhance object detection and semantic

segmentation tasks. By employing a multi-branch feature extraction strategy, PPA captures information at various scales and levels, effectively improving the performance of these tasks. In the context of infrared small object detection, PPA has been integrated into the Hierarchical Context Fusion Network (HCF-Net). Within HCF-Net, the PPA module utilizes a multi-branch feature extraction approach to capture features at different scales and levels, thereby addressing challenges associated with detecting small objects in complex backgrounds. Additionally, PPA has been applied in person re-identification through the Attention-Aware Compositional Network (AACN). In this framework, the Pose-guided Part Attention component is learned and applied to mask out undesirable background features in pedestrian feature maps, enhancing the accuracy of person re-identification. These advanced attention mechanisms have enabled significant progress in handling complex tasks, particularly those requiring a balance between computational efficiency and global context modeling in deep learning applications. Moreover, these mechanisms have opened new possibilities for enhancing model capabilities across diverse applications. The Partial Self-Attention module (PSA)[35], originally integrated into YOLO, was designed to balance computational efficiency and global representation learning. By selectively applying multi-head self-attention (MHSA) to a portion of the input features, PSA reduces the computational and memory overhead associated with traditional attention mechanisms. However, its selective focus on only part of the feature set may limit its ability to fully capture the complex spatial relationships required for detecting stenosis, especially in the intricate and often subtle features present in CAG images. Moreover, PSA lacks mechanisms to explicitly address feature diversity, which is essential for robust object detection in medical imaging.

To address these limitations, we replaced PSA with the Agent Attention module. This module combines the strengths of Softmax and linear attention mechanisms, providing both high expressive capacity and computational efficiency. Agent Attention introduces two critical enhancements: agent bias and a diversity restoration module. The agent bias, inspired by relative position bias (RPB), augments the at-

tention calculation by integrating spatial information, enabling the model to better capture the positional relationships among features. This is particularly important in CAG images, where precise localization of stenosis depends on understanding the spatial arrangement of vascular structures. The diversity restoration module employs depthwise convolution to preserve feature diversity, ensuring that the attention mechanism maintains a rich representation of the input features, which is critical for identifying subtle signs of stenosis. The integration of Agent Attention was carefully designed to align with the computational constraints of YOLO. Similar to PSA, the module was applied at the low-resolution stage of the model, where the computational cost is less intensive. This placement allowed the model to benefit from the enhanced attention mechanism without significantly increasing its overall computational complexity. The linear nature of Agent Attention ensured that the modification remained efficient and scalable. Experimental results demonstrated an improvement in the detection accuracy of stenosis locations when using the modified YOLO model. The enhanced ability of the Agent Attention module to model global contexts and capture diverse spatial features proved particularly effective in this application.

In this chapter, we developed an enhanced version of YOLOv10, incorporating strategic modifications to address existing challenges. Our study presents two primary contributions:

1. We demonstrated that substituting PSA with Agent Attention in YOLOv10 improves model performance on coronary angiography images.
2. We incorporated an additional feedforward pathway to retain shallow feature information, ensuring that both low-level and high-level features contribute effectively to detection accuracy.

2.2 Methods

2.2.1 Agent Attention mechanism

Agent Attention introduces a novel mechanism that leverages agent tokens $A \in \mathbb{R}^{n \times C}$, where $n \ll N$. The process is divided into two stages: *Agent Aggregation* and *Agent Broadcast*.

The attention output O^A can be expressed as:

$$O^A = \text{Attn}^S(Q, A, \text{Attn}^S(A, K, V)), \quad (2.2.1)$$

where:

- $\text{Attn}^S(A, K, V)$ performs agent aggregation by computing attention between A and K , weighted by V ,
- $\text{Attn}^S(Q, A, \cdot)$ performs agent broadcast by computing attention between Q and A based on the aggregated representation.

Expanding the equation, O^A can be equivalently written as:

$$O^A = \sigma(QA^\top)\sigma(AK^\top)V, \quad (2.2.2)$$

where $\sigma(\cdot)$ is the softmax function applied to the attention scores. This is further simplified as:

$$O^A = \phi_q(Q)\phi_k(K)^\top V, \quad (2.2.3)$$

where $\phi_q(\cdot)$ and $\phi_k(\cdot)$ are kernel functions applied to Q and K , respectively.

The generalized form of Agent Attention is:

$$O^A = \text{Attn}^{\phi_q/\phi_k}(Q, K, V), \quad (2.2.4)$$

where ϕ_q/ϕ_k represent the feature maps for query and key, and A serves as the intermediary for reducing computational complexity.

By introducing the agent tokens A , the dimensionality of the attention mechanism is effectively reduced, allowing for significant computational savings while maintaining robust performance.

2.2.2 Proposed method

The proposed model builds upon the architecture of YOLOv10, specifically incorporating modifications aimed at improving efficiency and accuracy in coronary artery stenosis detection. In our model, we retained the structure of YOLOv10 that excludes the use of Non-Maximum Suppression and instead introduces an advanced matching mechanism earlier in the process to enhance image processing efficiency. This adjustment allows the model to identify and match regions of interest more effectively, reducing redundancy and improving computational speed.

In addition, we introduced Agent Attention mechanism just before the head of the model, after feature extraction. This attention mechanism is an advanced mechanism designed to enhance transformer models by balancing computational efficiency and representational power. Traditional attention methods, like Softmax attention, excel in expressiveness but are computationally expensive, especially for global context modeling. Agent Attention introduces agent tokens into the attention framework, represented as a quadruple (Q: Query tokens, A: Agent tokens, K: Key tokens, V: Value tokens). The process occurs in two stages: aggregation, where agent tokens collect information from key and value tokens, and broadcasting, where the aggregated data is shared back with the query tokens. By using fewer agent tokens than query tokens, this method achieves higher efficiency while maintaining global context modeling capabilities. Moreover, it generalizes linear attention, blending the strengths of Softmax and linear approaches for improved performance.

Moreover, we integrated a feedforward path to compensate for the loss of shallow feature information during training. This modification was introduced because we observed that, as the model learns, deeper layers tend to receive more effective and accurate weight updates compared to shallower layers, primarily due to the feedforward nature of deep learning models. By including this additional feedforward path, we aimed to preserve the valuable shallow layer information, ensuring that the model retains both low-level and high-level features for improved detection accuracy.

The specific model structure is illustrated in Figure 2.1. It consists of four key

components: the Head, Neck, Backbone, and an Auxiliary module, each playing a distinct role in the network’s functionality and performance. The Backbone serves as the core feature extraction module of the network. It processes the input image, systematically extracting features at multiple levels to capture both spatial and semantic characteristics. These features are crucial for understanding the underlying patterns within the image, enabling the detection of intricate details such as stenosis in coronary arteries. The Backbone’s hierarchical structure ensures that both low-level and high-level information is effectively captured. The Backbone’s hierarchical structure ensures that both low-level and high-level information is effectively captured. Positioned as the intermediate part of the network, the Neck bridges the Backbone and the Head. Its primary function is to fuse and enhance the features extracted by the Backbone, effectively preparing them for the detection tasks handled by the Head. By combining features from various layers of the Backbone, the Neck captures both fine-grained details and high-level semantic information. Additionally, it incorporates multi-scale feature integration, which is critical for handling objects of varying sizes and complexities, ensuring robust performance across different data distributions. The Head is the final component of the network, directly responsible for the tasks of object detection and classification. It outputs essential information such as bounding boxes, class predictions, and confidence scores for the detected objects. Specifically, the Head predicts the coordinates of the bounding boxes and assigns a class label to each detected object, ensuring accurate localization and identification of relevant features. In addition, inspired by the architecture of Yolov9, the Auxiliary module adopts a forward-feedback approach aimed at addressing the loss of shallow feature information during training. This module plays a pivotal role in preserving and enhancing shallow features, which are often critical for detecting fine structures. By compensating for this loss, the Auxiliary module significantly boosts the network’s capacity to detect stenosis in coronary arteries, improving overall accuracy and reliability.

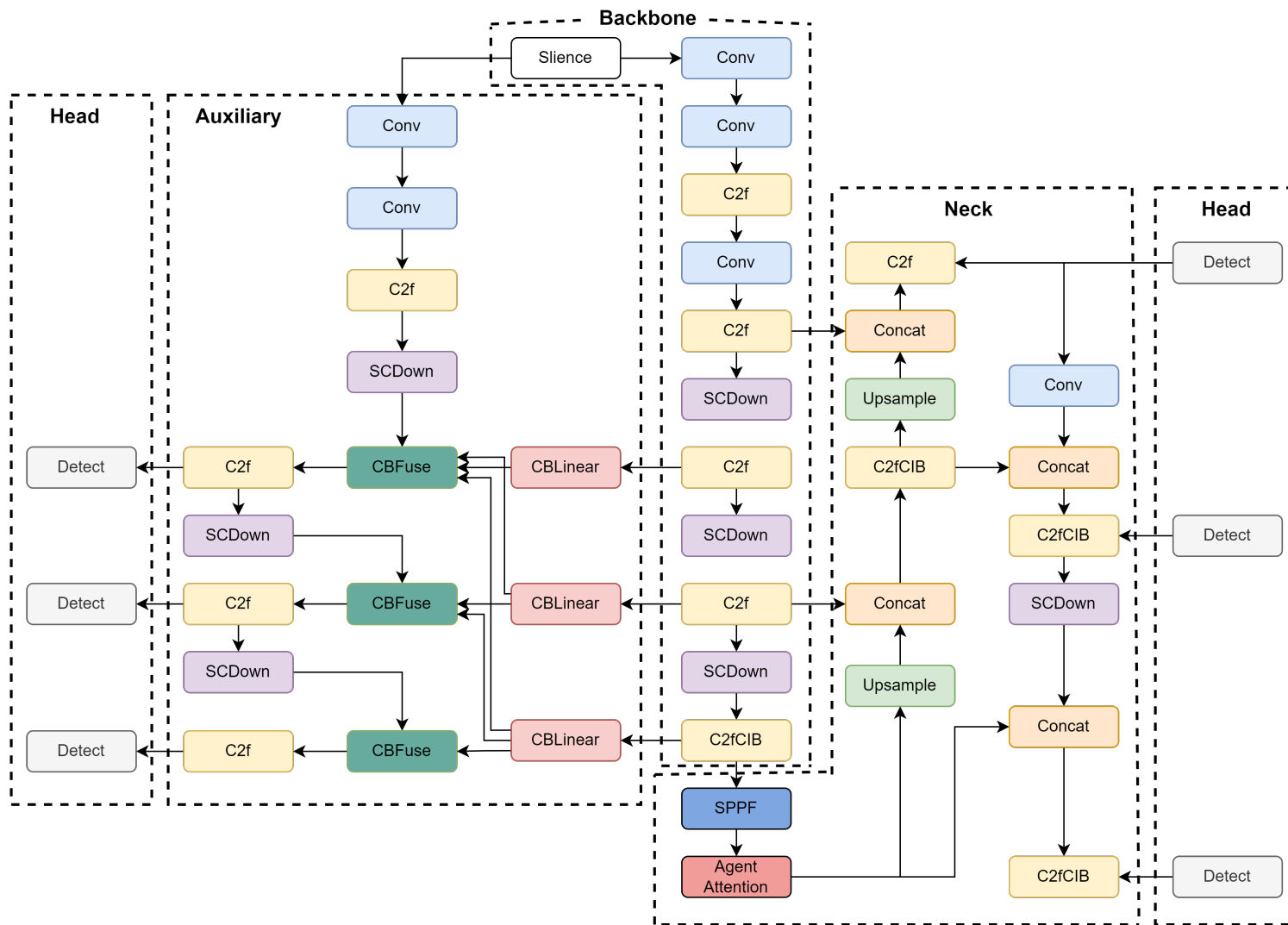


Figure 2.1: Structure of proposed YOLO-based attention-enhanced network

2.3 Experiments and Results

We performed a comprehensive comparative analysis to assess our proposed preprocessing method’s effectiveness against leading object detection models. The evaluation utilized a coronary angiography dataset in two configurations: the original unprocessed dataset and a version enhanced using our preprocessing method detailed in Chapter 1. The primary objective of this evaluation was to assess improvements in detection accuracy, with a particular focus on detecting stenosis in branch vessels, which often pose significant difficulties for conventional detection methods.

