

Article

Not peer-reviewed version

---

# Morphology-Integrated Segmentation for Coronary Interventional Imaging

---

[Yue Du](#), [Zhenxi Zhang](#), [Hongzhen Chen](#), [Hongxiang Zhou](#), [Yifei Han](#), [Ziqiao Qu](#), [Bin Li](#)<sup>\*</sup>, [Shoujun Zhou](#)<sup>\*</sup>

Posted Date: 22 July 2025

doi: 10.20944/preprints202507.1854.v1

Keywords: multi-scale feature fusion; attention mechanism; guidewire segmentation; coronary vessel segmentation; deep learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Morphology-Integrated Segmentation for Coronary Interventional Imaging

Yue Du<sup>1,2,†</sup>, Zhenxi Zhang<sup>1,2,†</sup>, Hongzhen Chen<sup>1,†</sup>, Hongxiang Zhou<sup>1</sup>, Yifei Han<sup>1,2</sup>, Ziqiao Qu<sup>1</sup>, Bin Li<sup>1,2,\*</sup> and Shoujun Zhou<sup>1,\*</sup>

<sup>1</sup> Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Nanshan District, Shenzhen, 518055, China

<sup>2</sup> Linying Medical Technology (Shenzhen) Co., LTD, Guangdong 518055

\* Correspondence

† These authors contributed equally to this work: Yue Du (Co-first, ranked 1st), Zhenxi Zhang (Co-first, ranked 2nd), Hongzhen Chen (Co-first, ranked 3rd).

‡ This work was supported by the Key-Area Research and Development Program of Guangdong Province (No. 2025B1111020001), in part by the Natural Science Foundation of Guangdong Province (No. 2023A1515010673), in part by the Shenzhen Science and Technology Innovation Bureau key project (No. JSGG20220831110400001, No. CJGJZD20230724093303007, No. KJZD20240903101259001), in part by Shenzhen Medical Research Fund (No. D2404001), in part by Shenzhen Engineering Laboratory for Diagnosis & Treatment Key Technologies of Interventional Surgical Robots (No. XMHT20220104009), and the Key Laboratory of Biomedical Imaging Science and System, CAS, for the Research platform support.

## Abstract

Robotic-assisted percutaneous coronary intervention (R-PCI) represents a significant technological advancement in the field of cardiovascular intervention. The combined segmentation of guidewires and vessels serves as the perceptual foundation for R-PCI systems, providing critical anatomical and instrumental positioning data to the robotic platform. Coronary interventional imaging relies on two core modalities: (1) 2D DSA (Digital Subtraction Angiography) acquired during contrast opacification for precise vessel mapping, and (2) 2D X-ray fluoroscopy captured after contrast clearance for real-time instrument tracking. However, both coronary vessels and guidewires typically exhibit slender, elongated linear structures that are susceptible to imaging challenges such as noise, low contrast, motion artifacts, and complex anatomical backgrounds, severely hindering segmentation accuracy and structural continuity. To address these challenges, we propose a coronary intervention image segmentation method with integrated morphological features, specifically two task-specific deep neural network architectures. For guidewire segmentation, we introduce GS-UNet (Guidewire Segmentation U-Net), an enhanced U-Net-based model tailored for detecting fine linear structures. GS-UNet incorporates a line-sensitive enhancement module and a guided feature augmentation mechanism, significantly improving the continuity and completeness of guidewire segmentation. The model is extensively validated on a self-constructed guidewire dataset, achieving superior performance in maintaining connectivity and accurately identifying guidewire endpoints. Building upon the architectural insights from GS-UNet, we further propose MSFNet (Multi-Scale Feature Fusion Network) for coronary vessel segmentation. MSFNet integrates multi-scale feature fusion modules and a composite attention mechanism to capture local fine details and global vascular topology jointly. Moreover, it enhances robustness against modality variation and complex backgrounds. Extensive experiments conducted on the authoritative ARCADE dataset demonstrate that MSFNet consistently outperforms state-of-the-art methods in terms of vessel continuity, boundary precision, and detection of fine vascular branches, highlighting its strong potential for real-world clinical applications. The morphology-integrated segmentation framework proposed in this study effectively enhances guidewire-vessel co-segmentation precision, enabling more precise anatomical localization for R-PCI navigation in complex coronary interventions.

**Keywords:** multi-scale feature fusion; attention mechanism; Guidewire segmentation; Coronary vessel segmentation; deep learning

## 1. Introduction

Percutaneous Coronary Intervention represents a cornerstone technique in cardiovascular medicine [1], utilizing real-time imaging guidance and minimally invasive catheter-based delivery systems to diagnose and treat coronary lesions with high precision [2,3]. This approach substantially reduces surgical trauma, accelerates patient recovery, and improves long-term clinical outcomes, thereby establishing itself as a primary therapeutic strategy for coronary artery disease. However, conventional coronary interventional techniques exhibit several significant limitations: (1) Prolonged exposure to X-ray radiation during procedures poses substantial health risks to physicians, even with protective measures in place [4]; (2) Current techniques demand highly specialized skills and extensive experience, resulting in a steep learning curve and prolonged training periods [5]; (3) Operator fatigue during lengthy procedures may lead to instability in manual operations and potential errors, directly compromising surgical outcomes and patient quality of life [6]. These clinical challenges impose a substantial burden on coronary interventional practice.

Robotic-assisted percutaneous coronary intervention (R-PCI) represents a significant technological advancement in the field of cardiovascular intervention [7]. Compared to traditional manual percutaneous coronary intervention (M-PCI), R-PCI demonstrates substantial advantages in operational precision, radiation protection, procedural standardization, and optimization of the learning curve [8]. The combined segmentation of guidewires and vessels serves as the perceptual foundation for R-PCI systems, providing critical anatomical and instrumental positioning data to the robotic platform. Given the radiolucent nature of vascular tissues, the standard clinical protocol involves intermittent injection of radiopaque contrast agents [9]: (1) During the contrast-filling phase, digital subtraction angiography (DSA) provides complete vascular morphological information on fluoroscopic images; (2) In the contrast-free phase, interventional instruments (e.g., guidewires and stents) become visible under X-ray, enabling real-time device positioning by operators. However, existing segmentation methods demonstrate three prevalent limitations when applied to this intermittent imaging paradigm: (1) Insufficient tubular feature extraction, where conventional convolution kernels cause discontinuities in slender structures; (2) Poor dynamic adaptability, with significant accuracy fluctuations due to contrast/device variations and reliance on manual motion compensation; (3) Limited clinical practicality, where algorithm latency exceeds real-time requirements.

To address these issues, we propose GS-UNet, a dedicated segmentation network tailored for guidewire segmentation, which demonstrates robust connectivity and structural integrity on a specialized guidewire dataset. Building upon this, we design a Multi-Scale Feature Fusion Network (MSFNet) specifically for coronary vessel segmentation, evaluated on the benchmark ARCADE dataset [10]. MSFNet enhances connectivity and detail preservation by fusing multi-scale features and modeling both local and global contexts. Comprehensive experiments demonstrate that MSFNet outperforms several state-of-the-art methods across multiple metrics, especially excelling in fine vessel identification, boundary preservation, and structural connectivity. As shown in Figure 1, the combined use of GS-UNet and MSFNet offers a robust and comprehensive segmentation solution for coronary interventional procedures. Moreover, as shown in Figure 2, MSFNet achieves superior segmentation performance while maintaining lower model complexity in terms of parameter count and computational cost compared to other segmentation models.

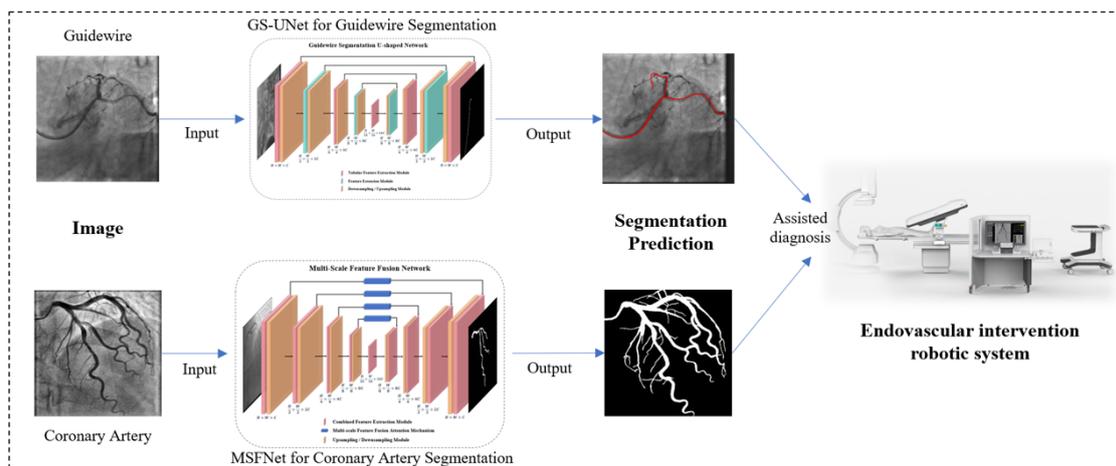


Figure 1. Overall architecture flow about guidewire and coronary artery segmentation.

