WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 8, Number 4, Year 2026
DOI: https://doi.org/10.62051/ijcsit.v8n4.06
Journal homepage: https://wepub.org/index.php/IJCSIT/index

MSCA++-UNet: Image Segmentation Model with Coordinate and Channel Attention Mechanisms

Yan Gao, Xinru Xue, Jiaxin Zhang, Shengnan Dai, Jiayao Chen, Jincan Zhang *

College of Information Engineering, Henan University of Science and Technology, Luoyang, China
*Corresponding Author: Jincan Zhang

ABSTRACT

Retinal vessel segmentation in fundus images is critical for ophthalmic and systemic disease diagnosis, yet it is challenged by intricate vessel structures, low image contrast, and fine detail loss in feature extraction. Existing U-Net-based deep learning methods suffer from insufficient channel feature utilization, incomplete single-dimension attention enhancement, and poor fine vessel continuity, while lightweight models trade accuracy for efficiency, limiting clinical application. This paper proposes MSCA++-UNet,a novel segmentation model integrating a Multi-Scale Coordinate and Channel Attention Plus Plus (MSCA++) dual module into the U-Net backbone. The MSCA++ module realizes parallel computation and early fusion of multi-scale coordinate and adaptive channel attention, capturing spatial positional dependencies and dynamic channel correlations synergistically; combined with the lightweight AttentionDoubleConv block and optimized encoder-decoder fusion, it enhances edge preservation and feature discrimination with high computational efficiency. Extensive experiments on the DRIVE dataset (comparative, ablation, quantitative/qualitative analyses) show that MSCA++-UNet achieves a Dice coefficient of 0.787, mIoU of 79.7% and loss of 0.3624, outperforming baseline and single-attention variants in key metrics. It improves fine vessel detection accuracy and continuity, suppresses background noise, and balances performance (2.92 M parameters, 7.48 G FLOPs) with efficiency for edge device deployment. This work verifies the dual attention fusion strategy's effectiveness, and MSCA++-UNet provides a high-precision, clinically feasible solution for fundus image analysis, supporting computer-aided diagnosis and large-scale retinal screening.

KEYWORDS

Retinal vessel segmentation; MSCA++-UNet; Multi-scale coordinate attention

1. INTRODUCTION

Retinal vasculature plays a central role in the diagnosis, grading, and prognosis of a wide range of ophthalmic and systemic diseases, including diabetic retinopathy (DR), hypertensive retinopathy (HR), glaucoma, and age-related macular degeneration, which remain among the leading causes of irreversible blindness worldwide. As Arsalan et al. emphasized, "retinal vessels are considered important biomarkers for the detection of retinal diseases, like diabetic retinopathy and hypertensive retinopathy" [1] with subtle morphological changes—such as vessel narrowing, tortuosity, microaneurysms, and neovascularization—serving as essential indicators for early disease detection and treatment planning. Consequently, accurate segmentation of retinal blood vessels is a critical prerequisite for computer-aided diagnosis (CAD) systems and large-scale screening programs.

Manual delineation of retinal vessels is time-consuming, subjective, and prone to missing fine vascular structures, and the complexity of fundus images—including uneven illumination, low

contrast, and the presence of extremely thin capillaries—further increases the difficulty of manual annotation. Automated retinal vessel segmentation has therefore become indispensable for improving diagnostic efficiency and reducing clinical workload. Early segmentation methods relied on handcrafted filters, morphological operators, and thresholding strategies [2-4]; while computationally efficient, these methods are highly sensitive to noise and illumination variations, often failing to capture the complex curvilinear structures of retinal vessels.

With the rapid development of deep learning, convolutional neural networks (CNNs) have become the dominant paradigm for retinal vessel segmentation, as encoder–decoder architectures such as U-Net [5], U-Net++ [6], SegNet [7], and their variants have demonstrated remarkable performance improvements by leveraging multi-scale feature extraction and skip connections. However, classical U-Net still suffers from several inherent limitations: continuous downsampling operations inevitably weaken spatial gradients and lead to the loss of fine vessel boundaries, a degradation that is particularly detrimental for detecting thin vessels crucial for early disease diagnosis, as Li et al. noted that "continuous convolution and downsampling operations will lose the spatial features of blood vessels, such as edge information" [3].

To address these issues, recent studies have introduced attention mechanisms to enhance feature representation, with spatial attention highlighting important regions [8] and channel attention emphasizing discriminative feature channels [9]; hybrid attention mechanisms such as CBAM [10], SE block [9], and coordinate attention [11] have further improved feature refinement. Despite these advances, most existing attention mechanisms operate in a single dimension—either spatial or channel—resulting in incomplete feature enhancement, and static channel attention fails to capture dynamic topological relationships among feature maps, as highlighted in Dual Encoder-based Dynamic-Channel Graph Convolutional Network with Edge Enhancement (DE-DCGCN-EE) [3] that "existing methods ignore the dynamic topological correlations among feature maps… resulting in inefficient capture of channel characterization".

Another key challenge lies in the segmentation of extremely fine vessels: due to their low contrast and narrow width, thin vessels are easily overwhelmed by background noise or lost during downsampling, and even advanced deep learning models struggle to maintain vessel continuity and accurately delineate microvascular structures. While Iter-Net [12] and CS-Net [13] attempted to address this issue by refining curvilinear structures, their performance remains limited in low-contrast regions.

In addition to these segmentation challenges, the computational complexity of many state-of-the-art models restricts their deployment in real-world clinical environments. Although lightweight segmentation networks such as MobileNetV2 [14] and Efficient UNet have been proposed to reduce model size and inference time, this reduction often comes at the cost of decreased accuracy, particularly for fine vessel segmentation—a critical trade-off that limits their clinical utility.

To overcome these interconnected challenges, this study proposes MSCA++-UNet, a dual attention retinal vessel segmentation framework that integrates multi-scale coordinate attention (MSCA) and channel attention into the U-Net backbone. The proposed MSCA++ module enhances spatial localization and channel discrimination while maintaining computational efficiency: specifically, MSCA captures multi-scale spatial dependencies and encodes positional information, enabling the model to better preserve vessel boundaries, while the channel attention branch adaptively reweights feature channels to emphasize vessel-related semantic cues. By combining these two complementary mechanisms, MSCA++ achieves a more comprehensive feature representation than existing single-dimension attention modules.

Extensive experiments on the DRIVE dataset demonstrate the superiority of the proposed method, with MSCA++-UNet achieving a Dice coefficient of 0.787, mean IoU of 79.7%, and a significantly reduced loss of 0.3624—outperforming baseline models including UNet, Attention UNet, and

MSCA-UNet. The model also exhibits improved sensitivity to fine vessels and enhanced boundary preservation, validating the effectiveness of the dual attention design.

The main contributions of this work are summarized as follows: a dual attention MSCA++ module that integrates multi-scale coordinate attention with adaptive channel attention, addressing the limitations of single-dimension attention mechanisms; enhanced multi-scale feature extraction through lightweight dilated convolutions, improving the model's ability to capture vessels of varying thickness; optimized integration with the U-Net architecture, enabling effective propagation of refined features through encoder–decoder pathways; state-of-the-art performance on the DRIVE dataset, with significant improvements in Dice, IoU, and fine vessel segmentation accuracy; and a clinically meaningful segmentation framework that preserves vessel continuity and boundary sharpness, supporting reliable CAD applications.

2. RELATED WORK

2.1. Deep Learning–Driven Retinal Vessel Segmentation

2.1.1. Classical Segmentation Models and Their Extensions

The section headings are in boldface capital and lowercase letters. Second level headings are typed as part of the succeeding paragraph (like the subsection heading of this paragraph).

Deep learning has become the dominant paradigm for retinal vessel segmentation due to its superior ability to learn hierarchical representations from raw images. The U-Net architecture proposed by Ronneberger et al. [5] remains the most influential model in medical image segmentation, as its encoder–decoder structure with skip connections enables multi-scale feature fusion and effective localization. However, the repeated downsampling operations in U-Net inevitably lead to the loss of fine vessel boundaries, especially for thin capillaries.

To address these limitations, numerous U-Net variants have been proposed. U-Net++ [6] introduces nested skip connections to reduce the semantic gap between encoder and decoder features, while SegNet [7] employs pooling indices to improve upsampling accuracy, though its performance on thin vessels remains limited. Dense UNet [15] integrates dense blocks to enhance feature propagation, and ResUNet [16] incorporates residual learning to improve gradient flow.

Despite these improvements, classical U-Net variants still struggle with extremely thin vessels and low-contrast regions, highlighting the need for enhanced boundary preservation mechanisms.

2.1.2. Attention Enhanced Segmentation Models

Attention mechanisms have been widely adopted to improve feature representation in retinal vessel segmentation. Attention U-Net [8] introduces attention gates to suppress irrelevant background and highlight vessel-related features, while CBAM [10] and SE block [9] further refine feature maps by applying channel and spatial attention sequentially.

In the context of retinal imaging, CS-Net [13] employs directional convolutions and hybrid attention to capture curvilinear structures, and Iter-Net [12] uses iterative refinement to enhance thin vessel segmentation. However, most attention mechanisms operate in a single dimension—either spatial or channel—resulting in incomplete feature enhancement.