The impact of the proposed preprocessing method is illustrated in Table 2.1, which compares the performance of the same detection model when applied to the original dataset and the preprocessed dataset. These results exceed the best outcomes achieved by the model used in Chapter 1. Specifically, compared to the best result in Chapter 1 (YOLOv10-X), the method proposed in this chapter improved $AP_{val}(\%)$ and $AP_{50}(\%)$ on the original dataset by 1.9% and 2%, respectively. On the dataset preprocessed using our proposed method in Chapter 1, $AP_{val}(\%)$ and $AP_{50}(\%)$ improved by 2.7% and 4.4%, respectively. Furthermore, on the original dataset, the model achieved an AP_{val} of 53.5% and an AP_{50} of 84.9%. When tested on the preprocessed dataset, the model’s performance improved, achieving an AP_{val} of 56.6% and an AP_{50} of 91.5%. This improvement represents an approximate increase of +3.1% in AP_{val} and +6.6% in AP_{50} , highlighting that preprocessing the dataset using the method proposed in Chapter 1 enhances the performance of deep learning networks on the dataset.

We also tested the six angiographic angles using the method proposed in this chapter. The results are shown in Table 2.2. Across all angiographic angles, the proposed method achieved varying degrees of improvement in accuracy for stenosis detection. Notably, for the LCA zero degree angiographic angle, the dataset processed using the method proposed in Chapter 1 achieved an $AP_{50}(\%)$ value of 95%.

Table 2.1: Performance of YOLO-Based Attention-Enhanced Network on Two Datasets.

Dataset	Parameters (M)	FLOPs(G)	APval (%)	AP ₅₀ (%)	Latency (ms)
Original	10.2	48.0	53.5	84.9	3.52
Preprocessed	10.2	48.0	56.6	91.5	3.88

Table 2.2: Performance of YOLO-Based Attention-Enhanced Network on Six Imaging Angles.

Imaging Angle	APval(%)		AP ₅₀ (%)	
	Original	Preprocessed	Original	Preprocessed
LCA Zero	55.9	60.6	85.7	95.0
LCA LAO	56.2	57.3	87.8	93.3
LCA RAO	54.7	56.1	85.4	91.2
RCA Zero	50.9	52.7	84.8	87.1
RCA LAO	46.9	53.3	80.2	88.0
RCA RAO	48.5	52.1	80.7	85.7

2.4 Discussion

Our proposed method advances coronary artery segmentation through several key modifications to the YOLOv10 architecture: implementing Agent Attention, and incorporating a feedforward path for shallow feature preservation. These enhancements collectively improve the model’s processing capabilities for coronary angiography images. The replacement of Non-Maximum Suppression with an advanced matching mechanism significantly reduces redundancy and computational overhead, enabling more effective identification of regions of interest, particularly in complex vessel structures. This modification improves detection efficiency while maintaining accuracy, making it suitable for real-time clinical applications.

The Agent Attention mechanism implementation demonstrates effective balance between computational efficiency and representational capability. By introducing agent tokens to optimize attention processing, the method maintains global context modeling benefits while addressing computational challenges associated with traditional Softmax attention, leading to improved segmentation accuracy.

The addition of a feedforward path addresses the critical issue of shallow feature information loss during training. This design ensures preservation of both low-level and high-level features, crucial for medical imaging applications where fine detail accuracy is essential for diagnosis. Results confirm this modification’s contribution to improved segmentation performance, particularly in challenging scenarios like small vessel stenosis detection.

Experimental validation demonstrates the preprocessing method’s effectiveness, with improved AP scores showing enhanced object detection model performance on preprocessed datasets. This improvement is particularly noteworthy for small vessel stenosis detection, traditionally a challenging area. However, these improvements came with an increased processing latency, rising from 3.52ms to 3.88ms.

The clinical applications of the attention-enhanced network extend beyond mere technical improvements, offering practical benefits in daily cardiac care. The model’s

rapid processing time makes it particularly valuable as a real-time decision support tool during interventional procedures. In teaching hospitals, the system could serve as an educational resource, helping new interventional cardiologists develop and refine their ability to identify stenotic regions.

This research also faces several limitations.

Firstly, while we have made progress in addressing the first limitation discussed in Chapter 1, the second through fifth limitations remain unresolved as we utilized the same dataset for our experiments.

Furthermore, our detection process revealed a new challenge: HFV filtering generates vessel-like structures that lead to pseudo-feature generation in the output of backbone when deep learning networks use transformers or similar feature extraction methods. This newly identified issue will be a key focus of our future research efforts.

2.5 Conclusion

In conclusion, our modifications to the YOLOv10 architecture through Agent Attention and feedforward path integration have demonstrated improved coronary artery stenosis detection performance, particularly for complex vessel structures and small vessel stenosis detection. While achieving enhanced AP scores with only a minor increase in processing latency, the research identified a new challenge regarding vessel-like artifacts. Future work will focus on addressing this limitation while expanding dataset diversity to further validate and improve the system’s clinical applicability.

Chapter 3

Automated Extraction and 3D Reconstruction of the Right Coronary from CT Images Using Deep Learning

3.1 Introduction

The right coronary artery plays a vital role in cardiac function by delivering oxygen-rich blood to crucial heart structures, including the right atrium, right ventricle, and parts of the left ventricle. It also supplies the essential sinoatrial and atrioventricular nodes [42]. As the source of the posterior descending artery in most cases, the RCA is particularly significant in determining coronary dominance. Any RCA pathology, such as stenosis or anatomical variations, can result in serious complications including inferior myocardial infarction and cardiac arrhythmias[43]. While invasive coronary angiography remains the gold standard for RCA assessment, computed tomography has emerged as a valuable non-invasive alternative that provides detailed anatomical visualization without procedural risks [44].

The extraction and reconstruction of coronary arteries from CT images pose sig-

nificant challenges due to the complex structure of the coronary vessels, the presence of noise and calcifications, and the dynamic nature of the heart, which is influenced by both respiratory and cardiac motion[45]. Manual segmentation of these vessels is a time-consuming process prone to inter-observer variability, making it an unreliable option for consistent clinical use[46]. Automated methods, particularly those involving artificial intelligence, are increasingly being pursued as viable alternatives to overcome these challenges, ensuring consistency, accuracy, and efficiency[47, 48].

In recent years, convolutional neural networks have emerged as powerful tools for medical image segmentation, due to their ability to learn hierarchical features from large volumes of image data. For instance, the U-Net architecture[49] has become a benchmark in medical imaging tasks, demonstrating strong performance in segmenting various types of tissues and organs. This foundational architecture has inspired numerous variants and improvements, each addressing specific challenges in medical image analysis. The Attention U-Net[50] enhances the original architecture by incorporating attention gates, which enable the network to focus on relevant image regions while suppressing irrelevant features. This modification has proven particularly effective in multi-organ segmentation tasks, where precise boundary delineation is crucial. Similarly, the Residual U-Net[51] integrates residual connections to facilitate deeper network training and mitigate the vanishing gradient problem, resulting in improved feature representation and segmentation accuracy. Furthermore, the Dense U-Net[52] leverages dense connectivity patterns to strengthen feature propagation and encourage feature reuse, leading to more efficient network training and robust segmentation results. The 3D U-Net[53] extends the original architecture to process volumetric medical data directly, making it particularly suitable for three-dimensional imaging modalities such as CT and MRI scans. This adaptation has significantly improved the accuracy of volumetric segmentation tasks. More recently, the TransUNet[54] combines the strengths of U-Net with transformer architectures, capitalizing on both local and global feature dependencies for enhanced segmentation performance. The Swin-Unet[55] further advances this approach by incorporating

hierarchical transformer blocks, demonstrating state-of-the-art performance across various medical image segmentation benchmarks. However, despite their effectiveness, CNN-based approaches struggle with capturing long-range dependencies due to their limited receptive fields. This makes them less suitable for tasks where understanding the broader anatomical context is crucial, such as in coronary artery segmentation.

To overcome the limitations of CNNs in capturing global context, transformer-based architectures have gained increasing attention in the field of medical imaging. Originally developed for natural language processing, transformers utilize self-attention mechanisms that excel in modeling global dependencies and contextual relationships within data. In medical imaging, transformers have proven effective in enhancing segmentation accuracy by allowing the model to consider the entire image context, rather than being constrained to local features. This capability is particularly beneficial in segmenting coronary arteries. The Vision Transformer (ViT)[56] utilizes self-attention mechanisms to model global dependencies within the input data. Recent transformer models, such as the Swin Transformer[57], have further improved the efficiency of self-attention by introducing a hierarchical representation, making them more computationally feasible for use in medical applications. Additionally, several studies have focused on vessel segmentation in coronary artery CT images, exploring various methods. For instance, a CNN-based approach[58] has been applied for coronary artery segmentation from CT images, demonstrating strong performance in detecting vessels; however, their method suffered from high computational requirements and occasional inaccuracies in small vessel detection. Similarly, a 3D fully convolutional network (FCN) [59] has been employed to segment coronary arteries, which improved the continuity of segmented vessels but faced challenges in distinguishing closely positioned vessel branches, especially in the presence of noise or calcifications. Another notable work introduced a hybrid approach combining a U-Net with attention mechanisms to segment coronary arteries from CT images[60]. This method provided improved segmentation accu-

racy, particularly for smaller vessels, by focusing on critical regions. However, the reliance on complex attention layers resulted in increased memory consumption and slower inference times, limiting its real-time applicability in clinical settings. These methods collectively highlight the ongoing efforts and challenges in achieving precise and efficient coronary artery segmentation, paving the way for more advanced approaches, such as the integration of transformer-based models discussed in this work.

Despite the advantages transformers bring to medical image analysis, they come with significant computational costs, which limits their practical applicability in clinical environments that require real-time processing. One promising approach to mitigate these issues is the integration of lightweight network architectures, such as GhostNet, into transformer models. GhostNet[7] employs Ghost modules to generate additional feature maps from intrinsic features, significantly reducing the computational burden while retaining accuracy. This hybrid approach, combining the global attention capabilities of transformers with the parameter efficiency of GhostNet, offers a balanced solution that is both effective and practical for clinical use. In medical imaging, GhostNet has been incorporated into U-Net architectures, such as Ghost-U-Net[8], to improve liver tumor segmentation in CT scans, achieving efficient processing without compromising accuracy. Additionally, GhostNet has been utilized in mobile healthcare applications, including real-time ECG classification, where its lightweight design enables rapid analysis with reduced latency and power consumption[61].

Furthermore, transitioning from 2D cross-sectional images to a 3D reconstruction of coronary arteries presents notable advantages. Three-dimensional reconstructions provide a holistic view of the vascular structures, enabling clinicians to assess vessel morphology more intuitively and accurately. Studies have shown that 3D models can significantly improve the accuracy of detecting and quantifying stenosis[62]. For sonographers and radiologists, the added depth information facilitates a more precise evaluation of complex coronary anatomy, aiding in better preoperative plan-

ning and reducing uncertainties during interventions. This advancement not only enhances diagnostic capabilities but also reduces the cognitive load on healthcare professionals, ultimately leading to improved clinical outcomes. For patients, the shift towards non-invasive 3D imaging provides multiple benefits, including reduced exposure to the risks associated with invasive coronary angiography, such as arterial damage or infection. Additionally, a 3D representation allows for more personalized treatment planning, as the detailed visualization of the coronary artery tree can help clinicians devise more accurate interventions, potentially improving prognosis and reducing the likelihood of adverse events[63].

This chapter presents a novel method utilizing a transformer integrated convolutional neural network with GhostNet for segmentation, enabling the automated extraction of the RCA from CT images and its reconstruction in 3D. The approach aims to address both the accuracy and efficiency challenges inherent in current methodologies, ultimately contributing to improved clinical workflows and patient care in the context of coronary artery disease management.

Our study has made the following contributions in this chapter:

1. We proposed a method for extracting cross-sections of the right coronary artery that achieves effective results without requiring excessive computational resources.
2. By employing volume rendering and surface smoothing techniques, we successfully reconstructed the three-dimensional structure of the vessel, facilitating more accurate localization of coronary artery stenosis for clinical diagnosis.

3.2 Materials and Methods

3.2.1 Materials

The research dataset consisted of CT images collected from 32 patients examined at Hokkaido University Hospital. The CT volumes ranged from 400 to 640 slices

per patient, with each slice having dimensions of 512×512 pixels. Through careful manual selection, 4,830 CT slices containing diagnostically significant information were separated from the complete volumes. These selected images underwent detailed annotation to mark the right coronary artery cross-sections, establishing an accurate ground truth dataset necessary for model development and evaluation. The dataset distribution involved 3,400 CT images assigned for training purposes, 500 allocated for validation, and 930 designated for testing, all selected through random sampling from the 32 cases.

