CoroJA_RetinaNet: A Multiscale Attention-Guided Framework for Automated Coronary Plaque Detection in CTA Images

Xuan Nie, Teng Li, Yinan Yuan, Zichen Yan, Yiwen Liu, Guangpu Zhou, and Bosong Chai

***Abstract*—Coronary heart disease is one of the most common cardiovascular diseases. Currently, CTA imaging has become the most widely used modality for its diagnosis. The detection of coronary plaque is an important basis for accurate diagnosis. In order to further improve the accuracy and efficiency of coronary plaque detection, this study proposes a series of automatic detection methods for coronary plaque based on deep learning. The approach begins by segmenting coronary arteries using a Transformer model integrated with an multi-resolution overlapping attention mechanism, thereby reducing interference in plaque detection. Subsequently, a two-stage hybrid strategy is employed for centerline extraction and optimization, and a multi-angle straightened surface reconstruction method is proposed to generate high-quality data for plaque detection. An improved RetinaNet (CoroJA_RetinaNet) is developed, integrating an attention mechanism, enhanced feature pyramid networks, and optimized post-processing strategies. Experimental results demonstrate that the proposed method significantly improves the accuracy and efficiency of coronary plaque detection compared to traditional approaches.**

Link to graphical and video abstracts, and to code: https://latamt.ieeer9.org/index.php/transactions/article/view/9889

***Index Terms*—Coronary heart disease, Coronary plaque detection, Surface reconstruction, Attention mechanism.**

I. Introduction

CORONARY heart disease is among the most common cardiovascular conditions today [1]. Its primary cause is the accumulation of plaques on vessel walls, leading to the narrowing of coronary arteries and subsequently impairing blood flow. Therefore, accurate detection and identification of coronary plaques provide a reliable basis for clinical diagnosis and pathological research, and its clinical value has significant importance in the medical field.

From the earliest rule-based and statistical models [2] to today's deep neural networks, the application of computer vision in various fields has continuously introduced new developments [3], [4], [5], driving the rapid advancement of artificial intelligence technology [6], [7]. With the development of medical imaging technologies, various modalities are available for detecting coronary artery plaques, including electrocardiography, nuclear myocardial imaging, hematological tests, IVUS, OCT, coronary CTA and CAG. Each method has its own advantages and disadvantages. In recent years, CTA has emerged as the preferred diagnostic modality due to its non-invasiveness, efficiency, low cost, and ease of operation [8]. However, clinicians often rely on visual assessment or semi-automatic evaluation after lumen segmentation [9]. These approaches are susceptible to observer bias and misjudgment or depend heavily on segmentation performance [10], making them time-consuming and labor-intensive. With the rapid development of deep learning in the field of computer vision, the use of deep learning technology to assist clinical diagnosis has been widely applied in various fields of medical image analysis [11]. More and more researchers are applying new technologies for the automated detection of coronary artery plaques [12], [13], [14].This study focuses on coronary CTA images and explores a deep learning–based approach for plaque detection, aiming to improve the accuracy and consistency of diagnostic results, enhance clinical efficiency, and reduce the risk of missed diagnoses or misdiagnoses. The proposed method holds practical significance for early disease detection and the standardization of diagnostic criteria.

Most current studies on coronary plaque detection consist of three key parts: coronary segmentation, surface reconstruction, and plaque detection [15], each of which significantly influences the final detection accuracy. The structure of the coronary arteries is complex and the images have strong background noise interference. This study employs a Transformer model integrated with an MRA mechanism to address data limitations and enhance multi-scale feature extraction. This model improves global semantic representation and segmentation performance. In the surface reconstruction stage, to overcome issues such as image blurring, single reconstruction angle, and difficulty in directly identifying plaques, a multi-angle straightening technique based on centerline extraction is introduced. This approach reduces information loss associated with single-angle reconstructions and produces clearer coronary surface views. Finally, to address the challenges posed by heterogeneous plaque features and varying plaque sizes, this study proposes a coronary plaque detection method based on improved feature pyramid and joint attention on the basis of the RetinaNet network. This enhances detection accuracy across different spatial scales. Furthermore, a novel anchor generation strategy is proposed, using maximum likelihood estimation to optimize the matching of features with various geometric layouts of objects, which can greatly alleviate the problem of data distribution without a center. In addition, a hard negative mining strategy is incorporated, selecting only a small number of challenging negative samples and all positive samples to participate in the loss calculation,

The associate editor coordinating the review of this manuscript and approving it for publication was Samuel Ortega (*Corresponding author: Teng Li*).

X. Nie, Teng Li, Y. Yuan, Z. Yan, Y. Liu, and G. Zhou are with School of Software, Northwestern Polytechnical University, Xian, China (e-mails: xnie@nwpu.edu.cn, liteng1@mail.nwpu.edu.cn, yinanyuan@mail.nwpu.edu.cn, liuyiwen@mail.nwpu.edu.cn, and zhurong@mail.nwpu.edu.cn).

B. Chai is with Zhejiang University, Zhejiang, China (e-mail: chaibosong@mail.zju.edu.cn).

effectively alleviating the problem of imbalanced sample data collection under a single background.

The main contributions of this work can be summarized as follows:

- A Transformer model with an MRA mechanism is proposed and a composite knowledge distillation model is constructed. This effectively improves the segmentation accuracy of coronary vessels.
- A multi-angle straightening method for coronary surface reconstruction based on the centerline, solving the issues of blurred images in coronary surface reconstruction, a single reconstruction angle, and the inability to directly detect plaques after segmentation of the coronary vessels.
- Addressing the issue of non-homogeneous and varying sizes of coronary artery plaques, a coronary artery plaque detection approach is proposed, based on an improved feature pyramid and joint attention, built upon the RetinaNet network, to enhance the accuracy of coronary artery plaque detection tasks.

TABLE I
LIST OF ABBREVIATIONS

Abbreviation	Full Term
CTA	Computed Tomography Angiography
IVUS	Intravascular Ultrasound
OCT	Optical Coherence Tomography
CAG	Coronary Angiography
ROI	Region Of Interest
CNN	Convolutional Neural Network
MRA	Multi-Resolution Overlapping Attention
SW-MSA	Shift Window Partitioning Multi-head Attention
MLP	Multi-Layer Perceptron
KDF	Knowledge Distillation Framework
SDF	Spatial Distillation Feature
CDF	Channel-wise Distillation Feature
MST	Minimum Spanning Tree
MPR	Multi-Planar Reconstruction
CPR	Curved Planar Reformation
NMS	Non-Maximum Suppression
FPN	Feature Pyramid Networks
SGD	Stochastic Gradient Descent

II. METHODS

A. Coronary Artery Segmentation

Coronary artery plaques are typically manifested as abnormal lesions located within the vessel wall. In medical imaging, they often appear blurred due to low contrast, noise interference, and the presence of artifacts. Directly performing plaque detection over the entire image is susceptible to interference from surrounding non-target tissues, which can result in a high false positive rate. Therefore, conducting accurate coronary artery segmentation prior to plaque detection not only effectively defines the ROI but also significantly enhances the accuracy and robustness of subsequent detection tasks [16], [17]. To address this, we propose a Transformer-based model that incorporates an MRA mechanism (shown in Fig. 1), aiming to comprehensively capture the multi-scale features of coronary structures from local to global levels, which in turn enhances segmentation accuracy [18], [19]. Unlike standard Transformer models, the proposed approach initially partitions the input feature map into fixed-size patches. Specifically, the patches used for generating keys and values are slightly larger and overlapping, whereas the query patches are non-overlapping. This design improves the ability of the attention module to better perceive spatial contextual information, enabling more effective discrimination between adjacent vascular structures and background regions. Furthermore, the attention module integrates a multi-scale shifted window structure (SW-MSA) to extract local feature representations at three different resolutions. The attention features from each scale are subsequently aligned and fused via bilinear upsampling, followed by an MLP layer to aggregate global semantic features. This mechanism not only improves the modeling of small vascular branches, but also expands the receptive field, in contrast to the fixed-scale and globally uniform attention computation in the vanilla Transformer. As a result, the model exhibits superior adaptability to structural variations and stronger global representation capabilities.