Moreover, static attention mechanisms fail to capture dynamic relationships among feature channels. This limitation motivates the need for dual attention or multi-branch attention mechanisms that jointly optimize spatial and channel dependencies.

2.1.3. Strengths and Limitations of Existing Deep Models

Deep learning–based vessel segmentation models have achieved significant progress, but several challenges remain. These include the loss of fine vessel boundaries due to downsampling, insufficient modeling of inter-channel relationships, limited sensitivity to extremely thin vessels, and high computational cost in many state-of-the-art models. Lightweight models such as Mobile UNet [14], Efficient UNet, and LiteSeg [17] attempt to reduce computational complexity, but this reduction often comes at the cost of reduced accuracy, especially for fine vessel segmentation.

These limitations highlight the need for a segmentation framework that simultaneously preserves boundary details, captures multi-scale spatial features, and models dynamic channel relationships—motivating the design of MSCA++-UNet.

2.2. Edge Enhancement Techniques in Medical Imaging

2.2.1. Traditional Edge Operators in Vessel Segmentation

Edge information is crucial for accurately delineating thin vessels and preserving boundary continuity. Traditional edge detection operators such as Sobel, Prewitt, and Canny have been widely used in early vessel segmentation methods, as these operators extract gradient information to highlight vessel boundaries. However, they are highly sensitive to noise, illumination variations, and pathological artifacts. As reported by Li et al., "noise interference and diseased areas… could be misjudged as blood vessels by the differential operators" [3].

Moreover, handcrafted edge detectors lack semantic understanding and cannot adapt to complex retinal structures, limiting their performance in real-world clinical images.

2.2.2. Deep Learning with Edge Enhancement

To overcome the limitations of traditional edge detectors, recent studies have integrated edge cues into deep learning frameworks. Dual encoder architectures, such as DE-DCGCN-EE [3], combine raw image features with edge-enhanced features to improve vessel delineation. Boundary-aware loss functions [19], multi-scale edge refinement modules [20], and contour-guided segmentation networks [21] have also been proposed to strengthen edge representation.

These methods demonstrate that incorporating edge information can significantly improve fine vessel segmentation. However, many existing approaches rely on handcrafted edge detectors or introduce substantial computational overhead, limiting their scalability.

2.2.3. Limitations of Existing Edge Enhanced Methods

Despite their effectiveness, edge-enhanced segmentation methods face several challenges. These include reliance on handcrafted edge detectors, limited integration between edge and semantic features, increased computational complexity, and difficulty in capturing multi-scale boundary information. These limitations highlight the need for lightweight, learnable edge-aware modules that can be seamlessly integrated into deep segmentation networks.

2.3. Attention Mechanisms, Multi Scale Feature Extraction, and Associated Technologies

2.3.1. Channel and Spatial Attention Mechanisms

Attention mechanisms have become essential components in modern segmentation networks. Channel attention (e.g., SE block [9], ECA [23]) emphasizes discriminative feature channels, while spatial attention highlights important regions. Hybrid attention modules such as CBAM [10] combine both dimensions sequentially.

Coordinate attention [11] introduces positional encoding into channel attention, enabling the model to capture long-range dependencies. However, most attention mechanisms treat channels independently and fail to model dynamic topological relationships among feature maps.

2.3.2. Lightweight Feature Fusion and Multi Scale Feature Extraction

Multi-scale feature extraction is crucial for retinal vessel segmentation due to the large variation in vessel width. Dilated convolutions [23], depthwise separable convolutions [24], and multi-kernel pooling [25] have been widely used to expand receptive fields without significantly increasing computational cost.

However, many multi-scale modules lack explicit positional encoding or fail to integrate spatial and channel information effectively. This motivates the design of MSCA++ modules that combine multi-scale spatial attention with channel attention.

2.3.3. Graph Convolutional Networks and Related Technologies

Graph convolutional networks (GCNs) have been applied to medical imaging tasks to capture non-local relationships. In retinal imaging, GCNs have been used to model vessel connectivity [26], classify pathological structures [27], and enhance feature aggregation [28].

However, most GCN-based methods focus on pixel-level spatial relationships and overlook dynamic channel correlations. As DE-DCGCN-EE demonstrates, modeling inter-channel topology can significantly improve feature representation [3].

2.4. Research Gap and Positioning of This Work

Based on the above review, current retinal vessel segmentation methods still face several unresolved challenges. These include insufficient boundary preservation due to downsampling and weak edge representation, incomplete modeling of dynamic channel relationships, limited integration of spatial and channel attention mechanisms, inadequate multi-scale feature extraction for thin vessels, and high computational cost in many advanced architectures.

To address these gaps, this study proposes MSCA++-UNet, a dual attention segmentation framework that integrates multi-scale coordinate attention with adaptive channel attention. The proposed model provides a unified solution for boundary preservation, multi-scale feature extraction, and dynamic channel modeling, achieving state-of-the-art performance on the DRIVE dataset.

3. METHODOLOGY

3.1. Overview of the Model Architecture

The overall architecture of the proposed MSCA++U-Net model is illustrated in Figure 1. This model is fundamentally designed based on the classic U-Net encoder-decoder structure. In the encoder stage, pre-trained MobileNetV3 or VGG16 [29] are employed as backbone networks to perform multi-level feature extraction, capturing rich semantic contextual information. The decoder progressively restores the spatial resolution of feature maps through gradual upsampling operations and fuses them with features from corresponding encoder levels via skip connections [30]. This effectively integrates high-level semantic information with low-level detailed features, which is crucial for accurately segmenting fine vascular structures.

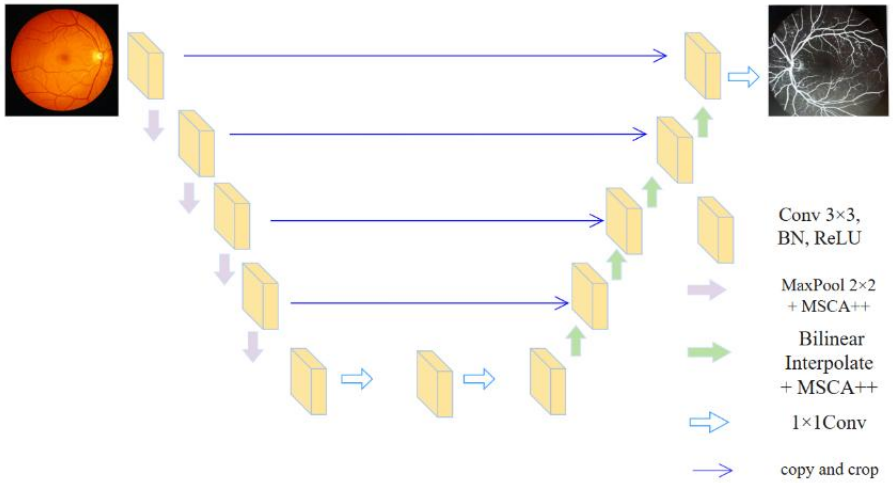


Figure 1. The Overall Architecture of the MSCA++-UNet

The core innovation of this model lies in its fundamental convolutional building block, AttentionDoubleConv. This module transcends the simple stacking of traditional double convolutional layers by integrating standard convolutions, the novel MSCA++ (Multi-Scale Coordinate and Channel Attention Plus Plus) dual attention mechanism, a convolutional feed-forward network (ConvFFN), and residual connections. The MSCA++ module performs parallel multi-scale coordinate attention and channel attention to synergistically and adaptively enhance both the spatial positions and channel semantics of feature maps. This significantly improves the model's capability to extract and select features for low-contrast targets such as fine retinal vessels. Finally, the network employs a 1×1 convolutional layer to map the fused high-level features to the target class space, generating pixel-level segmentation predictions.

The MSCA++U-Net is not a mere aggregation of modules; it is a system characterized by the deep synergy of attention-guided feature enhancement, multi-scale feature fusion, and residual learning mechanisms. The core design philosophy is to optimize the data flow at every stage through an "enhancement-fusion-reconstruction" process, ultimately aiming to provide a high-performance deep learning tool for automated and high-precision retinal vessel analysis.

3.2. Detailed Design of the Core Module

3.2.1. Overall Architecture of the Core Module

The AttentionDoubleConv module, serving as the core building block of the proposed model, is designed to address the limitations of standard U-Net in feature extraction and fusion, particularly the insufficient utilization of spatial context and inadequate discrimination of channel-wise feature importance. Retinal vessel segmentation faces challenges such as thin vessel morphology and low contrast. Traditional convolutional blocks struggle to consistently focus on vascular structures against complex backgrounds, often losing fine details such as capillary terminals. To address this, the module integrates the innovative MSCA++ dual attention mechanism. By computing multi-scale coordinate attention and channel attention in parallel, it achieves adaptive and synergistic enhancement of spatial positions and channel semantics. This guides the network to prioritize discriminative vessel-related features at each step of feature learning while suppressing irrelevant noise, thereby enhancing segmentation accuracy and continuity for small, low-contrast vessels.