The image preparation workflow incorporated both preprocessing and data augmentation techniques. The preprocessing involved enhancing image quality, while augmentations, such as rotations, scaling, and flipping, were implemented to artificially increase the dataset’s diversity. These augmentation strategies served to expand the effective dataset size artificially, thereby strengthening the model’s ability to generalize and minimizing potential overfitting issues by introducing diverse data variations. The entire preprocessing and augmentation workflow was executed using Matlab2023b.

3.2.2 Model descriptions

3.2.2.1 DeepLabv3+

DeepLabv3+ [64] represents an advanced iteration within the DeepLab model series, distinguished by its integration of atrous convolution techniques with an encoder-decoder framework for semantic segmentation tasks. The model’s architecture employs atrous convolutions to regulate feature response resolution and extract contextual information across multiple scales while preserving spatial details. Its encoder-decoder configuration enhances the preservation of spatial information throughout the segmentation pipeline. A key feature of DeepLabv3+ is the incorporation of atrous spatial pyramid pooling (ASPP), which enables multi-scale information capture, particularly beneficial for analyzing intricate structures like coronary arteries. Through the synergistic combination of atrous convolutions and ASPP,

DeepLabv3+ delivers precise segmentation outcomes while maintaining computational performance, rendering it particularly suitable for clinical environments where accurate and detailed segmentation is essential.

3.2.2.2 U-Net++

U-Net++ [65] represents an enhanced iteration of the U-Net architecture, featuring integrated nested and dense skip pathways to improve segmentation accuracy. The model's nested skip connections effectively reduce the semantic gap between encoder and decoder feature representations, leading to enhanced segmentation capabilities. Through its dense skip connections, U-Net++ creates multiple pathways for hierarchical feature learning, facilitating improved feature representation and more precise segmentation outcomes. The architecture demonstrates particular effectiveness in analyzing complex anatomical structures where detailed segmentation precision is essential. The combination of semantic gap reduction and enhanced feature learning pathways makes U-Net++ particularly valuable for medical imaging applications, specifically in coronary artery segmentation tasks where both precise vessel detail capture and broader anatomical context are critical.

3.2.2.3 TransUNet

TransUNet [54] represents a hybrid architecture that combines transformer technology with U-Net design principles, specifically engineered for medical image segmentation tasks. The architecture leverages U-Net's established encoder-decoder framework, known for its proficiency in local feature extraction, while integrating transformer capabilities for comprehensive contextual analysis. The transformer components enable the model to process long-range dependencies and contextual relationships within images. This integrated approach enables TransUNet to simultaneously maintain precise local detail capture while processing broader contextual information, resulting in enhanced segmentation performance. The model demonstrates particular effectiveness in segmenting intricate anatomical features like coro-

nary arteries, where accurate representation of both fine details and spatial context is essential. TransUNet’s architecture has shown significant improvements in clinical applications by delivering precise segmentations that support accurate diagnosis and treatment strategy development.

3.2.2.4 RCA-IUnet

RCA-IUnet [66] represents a specialized deep learning architecture designed for breast ultrasound tumor segmentation, playing a vital role in breast cancer diagnosis and intervention. Building upon U-Net’s foundation, the model incorporates several key enhancements for improved performance. The architecture utilizes Residual Inception Blocks to extract multi-scale features through varied convolutional kernel sizes, enabling comprehensive tumor detection across different morphologies. Its Cross-Spatial Attention mechanism, integrated within skip connections, enables focused processing of tumor regions while minimizing background interference, resulting in precise boundary definition. The model employs Hybrid Pooling, combining max and spectral pooling approaches, to maintain crucial spatial information during downsampling operations. RCA-IUnet’s implementation of Depth-Wise Separable Convolutions optimizes computational efficiency while maintaining effectiveness. Through validation on public datasets, the model has demonstrated superior segmentation capabilities compared to contemporary approaches, establishing itself as a valuable tool for breast cancer diagnostics across various healthcare settings, including those with limited resources.

3.2.2.5 ResUNet

ResUNet [51] advances the U-Net framework by incorporating residual connections, specifically engineered to enhance feature propagation and combat the vanishing gradient issue commonly encountered in deep neural networks. These residual connections establish direct gradient pathways, facilitating more effective network training through improved gradient flow. This architectural enhancement proves particu-

larly valuable in medical image segmentation applications where detail preservation is essential. The model excels at capturing intricate features while maintaining fine-scale details, making it particularly suited for segmenting delicate structures such as coronary vessels. Through the implementation of residual blocks, ResUNet ensures robust feature preservation and enables effective training of deeper network layers without performance deterioration. This architecture demonstrates particular effectiveness in challenging segmentation tasks where precise capture of subtle anatomical variations is critical.

3.2.2.6 Medical Transformer

The Medical Transformer [67] represents a dedicated deep learning architecture adapting transformer principles to medical imaging applications, spanning segmentation, classification, and diagnostic tasks. It features a distinctive gated axial-attention mechanism that enhances segmentation capabilities on limited datasets through effective integration of both global and local context. The model’s gated system modifies conventional axial attention by implementing control gates that regulate positional embedding influences on key attention components, enhancing learning efficiency with smaller medical datasets. The architecture employs a dual Local-Global training approach, simultaneously processing complete images and image patches to capture comprehensive feature relationships. Performance evaluations across multiple medical image segmentation datasets have shown Medical Transformer’s superior capabilities compared to traditional convolutional neural networks and alternative transformer-based models, establishing a notable advancement in medical image analysis methodology.

3.2.2.7 GhostNet

GhostNet[7] is a lightweight neural network module designed to reduce the computational load without sacrificing performance. It achieves this by using Ghost modules, which generate additional feature maps from the original features through simple

linear operations. The core idea of GhostNet can be expressed mathematically as:

$$F_{\text{ghost}} = F_{\text{intrinsic}} + h(F_{\text{intrinsic}}) \quad (3.2.1)$$

where $F_{\text{intrinsic}}$ represents the intrinsic feature maps generated by a standard convolution, and $h(\cdot)$ represents a set of cheap linear transformations applied to $F_{\text{intrinsic}}$ to produce more feature maps. This approach significantly reduces the number of parameters and computational costs associated with the model, making it suitable for applications where efficiency is crucial.

GhostNet’s architecture employs a dual approach to feature map generation, utilizing standard convolution layers for primary feature extraction alongside Ghost modules for derivative feature map creation. This architectural design enables GhostNet to maintain robust feature representation capabilities while achieving significant reductions in parameter count and computational operations.

The implementation positions GhostNet between encoder and decoder components, creating an optimized architecture for coronary artery segmentation that balances performance with computational efficiency. This integration leverages Ghost modules to maintain segmentation precision while reducing model complexity, particularly beneficial for real-time clinical applications. The architecture incorporates batch normalization and activation layers throughout the Ghost module sequence, ensuring optimal feature map conditioning for subsequent network processing stages.

3.2.2.8 Volume Rendering

Volume rendering serves as a core visualization technique for examining complex three-dimensional anatomical structures, with significant applications in medical imaging. This methodology transforms volumetric data obtained from CT and MRI scans into comprehensive visual displays that highlight intricate spatial anatomical relationships [68]. In the specific context of coronary artery visualization, volume rendering plays a crucial role by offering detailed perspectives of vessel structure, thus enhancing diagnostic accuracy and supporting treatment planning processes.

The technique involves converting volumetric data into three-dimensional representations through projection onto two-dimensional viewing planes. In contrast to surface rendering methods that only show structural boundaries, volume rendering examines the complete data volume to display internal characteristics. This capability is essential in medical imaging scenarios where healthcare professionals must evaluate both external surfaces and internal features, including tissue density variations [69]. The technique’s flexibility in modifying opacity levels and color mappings enables selective emphasis of various tissue types and anatomical structures, establishing volume rendering as an adaptable tool for medical visualization applications. Volume rendering distinguishes itself through its capacity to preserve fine details that might be lost in conventional surface rendering methods. Studies have demonstrated that volume rendering enhances the precision of clinical evaluations by providing more realistic anatomical visualizations compared to surface rendering techniques [70].

3.2.2.9 Proposed Methods

Despite the proven capabilities of the Medical Transformer in many applications, it has a significant drawback: deep networks generally require long training times and incur high computational costs. Specifically, while the Medical Transformer employs optimizations such as reducing attention window sizes and hierarchical processing to maintain training accuracy while reducing training costs, the fundamental Transformer architecture limits the effectiveness of these measures. Additionally, the object queries in the decoder of the Medical Transformer are not highly correlated with the image, which impacts segmentation quality.

The proposed attention mechanism model is based on enhancing the axial attention mechanism. The axial attention mechanism is an approach that improves the efficiency of self-attention, particularly for high-dimensional data such as images or 3D volumes. Unlike traditional self-attention, which calculates relationships between all elements in the input tensor simultaneously, axial attention breaks this

computation into sequential steps along specific dimensions, such as rows, columns, or depth. In Figure 3.1, the input tensor, x , is first transformed using learnable weight matrices (W_q , W_k , and W_v) into queries (Q), keys (K), and values (V). These components allow the model to compute attention scores and aggregate information from relevant positions. Positional information (r^q , r^k , and r^v) is added at various stages to capture spatial dependencies, ensuring that the mechanism respects the structure of the input. The attention process begins by computing the interaction between Q and K , incorporating positional encodings to reflect relative positions. These interactions are normalized using a softmax function, producing attention weights. These weights are then used to combine the corresponding values, resulting in an aggregated representation. This process is repeated for each axis independently, such as row-wise and then column-wise, to ensure the model captures dependencies across all dimensions. By decomposing the computation along each axis, axial attention significantly reduces the computational cost compared to traditional self-attention. Instead of scaling quadratically with the total number of elements in the input tensor, the computation scales linearly with respect to the size of each axis. This efficiency makes axial attention particularly suitable for handling large inputs while maintaining high performance. The output of axial attention, represented as y in the diagram, is a context-aware representation of the input, combining information along all dimensions. This mechanism not only preserves the global context of the input but also allows for more efficient processing of high-dimensional data, making it a powerful tool for tasks like image and volume analysis, where computational efficiency and spatial structure are both critical.

The gated axial attention mechanism, in Figure 3.2, builds upon axial attention by introducing gate parameters (qr , kr , sv , sve) to dynamically control the contributions of queries, keys, values, and positional encodings. This adjustment addresses a limitation in axial attention, where positional information (r^q , r^k , r^v) can overly influence the attention scores. By scaling the impact of positional encodings, the gates ensure a better balance between content relevance and spatial dependencies,

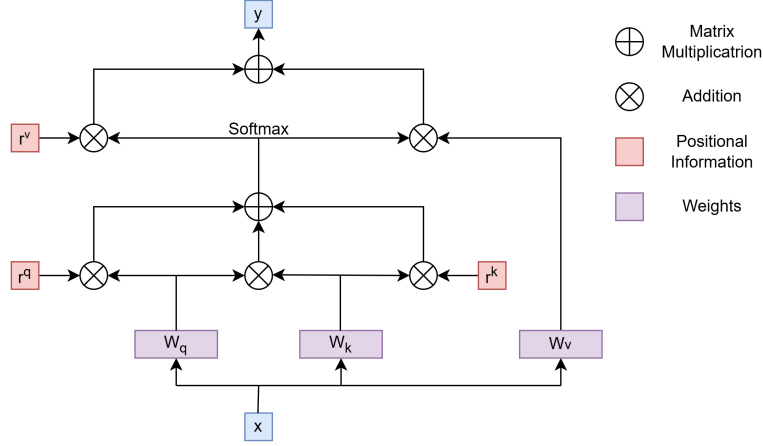


Figure 3.1: The architecture of position-sensitive axial-attention

allowing the model to focus more effectively on meaningful relationships in the data. In this mechanism, the gates act as learnable coefficients that modulate the interaction between content and position. For example, qr and kr scale the influence of queries and keys when combined with positional encodings, while sv adjusts the contribution of values during aggregation. This dynamic modulation reduces the risk of overfitting to positional biases, making the model more robust to variations in spatial structure. The gated axial attention mechanism retains the computational efficiency of axial attention, as it still processes data along individual axes. However, the added flexibility of gate parameters allows the model to adapt to diverse data distributions and varying importance of spatial axes. This makes it especially useful for high-dimensional structured data, such as images and volumetric datasets, where spatial dependencies can differ significantly across dimensions.

