


ORIGINAL RESEARCH OPEN ACCESS

EANet: Integrate Edge Features and Attention Mechanisms Multi-Scale Networks for Vessel Segmentation in Retinal Images

Jiangyi Zhang^{1,2,3} | Yuxin Tan² | Duantengchuan Li⁴  | Guanghui Xu¹ | Fuling Zhou^{2,3,5}

¹School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan, China | ²Department of Hematology, Zhongnan Hospital of Wuhan University, Wuhan, China | ³Research Center for Lifespan Health, Wuhan University, Wuhan, China | ⁴School of Information Management, Wuhan University, Wuhan, China | ⁵School of Nursing, Wuhan University, Wuhan, China

Correspondence: Duantengchuan Li (dtcleel222@whu.edu.cn) | Guanghui Xu | Fuling Zhou

Received: 28 July 2024 | **Revised:** 14 February 2025 | **Accepted:** 18 March 2025

Funding: This work is supported by the National Natural Science Foundation of China (No. 82370176), the Key Research and Development Program of Hubei Province (No. 2023BCB025) and the Fundamental Research Funds for the Central Universities (No. 2042024kf0024).

ABSTRACT

Accurately extracting blood vessel structures from retinal fundus images is critical for the early diagnosis and treatment of various ocular and systemic diseases. However, retinal vessel segmentation continues to face significant challenges. Firstly, capturing the boundary information of small vessels is particularly difficult. Secondly, uneven vessel thickness and irregular distribution further complicate the multi-scale feature modelling. Lastly, low-contrast images lead to increased background noise, further affecting the segmentation accuracy. To tackle these challenges, this article presents a multi-scale segmentation network that combines edge features and attention mechanisms, referred to as EANet. It demonstrates significant advantages over existing methods. Specifically, EANet consists of three key modules: the edge feature enhancement module, the multi-scale information interaction encoding module, and the multi-class attention mechanism decoding module. Experimental results validate the effectiveness of the method. Specifically, EANet outperforms existing advanced methods in the precise segmentation of small and multi-scale vessels and in effectively filtering background noise to maintain segmentation continuity.

1 | Introduction

Retinal fundus diseases are closely linked to visual health, with retinal lesions caused by leukaemia [1] and tumours [2] being of particular concern. For instance, leukemia can cause symptoms such as cotton-wool spots, retinal vein tortuosity, and retinal infiltration [3], while tumours can cause tumour-associated retinal lesions [4]. These lesions can result in irreversible visual impairment. In addition, conditions such as diabetes [5], heart disease [6], and hypertension [7] can also cause retinal lesions, leading to visual impairment. Therefore, accurate and rapid retinal vessel

lesion screening can provide critical clinical information for disease diagnosis and effectively prevent the rapid deterioration of associated conditions. Healthcare professionals typically rely on visual inspection to screen retinal vessels in clinical practice. This screening approach is extremely sensitive to external influences, including collection equipment and lighting, as well as variations in the subjective diagnosis of physicians, which significantly reduces the accuracy and efficiency of detection [8]. Consequently, automated retinal vessel segmentation technology plays a crucial role in mitigating these problems and is of great importance for intelligent screening of retinal-related diseases [9].

Abbreviations: ANA, anti-nuclear antibodies; APC, antigen-presenting cells; IRF, interferon regulatory factor.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *IET Image Processing* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

It is significant to note that the models above focused exclusively on modelling either shallow features or deep semantic features in the process of feature capture, neglecting to incorporate contextual information effectively. Furthermore, due to the irregular distribution of vessels and the indistinct edge features in vascular images, networks struggle to identify channel information accurately. This, in turn, limits the model's ability to extract detailed vessel features [10]. Owing to the inherent intricacy of vessel shapes and their low contrast against the background, it is crucial to effectively extract vessel edge details for successful segmentation of retinal vessels, mainly when dealing with fine vessels.

Early mainstream approaches to retinal image processing often used techniques such as thresholding and filtering [11, 12]. These techniques are easy to implement and can provide segmentation results with a degree of accuracy. However, these approaches fail to segment vessels in densely populated areas of retinal images accurately and somewhat rely on manual intervention for interpretation. Recently, convolutional neural network based techniques have achieved commendable success in image segmentation. For instance, Wang et al. [13] integrated CNN with random forests for vessel segmentation, thereby improving the accuracy of segmentation. Given the abundance of edge information in vascular images, López-Linares et al. [14] developed a fully convolutional network that exploits edge detection for automated vessel feature modelling. Specifically, the U-Net model [15] achieves commendable segmentation performance due to its simplicity of a wide range of U-Net-enabled variants already extensively deployed. The principles of residual connections, multi-scale fusion, and transfer learning were applied, which increased the segmentation efficiency of the U-Net [16]. Recognizing the importance of feature fusion, Khened et al. [17] merged the U-Net architecture to introduce a novel fully convolutional neural network (FCN) framework.

While the aforementioned methods have yielded satisfactory results, many are constrained by fixed receptive fields, which limits their ability to capture features at varying scales. Moreover, these methods are unable to effectively model the vessel edge features. As a result, accurately restoring the characteristics of fine vessels in vascular images is challenging, making it difficult to increase the final segmentation quality [18]. Acknowledging the importance of integrating multi-scale contextual features, Wu et al. [19, 20] have advanced the network's segmentation capabilities for scenarios with substantial scale variations via multi-scale interactions, feature aggregation, and multi-level semantic supervision modules. To improve the network's capacity for modelling global features, Guo et al. [21, 22] introduced an attention mechanism. This mechanism helps the network learn crucial spatial and channel information, improving its resistance to interference. Furthermore, Qi et al. [23] utilized deformable convolutions to detect retinal vessels of various shapes and sizes, enhancing the network's capability of capturing finer details. The focus was on the pronounced tubular characteristics of the vessels.

Considering that the current vessel segmentation task continues to face challenges, including the difficulty in capturing small vessels, the high complexity of multi-scale vessel feature modelling, and the significant impact of background noise, this article introduces a network architecture, designated as EANet. The

network comprises three modules: an edge feature enhancement (EFE) module, a multi-scale feature exchange encoding (MSE) module, and a multi-category attention mechanism decoding (MAD) module. The first stage of the process is the extraction of the vessel edge features, which is carried out effectively by the EFE module. Secondly, the MSE module enhances the network's capacity to extract features from vessels of varying thickness, minimizes the loss of tubular feature information and integrates the contextual information extracted by the network. Lastly, the MAD module enhances the network's capacity to model global features and improves the overall performance of feature extraction. The validity as well as feasibility of EANet were demonstrated on three mainstream retinal vessel datasets. Comparative experimental findings demonstrate that the EANet model surpasses current mainstream approaches across the metrics of accuracy, mean intersection over union, and F1 score. The four major contributions summarised are as follows:

- We present a multi-scale network for retinal vessel segmentation that integrates edge features with attention mechanisms, which effectively learns edge features of vessel images and exhibits robust multi-scale feature fusion capabilities.
- Given the challenges in extracting small vessel boundary features, we propose an edge feature enhancement (EFE) module designed to improve the network's ability to express small vessel features with both completeness and precision.
- To tackle the issue of uneven vessel thickness and irregular distribution, we introduce a multi-scale information interaction encoding (MSE) module. It dynamically models the tubular characteristics of vessels while integrating multi-scale contextual information.
- We introduce a multi-category attention mechanism decoding (MAD) module, which is designed to enhance the network's capacity to model global features during the decoding phase. It retains more details while minimizing the impact of background noise in low-contrast images.

2 | Related Work

2.1 | Traditional Retinal Image Processing Methods

Traditional approaches to retinal image processing predominantly involve matched filtering and machine learning techniques. The use of matched filtering in image segmentation enables the incorporation of various techniques such as thresholding, edge detection, and morphological operations. Chaudhuri et al. [24] introduced the feature selection method designed explicitly for fundus vessel segmentation. This method combines the computational simplicity of threshold-based edge operators with the advantages of model-based edge detectors. To mitigate the effects of image background noise on the segmentation results, Moghimirad et al. [25] employed techniques weighted by mean deviation and median function, acknowledging the influence of background noise on segmentation outcomes. Accurately segmenting vascular images can be significantly complicated by the abundance of small vessels. To address this issue, Kar et al. [26] introduced a matched filtering technique for enhanced

extraction from small blood. The techniques mentioned above are effective in segmenting images with clear and specific characteristics. However, they often require manual parameter adjustments and rely on prior knowledge and specific assumptions. These methods address challenges such as high computational complexity, sensitivity to noise, and poor segmentation performance in images with low contrast or greyscale overlap. Consequently, when addressing complex or low-contrast vascular images, their efficacy falls short of the desired level.

To address these challenges, researchers have developed a variety of machine learning-based approaches to image processing. These techniques initially extract features from input images, utilizing gold standards to train networks specifically for vascular image segmentation. Marín et al. [27] introduced a supervised learning technique that uses neural networks for pixel classification, recognizing the importance of labelled data. To automate vessel segmentation while minimizing false positives in pathological areas, a semi-supervised self-training method and AdaBoost classifier [28] were utilized for training on vessel samples, enabling automatic vessel segmentation. In contrast to the previously mentioned supervised training methods, [29] adopted an unsupervised approach that integrates line detectors with image enhancement techniques for vessel segmentation. Recognizing the importance of extracting precise edge features to enhance segmentation accuracy, studies by [30] have focused on vessel edge detection. They have utilized random decision forests and heuristic algorithms to extract multi-scale edge features, resulting in improved segmentation accuracy.

In comparison to methods such as matched filtering, machine learning techniques can learn features from data independently and decrease operator involvement while improving the accuracy and adaptability of segmentation. However, these methods may have limited generalization capabilities, and for data with complex features, the segmentation outcomes may not meet expectations.

2.2 | Deep Learning Retinal Image Processing Methods

Deep learning, which uses backpropagation and multilayer neural networks, is highly effective at automatically learning complex features. Its ability to generalize models robustly has led to its widespread application across numerous domains [31–35].

Over the past few years, many researchers using deep learning techniques have tackled challenges in vessel segmentation. Existing research mainly focuses on the following four areas.

Encoder–decoder architecture: Ronneberger et al. [15] introduced the U-Net, which has been successful in medical image segmentation. Due to U-Net's exceptional performance in medical image segmentation, there has been a growing effort to improve and refine its network structure. Wu et al. [36] integrated cascaded dilated convolutions and SE attention residual blocks into the U-Net framework to enhance the ability to extract global multi-scale features. This method facilitates effective segmentation of retinal vessels while enhancing the extraction of multi-scale characteristics. Owing to the abundance of small vessels in

vascular images, Dong et al. [37] developed a cascaded residual attention U-Net, focusing on the detailed segmentation of retinal vessels from coarse to fine details.

Edge feature perception and optimization: Considering the abundance of boundary feature information in retinal vessel images, effectively utilizing this data can greatly improve the final segmentation accuracy. Consequently, [38] used boundary feature detection techniques in conjunction to refine edge features. Zhu et al. [39] introduces a boundary shape correction module to enhance the accuracy of retrieving boundary position information. Zhu et al. [40] utilizes an edge spatial attention module to model edge features.