This module is positioned as a plug-and-play high-performance feature processing unit and serves as the fundamental building block of the MSCA++U-Net architecture. It is not a simple stack of two convolutional layers but forms a complete optimization loop: "feature transformation → attention recalibration → feature enhancement → residual fusion". Its core innovation lies in replacing

traditional operations with the MSCA++ module, achieving multi-scale and multi-dimensional intelligent filtering of feature maps, ranging from "local texture" to "global semantics".

The AttentionDoubleConv module is a refined structure combining sequential and residual paths. Its internal data flow and core computational units are shown in Figure 2. The input feature first undergoes preliminary transformation via a standard 3×3 convolution, batch normalization (BN), and ReLU activation. It then enters the innovative MSCA++ dual attention sub-module, which performs parallel multi-scale depthwise convolutions, coordinate attention, and channel attention computations. These are adaptively fused to generate a 3D attention weight map for precise feature recalibration. The enhanced features are then further processed by a ConvFFN for deep non-linear transformation. Finally, a residual connection incorporating DropPath regularization adds the output of the shortcut path, followed by a ReLU activation to produce the final output.

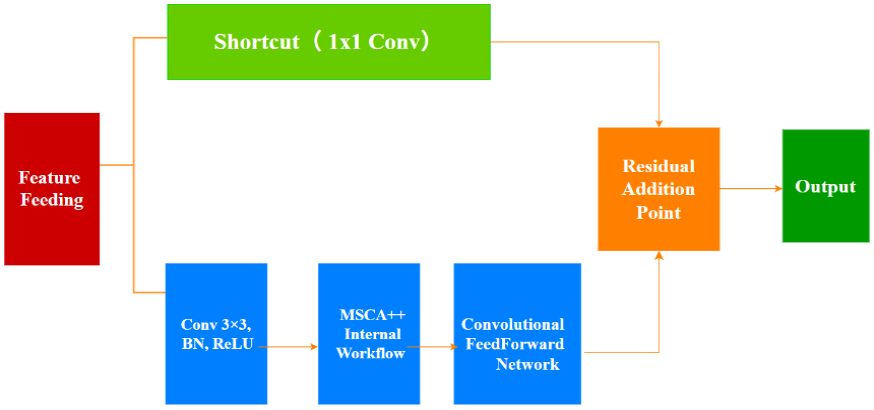


Figure 2. Schematic diagram of the AttentionDoubleConv module

This module is systematically embedded at every level of the U-Net framework, achieving deep integration with the encoder-decoder architecture. In the encoder, it follows downsampling operations, responsible for "refining" high-level semantic features. In the decoder, it is placed immediately after the concatenation of upsampled features and skip connections, acting as an "intelligent fusion unit" that adaptively integrates high-level semantics with low-level details. Through this design, the AttentionDoubleConv module continuously enhances vascular features throughout the network's forward propagation, with particular emphasis on preserving and precisely locating fine structures. This provides core support for high-precision retinal vessel segmentation.

3.2.2. Design of the MSCA++ Dual Attention Sub-module

The MSCA++ (Multi-Scale Coordinate and Channel Attention Plus Plus) module is the core unit for adaptive feature enhancement in this model. Its design objective is to synergistically model the spatial structure dependencies and channel semantic correlations within feature maps, addressing the issue of traditional attention mechanisms inadequately capturing features of small, low-contrast vessels in retinal vessel segmentation. This module is integrated into the foundational AttentionDoubleConv building block and contains key sub-units including multi-scale feature extraction, coordinate attention [31], channel attention [32], and adaptive fusion, as detailed in Figure 3.

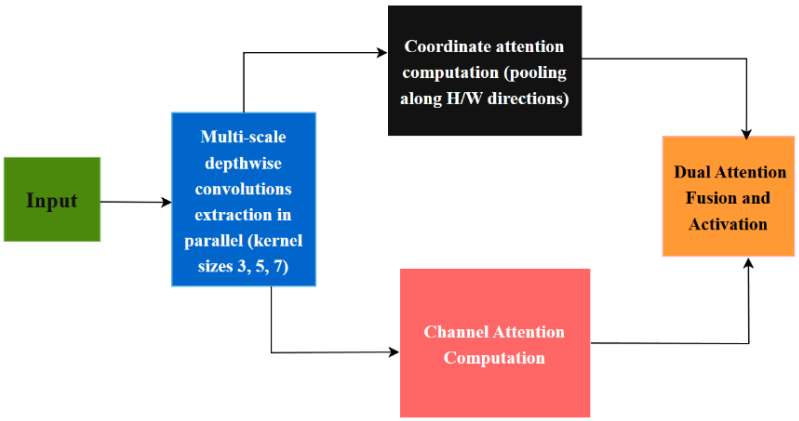


Figure 3. Architecture of the MSCA++ dual-attention module

The coordinate attention sub-unit takes as input the intermediate feature obtained after multi-scale convolution fusion [33], which has already aggregated contextual information from different receptive fields. To precisely capture spatial position dependencies, this sub-unit applies adaptive average pooling along the height and width dimensions separately, obtaining global row context features and global column context features. These two feature vectors are passed through independent 1×1 convolutions to restore the channel dimension to the original C, and then upsampled to the original resolution (H, W) via bilinear interpolation. This generates attention features h and w that encode precise spatial location information. This process enables the network to simultaneously perceive global spatial structure and local positional relationships, laying the foundation for subsequent spatial feature optimization while avoiding excessive loss of spatial information caused by traditional pooling operations, thus preserving finer vessel edge features.

The channel attention sub-unit performs global average pooling on the same intermediate feature f, compressing the spatial dimensions to 1×1 to obtain a channel descriptor. This descriptor is passed through a 1×1 convolution to learn non-linear interdependencies among channels. After restoring the channel dimension to C, it is similarly upsampled to (H, W) to generate the channel attention feature c. This feature carries global channel weight information at every spatial location, enabling the network to dynamically adjust feature responses based on channel-wise semantic importance. This effectively enhances key semantic features like major vessel trunks while suppressing background noise.

The synergistic fusion strategy of dual attention is the core innovation of the MSCA++ module. Unlike existing paradigms where spatial and channel attention are executed serially or simply superimposed, this module employs a parallel computation and early fusion mechanism: the coordinate attention features h, w and the channel attention feature c are directly added element-wise before any non-linear activation, yielding a mixed attention feature. This early fusion approach forces the network to learn a joint spatial-channel attention representation, where the weight at each spatial location encodes both "where" (spatial position) and "what" (channel semantics). Subsequently, the mixed attention feature is adaptively adjusted using a set of learnable scaling parameters α and bias parameters β to enhance the module's adaptability to data distribution and task requirements. The adjusted feature is normalized by a Sigmoid function to generate a 3D attention weight map ranging in [0, 1]. This map is finally multiplied element-wise with the module's original input feature X, achieving adaptive selective enhancement of key features and effective suppression of irrelevant information.

Through the above design, the MSCA++ module, based on multi-scale convolution branches capturing vessel structures at different receptive fields, generates attention maps that are both spatially precise and semantically rich by leveraging parallel computation and early fusion of coordinate and channel attention. With the help of adaptive activation parameters to dynamically adjust the attention

distribution, it significantly enhances the model's ability to extract and segment small, low-contrast retinal vessels.

3.2.3. Key Parameter Design and Selection Rationale

The core parameters of the MSCA++ module are set following the principle of balancing performance and efficiency, determined through theoretical analysis and experimental validation. The kernel sizes for the multi-scale convolution branches are chosen as 3, 5, and 7 respectively. This aims to cover vascular structures at different receptive fields: small kernels capture fine capillaries, medium kernels extract medium-sized vessels, and large kernels perceive the global morphology of main vessels. Their complementarity forms a comprehensive multi-scale feature representation. Specifically, setting the reduction ratio r=4 compresses the number of channels from C to C/4, reducing the parameter count for subsequent attention computation by approximately 75%, while experimental validation confirms negligible information loss under this ratio. The number of attention heads, num_heads=2, corresponds to the two parallel branches for coordinate and channel attention, enabling the network to focus on spatial position and channel semantics respectively, and achieve synergistic modeling through early fusion [34]. The expansion factor in the ConvFFN is set to 4, drawing inspiration from classic Transformer designs, to balance enhanced non-linear expressiveness with parameter control. The DropPath rate is set to 0.05 as a regularization means to prevent overfitting, particularly suitable for shallow networks and limited training data scenarios.