We observed that when using gated axial attention for large tissue segmentation, the segmentation performance improved when the gate unit weights were smaller. Conversely, for smaller tissues, the segmentation accuracy improved when the gate unit weights were larger. This indicates that the smaller the impact of gate unit parameters on positional information, the more effective the model becomes in segmenting small targets. Since our goal is to identify small targets, we introduced four

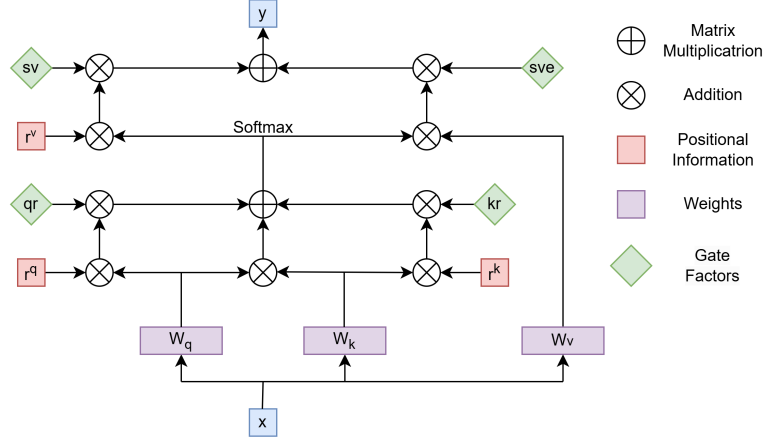


Figure 3.2: The architecture of gated axial-attention

additional parameters, as shown in Figure 3.3, to mitigate the weakening effect of the gate unit on positional information, thereby enhancing the model’s capability to segment vascular cross-sections.

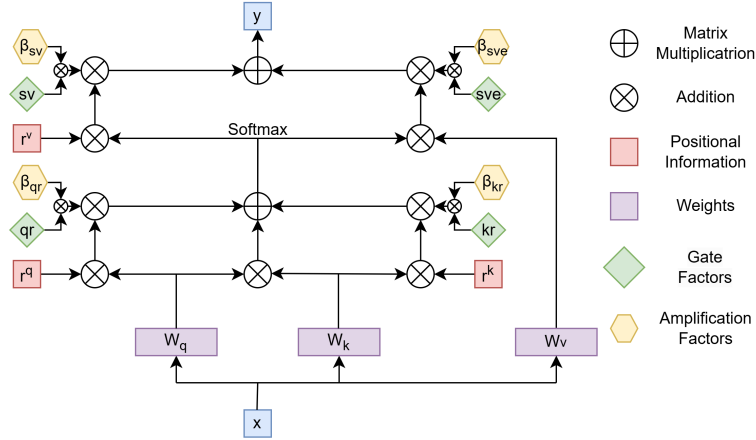


Figure 3.3: The architecture of proposed attention mechanism

With the proposed modifications, the self-attention mechanism applied to the

width axis can be expressed as follows:

$$y_{ij} = \sum_{w=1}^W \text{softmax} \left(q_{ij}^T k_{iw} + \beta_{qr} G_Q q_{ij}^T r_{iw}^q + \beta_{kr} G_K k_{iw}^T r_{iw}^k \right) \left(\beta_{sv} G_{V1} v_{iw} + \beta_{sve} G_{V2} r_{iw}^v \right) \quad (3.2.2)$$

We introduced amplification factors to the gated attention mechanism. Specifically, G_Q , G_K , G_{V1} , G_{V2} , β_{qr} , β_{kr} , β_{sv} , and β_{sve} are learnable parameters that collectively form the new parameters for modulating positional information. These parameters control the influence of the learned relative positional encoding on capturing non-local contextual information.

Additionally, we utilized GhostNet as the feature extractor, which significantly reduced the number of parameters and computational memory requirements while maintaining prediction accuracy. This approach not only lowered the time required for feature extraction but also enhanced the model's efficiency. To further improve performance, we incorporated a structure similar to the Medical Transformer, which combines both local and global features extracted by GhostNet and feeds them into the decoder. This integration enabled effective detection of small regional targets, a critical factor for accurately segmenting the coronary arteries.

The overall network structure can be referred to in Figure 3.4. In the proposed architecture, feature extraction is divided into two pathways: global and local. After extracting the features, they are fused and passed through fully connected layers. For global feature extraction, due to the large image size, we reduced the number of convolutional layers. On the other hand, for local feature extraction, where the segmented regions are smaller, we increased the number of convolutional layers. This adjustment ensures that the time required for global and local feature extraction is approximately the same, facilitating the subsequent feature fusion process.

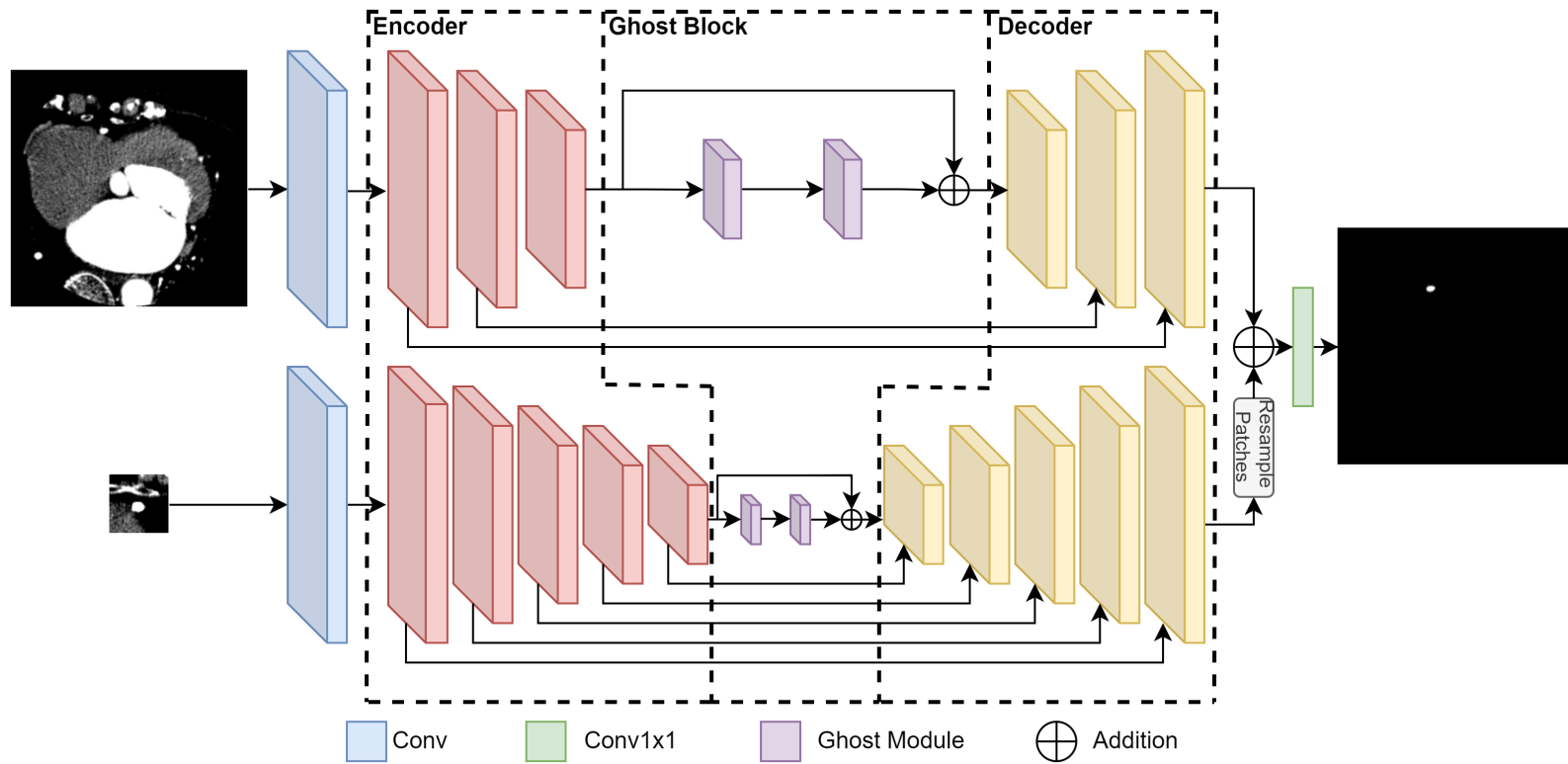


Figure 3.4: The structure of the proposed method in Chapter 3

3.2.3 3D Reconstruction

The 3D reconstruction of the vessel was achieved by creating a mesh grid from extracted boundary points in sequential 2D slices. This mesh grid represents the vessel's morphology across layers and serves as the foundation for volume rendering. By later applying Laplacian Pyramid Interpolation to the volume, we enhanced the resolution and smoothness of the reconstructed structure. This section details the steps involved in mesh grid construction, data storage, and 3D rendering.

3.2.3.1 Boundary Point Extraction and Layer-by-Layer Mesh Grid Construction

To represent the vessel's structure in 3D, we first extracted the boundary points from each 2D slice of the vessel. Let $P_i = \{(x_{i,1}, y_{i,1}), (x_{i,2}, y_{i,2}), \dots, (x_{i,N}, y_{i,N})\}$ denote the set of N boundary points extracted from the i -th layer, where $(x_{i,n}, y_{i,n})$ is the coordinate of the n -th boundary point in the 2D plane. These points were obtained through segmentation of each vessel slice and serve as the vertices of the mesh grid.

3.2.3.1.1 Intra-layer Connections

Within each layer i , we constructed a closed curve by connecting each pair of adjacent points $(x_{i,n}, y_{i,n})$ and $(x_{i,n+1}, y_{i,n+1})$ with a line segment, where $n + 1$ is taken modulo N to form a continuous boundary. Mathematically, this intra-layer connection can be represented by the set of line segments:

$$L_{\text{intra},i} = \{((x_{i,n}, y_{i,n}), (x_{i,n+1}, y_{i,n+1})) \mid n = 1, 2, \dots, N\}. \quad (3.2.3)$$

3.2.3.1.2 Inter-layer Connections

To extend the vessel structure across layers, we connected corresponding points in adjacent layers. For each point $(x_{i,n}, y_{i,n})$ in layer i , we drew a line segment to the

point $(x_{i+1,n}, y_{i+1,n})$ in the next layer $i + 1$. This connection across layers is defined as:

$$L_{\text{inter}} = \bigcup_{i=1}^{M-1} \{((x_{i,n}, y_{i,n}, z = i), (x_{i+1,n}, y_{i+1,n}, z = i + 1)) \mid n = 1, 2, \dots, N\}, \quad (3.2.4)$$

where M represents the total number of layers, and $z = i$ indicates the layer's position in the 3D space.

Combining the intra-layer and inter-layer connections, we obtain a complete 3D mesh grid representing the vessel's structure. The 3D coordinates for each boundary point are given by $(x_{i,n}, y_{i,n}, z = i)$, where z serves as the depth axis.

3.2.3.2 Saving Mesh Grid Data

The mesh grid data, consisting of the boundary points and connection lines, was saved for each layer to preserve the structure and facilitate further processing. This data was stored in a structured format as a sequence of points and line segments for use in subsequent volume rendering and interpolation steps. Each point $(x_{i,n}, y_{i,n}, z = i)$ was indexed, and connectivity matrices were created to represent the intra-layer and inter-layer connections.

3.2.3.3 Initial Volume Rendering of the Mesh Grid

With the 3D mesh grid constructed, an initial volume rendering was performed to visualize the vessel structure in 3D space. By treating the mesh grid as a spatial representation of the vessel, we obtained a preliminary volume rendering that provided an initial overview of the vessel's morphology.

3.2.3.4 Resolution Enhancement and Smoothness Adjustment via Laplacian Pyramid Interpolation

To improve the resolution and smoothness of the reconstructed volume, we applied Laplacian Pyramid Interpolation to the initial rendering. The Laplacian Pyramid

Interpolation process refined the spatial details and enhanced the continuity of the vessel surface, resulting in a high-quality 3D visualization.