Additionally, in consideration of the practical requirements for model efficiency and inference speed in clinical settings, we further propose a lightweight segmentation framework that incorporates a KDF. As shown in Fig. 2, this framework adopts a teacher-student architecture, in which a high-capacity teacher network guides a lightweight student network to learn more robust feature representations, thus enhancing the model's generalization ability under challenging conditions, such as low-quality or partially missing data. To further improve the effectiveness of feature transfer between teacher and student networks, we introduce a novel collaborative distillation strategy that combines SDF and CDF [20], [21], [22]. In the SDF module (shown in Fig. 3b), the activation maps of the teacher and student networks are spatially aligned, and the activation values at each spatial position are normalized. The strategy minimizes both point-wise and pair-wise discrepancies to enable smoother and more effective knowledge transfer. To avoid overly rigid constraints that could lead to suboptimal solutions, we incorporate the CDF module (shown in Fig. 3a), which performs normalization at the channel level to produce soft probability maps. By minimizing the Kullback-Leibler (KL) divergence between the teacher and student probability distributions at the channel level [23], the distillation process focuses more effectively on the salient regions within each channel, thus improving the model's attention to coronary structures and optimizing diagnostic performance [24].

B. Centerline Extraction and Surface Reconstruction

1) A Hybrid Approach for Coronary Centerline Extraction and Optimization: Traditional methods such as topological thinning and the MST algorithm have been widely applied in centerline extraction, ensuring topological integrity and branch continuity of vessels [25]. However, these methods exhibit limitations when handling complex vascular structures, particularly in smoothness, irregularities, and small discontinuities. To address these shortcomings, a two-stage hybrid framework is proposed that combines classical image analysis techniques with deep learning (shown in Fig. 4). In the first stage, a topological thinning algorithm extracts the coronary

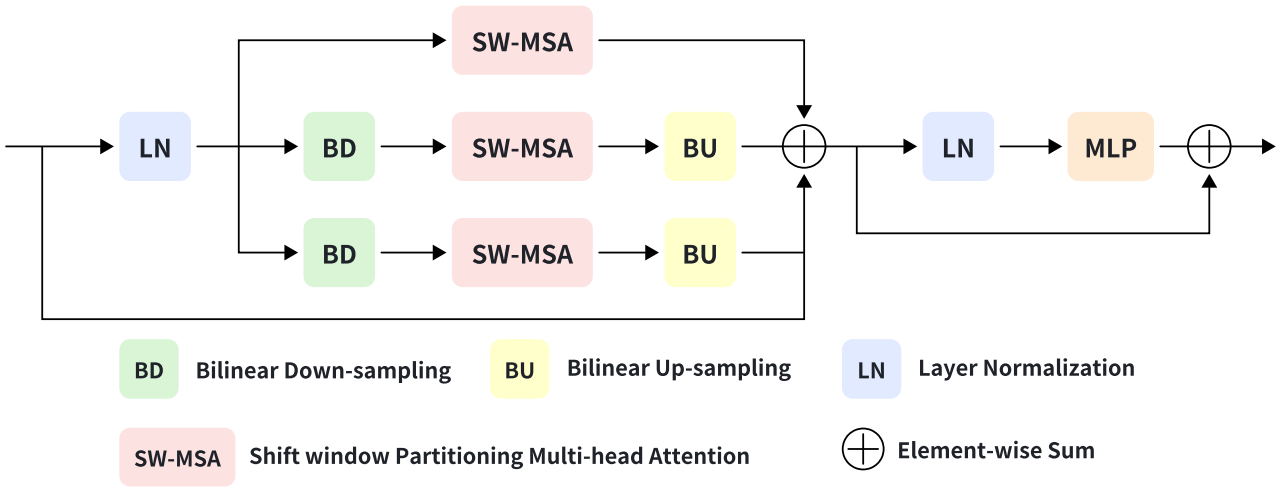


Fig. 1. A novel transformer model incorporating an MRA mechanism.

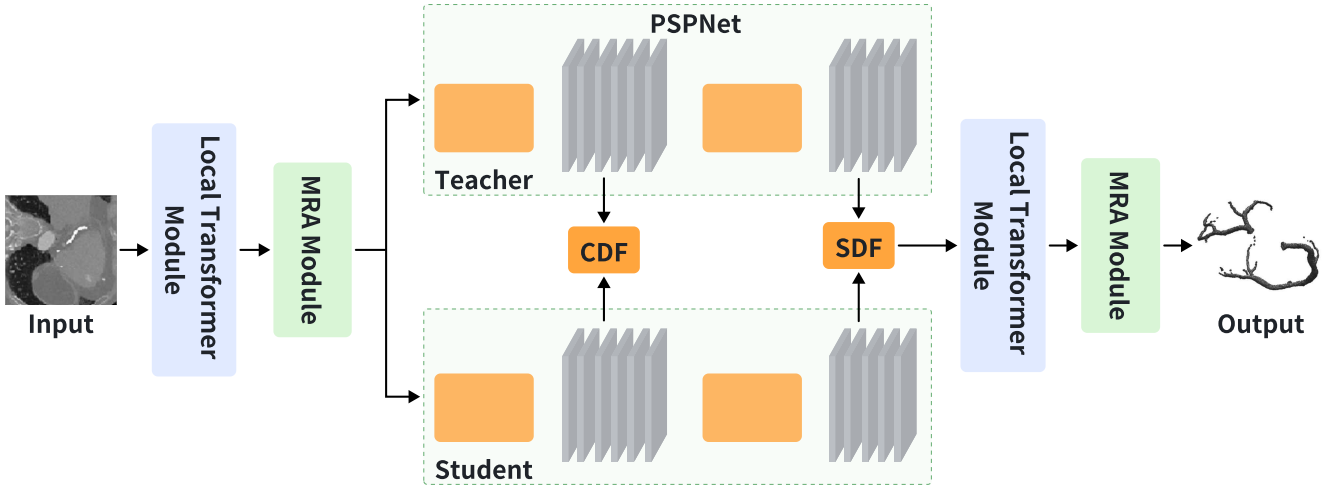


Fig. 2. The coronary artery segmentation model based on MRA proposed in this paper.

vessel skeleton by removing pixels that don't meet centerline criteria based on local neighborhood topology, resulting in a stable and continuous preliminary centerline. However, due to the anatomical complexity of coronary arteries, small gaps and local irregularities often remain, especially at bifurcations and intricate structures. Thus, the MST algorithm is also employed to connect discontinuous vessel branches by calculating Euclidean distances, ensuring centerline continuity [26]. Despite these enhancements, the skeleton line may still exhibit local roughness and irregularities. To further refine this, a deep learning-based optimization method is introduced in the second stage. A CNN is employed to predict the vessel radius precisely, utilizing convolutional layers to extract multi-scale features and fully connected layers to output radius predictions for each point. To enhance performance, dilated convolutions are integrated to expand the receptive field, allowing the aggregation of multi-scale features while maintaining a manageable number of trainable parameters [27]. This approach enables the refined network to smooth traditional skeleton outputs, eliminate small discontinuities and irregularities, and generate a more precise and smooth centerline.

