# FMA-WGAN: A Frequency-Aware Multi-Scale Adversarial Network for Retinal Vessel Segmentation

Xiaoyu Wang, Qichen Wang, Yiying Wang and Qing Ai*

***Abstract*—Retinal vessel segmentation plays a vital role in fundus medical image analysis and offers essential diagnostic insights into conditions such as diabetic and hypertensive retinopathy. However, the presence of complex vascular structures—such as tortuous microvessels—combined with low-contrast backgrounds, presents significant challenges for automated segmentation techniques. To overcome these limitations, we propose a Frequency-aware Multi-scale Attention Wasserstein Generative Adversarial Network (FMA-WGAN) to achieve high-precision retinal vessel segmentation. Our method leverages an Attention Gate-enhanced U-Net as the generator backbone and incorporates multi-scale input features to improve sensitivity to delicate vessel structures. In addition, a spectral-normalized WGAN-GP discriminator is adopted to strengthen the model's ability to distinguish fine-grained vascular features. The proposed loss function strategically combines Wasserstein adversarial loss, focal frequency loss, and L1 loss, harnessing their complementary strengths to preserve global vessel continuity while improving boundary accuracy for microvessels. The proposed method was evaluated on three publicly available datasets: DRIVE, STARE, and CHASE_DB1. Experimental results show that our model achieved the highest accuracy on all three benchmarks, recording values of 0.9652, 0.9743, and 0.9738, respectively. Moreover, the model demonstrated strong performance in terms of specificity, sensitivity, and area under the ROC curve (AUC) across most datasets. Overall, FMA-WGAN consistently outperforms existing state-of-the-art approaches in retinal vessel segmentation tasks.**

***Index Terms*—Retinal vessel segmentation; Generative adversarial network ; Frequency-aware ; Multi-scale ; Deep learning**

## I. Introduction

MORPHOLOGICAL, alterations in retinal vasculature are essential diagnostic biomarkers for a variety of fundus-related diseases, including diabetic retinopathy, glaucoma, and age-related macular degeneration, all of which are characterized by changes in vessel diameter and tortuosity [1], [2]. However, manual annotation of retinal vessels is both time-consuming and susceptible to inter-observer variability, primarily due to the low contrast, anatomical complexity, and dense branching patterns of retinal images. Automated retinal vessel segmentation techniques not only reduce the workload of ophthalmologists but also enhance diagnostic consistency and operational efficiency. Therefore, the development of robust and highly accurate vessel segmentation algorithms holds considerable clinical significance.

Early retinal vessel segmentation methods primarily depended on handcrafted features and conventional image processing techniques [3], such as matched filtering, thresholding, morphological operations, and classical machine learning classifiers. While these methods achieved reasonable performance in specific scenarios, they generally involved intricate preprocessing pipelines, were highly sensitive to noise and image variability, failed to consistently capture fine vascular structures, and often resulted in fragmented segmentation outputs. With the advent of deep learning, convolutional neural networks (CNNs) have demonstrated remarkable capabilities in semantic segmentation tasks [4]. As a result, a wide range of CNN-based approaches have been proposed for retinal vessel segmentation, including Fully Convolutional Networks (FCNs), U-shaped Networks (U-Net), and their numerous improved variants. Among these architectures, U-Net has emerged as the dominant framework owing to its symmetrical encoder-decoder structure and skip connections, which facilitate effective integration of multi-scale contextual information. Various studies have further refined the U-Net architecture by incorporating residual connections, dense blocks, and attention mechanisms to enhance vessel feature representation, thereby substantially improving segmentation accuracy and recall.