Multi-scale feature interaction strategy: However, the original U-Net structure's inability to effectively merge extracted shallow edge and deep semantic information led to the introduction of feature fusion [41–43], which aims at a high aggregation of contextual data within that network to efficiently acquire global relations as well as the nuances of the underlying feature space. Zhu et al. [44] integrates heterogeneous features through multi-level edge feature fusion. Zhu et al. [45] reduces computational complexity while enabling multi-scale feature interaction.

Attention mechanism collaboration: Sun et al. [46] used multiple attention mechanisms within an encoder-decoder architecture to capture a rich array of multi-scale contextual and advanced semantic features. Acknowledging the prominent tubular characteristics present in vascular images, Deng and Jin et al. [47] applied the concept of deformable convolution integrated with multi-scale attention modules. This enabled the network to adaptively discern more tubular structural features, consequently improving the network's segmentation ability.

Deep learning approaches are known for superior feature extraction and generalization capabilities compared to traditional techniques. However, most of these methods ignore the significance of edge information and do not effectively integrate it with other types of features. These methods inadequately integrate multi-scale contextual features, resulting in poor segmentation precision for small vessels. The FPN network, for instance, has not been specifically designed for the vascular segmentation task. Therefore, extracting edge features effectively and integrating them with multi-scale contextual features are critical for further research.

3 | Methodology

3.1 | Overall Architecture

This study presents EANet, a multi-scale network that integrates edge features with attention mechanisms for retinal vessel segmentation, as shown in Figure 1. EANet consists of three primary modules: an EFE, an MSE, and an MAD.

To extract more precise contextual information, integrate multi-scale data, and enhance the network's ability to capture fine vessel features, this article proposes an EFE. Focusing on the local and global composite features of vascular images extracted by VGG-16, the EFE enhances edge information on feature maps

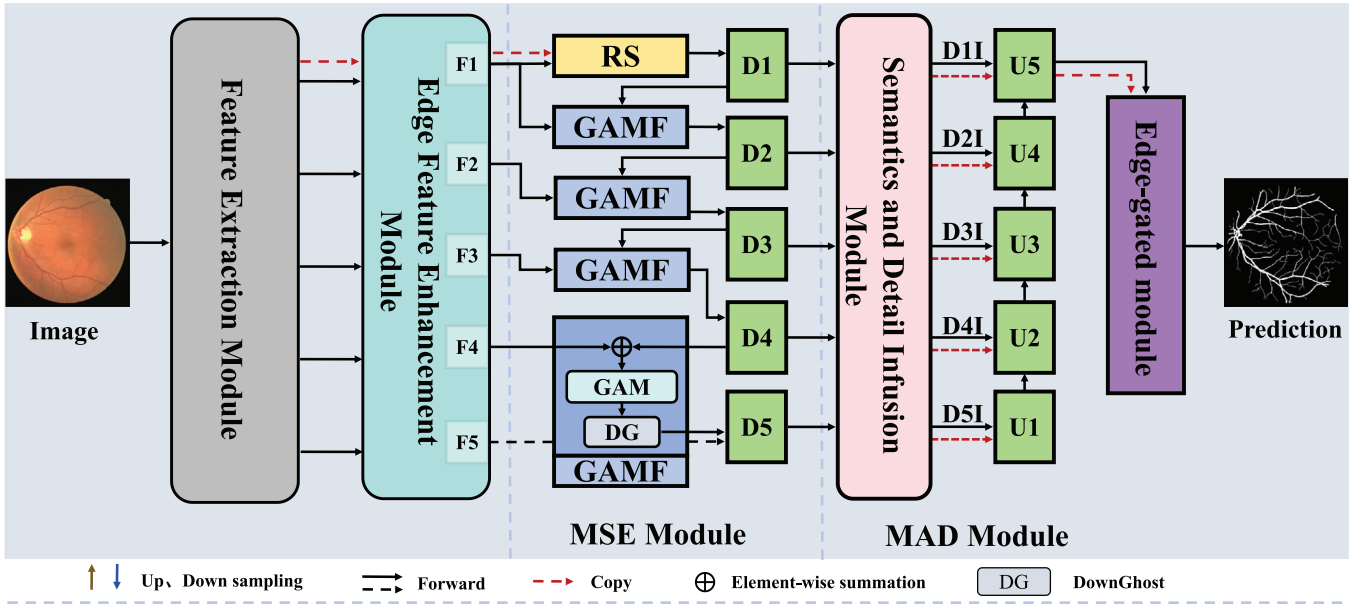


FIGURE 1 | The EANet's overall architecture comprises three primary modules: the edge feature enhancement module (EFE), the multi-scale information interaction encoding module (MSE), and the multi-attention mechanism decoding module (MAD). The MSE is composed of the RS and GAMF sections, as shown in the diagram. MAD, on the other hand, is composed of the U1 to U5 units and the CHG component.

to facilitate feature fusion during the encoding phase, enabling a more comprehensive capture of edge feature details.

Considering the presence of numerous tubular structures in vascular images and the uneven distribution of vessels with varying thicknesses, this article proposes an MSE for the encoding stage. The method utilizes dynamic snake convolution to capture abundant tubular structural information in vascular images, adapting to changes in vessel structure. Subsequently, a global attention mechanism is introduced to minimize information attenuation and enhance global interactions.

In the final decoding module, to address the issue of low contrast and significant background noise in the original images, which leads to the loss of detailed information, this article introduces an MAD. The MAD module aims to diminish the amount for upsampling and to improve integration between global and local data. MAD comprises a multi-attention upsampling machine that uses an edge element fusion machine. An upsampling module with a multi-attention mechanism can capture a wide range of global and local information while avoiding detailed feature loss upsampling. The module for edge feature fusion merges the original edge features with those obtained via the decoding module. This accentuates the edge features while enhancing the module's capacity to capture information across various scales and abstraction levels. As a result, the precision of the final segmentation is elevated.

3.2 | Edge Feature Enhancement Module

3.2.1 | Feature Extraction

The feature extraction section uses a pre-trained VGG-16 model [48]. The model has a simple and repetitive architecture, con-

sisting of 13 convolutional layers and five max pooling layers. In addition, this architecture uses several 3×3 convolutional layers instead of a larger convolutional kernel. Convolutional layers are primarily designed to extract features within images. Pooling layers decrease feature map size, increasing network receptive field. The structure's repetitive nature simplifies the network's comprehension and training while deepening it effectively. Beginning with the input feature map $F_{\text{Im}} \in \mathbb{R}^{3 \times H \times W}$, VGG-16 extracts the feature map identified as $(F_1^i, F_2^i, F_3^i, F_4^i, F_5^i)$. This fragment includes components $F_1^i \in \mathbb{R}^{C \times H \times W}$, $F_2^i \in \mathbb{R}^{2C_{\text{in}} \times H \times W}$, $F_3^i \in \mathbb{R}^{4C_{\text{in}} \times H \times W}$, and $F_4^i, F_5^i \in \mathbb{R}^{8C_{\text{in}} \times H \times W}$.

3.2.2 | Feature Enhancement

Due to the abundance of vascular boundaries in retinal vessel images, it is crucial to effectively extract and utilize comprehensive and rich boundary information to enhance segmentation accuracy. Therefore, this research uses the Canny algorithm [49] for feature enhancement to extract effective vessel boundary information. The Canny algorithm extracts a feature map, denoted as $(F_C^0, F_C^1, F_C^2, F_C^3)$, which includes $F_C^0 \in \mathbb{R}^{1 \times H \times W}$, $F_C^1 \in \mathbb{R}^{4C_{\text{in}} \times H \times W}$, $F_C^2 \in \mathbb{R}^{8C_{\text{in}} \times H \times W}$, and $F_C^3 \in \mathbb{R}^{8C_{\text{in}} \times H \times W}$. VGG-16 has numerous convolutional and pooling layers with larger kernel sizes and strides. This design may cause small features to be overlooked or blurred, hindering the extraction of rich minute features in vascular images. Consequently, the ParNet block [50] is introduced to enhance VGG-16's feature extraction capabilities Figure 2. This module enhances the network's receptive field without increasing its depth, thereby significantly improving its capability for obtaining valuable information within feature maps. Furthermore, the concept of feature fusion is embraced, merging boundary feature information with extracted feature information. The deep network feature information is combined with the shallow network feature information to enable

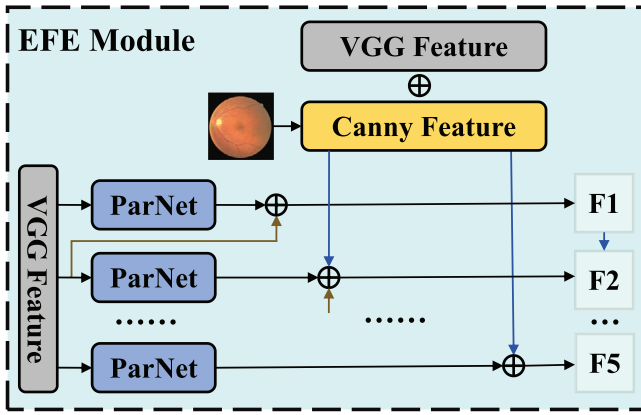


FIGURE 2 | The edge feature enhancement module effectively fuses the extracted edge features with the multi-scale information extracted by the ParNet module to provide more effective information for the encoding module.

multi-scale feature fusion. This enhances the semantic clarity of the vascular boundary region. This study introduces an edge feature enhancement module that can extract feature maps F_i at five distinct depths based on the preceding analysis. Model local semantic information, such as boundary textures and the subtle contrast between blood vessels and the background. If $i \in [2, 4]$ is the case, the output feature map F_i is represented as:

$$F_i = f_p(F_i^i) \oplus Up(F_i^{i+1}) \oplus (\circ(F_i^{i-1})) \oplus (\circ(F_C^{i-1})), \quad (1)$$

if $i = 1$ is the case, then $F_i = f_p(F_i^1) \oplus Up(F_i^2)$ follows. \oplus denotes the element-wise summation operation, \circ denotes the MaxPooling operation, $Up()$ denotes the upsampling operation, and f_p denotes the ParNet operation.

3.3 | Encoding and Decoding Module

To enhance segmentation accuracy, it is essential to effectively integrate contextual information for multi-scale feature fusion, considering the significant variation in scale information among retinal vessels. The U-Net architecture's original encoding and decoding modules execute convolution and down/up-sampling operations to extract pertinent information from feature maps and assimilate contextual data. However, standard convolution operations have a relatively fixed receptive field, which can cause fine vessels to be missed and, therefore, cannot accurately capture the features of complex vascular structures. As the network depth increases, shallow and minute feature information is lost, resulting in inadequate feature restoration. Therefore, this study introduces an MSE. In the encoding stage, we first introduce the MSE module, which consists of two parts: the residual dynamic snake convolution (RS) module and the fusion attention downsampling mechanism (GAMF). It captures global semantics, such as the overall vascular morphology and spatial relationships of large vessels, while encoding local semantics, including fine-grained details of small vessels. The RS module can incorporate more detailed information about tubular structures, while the GAMF better integrates contextual information at multiple scales. Consequently, it provides the decoding module with feature information at different scales, demonstrating effective

feature fusion. The dynamic serpentine convolution [23] can adaptively concentrate on slender and tortuous local structures. Hence, this structure is well-suited for the extraction of vascular features. Simultaneously, the residual structure lightens the network training load and preserves essential information from the original input, accomplishing local feature fusion. Regarding the GAM [51] attention mechanism, it filters out redundant information, mitigates the loss of crucial data, and enhances global interaction capacity. This compensates for the information loss resulting from downsampling operations. However, the use of the original convolution operations introduces some degree of network complexity. Therefore, we substituted them with the ghost module [52]. This module collects additional feature information from less complex operations, reducing the burden on network training.