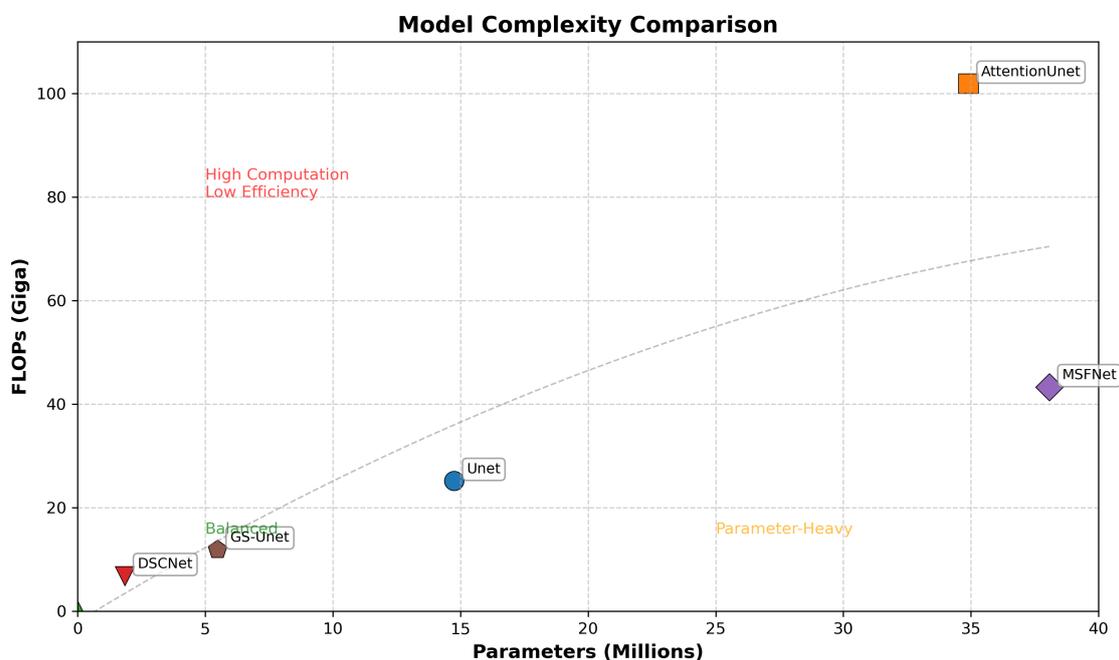


Figure 2. Visualization of model complexity for all comparison models, GS-UNet, and MSFNet.

The main contributions of our work are summarized as follows:

- To accurately identify the elongated morphological structures of guidewires and improve structural integrity, we design a Tubular Feature Extraction Module (TFEM) that significantly enhances the model's ability to capture tubular features. Building upon this, we develop a Combined Feature Extraction Module (CFEM) tailored to the elongated and tortuous characteristics of vascular morphology. This module combines local detail perception with global structural modeling capabilities, effectively strengthening the response to small-scale vessels and multi-branch structures.
- We propose a Feature Extension Module (FEM) that, while reducing the overall computational complexity of the model, enhances the network's feature generalization ability, enabling robust discrimination of guidewires with varying shapes and positions. Based on this module, for the task of vessel segmentation, we further design a Feature Generalization Enhancement Module (FGEM) that effectively improves the model's generalization capacity to image features, enhancing robustness across different imaging modalities and complex backgrounds.

- We introduce a multi-scale feature fusion attention mechanism (MFFA), which aggregates rich spatial contextual information from multiple receptive field scales between the encoder and decoder. Through attention-guided weighting, the model adaptively emphasizes regions relevant to coronary vessels in the image, thereby enhancing attention to fine-grained details and target areas. This leads to more accurate segmentation of vascular regions.

The remainder of this paper is organized as follows. Section 2 reviews traditional and deep learning-based methods for coronary artery segmentation. Section 3 presents our proposed framework in detail, including architectural innovations. Section 4 describes the datasets, implementation details, comprehensive experiments, and ablation studies. Section 5 discusses the limitations of our method. Finally, Section 6 concludes the study.

## 2. Related Work

### 2.1. Traditional Methods for Guidewires and Coronary Artery Segmentation

Before the widespread adoption of deep learning, the segmentation of guidewires and coronary arteries primarily relied on classical image processing techniques and traditional machine learning approaches [11,12]. Common methods include thresholding [13], edge detection [14], region growing [15,16], morphological operations [17], and conventional machine learning algorithms [18]. These techniques typically depend on handcrafted image features and heuristic rules to extract elongated tubular structures from medical images.

Thresholding is one of the earliest segmentation techniques, which separates vessel or guidewire regions from the background by setting intensity thresholds. For instance, the Otsu method automatically determines an optimal threshold based on the image histogram [19]. However, it performs poorly in scenarios with intensity inhomogeneity, low contrast, or when the guidewire or vessel shares similar grayscale values with the background. Edge detection methods use image gradient information to extract structural boundaries, with common operators including Canny, Sobel, and Prewitt [20]. While effective when edges are clear, these techniques are highly sensitive to noise and often fail to capture fine structures like thin guidewires or distal vessels, resulting in discontinuous or broken edges. Region growing methods start from seed points and expand regions based on intensity or texture similarity, such as Region Growing and Fast Marching algorithms [21]. However, these methods strongly depend on the choice of initial seeds and growth criteria, making them prone to over-segmentation or missed detection in the presence of tortuous branches or complex anatomy. Morphological operations [22] (e.g., erosion, dilation, opening, skeletonization) are widely used to refine vessel and guidewire structures or connect fragmented regions. Although they improve edge continuity, they rely heavily on manually designed structuring elements and show limited adaptability to varying anatomical shapes. Traditional machine learning methods, such as k-nearest neighbors (KNN) and support vector machines (SVM), typically classify vessel and guidewire regions based on handcrafted features like local texture, grayscale variation, or edge gradients. Compared with purely rule-based approaches, these models offer some robustness, but still suffer from limited generalization due to their reliance on feature quality, image conditions, and annotated training data.

In summary, while traditional methods play a foundational role in early guidewire and coronary artery segmentation research, they continue to face challenges in accurately and robustly handling elongated structures, maintaining connectivity, and adapting to diverse imaging modalities or complex backgrounds.

### 2.2. Deep Learning Methods for Guidewires and Coronary Artery Segmentation

With the rapid advancement of deep learning, particularly its success in computer vision tasks, deep neural network-based methods have become the dominant approach for segmenting guidewires and coronary arteries [23]. Compared with traditional techniques, deep learning offers stronger feature representation capabilities and enables end-to-end optimization, significantly improving segmentation accuracy and robustness.

Convolutional Neural Networks (CNNs) [24] represent one of the earliest deep learning models applied to medical image segmentation, capable of automatically extracting local spatial features [25]. Classical architectures such as VGG [26] and ResNet [27] are used to encode textural and boundary features of vascular and guidewire structures. However, their limited receptive fields restrict their ability to capture long-range dependencies and model complex topologies such as thin, curved guidewires and multi-branch vessels. The U-Net architecture [28], a milestone in medical segmentation, adopts an encoder-decoder structure with skip connections to preserve spatial information and integrate multi-scale features, showing excellent performance in segmenting fine vascular and guidewire structures. Its variants—U-Net++ [29], Attention U-Net [30], and ResU-Net [31]—introduce nested skip paths, attention modules, and residual connections, further enhancing the model’s ability to detect small and complex structures under challenging conditions. Attention mechanisms are widely adopted to help the network focus on task-relevant information. For example, the Squeeze-and-Excitation (SE) [32] and CBAM (Convolutional Block Attention Module) [33] modules dynamically reweight channel and spatial features, allowing the network to emphasize vessel and guidewire regions while suppressing irrelevant background noise. In recent years, Transformer-based models have gained popularity in medical image analysis due to their ability to model global dependencies [34]. TransUNet [35] integrates Vision Transformer (ViT) modules into a U-Net-like structure, combining local convolution with global attention to improve segmentation precision. Swin-UNet [36] introduces a shifted window attention mechanism to enhance multi-scale representation and global context modeling.

In summary, deep learning methods—particularly those integrating multiple architectural innovations—demonstrate strong performance in guidewire and coronary artery segmentation. However, challenges such as limited annotated datasets, model interpretability, and high computational costs persist [37]. Future work may focus on lightweight model design, domain adaptation, few-shot learning, and multimodal data fusion to enhance the clinical applicability of these models.

### 3. Methodology

We sequentially introduce our proposed GS-UNet network for guidewire segmentation and the MSFNet network for coronary artery segmentation.

#### 3.1. Overall Architecture of the Guidewire Segmentation Network

As shown in Figure 3, the overall framework of our proposed Guidewire Segmentation Network (GS-UNet) consists of three parts: the encoder stage, the decoder stage, and the skip connection stage. Specifically, in GS-UNet, the Tubular Feature Extraction Module (TFEM), Feature Extension Module (FEM), and Sampling Module are designed for feature extraction, feature enhancement and generalization, and feature sampling within the encoder-decoder pathway.

##### 3.1.1. Tubular Feature Extraction Module in Guidewire Segmentation Network

The structure of the Tubular Feature Extraction Module (TFEM) designed in this chapter is shown in Figure 4. The module is mainly based on a CNN architecture [24], integrating the scaling architecture and dynamic snake convolution [38]. In particular, large-kernel depthwise separable convolutions are used to provide a larger receptive field to effectively capture long-range dependencies in the image.

Depthwise Separable Convolution [39] is a convolutional structure that effectively reduces the computational complexity and number of parameters in convolutional neural networks. It mainly decomposes the standard convolution operation into two independent steps: depthwise convolution and pointwise convolution. Standard convolution jointly computes all input channels with the convolution kernel, resulting in high computational costs, whereas depthwise separable convolution breaks this process down.