Despite the evident superiority of deep learning-based methods over traditional approaches, several key challenges remain unresolved. Notably, fine vessels are often missed due to foreground-background class imbalance, and noise or pathological artifacts are frequently misclassified as vessels, resulting in increased false positives [5]. Statistically, microvessels—defined as vessels with widths less than three pixels—account for merely 30% of the total vessel pixels [6]. This class imbalance often leads conventional CNNs to focus disproportionately on larger vessels while overlooking sparse microvascular structures, thereby resulting in incomplete segmentation of fine vessels. Therefore, improving the detection and structural continuity of microvessels, without compromising overall segmentation performance, remains a critical challenge and central research objective in retinal vessel segmentation.

To address these challenges, this study proposes a Frequency-aware Multi-scale Attention Wasserstein Generative Adversarial Network (FMA-WGAN), integrating

Manuscript received June 18, 2025; revised August 23, 2025.

The research work was supported in part by the Basic Research Project of Education Department of Liaoning Province in China (JYTMS20230929).

Xiaoyu Wang is a Postgraduate of University of Science and Technology Liaoning, Anshan, Liaoning, China (e-mail: xiaoyu925910@163.com).

Qichen Wang is a Postgraduate of University of Science and Technology Liaoning, Anshan, Liaoning, China (e-mail: qichen031017@163.com).

Yiying Wang is a Postgraduate of University of Science and Technology Liaoning, Anshan, Liaoning, China (e-mail: yiyingWang2024@126.com).

Qing Ai* is a Professor of University of Science and Technology Liaoning, Anshan, Liaoning, China (corresponding author to provide phone: +086-155-0492-8887; e-mail: lyaiqing@126.com).

frequency domain information and a multi-scale attention mechanism tailored for retinal vessel segmentation. The primary contributions of this work are summarized as follows:

(1) Multi-scale attention feature fusion: By integrating multi-scale input branches with an attention mechanism into the Attention Gate U-Net architecture, the model achieves synergistic modeling and adaptive fusion of global and local scale features. This strategy substantially enhances the network's responsiveness to fine-grained vascular patterns and its ability to continuously capture delicate vessel structures, thereby improving the accuracy and robustness of vascular segmentation.

(2) Adversarial discriminator constraint on details: Within the WGAN-GP [7] framework, the discriminator enforces dual-level constraints on the generated segmentation results, simultaneously regulating global distribution consistency and local structural fidelity. This mechanism markedly enhances the realism and continuity of vascular details, ensuring structurally coherent reconstructions and facilitating the accurate delineation of fine vessel networks.

(3) Complementary loss integration from frequency and spatial domains: We combine focal frequency loss and L1 loss to leverage their complementary strengths. Specifically, focal frequency loss promotes alignment between the spectral characteristics of generated results and actual vascular structures, capturing global morphology and intricate textures, whereas L1 loss directly enhances the regional consistency between segmentation results and ground truth. Together, these losses ensure both global structural fidelity and precise local segmentation accuracy.

Comprehensive experimental evaluations performed on three widely used benchmark datasets, namely DRIVE, STARE, and CHASE_DB1, demonstrate that our method markedly surpasses current cutting-edge techniques, delivering superior quantitative performance and enhancing the reconstruction of fine vascular details. These results confirm the efficacy of the proposed methodological strategies.