2) Multi-Angle Straightened Surface Reconstruction: Although MPR provides comprehensive anatomical information from axial, coronal, and sagittal views, it cannot display curved vessels or organs in a single slice. Due to the prominent curvature and torsion of the coronary arteries, CPR is used. The accuracy of CPR reconstruction is determined by the centerline and the orientation of interest vectors. Adequate centerline sampling density is critical for precise reconstruction to avoid missing plaques due to eccentric growth or not accurately displaying plaque size and morphology in a single slice [28]. Therefore, multi-angle sampling is performed (shown on the

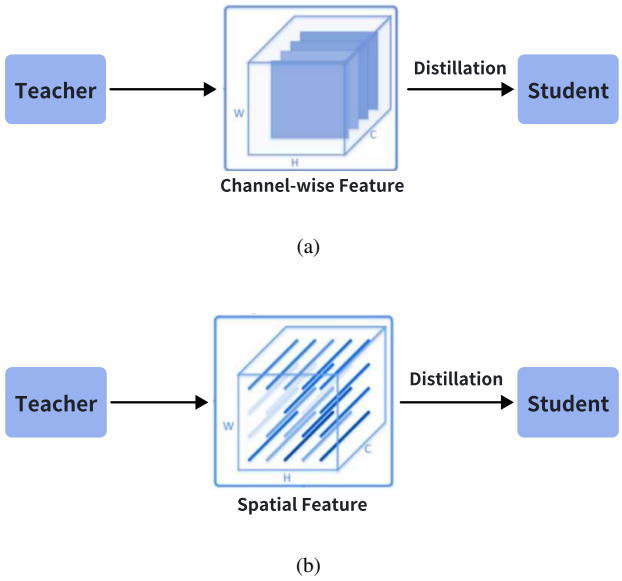


Fig. 3. Knowledge distillation modules in the proposed framework. (a) CDF module: Calculate soft probability maps based on teacher feature channels and shift semantic focus by minimizing channel level KL divergence. (b) SDF module: Align feature maps of teacher and student in spatial space, and guides learning by reducing point-wise and pair-wise differences.

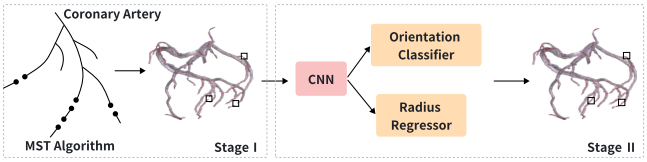


Fig. 4. A two-stage hybrid framework combining classical image analysis techniques with deep learning methods.

left side of Fig. 5), using four angles -0°, 45°, 90°, and 135°- to ensure plaques are captured regardless of growth direction, enhancing reconstruction accuracy [29], [30]. Traditional CPR methods suffer from limited orientation vectors, which don't preserve vessel geometric trends. To address this, an improved CPR method with three specific interest vectors is proposed at each center point (shown on the right side of Fig. 5). Three imaging planes are established for each center point, mapping sampled data along these three vector directions. This linear sampling strategy in multiple directions comprehensively captures the spatial features of vessels in complex coronary structures, reducing information loss. The results are shown in Fig. 6, where the plaques are clearly visible from multiple angles.

C. Coronary Plaque Detection

1) An Improved RetinaNet with Attention Mechanism: RetinaNet is a single-stage object detection model that introduces Focal Loss to address severe class imbalance [31]. It employs multiple prediction heads to output class probabilities and bounding box coordinates for each candidate object, achieving high accuracy and efficiency compared to two-stage detectors

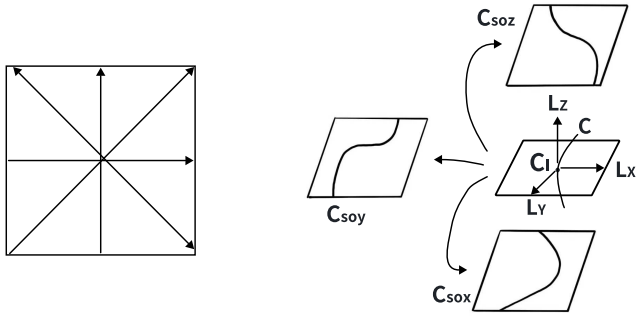


Fig. 5. Multi-angle straightened surface reconstruction.

such as Faster R-CNN [32]. This study constructs a coronary plaque detection network: CoroJA_RetinaNet (Coronary plaque detection network with improved feature pyramid and joint attention) for plaque detection tasks (shown in Fig. 7). The network uses ResNet50 as the backbone, coupled with Attention-ResNet for feature enhancement. During training, to balance classification and regression loss, a hard negative mining strategy is applied, and the classification loss is multiplied by a dynamically computed weight factor λ, ensuring balanced optimization Equation 1.The traditional Smooth L1 Loss used in RetinaNet for bounding box regression lacks sensitivity to small positional changes critical to detect small plaques. Therefore, a combined loss function integrating GIoU Loss with dynamic weighting is proposed to enhance localization precision for small plaque targets [33]. The dynamic weight factor adjusts automatically based on the target size and spatial distribution within the image [34], improving the sensitivity in small or densely packed regions Equation 2.

$$\lambda = Pow(10, E(N(reg_{\text{loss}})) - E(N(cls_{\text{loss}}))) . \quad (1)$$

$Pow()$ is a power function, $N()$ is a conversion of scientific notation, $E()$ refers to the index, reg_{loss} is regression loss and cls_{loss} is classification loss.

$$L_{GIOU} = 1 - IoU + \frac{\propto |C \setminus (A \cup B)|}{|C|}. \quad (2)$$

$\propto$ is the weight factor, A is the actual box area, B is the predicted box area, and C is the minimum rectangular box area that combines A and B; IoU is the intersection ratio of A and B. Then, an improved feature pyramid module is adopted to fuse features from different levels using top-down and horizontal connections. Finally, each feature layer is connected to a classification branch (class subnet) and a regression branch (box subnet) to predict the target category and bounding box. Use NMS to sort these suggestion boxes based on classification scores, remove redundant suggestion boxes, and select the suggestion boxes with higher scores as output.

2) Joint Attention Mechanism: In coronary CPR images, most areas are occupied by background or surrounding tissues, with plaques comprising only a small portion, making them visually subtle and difficult to distinguish. In addition,

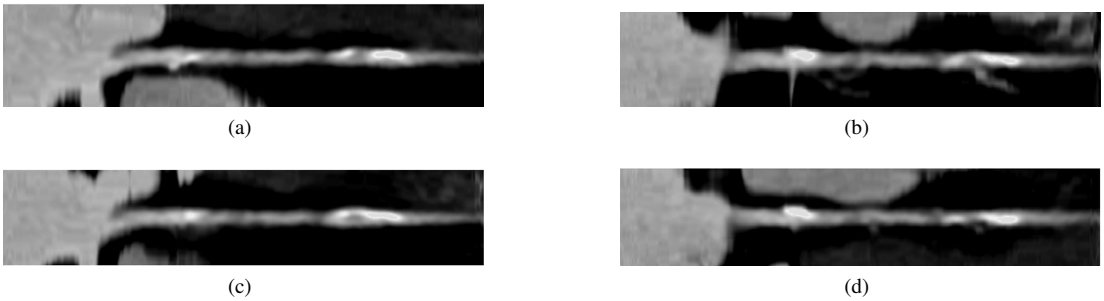


Fig. 6. The result of multi-angle curved planar reconstruction of the coronary arteries. (a) CPR straight in the 0°. (b) CPR straight in the 45°. (c) CPR straight in the 90°. (d) CPR straight in the 135°.