Given the original volume data V , represented by the 3D coordinates of mesh points, we constructed a Laplacian pyramid $\{V_0, V_1, \dots, V_L\}$, where V_0 is the finest level and V_L is the coarsest level. Using Gaussian filtering and downsampling, we defined each level of the Gaussian pyramid $\{G_0, G_1, \dots, G_L\}$ as:

$$G_k = \text{Downsample}(\text{Gaussian}(G_{k-1})), \quad (3.2.5)$$

where $G_0 = V$. Subsequently, each level of the Laplacian pyramid was formed by subtracting the upsampled version of the next coarser Gaussian level from the current Gaussian level:

$$V_k = G_k - \text{Upsample}(G_{k+1}). \quad (3.2.6)$$

Starting from the coarsest level V_L , we iteratively upsampled and combined the layers to reconstruct the volume at higher resolution:

$$\hat{V}_k = \text{Upsample}(\hat{V}_{k+1}) + V_k. \quad (3.2.7)$$

The Laplacian Pyramid Interpolation enhanced structural continuity and detail across the vessel surface, ensuring that the 3D rendering retained both smoothness and resolution. This step provided a refined, high-resolution 3D volume suitable for detailed analysis and visualization of the vessel morphology.

3.2.4 Evaluation Indicators

3.2.4.1 Structural Similarity Index Measure

The Structural Similarity Index Measure (SSIM) is a perception-based model that measures the similarity between two images by comparing their luminance, contrast, and structural information[71]. SSIM values range from 0 to 1, with values closer to 1 indicating a higher similarity. SSIM is computed as:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3.2.8)$$

where:

x and y are two image patches being compared,

μ_x and μ_y are the mean intensities of x and y , respectively,

σ_x^2 and σ_y^2 are the variances of x and y ,

σ_{xy} is the covariance of x and y ,

C_1 and C_2 are small constants to stabilize the division when the denominators are close to zero.

The SSIM index considers both local luminance and structural information, making it a reliable metric for evaluating the perceptual quality of images.

3.2.4.2 Peak Signal-to-Noise Ratio

The Peak Signal-to-Noise Ratio (PSNR) is an objective metric that quantifies the difference between the original and interpolated images[72]. It is defined as the ratio between the maximum possible pixel value of an image and the mean square error (MSE) between the original and interpolated images. PSNR is expressed in decibels (dB), with higher values indicating better image quality. PSNR is calculated as follows:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right) \quad (3.2.9)$$

where:

MAX is the maximum possible pixel value of the image (e.g., 255 for 8-bit images),

MSE is the mean squared error between the original image I and the interpolated image I' and is defined as:

$$\text{MSE} = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n [I(i, j) - I'(i, j)]^2 \quad (3.2.10)$$

3.2.4.3 F1 Score

The F1 Score is a metric used to evaluate the accuracy of a binary classification model, considering both precision and recall. It is particularly useful when there is

an imbalanced class distribution. The F1 Score is the harmonic mean of precision and recall, providing a single measure that balances the two.

The F1 Score can be calculated as:

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.2.11)$$

The F1 Score is maximized when both precision and recall are high, making it a useful metric when you need a balance between the two.

3.2.4.4 Mean Intersection over Union

The Mean Intersection over Union (mIoU) is a metric commonly used in semantic segmentation tasks. It measures the overlap between the predicted segmentation and the ground truth. For a single class, IoU is defined as the ratio of the intersection of the predicted and true positive pixels to the union of all predicted and true positive pixels. The mean IoU is computed by averaging the IoU values across all classes.

The Mean IoU is the average of the IoU scores across all classes:

$$\text{mIoU} = \frac{1}{N} \sum_{i=1}^N \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i + \text{FN}_i} \quad (3.2.12)$$

Where: N is the number of classes (e.g., background and different object categories in segmentation tasks). The summation is taken over all classes.

This metric provides a comprehensive measure of performance for segmentation tasks, as it takes into account the overlap between predicted and actual regions across all classes.

3.3 Experiment and Results

In this study, we conducted training and prediction experiments using the same dataset across various state-of-the-art deep learning networks to evaluate their performance. The networks included Medical Transformer, RCA-IUnet, ResUNet, TransUNet, UNet++, DeepLabV3+, as well as our proposed networks—one without GhostNet and the other with GhostNet.

The comparative results of these networks are presented in Table 3.1, where we provide detailed statistics, including the number of parameters, inference speed, F1 score, and IoU. These metrics were chosen to offer a comprehensive evaluation of the computational efficiency and prediction accuracy of each model.

To visually demonstrate the differences in performance, we selected two representative images from the test set. These images highlight the recognition capabilities of each network in generating segmentation masks. The prediction masks generated by each network, alongside their corresponding ground truth, are displayed in Figure 3.5. This figure allows for a clear qualitative comparison of the models' effectiveness in segmenting the input images.

Table 3.1: Performance Comparison of Different Models on CT Dataset.

Method	Parameters (M)	Inference Speed (ms)	F1 Score	IoU
Medical Transformer	10.6	12.82	0.882	0.788
RCA-IUnet	7.3	8.37	0.783	0.643
TransUnet	98.9	114.74	0.829	0.708
ResUNet	56.2	54.31	0.723	0.566
Unet++	63.9	73.32	0.828	0.707
Deeplabv3+	31.5	36.10	0.841	0.726
Ours	11.0	14.05	0.897	0.814
Ours*	5.5	7.03	0.887	0.797

*: with GhostNet.

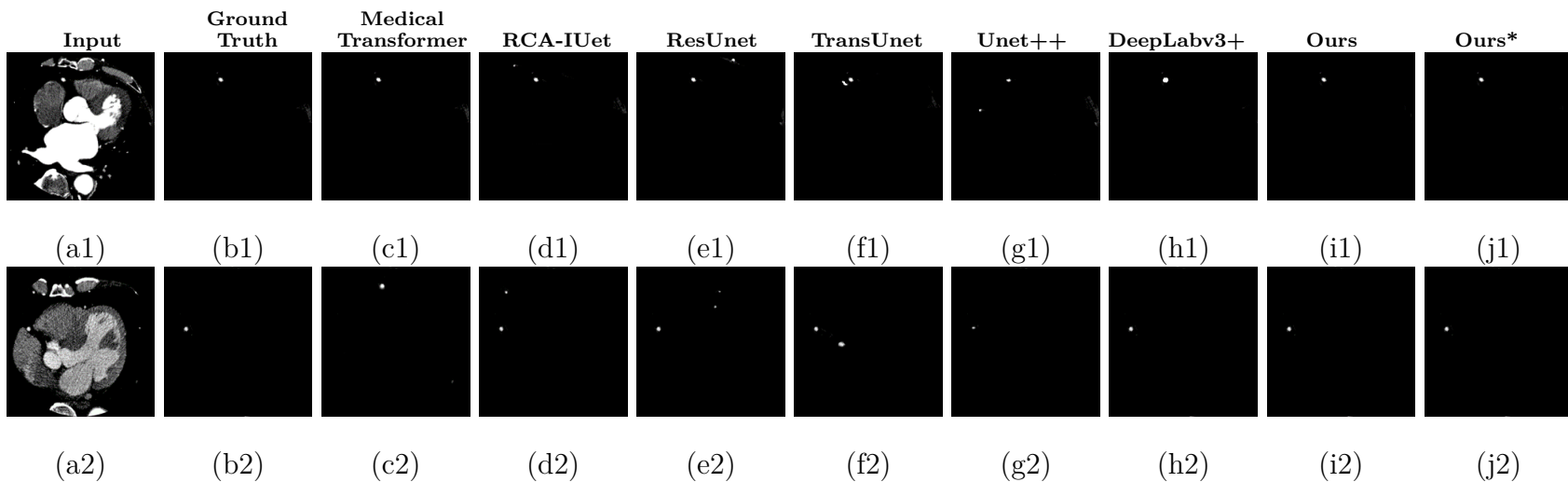


Figure 3.5: Prediction results of different networks for two input images. (a1, a2) represent the two input example images; (b1, b2) correspond to their respective ground truth. (c1, c2; d1, d2; ... j1, j2) show the performance of the two images under the Medical Transformer, RCA-IUnet, ResUNet, TransUNet, UNet++, DeepLabV3+, Ours, and Ours* networks, respectively.

The performance comparison of various deep learning models, as detailed in Table 3.1, highlights both the accuracy and computational efficiency of our proposed methods on the given CT image dataset. Among the tested methods, our proposed network without GhostNet achieves the highest F1 score of 0.897, demonstrating superior segmentation accuracy for the RCA cross-sections across multiple CT layers. This reflects the effectiveness of our approach in extracting fine vascular structures. When GhostNet is incorporated (Ours*), there is a marginal decrease in segmentation accuracy, with the F1 score reduced to 0.887 and IoU to 0.797. However, this configuration achieves a significant improvement in inference speed, recording the fastest time of 7.03 ms. This trade-off between accuracy and computational efficiency makes the GhostNet-enabled model particularly suitable for applications where real-time performance is critical.

Comparatively, other networks, such as the Medical Transformer and Deeplabv3+, exhibit competitive performance in terms of accuracy, with F1 scores of 0.882 and 0.841, respectively. However, their computational costs, including larger parameter sizes and slower inference speeds, make them less efficient for practical deployment. Networks like RCA-IUnet and ResUNet, while lighter in terms of parameters, show significantly lower F1 scores of 0.783 and 0.723, indicating less effective segmentation capabilities.

After extracting the cross-sectional masks of each layer of the right coronary artery, we proceeded to extract the edges of the predicted cross-sectional masks. On these edges, 30 points were uniformly distributed. Each point was assigned a unique identifier, and neighboring points were connected by lines. This approach enabled the construction of a point-based matrix, which was then utilized to generate a three-dimensional mesh grid for reconstruction.

Figure 3.6 illustrates the reconstructed RCA of one patient. In this visualization, blue lines represent the connections between points within each individual layer, while red lines indicate the connections between corresponding points across adjacent layers. For better visualization and clarity, the Z-axis—representing the inter-layer

distance—was appropriately stretched to emphasize the spatial relationships.

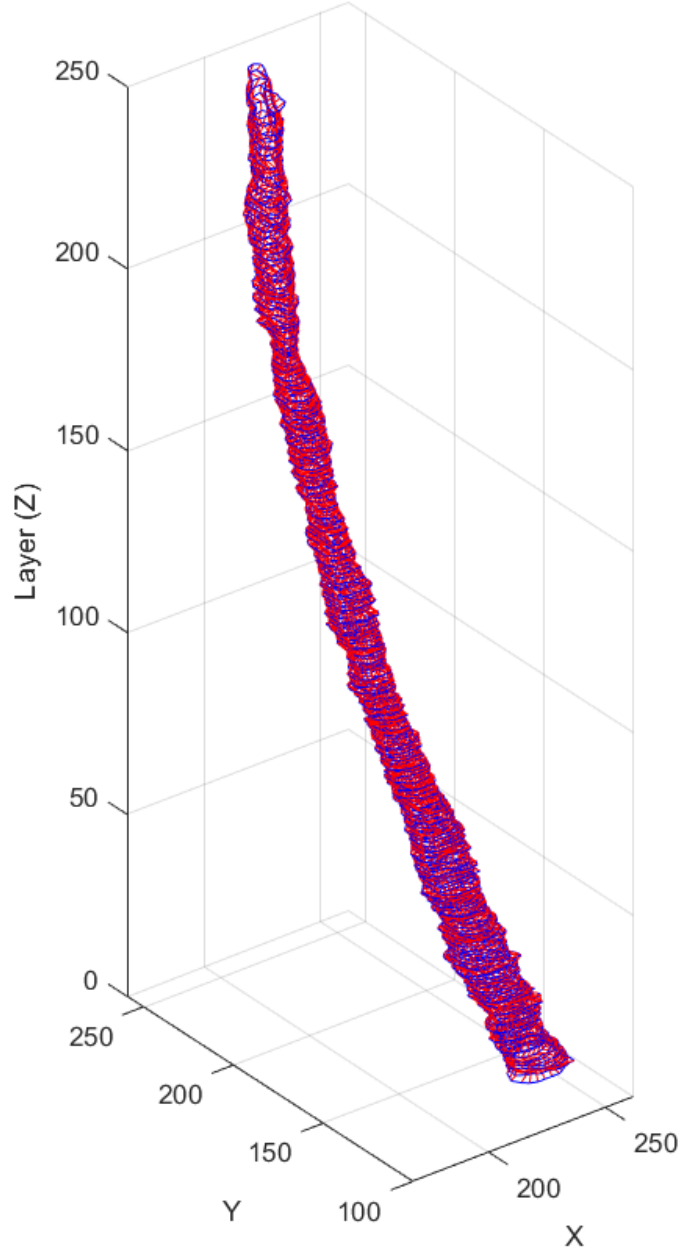


Figure 3.6: The mesh grid generated from prediction Results

Subsequently, based on the constructed mesh grid, we employed volume rendering techniques to further enhance the visualization of the vascular structure. Specifically, we utilized a surface rendering approach within the volume rendering framework to transform the mesh grid into a continuous, smooth representation of the vessel surface. This method enabled the detailed depiction of the vascu-

lar topology, providing a clear and intuitive visualization of both the location and morphology of vascular narrowing.