## II. Related work

### A. Early methods

Over the past few decades, numerous retinal vessel segmentation methods have been developed, primarily categorized into unsupervised and supervised approaches. Unsupervised methods do not require manually annotated ground-truth vessel data and mainly utilize intrinsic low-level image features, such as brightness, texture, edges, and shapes, thereby eliminating the necessity of a training process. Chaudhuri et al. [8] were the first to apply Gaussian-based matched filtering for retinal vessel segmentation, encoding inherent fundus image features to achieve automated segmentation. Following this pioneering work, traditional digital image processing-based unsupervised segmentation became a predominant research methodology in early retinal vessel analysis. Kumar et al. [9] employed matched filters combined with Laplacian-of-Gaussian kernels to detect vessel-like structures in retinal images. Li et al. [10] proposed a multi-scale matched filter multiplication approach, effectively enhancing vessel responses across multiple scales while suppressing background noise, thus achieving efficient extraction and accurate width estimation of retinal vessels. Jiang et al. [11] introduced a hybrid method combining morphological global threshold segmentation and centerline detection, significantly reducing computational complexity and enabling simultaneous processing of multiple independent segmentation tasks. Zhao et al. [12] proposed an active contour model incorporating infinite perimeter regularization and hybrid region information, finely delineating vessel boundaries through integrated image intensity and local phase-enhancement features. Nguyen et al. [13] developed an unsupervised method using a linear combination of multi-scale line detectors, effectively segmenting central reflex vessels and closely adjacent structures by fusing responses from different scales. Fraz et al. [14] presented a feature vector that integrated gradient field orientation analysis, morphological features, linear responses, and Gabor filter responses, subsequently classified using Bagging and Boosting decision tree ensembles, thus substantially enhancing segmentation accuracy. Although traditional unsupervised methods are advantageous in their simplicity and annotation-free implementation, their sensitivity to image quality, lack of contextual awareness, poor microvessel segmentation capability, and heavy reliance on manual parameter tuning considerably limit their accuracy and robustness in segmenting complex or low-quality fundus images.

### B. Supervised vessel segmentation methods

In recent years, deep learning–driven image segmentation has witnessed remarkable advancements, delivering outstanding performance in medical imaging tasks such as retinal vessel segmentation. The majority of these methods are based on convolutional neural networks (CNNs). Since the introduction of the U-Net architecture by Ronneberger et al. [15]in 2015, which features an encoder–decoder structure with skip connections, researchers have proposed numerous improvements incorporating dense connections [16], attention mechanisms [17], and expansion modules [18]. These enhancements have led to a variety of U-Net derivatives that excel in segmenting retinal vasculature. For instance, Cameron et al. [19]proposed Super U-Net by adding dynamic receptive field modules and fused upsampling strategies, significantly boosting performance across multiple modalities. Zhang et al. [20]developed MCSE-U-Net by introducing multi-convolution modules and squeeze-and-excitation blocks, while employing YCbCr color space transformation to improve vessel visibility. Luo et al. [21]presented PA-Net, combining CNNs with lightweight parallel Transformers to reduce data and computational demands. Xie et al. [22]introduced ARSA-UNet, which incorporates structure-adaptive convolutions, dilated residual paths, feature filtering, and a deep supervision mechanism to enhance multi-scale learning and mitigate network degradation.Nevertheless, these methods still face difficulties in segmenting intricate microvascular structures, particularly under low contrast, where failure to detect edge vessels can lead to segmentation gaps or omissions.