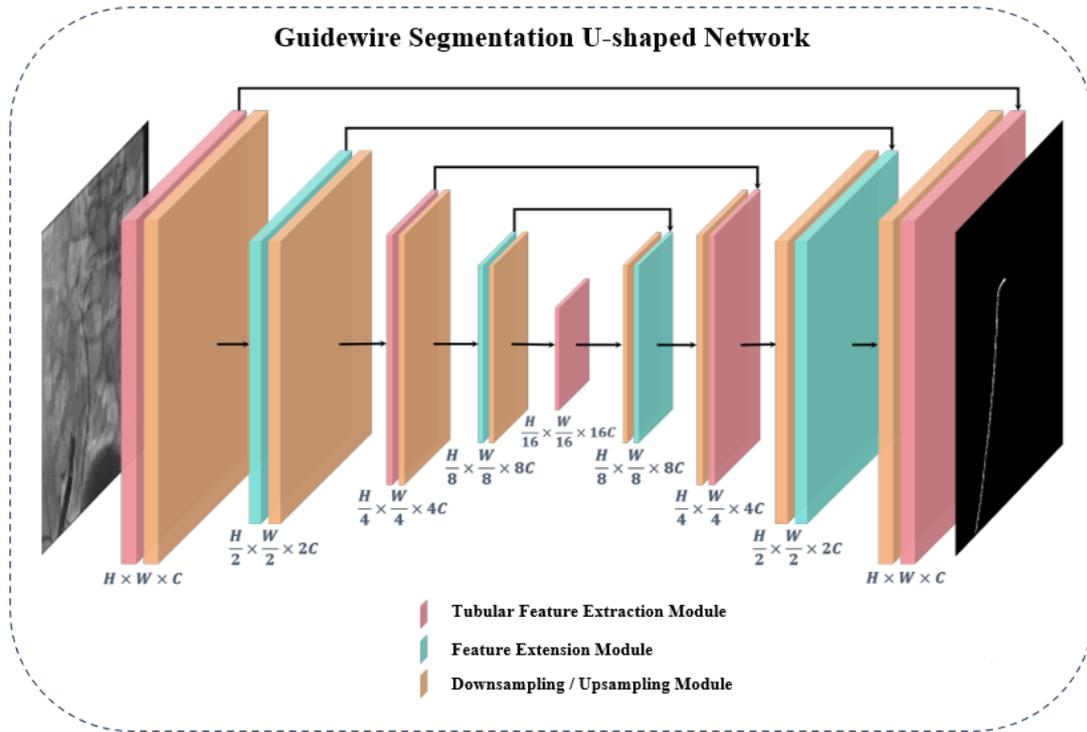


Figure 3. Illustration of the overall architecture of GS-UNet.

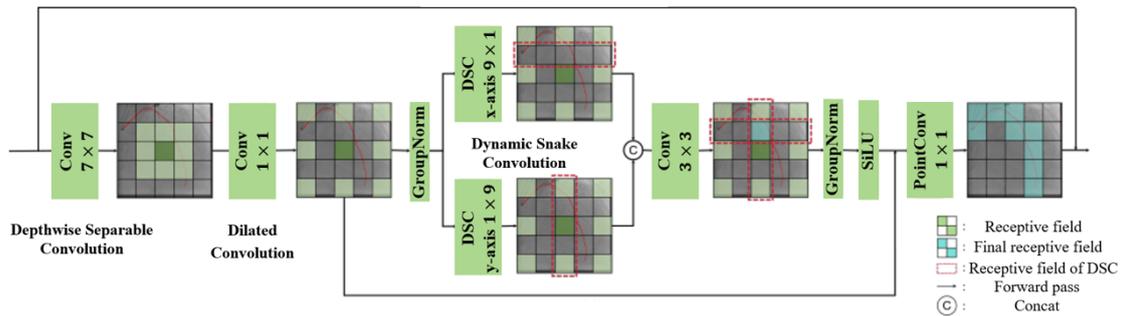


Figure 4. Illustration of the Tubular Feature Extraction Module.

In depthwise convolution, each input channel is convolved with its independent convolution kernel without channel fusion. Suppose the depthwise convolution kernel size is  $K \times K$ , and the number of input channels is  $C$ ; then the computation process can be expressed as:

$$Z_m(i, j) = \sum_{u=-k}^k \sum_{v=-k}^k K_{u,v,m}^{(d)} \cdot X_m(i+u, j+v), m = 1, 2, \dots, C_{in}, \quad (1)$$

where  $Z_m(i, j)$  represents the output value of the  $m$ -th channel after the depthwise convolution at position  $(i, j)$ ;  $X_m(i+u, j+v)$  denotes the pixel value at position  $(i+u, j+v)$  on the  $m$ -th input channel;  $u$  and  $v$  are the spatial indices of the convolution kernel, controlling the offset of the sliding window in the spatial dimensions. Since the output still has  $C_{in}$  channels, the spatial structural features of each channel are preserved. Pointwise convolution uses a  $1 \times 1$  convolution kernel to perform a linear combination along the channel dimension, enabling cross-channel information fusion. Compared to standard convolution, it greatly reduces the computational cost and number of parameters while maintaining a similar receptive field.

In the dynamic snake convolution, the convolution kernel is unfolded into linear forms along the x-axis and y-axis. Given a kernel size of  $K$ , an original kernel center index  $i$ , and a learned offset  $c$ , the cumulative offset along the x-axis can be formulated as:

$$K_{i\pm c} = \begin{cases} (x_{i+c}, y_{i+c}) = \left( x_i + c, y_i + \sum_{i}^{i+c} \Delta y \right) \\ (x_{i-c}, y_{i-c}) = \left( x_i - c, y_i + \sum_{i}^{i-c} \Delta y \right) \end{cases} . \quad (2)$$

Accordingly, the cumulative offset along the y-axis can be expressed as:

$$K_{j\pm c} = \begin{cases} (x_{j+c}, y_{j+c}) = \left( x_j + \sum_{j}^{j+c} \Delta x, y_j + c \right) \\ (x_{j-c}, y_{j-c}) = \left( x_j + \sum_{j}^{j-c} \Delta x, y_j - c \right) \end{cases} . \quad (3)$$

In the TFEM module, assuming the input feature map is  $X$ , it is first passed through a  $7 \times 7$  depthwise convolution layer for global feature extraction. Considering the slender shape of the guidewire, the processed feature map is then input into an expansion convolution layer with a dilation rate of 4 and a kernel size of  $1 \times 1$ . This enlarges the receptive field by 4 times, allowing the layer to increase the receptive field while keeping computational cost limited. The resulting feature map is denoted as  $X'$ . Next, to extract the tubular features of the guidewire,  $X'$  is split into two parts,  $X'_x$  and  $X'_y$ , representing the feature maps for tubular feature extraction along the x-axis and y-axis, respectively. These are processed by dynamic serpentine convolutions with kernel sizes  $9 \times 1$  and  $1 \times 9$  along the x and y directions. The outputs are then concatenated.

After that, a  $3 \times 3$  convolution layer is applied for feature fusion, enabling the module to fully capture the guidewire's tubular features across the entire 2D plane. Finally, the SiLU activation function is applied, followed by a pointwise convolution with a  $1 \times 1$  kernel to produce the output feature map, which retains the same size as the input feature map. The main operations of the TFEM module are as follows:

$$Y = Conv(C(F(X'_x), F(X'_y))), \quad (4)$$

where  $Conv$  denotes the convolution operation,  $C$  represents concatenation, and  $F(X'_i)$  denotes the dynamic deformable convolution along the  $i$ -axis (where  $i$  is either  $x$  or  $y$ ).

### 3.1.2. Feature Extension Module in Guidewire Segmentation Network

Considering the relatively high computational cost of deformable convolutions, we specifically design a Feature Extension Module (FEM), which is symmetrically and alternately placed with the TFEM blocks within the U-shaped network. This arrangement aims to enhance the network's feature generalization capability while reducing the overall computational load.

Specifically, the structure of the FEM module is shown in Figure 5. Since matrix multiplication can significantly enhance the generalization ability of the model without substantially altering its overall structure, the FEM block integrates matrix multiplication with a convolutional scaling architecture to construct the feature extension component. The upper-level structure of the FEM module is similar to that of the TFEM module, consisting of a depthwise convolution layer and an extension layer of the same size. The input feature map  $X$  of the model is first processed by a  $7 \times 7$  depthwise convolution layer for feature extraction and group normalization, transforming it into  $X'$ . It is then split into two branches, each fed into a  $1 \times 1$  expansion convolution layer with a dilation rate of 4. One branch applies the GELU activation function after the expansion convolution, while the other remains unactivated. The outputs of the two branches are then multiplied using matrix multiplication, expanding the spatial dimensions and enabling the module to reconstruct more diverse feature representations. Finally, the resulting feature map is passed through another  $7 \times 7$  depthwise convolution to ensure the output feature map maintains the same dimensionality as the input  $X$ .

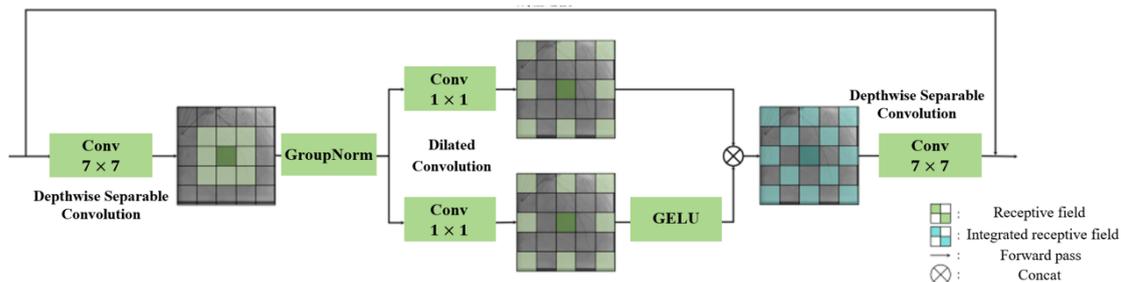


Figure 5. Illustration of the Feature Extension Module.

The main processing flow of the Feature Extension Module (FEM) is as follows:

$$Y = D(\sigma(X'_1) * X'_2), \quad (5)$$

where  $D$  denotes depthwise convolution,  $\sigma$  represents the GELU activation function, and  $*$  stands for matrix multiplication. Similar to the TFEM module, a residual connection is also introduced in the outer layer of the FEM module to ensure the robustness of the generalized feature representation.

### 3.1.3. Sampling Module in the Encoder-Decoder Architecture

In the U-shaped symmetrical encoder-decoder architecture, to effectively capture multi-scale feature information and avoid mutual interference between the upsampling module, feature extraction module, and downsampling module, we design a novel Downsampling-Upsampling Block, as illustrated in Figure 6.