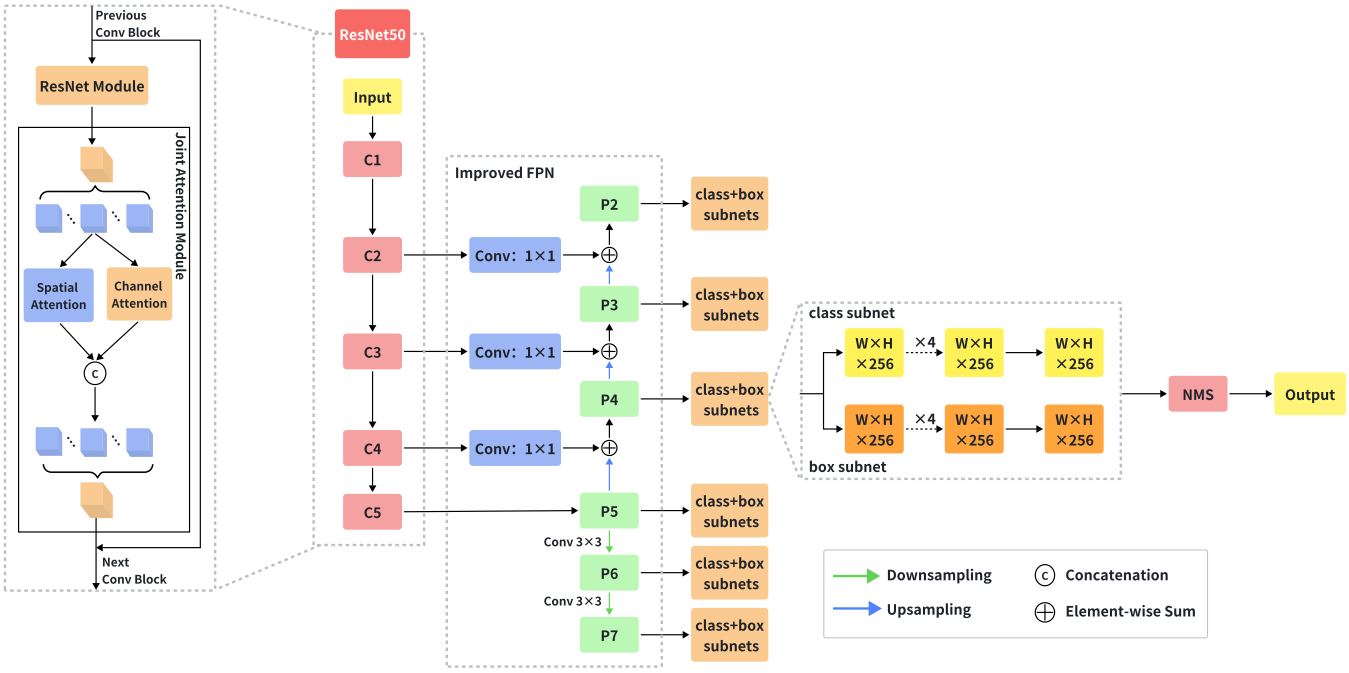


Fig. 7. The structure of CoroJA_RetinaNet. The model is based on ResNet50 and an improved FPN to extract multi-scale features (P2–P7). A joint spatial-channel attention module is integrated to enhance feature representation. Detection is performed through shared classification and box regression subnets, and final results are obtained after NMS.

plaque appearance varies with severity, and detecting mild lesions has significant clinical value. To enhance attention to bright plaque features while suppressing background noise and motion artifacts [35], [36], a joint attention mechanism combining channel and spatial attention is integrated into residual structures (Fig. 7). Unlike the sequential structure of CBAM [37], the proposed joint attention module uses a parallel fusion of channel and spatial attention modules for a more effective feature interaction, preserving the critical semantic information of the plaques. Channel attention functions as a "commander" coordinating feature flow between channels to strengthen semantic representation [38]. Spatial attention acts as a "scout" capturing key spatial pixels to enhance positional understanding and feature extraction.

3) Improved FPN: Detecting small plaques remains a significant challenge in coronary artery analysis. These tiny lesions typically exhibit weak texture and edge characteristics, which are prone to being suppressed during the extraction of high-level semantic features in deep convolutional networks. Although joint attention mechanisms have been introduced to improve feature representation, essential low-level details, such as fine textures and boundary cues, may still be lost due to progressive feature abstraction, thereby degrading the performance in small plaque detection.

To address multiscale detection tasks, FPN have been widely adopted due to their ability to integrate features at different semantic levels [39]. However, the original RetinaNet implementation only incorporates features from layers C3 and above in the ResNet backbone, omitting layer C2 that contains rich low-level spatial information crucial for small object detection [40]. In addition, the anchor matching strategy in RetinaNet is not specifically optimized for tiny targets, further limiting the detection accuracy for small plaques.

Specifically, RetinaNet generates anchor boxes from the

FPN layers P3, P4, P5, P6, P7, corresponding to strides of 8, 16, 32, 64, 128. For small plaques of approximately 16×16 pixels in the input image, the smallest anchor size (32×32) often leads to low IoU scores, falling below the common threshold of 0.5. During training, a target is assigned as a positive sample only if its IoU with an anchor exceeds the threshold. Otherwise, the anchor with the highest IoU is chosen. However, due to the limited overlap between small plaques and standard anchors, the number of positive samples is drastically reduced, compromising detection precision and training stability.

To overcome these limitations, we propose an improved FPN architecture that introduces low-level features to improve the network's ability to detect small-scale lesions. Concretely, we incorporate the C2 feature map from the backbone into the FPN and apply a 1×1 convolution to adjust its channel dimensions, resulting in a new P2 feature map with a finer spatial resolution (stride = 4). This higher resolution feature map better preserves edge and texture details, providing crucial support for detecting small targets [41]. The improved FPN structure is illustrated in Fig. 7.

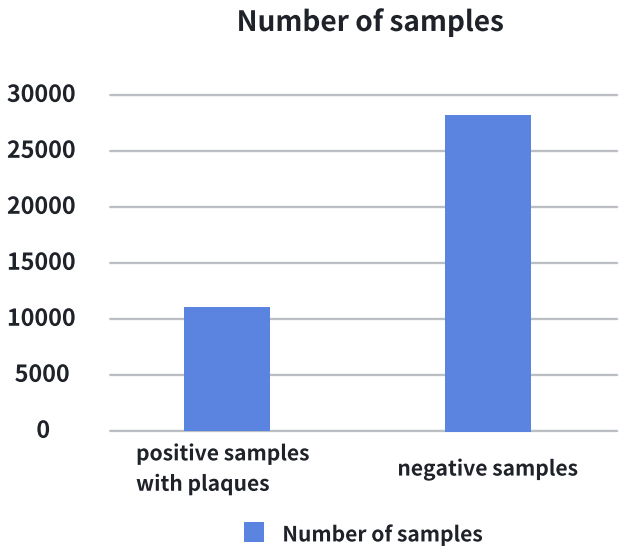


Fig. 8. Dataset sample distribution.

III. Results

A. Experimental Setup

1) Dataset: The data used in this study were derived from coronary CTA scans of 1,021 patients at Shanxi Provincial People's Hospital, with informed consent and ethical approval obtained for all cases. The raw images were acquired using a Siemens dual source spiral CT system with the following parameters: tube voltage 70–150 kV (routine 120 kV), tube current 210–300 mAs, reconstructed slice thickness 0.4 mm and a 512×512 pixel resolution. CT values ranged from -2048 to 2048 HU. During preprocessing, cases with severe metal artifacts or insufficient reconstruction quality were excluded. Using our proposed CPR algorithm (see Section 2), we extracted 7,943 complete coronary centerlines and generated 39,761 CPR images.

Three cardiovascular specialists annotated all CPR images using our in-house NPUCTA software, recording the start and end coordinates of the plaque regions. A three-tier quality control protocol ensured annotation accuracy: initial annotations by specialists, cross-validation to resolve discrepancies, and final approval by a senior physician. To prevent data bias, all CPR images—including plaque-negative cases—were systematically labeled. The dataset was partitioned into training and test sets at a 7:3 ratio, corresponding to 715 patients (27,833 CPR images) and 306 patients (11,928 CPR images), respectively.