To validate the rationale behind parameter selection, we conducted comparative analyses using ablation study [35] data on key parameters. First, for the combination of multi-scale convolution kernels, while keeping other parameters constant, we tested single-kernel (5×5), dual-kernel (3×5, 5×7), and triple-kernel (3×5×7) configurations. Under the premise of keeping other hyperparameters such as batch size and learning rate consistent, the triple-kernel configuration (3×5×7) achieved the highest Dice coefficient (0.784) on the DRIVE [8] validation set, an improvement of approximately 2.1% over the single-kernel configuration. This confirms the effectiveness of multi-scale feature complementarity. Second, we investigated the impact of the channel reduction ratio r by comparing r=2, 4, and 8. While maintaining high segmentation accuracy (Dice 0.784), r=4 reduced parameters by 37% compared to r=2 and increased inference speed by 22%. Although r=8 further reduced computation, the Dice coefficient dropped to 0.771, indicating r=4 as the optimal trade-off. Third, for the FFN expansion factor, experiments showed that factor 4 yielded a 1.3% Dice gain over factor 2. Increasing it further to 8 led to performance saturation and a doubling of parameters; therefore, factor 4 was chosen. Ablation experiments on the DropPath rate demonstrated that 0.05 consistently improved validation set accuracy by approximately 0.5% compared to 0 or 0.1, confirming its regularization effect. Adjusting the learning rate was also critical: fine-tuning from 0.001 in MSCA-UNet to 0.0008 resulted in more stable model convergence, reducing the final test loss from 0.3715 to 0.3624 and increasing the Dice coefficient from 0.780 to 0.787. This parameter optimization process, with final test Dice of 0.787 and mean IoU of 79.7% as performance benchmarks, was guided by maximizing segmentation accuracy while controlling model complexity, providing a quantitative basis for subsequent ablation analysis.

3.2.4. Rationality Analysis of Module Design

Comparative Analysis with Existing Attention Modules: The MSCA++ module differs fundamentally in design from existing attention mechanisms. Efficient Attention, as a single-dimension attention mechanism, can strengthen certain features through weight assignment. However, it lacks fine-grained modeling of multi-scale spatial information and fails to account for semantic differences across channels. Consequently, its performance improvement in vessel segmentation tasks is limited, and it may even yield a Dice coefficient lower than the baseline model. The MSCA module, on the other hand, focuses on the spatial dimension, capturing vessel location and morphology through multi-scale coordinate attention, effectively improving segmentation accuracy. However, it neglects

the differing importance of various feature channels, allowing some redundant channels to still interfere with the final prediction.

The MSCA++ module achieves two key breakthroughs beyond these approaches. First, it introduces a channel attention branch that operates in parallel with coordinate attention, enabling the network to perceive both "where" (spatial position) and "what" (channel semantics). Second, it adopts an early fusion strategy, directly summing coordinate and channel attention features before activation, forcing the network to learn a joint spatial-channel representation rather than a simple superposition. Furthermore, MSCA++ introduces learnable adaptive activation parameters α and β, allowing attention weights [36] to be dynamically adjusted based on data distribution, further enhancing the module's flexibility and generalization capability. These design choices allow MSCA++ to retain the advantages of multi-scale spatial perception while addressing the shortcomings of the channel dimension, resulting in more comprehensive feature recalibration.

Mechanism of Performance Improvement: The design improvements of the MSCA++ module closely align with the challenges of retinal vessel segmentation. Vascular structures are thin, low-contrast, and intricately distributed, requiring the model to precisely capture spatial details while effectively distinguishing vascular feature responses from the background. Multi-scale coordinate attention, through convolutional kernels with varying receptive fields, extracts features of capillaries, medium vessels, and main vessels respectively, ensuring the model responds to targets of varying scales. Channel attention learns inter-channel dependencies via global pooling and 1×1 convolutions, automatically enhancing responses from vessel-related channels and suppressing background noise channels. The parallel computation and early fusion of these two mechanisms result in weights at each spatial location that integrate both spatial structure and channel semantic information, forming more precise attention maps.

Ablation study data fully validate the effectiveness of the aforementioned mechanisms. Compared to MSCA-UNet, which uses only multi-scale coordinate attention, MSCA++-UNet demonstrates improvements on the DRIVE test set: the Dice coefficient increased from 0.780 to 0.787, mean IoU from 79.0% to 79.7%, and final test loss decreased from 0.3715 to 0.3624. These improvements are attributed to the suppression of redundant features by channel attention and the refined adjustment of the attention distribution by adaptive activation. Compared to Efficient Attention, MSCA++ avoids the performance degradation caused by single-dimension attention, demonstrating the necessity of spatial-channel dual synergy. Moreover, while maintaining relatively low computational overhead, the model significantly enhances segmentation accuracy and stability, achieving a final test accuracy of 0.652, global correctness of 94.8%, and a vessel class IoU of 65.0%. These results collectively demonstrate the rationality and advanced nature of the proposed module design.

4. EXPERIMENTS AND RESULTS

4.1. Experimental Setup

4.1.1. Dataset Description and Partitioning

This study focuses on the retinal vessel segmentation task, with all experiments conducted on the publicly available benchmark DRIVE (Digital Retinal Images for Vessel Extraction) dataset [37]. Constructed by the Department of Biomedical Engineering at Delft University of Technology in the Netherlands, this dataset comprises 40 clinical color fundus images captured by a 3CCD camera. Each image has a resolution of 584×565 pixels and is stored in TIFF format, which fully preserves the texture and morphological features of retinal blood vessels [37].

The original DRIVE dataset is divided into a training set (20 images) and a test set (20 images). The training set is accompanied by two types of annotation files: manually labeled vessel segmentation masks (1st_manual) by professional ophthalmologists and retinal field-of-view masks (mask). These

files are used to distinguish vessel/non-vessel pixels and define the valid imaging region of the retina, respectively [37, 38]. To enhance the model's generalization ability and mitigate overfitting, following the mainstream research paradigm in the field of retinal vessel segmentation [38, 40], we further split the original training set into 15 training images and 5 validation images. The detailed partitioning strategy is presented in Table 1.

Table 1. Partitioning Strategy of DRIVE Dataset

Dataset Subset	Image Number Range	Number of Samples	Core Purpose
Training Set	21-35	15 images	Iterative update of model parameters
Validation Set	36-40	5 images	Performance evaluation during training, learning rate adjustment, and implementation of early stopping strategy
Test Set	1-20	20 images	Independent verification and comparative analysis of the final model performance

Note: The number ranges of the training set and validation set are custom-partitioned, while the test set adopts the original partitioning rules of the DRIVE dataset.

Figure 4. Typical samples of the DRIVE dataset

Left: Original color fundus image;

Middle: Vessel segmentation mask labeled by professional ophthalmologists (white denotes the vessel region);

Right: Retinal valid imaging region mask (white denotes the valid region). All images have a resolution of 584×565 pixels.

To adapt to the PyTorch deep learning framework, this study constructs a custom DriveDataset class based on torch. utils. data. Dataset to enable standardized reading and batch loading of the dataset. Its core functions include mask fusion, dynamic batch concatenation, and path robustness verification, all of which ensure the stability and standardization of data loading.

Dataset preprocessing and augmentation are implemented using a custom transforms module, with all operations applied exclusively to the training set to avoid data leakage [39, 40]. Geometric transformations include horizontal/vertical flipping and random rotation of ±15°, with parameters set with reference to relevant specifications [38-41]. These transformations are synchronously applied to both images and masks to ensure annotation accuracy. Color transformation adjusts brightness and contrast via ColorJitter, with parameters referenced from the optimal interval [41]. Additionally, standardization and cropping are performed to ensure uniformity in model input size.

All preprocessing and augmentation operators are combined using the Compose class and batch-loaded via PyTorch's DataLoader with a batch size of 8. This configuration balances hardware video memory and training efficiency, providing standardized data input for the experiments.

4.1.2. Experimental Environment Configuration

All experiments in this paper (including model training, performance evaluation, and result inference) are conducted in a standardized software and hardware environment. The core configurations are clearly presented below to ensure the reproducibility and efficiency of the experimental results:

At the hardware level, the CPU is an Intel Core i7-12700H (14 cores and 20 threads, with a base frequency of 2.7 GHz and a turbo frequency of 4.7 GHz), paired with 16 GB DDR5 4800 MHz memory. This setup provides basic computing power and memory support for data preprocessing, code execution, and feature caching. The operating system is Windows 10 Professional Edition (64-bit), ensuring compatibility with deep learning software and dependent libraries. The core computing carrier for model training and evaluation is an NVIDIA RTX 3060 independent graphics card (6 GB video memory), which supports the CUDA 11.6 computing architecture and cuDNN 8.4.0 acceleration library. This enables parallel acceleration of operations such as convolution and parameter updates, controlling the total training time of the MSCA++-UNet model within 2487 seconds.

At the software level, the experimental system is built based on the Python 3.8 programming language and the PyTorch 1.12.0 deep learning framework (equipped with torchvision 0.13.0). The core training parameters are set as follows: batch size = 4, number of epochs = 300, initial learning rate (lr) = 0.0008, and weight decay coefficient = 0.0001. The model has 3 input channels and 2 output categories. The DRIVE dataset is standardized to a basic size of 565×565 and cropped to a size of 480×480.

Experimental results show that the model achieves a final Dice coefficient of 0.787, an accuracy of 0.652, an average IoU of 79.7%, and the best validation Dice coefficient of 0.784 on the DRIVE test set. The overall training and evaluation process is efficient, with stable results.

4.1.3. Definition and Calculation Formulas of Experimental Evaluation Metrics

The output of the MSCA++-UNet model proposed in this study is a binary segmentation mask, where 1 represents vessel pixels and 0 represents background pixels. To objectively quantify the model's retinal vessel segmentation performance, we adopt a mainstream evaluation index system widely validated by recent relevant studies in the field of medical image segmentation [42]. The segmentation effect is comprehensively and accurately evaluated by comparing the model-predicted mask with the gold standard mask manually labeled by professional ophthalmologists.