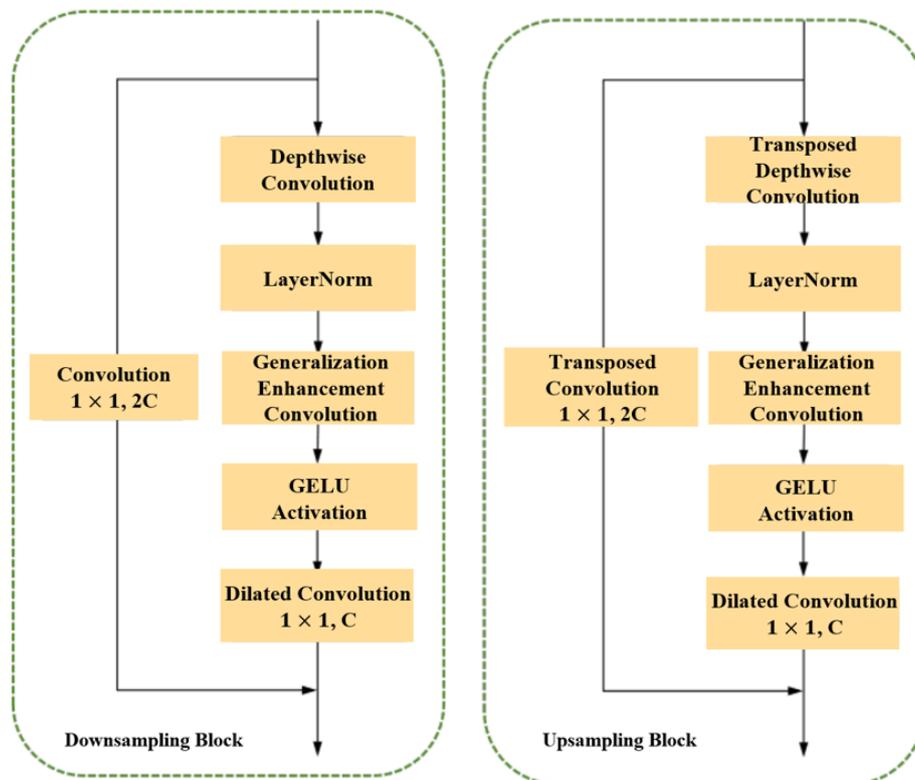


Figure 6. Illustration of the Sampling Module.

Specifically, when a feature map with  $C$  channels enters the downsampling block, it first passes through a depthwise convolution layer with a kernel size of  $7 \times 7$  and a stride of 2, followed by Group Normalization. Next, the feature map is processed by a  $1 \times 1$  convolution layer with a dilation rate of 4, effectively expanding the receptive field to four times its original size, and increasing the

channel number to  $4C$ . After activation with the GELU function, the feature map is passed through another  $1 \times 1$  convolution layer (compression layer), reducing the number of channels to  $2C$  and the spatial dimensions to half of the original, thus completing the downsampling operation. To further enhance the robustness of the sampling module, an additional residual path is introduced in the downsampling block. This residual branch consists of a  $1 \times 1$  convolution layer with a stride of 2, directly converting the input to  $2C$  channels and halving the spatial resolution, ensuring better feature preservation and stable gradient flow. The upsampling module is designed in contrast to the downsampling block, where each convolution operation is replaced by a transposed convolution. The output of the upsampling block has half the number of channels of the input feature map, and the spatial resolution is doubled, effectively restoring the feature map to its previous scale.

### 3.2. Overall Architecture of the Multi-Scale Feature Fusion Network

As shown in Figure 7, the overall framework of the Multi-Scale Feature Fusion Network (MSFNet) both mainly of three parts: the encoder stage, the decoder stage, and the skip connection stage. Specifically, in MSFNet, the Combined Feature Extraction Module (CFEM) and Sampling Module are designed for feature extraction and enhancement along the encoder-decoder path. In the skip connection stage, we introduce a Multi-scale Feature Fusion Attention Mechanism (MFFA) before the simple concatenation operation typically used in U-Net. This design further strengthens the overall capability of the symmetric U-shaped network in terms of feature extraction and feature fusion.

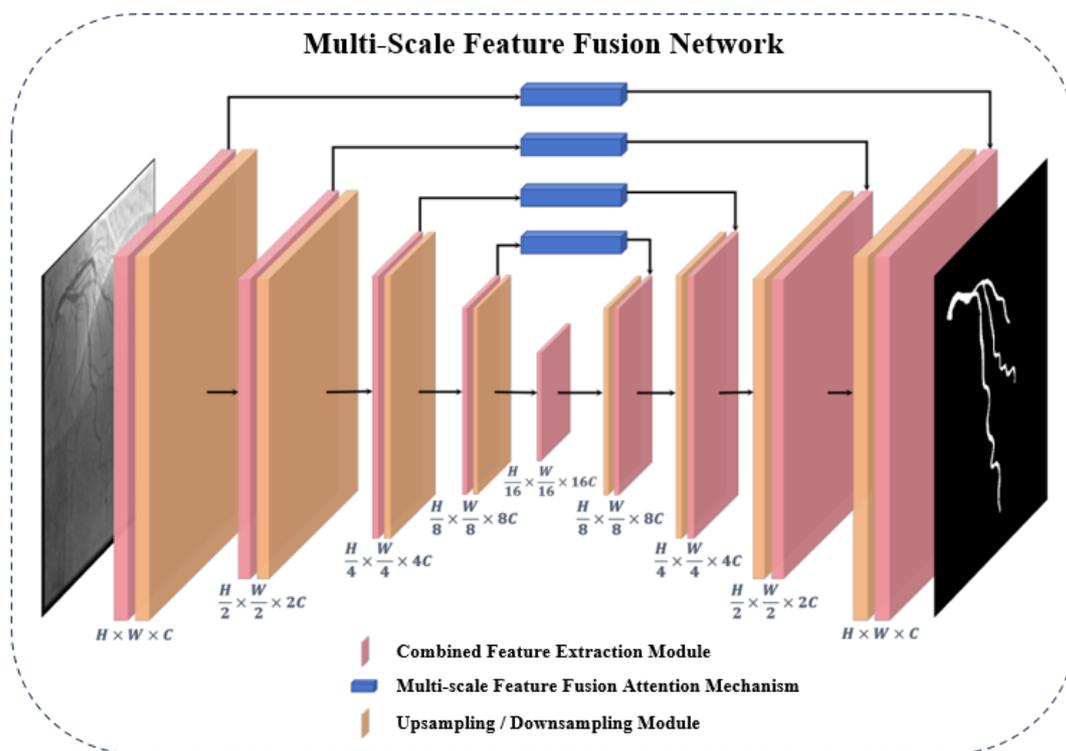


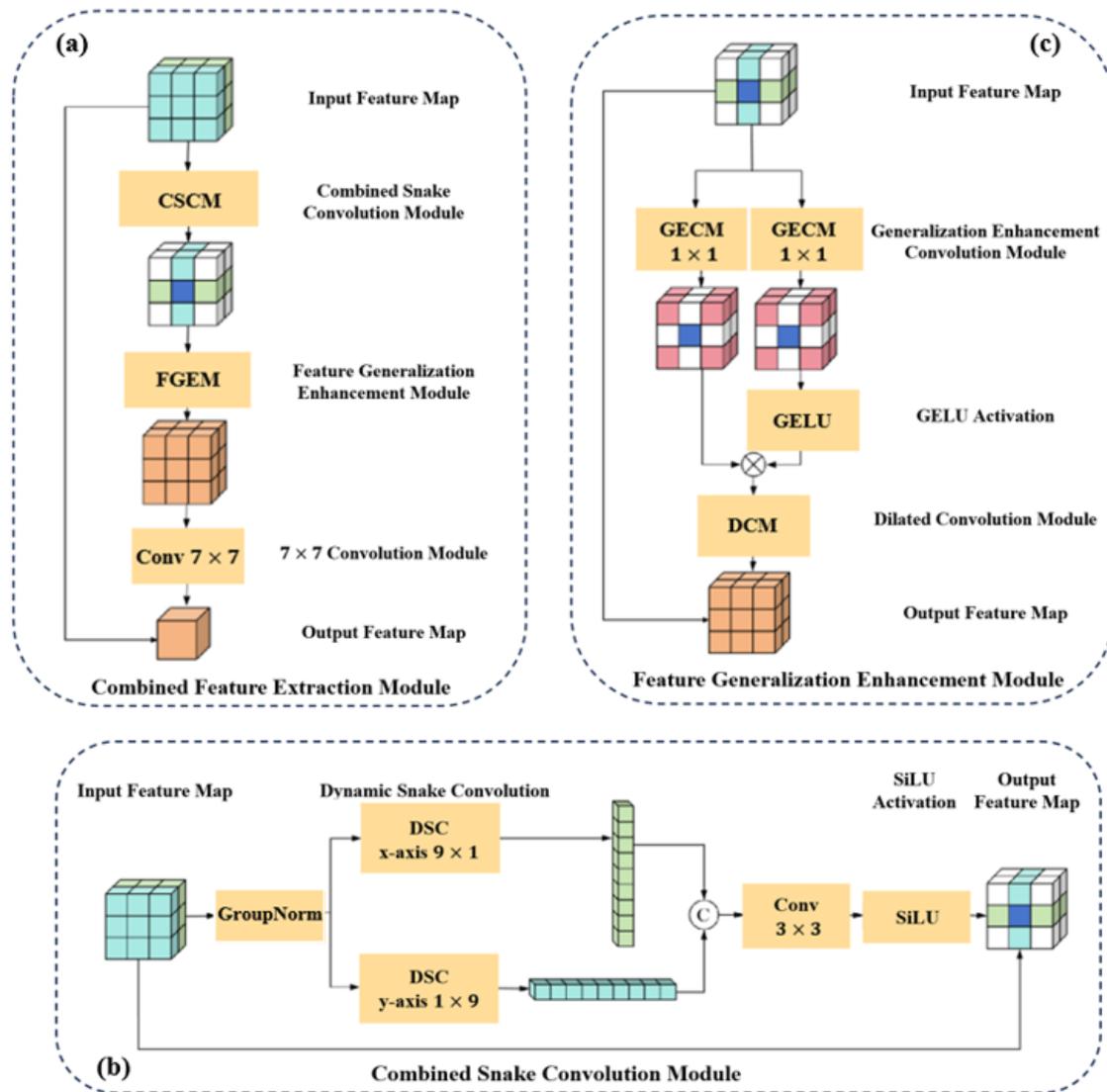
Figure 7. Illustration of the overall architecture of MSFNet.