Efficiently restoring the detailed information of acquired features in the decoding module is crucial for enhancing the segmentation accuracy. However, the original U-Net architecture's decoding module fails to fully and accurately restore relatively minor details. Simultaneously, its ability to integrate both global and local information is limited. Thus, this study introduces a multi-attention mechanism decoding structure. Within the skip connection segment, we have retained the SDI module from U-Netv2 [53]. A multi-attention mechanism and a context feature fusion module were incorporated during the decoding phase. During the layer-by-layer sampling process, distinct decoding sub-modules were crafted for different structural layers. The model employs the position attention module, channel attention module, and spatial attention module for layers U1 to U3 to combine associated features from spatial and channel dimensions. Moreover, to counteract the gradient vanishing issue stemming from the repetitive stacking of identical modules, we implemented structural alterations in each layer. In layers U2 and U3, we integrated edge features to give it the ability to capture the detail from various perspectives. Thereby boosting the network's generalizability. We acknowledge the significance of edge features and the potential loss of minute features. As layer U4 is located in the midsection of the decoder, it contains both abundant shallow detail information and extensive deep semantic information. Consequently, we added the context feature extraction module at layer U4. This minimizes the loss of detailed information without significantly increasing the network's computational demand. The module for extracting context features consists of two parts: a dense atrous convolution module and a residual multi-kernel pooling module [54]. Finally, we use the concept of gated convolution [55] to combine edge features with others to achieve accurate segmentation predictions. The techniques for obtaining edge features include Canny edge detection and holistic nested edge detection [56]. Due to the complexity and richness of detail in vascular images, the holistic nested edge detection method can extract deeper features and contextual information. This method blends two edge detection techniques to accentuate vascular edge features. The following section provides specific details.

3.3.1 | Residual Dynamic Serpentine Convolution

Vascular images contain many tubular structural features. Therefore, by incorporating these features, the encoding module can

better integrate multi-scale contextual information. This study introduces a residual dynamic serpentine convolution module, which consists of two components: a residual structure and dynamic snake convolution. Using dynamic serpentine convolution allows for acquiring additional tubular feature information \mathbf{F}_S . Additionally, this technique can extract richer spatial information from input images, resulting in more diverse features. Simultaneously, we aim to diversify the features by exploiting input images that possess richer spatial information. Following the principle of residual structures and feature fusion, we amalgamate the original image feature information, the features marked as \mathbf{F}_{im} and \mathbf{F}_1 , and the tubular feature information as \mathbf{F}_S to obtain more comprehensive tubular feature information. The technique of dynamic snake convolution focuses on slender and curved local structures, allowing for accurate capture of tubular structural features. Dynamic snake convolution is an extension of deformable convolution [57]. It incorporates deformation offsets and uses an iterative strategy to select observation positions for each target under consideration sequentially. This approach preserves the continuity of attention without excessively dispersing the receptive field, even when significant deformation offsets are present [23, 58]. Secondly, to optimize the utilization of edge information in the original images, the structure incorporates residual connections to facilitate feature fusion. This method provides the GAMF module with more accurate information on tubular features, thereby improving the integration of multi-scale contextual information.

Initially, the tubular structural features \mathbf{F}_S of the input image \mathbf{F}_{im} are captured using dynamic snake convolution. Subsequently, a 1×1 convolution is applied to \mathbf{F}_{im} to align the channel count with \mathbf{F}_1 . Finally, a feature map enriched with tubular features is obtained by performing an element-wise summation on all acquired feature maps \mathbf{F}_{RS} .

$$\mathbf{F}_S = \phi_S(\mathbf{F}_{im}), \quad (2)$$

$$\mathbf{F}_{RS} = \phi_1(\mathbf{F}_{im}) \oplus \mathbf{F}_1 \oplus \mathbf{F}_S, \quad (3)$$

where ϕ_S represents the dynamic snake convolution operation and ϕ_1 represents the 1×1 convolution.

3.3.2 | Fusion Attention Downsampling Mechanism

The EFE assimilates contextual information from various scales of the input image. Concurrently, the RS module captures richer tubular structural details, thereby enhancing the integration of multi-scale contextual information. This supplies the decoding module with more precise vascular feature information and more precise spatial relationships. Based on the abovementioned analysis, this study presents a fusion attention downsampling mechanism. The mechanism consists two parts: GAM mechanism and down ghost mechanism as shown in Figure 3. The GAM attention mechanism is a global attention framework that combines spatial features from \mathbf{F}_S with semantic features that increase progressively from \mathbf{F}_2 to \mathbf{F}_5 . This process provides the decoding module with more precise and enriched detail features, enhancing accuracy during detail restoration. Moreover, it can improve model generalization, allowing it to adapt to

various datasets. However, this architecture increases network complexity and raises the risk of overfitting. Therefore, we implemented the ghost module, which replaces the original convolution operations. It achieves equivalent or superior feature extraction outcomes compared to traditional convolutions using more straightforward arithmetic, reducing the overall number of model coefficients.

The fusion attention downsampling mechanism's GAM integrates the output feature from the preceding encoding module with the input feature map of the same encoding module \mathbf{F}_i . Thus, by performing an element-wise summation on both input features and then using the GAM attention mechanism, rich contextual information \mathbf{F}_G^i is obtained in the new feature map. The expression for this is as follows:

$$\mathbf{F}_G^i = \circ(f_G(\mathbf{F}_D^{i-1} \oplus \mathbf{F}_{i-1})). \quad (4)$$

Concurrently, we use the down ghost module to extract features for the obtained value of \mathbf{F}_G^i . The expression is as follows:

$$\mathbf{F}_{GF}^i = f_{DG}(f'_{DG}(\mathbf{F}_G^i)), \quad (5)$$

where

$$f'_{GB} = \delta(B(\phi_G(\mathbf{F}_G^i))), \quad (6)$$

$$f_{GB} = \delta(B(\phi_G(f'_{GB}(\mathbf{F}_G^i)))), \quad (7)$$

where f_G denotes the GAM attention operation, B denotes batch normalization operation, δ denotes ReLu operation, ϕ_G denotes ghost convolution, f_{GB} denotes the ghost module, $i \in [2, 5]$.

3.3.3 | Semantics and Detail Infusion

Utilizing detailed lower-layer spatial features to refine higher-layer semantic features with greater abstraction can provide the decoding module with more precise features across different scales, thereby improving segmentation accuracy. Therefore, we have included the semantics and detail infusion skip connection module from U-Netv2. The architecture utilizes attention mechanisms to combine information from the five feature maps ($\mathbf{F}_D^1, \mathbf{F}_D^2, \mathbf{F}_D^3, \mathbf{F}_D^4, \mathbf{F}_D^5$) produced by the encoder. A 3×3 convolution is then used to smooth the resulting feature maps. Finally, the Hadamard product is used to integrate all feature mappings, leveraging more semantic information and finer details to enhance the features. The ultimate feature map \mathbf{F}_{DI}^i is acquired, expressed as follows:

$$\mathbf{F}_{DI}^i = f_S(\mathbf{F}_D^i), \quad (8)$$

where f_S denotes the SDI operation, $i \in [1, 5]$

3.3.4 | Decoding Module

While deep features excel in capturing semantic information, they fall short in providing spatial details. On the other hand, shallow features offer abundant spatial information but lack the capacity to convey global semantics. Consequently, by merging

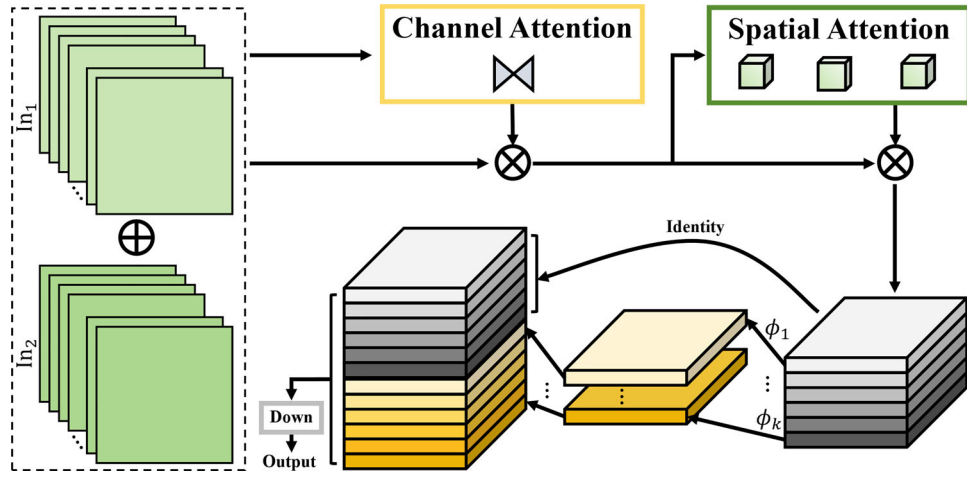


FIGURE 3 | The GAMF module is composed of GAM and down ghost. The fused features serve as input, and the global attention mechanism effectively combines spatial and semantic information. Next, down ghost abstracts the features while preserving the original information. This, in turn, reduces the model complexity.

these two features, the model can improve its ability to comprehend image content and capture precise details, resulting in higher segmentation accuracy. It is composed of a multi-class attention mechanism and an edge gating module (CHG). For low-contrast regions, MAD enhances semantic representation by using global attention, while leveraging edge semantics to suppress noise and refine vascular details. Based on the analysis mentioned above, this study presents a decoding module that utilizes a multi-attention mechanism. Spatial and channel attention mechanisms have been incorporated for layers U1, U2, and U3 to capture a significant amount of high-level and detailed feature information ($\mathbf{F}_D^3, \mathbf{F}_D^4, \mathbf{F}_D^5$). The spatial attention mechanism encodes broader contextual information into local features, while the channel attention mechanism combines inter-channel related features. Thus, merging these two mechanisms can effectively consolidate contextual information. In addition, we included boundary features to aid in multi-level feature fusion.

The process for layer U2 will be detailed due to the structural similarities among layers U1, U2, and U3. For layer U2, we use the multi-attention mechanism to extract pivotal features in \mathbf{F}_D^4 . We also apply an element-wise summation to the boundary features \mathbf{F}_C^3 and \mathbf{F}_{DI}^4 to obtain the combined feature map \mathbf{F}_U^2 . The expression is as follows:

$$\mathbf{F}_{\text{Sum}}^k = f_{\text{Sum}}(f_{\text{PA}}(\mathbf{F}_D^{A,k}), f_{\text{CA}}(\mathbf{F}_D^{A,k}), f_{\text{SA}}(\mathbf{F}_D^{A,k}), \mathbf{F}_{DI}^{A,k}, \mathbf{F}_C^{3,k}), \quad (9)$$

$$\mathbf{F}_U^2 = Up(\mathbf{F}_U^1) \odot \mathbf{F}_{\text{Sum}}^k. \quad (10)$$

The up ghost module is used to perform upsampling operations on the acquired data \mathbf{F}_U^2 , gradually restoring feature details and obtaining the output feature map \mathbf{F}_O^2 , expressed as follows:

$$\mathbf{F}_O^2 = f_{\text{GB}}(f'_{\text{GB}}(\mathbf{F}_U^2)), \quad (11)$$

where

$$f'_{\text{GB}} = \delta(B(\phi_G(\mathbf{F}_U^2))), \quad (12)$$

$$f_{\text{GB}} = \delta(B(\phi_G(f'_{\text{GB}}(\mathbf{F}_U^2)))), \quad (13)$$

where \odot denotes concatenation, f_{Sum} denotes multi-input element-wise summation, f_{PA} denotes the spatial attention mechanism, f_{CA} denotes the channel attention mechanism, f_{SA} denotes the spatial attention mechanism, k denotes the number of layers in the decoder, $k \in [1, 5]$.