The dataset contains approximately 30% positive samples with plaques and 70% negative samples without plaques, reflecting the natural distribution in the clinical population, shown in Fig. 8. Although detailed stratification by plaque type and clinical severity was not fully performed due to annotation limitations, efforts are underway to further analyze these aspects in future work. Basic demographic information such as patient age and gender was collected, and the dataset represents a balanced demographic distribution as far as available clinical records allow.

2) Implementation Details: Experiments were conducted using the CoroJA_RetinaNet architecture. To accelerate convergence, the model was initialized with weights pre-trained on the ImageNet classification task. Training was performed for 12 epochs with a batch size of 8. The SGD optimizer was used with an initial learning rate of 1e-3 and a weight decay of 1e-4. A linear warmup learning rate schedule was applied over the first 2,000 iterations, during which the learning rate gradually increased from a low value to the initial learning rate. After warmup, the learning rate remained constant for the remainder of training. For loss computation, a hard negative mining strategy was adopted to selectively retain challenging negative samples while including all positive samples. All annotations followed the VOC-2012 format, with plaque coordinates and vascular identifiers encoded in XML files. Training stability was monitored throughout and the final model achieved convergence with a loss of approximately 0.05. The average training time per epoch was approximately 15–20 minutes, resulting in a total training time of about 3.5 to 4 hours.

To ensure robust performance evaluation, the dataset was divided into separate training, validation and test sets, preserving the original class distribution, approximately 30% positive samples with plaques and 70% negative samples, across all subsets. This split strategy maintains consistency and mitigates bias in model assessment. Model performance was reported using key evaluation metrics, including AR, AP, and IoU, based on the VOC-2012 annotation style.

To address minor class imbalance and improve generalization, we applied loss weighting and data augmentation during training. The augmentation pipeline included random horizontal flipping and rotation within ±10°, enhancing the representation of underrepresented categories while preserving anatomical integrity.

TABLE II
ABLATION STUDY ON THE EFFECT OF CDF, SDF, AND MAR MODULES IN CORONARY ARTERY SEGMENTATION

Model	ACC	Se	Sp	Pre	IOU	F1	FLOPs
Teacher Model	0.9835	0.8774	0.9824	0.9671	0.8362	0.7905	86.26G
Student Model	0.9496	0.8239	0.9667	0.9604	0.7623	0.7438	2.49G
Student Model w/ CDM	0.9648	0.8629	0.9715	0.9652	0.8025	0.7781	2.49G
Student Model w/ CDF+SDF	0.9742	0.8735	0.9887	0.9681	0.8349	0.7993	2.49G
Student Model w/ CDF+SDF+MRA	**0.9852**	**0.8763**	**0.9891**	**0.9723**	**0.8358**	**0.8016**	**2.49G**

B. Experimental Analysis of Ablation

To validate the effectiveness of the proposed coronary vessel segmentation method, we systematically designed ablation experiments to evaluate the contributions of the MRA, CDF, and SDF modules. The baseline for comparison was a student network without any of the proposed modules, and the evaluation metrics included accuracy (ACC), sensitivity (Se), specificity (Sp), precision (Pre), IoU and F1 score. The results are summarized in Table II.

By introducing the CDF module, the student network achieved significant improvements in vessel continuity and background suppression. The CDF module increased ACC, Se, and IoU by 1.52%, 3.9%, and 4.02%, respectively. However, some noise artifacts remained in the distal vessel regions. Further integration of the SDF module enabled pixel-level feature refinement, correcting noisy regions and optimizing vessel edges. Compared to the baseline model, the combined CDF and SDF modules improved IoU and F1 score by 7.26% and 5.55%, respectively, while surpassing the teacher model in Sp and Pre by 0.63% and 0.1%. Note that the proposed modules (CDF, SDF, and MRA) are utilized only during training, and the final deployed student model excludes these components, resulting in reduced computational overhead and faster inference while maintaining high segmentation accuracy.

Finally, the complete model incorporating the MRA module achieved optimal performance in vessel continuity and fine-grained segmentation. The complete model (CDF + SDF + MRA) demonstrated improvements of 3.56%, 2.24%, 1.19%, and 5.78% in ACC, Sp, Pre, and F1 score, respectively, compared to the baseline model, with an IoU of 0.8358, only 0.04% lower than that of the teacher model (0.8362). Notably, the full model outperformed the teacher network in Sp (0.9891) and Pre (0.9723) by 0.67% and 0.52%, respectively, demonstrating the effectiveness of the multi-module collaboration in extracting multi-scale vessel features. The experimental results indicate that the cascaded design of channel-spatial distillation and multi-resolution attention mechanisms is crucial for improving the accuracy of coronary vessel segmentation.

To validate the effectiveness of the proposed coronary plaque detection model integrating an enhanced feature pyramid and joint attention mechanisms, we systematically conducted ablation experiments. As summarized in Table III, the baseline RetinaNet model (Row 2) achieved an average precision (AP), recall (AR) and IoU of 0.6194, 0.5403, and 0.5929, respectively. Upon integrating the fusion attention module (Row 3), these metrics improved by 3.43%, 8.97% and 3.09%, with the attention mechanism notably enhancing focus on plaque regions in complex backgrounds, shown in Fig. 9.

Further incorporating the joint attention module and improved FPN (Row 4) yielded additional gains of 3.57%, 4.2%, and 4.01%, attributed to the FPN's enhanced multi-scale feature representation, particularly for small plaques (shown in Fig. 10).

The addition of the enhanced NMS module (Row 5) further boosted performance by 6.19%, 5.67%, and 5.79%, effectively reducing missed and false detections in densely clustered plaque regions (shown in Fig. 11). The introduction of hard negative mining (Row 6) improved metrics by 6.46%, 4.87% and 6.07%, optimizing the ability of the model to handle challenging samples. The complete model (Row 7), combining all five enhancements, achieved final improvements of 5.9%, 5.25%, and 6.68%, with GPU utilization increasing from 103.62 to 302.58 and detection time decreasing from 11.4 to 5.2 s. These results demonstrate that the progressive integration of attention mechanisms, multi-scale feature refinement, and post-processing optimization synergistically improve both detection accuracy and efficiency (shown in Fig. 12). Visual example of comparing the predictions of CoroJA_RetinaNet with those of the baseline model in challenging situations is shown in Fig. 13

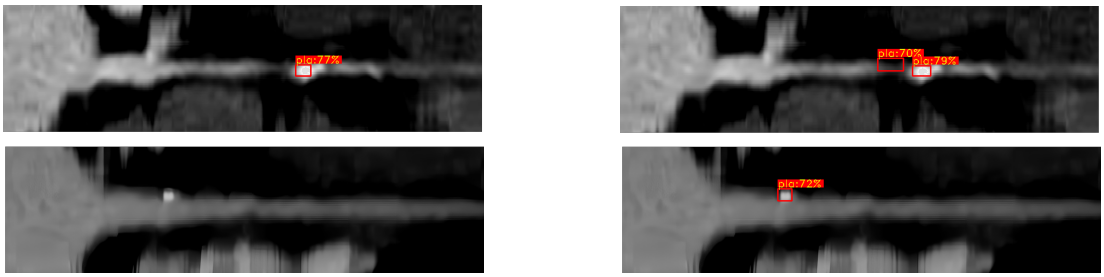
To further validate our approach, we compared the proposed method with several classical object detection models, including R-CNN [42], Fast R-CNN [43], Faster R-CNN [32], YOLOv5 [44] and CenterNet [45]. The results are summarized in Table IV.

To ensure fair and controlled comparisons, all models were trained and evaluated on the same fixed training, validation, and test splits described earlier. The same preprocessing pipeline was consistently applied in all models, including intensity normalization and data augmentation techniques, specifically random horizontal flipping and small-angle rotation (within ±10°). All models were trained for 12 epochs with early stopping based on validation loss. Hyperparameters such as learning rate and optimizer settings were kept consistent when applicable, or individually tuned according to official implementation guidelines. All experiments were conducted on the same hardware and software environment to ensure consistent measurement of computational efficiency.