#### 3.2.1. Combined Feature Extraction Module in the Encoder-Decoder Architecture

To effectively model the complex structural characteristics of coronary vessels in medical images, we design a Combined Feature Extraction Module (CFEM). As shown in Figure 8, this module is aimed at enhancing the model's perception of elongated, tortuous, and directionally varied vessel structures, while improving spatial contextual representation and overall generalization capability.

The module first applies group normalization [40] to the input features to mitigate the instability of feature distributions caused by small-batch training, thereby improving the stability of model convergence. Subsequently, the module introduces dynamic snake convolutions [38] along the x-axis

and y-axis to extract features. This operation incorporates a learnable spatial offset mechanism that allows the convolution kernels to dynamically adapt to the orientation of vessel structures. It enables irregular sampling along the centerline of the vessels, thereby enhancing the model's ability to capture structural deformations, particularly in handling intricate vessel branches, curves, and tapering with greater flexibility and representational power.



**Figure 8.** (a) Illustration of the Combined Feature Extraction Module. (b) Illustration of the Combined Snake Convolution Module. (c) Illustration of the Feature Generalization Enhancement Module.

The input feature map  $X$  is first processed by Group Normalization to mitigate the instability caused by small batch training. It is then fed into two separate paths corresponding to the x-axis and y-axis, where dynamic deformable convolutions of sizes  $9 \times 1$  and  $1 \times 9$ , respectively, are applied, yielding feature maps  $X'_x$  and  $X'_y$ .

Subsequently, the features  $X'_x$  and  $X'_y$  are concatenated along the channel dimension to fuse multi-directional spatial information. A  $3 \times 3$  convolutional layer is then employed to reconstruct contextual dependencies and compress the features, thereby enhancing their representational capacity.

For non-linear transformation, the SiLU activation function is utilized in the CFEM. Its smooth response curve helps preserve edge details in the image, improves the stability of gradient propagation, and accelerates model convergence.

The entire process can be formally expressed as:

$$X_i = D_i(N(X)), \quad (6)$$

$$Y = \sigma\left(\text{Conv}\left(C(X'_x, X'_y)\right)\right). \quad (7)$$

Overall, the CFEM achieves collaborative modeling of directional awareness, spatial adaptability, and feature fusion through a lightweight architecture. This design enhances the network's ability to identify complex vascular topologies. By replacing the fixed kernel sampling strategy with dynamic modeling, CFEM strengthens the model's performance in fine-grained segmentation tasks, particularly for target structures like coronary arteries that exhibit pronounced directionality and morphological variation.

In medical image segmentation tasks—especially those involving structures with complex spatial distributions, such as coronary arteries—relying solely on local convolutional features often fails to capture long-range dependencies and contextual relationships between regions. This limitation may lead to blurred structural boundaries or missed detections of distal vessels. To address this, we introduce the Feature Generalization Enhancement Module (FGEM) into the encoder-decoder structure of the backbone network. FGEM is designed to expand the model's feature perception range and semantic modeling capability, thereby improving its capacity to analyze vascular structures with complex topologies.

As shown in Figure 7, FGEM receives the output feature map from the previous feature extraction module and splits it into two parallel paths. Each path is then processed by a dilated convolution with a kernel size of  $1 \times 1$  and a dilation rate of 4, effectively expanding the receptive field to four times the original size. This design enables the acquisition of broader contextual information while maintaining a relatively small parameter count, which is especially beneficial for detecting long-range but semantically related pixel responses in elongated structures.

Following the dilated convolution, one of the two branches undergoes a GELU non-linear activation, enhancing its non-linear modeling capability and gradient response strength. The output of this activated branch is then element-wise multiplied with the output of the other (non-activated) branch, facilitating inter-branch feature interaction and enhancement. This mechanism effectively introduces a dynamic modulation mechanism that allows the activated branch to spatially regulate the feature representations of the original branch. As a result, it reinforces responses in key regions while suppressing redundant or irrelevant features.

The fused feature map is subsequently passed through a  $1 \times 1$  pointwise convolution to reduce the channel dimension back to its original size, restoring the receptive field to its initial state. This helps prevent computational redundancy while maintaining compact and expressive feature representations. The entire process can be viewed as a feature generalization-compression mechanism that first expands and then contracts the features. Without introducing significant computational overhead, it enables collaborative modeling of both local and global semantics.

The above operations can be formally expressed as:

$$Y = \text{Com}(\sigma(\text{Exp}(X) * \text{Exp}(X))), \quad (8)$$

where  $\text{Exp}$  denotes dilated convolution,  $\text{Com}$  represents compression convolution and  $\sigma$  refers to the GELU activation function.

In summary, the Feature Generalization Enhancement Module (FGEM) balances contextual awareness, inter-channel interaction, and lightweight computation. It not only enhances the model's representation capability for complex vascular structures but also achieves stronger feature selectivity and branch generalization through inter-path modulation. Experimental results demonstrate that the introduction of FGEM significantly improves the model's precision in capturing the detailed structures and boundaries of coronary artery branches.

### 3.2.2. Multi-scale Feature Fusion Attention Mechanism in Skip Connections

Attention mechanisms, due to their ability to model long-range dependencies, have been widely applied in image segmentation tasks [41]. In coronary vessel segmentation, different branches often exhibit high morphological similarity. Introducing an appropriate attention mechanism helps the model effectively capture vascular features that are located in different spatial regions but share common characteristics, thereby improving overall segmentation accuracy. Therefore, this chapter introduces a Multi-scale Feature Fusion Attention (MFFA) Module into the skip connections of the U-shaped symmetric structure, aiming to enhance the feature extraction and fusion capabilities between the encoder and decoder.

Common attention mechanisms include spatial attention, channel attention, and multi-head attention. Although multi-head attention performs well in Transformer architectures, it suffers from high computational complexity and training costs, making it unsuitable for real-time intraoperative applications in clinical settings. In contrast, spatial and channel attention mechanisms are more lightweight and computationally efficient, making them better suited for real-time coronary artery segmentation in DSA images.

To improve model performance without significantly increasing computational cost, we propose a Multi-scale Feature Fusion Attention Module that integrates multi-scale convolution structures, spatial attention, and channel attention mechanisms to achieve efficient and lightweight feature enhancement. The module architecture is shown in Figure 9.

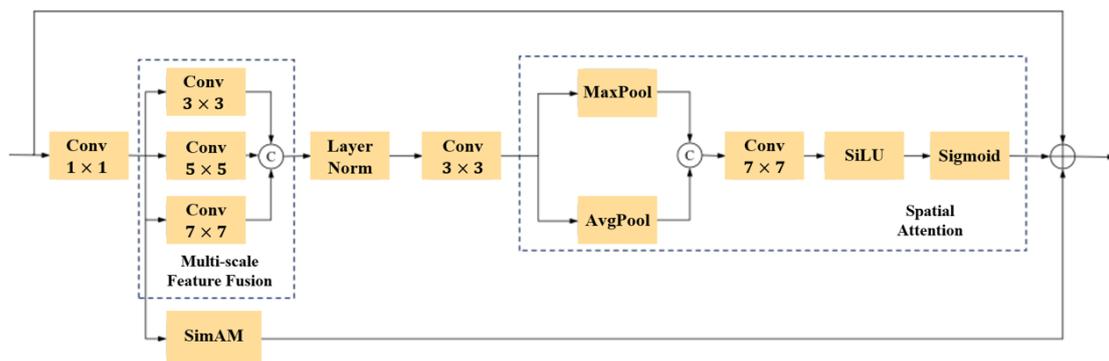


Figure 9. Illustration of the Multi-scale Feature Fusion Attention Mechanism.

Specifically, the input feature map is first passed through a  $1 \times 1$  convolution layer to reduce the number of channels to one-fourth of the original. The output is then processed in four parallel branches. Among these, three branches apply convolution operations with kernel sizes of  $1 \times 1$ ,  $3 \times 3$ , and  $5 \times 5$ , respectively, to extract spatial features at multiple scales. The resulting three feature groups are concatenated, followed by Group Normalization, and then passed through another  $1 \times 1$  convolution layer to fuse the features.

The fused feature map is then fed into a spatial attention module. This module applies both average pooling and max pooling along the channel dimension, producing two 2D spatial maps that are concatenated. The concatenated map is then passed through a  $7 \times 7$  convolution with 2 input channels and 1 output channel, followed by SiLU and sigmoid activations to generate a normalized spatial attention map. Finally, this attention map is element-wise multiplied with the original feature map to enhance the response in significant spatial regions.

The entire process can be represented as:

$$F' = F \otimes \sigma \left( f^{7 \times 7} [\text{AvgPool}(F); \text{MaxPool}(F)] \right), \quad (9)$$

where  $F'$  denotes the output feature map,  $F$  is the input feature map, AvgPool represents average pooling, MaxPool represents max pooling,  $\otimes$  denotes element-wise multiplication,  $\sigma$  is the sigmoid function, and  $f^{7 \times 7}$  represents a  $7 \times 7$  convolution operation.