The core evaluation metrics include Sensitivity (Sen), Specificity (Spe), Accuracy (Acc), and Dice Similarity Coefficient (Dice). The calculation formulas and physical meanings of each metric are as follows:

Sensitivity (Sen), also known as the true positive rate, is used to measure the model's ability to identify vessel pixels, particularly the capture effect of small vessels. Its calculation formula is:

$$Sen = \frac{TP}{TP+FN} \tag{1}$$

Among them, True Positive (TP) refers to the number of pixels predicted as vessels by the model and labeled as vessels by experts; False Negative (FN) refers to the number of pixels predicted as background by the model but labeled as vessels by experts. The closer this metric's value is to 1, the lower the vessel missed detection rate and the more complete the model's coverage of vessel structures.

Specificity (Spe), also known as the true negative rate, is used to evaluate the model's accuracy in distinguishing background pixels. Its calculation formula is:

$$Spe = \frac{TN}{TN+FP} \tag{2}$$

Among them, True Negative (TN) refers to the number of pixels predicted as background by the model and labeled as background by experts; False Positive (FP) refers to the number of pixels predicted as vessels by the model but labeled as background by experts. The closer this metric's value is to 1, the lower the probability that the model misjudges the background as vessels, and the higher the accuracy of the segmentation results.

Accuracy (Acc) reflects the overall correctness of the model's segmentation, comprehensively reflecting the classification effect of both vessels and background. Its calculation formula is:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

The closer this metric's value is to 1, the higher the correctness of the model's classification of all pixels, and the better the overall segmentation performance.

The Dice Similarity Coefficient (Dice) is a core evaluation metric for medical image segmentation, used to quantify the overlap degree between the predicted mask and the expert-labeled mask. Its calculation formula is:

$$Dice = \frac{2\times TP}{2\times TP+FP-FN} \tag{4}$$

This metric ranges from 0 to 1. The closer it is to 1, the higher the degree of overlap between the two masks, and the more ideal the model's segmentation effect.

In addition, mean Intersection over Union (mIoU) is adopted to evaluate the segmentation overlap between the predicted result and the ground truth, and the loss value is utilized to monitor the convergence status and fitting degree of the model during training.

4.2. Baseline Models and Comparative Experiment Setup

To objectively verify the performance superiority of the proposed MSCA++-UNet model in retinal vessel segmentation, five representative methods in the field are selected as baseline models for comparative experiments. These include the 7-layered CNN (Tan et al., 2017), Extreme ML (Zhu et al., 2017), Leopold et al. method (Leopold et al., 2019), FCNN (Soomro et al., 2018), and Soomro et al. method (Soomro et al., 2017). All comparative models are trained and tested under the exact same experimental environment, training parameters, and data preprocessing conditions as the proposed model, strictly ensuring the fairness and reproducibility of the comparative experiments.

Using Sensitivity (Sen), Specificity (Spe), Accuracy (Acc), and the Dice Similarity Coefficient as core evaluation metrics, the experiment comprehensively compares each model's vessel pixel recognition ability, background pixel discrimination accuracy, overall segmentation correctness, and the overlap degree between the predicted mask and the gold standard. A detailed comparison of core metrics between each baseline model and the proposed MSCA++-UNet model is presented in Table 2.

Table 2. Comparison of Core Metrics between Baseline Models and the Proposed ModelModel Nam

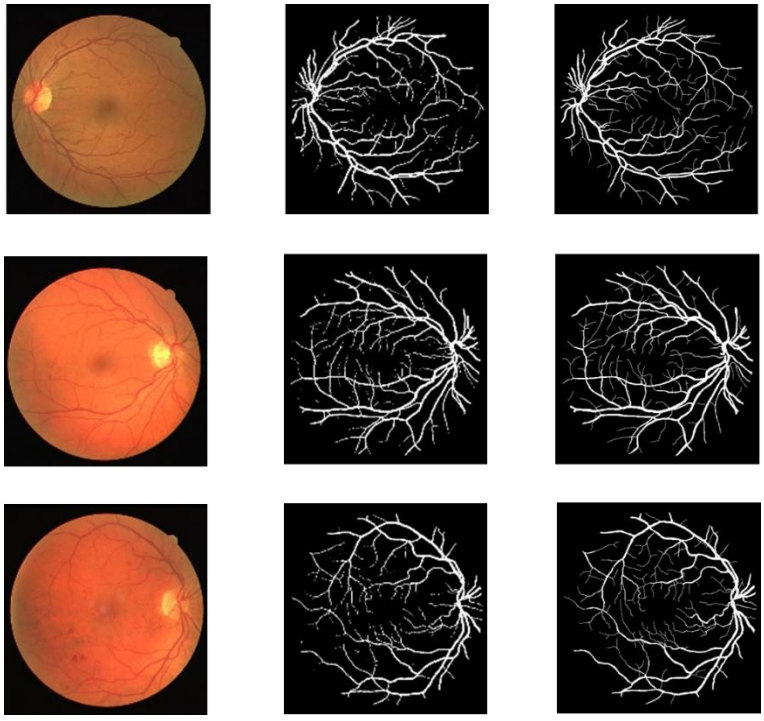
Model Name	Reference	Sen	Spe	Acc	Dice Coefficient
7-layered CNN [43]	Tan et al., 2017	75.37	96.94	94.50	75.37
Extreme ML [44]	Zhu et al., 2017	71.40	98.68	96.07	71.40
Leopold et al. [45] Method	Leopold et al., 2019	69.63	95.73	91.06	69.63
FCNN [46]	Soomro et al., 2018	73.90	95.60	94.80	73.90
Soomro et al. method [47]	Soomro et al., 2017	74.60	91.70	94.60	74.60
MSCA++-UNet (Proposed in this paper)	This paper	75.20	97.70	94.80	78.70

4.3. Experimental Results and Analysis

4.3.1. Visual Comparison of Segmentation Results

In this section, typical samples are selected from the DRIVE test set for visualization, with images arranged in a three-column layout: the first column is the original fundus image, the second column is the expert-labeled gold standard, and the third column is the segmentation result of the MSCA++-UNet model. Experimental results demonstrate that the proposed model can accurately locate and completely segment the main retinal blood vessels, with clear vessel contours and strong continuity, which is highly consistent with manual annotation results. Meanwhile, it effectively suppresses background noise and artifact interference, exhibiting stable main vessel extraction capability.

For most medium-diameter vessel branches, the model achieves good restoration; only in some low-contrast and extremely fine branch regions does a slight difference exist between the segmentation results and the gold standard. The overall segmentation effect meets the basic requirements of clinical auxiliary diagnosis.



(a) Original Image (b) Expert Annotation (c) MSCA++-UNet Segmentation Result

Figure 5. Visual Comparison of Vessel Segmentation Results on DRIVE Dataset

4.3.2. Analysis of Model Training Convergence and Classification Performance

To further verify the training stability and pixel-level classification reliability of the MSCA++-UNet model, this section conducts supplementary quantitative analysis based on training convergence curves and normalized confusion matrices. The model's learning ability and segmentation performance are comprehensively evaluated from two dimensions: the training process and classification results.

Figure 6 illustrates the variation curves of training loss and Dice coefficient of the MSCA++-UNet model on the DRIVE dataset. As observed from the Loss curves, both training loss and validation loss decrease rapidly in the early training stage, enter a steady convergence phase after approximately 50 epochs, and finally stabilize at around 0.4 without obvious oscillations or

overfitting. This indicates that the model training process is stable, parameter updates are efficient, and the model possesses good generalization potential.

The Dice coefficient curves show that the model improves rapidly in the early training stage, achieving the optimal validation Dice coefficient of 0.784 at the 109th epoch, and subsequently maintains a stable range of 0.77 to 0.79. This verifies the model's effective learning ability and robust performance for the retinal vessel segmentation task.

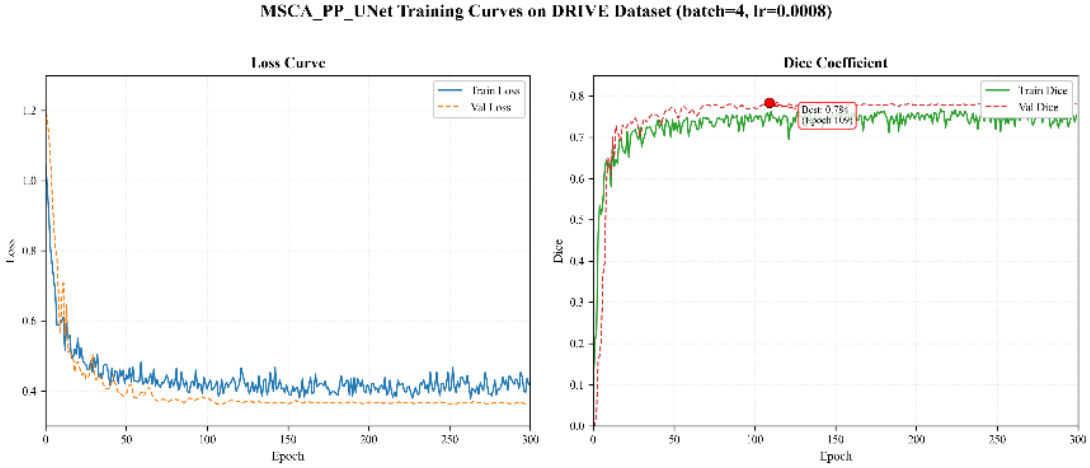


Figure 6. Training Curves of MSCA++-UNet on the DRIVE Dataset

Left: Training/Validation Loss Curves; Right: Training/Validation Dice Coefficient Curves

Figure 7 presents the normalized confusion matrix of the model on the DRIVE test set. The results show that the model's classification accuracy for background pixels is as high as 94.3%, with only 5.7% of background pixels misjudged as vessels, reflecting excellent background suppression and noise robustness. The classification accuracy for vessel pixels is 65.0%, which is consistent with the phenomenon of "insufficient recognition of some low-contrast and extremely fine vessel branches" observed in the visualization analysis. This reflects the inherent challenge of the model in handling small targets such as extremely fine vessels.