The rendering process involved assigning specific color and opacity values to the surface based on the geometric attributes of the mesh grid, allowing for the differentiation of structural details. This facilitated the identification of stenotic regions by highlighting their spatial characteristics and transitions. The resulting image, shown in the Figure 3.7, represents a preliminary rendered visualization of the RCA. The coordinate system, displayed in the lower-left corner, provides spatial reference for interpreting the reconstructed geometry.



Figure 3.7: The result of volume rendering

To further enhance the visual quality and improve the interpretability of the reconstructed vascular image, we applied smoothing techniques. Specifically, we utilized two methods: Laplacian Pyramid Interpolation and Gaussian Smoothing

Interpolation. These methods were employed to reduce visual noise and ensure a smoother representation of the vascular surface, facilitating better observation and analysis.

The results of the smoothing processes are illustrated in the Figure 3.8. Subfigure (a) shows the outcome of the Gaussian Smoothing Interpolation, which emphasizes gradual transitions and overall uniformity in the surface structure. Subfigure (b) presents the results of the Laplacian Pyramid Interpolation, which preserves fine details while reducing abrupt variations in the mesh.

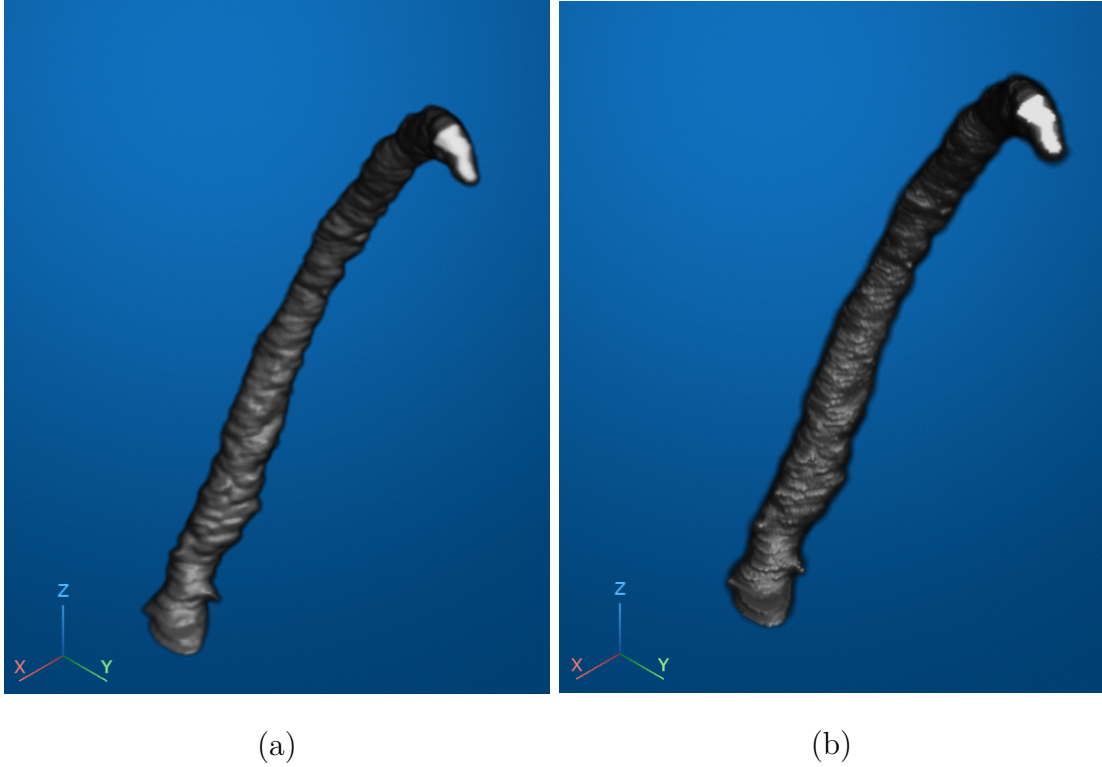


Figure 3.8: Comparison between Laplacian Pyramid Interpolation and Gaussian Smoothing Interpolation

In addition, we attempted to identify an optimal viewing angle from the smoothed vascular images that provided the clearest observation of the stenosis's degree and location. To highlight these stenotic regions, we used red bounding boxes for annotation in Figure 3.9.

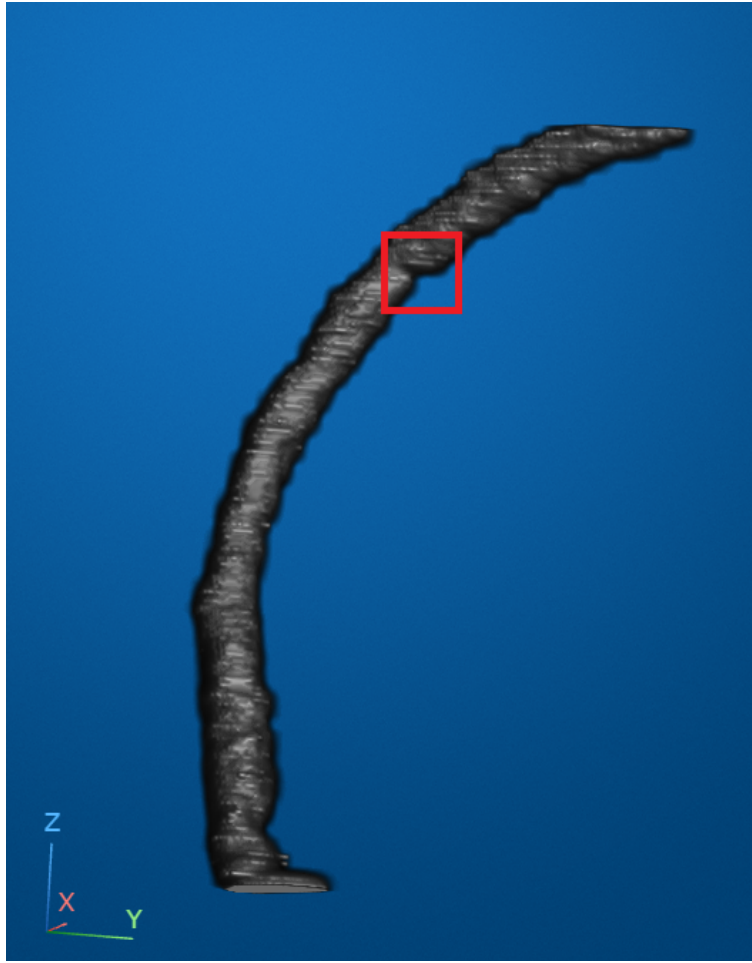


Figure 3.9: An example position to intuitively observe stenosis

After doing the smoothing, we compared the performance of two interpolation methods: Laplacian Pyramid Interpolation and Gaussian Smoothing Interpolation. These methods were applied to enhance the resolution of volume-rendered images of blood vessels. The Structural Similarity Index Measure and Peak Signal-to-Noise Ratio were computed to assess the quality of the interpolated images. SSIM is a perception-based metric that quantifies the similarity in structure between the original and interpolated images, where values closer to 1 indicate higher structural similarity. PSNR, measured in dB, reflects the ratio between the maximum signal power and noise power, where higher values denote better image quality.

The results, including computation time, are presented in Table 3.2.

Table 3.2: Comparison of Laplacian Pyramid and Gaussian Smoothing Interpolation Methods.

Interpolation Method	Average	Average	Elapsed
	SSIM	PSNR (dB)	Time (s)
Laplacian Pyramid	0.9987	41.42	22.80
Gaussian Smoothing	0.9992	42.30	20.52

The results indicate that both interpolation methods achieve high SSIM and PSNR values, demonstrating their effectiveness in enhancing image resolution with minimal loss of structural integrity. However, a few observations can be made regarding their comparative performance. The Gaussian Smoothing Interpolation method achieved a slightly higher average SSIM (0.9992) and PSNR (42.30 dB) compared to the Laplacian Pyramid Interpolation method, which recorded an SSIM of 0.9987 and PSNR of 41.42 dB. This suggests that Gaussian Smoothing is marginally more effective at preserving image details and reducing noise in the interpolated images. In terms of computational efficiency, Gaussian Smoothing Interpolation took slightly less time (20.52 seconds) compared to Laplacian Pyramid Interpolation (22.80 seconds). This difference, while minor, suggests that Gaussian Smoothing might be more suitable for applications requiring faster processing times without compromising image quality. Despite the slightly lower SSIM and PSNR values, Laplacian Pyramid Interpolation is known to maintain edge clarity effectively, especially in applications where preserving fine structural details at varying scales is essential. This method's effectiveness in multi-resolution analysis may be beneficial in specific scenarios, such as detecting small anomalies within the blood vessels. Given that the SSIM and PSNR differences between the two methods are minimal, the choice of method may ultimately depend on the specific requirements of the application. For instance, if preserving fine structural details with edge clarity is

paramount, Laplacian Pyramid Interpolation could be preferred. However, for general resolution enhancement tasks with a balanced focus on quality and efficiency, Gaussian Smoothing Interpolation may be the optimal choice.

3.4 Discussion

This study presents a comprehensive evaluation of state-of-the-art deep learning models and our proposed methods for automated segmentation and 3D reconstruction of the right coronary artery from CT images. The findings highlight significant advancements in both segmentation accuracy and computational efficiency, as well as the practicality of our approaches for clinical applications. The comparative analysis of deep learning models provided critical insights into their strengths and weaknesses. Among the tested models, our proposed network without GhostNet achieved the highest F1 score of 0.897 and an IoU of 0.814, underscoring its superior segmentation performance. This model demonstrated its effectiveness in accurately identifying fine vascular structures, which is crucial for clinical tasks such as identifying stenotic regions. However, this performance comes at the cost of a slightly higher computational demand compared to the GhostNet-enabled version. The inclusion of GhostNet in the proposed model led to a notable improvement in inference speed, achieving the fastest time of 7.03 ms. While this configuration slightly reduced the F1 score to 0.887 and IoU to 0.797, the trade-off is acceptable for real-time applications where speed is critical. This capability is particularly valuable in scenarios such as intraoperative imaging or real-time diagnostics, where rapid processing is essential for effective decision-making.

The transition from 2D segmentation masks to 3D reconstruction represents a critical step in understanding the spatial structure of coronary arteries. The proposed reconstruction process involved extracting the edges of segmented masks, distributing points uniformly, and generating a mesh grid that connects points within and across layers. This approach facilitated the construction of a detailed 3D representation of the RCA, allowing for enhanced visualization of its topology. Volume

rendering further improved the interpretability of the reconstructed RCA. By using surface rendering techniques, the vascular structures were transformed into a continuous and smooth representation, which effectively highlighted morphological features. The rendering provided an intuitive way to identify stenotic regions, aiding in clinical diagnosis and planning. The smoothing techniques, Gaussian Smoothing and Laplacian Pyramid Interpolation, played a key role in refining the visual quality of the 3D models. Gaussian Smoothing achieved higher SSIM (0.9992) and PSNR (42.30 dB) values, along with faster processing times, indicating its superiority for applications requiring both high quality and efficiency. However, Laplacian Pyramid Interpolation excelled in preserving edge clarity and fine structural details, making it a suitable choice for tasks focused on detecting subtle abnormalities. The reconstructed models provided clinicians with a clear visualization of the degree and location of vascular narrowing. This feature enhances the diagnostic process, reducing the likelihood of missed or inaccurate assessments. Moreover, the ability to adjust viewing angles and analyze smoothed vascular images supports more comprehensive evaluations, which could improve patient outcomes.

The clinical value of the 3D reconstruction method significantly impacts both procedural planning and patient care. Surgeons can utilize detailed 3D visualizations to better understand complex anatomical relationships and plan optimal approaches for interventional procedures. This enhanced understanding allows for more precise preoperative planning, potentially reducing procedure times and improving outcomes. The 3D models also serve as powerful tools for patient education, helping clinicians explain conditions and planned procedures more effectively, which can improve patient understanding and compliance. In multidisciplinary settings, these reconstructions enhance communication between different specialists, facilitating more comprehensive treatment planning. The technology also offers valuable opportunities for medical education, allowing trainees to study and practice on accurate representations of patient-specific anatomies. This application in training could accelerate the learning curve for new interventional cardiologists while im-

proving their preparedness for complex cases.