The importance of abundant edge detail features in shallow layers for detail restoration is acknowledged. The U4 layer incorporates a context feature extraction module. This module effectively combines surface-level details with in-depth semantic features to maintain crucial semantic context information. The dense atrous convolution module uses convolutions with varied receptive fields to extract features of different abstraction levels. These features are then integrated by the residual multi-kernel pooling module to combine context features from various layers.

The combination with context feature extraction has improved the ability to interact with contextual information in layer U4. The resulting feature map \mathbf{F}_O^4 , further refined for detail, is expressed as follows:

$$\mathbf{F}_U^4 = Up(\mathbf{F}_O^3) \odot [f_R((f_D(\mathbf{F}_D^2))) \oplus \mathbf{F}_{DI}^2], \quad (14)$$

$$\mathbf{F}_O^4 = f_{\text{GB}}(f'_{\text{GB}}(\mathbf{F}_U^4)). \quad (15)$$

Ultimately, to obtain additional information without relying too heavily on deep semantic analysis. This article uses the concept of gated convolution to integrate boundary features across different levels with refined and enriched semantic features to achieve effective feature fusion. as shown in Figure 4. The holistic nested edge detection method is used to acquire edge features \mathbf{F}_H . These features are then concatenated with \mathbf{F}_H , \mathbf{F}_C^0 , and previous features along the channel dimension \mathbf{F}_D^1 , resulting in the concatenated feature map \mathbf{F}_C' , expressed as follows:

$$\mathbf{F}_C' = \mathbf{F}_H \odot \mathbf{F}_C^0 \odot \mathbf{F}_D^1. \quad (16)$$

A new gated convolution module has been defined, the output layer of this module is subject to a Sigmoid operation, which results in a weight scaling factor α between 0 and 1. Using α

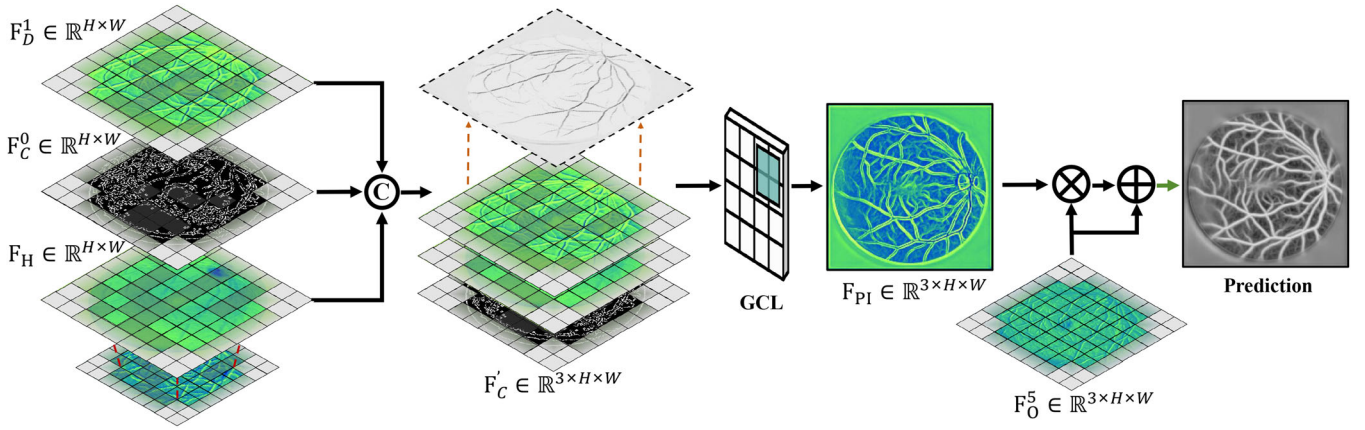


FIGURE 4 | The structural diagram of the CHG module is depicted. Edge information serves as a positional prior, leveraging the gating layer to guide the extraction of low-level features. This design retains essential regional information and mitigates noise interference.

to highlight the key features in F_O^5 , we obtain the final output F_P . It is akin to adding a filter to F'_C , permitting only the features related to semantics to pass through. We treat F'_C as low-level input and effectively control its activation using higher-level activations derived from the decoding phase. This allows a substantial amount of valuable edge information to be preserved. Expression is shown below:

$$\alpha = f_{GC}(F'_C), \quad (17)$$

$$F_P = \phi_1(F_O^5 \otimes (\alpha + 1)), \quad (18)$$

$$f_{GC} = \sigma(B(\phi_1(\delta(\phi_1(B(F'_C)))))), \quad (19)$$

where \otimes denotes the element-wise multiplication operation.

3.4 | Loss Function

In order to account for the potential disequilibrium between vascular and non-vascular distributions within vascular images, a weighted cross-entropy loss function is used in this study:

$$L(y_i, \hat{y}_i) = -\frac{1}{N} \sum_{i=1}^N [w_p \cdot y_i \cdot \log(\hat{y}_i) + w_n \cdot (1 - y_i) \cdot \log(1 - \hat{y}_i)], \quad (20)$$

where N represents the batch size, y_i denotes the true label of the i th sample, with $y_i = 1$ for positive samples and $y_i = 0$ for negative samples. The variable \hat{y}_i represents the probability of the model predicting the i th image as a positive sample, while w_p and w_n are the weights assigned to positive and negative samples, respectively (Algorithm 1).

4 | Experiment Setup

4.1 | Datasets

This study uses three publicly available datasets: DRIVE, STARE, and CHASEDB1 for retinal vessel segmentation experiments to evaluate the effectiveness of our method, as shown in Table 1.

Algorithm 1 | EANet model training process.

- 1: **Input:** Training set F_{lm}^{tra} , Test set F_{lm}^{tes} , Training rounds N , Various parameters
- 2: **Output:** The predictive feature matrix F_{pr}
- 3: Initialize optimum mIoU bmIoU
- 4: Initialize weight S
- 5: Initialize the performance indicator matrix R_M
- 6: **for** $n = 1$ **to** N **do**
- 7: Using the EFE module to acquire the encoded feature matrix F_i (**refer to Equation 1**)
- 8: Using MSE to obtain encoded feature matrices F_{RS} , F_{GF}^i (**refer to Equations 3 and 5**)
- 9: Using SDI to obtain the fused feature matrix F_{DI}^i (**refer to Equation 8**)
- 10: Using MAD to progressively restore detailed features and acquire the predicted feature matrix F_{pr}^n (**refer to Equations 9 to 19**)
- 11: Using the loss function (**refer to Equation 20**) to calculate the loss L_s and obtain current weights S
- 12: Computing performance metrics: Acc, Se, F1, and mIoU (**refer to Equations 21 and 25**)
- 13: **if** bmIoU > mIoU **then**
- 14: Set bestS equal to S
- 15: Document the current performance metrics matrix $R_M = [Acc, Se, F1, mIoU]$
- 16: Set best F_{pr} equal to F_{pr}^n
- 17: **end if**
- 18: Updating various types of parameters and utilizing L_s to update model parameters
- 19: **end for**
- 20: **return** F_{pr} , bestS, R_M

TABLE 1 | Detailed information on three public datasets.

Dataset	Image numbers	Train/Test	Image size
DRIVE	40	20/20	584 × 565
CHASEDB1	28	20/8	999 × 960
STARE	20	10/10	700 × 605

DRIVE: The DRIVE dataset was released in 2004 as part of a Dutch study on diabetic retinopathy screening by [59]. It consists of 40 colour fundus images with resolutions of 584 × 565 and corresponding annotated images. The dataset consists of training and testing datasets, which contain 20 images in total. Annotated images are categorized into vessel annotations and mask images, with two sets of vessel annotations available. Researchers usually consider the set with a greater number of vessel annotations as the gold standard.

CHASEDB1: The CHASEDB1 dataset was released in 2009 for a study on child heart and health in England by [60]. It comprises 28 colour fundus images, each with a resolution of 999 × 960, taken from 14 children and two sets of corresponding vessel annotation images. Our research uses the first set of annotations as the gold standard. No mask images are provided in this dataset.

STARE: The STARE dataset consists of 20 colour fundus images with a resolution of 700 × 605 and two sets of corresponding vascular annotations, introduced in 2000 and developed by [61] using a region-based detection method. The fundus images include 10 with lesions and 10 without. The first set of vessel annotations is the gold standard in our research. No mask images are provided with this dataset.

4.2 | Data Preprocessing

Regarding dataset division, the DRIVE dataset adheres to the official partitioning, with 20 datasets intended to train model while another 20 datasets to test. This study uses a commonly employed dataset-splitting approach because there are no official partitions for the CHASEDB1 and STARE datasets. The CHASEDB1 dataset consists of 20 images for its training dataset and 8 images for its testing dataset. The STARE dataset includes 10 images in the training and test sets, with 5 images of vessels without lesions and 5 images with lesions in each set. The DRIVE dataset includes field of view (FOV) masks, absent in the other two datasets. For datasets without masks, mask images were manually annotated using the Labelme tool. This ensures that each dataset contains two types of annotations.

Given the small number of images in all three datasets, which could potentially lead to overfitting during training, we employed data augmentation techniques to expand the datasets. The translation range is defined as $[-10\%, 10\%]$, the scaling factor varies within $[0.8, 1.1]$, and brightness adjustment is confined to $[-10\%, 10\%]$. These techniques included flipping, rotation, translation, scaling, adding Gaussian noise, and making random brightness adjustments. Concurrently, the input image sizes were standardized to 380 × 380 to align with the network's output size specifications.

4.3 | Evaluation Metrics

Retinal vessel images consist of two types: vessels (foreground) and non-vessels (background). Vessels, which are the focus of detection and segmentation, are identified as the positive class, while non-vessels are designated as the negative class. The results of network segmentation are compared to the true values to determine true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). TP represents the number of pixels accurately identified as the vessel class, while FP refers to the number of pixels mistakenly identified as background. FN is the number of pixels misclassified as the vessel class, and TN is the number of pixels correctly identified as background. This study uses accuracy (Acc), sensitivity (Se), the F1 score, and mean intersection over union (mIoU) to evaluate the effectiveness of EANet. Acc represents the probability of correctly classifying both vessel and background categories, while Se represents the probability of correctly classifying the vessel category. The F1 score is the harmonic mean of precision and sensitivity. The mean intersection over union (mIoU) metric is calculated by averaging the IoU across all categories. The IoU is the ratio of the intersection over the union between the actual labels and predicted outcomes for each category.

$$Acc = \frac{TP + TN}{TP + FN + TN + FP}, \quad (21)$$

$$Se = \frac{TP}{TP + FN}, \quad (22)$$

$$F1 = \frac{2TP}{2TP + FP + FN}, \quad (23)$$

$$IoU = \frac{TP}{TP + FP + FN}, \quad (24)$$

$$mIoU = \frac{1}{M+1} \sum_{i=1}^M IoU_i, \quad (25)$$

where M denotes the class total and IoU_i denotes the IoU value for the i^{th} category.