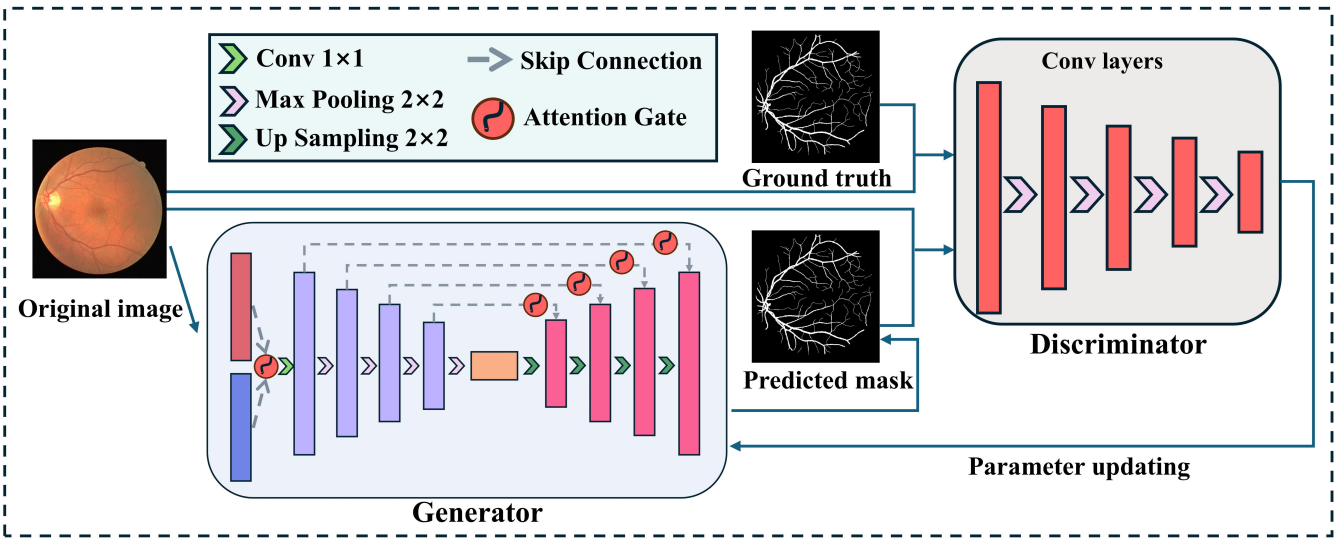


Fig. 1. Overall architecture of the proposed FMA-WGAN.

### C. Visual transformer

Since its introduction by Vaswani et al. in 2017 [23], the Transformer architecture has rapidly become a cornerstone in deep learning, achieving outstanding success particularly in natural language processing. In recent years, its application scope has expanded into computer vision, giving rise to influential models such as the Vision Transformer (ViT) [24]and Swin Transformer [25]. Compared to traditional convolutional neural networks (CNNs), Transformers possess a natural advantage in modeling long-range dependencies, making them highly effective for complex visual tasks. Dosovitskiy et al. introduced ViT by dividing images into fixed-size patches, applying linear projection, and leveraging self-attention mechanisms to capture global semantic features, achieving superior classification performance over CNNs [24]. Subsequently, Liu et al. proposed the Swin Transformer, which introduces window-based self-attention and a hierarchical representation to improve efficiency and scalability, setting new performance benchmarks in image classification, object detection, and semantic segmentation [25]. In the field of medical image segmentation, Vision Transformers have attracted considerable attention. Xu et al. proposed TransUNet, a hybrid architecture that combines the local feature modeling capabilities of CNNs with the global context learning of Transformers, achieving state-of-the-art performance across multiple segmentation tasks [26]. Wang et al. further introduced Swin-UNet, which utilizes a hierarchical Transformer backbone and skip connections to effectively extract both global and fine-grained features from medical images, substantially enhancing segmentation accuracy and generalization [27]. Recently, Vision Transformers have also shown great promise in retinal vessel segmentation. Lv et al. developed TCDDU-Net, which integrates CNN and Transformer structures via selectively dense-connected Swin Transformer blocks, resulting in improved segmentation accuracy and boundary detection [28]. Li et al. designed GViT-RSNet, which combines graph convolutional attention with multi-scale Vision Transformers to further enhance representation capabilities for complex vascular structures [29]. Despite these advances, Transformer-based models are often computationally intensive, require large-scale datasets, and are sensitive to hyperparameter settings—factors that can limit their effectiveness in retinal image analysis, where training data is typically scarce.

## III. Proposed Method

In this section, we present a detailed account of the proposed method. Section 3.1 outlines the overall structure of FMA-WGAN. Section 3.2 explains the designs of the generator and discriminator. Finally, Section 3.3 presents the loss functions that guide the training process.

### A. Overall architecture

As illustrated in Figure 1, the proposed FMA-WGAN is established upon the classical generative adversarial network (GAN) [30] framework, where iterative adversarial learning is leveraged to model the complex and highly nonlinear mapping between input images and target outputs. This adversarial mechanism enables the generated segmentation maps to progressively approximate the statistical distribution and structural characteristics of the ground truth. Within this framework, the generator is optimized to produce segmentation outputs that exhibit both global distributional consistency and local structural fidelity, rendering them increasingly indistinguishable from real annotations. Conversely, the discriminator is trained to enhance its capability of differentiating authentic samples from synthesized ones, thereby providing continuous feedback to guide the generator toward improved realism and coherence. When the discriminator fails to reliably discriminate between generated predictions and ground-truth labels, the synthetic outputs can be considered perceptually realistic and structurally consistent.