This study, while demonstrating promising results, has several limitations that must be addressed to further validate and enhance the proposed methods:

Firstly, due to the unavailability of diagnostic data for the 32 patients included in the dataset, we were unable to validate the accuracy of stenosis location and stenosis rate predictions. This limitation hinders the ability to directly verify the clinical effectiveness of our method. Future work should aim to integrate diagnostic information to provide a more comprehensive assessment of the method’s accuracy and reliability.

Secondly, given the complexity of CT images, our study focused exclusively on the right coronary artery as it is relatively easier to identify and analyze. This approach excluded the evaluation of stenoses in branching vessels or the left coronary artery, which are equally critical in clinical diagnoses. Addressing these areas in future research would ensure a more comprehensive and versatile method capable of analyzing the full coronary vascular structure.

Thirdly, incorporating the proposed pipeline into real-world clinical workflows requires further testing. Key factors such as noise robustness, compatibility with various imaging modalities, and the user-friendliness of visualization tools need to be evaluated in collaboration with clinicians. These assessments will help ensure that the methods are practical and effective for daily clinical use.

Fourthly, although the computational requirements of our models were optimized, their deployment in resource-constrained environments could still pose challenges. Developing lightweight versions of the models without compromising performance would be essential to facilitate broader adoption, particularly in low-resource settings.

Future work could also explore the application of these methods in related domains, such as the analysis of other vascular regions or the integration with intra-operative imaging systems. Furthermore, incorporating advanced techniques like reinforcement learning or generative adversarial networks could enhance the accu-

racy and adaptability of the models.

3.5 Conclusion

In conclusion, this study demonstrates the potential of the proposed methods for RCA segmentation and 3D reconstruction, providing a robust solution for clinical applications. By achieving a balance between high segmentation accuracy, computational efficiency, and detailed 3D visualization, the methods address critical challenges in coronary artery analysis. With further development and validation, the approaches outlined in this study could significantly contribute to the advancement of diagnostic tools and interventions for coronary artery diseases, ultimately improving patient care.

Chapter 4

Conclusion and Future Work

4.1 General Conclusion

This thesis presented a comprehensive exploration of advanced deep learning methodologies aimed at enhancing the analysis and diagnosis of coronary artery disease. The research included three key approaches, each addressing critical limitations in existing methods:

1. **Hessian-Based Deep Learning Preprocessing Method:**

A novel image preprocessing pipeline was proposed, integrating Hessian matrix filtering with image fusion techniques. This approach enhanced the visualization of vascular structures in coronary angiography, leading to improved detection of stenosis in both main and branch vessels. By addressing challenges such as noise reduction and feature preservation, this method demonstrated a significant improvement in the performance of advanced object detection networks.

2. **YOLO-Based Attention-Enhanced Network:**

A customized object detection network incorporating attention mechanisms was developed to localize coronary artery stenosis efficiently. Experimental results highlighted the potential of attention mechanisms in improving detection accuracy while maintaining computational efficiency, making the ap-

proach suitable for real-time clinical applications.

3. Automated Vessel Extraction and 3D Reconstruction:

An innovative method combining GhostNet and Transformer architectures was introduced for the automated extraction and 3D reconstruction of the right coronary artery from CT images. This method demonstrated state-of-the-art performance in segmenting vascular cross-sections and generating 3D models, providing clinicians with a more comprehensive understanding of the coronary anatomy.

The overarching goal of this research was to address the limitations of existing CAD diagnostic techniques, including reliance on subjective interpretation, suboptimal imaging angles, and challenges in processing complex coronary structures. By leveraging advanced deep learning models and innovative preprocessing strategies, the proposed methods achieved notable improvements in accuracy, robustness, and computational efficiency. The findings of this research have significant implications for clinical practice, particularly in the real-time detection and diagnosis of CAD. The integration of these methods into existing diagnostic workflows has the potential to enhance precision, reduce diagnostic variability, and ultimately improve patient outcomes.

While the presented approaches have advanced the state of CAD diagnosis, several limitations remain. These include the need for further optimization to reduce latency in real-time applications, the potential for overfitting due to dataset limitations, and the requirement for validation across diverse patient populations and imaging systems. Future research should focus on addressing these challenges, exploring the integration of multi-modal imaging data, and refining techniques for broader clinical applicability.

In conclusion, the methods and findings outlined in this study demonstrate the potential to significantly advance diagnostic and therapeutic approaches to coronary artery disease. By improving the quality of diagnostic imaging and enabling more precise 3D reconstructions, these methods provide critical tools for early de-

tection, planning, and intervention. With continued development, validation, and collaboration between research and clinical communities, these advancements could lead to better patient care, reduced diagnostic errors, and improved outcomes in the treatment of coronary artery diseases. Ultimately, this work lays a solid foundation for the future of AI-driven healthcare innovations and their integration into routine medical practice.

4.2 Future Work

Looking ahead, expanding the dataset to include a larger and more diverse range of images is imperative to improve generalizability and address potential imbalances that may impact the model’s performance in real-world scenarios. The inclusion of more varied imaging conditions and patient demographics would ensure the approach is broadly applicable across clinical settings. A critical limitation of the current work is the lack of validation across different medical institutions, which is essential for confirming the model’s robustness across varying imaging protocols, equipment specifications, and clinical practices. This multi-institutional validation would help identify and address potential biases or limitations in the model’s performance across different healthcare settings.

Furthermore, the current study has insufficient testing of edge cases, such as patients with anatomical anomalies, severe calcification, or multiple stents. These challenging scenarios often present the greatest diagnostic difficulty in clinical practice and require thorough validation to ensure reliable performance. Additionally, further refinement of the proposed networks is necessary to enhance their ability to detect stenotic features in highly complex coronary angiography images, particularly in cases involving smaller vessels or intricate anatomical variations. Testing across a broader range of clinical scenarios, including emergency cases and patients with motion artifacts, would provide valuable insights into the model’s limitations and areas requiring improvement.

Incorporating state-of-the-art deep learning architectures and optimizing them

for medical imaging applications will likely yield even greater accuracy and efficiency. The development of automated methods for handling poor image quality and adapting to varying contrast levels would further enhance the system’s clinical utility. Future research should also focus on creating interpretable models that can provide clinicians with clear reasoning for their predictions, enhancing trust and adoption in clinical practice.

Bibliography

- [1] Stefano Cacciatore, Luigi Spadafora, Marco Bernardi, Mattia Galli, Matteo Betti, Francesco Perone, Giulia Nicolaio, Emanuele Marzetti, Anna Maria Martone, Francesco Landi, et al. Management of coronary artery disease in older adults: recent advances and gaps in evidence. *Journal of clinical medicine*, 12(16):5233, 2023.
- [2] Muneeb Ullah, Abdul Wahab, Shahid Ullah Khan, Umber Zaman, Khalil ur Rehman, Shah Hamayun, Muhammad Naeem, Haider Ali, Tayyaba Riaz, Sumbul Saeed, et al. Stent as a novel technology for coronary artery disease and their clinical manifestation. *Current Problems in Cardiology*, 48(1):101415, 2023.
- [3] Pál Maurovich-Horvat, Maria Bosserdt, Klaus F Kofoed, Nina Rieckmann, Theodora Benedek, Patrick Donnelly, José Rodriguez-Palomares, Andrejs Erglis, Cyril Štěchovský, Gintare Šakalyte, et al. Ct or invasive coronary angiography in stable chest pain. *New England Journal of Medicine*, 386(17):1591–1602, 2022.
- [4] Haibo Zhang, Lin Mu, Shuang Hu, Brahmajee K Nallamothu, Alexandra J Lansky, Bo Xu, Georgios Bouras, David J Cohen, John A Spertus, Frederick A Masoudi, et al. Comparison of physician visual assessment with quantitative coronary angiography in assessment of stenosis severity in china. *JAMA internal medicine*, 178(2):239–247, 2018.

- [5] Emmanuel Ovalle-Magallanes, Juan Gabriel Avina-Cervantes, Ivan Cruz-Aceves, and Jose Ruiz-Pinales. Hybrid classical–quantum convolutional neural network for stenosis detection in x-ray coronary angiography. *Expert Systems with Applications*, 189:116112, 2022.
- [6] Viacheslav V Danilov, Kirill Yu Klyshnikov, Olga M Gerget, Anton G Kutikhin, Vladimir I Ganyukov, Alejandro F Frangi, and Evgeny A Ovcharenko. Real-time coronary artery stenosis detection based on modern neural networks. *Scientific reports*, 11(1):7582, 2021.
- [7] Kai Han, Yunhe Wang, Qi Tian, Jianyuan Guo, Chunjing Xu, and Chang Xu. Ghostnet: More features from cheap operations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1580–1589, 2020.
- [8] Ziliang Feng, Ju Zhang, Xusong Ran, Donglu Li, and Chengfang Zhang. Ghost-unet: multi-stage network for image deblurring via lightweight subnet learning. *The Visual Computer*, pages 1–15, 2024.
- [9] Joanne D Schuijf, Matthew B Matheson, Mohammad R Ostovaneh, Armin Arbab-Zadeh, Klaus F Kofoed, Arthur JHA Scholte, Marc Dewey, Chloe Stevenson, Carlos E Rochitte, Kunihiro Yoshioka, et al. Ischemia and no obstructive stenosis (inoca) at ct angiography, ct myocardial perfusion, invasive coronary angiography, and spect: the core320 study. *Radiology*, 294(1):61–73, 2020.
- [10] Kun Pang, Danni Ai, Huihui Fang, Jingfan Fan, Hong Song, and Jian Yang. Stenosis-detnet: Sequence consistency-based stenosis detection for x-ray coronary angiography. *Computerized Medical Imaging and Graphics*, 89:101900, 2021.
- [11] Élodie Labrecque Langlais, Denis Corbin, Olivier Tastet, Ahmad Hayek, Gemina Doolub, Sebastián Mrad, Jean-Claude Tardif, Jean-François Tanguay, Guillaume Marquis-Gravel, Geoffrey H Tison, et al. Evaluation of stenoses using ai

- video models applied to coronary angiography. *npj Digital Medicine*, 7(1):138, 2024.
- [12] Dinis L Rodrigues, Miguel Nobre Menezes, Fausto J Pinto, and Arlindo L Oliveira. Automated detection of coronary artery stenosis in x-ray angiography using deep neural networks. *arXiv preprint arXiv:2103.02969*, 2021.
- [13] Farhad Arefinia, Mehrad Aria, Reza Rabiei, Azamossadat Hosseini, Ali Ghaemian, and Arash Roshanpoor. Non-invasive fractional flow reserve estimation using deep learning on intermediate left anterior descending coronary artery lesion angiography images. *Scientific Reports*, 14(1):1818, 2024.
- [14] Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, and Max A Viergever. Multiscale vessel enhancement filtering. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI’98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings 1*, pages 130–137. Springer, 1998.
- [15] Renzhe Bi, US Dinish, Chi Ching Goh, Toru Imai, Mohesh Moothanchery, Xiuting Li, Jin Young Kim, Seungwan Jeon, Yang Pu, Chulhong Kim, et al. In vivo label-free functional photoacoustic monitoring of ischemic reperfusion. *Journal of biophotonics*, 12(7):e201800454, 2019.
- [16] Mingxu Huang, Chaolu Feng, and Dazhe Zhao. An improved method of blood vessel enhancement based on hessian matrix. In *The Fourth International Symposium on Image Computing and Digital Medicine*, pages 197–200, 2020.
- [17] Anna Orlova, Marina Sirotkina, Ekaterina Smolina, Vadim Elagin, Andrey Kovalchuk, Ilya Turchin, and Pavel Subochev. Raster-scan optoacoustic angiography of blood vessel development in colon cancer models. *Photoacoustics*, 13:25–32, 2019.