Along the channel dimension, to further reduce model complexity, this module adopts the SimAM (Simple Attention Module) [42] mechanism instead of the conventional channel attention used in CBAM [33]. SimAM leverages a parameter-free energy function to estimate the importance of each pixel by computing its energy relation with other pixels in the same channel, thereby generating channel-wise attention weights. This design is lightweight and computationally efficient, making it particularly suitable for skip connections where minimizing overhead is essential.

The SimAM [42] process can be described as follows: for a given feature map  $x \in \mathbb{R}^{H \times W}$  in a certain channel, it is first flattened into a vector  $x = [x_1, x_2, \dots, x_n]$ , where  $n = H \times W$ . For each pixel  $x_i$ , SimAM calculates an energy function to evaluate how suppressing neuron  $x_i$  affects the overall activation level of the feature map. The formulation is expressed as:

$$E(x_i) = (x_i - \mu)^2 + \lambda \cdot \sigma^2, \quad (10)$$

where  $\mu$  denotes the mean of the current channel,  $\sigma^2$  represents the variance of the current channel, and  $\lambda$  is a hyperparameter with a value of  $1e-4$ .  $E(x_i)$  denotes the energy of the pixel  $x_i$ —the smaller the value of  $E(x_i)$ , the higher the cost of suppressing this neuron. Therefore, the attention weight is defined as:

$$\alpha_i = \frac{1}{1 + E(x_i)} = \frac{1}{1 + (x_i - \mu)^2 + \lambda \cdot \sigma^2}, \quad (11)$$

where the final weighted output feature map is  $y_i = \alpha_i \cdot x_i$ .

To further simplify computation, an approximate behavior of SimAM is adopted here, converting the per-pixel weight computation into a channel-wise weight calculation, making the model more lightweight. The weighting constant  $\alpha$  is calculated as follows:

$$\alpha = \frac{\sigma^2}{4(\sigma^2 + \lambda)}. \quad (12)$$

Finally, the feature maps enhanced by spatial and channel attention are fused with the original input feature maps through a residual connection. This fusion preserves low-level detail information while reinforcing high-level discriminative features, thereby improving the model's ability to detect and localize vascular targets.

With the proposed Multi-scale Feature Fusion Attention Module, our proposed model can enhance feature retention and generalization capabilities within the encoder-decoder architecture without significantly increasing the number of parameters or computational cost. This makes it particularly suitable for real-time processing of DSA images during surgery, effectively improving both the efficiency and accuracy of coronary vessel segmentation.

## 4. Experiments

### 4.1. Datasets

This study utilizes a guidewire dataset to support model training and evaluation for the guidewire segmentation task. The dataset contains 3,669 pairs of two-dimensional X-ray fluoroscopy guidewire images along with corresponding segmentation labels, with each image sized  $512 \times 512$  pixels. The dataset includes guidewire images of various categories, shapes, and quantities. In addition, this study also employs a high-quality coronary artery imaging dataset to support model training and evaluation for computer-aided diagnosis of coronary artery disease (CAD). The ARCADE [10] dataset is based on X-ray coronary angiography (XCA) images and primarily focuses on multi-region segmentation and plaque detection tasks. It consists of two tasks: coronary vessel tree region segmentation and atherosclerotic plaque annotation, each containing 1,000 training images and 200 validation images. The segmentation of cardiac regions follows the Syntax Score methodology.

#### 4.2. Implementation Detail

In the experiments, five models with strong performance in coronary artery or tubular structure segmentation tasks are selected for comparison: U-Net [28], Attention U-Net [30], MedNeXt [43], DSCNet [44], and SPNet [45]. All models and experiments are implemented using the PyTorch framework and executed on an NVIDIA RTX 3090 GPU. The Adam optimizer is used during training, with a learning rate set to  $1e-4$ . For the ARCADE dataset, all models are trained for 200 epochs with a batch size of 4; for the Guidewires dataset, the number of epochs remains the same (200), while the batch size is set to 24.

#### 4.3. Evaluation Metrics

Given that the segmentation task in this study involves a single class, the evaluation metrics used differ slightly from the traditional practice in medical image segmentation, which typically relies solely on the Dice Similarity Coefficient (DSC) and the 95% Hausdorff Distance (HD95). To comprehensively assess the performance of the proposed models, this study employs four metrics: DSC, HD95, Recall, and Precision. Among these, DSC measures the overlap between the predicted segmentation and the ground truth, reflecting the overall segmentation accuracy. HD95 evaluates the 95th percentile of the maximum distance between the predicted and actual boundaries, providing insight into the model's boundary accuracy and localization performance. Recall indicates the model's ability to correctly identify the actual target regions, while Precision reflects the proportion of correctly predicted target regions among all predicted positive regions. Together, these four metrics offer a multi-faceted evaluation of both detection accuracy and boundary delineation, enabling a more thorough and precise assessment of coronary artery segmentation performance.

#### 4.4. Results

We validate the performance of GS-UNet on the Guidewire dataset and the performance of MSFNet on the ARCADE dataset using five advanced methods: U-Net [28], Attention U-Net [30], MedNeXt [43], DSCNet [44], and SPNet [45]. Figure 10 presents the visualized segmentation results of all methods on the Guidewire dataset, while Figure 11 shows the results on the ARCADE dataset. Mis-segmented regions are highlighted with red boxes, and target regions where GS-UNet or MSFNet demonstrates superior segmentation are highlighted with yellow boxes. The specific qualitative and quantitative analysis results are as follows:

Attention U-Net [30] incorporates attention mechanisms into the U-Net architecture, achieving strong performance across multiple tasks, including vascular segmentation. MedNeXt [43], introduces a scalable architecture that significantly enhances the capabilities of symmetric U-shaped networks in both 2D and 3D medical image segmentation tasks, demonstrating excellent performance across various categories and modalities of medical image segmentation datasets. DSCNet [44] integrates dynamic snake convolutions into the basic U-Net architecture, enabling highly accurate vascular segmentation. SPNet [45] introduces strip pooling operations, which leverage long kernel shapes to capture contextual information of elongated structures while suppressing interference from irrelevant structures, achieving promising results in the segmentation of elongated objects in natural scenes. In this section, SPNet is introduced as a baseline method to explore the feasibility of its strip pooling operations in addressing the tubular features and long-range dependencies of coronary vessels.

##### 4.4.1. Segmentation Performance on the Guidewire dataset

The comparative experimental results on the Guidewire dataset are shown in Table 1. As illustrated in Table 1, the model proposed in this chapter outperforms all comparison models across four key metrics: DSC, HD95, Recall, and Precision. This indicates that the proposed model achieves the best overall performance in terms of segmentation accuracy, boundary precision, recall, and precision.

Table 1. COMPARISONS WITH STATE-OF-THE-ART MODELS ON THE GUIDEWIRE DATASET.

Method	DSC $\uparrow$ (% mean)	HD95 $\downarrow$ (mm, mean)	Recall $\uparrow$ (%, mean)	Precision $\uparrow$ (%, mean)
U-Net	85.56	11.956	86.69	85.41
Attention U-Net	85.87	10.113	84.64	86.16
DSCNet	85.90	11.350	86.86	85.75
MedNeXt	86.11	10.179	86.81	86.01
SPNet	82.91	20.995	83.63	83.65
<b>GS-UNet</b>	<b>86.31</b>	<b>8.685</b>	<b>86.88</b>	<b>86.69</b>

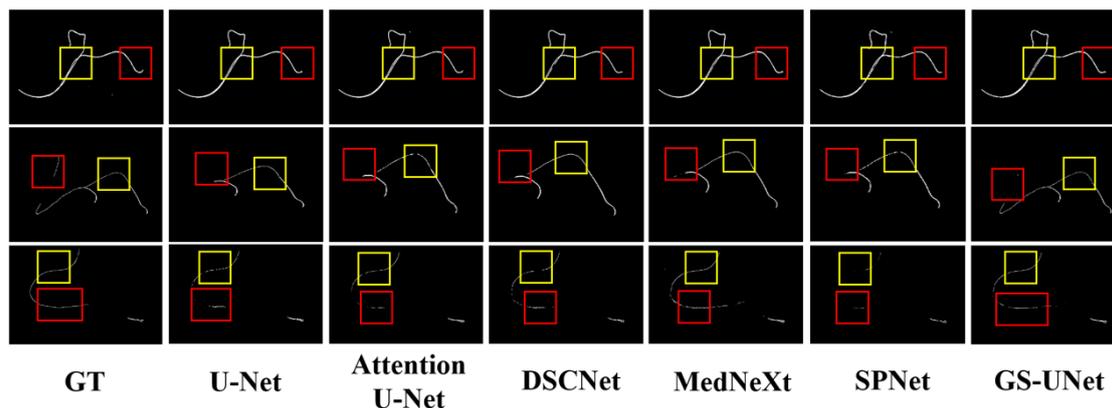


Figure 10. Visual examples of segmentation on the Guidewire dataset.

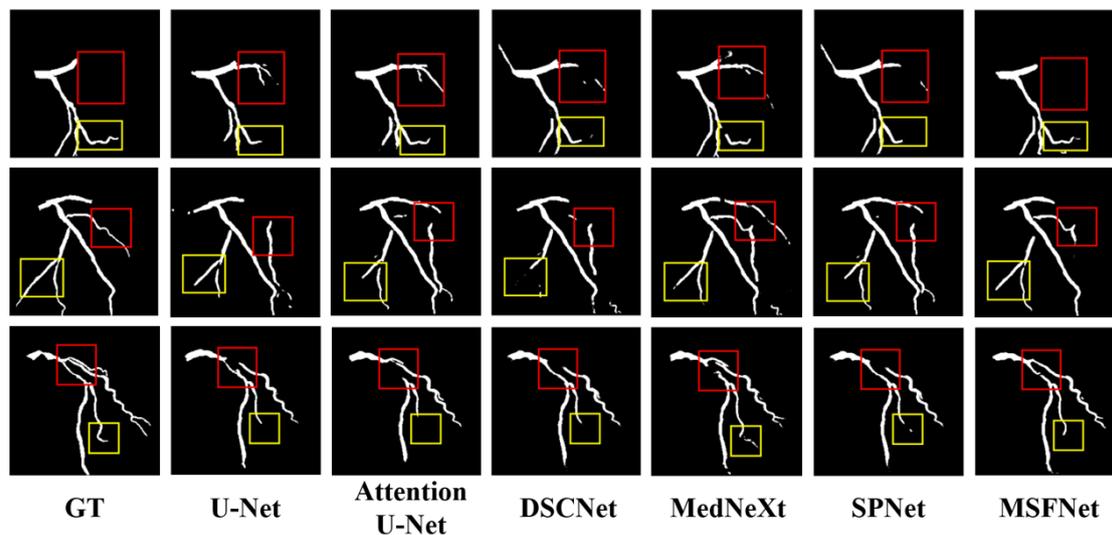


Figure 11. Visual examples of segmentation on the ARCADE dataset.