It is important to note that traditional generative adversarial networks (GANs) generate images based solely on random noise inputs, which inherently lack explicit supervisory signals. This absence of direct guidance makes it challenging to precisely control the content of generated outputs—an issue that becomes particularly critical in

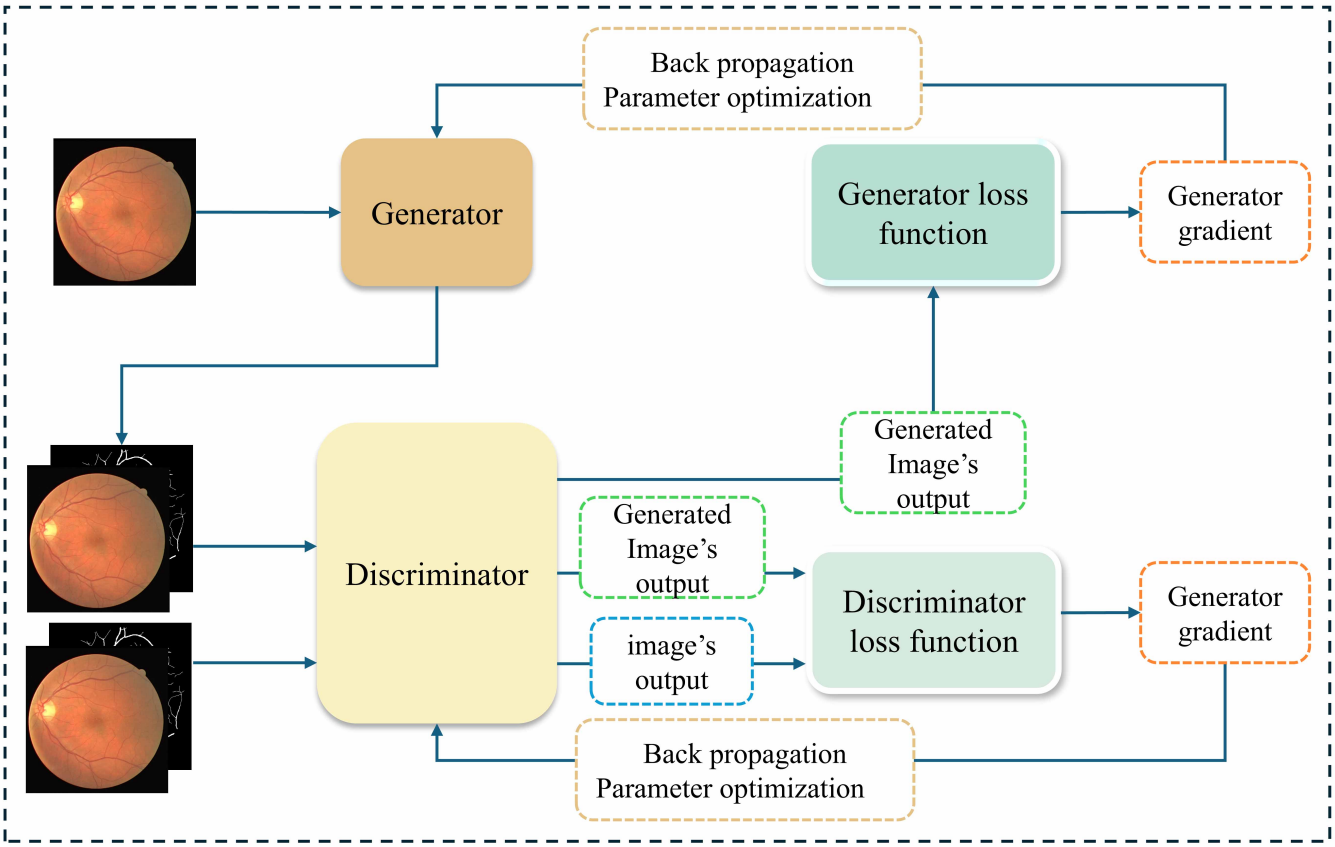


Fig. 2. Training workflow of the proposed FMA-WGAN.

medical image segmentation tasks, where accurate pixel-level correspondence is essential. To overcome this limitation, we extend the standard GAN architecture into a conditional GAN (cGAN) [31] framework. In this configuration, the discriminator is conditioned not only on the generated image but also on additional contextual information—in this case, the corresponding retinal segmentation label map. Moreover, conventional GANs often suffer from gradient instability during training, which can impede convergence and result in suboptimal performance. To address this problem, we employ the Wasserstein GAN with Gradient Penalty (WGAN-GP) as the backbone of FMA-WGAN. WGAN-GP applies the Wasserstein distance—commonly referred to as the Earth Mover's Distance—as a more reliable and informative criterion to evaluate the discrepancy between real and generated distributions. The loss function of WGAN-GP can be expressed as:

$$L_D = \mathbb{E}_{(x,\tilde{y})\sim p_g}\left[D(x,\tilde{y})\right] - \mathbb{E}_{(x,y)\sim p_r}\left[D(x,y)\right] + \lambda_{\text{gp}}\,\mathbb{E}_{\hat{y}\sim p_{\hat{y}}}\left(\left\|\nabla_{\hat{y}} D(x,\hat{y})\right\|_2 - 1\right)^2 \quad (1)$$

Let $(x, y)$ represent a sample pair drawn from the real data distribution $p_r$, and $(x, \tilde{y})$ denote a fake sample pair where $\tilde{y} = G(x)$ is generated from the distribution $p_g$. To enforce the Lipschitz constraint in WGAN-GP, a set of interpolated samples $\hat{y}$ is introduced between real and generated samples. These interpolations are sampled from a distribution $p_{\hat{y}}$, and are computed