4.4 | Implementation Details

The EANet implementation presented in this study is built on the Pytorch platform using an NVIDIA RTX 3090 graphics card with 24 GB of memory and a 12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50 GHz. The optimization technique used is the Adam algorithm, which is initialized with a learning rate of 1×10^{-3} , a weight decay of 1.5×10^{-4} , and a batch size of 2. The DRIVE and CHASEDB1 datasets were trained for 40 epochs during our experiments. The STARE dataset was trained for 80 epochs. Additionally, to reduce overfitting, Dropout operations were implemented.

5 | Results and Discussion

The trials performed to evaluate the performance of the proposed model in retinal vessel segmentation tasks are described in the following subsection. Initially, the methodology introduced in this study was comprehensively compared with advanced

approaches across three public retinal vessel datasets, underscoring its effectiveness. Subsequently, ablation studies focused on three critical components of the EANet architecture were carried out to ascertain the distinct effectiveness of each module. In addition, we conducted comparative experiments on the critical parameters of the EFE module to determine its optimal performance. Finally, we compared the iteration speeds of various methods to validate the high efficiency of EANet.

This section describes the experiments carried out on the proposed model to evaluate its performance in retinal vessel segmentation tasks. Initially, the methodology introduced in this study was comprehensively compared with advanced approaches across three public retinal vessel datasets, underscoring its effectiveness. Subsequently, ablation studies focused on three key components of the EANet architecture were carried out to ascertain the distinct effectiveness of each module. Finally, we conducted comparative experiments on the key parameters of the EFE module to determine its optimal performance.

5.1 | Tested Models

The 10 established methods were compared with the 5 variant methods proposed in this study to assess the feasibility of the method described here. All methods adopted identical experimental configurations and training strategies to guarantee a fair comparison. The methods compared are as follows:

- U-Net [15]: The U-Net model enables efficient medical image segmentation with minimal annotated samples.
- R2U-Net [62]: R2U-Net is a segmentation model that utilizes residual structures and recurrent residual convolution modules to enhance feature fusion, resulting in improved segmentation performance.
- AttU-Net [63]: AttU-Net uses attention gate modules to identify different target structures and highlight relevant features while reducing the impact of irrelevant ones.
- SAU-Net [22]: SAU-Net is a variant of the lightweight spatial attention U-Net that adaptively refines input feature maps.
- CS2-Net [64]: CS2-Net utilizes various attention modules to comprehensively integrate the multi-layered information in curve structures.
- FRU-Net [42]: This is a full-resolution network that integrates multi-scale contextual information.
- ConvMixer [65]: ConvMixer uses self-attention mechanisms in conjunction with U-Net for modelling.
- IterMiU-Net [66]: This article proposes a lightweight convolutional model, IterMiU-Net, which combines the Iternet and MiU-Net architectures.
- WS-DMF [67]: The WS-DMF network introduces a novel approach by combining anisotropic awareness convolution with anisotropic enhancement modules.
- IMFF-Net [68]: IMFF-Net is a segmentation network that fuses features at multiple scales.
- EANet_B: The basic network of the model presented in this work. This is a network that integrates local feature fusion with VGG-16, using VGG-16's superior feature extraction capabilities to enhance U-Net's global and local feature extraction.
- EANet_B + EFE: It proposes a network that combines local features with edge feature enhancement, using the EFE module to enhance the network's edge feature detection capability.
- EANet_B + MSE: This network combines local features with multi-scale information exchange, using the MSE module to exploit the abundant tubular features in images and increase the global interactive capacity of the network.
- EANet_B + MAD: This network integrates local features with a multi-attention mechanism. The MAD module is deployed to strengthen the fusion of context information and mitigate the loss of detailed feature information.
- EANet: This method is proposed in this article. EANet introduces a multi-scale network that combines edge features with an attention mechanism. It adeptly incorporates the EFE, MSE, and MAD modules within the EANet_B.

5.1.1 | Visual Comparison

Our approach was visually compared to other methodologies across three public datasets, as shown in Figures 5 and 6. Although U-Net is a standard in segmentation benchmarks, its performance was unsatisfactory. The high prevalence of yellow and red pixels indicates a higher occurrence of false negatives and positives in its segmentation outcomes. R2U-Net adopts the principle of feature fusion through the integration of recursive residual convolution blocks. Figure 6 shows that, compared to U-Net, R2U-Net has fewer red pixels and more yellow pixels. This indicates that the method's ability to detect fine vessels may have been compromised due to the extensive feature fusion operations. As shown in Figure 5, R2U-Net has difficulty accurately segmenting small blood vessels. AttU-Net, which incorporates attention gates, improves the network's focus on targets of varying sizes and shapes, outperforming R2U-Net. However, incomplete capture of spatial information leads to reduced continuity in vessel segmentation. Upon comparison with the ground truth, it is evident that although AttU-Net can detect fine vessels, the segmentation results lack continuity. This indicates a deficiency in effectively capturing and utilizing deep semantic information. By incorporating a spatial attention mechanism, SAU-Net has improved its capability of acquiring spatial features. The segmentation outcomes of this method are more precise than the previously mentioned three methods. However, increased red pixels within its segmentation results indicate a reduced ability to resist interference. CS2-Net uses different attention mechanisms to enhance its ability to capture curvilinear structural features. However, as shown in Figure 5, it cannot accurately segment fine vessels. FRU-Net integrates features from all network layers to improve contextual feature capture. However, it is susceptible to disruptions caused by image blur and complex structures, resulting in increased red pixels in the imagery. ConvMixer, which employs transfer learning principles, has improved the network's ability to resist interference, significantly reducing red

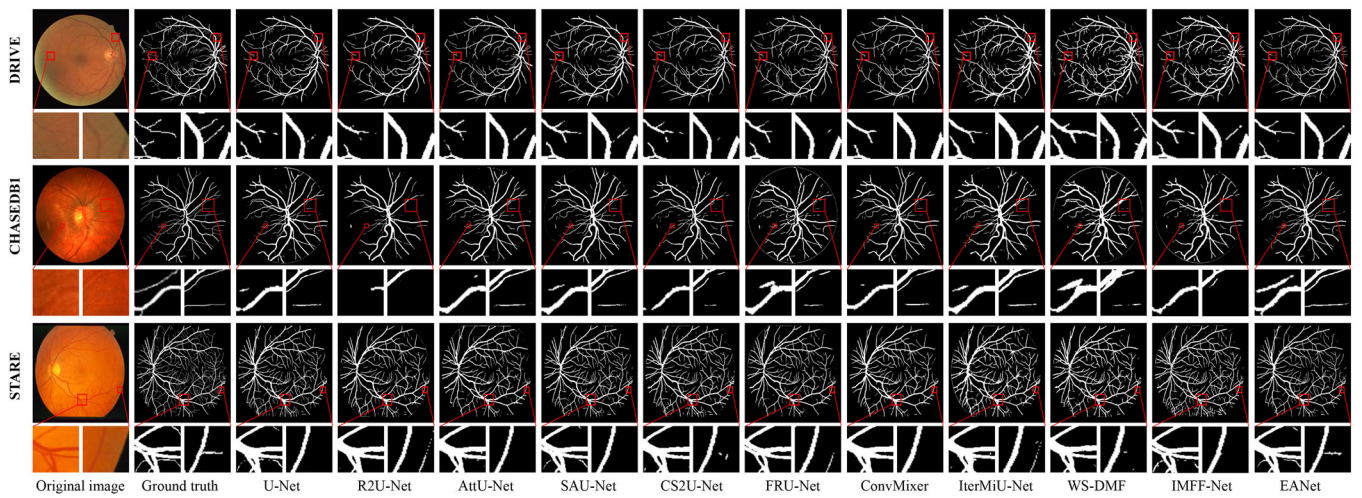


FIGURE 5 | The segmentation outcomes for all comparative methods across the three datasets are as follows: DRIVE (top), CHASEDB1 (middle), and STARE (bottom).

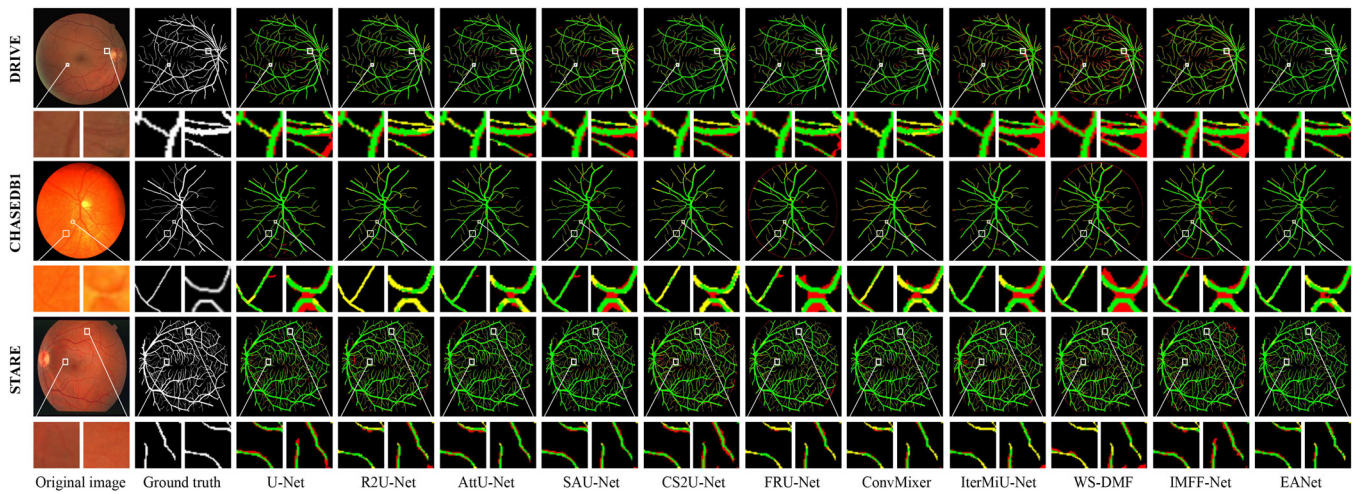


FIGURE 6 | The segmentation outcomes for all comparative methods across the three datasets are as follows: DRIVE (top), CHASEDB1 (middle), and STARE (bottom). In the segmentation results, red indicates false positives, yellow false negatives, and green true positives.

pixels in the images. However, due to structural limitations, the method cannot extract more comprehensive vessel details. IterMiU-Net is a lightweight design that simplifies network complexity, building on the Iternet architecture. However, its adaptability to complex vascular images is limited, demonstrating weak interference resistance. Although WS-DMF can enhance vascular features with anisotropic awareness convolution, its structural limitations hinder its adaptability to complex vascular structures, which reduces the model's generalization capabilities. IMFF-Net is a multi-scale feature fusion network that demonstrates improved fine vessel segmentation capabilities compared to previous methods. However, this approach is still susceptible to interference from background noise.