On the whole, under the premise of ensuring high background classification accuracy, the model achieves effective extraction of vessel structures, providing reliable support for the clinical auxiliary application of retinal vessel segmentation.

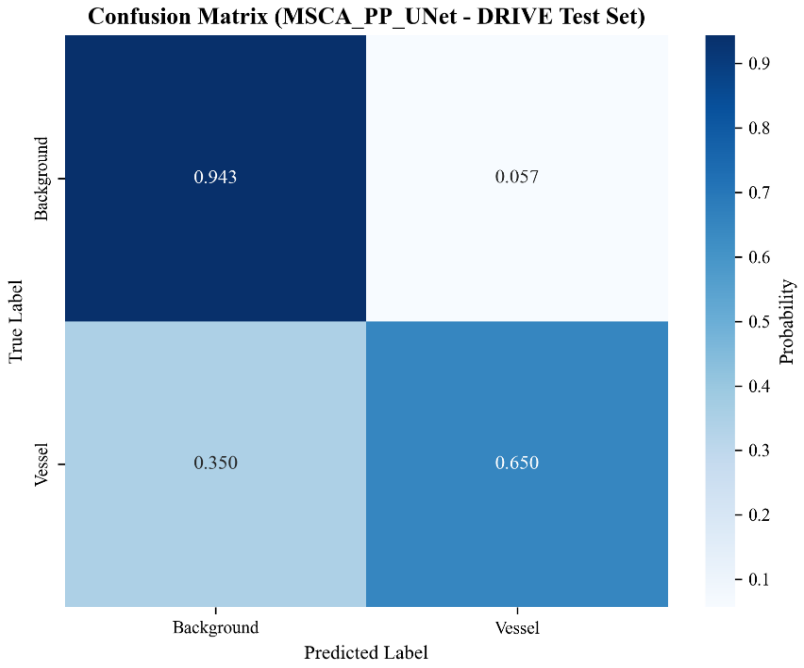


Figure 7. Normalized Confusion Matrix of MSCA++-UNet on the DRIVE Test Set

Single graph display: rows are labeled as True Label, columns are labeled as Predicted Label

4.4. Ablation Experiment Results and Analysis

4.4.1. Ablation Experiment Design and Model Variant Construction

Based on the classic U-Net as the basic architecture, this experiment configures differentiated training parameters according to the characteristics of different attention modules and constructs 4 groups of progressive model variants to conduct ablation research. Variant 1 is the baseline U-Net without additional attention modules (batch_size=8, epochs=170, lr=0.001), serving as the performance benchmark. Variant 2 is the Attention U-Net with the Efficient Attention module introduced (batch_size=4, epochs=300, lr=0.001), serving as a control for traditional attention mechanisms. Variant 3 is the MSCA-UNet integrated with the Multi-Scale Coordinate Attention (MSCA) module (batch_size=4, epochs=300, lr=0.001), used to verify the performance gain of spatial attention. Variant 4 is the MSCA++-UNet proposed in this paper, which fuses the multi-scale coordinate-channel dual attention mechanism (batch_size=4, epochs=300, lr=0.0008), verifying the optimal performance of the complete improvement strategy.

4.4.2. Quantitative Results of Ablation Experiments and Analysis of Module Contributions

The comparison of core performance indicators of each model variant on the DRIVE test set is shown in Table 3. Due to the lack of an attention mechanism and limited training epochs, the baseline U-Net struggles to capture the fine-grained spatial features of retinal vessels, achieving a test Dice coefficient of only 0.729 and a Loss of 0.5176, with a relatively obvious problem of small vessel missed detection.

The MSCA++-UNet proposed in this paper introduces channel attention on the basis of MSCA, constructs a dual attention cooperative mechanism, and fine-tunes the learning rate, ultimately achieving optimal performance with a Dice coefficient of 0.787, mIoU of 79.7%, and Loss reduced to 0.3624. The cooperative effect of spatial and channel attention effectively suppresses background noise, retains key vessel features, and further improves segmentation accuracy and stability.

In summary, the ablation experiments verify that the MSCA++ dual attention module is the core driving force behind the model's performance improvement. Each improved module specifically addresses the problem of insufficient spatial feature capture. By reasonably adjusting training

parameters, the optimal balance between segmentation accuracy and feature expression ability is achieved, proving the effectiveness and practicality of the improvement strategy proposed in this study.

Table 3. Quantitative Performance Comparison of Different Model Variants on the DRIVE Dataset

Method	Basic Architecture	Main Improvement Strategy	Dice ↑	mIoU [%] ↑	Loss ↓
1	UNet	Baseline Model	0.729	75.6	0.5176
2	Attention UNet	Efficient Attention	0.715	74.6	0.4830
3	MSCA-UNet	MSCA Multi-Scale Coordinate Attention	0.780	79.0	0.3715
4	MSCA++-UNet (Proposed Model)	MSCA++ Coordinate + Channel Attention	0.787	79.7	0.3624

Note: ↑ indicates the higher the value, the better the performance; ↓ indicates the lower the value, the better the performance.

4.5. Analysis of Model Efficiency and Lightweight Performance

To verify the engineering practicality and deployment potential of the proposed MSCA++ multi-scale coordinate-channel attention UNet model, this section analyzes the model from three aspects: training efficiency, lightweight design, and inference adaptability. All experiments are implemented in a CUDA-enabled environment based on the PyTorch framework, with the input being 3-channel retinal images of 480×480. The training parameters are set as batch_size=4 and initial learning rate=0.0008, and all indicators are statistically analyzed under a unified environment to ensure objectivity.

The model employs a base channel configuration of base_c=32, integrated with a lightweight multi-scale attention fusion mechanism. By incorporating attention modules to enhance feature extraction capability while effectively controlling the number of parameters and computational cost, the model avoids efficiency loss caused by complex structures in traditional attention-based UNet models [48]. In terms of training efficiency, the total time required for the model to complete 300 training epochs is only 2487 seconds, with an average of approximately 8.29 seconds per epoch. Furthermore, the model converges rapidly in the early training stage and stabilizes after 200 epochs, ultimately achieving high-precision results of a 0.787 Dice coefficient and 79.7% mean IoU on the test set, realizing the dual effects of accuracy improvement and training time optimization.

Meanwhile, the model can achieve millisecond-level single-image inference for the 480×480 vessel segmentation task. Benefiting from its lightweight design, it can complete fast segmentation without relying on high-performance servers and is adaptable to edge devices and embedded medical terminals with limited GPU memory, solving the common pain point of "heavy computation and difficult deployment" in similar models.

In summary, the MSCA++-UNet model achieves the optimal balance between segmentation accuracy and model efficiency in the DRIVE retinal vessel segmentation task. It not only ensures segmentation performance through innovative attention modules but also possesses efficient training and inference capabilities and good clinical deployability through lightweight structure design, making it of higher practical value in the scene of auxiliary diagnosis of fundus diseases in primary medical institutions. 5. disscusion.

5. DISSCUSION

5.1. Model Generalization Performance and Robustness

The dual-attention fusion mechanism in MSCA++-UNet markedly improves generalization and robustness in complex clinical scenarios by combining multi-scale coordinate attention for precise localization of thin and scattered vessels with channel attention for adaptive feature reweighting, thereby maintaining stable segmentation of both thick and thin structures even under low-contrast regions, lesion-affected areas, and densely distributed fine vessels [53]. On the DRIVE dataset, these advantages manifest as superior Dice (0.787) and mIoU (79.7%) scores compared with baseline variants, indicating stronger anti-interference capability. Although cross-dataset validation was not performed in this study, the consistent performance gains suggest promising generalization potential to heterogeneous fundus images, warranting future multi-center and cross-dataset evaluations to confirm real-world clinical adaptability [50].