Compared to the current state-of-the-art segmentation model DSCNet [44], GS-UNet achieves improvements of 0.65% in DSC, 1.44% in Recall, and 1.44% in Precision, while further reducing the HD95 value by 0.108 mm. These results clearly demonstrate that GS-UNet offers superior capabilities in extracting and preserving tubular structures such as guidewires. Leveraging its dedicated tubular feature extraction module and enhanced feature generalization components, GS-UNet significantly improves the network's capacity to adapt to varying guidewire shapes, sizes, and imaging conditions. The notable gain in Recall underscores the model's effectiveness in detecting challenging, small-scale, or low-contrast guidewire segments, while the reduction in HD95 reflects improved boundary localization accuracy. Collectively, these findings validate the robustness and efficacy of the proposed GS-UNet framework for guidewire segmentation tasks in clinical settings.

Figure 10 shows the visualization results of guidewire segmentation on the guidewire dataset using different models in the comparative experiment. Each row represents a case of guidewire segmentation.

In the first row, as highlighted in the yellow box, DSCNet [44] exhibits guidewire breakage, while the other models achieve relatively good segmentation results. In the red box, only Attention U-Net [30], MedNeXt [43], and GS-UNet manage to accurately identify and segment the guidewire tip.

In the second row, the yellow box shows segmentation errors of varying degrees in both Attention U-Net [30] and SPNet [45]. In the red box, only GS-UNet successfully identifies the corresponding guidewire segment, whereas the other models treat it as background and fail to segment it. This highlights the importance of tubular feature extraction modules: models with weaker feature extraction capabilities tend to perform poorly, whereas those with stronger feature extraction perform better. However, since these models lack modules specifically designed to extract tubular features, their segmentation results remain suboptimal.

In the third row, as shown in the yellow box, all models except DSCNet [44] and GS-UNet show varying degrees of guidewire breakage in their segmentation. In the red box, in cases where the ground truth is inherently discontinuous, U-Net [28], Attention U-Net [30], and SPNet [45] fail to effectively handle the discontinuous background and guidewire, while MedNeXt [43] and GS-UNet demonstrate relatively ideal segmentation performance.

#### 4.4.2. Segmentation Performance on the ARCADE dataset

The comparative experimental results on the ARCADE dataset are shown in Table 2. To validate the effectiveness of the multi-scale composite attention module, ablation experiments were also conducted. In the ablation setting, all skip connections containing the multi-scale combined attention module were replaced with standard skip connections, where the output from the encoder block is directly concatenated to the input of the corresponding decoder block. As shown in Table 2, the proposed model outperforms all baseline models in four key metrics: DSC, HD95, Recall, and Precision, indicating superior performance in overall segmentation accuracy, boundary accuracy, recall, and precision. When comparing the network without the attention module to other models, a significant improvement is still observed, further demonstrating that the composite tubular feature extraction module effectively contributes to the segmentation task by leveraging the morphological characteristics of coronary vessels.

**Table 2.** COMPARISONS WITH STATE-OF-THE-ART MODELS ON THE ARCADE DATASET.

Method	DSC $\uparrow$ (% mean)	HD95 $\downarrow$ (mm, mean)	Recall $\uparrow$ (%, mean)	Precision $\uparrow$ (%, mean)
U-Net	72.26	76.013	69.63	77.88
Attention U-Net	72.51	72.021	70.01	78.12
DSCNet	73.48	69.527	73.60	75.24
MedNeXt	75.10	61.960	72.50	80.02
SPNet	55.90	79.391	59.92	54.17
MSFNet(without MFFA)	75.39	61.877	73.20	79.64
<b>MSFNet</b>	<b>76.74</b>	<b>57.836</b>	<b>74.87</b>	<b>80.66</b>

Meanwhile, the proposed model in this study achieves significantly higher DSC (76.74%) and Precision (80.66%) values compared to DSCNet [44], which also possesses tubular feature extraction capabilities. This indicates that the model's feature expansion module contributes effectively to enhancing feature generalization. Furthermore, the HD95 value is markedly reduced to 57.836 mm, suggesting that the improved generalization also enhances the accuracy of boundary segmentation, particularly in regions where vessel edges are difficult to identify. A further comparison between the complete model with the attention module and its counterpart without attention shows that the

attention mechanism contributes positively to improvements across nearly all evaluation metrics. The improvements in DSC and HD95 are especially notable. This demonstrates that the proposed multi-scale composite attention module enables the network to learn more comprehensive multi-scale vascular features, resulting in more complete vessel structure segmentation. Among the baseline models, MedNeXt [43] and DSCNet perform relatively well, which confirms that large receptive fields in scalable architectures and tubular feature extraction via dynamic snake convolution are effective in capturing the branching and tubular morphological features of blood vessels. However, SPNet [45] performs significantly worse than the other models, with a DSC of only 55.90%. This may be attributed to its use of strip pooling, which is specifically designed for horizontal and vertical patterns in natural scenes and linear objects. Such an approach is unsuitable for blood vessels, which often exhibit branching and highly curved tubular structures.

Figure 11 presents the visualization results of coronary vessel segmentation from different models in the comparative experiments, including several well-performing models—U-Net, Attention U-Net, DSCNet [44], and MedNeXt [43]—as well as the model proposed in this chapter. Each column corresponds to the segmentation mask of the same coronary vessel. In the first row, the red box shows that only the proposed model correctly identifies the target region, while the other models exhibit large areas of missegmentation. In the yellow box, the proposed model demonstrates the most complete segmentation and the best connectivity. Although U-Net [28] and Attention U-Net [30] achieve relatively good connectivity, they lack completeness in segmenting the vessel segment. MedNeXt and DSCNet, on the other hand, suffer from vessel disconnection, resulting in poor connectivity. In the second row, the red box highlights that all models suffer from varying degrees of missegmentation. However, the missegmentation in the comparison models is significantly more severe than that of the proposed model, indicating its superior ability to reduce such errors. The yellow box shows that the proposed model achieves the most complete segmentation of the vessel segment, closely matching the ground truth. This demonstrates its excellent performance in connectivity, completeness, and its stronger capability in extracting tubular features and generalizing across variations. In the third row, the red box indicates that all comparison models exhibit severe vessel breakage. Although the proposed model also shows some disconnection, the breakage is much smaller, indicating greatly improved connectivity. In the yellow box, the proposed model misses only a small segment at the distal end of the vessel. In contrast, U-Net, Attention U-Net, and DSCNet miss significantly larger areas, while MedNeXt shows a missegmentation.

In summary, the model proposed in this chapter demonstrates superior morphological feature extraction and generalization capabilities, significantly improving the accuracy and connectivity in coronary vessel segmentation.

## 5. Limitation

Although MSFNet demonstrates promising performance on the ARCADE coronary artery segmentation dataset [10], several limitations remain. Firstly, while the model effectively integrates multi-scale features and captures both local and global vascular structures, its segmentation accuracy may deteriorate notably in images characterized by extremely low contrast, high levels of noise, or imaging artifacts. In such scenarios, vessel boundaries become indistinct and the intensity distribution between target vessels and the background converges, thereby increasing the risk of false negatives and false positives. Secondly, the current evaluation is predominantly conducted on publicly available benchmark datasets that exhibit a certain degree of homogeneity in image quality, acquisition protocols, and pathological variations, which may limit the model's generalizability to the heterogeneous and complex cases encountered in real-world clinical environments. Moreover, MSFNet has yet to undergo extensive validation within actual clinical workflows, including assessments involving expert annotations and interactive feedback.

Similarly, while GS-UNet achieves competitive performance on the guidewire segmentation task, it also faces several challenges. Guidewire structures in fluoroscopic images are often extremely thin

and low in contrast, making them particularly vulnerable to occlusions, overlaps with anatomical structures, and imaging noise. GS-UNet, though designed to enhance fine structural detail, may still suffer from missed detections or inaccurate boundary localization in complex backgrounds. Furthermore, due to the lack of diverse publicly available guidewire datasets, the model's performance evaluation is limited in scope, raising concerns about its robustness across different imaging devices, patient anatomies, and surgical contexts.

Future research should focus on incorporating advanced domain adaptation techniques, active learning frameworks, and model compression strategies to further enhance robustness, generalizability, and computational efficiency of both MSFNet and GS-UNet, thereby facilitating their translation from research settings to clinical practice.

## 6. Discussion

As illustrated in Figure 1 (Model Complexity Comparison), our proposed GS-UNet and the reference MSFNet exhibit fundamentally different efficiency-accuracy trade-offs, each demonstrating unique advantages for medical image segmentation tasks.

GS-UNet achieves remarkable parameter efficiency with only 5.48 million parameters (7.0× fewer than MSFNet's 38.07M) and 11.92 GFLOPs computational cost (3.6× lower than MSFNet's 43.28G). This efficiency stems from two architectural innovations: The TFEM module employs axis-separable dynamic snake convolutions (9×1 and 1×9 kernels) followed by 3×3 feature fusion, reducing redundant computations while preserving vascular structure sensitivity. Our downsampling/upsampling blocks integrate depthwise separable convolutions with residual connections, achieving 84.3% parameter reduction compared to standard U-Net [28] blocks.