Our method benefits from the EFE module, which adeptly merges vessels' edge features with multi-scale features, thus augmenting EANet's edge detection and interference resistance capabilities. Figure 5 demonstrates EANet's ability to segment a larger number of fine vessels while effectively suppressing irrelevant

background noise, as shown by the red pixels in Figure 6. The MSE module proposed here enhances capabilities of the model for tubular feature acquisition and promotes interactions within the network context. Figures 5 and 6 demonstrate that EANet captures vascular features with more excellent continuity. Furthermore, our approach achieves a more precise segmentation of fine vessels due to the proposed MAD structure. This architecture minimizes the loss of crucial information in the network and amplifies its capacity to restore detailed features.

5.1.2 | Statistical Evaluation

To evaluate results quantitatively, we used Acc, mIoU, F1, and Se metrics for statistical comparison to assess the efficacy of our proposed EANet. Tables 2–4 present the data indicating that EANet attained the highest scores in Acc, mIoU, and F1 across the DRIVE, CHASEDB1, and STARE datasets. The DRIVE dataset achieved Acc of 95.48%, a mIoU of 82.40%, and an

TABLE 2 | Comparison results of different segmentation methods on the DRIVE dataset.

Dataset	Method	Acc (%)	Se (%)	F1 (%)	mIoU (%)
DRIVE	U-Net [15]	95.18	82.89	81.42	81.64
	R2U-Net [62]	95.12	80.83	80.84	81.20
	AttU-Net [63]	95.25	80.72	81.21	81.54
	SAU-Net [22]	94.96	85.83	81.25	81.38
	CS2-Net [64]	95.26	79.77	81.08	81.45
	FRU-Net [42]	94.98	82.21	80.65	80.98
	ConvMixer [65]	95.14	79.96	80.73	81.14
	IterMiU-Net [66]	95.29	75.68	80.36	80.98
	WS-DMF [67]	94.69	80.99	79.54	80.06
	IMFF-Net [68]	95.08	69.73	78.29	79.46
	EANet	95.48	82.33	82.25	82.40

TABLE 3 | Comparison results of different segmentation methods on the CHASEDB1 dataset.

Dataset	Method	Acc (%)	Se (%)	F1 (%)	mIoU (%)
CHASEDB1	U-Net [15]	94.86	86.77	80.33	80.70
	R2U-Net [62]	95.16	75.78	77.64	79.08
	AttU-Net [63]	95.21	86.31	79.69	80.48
	SAU-Net [22]	94.99	87.44	79.91	80.49
	CS2-Net [64]	95.05	76.91	79.23	80.07
	FRU-Net [42]	94.23	78.44	75.96	77.45
	ConvMixer [65]	95.24	86.93	81.19	81.51
	IterMiU-Net [66]	95.19	73.64	77.08	78.74
	WS-DMF [67]	94.40	82.85	77.64	78.62
	IMFF-Net [68]	94.67	68.81	75.61	77.49
	EANet	95.34	87.72	82.33	82.37

TABLE 4 | Comparison results of different segmentation methods on the STARE dataset.

Dataset	Method	Acc (%)	Se (%)	F1 (%)	mIoU (%)
STARE	U-Net [15]	94.74	81.54	83.08	82.51
	R2U-Net [62]	94.21	73.03	79.62	79.81
	AttU-Net [63]	94.79	80.07	83.06	82.53
	SAU-Net [22]	94.87	81.99	83.28	82.73
	CS2-Net [64]	94.35	78.22	80.83	80.71
	FRU-Net [42]	94.27	77.82	81.02	80.78
	ConvMixer [65]	93.69	74.26	78.57	78.78
	IterMiU-Net [66]	93.51	61.33	74.19	75.90
	WS-DMF [67]	94.34	70.77	79.27	79.66
	IMFF-Net [68]	92.46	53.58	69.08	72.26
	EANet	95.11	79.93	83.56	83.09

TABLE 5 | Ablation studies of proposed EANet on DRIVE dataset.

Dataset	Method	Acc (%)	Se (%)	F1 (%)	mIoU (%)
DRIVE	EANet_B	95.11	83.66	81.34	81.54
	EANet_B+EFE	95.42	81.62	81.95	82.16
	EANet_B+MSE	95.25	84.61	81.95	82.05
	EANet_B+MAD	95.25	82.14	81.49	81.72
	EANet	95.48	82.33	82.25	82.40

F1 score of 82.25%. The CHASEDB1 dataset achieved an Acc of 95.34%, a mIoU of 82.37%, and an F1 score of 87.72%. The STARE dataset recorded the following performance metrics: Acc at 95.11%, mIoU at 83.09%, and F1 score at 83.56%. This suggests that EANet outperforms the most advanced methodologies in terms of comprehensive performance. EANet achieved the highest Se score of 87.72% in the CHASEDB1 dataset. This suggests it has an enhanced capacity to identify fine vessels and edge features, as demonstrated in numerous studies [19, 42]. These results highlight EANet's proficiency in extracting microvascular structures.

SAU-Net's integration of a spatial attention mechanism concurrently boosted its fine vessel extraction capacity, resulting in the highest Se scores in both the DRIVE and STARE datasets. Furthermore, a comparative analysis of performance metrics on CHASEDB1 demonstrates that ConvMixer closely matches our approach in terms of Acc and F1 score. This can be attributed to ConvMixer's capability to combine features across spatial and channel dimensions, allowing the network to capture a more comprehensive feature information. Consequently, it achieved commendable results about Acc and F1 scores.

5.2 | Ablation Studies

Number of ablation experiments were conducted across publicly retinal vessel segmentation datasets to determine the effectiveness of the EANet and the three new modules introduced in this study. The EANet_B model served as the foundational network, with three innovative modules subsequently integrated for ablation analysis. Results are shown in Tables 5–7 for ablation tests for each dataset.

Table 5 shows that the baseline network achieved an Acc of 95.11%, a mIoU of 81.54%, a Se of 83.66%, and an F1 score of 81.34% for the DRIVE dataset. Subsequently, each of the three new modules was sequentially integrated with the baseline network, and their performance was compared to that of the baseline network. The integration of the EFE module initially resulted in improved performance metrics. Specifically, Acc increased by 0.31%, F1 by 0.61% and mIoU by 0.62%. After including the MSE structure in the baseline network, there was an increase of 0.14% in Acc, 0.51% in mIoU, 0.95% in Se, and 0.61% in F1 score. Finally, the Acc, mIoU, and F1 were improved by 0.14%, 0.18%, and 0.15%, respectively, with the incorporation of the MAD structure into the baseline network. In conclusion, the three new modules have greatly enhanced the segmentation capabilities within

TABLE 6 | Ablation studies of proposed EANet on CHASEDB1 dataset.

Dataset	Method	Acc (%)	Se (%)	F1 (%)	mIoU (%)
CHASEDB1	EANet_B	94.72	87.41	80.66	80.83
	EANet_B+EFE	94.79	87.17	80.67	80.81
	EANet_B+MSE	94.83	87.98	81.47	81.45
	EANet_B+MAD	95.15	82.33	80.74	81.15
	EANet	95.34	87.72	82.33	82.37

TABLE 7 | Ablation studies of proposed EANet on STARE dataset.

Dataset	Method	Acc (%)	Se (%)	F1 (%)	mIoU (%)
STARE	EANet_B	94.96	81.99	83.29	82.08
	EANet_B+EFE	94.95	79.64	83.24	82.76
	EANet_B+MSE	95.00	76.84	82.97	82.61
	EANet_B+MAD	94.75	80.96	82.63	82.20
	EANet	95.11	79.93	83.56	83.09

baseline networks. Additionally, EANet effectively integrated these three modules, resulting in optimal overall performance. The segmentation efficacy of the baseline was improved variably by incorporating the three modules in the CHASEDB1 and STARE datasets. Results are shown in Tables 6 and 7. The integration of the EFE module enabled the baseline network to capture more comprehensive feature information, resulting in significant improvements in both Acc and F1 score. Furthermore, the MSE module enhanced the baseline network's ability to delineate minor features, resulting in peak Se. Additionally, the MAD module somewhat improved the baseline network's generalization capability.

5.3 | Key Parameter Analysis

From the above, it is clear that within our proposed EFE, capability for features extraction can be modulated by varying the number of layers dedicated to feature fusion. Therefore, we performed some experiments with this parameter to measure its impact on model performance, implementing different numbers of fusion layers on three publicly available datasets.

Figure 7a illustrates that the EFE module's initial configuration had three fusion layers for the DRIVE dataset. This yielded performance metrics of 94.43% Acc, 86.14% Se, 81.45% F1 score, and 81.18% mIoU. Adjusting the fusion layer count to two resulted in improvements: an increase of 0.64% in Acc, 3.63% in Se, 0.70% in F1, and 0.90% in mIoU. Finally, reducing the number of fusion layers from two to one resulted in an increase in Acc of 0.41%, F1 of 0.10%, and mIoU of 0.32%. Compared to the initial setting of three layers, the performance metrics showed an increase in Acc of 1.05%, F1 of 0.80%, and mIoU of 1.22%. It is important to note that Se is at its peak when two layers are fused in the EFE module, demonstrating improved detection of fine vascular

TABLE 8 | Model performance and computational complexity.

Layers	Acc (%)	FLOPs (G)
1	95.48	316.005
2	95.07	317.872
3	94.43	319.638

features, albeit at the expense of overall network performance. Figure 7b,c shows that the EFE fusion layer count of one achieved optimal results for Acc, F1, and mIoU in the CHASEDB1 and STARE datasets, highlighting the network's superior edge feature extraction capability in this configuration. Furthermore, Table 8 demonstrates that having only one fusion layer not only decreases the computational requirements of the model but also improves its accuracy. Consequently, we've set the number of fusion layers in the EFE module to one, improving EANet's edge feature extraction efficiency and accelerated processing speed.

5.4 | Comparison of Iteration Speed

Thoroughly evaluating the effectiveness of our proposed method, we examined the convergence rates of EANet versus competing methodologies across three datasets. We specifically evaluated the Acc, F1, and mIoU metrics. Figure 8 shows that EANet achieves superior metric values on the DRIVE and STARE datasets at an early speed during the onset of convergence. Furthermore, in these datasets, EANet's performance indicators show smoother convergence trajectories than alternative strategies. This demonstrates that our methodology converges more quickly and exhibits greater stability. For the CHASEDB1 dataset, the convergence curve fluctuations between EANet and ConvMixer are similar, indicating that ConvMixer is equally stable in this dataset. However, the data on convergence demonstrate that EANet achieves higher F1 and mIoU values. In conclusion, the proposed EANet demonstrates accelerated convergence speed and enhanced stability. Its superior performance across various datasets illustrates the model's robust generalization capacity, ensuring unparalleled overall efficacy.