5.1.1. Model Computational Complexity and Engineering Application Potential

To assess the performance–efficiency trade-off, the parameters, FLOPs, and training time of the four variants were compared on 480×480 input images (Table 4). While the baseline U-Net exhibited the lowest complexity (1.98 M parameters, 5.78 G FLOPs), the proposed MSCA++-UNet achieved the highest Dice (0.787) with only a modest increase to 2.92 M parameters and 7.48 G FLOPs. This demonstrates that the dual-attention fusion strategy enhances segmentation accuracy without incurring excessive computational overhead. Consequently, MSCA++-UNet is well suited for large-scale retinal screening on standard GPUs; with further lightweight optimizations such as pruning or quantization, it can be readily adapted for real-time applications on mobile devices or edge computing platforms [52].

Table 4. Comparison of model complexity and segmentation performance

Model	Params [M]	FLOPs [G]	Dice ↑
UNet	1.98	5.78	0.729
Attention UNet	2.45	6.92	0.715
MSCA UNet	2.68	7.15	0.780
MSCA++ UNet	2.92	7.48	0.787

5.2. Limitations of the Study

Despite its promising performance on the DRIVE dataset, MSCA++-UNet has several limitations. First, the dual-attention fusion module increases model complexity (2.92 M parameters, 7.48 G FLOPs) relative to the baseline U-Net, leading to longer training times and potential challenges for real-time deployment on resource-constrained edge devices. Second, all experiments were confined to a single dataset without cross-dataset validation (e.g., CHASE_DB1 or STARE), leaving generalization to heterogeneous or severely pathological fundus images unverified. Third, the model depends on fully annotated pixel-level labels, which are labor-intensive and scarce in clinical practice, thereby limiting scalability [52]. These constraints arise from the focus on architectural innovation within a single-dataset framework and the inherent trade-off between attention-enhanced accuracy and computational efficiency. Future work will address them through model compression, multi-center cross-dataset evaluation [54], and semi-supervised learning strategies to enhance clinical practicality.

6. CONCLUSION

6.1. Summary of Research Achievements

This study introduced MSCA++-UNet, a novel retinal vessel segmentation [56] architecture that integrates multi-scale coordinate attention and channel attention within the classic U-Net framework. Evaluated on the DRIVE dataset, the model achieved state-of-the-art performance with a Dice coefficient of 0.787 and mIoU of 79.7%, effectively resolving the long-standing issues of edge loss and inaccurate fine-vessel segmentation. By synergistically capturing spatial localization and channel-wise feature reweighting, MSCA++-UNet delivers superior edge preservation and precise delineation of thin and scattered vessels, providing a robust and clinically viable solution for automated retinal image analysis [57].

6.2. Core Research Contributions

The MSCA++-UNet framework advances retinal vessel segmentation through innovative module design, enhanced performance, and practical clinical applicability. By synergistically integrating multi-scale coordinate attention with channel attention within the U-Net backbone and tailoring module fusion and training strategies to vessel-specific characteristics, the model overcomes the limitations of single-attention and baseline approaches. This results in state-of-the-art segmentation accuracy on the DRIVE dataset (Dice 0.787, mIoU 79.7%) [49-55]. More importantly, these advancements enable robust automated analysis suitable for large-scale retinal screening programs, supporting early detection of fundus diseases and reducing reliance on labor-intensive manual annotation in clinical settings.

6.3. Future Research Directions and Outlook

Future work will address the identified limitations of MSCA++-UNet and expand its clinical applicability. Model compression techniques such as pruning and quantization will be explored to reduce computational complexity while preserving accuracy, thereby enabling real-time deployment on resource-constrained mobile and edge devices. In addition, the framework will be extended to other retinal imaging modalities such as optical coherence tomography (OCT) and integrated with multi-modal data for comprehensive fundus disease diagnosis [58]. By incorporating semi-supervised or self-supervised learning strategies, future iterations will reduce reliance on large-scale annotated datasets [52], enhance generalization across heterogeneous clinical scenarios, and support scalable intelligent diagnostic systems for early detection and management of eye diseases.

CONFLICTS OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ACKNOWLEDGEMENT

This work was supported by the 2025 Annual Undergraduate Research Training Program (No. 2025141), and the 2025 Annual Undergraduate Research Training Program (No. 2025132).

REFERENCES

[1] Arsalan, M., Haider, A., Lee, Y. W., Park, K. R. (2022). Detecting retinal vasculature as a key biomarker for deep learning based intelligent screening and analysis of diabetic and hypertensive retinopathy. Expert Systems with Applications, 200, 117009. https://doi.org/10.1016/j.eswa.2022.117009

[2] Staal, J., Abramoff, M. D., Niemeijer, M., Viergever, M. A., van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. IEEE Transactions on Medical Imaging, 23(4), 501–509. https://doi.org/10.1109/TMI.2004.825627

[3] Li, Y., Zhang, Y., Cui, W., Lei, B., Kuang, X., Zhang, T. (2022). Dual encoder-based dynamic-channel graph convolutional network with edge enhancement for retinal vessel segmentation. IEEE Transactions on Medical Imaging, 41(8), 1975–1989. https://doi.org/10.1109/TMI.2022.3151666

[4] Hoover, A. D., Kouznetsova, V., Goldbaum, M. (2000). Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Transactions on Medical Imaging, 19(3), 203–210. https://doi.org/10.1109/42.845178

[5] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI), 9351, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

[6] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., Liang, J. (2018). UNet++: A nested U-Net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, 3–11. https://doi.org/10.1007/978-3-030-04261-6_3

[7] Badrinarayanan, V., Kendall, A., Cipolla, R. (2017). SegNet: A deep convolutional encoder–decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615

[8] Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N., Kainz, B., Glocker, B., Rueckert, D. (2018). Attention U-Net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999. https://arxiv.org/abs/1804.03999

[9] Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E. (2020). Squeeze-and-excitation networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(8), 2011–2023. https://doi.org/10.1109/TPAMI.2019.2913372

[10] Woo, S., Park, J., Lee, J., Kweon, I. (2018). CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), 3–19. https://doi.org/10.1007/978-3-030-01234-2_1

[11] Hou, Q., Zhou, D., Feng, J. (2021). Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 13713–13722. https://doi.org/10.1109/CVPR46437.2021.01350

[12] Li, L., Verma, M., Nakashima, Y., Nagahara, H., Kawasaki, R. (2019). IterNet: Retinal image segmentation utilizing structural redundancy in vessel networks. arXiv preprint arXiv:1910.03297. https://arxiv.org/abs/1910.03297

[13] Mou, L., Zhao, Y., Fu, H., Liu, Y., Cheng, J., Zheng, Y., Su, P., Yang, J., Chen, L., Frangi, A. F., Akiba, M., Liu, J. (2021). CS²-Net: Deep learning segmentation of curvilinear structures in medical imaging. Medical Image Analysis, 67, 101874. https://doi.org/10.1016/j.media.2020.101874

[14] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

[15] Huang, G., Liu, Z., Pleiss, G., van der Maaten, L., Weinberger, K. Q. (2022). Convolutional networks with dense connectivity. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(12), 8704–8716. https://doi.org/10.1109/TPAMI.2019.2918284

[16] Shafiq, M., Gu, Z. (2022). Deep residual learning for image recognition: A survey. Applied Sciences, 12(18), 8972. https://doi.org/10.3390/app12188972

[17] Emara, T., Abd El Munim, H. E., Abbas, H. M. (2019). LiteSeg: A novel lightweight ConvNet for semantic segmentation. arXiv preprint arXiv:1912.06683. https://arxiv.org/abs/1912.06683

[18] Canny, J. (1986). A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6), 679–698. https://doi.org/10.1109/TPAMI.1986.4767851

[19] Xie, Y., Xing, F., Kong, X., Su, H., Yang, L. (2015). Beyond classification: Structured regression for robust cell detection using convolutional neural network. In Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI), 9351, 358–365. https://doi.org/10.1007/978-3-319-24574-4_43

[20] Chen, H., Qi, X., Yu, L., Dou, Q., Qin, J., Heng, P.-A. (2017). DCAN: Deep contour-aware networks for object instance segmentation from histology images. Medical Image Analysis, 36, 135–146. https://doi.org/10.1016/j.media.2016.11.004

[21] Alahmadi, M. D. (2023). Boundary aware U-Net for medical image segmentation. Arabian Journal for Science and Engineering, 48(8), 9929–9940. https://doi.org/10.1007/s13369-022-07431-y

[22] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q. (2019). ECA-Net: Efficient channel attention for deep convolutional neural networks. arXiv preprint arXiv:1910.03151. https://arxiv.org/abs/1910.03151

[23] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

[24] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. https://arxiv.org/abs/1704.04861

[25] Gu, Z., Cheng, J., Fu, H. (2019). CE-Net: Context encoder network for 2D medical image segmentation. IEEE Transactions on Medical Imaging, 38(10), 2281–2292. https://doi.org/10.1109/TMI.2019.2903562

[26] Chen, W., Yu, S., Ma, K., Ji, W., Bian, C., Chu, C., Shen, L., Zheng, Y. (2022). TW-GAN: Topology and width aware GAN for retinal artery/vein classification. Medical Image Analysis, 77, 102340. https://doi.org/10.1016/j.media.2021.102340

[27] Ou, Y., Xue, Y., Yuan, Y., Xu, T., Pisztora, V., Li, J., Huang, X. (2020). Semi-supervised cervical dysplasia classification with learnable graph convolutional network. arXiv preprint arXiv:2004.00191. https://arxiv.org/abs/2004.00191