The proposed method, based on an enhanced RetinaNet, achieves a strong balance between accuracy and efficiency for coronary plaque detection. It outperforms Faster R-CNN in AP (0.874 vs. 0.800) and AR (0.815 vs. 0.770) while

TABLE III
ABLATION STUDY ON THE EFFECT OF DIFFERENT MODULES (RETINANET AS THE BASELINE MODEL) IN CORONARY PLAQUE DETECTION

Joint Attention Module	Improved Feature Pyramid	Improved NMS	Hard Negative Mining	GIoU Loss	AP	AR	IoU
×	×	×	×	×	0.6194	0.5403	0.5929
✓	×	×	×	×	0.6537	0.5927	0.6238
✓	✓	×	×	×	0.6894	0.6347	0.6639
✓	✓	✓	×	×	0.7513	0.6914	0.7218
✓	✓	✓	✓	×	0.8159	0.7401	0.7825
✓	✓	✓	✓	✓	**0.8749**	**0.7926**	**0.8493**



(a) (b)

Fig. 9. Comparison of plaque detection results with and without the joint attention mechanism. (a) Without attention. (b) With attention.

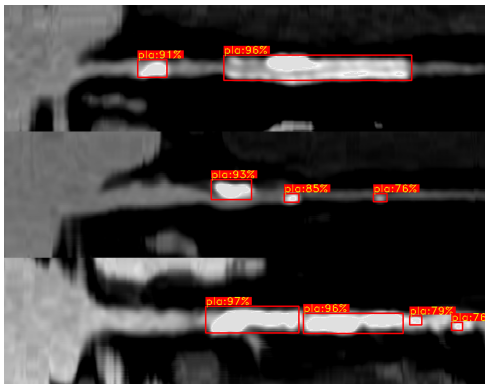


Fig. 10. Detection performance on plaques of different scales after incorporating the improved FPN.

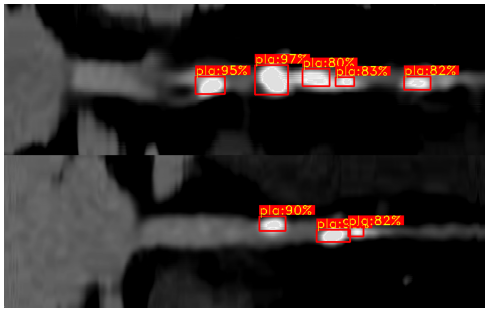


Fig. 11. Detection performance on densely clustered plaques after integrating the enhanced NMS module.

maintaining a lower computational cost (302 GFLOPs vs. 220 GFLOPs) and faster inference speed (9.7 FPS vs. 8.2 FPS). Compared to other single-stage models like YOLOv5 and CenterNet, it delivers higher accuracy with moderate GFLOPs, demonstrating its suitability for real-time clinical applications.

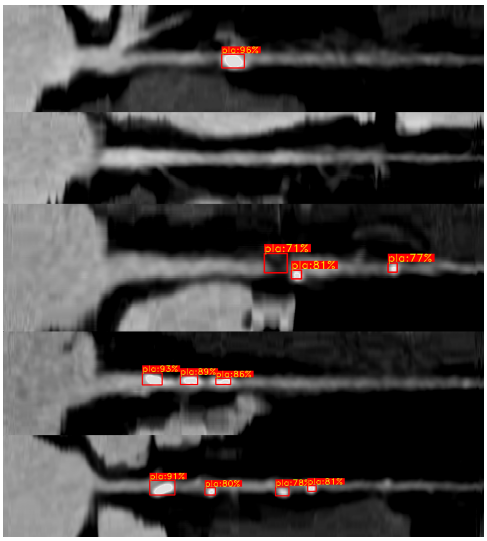
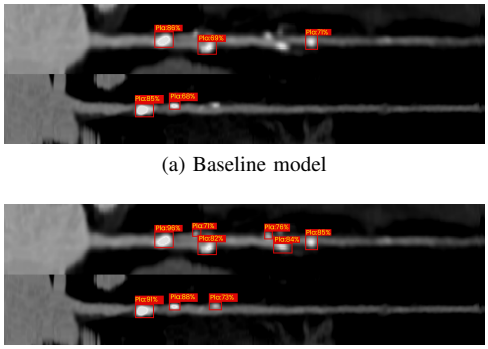


Fig. 12. Detection performance of coronary plaques across multiple samples using the proposed method.

TABLE IV
COMPARISON OF CORONARY PLAQUE DETECTION PERFORMANCE WITH REPRESENTATIVE OBJECT DETECTION MODELS

Model	AP	AR	IoU	GFLOPs	FPS
R-CNN	0.550	0.520	0.550	520	0.3
Fast R-CNN	0.705	0.680	0.670	250	5.4
Faster R-CNN	0.800	0.770	0.790	220	8.2
Nie 2025	**0.892**	**0.812**	**0.863**	300	6.2
YOLOv5	0.730	0.710	0.690	130	**35.4**
CenterNet	0.795	0.690	0.700	265	13.4
Proposed Method	0.874	**0.815**	0.850	**302**	9.7



(a) Baseline model

(b) CoroJA_RetinaNet

Fig. 13. Comparison of prediction results between CoroJA_RetinaNet and the baseline model.

IV. Conclusion

This study proposes a deep learning-based method for automatic coronary plaque detection, aiming to address the limitations of traditional approaches that rely on manual feature extraction and complex post-processing. Unlike previous methods, our approach significantly improves the accuracy and efficiency of coronary plaque detection through an enhanced FPN, a joint attention mechanism, and optimized post-processing strategies. Experimental results demonstrate that the proposed method performs exceptionally well in coronary plaque detection tasks, providing reliable technical support for the prevention and diagnosis of coronary artery disease.

In terms of coronary segmentation and CPR, this study introduces a coronary vessel segmentation method based on composite knowledge distillation and a multi-angle straightened CPR method based on centerline extraction. The former constructs a lightweight and deployable deep learning model by integrating spatial and channel-wise knowledge distillation, effectively addressing the high complexity and time-consuming nature of traditional segmentation methods. The latter extracts coronary skeletons using a thinning algorithm, optimizes centerline positions with an MST algorithm and CNNs, and achieves multi-angle CPR to mitigate information loss in single-angle reconstruction. These contributions provide a high-quality data foundation for plaque detection.

The core contribution of this study lies in the proposed improved RetinaNet model for automatic coronary plaque detection. By incorporating a joint attention mechanism, an enhanced FPN, optimized NMS, hard negative mining, and GIoU loss, the model significantly improves detection performance. Specifically, the joint attention mechanism enhances the model's focus on plaque regions, particularly improving detection in complex backgrounds. The improved FPN effectively boosts the model's ability to detect plaques of varying scales, especially improving the recognition accuracy of small plaques. The optimized NMS method excels in dense plaque regions, effectively reducing missed and false detections and enhancing detection robustness. Experimental results show that, compared to traditional methods, the proposed method eliminates the need for manual feature extraction, automatically learns plaque features, and achieves significant improvements in accuracy, recall, and IoU.

In summary, the proposed method represents a significant advancement in computer-aided diagnosis of coronary artery disease. Through improved feature extraction and post-processing strategies, this study not only enhances the accuracy of coronary plaque detection but also provides an efficient and reliable tool for early screening and diagnosis of coronary artery disease. The diagnosis of plaque-free results offers valuable references for clinicians to assess the health of patients' coronary arteries, thereby reducing unnecessary further treatments. In the future, with the accumulation of more high-quality data and further algorithmic optimization, the proposed method is expected to play an even greater role in clinical practice.