6 | Conclusion

This article introduces a multi-scale network, EANet, that integrates edge features and attention mechanisms for precise segmentation of retinal vessels. The EANet proposed in this

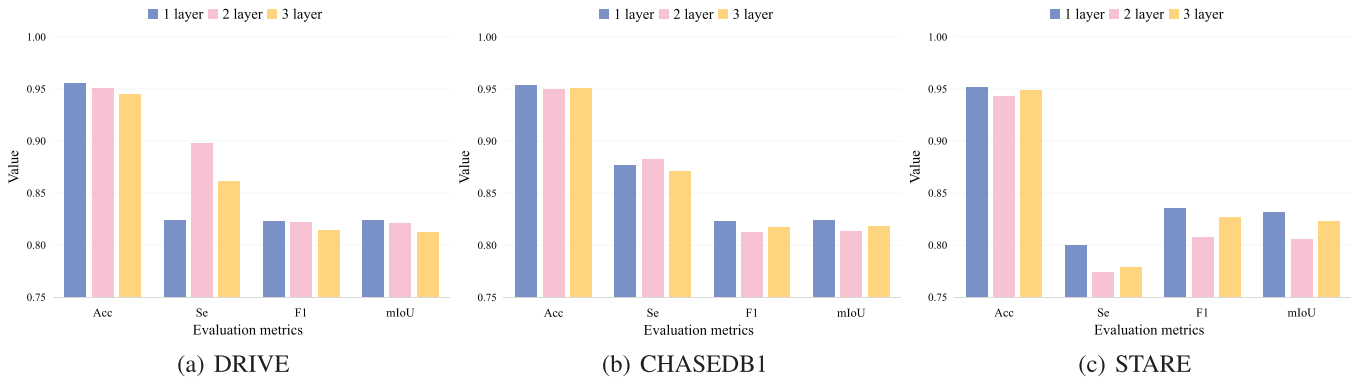


FIGURE 7 | Analysis of key parameters on three datasets.

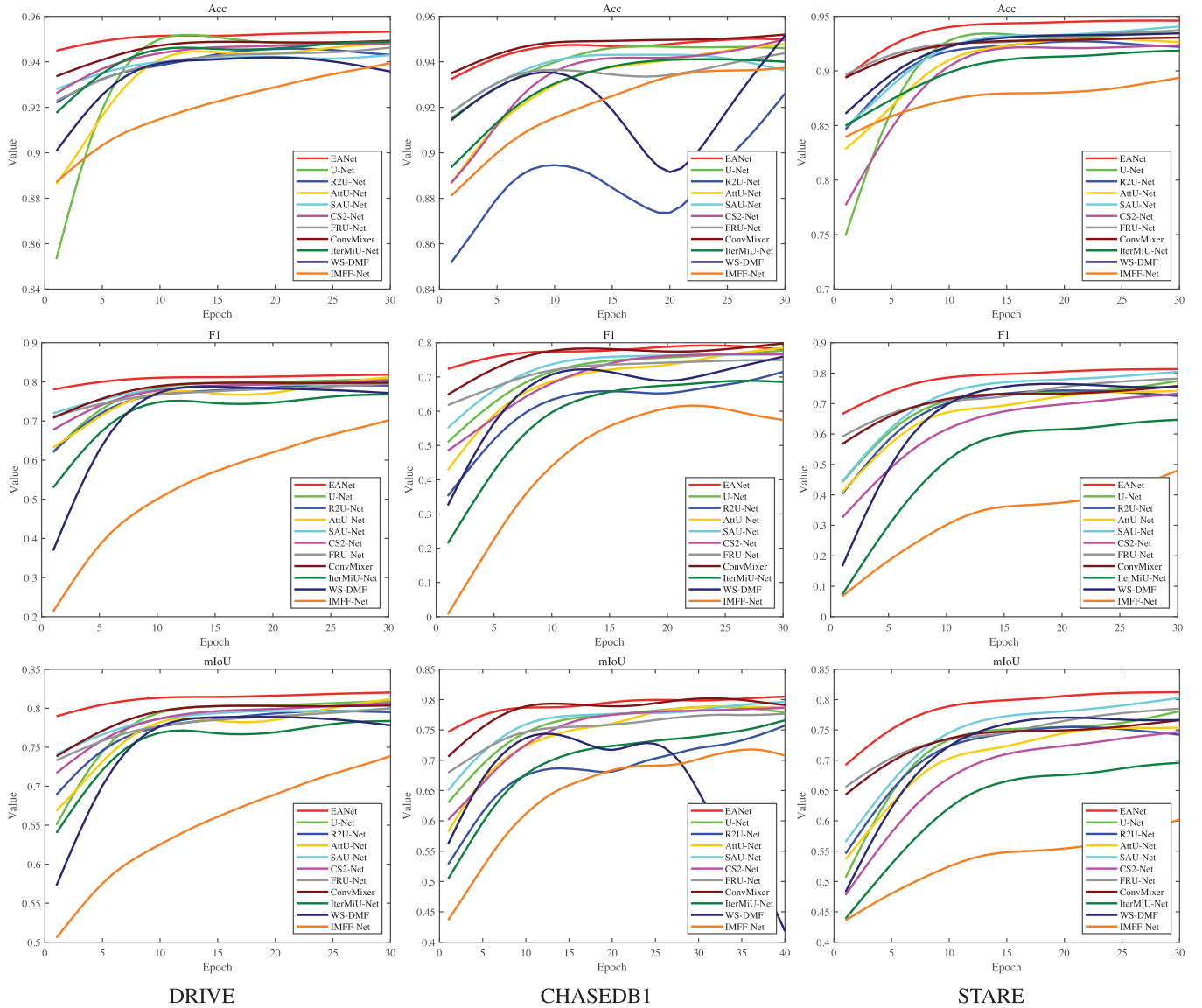


FIGURE 8 | Analysis of iteration speed on three datasets.

study enables the effective integration of shallow edge attributes with multi-level features by efficiently harvesting more precise contextual details. Additionally, EA-Net amplifies the capability to restore detailed characteristics. EA-Net consists of three key modules: EFE, MSE, and MAD. During the feature extraction and

encoding phases, the EFE and MSE modules are introduced. The EFE module uses the ParNet module to expand the network's receptive field, capturing extensive contextual data. Enhanced the network's accuracy in extracting small vessels. In MSE, dynamic serpentine convolution is employed to capture the rich

tubular features of vessels, while GAMF is used to integrate multi-scale contextual information, reducing the loss of detailed information and improving the network's ability to extract multi-scale features. During the encoding phase, we introduced MAD. This module leverages multiple attention mechanisms to improve the decoder's ability to fuse both global and local features, while minimizing background noise and preserving more details. EANet demonstrates superior segmentation performance on three public datasets. However, this study has several limitations: the large number of model parameters results in lower training efficiency and increased resource consumption, the model struggles with complex pathological conditions, and there is a risk of overfitting. Consequently, in future research, we will focus on model simplification, enhance data augmentation techniques to improve the model's generalization capability, and optimize the hyperparameter tuning process to mitigate the risk of overfitting.

Author Contributions

Jiangyi Zhang: conceptualization, software, formal analysis, writing – original draft, writing – review & editing. **Yuxin Tan:** resources, conceptualization, formal analysis, data curation. **Duantengchuan Li:** conceptualization, methodology, writing – original draft, writing – review & editing, project administration, funding acquisition, supervision. **Guanghui Xu:** conceptualization, supervision, validation. **Fuling Zhou:** resources, methodology, writing – review & editing, project administration, funding acquisition.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 82370176), the Key Research and Development Program of Hubei Province (No. 2023BCB025) and the Fundamental Research Funds for the Central Universities (No. 2042024kf0024).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in <https://github.com/dacilab/EANet>.

References

1. A. Ralph Rosenthal, "Ocular Manifestations of Leukemia: A Review," *Ophthalmology* 90, no. 8 (1983): 899–905.
2. R. A. Sawyer, J. B. Selhorst, L. E. Zimmerman, and W. F. Hoyt, "Blindness Caused by Photoreceptor Degeneration as a Remote Effect of Cancer," *American Journal of Ophthalmology* 81, no. 5 (1976): 606–613.
3. T. Sharma, J. Grewal, S. Gupta, and P. I. Murray, "Ophthalmic Manifestations of Acute Leukaemias: The Ophthalmologist's Role," *Eye* 18, no. 7 (2004): 663–672.
4. G. Adamus, "Are Anti-Retinal Autoantibodies a Cause or a Consequence of Retinal Degeneration in Autoimmune Retinopathies?," *Frontiers in Immunology* 9 (2018): 765.
5. S. Sivaprasad, B. Gupta, R. Crosby-Nwaobi, and J. Evans, "Prevalence of Diabetic Retinopathy in Various Ethnic Groups: A Worldwide Perspective," *Survey of Ophthalmology* 57, no. 4 (2012): 347–370.
6. D. S. W. Ting and T. Y. Wong, "Eyeing Cardiovascular Risk Factors," *Nature Biomedical Engineering* 2, no. 3 (2018): 140–141.
7. C. Y. Cheung, D. Xu, C. Y. Cheng, C. Sabanayagam, and T. Y. Wong, "A Deep-Learning System for the Assessment of Cardiovascular Disease Risk via the Measurement of Retinal-Vessel Calibre," *Nature Biomedical Engineering* 5, no. 6 (2020): 498–508.
8. X. Li, Y. Jiang, M. Li, and S. Yin, "Lightweight Attention Convolutional Neural Network for Retinal Vessel Image Segmentation," *IEEE Transactions on Industrial Informatics* 17, no. 3 (2021): 1958–1967.
9. J. Lin, X. Huang, H. Zhou, Y. Wang, and Q. Zhang, "Stimulus-Guided Adaptive Transformer Network for Retinal Blood Vessel Segmentation in Fundus Images," *Medical Image Analysis* 89 (2023): 102929.
10. W. Ding, Y. Sun, J. Huang, et al., "RCAR-UNet: Retinal Vessel Segmentation Network Algorithm Via Novel Rough Attention Mechanism," *Information Sciences* 657 (2024): 120007.
11. L. Câmara Neto, G. L. Ramalho, J. F. Rocha Neto, R. M. Veras, and F. N. Medeiros, "An Unsupervised Coarse-to-Fine Algorithm for Blood Vessel Segmentation in Fundus Images," *Expert Systems With Applications* 78 (2017): 182–192.
12. K. Rezaee, J. Haddadnia, and A. Tashk, "Optimized Clinical Segmentation of Retinal Blood Vessels by Using Combination of Adaptive Filtering, Fuzzy Entropy and Skeletonization," *Applied Soft Computing* 52 (2017): 937–951.
13. S. Wang, Y. Yin, G. Cao, B. Wei, Y. Zheng, and G. Yang, "Hierarchical Retinal Blood Vessel Segmentation Based on Feature and Ensemble Learning," *Neurocomputing* 149 (2015): 708–717.
14. K. López-Linares, N. Aranjuelo, L. Kabongo, et al., "Fully Automatic Detection and Segmentation of Abdominal Aortic Thrombus in Post-Operative CTA Images Using Deep Convolutional Neural Networks," *Medical Image Analysis* 46 (2018): 202–214.
15. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2015), 234–241.
16. Y. Zhang, M. He, Z. Chen, K. Hu, X. Li, and X. Gao, "Bridge-Net: Context-Involved U-Net With Patch-Based Loss Weight Mapping for Retinal Blood Vessel Segmentation," *Expert Systems With Applications* 195 (2022): 116526.
17. M. Khened, V. A. Kollerathu, and G. Krishnamurthi, "Fully Convolutional Multi-Scale Residual DenseNets for Cardiac Segmentation and Automated Cardiac Diagnosis Using Ensemble of Classifiers," *Medical Image Analysis* 51 (2019): 21–45.
18. Z. Zhang, X. Zhang, C. Peng, X. Xue, and J. Sun, "ExFuse: Enhancing Feature Fusion for Semantic Segmentation," in *Proceedings of the European Conference on Computer Vision* (Springer, 2018), 273–288.
19. H. Wu, W. Wang, J. Zhong, B. Lei, Z. Wen, and J. Qin, "SCS-Net: A Scale and Context Sensitive Network for Retinal Vessel Segmentation," *Medical Image Analysis* 70 (2021): 102025.
20. Y. Ye, C. Pan, Y. Wu, S. Wang, and Y. Xia, "MFI-Net: Multiscale Feature Interaction Network for Retinal Vessel Segmentation," *IEEE Journal of Biomedical and Health Informatics* 26, no. 9 (2022): 4551–4562.
21. R. Gu, G. Wang, T. Song, et al., "CA-Net: Comprehensive Attention Convolutional Neural Networks for Explainable Medical Image Segmentation," *IEEE Transactions on Medical Imaging* 40, no. 2 (2021): 699–711.
22. C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, and C. Fan, "SA-UNet: Spatial Attention U-Net for Retinal Vessel Segmentation," in *Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR)* (IEEE, 2021), 1236–1242.
23. Y. Qi, Y. He, X. Qi, Y. Zhang, and G. Yang, "Dynamic Snake Convolution Based on Topological Geometric Constraints for Tubular Structure Segmentation," in *Proceedings of the IEEE International Conference on Computer Vision* (IEEE, 2023), 6047–6056.
24. S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of Blood Vessels in Retinal Images Using Two-Dimensional