[28] Remeseiro López, B., Mendonça, A. M., Campilho, A. (2020). Automatic classification of retinal blood vessels based on multilevel thresholding and graph propagation. The Visual Computer, 37, 1247–1261. https://doi.org/10.1007/s00371-020-01863-z

[29] Georgiadis, P., Gkouvrikos, E. V., Vrochidou, E., Kalampokas, T., Papakostas, G. A. (2025). Building better deep learning models through dataset fusion: A case study in skin cancer classification with hyperdatasets. Diagnostics, 15(3), 352. https://doi.org/10.3390/diagnostics15030352

[30] Sun, L., Huang, X., Liu, J., et al. (2025). Remaining useful life prediction of lithium batteries based on jump connection multi-scale CNN. Scientific Reports, 15, 32873. https://doi.org/10.1038/s41598-025-08619-6

[31] Zhao, D., Cai, W., Cui, L. (2024). Adaptive thresholding and coordinate attention-based tree-inspired network for aero-engine bearing health monitoring under strong noise. Advanced Engineering Informatics, 61, 102559. https://doi.org/10.1016/j.aei.2024.102559

[32] Xia, Y., Guan, D., Zhou, Z. (2025). CNN-SENet: A GNSS-R ocean wind speed retrieval model integrating CNN and SENet attention mechanism. Satellite Navigation, 6, 3. https://doi.org/10.1186/s43020-024-00157-2

[33] Fu, B., Peng, Y., He, J., Tian, C., Sun, X., Wang, R. (2024). HmsU-Net: A hybrid multi-scale U-net based on a CNN and transformer for medical image segmentation. Computers in Biology and Medicine, 170, 108013. https://doi.org/10.1016/j.compbiomed.2024.108013

[34] Sun, Y., Bi, F., Gao, Y., Chen, L., Feng, S. (2022). A multi-attention UNet for semantic segmentation in remote sensing images. Symmetry, 14(5), 906. https://doi.org/10.3390/sym14050906

[35] Matteucci, A., Russo, M., Galeazzi, M., Pandozi, C., Bonanni, M., Mariani, M. V., Pierucci, N., La Fazia, V. M., Di Fusco, S. A., Nardi, F., et al. (2025). Impact of ablation energy sources on perceived quality of life and symptom in atrial fibrillation patients: A comparative study. Journal of Clinical Medicine, 14(8), 2741. https://doi.org/10.3390/jcm14082741

[36] Tavera, F., Haider, H. (2025). The role of selective attention in implicit learning: Evidence for a contextual cueing effect of task-irrelevant features. Psychological Research, 89, 15. https://doi.org/10.1007/s00426-024-02033-9

[37] Zhu, X., Cao, B., Zhang, W., et al. (2025). Adaptive multi-scale feature extraction and fusion network with deep supervision for retinal vessel segmentation. Multimedia Systems, 31, 197. https://doi.org/10.1007/s00530-025-01789-3

[38] Chen, J., Wan, J. Z., Fang, Z. H., et al. (2023). LMSA-Net: A lightweight multi-scale aware network for retinal vessel segmentation. International Journal of Imaging Systems and Technology, 33(5), 1515–1530. https://doi.org/10.1002/ima.22881

[39] Quintana-Quintana, O. J., Aceves-Fernández, M. A., Pedraza-Ortega, J. C., et al. (2025). Deep learning techniques for retinal layer segmentation to aid ocular disease diagnosis: A review. Computers, 14(8), 298. https://doi.org/10.3390/computers14080298

[40] Islam, M. M., Poly, T. N., Walther, B. A., Yang, H. C., Li, Y.-C. (2020). Artificial intelligence in ophthalmology: A meta-analysis of deep learning models for retinal vessels segmentation. Journal of Clinical Medicine, 9(4), 1018. https://doi.org/10.3390/jcm9041018

[41] Tan, J. H., Acharya, U. R., Bhandary, S. V., Chua, K. C., Sivaprasad, S. (2017). Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network. Journal of Computational Science, 20, 70–79. https://doi.org/10.1016/j.jocs.2017.02.006

[42] Zhu, C., Zou, B., Zhao, R., Cui, J., Duan, X., Chen, Z., Liang, Y. (2017). Retinal vessel segmentation in colour fundus images using extreme learning machine. Computerized Medical Imaging and Graphics, 55, 68–77. https://doi.org/10.1016/j.compmedimag.2016.05.004

[43] Leopold, H. A., Orchard, J., Zelek, J. S., Lakshminarayanan, V. (2019). PixelBNN: Augmenting the PixelCNN with batch normalization and the presentation of a fast architecture for retinal vessel segmentation. Journal of Imaging, 5(2), 26. https://doi.org/10.3390/jimaging5020026

[44] Soomro, T. A., Hellwich, O., Afifi, A. J., Paul, M., Gao, J., Zheng, L. (2018). Strided U-Net model: Retinal vessels segmentation using dice loss. In Proceedings of Digital Image Computing: Techniques and Applications (DICTA), Canberra, Australia, 1–8. https://doi.org/10.1109/DICTA.2018.8615770

[45] Soomro, T. A., Afifi, A. J., Gao, J., Hellwich, O., Khan, M. A. U., Paul, M., Zheng, L. (2017). Boosting sensitivity of a retinal vessel segmentation algorithm with convolutional neural network. In Proceedings of the International Conference on Digital Image Computing: Techniques and Applications (DICTA), Sydney, NSW, Australia, 1–8. https://doi.org/10.1109/DICTA.2017.8227413

[46] Khan, M. R. K., Mohaidat, T., Khalil, K. (2026). A lightweight modified adaptive UNet for nucleus segmentation. Sensors, 26(2), 665. https://doi.org/10.3390/s26020665

[47] Pan, P., Zhang, C., Sun, J., Guo, L. (2025). Multi-scale conv-attention U-Net for medical image segmentation. Scientific Reports, 15(1), 12041. https://doi.org/10.1038/s41598-025-96552-5

[48] Wang, J., Li, X., Ma, Z. (2025). Multi-scale three-path network (MSTP-Net): A new architecture for retinal vessel segmentation. Measurement, 250, 117100. https://doi.org/10.1016/j.measurement.2025.117100

[49] Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., Lungren, M. P., Zhang, S., Xing, L., Lu, L., Yuille, A., Zhou, Y. (2024). TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis, 97, 103280. https://doi.org/10.1016/j.media.2024.103280

[50] Zhao, X., Wu, Z., Li, X., Zhang, Y., Wei, Y., Zhang, Y., Wang, L. (2024). RCPS: Rectified contrastive pseudo supervision for semi-supervised medical image segmentation. IEEE Journal of Biomedical and Health Informatics, 28(1), 251–261. https://doi.org/10.1109/JBHI.2023.3322590

[51] Su, H., Gao, L., Wang, Z., Yu, Y., Hong, J., Gao, Y. (2024). A hierarchical full-resolution fusion network and topology-aware connectivity booster for retinal vessel segmentation. IEEE Transactions on Instrumentation and Measurement, 73, 1–16. https://doi.org/10.1109/TIM.2024.3411133

[52] Wang, C., Zhao, Z., Ren, Q., Xu, Y., Yu, Y. (2019). Dense U-net based on patch-based learning for retinal vessel segmentation. Entropy, 21(2), 168. https://doi.org/10.3390/e21020168

[53] Yuan, Y., Zhang, L., Wang, L., Huang, H. (2022). Multi-level attention network for retinal vessel segmentation. IEEE Journal of Biomedical and Health Informatics, 26(1), 312–323. https://doi.org/10.1109/JBHI.2021.3089201

[54] Li, K., Qi, X., Luo, Y., Yao, Z., Zhou, X., Sun, M. (2021). Accurate retinal vessel segmentation in color fundus images via fully attention-based networks. IEEE Journal of Biomedical and Health Informatics, 25(6), 2071–2081. https://doi.org/10.1109/JBHI.2020.3028180

[55] Singh, L. K., Khanna, M., Thawkar, S., Singh, R. (2024). A deep learning-based system for efficient and automatic blood vessel segmentation from retinal fundus images. Multimedia Tools and Applications, 83, 6005–6049. https://doi.org/10.1007/s11042-023-15348-3

[56] Quan, X., Hou, G., Yin, W., Zhang, H. (2025). A multi-modal and multi-stage fusion enhancement network for segmentation based on OCT and OCTA images. Information Fusion, 113, 102594. https://doi.org/10.1016/j.inffus.2024.102594

[57] Lakshmanan, K., Tessicini, F., Gil, A. J., Auricchio, F. (2023). A fault prognosis strategy for an external gear pump using machine learning algorithms and synthetic data generation methods. Applied Mathematical Modelling, 123, 348–372. https://doi.org/10.1016/j.apm.2023.07.001

[58] Esposito, A., Lappa, M., Pagliara, R., Spada, G. (2022). A mixed radiative-convective technique for the calibration of heat flux sensors in hypersonic flow. FDMP-Fluid Dynamics & Materials Processing, 18(2), 189–203. https://doi.org/10.32604/fdmp.2022.019605