Furthermore, our method is developed and validated on straightened CPR images, which reconstruct vascular longitudinal sections by dynamically following the vessel centerline. However, other CPR reconstruction strategies, such as projection CPR or stretched CPR, are also commonly used in practice. These variants provide alternative representations of vessel morphology, and it is plausible that our model could be adapted or directly applied to datasets using these different CPR types, facilitating broader applicability across imaging protocols.

In addition, since atherosclerotic plaques can occur in various arteries beyond the coronary vessels, the proposed centerline-based longitudinal analysis approach may be transferable to other vascular regions, such as carotid or peripheral arteries. Currently, our experiments are limited to a single internal dataset. Evaluation on external datasets is planned as future work to further validate the robustness and generalizability of the proposed method.

Acknowledgments

This work was supported by the 2024 Shaanxi Provincial Key R&D Program (Project No.2024SF-GJHX-35), the 2025 Shaanxi Provincial Key R&D Program (Project No.2025QCY-KXJ-068), and the 2024 Xi'an Science and Technology Program (Project No.24KGDW0009).

References

[1] W. Huang, L. Huang, Z. Lin, S. Huang, Y. Chi, J. Zhou, J. Zhang, R.-S. Tan, and L. Zhong, "Coronary artery segmentation by deep learning neural networks on computed tomographic coronary angiographic images," in *2018 40th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2018, pp. 608–611, doi:10.1109/EMBC.2018.8512328.

[2] P. Feng and X. Peng, "A note on monge–kantorovich problem," *Statistics & Probability Letters*, vol. 84, pp. 204–211, 2014, https://doi.org/10.1016/j.spl.2013.10.011.

[3] Z. Zhang, N. A. S. Janvekar, P. Feng, and N. BHASKAR, "Graph-based detection of abusive computational nodes," Feb. 11 2025, uS Patent 12,223,056.

[4] L. Sun, W. Shi, X. Tian, J. Li, B. Zhao, S. Wang, and J. Tan, "A plane stress measurement method for cfrp material based on array lcr waves," *NDT & E International*, vol. 151, p. 103318, 2025, https://doi.org/10.1016/j.ndteint.2024.103318.

[5] W. Lu, J. Wang, T. Wang, K. Zhang, X. Jiang, and H. Zhao, "Visual style prompt learning using diffusion models for blind face restoration," *Pattern Recognition*, vol. 161, p. 111312, 2025, https://doi.org/10.1016/j.patcog.2024.111312.

[6] Y. Zhou, H. Xia, D. Yu, J. Cheng, and J. Li, "Outlier detection method based on high-density iteration," *Information Sciences*, vol. 662, p. 120286, 2024, https://doi.org/10.1016/j.ins.2024.120286.

[7] H. Hao, E. Yao, L. Pan, R. Chen, Y. Wang, and H. Xiao, "Exploring heterogeneous drivers and barriers in maas bundle subscriptions based on the willingness to shift to maas in one-trip scenarios," *Transportation Research Part A: Policy and Practice*, vol. 199, p. 104525, 2025, https://doi.org/10.1016/j.tra.2025.104525.

[8] M. J. Budoff, D. Dowe, J. G. Jollis, M. Gitter, J. Sutherland, E. Halamert, M. Scherer, R. Bellinger, A. Martin, R. Benton *et al.*, "Diagnostic performance of 64-multidetector row coronary computed tomographic angiography for evaluation of coronary artery stenosis in individuals without known coronary artery disease: results from the prospective multicenter accuracy (assessment by coronary computed tomographic angiography of individuals undergoing invasive coronary angiography) trial," *Journal of the American College of Cardiology*, vol. 52, no. 21, pp. 1724–1732, 2008, https://doi.org/10.1016/j.jacc.2008.07.031.

[9] R. C. Cury, S. Abbara, S. Achenbach, A. Agatston, D. S. Berman, M. J. Budoff, K. E. Dill, J. E. Jacobs, C. D. Maroules, G. D. Rubin *et al.*, "Cad-radstm coronary artery disease–reporting and data system. an expert consensus document of the society of cardiovascular computed tomography (scct), the american college of radiology (acr) and the north american society for cardiovascular imaging (nasci). endorsed by the american college of cardiology," *Journal of cardiovascular computed tomography*, vol. 10, no. 4, pp. 269–281, 2016, https://doi.org/10.1016/j.jcct.2016.04.005.

[10] A. Arbab-Zadeh and J. Hoe, "Quantification of coronary arterial stenoses by multidetector ct angiography in comparison with conventional angiography: methods, caveats, and implications," *JACC: Cardiovascular Imaging*, vol. 4, no. 2, pp. 191–202, 2011, https://doi.org/10.1016/j.jcmg.2010.10.011.

[11] S.-C. Lo, S.-L. Lou, J.-S. Lin, M. T. Freedman, M. V. Chien, and S. K. Mun, "Artificial convolution neural network techniques and applications for lung nodule detection," *IEEE transactions on medical imaging*, vol. 14, no. 4, pp. 711–718, 1995, doi: 10.1109/42.476112.

[12] X. Nie, B. Chai, K. Zhang, C. Liu, Z. Li, R. Huang, Q. Wei, M. Huang, and W. Huang, "Improved cascade-rcnn for automatic detection of coronary artery plaque in multi-angle fusion cpr images," *Biomedical Signal Processing and Control*, vol. 99, p. 106880, 2025, https://doi.org/10.1016/j.bspc.2024.106880.

[13] M. M. Jawaid, A. Riaz, R. Rajani, C. C. Reyes-Aldasoro, and G. Slabaugh, "Framework for detection and localization of coronary non-calcified plaques in cardiac cta using mean radial profiles," *Computers in biology and medicine*, vol. 89, pp. 84–95, 2017, https://doi.org/10.1016/j.compbiomed.2017.07.021.

[14] A. Abdolmanafi, L. Duong, R. Ibrahim, and N. Dahdah, "A deep learning-based model for characterization of atherosclerotic plaque in coronary arteries using optical coherence tomography images," *Medical Physics*, vol. 48, no. 7, pp. 3511–3524, 2021, https://doi.org/10.1002/mp.14909.

[15] X. Nie, Z. Yan, Q. Wei, B. Chai, M. Huang, L. Li, H. Dang, and Y. Liu, "Attention res-unet for coronary cpr image stenosis assessment," in *Seventh International Conference on Intelligent Medicine and Image Processing (IMIP 2025)*, vol. 13658. SPIE, 2025, pp. 66–75, https://doi.org/10.1117/12.3071669.

[16] O. Petit, N. Thome, C. Rambour, L. Themyr, T. Collins, and L. Soler, "U-net transformer: Self and cross attention for medical image segmentation," in *Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 12*. Springer, 2021, pp. 267–276, https://doi.org/10.1007/978-3-030-87589-3_28.

[17] X. Huang, Z. Deng, D. Li, and X. Yuan, "Missformer: An effective medical image segmentation transformer," *arXiv preprint arXiv:2109.07162*, 2021, https://doi.org/10.48550/arXiv.2109.07162.

[18] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890, doi: 10.1109/CVPR.2017.660.

[19] H. Du, J. Wang, M. Liu, Y. Wang, and E. Meijering, "Swinpa-net: Swin transformer-based multiscale feature pyramid aggregation network for medical image segmentation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 4, pp. 5355–5366, 2022, doi: 10.1109/TNNLS.2022.3204090.

[20] Q. Cai, R. Chen, L. Li, C. Huang, H. Pang, Y. Tian, M. Di, M. Zhang, M. Ma, D. Kong *et al.*, "The application of knowledge distillation toward fine-grained segmentation for three-vessel view of fetal heart ultrasound images," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, p. 1765550, 2022, https://doi.org/10.1155/2022/1765550.

[21] B. Murugesan, S. Vijayarangan, K. Sarveswaran, K. Ram, and M. Sivaprakasam, "Kd-mri: A knowledge distillation framework for image reconstruction and image restoration in mri workflow," in *Medical imaging with deep learning*. PMLR, 2020, pp. 515–526, https://proceedings.mlr.press/v121/murugesan20a.html.