- Matched Filters," *IEEE Transactions on Medical Imaging* 8, no. 3 (1989): 263–269.
25. E. Moghimirad, S. Hamid Rezaatofghi, and H. Soltanian-Zadeh, "Retinal Vessel Segmentation Using a Multi-Scale Medialness Function," *Computers in Biology and Medicine* 42, no. 1 (2012): 50–60.
26. S. S. Kar and S. P. Maity, "Blood Vessel Extraction and Optic Disc Removal Using Curvelet Transform and Kernel Fuzzy C-Means," *Computers in Biology and Medicine* 70 (2016): 174–189.
27. D. Marín, A. Aquino, M. E. Gegundez-Arias, and J. M. Bravo, "A New Supervised Method for Blood Vessel Segmentation in Retinal Images by Using Gray-Level and Moment Invariants-Based Features," *IEEE Transactions on Medical Imaging* 30, no. 1 (2011): 146–158.
28. X. You, Q. Peng, Y. Yuan, m. Y. Cheung, and J. Lei, "Segmentation of Retinal Blood Vessels Using the Radial Projection and Semi-Supervised Approach," *Pattern Recognition* 44, no. 10 (2011): 2314–2324.
29. J. Dash and N. Bhoi, "An Unsupervised Approach for Extraction of Blood Vessels from Fundus Images," *Journal of Digital Imaging* 31, no. 6 (2018): 857–868.
30. P. Dollár and C. L. Zitnick, "Structured Forests for Fast Edge Detection," in *Proceedings of the 2013 IEEE International Conference on Computer Vision* (IEEE, 2013), 1841–1848.
31. D. Li, C. Deng, X. Wang, et al., "Joint Inter-Word and Inter-Sentence Multi-Relation Modeling for Summary-Based Recommender System," *Information Processing & Management* 61, no. 3 (2024): 103631.
32. D. Li, F. Shi, X. Wang, C. Zheng, Y. Cai, and B. Li, "Multi-Perspective Knowledge Graph Completion With Global and Interaction Features," *Information Sciences* 666 (2024): 120438.
33. D. Li, T. Xia, J. Wang, et al., "SDFormer: A Shallow-to-Deep Feature Interaction for Knowledge Graph Embedding," *Knowledge-Based Systems* 284 (2024): 111253.
34. W. Zhao, S. Wang, X. Wang, et al., "DADL: Double Asymmetric Distribution Learning for Head Pose Estimation in Wisdom Museum," *Journal of King Saud University - Computer and Information Sciences* 36, no. 1 (2024): 101869.
35. K. Lin, D. Li, Y. Li, S. Chen, and X. Wu, "FHCPL: An Intelligent Fixed-Horizon Constrained Policy Learning System for Risk-Sensitive Industrial Scenario," *IEEE Transactions on Industrial Informatics* 20, no. 4 (2024): 5794–5804.
36. J. Wu, Y. Liu, Y. Zhu, and Z. Li, "Atrous Residual Convolutional Neural Network Based on U-Net for Retinal Vessel Segmentation," *PLoS One* 17, no. 8 (2022): e0273318.
37. F. Dong, D. Wu, C. Guo, S. Zhang, B. Yang, and X. Gong, "CRAUNet: A Cascaded Residual Attention U-Net for Retinal Vessel Segmentation," *Computers in Biology and Medicine* 147 (2022): 105651.
38. Y. Yuan, Y. Zhang, L. Zhu, L. Cai, and Y. Qian, "Exploiting Cross-Scale Attention Transformer and Progressive Edge Refinement for Retinal Vessel Segmentation," *Mathematics* 12, no. 2 (2024): 264.
39. Z. Zhu, Z. Wang, G. Qi, N. Mazur, P. Yang, and Y. Liu, "Brain Tumor Segmentation in MRI With Multi-Modality Spatial Information Enhancement and Boundary Shape Correction," *Pattern Recognition* 153 (2024): 110553.
40. Z. Zhu, X. He, G. Qi, Y. Li, B. Cong, and Y. Liu, "Brain Tumor Segmentation Based on the Fusion of Deep Semantics and Edge Information in Multimodal MRI," *Information Fusion* 91 (2023): 376–387.
41. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation," *IEEE Transactions on Medical Imaging* 39, no. 6 (2020): 1856–1867.
42. W. Liu, H. Yang, T. Tian, et al., "Full-Resolution Network and Dual-Threshold Iteration for Retinal Vessel and Coronary Angiograph Segmentation," *IEEE Journal of Biomedical and Health Informatics* 26, no. 9 (2022): 4623–4634.
43. D. Chen, W. Yang, L. Wang, S. Tan, J. Lin, and W. Bu, "PCAT-UNet: UNet-Like Network Fused Convolution and Transformer for Retinal Vessel Segmentation," *PLoS One* 17, no. 1 (2022): e0262689.
44. Z. Zhu, M. Sun, G. Qi, Y. Li, X. Gao, and Y. Liu, "Sparse Dynamic Volume TransUNet With Multi-Level Edge Fusion for Brain Tumor Segmentation," *Computers in Biology and Medicine* 172 (2024): 108284.
45. Z. Zhu, K. Yu, G. Qi, et al., "Lightweight Medical Image Segmentation Network With Multi-Scale Feature-Guided Fusion," *Computers in Biology and Medicine* 182 (2024): 109204.
46. Q. Sun, M. Dai, Z. Lan, et al., "UCR-Net: U-Shaped Context Residual Network for Medical Image Segmentation," *Computers in Biology and Medicine* 151 (2022): 106203.
47. X. Deng and J. Ye, "A Retinal Blood Vessel Segmentation Based on Improved D-MNet and Pulse-Coupled Neural Network," *Biomedical Signal Processing and Control* 73 (2022): 103467.
48. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv:1409.1556* (2014).
49. J. Canny, "A Computational Approach to Edge Detection," in *Readings in Computer Vision*, ed. M. A. Fischler and O. Firschein (Morgan Kaufmann, 1987), 184–203.
50. A. Goyal, A. Bochkovskiy, J. Deng, and V. Koltun, "Non-Deep Networks," in *Advances in Neural Information Processing Systems*, Vol. 35, ed. S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Curran Associates, 2022), 6789–6801.
51. Y. Liu, Z. Shao, and N. Hoffmann, "Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions," *arXiv:2112.05561* (2021).
52. K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More Features From Cheap Operations," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2020), 1577–1586.
53. Y. Peng, M. Sonka, and D. Z. Chen, "U-Net v2: Rethinking the Skip Connections of U-Net for Medical Image Segmentation," *arXiv:2311.17791* (2024).
54. Z. Gu, J. Cheng, H. Fu, et al., "CE-Net: Context Encoder Network for 2D Medical Image Segmentation," *IEEE Transactions on Medical Imaging* 38, no. 10 (2019): 2281–2292.
55. T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, "Gated-SCNN: Gated Shape CNNs for Semantic Segmentation," in *Proceeding of the 2019 IEEE International Conference on Computer Vision* (2019), 5228–5237.
56. S. Xie and Z. Tu, "Holistically-Nested Edge Detection," in *Proceeding of the 2015 IEEE International Conference on Computer Vision (ICCV)* (IEEE, 2015), 1395–1403.
57. J. Dai, H. Qi, Y. Xiong, et al., "Deformable Convolutional Networks," in *Proceeding of the 2017 IEEE International Conference on Computer Vision (ICCV)* (IEEE, 2017), 764–773.
58. L. Niu, X. Chen, and S. Guo, "Bézier Curves Meet Deep Learning: A Novel Pretraining Method for Improved and Generalized Retinal Vessel Segmentation," in *Proceeding of the 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (IEEE, 2023), 2159–2164.
59. J. Staal, M. Abramoff, M. Niemeijer, M. Viergever, and v. B. Ginneken, "Ridge-Based Vessel Segmentation in Color Images of the Retina," *IEEE Transactions on Medical Imaging* 23, no. 4 (2004): 501–509.
60. C. G. Owen, A. R. Rudnicka, R. Mullen, et al., "Measuring Retinal Vessel Tortuosity in 10-Year-Old Children: Validation of the Computer-Assisted Image Analysis of the Retina (CAIAR) Program," *Investigative Ophthalmology & Visual Science* 50, no. 5 (2009): 2004–2010.
61. A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating Blood Vessels in Retinal Images by Piecewise Threshold Probing of a Matched Filter Response," *IEEE Transactions on Medical Imaging* 19, no. 3 (2000): 203–210.

62. M. Z. Alom, C. Yakopcic, T. M. Taha, and V. K. Asari, "Nuclei Segmentation With Recurrent Residual Convolutional Neural Networks Based U-Net (R2U-Net)," in *Proceedings of NAECON 2018 - IEEE National Aerospace and Electronics Conference* (IEEE, 2018), 228–233.
63. J. Schlemper, O. Oktay, M. Schaap, et al., "Attention Gated Networks: Learning to Leverage Salient Regions in Medical Images," *Medical Image Analysis* 53 (2019): 197–207.
64. L. Mou, Y. Zhao, H. Fu, et al., "CS2-Net: Deep Learning Segmentation of Curvilinear Structures in Medical Imaging," *Medical Image Analysis* 67 (2021): 101874.
65. A. Solano, K. N. Dietrich, M. Martínez-Sober, R. Barranquero-Cardenosa, J. Vila-Tomás, and P. Hernández-Cámara, "Deep Learning Architectures for Diagnosis of Diabetic Retinopathy," *Applied Sciences* 13, no. 7 (2023): 4445.
66. A. Kumar, R. K. Agrawal, and L. Joseph, "IterMiUnet: A Lightweight Architecture for Automatic Blood Vessel Segmentation," *Multimedia Tools and Applications* 82, no. 28 (2023): 43207–43231.
67. Y. Tan, K. F. Yang, S. X. Zhao, J. Wang, L. Liu, and Y. J. Li, "Deep Matched Filtering for Retinal Vessel Segmentation," *Knowledge-Based Systems* 283 (2024): 111185.
68. M. Liu, Y. Wang, L. Wang, S. Hu, X. Wang, and Q. Ge, "IMFF-Net: An Integrated Multi-Scale Feature Fusion Network for Accurate Retinal Vessel Segmentation From Fundus Images," *Biomedical Signal Processing and Control* 91 (2024): 105980.