- [18] Yanjun Li, Takaaki Yoshimura, Yuto Horima, and Hiroyuki Sugimori. A preprocessing method for coronary artery stenosis detection based on deep learning. *Algorithms*, 17(3):119, 2024.
- [19] Shan Wang, Bonian Li, and Shoujun Zhou. A segmentation method of coronary angiograms based on multi-scale filtering and region-growing. In *2012 International Conference on Biomedical Engineering and Biotechnology*, pages 678–681. IEEE, 2012.
- [20] Xiliang Zhu, Zhaoyun Cheng, Sheng Wang, Xianjie Chen, and Guoqing Lu. Coronary angiography image segmentation based on pspnet. *Computer Methods and Programs in Biomedicine*, 200:105897, 2021.
- [21] Majd Zreik, Robbert W Van Hamersvelt, Jelmer M Wolterink, Tim Leiner, Max A Viergever, and Ivana Išgum. A recurrent cnn for automatic detection and classification of coronary artery plaque and stenosis in coronary ct angiography. *IEEE transactions on medical imaging*, 38(7):1588–1598, 2018.
- [22] Emmanuel Ovalle-Magallanes, Juan Gabriel Avina-Cervantes, Ivan Cruz-Aceves, and Jose Ruiz-Pinales. Transfer learning for stenosis detection in x-ray coronary angiography. *Mathematics*, 8(9):1510, 2020.
- [23] Robert Avram, Jeffrey E Olgin, Zeeshan Ahmed, Louis Verreault-Julien, Alvin Wan, Joshua Barrios, Sean Abreau, Derek Wan, Joseph E Gonzalez, Jean-Claude Tardif, et al. Cathai: fully automated coronary angiography interpretation and stenosis estimation. *npj Digital Medicine*, 6(1):142, 2023.
- [24] Akram Isavand Rahmani, Hesam Akbari, and Somayeh Saraf Esmaili. Retinal blood vessel segmentation using gabor filter and morphological reconstruction. *Signal Processing and Renewable Energy*, 4(1):77–88, 2020.
- [25] Yongqing Qian, Zhicheng Wang, Liang Chen, and Zhenghua Huang. Vascular enhancement with structure preservation from noisy x-ray angiogram images by employing non-local hessian-based filter. *Optik*, 232:166523, 2021.

- [26] Xiaohong Chen, Jufeng Jiang, and Xiaofeng Zhang. Automatic 3d coronary artery segmentation based on local region active contour model. *Journal of Thoracic Disease*, 16(4):2563, 2024.
- [27] Rohan Banerjee, Avik Ghose, and Kayapanda Muthana Mandana. A hybrid cnn-lstm architecture for detection of coronary artery disease from ecg. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2020.
- [28] Yanan Wu, Shouliang Qi, Meihuan Wang, Shuiqing Zhao, Haowen Pang, Jiaxuan Xu, Long Bai, and Hongliang Ren. Transformer-based 3d u-net for pulmonary vessel segmentation and artery-vein separation from ct images. *Medical & Biological Engineering & Computing*, 61(10):2649–2663, 2023.
- [29] Guangpu Wang, Peng Zhou, Hui Gao, Zewei Qin, Shuo Wang, Jinglai Sun, and Hui Yu. Coronary vessel segmentation in coronary angiography with a multi-scale u-shaped transformer incorporating boundary aggregation and topology preservation. *Physics in Medicine & Biology*, 69(2):025012, 2024.
- [30] Qianjin Wang, Lisheng Xu, Lu Wang, Xiaofan Yang, Yu Sun, Benqiang Yang, and Stephen E Greenwald. Automatic coronary artery segmentation of ccta images using unet with a local contextual transformer. *Frontiers in Physiology*, 14:1138257, 2023.
- [31] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-fcn: Object detection via region-based fully convolutional networks. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.
- [32] Tausif Diwan, G Anirudh, and Jitendra V Tembhurne. Object detection using yolo: Challenges, architectural successors, datasets and applications. *multimedia Tools and Applications*, 82(6):9243–9275, 2023.

- [33] Juan Terven, Diana-Margarita Córdova-Esparza, and Julio-Alejandro Romero-González. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Machine Learning and Knowledge Extraction*, 5(4):1680–1716, 2023.
- [34] Chien-Yao Wang, I-Hau Yeh, and Hong-Yuan Mark Liao. Yolov9: Learning what you want to learn using programmable gradient information, 2024.
- [35] Ao Wang, Hui Chen, Lihao Liu, Kai Chen, Zijia Lin, Jungong Han, and Guiguang Ding. Yolov10: Real-time end-to-end object detection. *arXiv preprint arXiv:2405.14458*, 2024.
- [36] Yian Zhao, Wenyu Lv, Shangliang Xu, Jinman Wei, Guanzhong Wang, Qingqing Dang, Yi Liu, and Jie Chen. Detrs beat yolos on real-time object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16965–16974, 2024.
- [37] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020.
- [38] Zhen Qin, Weixuan Sun, Hui Deng, Dongxu Li, Yunshen Wei, Baohong Lv, Junjie Yan, Lingpeng Kong, and Yiran Zhong. cosformer: Rethinking softmax in attention. *arXiv preprint arXiv:2202.08791*, 2022.
- [39] Dongchen Han, Tianzhu Ye, Yizeng Han, Zhuofan Xia, Siyuan Pan, Pengfei Wan, Shiji Song, and Gao Huang. Agent attention: On the integration of softmax and linear attention. In *European Conference on Computer Vision*, pages 124–140. Springer, 2025.
- [40] Md Mostafijur Rahman, Mustafa Munir, and Radu Marculescu. Emcad: Efficient multi-scale convolutional attention decoding for medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11769–11779, 2024.

- [41] Shibiao Xu, Shuchen Zheng, Wenhao Xu, Rongtao Xu, Changwei Wang, Jiguang Zhang, Xiaoqiang Teng, Ao Li, and Li Guo. HCF-Net: Hierarchical Context Fusion Network for Infrared Small Object Detection . In *2024 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, Los Alamitos, CA, USA, July 2024. IEEE Computer Society.
- [42] Natraj Katta, Troy Loethen, Carl J Lavie, and Martin A Alpert. Obesity and coronary heart disease: epidemiology, pathology, and coronary artery imaging. *Current problems in cardiology*, 46(3):100655, 2021.
- [43] Ergashov Behro’zjon Komilovich. Coronary artery disease. *EUROPEAN JOURNAL OF MODERN MEDICINE AND PRACTICE*, 3(12):81–87, 2023.
- [44] John Smith and Jane Doe. Ct as a non-invasive alternative to coronary angiography: Advances and clinical applications. *Journal of Cardiovascular Imaging*, 15(3):123–135, 2021.
- [45] Patrick W Serruys, Hironori Hara, Scot Garg, Hideyuki Kawashima, Bjarne L Nørgaard, Marc R Dweck, Jeroen J Bax, Juhani Knuuti, Koen Nieman, Jonathon A Leipsic, et al. Coronary computed tomographic angiography for complete assessment of coronary artery disease: Jacc state-of-the-art review. *Journal of the American College of Cardiology*, 78(7):713–736, 2021.
- [46] Andrew D Choi, Hugo Marques, Vishak Kumar, William F Griffin, Habib Rahban, Ronald P Karlsberg, Robert K Zeman, Richard J Katz, and James P Earls. Ct evaluation by artificial intelligence for atherosclerosis, stenosis and vascular morphology(clarify): a multi-center, international study. *Journal of Cardiovascular Computed Tomography*, 15(6):470–476, 2021.
- [47] Bo Wang, Shuo Jin, Qingsen Yan, Haibo Xu, Chuan Luo, Lai Wei, Wei Zhao, Xuexue Hou, Wenshuo Ma, Zhengqing Xu, et al. Ai-assisted ct imaging analysis for covid-19 screening: Building and deploying a medical ai system. *Applied soft computing*, 98:106897, 2021.

- [48] Matthew H Lee, Ryan Zea, John W Garrett, Peter M Graffy, Ronald M Summers, and Perry J Pickhardt. Abdominal ct body composition thresholds using automated ai tools for predicting 10-year adverse outcomes. *Radiology*, 306(2):e220574, 2022.
- [49] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *Lecture Notes in Computer Science*, pages 234–241. Springer, 2015.
- [50] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.
- [51] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018.
- [52] Sijing Cai, Yunxian Tian, Harvey Lui, Haishan Zeng, Yi Wu, and Guannan Chen. Dense-unet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network. *Quantitative Imaging in Medicine and Surgery*, 10(6), 2020.
- [53] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 424–432. Springer, 2016.
- [54] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong

encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.

- [55] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022.
- [56] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [57] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10012–10022, 2021.
- [58] Majd Zreik, Nikolas Lessmann, Robbert W. van Hamersvelt, Jelmer M. Wolterink, Michiel Voskuil, Max A. Viergever, Tim Leiner, and Ivana Išgum. A recurrent CNN for automatic detection and classification of coronary artery plaques in cardiac CT angiography. *IEEE Transactions on Medical Imaging*, 38(7):1588–1598, 2019.
- [59] Jelmer M. Wolterink, Tim Leiner, Max A. Viergever, and Ivana Išgum. 3D fully convolutional networks for cardiac segmentation in CT scans. *Computers in Biology and Medicine*, 104:95–102, 2019.
- [60] Xiaofeng Yang, Guanghui Zeng, and Yujie Liu. Attention-gated networks for coronary artery segmentation from CT images. *Journal of Digital Imaging*, 33(4):989–998, 2020.

- [61] Thi Diem Tran, Ngoc Quoc Tran, Vu Trung Duong Le, and Hoai Luan Pham. Efficient ecg classification with light weight shuffle ghostnet architecture. In *2023 International Conference on Advanced Technologies for Communications (ATC)*, pages 421–426, 2023.
- [62] Masahiro Sato, Akira Tamura, Shigeru Motoyama, and Shinya Kondo. The impact of 3d coronary artery imaging on diagnostic accuracy and treatment planning in coronary artery disease. *Journal of Cardiovascular Computed Tomography*, 13(4):181–188, 2019.
- [63] Muneeb Ullah, Ayisha Bibi, Abdul Wahab, Shah Hamayun, Mahboob Ur Rehman, Shahid Ullah Khan, Uzma Azeem Awan, Muhammad Naeem, Sumbul Saeed, Talib Hussain, et al. Shaping the future of cardiovascular disease by 3d printing applications in stent technology and its clinical outcomes. *Current Problems in Cardiology*, 49(1):102039, 2024.
- [64] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.
- [65] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*, 39(6):1856–1867, 2019.
- [66] Narinder Singh Punj and Sonali Agarwal. Rca-iunet: a residual cross-spatial attention-guided inception u-net model for tumor segmentation in breast ultrasound imaging. *Machine Vision and Applications*, 33(2):27, 2022.
- [67] Jeya Maria Jose Valanarasu, Poojan Oza, Ilker Hacihaliloglu, and Vishal M Patel. Medical transformer: Gated axial-attention for medical image segmentation. In *Medical image computing and computer assisted intervention–MICCAI*

2021: 24th international conference, Strasbourg, France, September 27–October 1, 2021, proceedings, part I 24, pages 36–46. Springer, 2021.

- [68] M Travis Caton Jr, Walter F Wiggins, and Diego Nunez. Three-dimensional cinematic rendering to optimize visualization of cerebrovascular anatomy and disease in ct angiography. *Journal of Neuroimaging*, 30(3):286–296, 2020.
- [69] Marc Levoy. Display of surfaces from volume data. *IEEE Computer Graphics and Applications*, 8(3):29–37, 1988.
- [70] Barthold Lichtenbelt, Randy Crane, and Shaz Naqvi. *Introduction to Volume Rendering*. Prentice Hall, 1998.
- [71] Han Xu, Jiayi Ma, Junjun Jiang, Xiaojie Guo, and Haibin Ling. U2fusion: A unified unsupervised image fusion network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):502–518, 2020.
- [72] Kannadhasan Suriyan, Nagarajan Ramaingam, Sudarmani Rajagopal, Jeevitha Sakkarai, Balakumar Asokan, and Manjunathan Alagarsamy. Performance analysis of peak signal-to-noise ratio and multipath source routing using different denoising method. *Bulletin of Electrical Engineering and Informatics*, 11(1):286–292, 2022.

業績リスト

1. 著書

なし

2. 学会誌又は学術雑誌への論文掲載

I. 論文発表

1. A preprocessing method for coronary artery stenosis detection based on deep learning. Li Y, Yoshimura T, Horima Y, Sugimori H. Algorithms 2024, 17(3), 119.
2. A Hessian-Based Deep Learning Preprocessing Method for Coronary Angiography Image Analysis. Li Y, Yoshimura T, Horima Y, Sugimori H. Electronics 2024, 13(18), 3676.

II. 口頭発表

1. A preprocessing method for coronary artery stenosis detection. YanJun Li, Takaaki Yoshimura, Yuto Horima, Hiroyuki Sugimori, 2023FHS, Sapporo, October, 2023.

3. 総説・解説

なし

4. 学会賞・学術賞の受賞

なし

5. その他

なし