[22] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015, https://doi.org/10.48550/arXiv.1503.02531.

[23] L. Zhao, X. Qian, Y. Guo, J. Song, J. Hou, and J. Gong, "Mskd: Structured knowledge distillation for efficient medical image segmentation," *Computers in Biology and Medicine*, vol. 164, p. 107284, 2023, https://doi.org/10.1016/j.compbiomed.2023.107284.

[24] Y. Choi, M. A. Al-Masni, K.-J. Jung, R.-E. Yoo, S.-Y. Lee, and D.-H. Kim, "A single stage knowledge distillation network for brain tumor segmentation on limited mr image modalities," *Computer Methods and Programs in Biomedicine*, vol. 240, p. 107644, 2023, doi: 10.1016/j.cmpb.2023.107644.

[25] J. Han, D. Zhang, G. Cheng, N. Liu, and D. Xu, "Advanced deep-learning techniques for salient and category-specific object detection: a survey," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 84–100, 2018, doi: 10.1109/MSP.2017.2749125.

[26] J. M. Wolterink, R. W. van Hamersvelt, M. A. Viergever, T. Leiner, and I. Išgum, "Coronary artery centerline extraction in cardiac ct angiography using a cnn-based orientation classifier," *Medical image analysis*, vol. 51, pp. 46–60, 2019, https://doi.org/10.1016/j.media.2018.10.005.

[27] G. Xu, H. Cao, J. K. Udupa, Y. Tong, and D. A. Torigian, "Disegnet: A deep dilated convolutional encoder-decoder architecture for lymph node segmentation on pet/ct images," *Computerized Medical Imaging and Graphics*, vol. 88, p. 101851, 2021, https://doi.org/10.1016/j.compmedimag.2020.101851.

[28] H. Yang, X. Zhen, Y. Chi, L. Zhang, and X.-S. Hua, "Cpr-gcn: conditional partial-residual graph convolutional network in automated anatomical labeling of coronary arteries," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 3803–3811, doi: 10.1109/CVPR42600.2020.00386.

[29] Y. Wang, W. Zhou, T. Jiang, X. Bai, and Y. Xu, "Intra-class feature variation distillation for semantic segmentation," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16*. Springer, 2020, pp. 346–362, https://doi.org/10.1007/978-3-030-58571-6_21.

[30] A. Tejero-de Pablos, H. Yamane, Y. Kurose, J. Iho, Y. Tokunaga, M. Horie, K. Nishizawa, Y. Hayashi, Y. Koyama, and T. Harada, "Improving segmentation of calcified and non-calcified plaques on cctacpr scans via masking of the artery wall," in *Medical Imaging 2023: Computer-Aided Diagnosis*, vol. 12465. SPIE, 2023, pp. 460–468, https://doi.org/10.1117/12.2652895.

[31] Z. Gao, L. Wang, R. Soroushmehr, A. Wood, J. Gryak, B. Nallamothu, and K. Najarian, "Vessel segmentation for x-ray coronary angiography using ensemble methods with deep learning and filter-based features," *BMC Medical Imaging*, vol. 22, no. 1, p. 10, 2022, https://doi.org/10.1186/s12880-022-00734-4.

[32] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016, doi: 10.1109/TPAMI.2016.2577031.

[33] Y. Luo, X. Cao, J. Zhang, P. Cheng, T. Wang, and Q. Feng, "Dynamic multi-scale loss optimization for object detection," *Multimedia Tools and Applications*, vol. 82, no. 2, pp. 2349–2367, 2023, https://doi.org/10.1007/s11042-022-13164-9.

[34] D. Zhang and H. Ma, "Yolodcc: Improved yolov8 combined with dynamic confidence compensation for lightweight moving object detection," *IET Image Processing*, vol. 18, no. 12, pp. 3699–3715, 2024, https://doi.org/10.1049/ipr2.13207.

[35] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125, doi: 10.1109/CVPR.2017.106.

[36] R. Li, M. Li, J. Li, and Y. Zhou, "Connection sensitive attention u-net for accurate retinal vessel segmentation," *arXiv preprint arXiv:1903.05558*, 2019, https://doi.org/10.48550/arXiv.1903.05558.

[37] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19, https://doi.org/10.48550/arXiv.1807.06521.

[38] R. Gu, G. Wang, T. Song, R. Huang, M. Aertsen, J. Deprest, S. Ourselin, T. Vercauteren, and S. Zhang, "Ca-net: Comprehensive attention convolutional neural networks for explainable medical image segmentation," *IEEE transactions on medical imaging*, vol. 40, no. 2, pp. 699–711, 2020, doi: 10.1109/TMI.2020.3035253.

[39] Z. Jiang, C. Ou, Y. Qian, R. Rehan, and A. Yong, "Coronary vessel segmentation using multiresolution and multiscale deep learning," *Informatics in Medicine Unlocked*, vol. 24, p. 100602, 2021, https://doi.org/10.1016/j.imu.2021.100602.
[40] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 658–666, doi: 10.1109/CVPR.2019.00075.
[41] Y. Gong, X. Yu, Y. Ding, X. Peng, J. Zhao, and Z. Han, "Effective fusion factor in fpn for tiny object detection," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2021, pp. 1160–1168, doi: 10.1109/WACV48630.2021.00120.
[42] R. B. Girshick, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Columbus, OH, USA: IEEE, 2014, pp. 580–587, doi: 10.1109/CVPR.2014.81.
[43] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448, doi: 10.1109/TPAMI.2016.2577031.
[44] G. Jocher, "Yolov5: The next generation of object detection," https://github.com/ultralytics/yolov5, 2020, accessed: 2025-04-14.
[45] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "Centernet: Keypoint triplets for object detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6569–6578, doi: 10.1109/ICCV.2019.00667.

Xuan Nie holds a Ph.D. in Computer Application Technology from Northwestern Polytechnical University.His research interests include computer vision, intelligent car assisted driving ADAS3, artificial intelligence, and medical imaging big data processing and analysis(coronary heart disease, cerebral infarction, and Parkinson's disease assisted discrimination). He is the author of 2 textbooks and several publications in core journals and important international conferences both domestically and internationally.

Teng Li holds a B.S. degree in Internet of Things Engineering from Chongqing University of Education. She is currently pursuing a M.S. degree in Software Engineering at Northwestern Polytechnical University. Her research interests include computer vision, artificial intelligence, and medical imaging big data processing and analysis.

Yinan Yuan holds a B.S. degree in Computer Science and Artificial Intelligence from Zhengzhou University. He is currently pursuing a M.S. degree in Software Engineering at Northwestern Polytechnical University. His research interests include computer vision, artificial intelligence, and medical imaging big data processing and analysis.

Zichen Yan holds a B.S. degree in Computer and Financial Engineering from Nanjing University. He is currently pursuing a M.S. degree in Software Engineering at Northwestern Polytechnical University. His research interests include computer vision, artificial intelligence, and medical imaging big data processing and analysis.

Yiwen Liu holds a B.S. degree in Software Engineering from Northwestern Polytechnical University. He is currently pursuing a M.S. degree in Software Engineering at Northwestern Polytechnical University. His research interests include computer vision, artificial intelligence, and medical imaging big data processing and analysis.

Guangpu Zhou holds a B.S. degree from the School of Literature at Henan University and a M.S. degree in Software Engineering from Northwestern Polytechnical University. His research interests include computer vision, artificial intelligence, and medical imaging big data processing and analysis.

Bosong Chai holds a B.S. degree from Lanzhou University in 2019 and a M.S. degree from Northwestern Polytechnical University in 2023. He is currently pursuing a Ph.D. degree in the College of Computer Science and Technology at Zhejiang University. His research interests include Brain-inspired Computing and Computer Vision.