In contrast, MSFNet's higher complexity (38.07M/43.28G) originates from its multi-scale feature fusion mechanism, which processes 5 parallel resolution pathways. While this enhances segmentation precision, it imposes significant hardware requirements.

## 7. Conclusions

In summary, we present two novel deep learning frameworks—GS-UNet for guidewire segmentation and MSFNet for coronary vessel segmentation—each specifically tailored to address the unique challenges of their respective tasks.

GS-UNet introduces a Tubular Feature Extraction Module (TFEM) that significantly enhances the model's capability to capture the elongated and delicate morphological structures of guidewires, thereby preserving structural continuity. Furthermore, a Feature Extension Module (FEM) is proposed to improve generalization while reducing computational overhead, enabling robust discrimination of guidewires across varied shapes and positions.

For coronary vessel segmentation, MSFNet incorporates a Combined Feature Extraction Module (CFEM) designed to effectively model the elongated and tortuous vascular morphology by integrating local detail perception with global structural awareness. Additionally, a Feature Generalization Enhancement Module (FGEM) is proposed to further improve the model's adaptability to different imaging conditions and anatomical variations. The integration of a Multi-scale Feature Fusion Attention mechanism (MFFA) allows for the adaptive aggregation of spatial contextual information from multiple receptive field scales, guiding the network's focus toward vessel-relevant regions and improving fine vessel delineation and structural consistency.

Extensive experimental evaluations on both the self-constructed guidewire dataset and the benchmark ARCADE dataset [10] demonstrate that GS-UNet and MSFNet outperform several state-of-the-art methods across a variety of quantitative metrics. The results confirm their superior performance in fine structure identification, boundary accuracy, and topological integrity, highlighting their strong potential for enhancing automated segmentation in clinical interventional and diagnostic workflows.

## References

1. Khan, S.Q.; Ludman, P.F. Percutaneous coronary intervention. *Medicine* **2022**, *50*, 437–444.
2. Dimagli, A.; Spadaccio, C.; Myers, A.; Demetres, M.; Rademaker-Havinga, T.; Stone, G.W.; Spertus, J.A.; Redfors, B.; Fremes, S.; Gaudino, M.; et al. Quality of life after percutaneous coronary intervention versus coronary artery bypass grafting. *Journal of the American Heart Association* **2023**, *12*, e030069.
3. Alkhalil, M. An Overview of Current Advances in Contemporary Percutaneous Coronary Intervention. *Current Cardiology Reviews* **2022**, *18*, 5.
4. Alvandi, M.; Javid, R.N.; Shaghaghi, Z.; Farzipour, S.; Nosrati, S. An in-depth analysis of the adverse effects of ionizing radiation exposure on cardiac catheterization staffs. *Current Radiopharmaceuticals* **2024**, *17*, 219–228.
5. Ball, W.T.; Sharieff, W.; Jolly, S.S.; Hong, T.; Kutryk, M.J.; Graham, J.J.; Fam, N.P.; Chisholm, R.J.; Cheema, A.N. Characterization of operator learning curve for transradial coronary interventions. *Circulation: Cardiovascular Interventions* **2011**, *4*, 336–341.
6. Garcia, C.d.L.; Abreu, L.C.d.; Ramos, J.L.S.; Castro, C.F.D.d.; Smiderle, F.R.N.; Santos, J.A.d.; Bezerra, I.M.P. Influence of burnout on patient safety: systematic review and meta-analysis. *Medicina* **2019**, *55*, 553.
7. Patel, T.M.; Shah, S.C.; Pancholy, S.B. Long distance tele-robotic-assisted percutaneous coronary intervention: a report of first-in-human experience. *EclinicalMedicine* **2019**, *14*, 53–58.
8. Łajczak, P.; Ayesha, A.; Sahin, O.K.; Freeman, P.I.; Majeed, M.W.; Righetto, B.B.; Obi, O.; Moreno, G.J.; Krishna, M.M.; Mulenga, K.F.; et al. Comparison between robot-assisted and manual percutaneous coronary intervention—an updated systematic review, meta-analysis, propensity-matched investigation, and trial sequential analysis. *Cardiovascular Intervention and Therapeutics* **2025**, pp. 1–16.
9. Covello, B.; McKeon, B. *Fluoroscopic angiography assessment, protocols, and interpretation*; StatPearls Publishing, Treasure Island (FL), 2022.
10. Popov, M.; Amanturdieva, A.; Zhaksylyk, N.; Alkanov, A.; Saniyzbekov, A.; Aimyshev, T.; Ismailov, E.; Bulegenov, A.; Kuzhukeyev, A.; Kulanbayeva, A.; et al. Dataset for automatic region-based coronary artery disease diagnostics using X-ray angiography images. *Scientific data* **2024**, *11*, 20.
11. Kaur, D.; Kaur, Y. Various image segmentation techniques: a review. *International journal of computer science and Mobile Computing* **2014**, *3*, 809–814.
12. Xu, Y.; Quan, R.; Xu, W.; Huang, Y.; Chen, X.; Liu, F. Advances in medical image segmentation: A comprehensive review of traditional, deep learning and hybrid approaches. *Bioengineering* **2024**, *11*, 1034.
13. Senthilkumaran, N.; Vaithegi, S. Image segmentation by using thresholding techniques for medical images. *Computer Science & Engineering: An International Journal* **2016**, *6*, 1–13.
14. Al-Amri, S.S.; Kalyankar, N.; Khamitkar, S. Image segmentation by using edge detection. *International journal on computer science and engineering* **2010**, *2*, 804–807.
15. Adams, R.; Bischof, L. Seeded region growing. *IEEE Transactions on pattern analysis and machine intelligence* **1994**, *16*, 641–647.
16. Fan, J.; Zeng, G.; Body, M.; Hacid, M.S. Seeded region growing: an extensive and comparative study. *Pattern recognition letters* **2005**, *26*, 1139–1156.
17. Jalali, Y.; Fateh, M.; Rezvani, M.; Abolghasemi, V.; Anisi, M.H. ResBCDU-Net: a deep learning framework for lung CT image segmentation. *Sensors* **2021**, *21*, 268.
18. Seo, H.; Badieli Khuzani, M.; Vasudevan, V.; Huang, C.; Ren, H.; Xiao, R.; Jia, X.; Xing, L. Machine learning techniques for biomedical image segmentation: an overview of technical aspects and introduction to state-of-art applications. *Medical physics* **2020**, *47*, e148–e167.
19. Goh, T.Y.; Basah, S.N.; Yazid, H.; Safar, M.J.A.; Saad, F.S.A. Performance analysis of image thresholding: Otsu technique. *Measurement* **2018**, *114*, 298–307.
20. Ahmed, A.S. Comparative study among Sobel, Prewitt and Canny edge detection operators used in image processing. *J. Theor. Appl. Inf. Technol* **2018**, *96*, 6517–6525.
21. Forcadel, N.; Le Guyader, C.; Gout, C. Generalized fast marching method: applications to image segmentation. *Numerical Algorithms* **2008**, *48*, 189–211.
22. Srisha, R.; Khan, A. Morphological operations for image processing: understanding and its applications. *NCVSComs-13* **2013**, *13*, 19.
23. Mienye, I.D.; Swart, T.G.; Obaido, G.; Jordan, M.; Ilono, P. Deep convolutional neural networks in medical image analysis: A review. *Information* **2025**, *16*, 195.
24. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: an overview and application in radiology. *Insights into imaging* **2018**, *9*, 611–629.

25. Chen, C.; Qin, C.; Qiu, H.; Tarroni, G.; Duan, J.; Bai, W.; Rueckert, D. Deep learning for cardiac image segmentation: a review. *Frontiers in cardiovascular medicine* **2020**, *7*, 25.
26. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* **2014**.
27. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
28. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.
29. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In Proceedings of the International workshop on deep learning in medical image analysis. Springer, 2018, pp. 3–11.
30. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999* **2018**.
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
32. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.
33. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3–19.
34. Yao, W.; Bai, J.; Liao, W.; Chen, Y.; Liu, M.; Xie, Y. From cnn to transformer: A review of medical image segmentation models. *Journal of Imaging Informatics in Medicine* **2024**, *37*, 1529–1547.
35. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306* **2021**.
36. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-unet: Unet-like pure transformer for medical image segmentation. In Proceedings of the European conference on computer vision. Springer, 2022, pp. 205–218.
37. Urrea, C.; Vélez, M. Advances in Deep Learning for Semantic Segmentation of Low-Contrast Images: A Systematic Review of Methods, Challenges, and Future Directions. *Sensors* **2025**, *25*, 2043.
38. Qi, Y.; He, Y.; Qi, X.; Zhang, Y.; Yang, G. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2023, pp. 6070–6079.
39. Kaiser, L.; Gomez, A.N.; Chollet, F. Depthwise separable convolutions for neural machine translation. *arXiv preprint arXiv:1706.03059* **2017**.
40. Wu, Y.; He, K. Group normalization. In Proceedings of the Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3–19.
41. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62.
42. Yang, L.; Zhang, R.Y.; Li, L.; Xie, X. Simam: A simple, parameter-free attention module for convolutional neural networks. In Proceedings of the International conference on machine learning. PMLR, 2021, pp. 11863–11874.
43. Roy, S.; Koehler, G.; Ulrich, C.; Baumgartner, M.; Petersen, J.; Isensee, F.; Jaeger, P.F.; Maier-Hein, K.H. Mednext: transformer-driven scaling of convnets for medical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2023, pp. 405–415.
44. Qi, Y.; He, Y.; Qi, X.; Zhang, Y.; Yang, G. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2023, pp. 6070–6079.
45. Hou, Q.; Zhang, L.; Cheng, M.M.; Feng, J. Strip pooling: Rethinking spatial pooling for scene parsing. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 4003–4012.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s)

disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.