as $\hat{y} = \epsilon y + (1-\epsilon)\tilde{y}$, where $\epsilon \sim \mathcal{U}(0,1)$. The gradient penalty is applied based on $\hat{y}$, encouraging the norm of the discriminator's gradient with respect to $\hat{y}$ to be close to 1. This strategy avoids the limitations of weight clipping in the original WGAN and leads to improved training stability and higher-quality generation.

The overall architecture and workflow of FMA-WGAN are depicted in Figure 2. The original retinal image is first processed by the generator, which outputs a predicted segmentation mask. This prediction, along with the ground truth mask, is respectively paired with the original image to form two input combinations for the discriminator. The discriminator evaluates these inputs by assigning scores to the predicted segmentation masks and optimizing a loss function composed of the Wasserstein distance and a gradient penalty. Through this process, the discriminator learns to effectively distinguish between real and generated samples. Its parameters are iteratively updated via gradients derived from the loss function using an optimizer. Meanwhile, the generator computes its own loss based on the feedback received from the discriminator. Gradient signals from the discriminator's output are propagated backward through the generator network, enabling the adjustment of its parameters. This adversarial training scheme gradually guides the generator toward producing segmentation masks that more closely approximate the ground truth annotations.

### *B. Generator and discriminator*

The generator in FMA-WGAN is based on an enhanced U-Net architecture, incorporating a multi-scale input feature fusion strategy to improve the segmentation of fine retinal vessels through the integration of attention mechanisms and scale-aware representation. As illustrated in Figure 3, the generator comprises three core components: an encoder, a decoder, and skip connections. The encoder follows the standard U-Net configuration, consisting of several

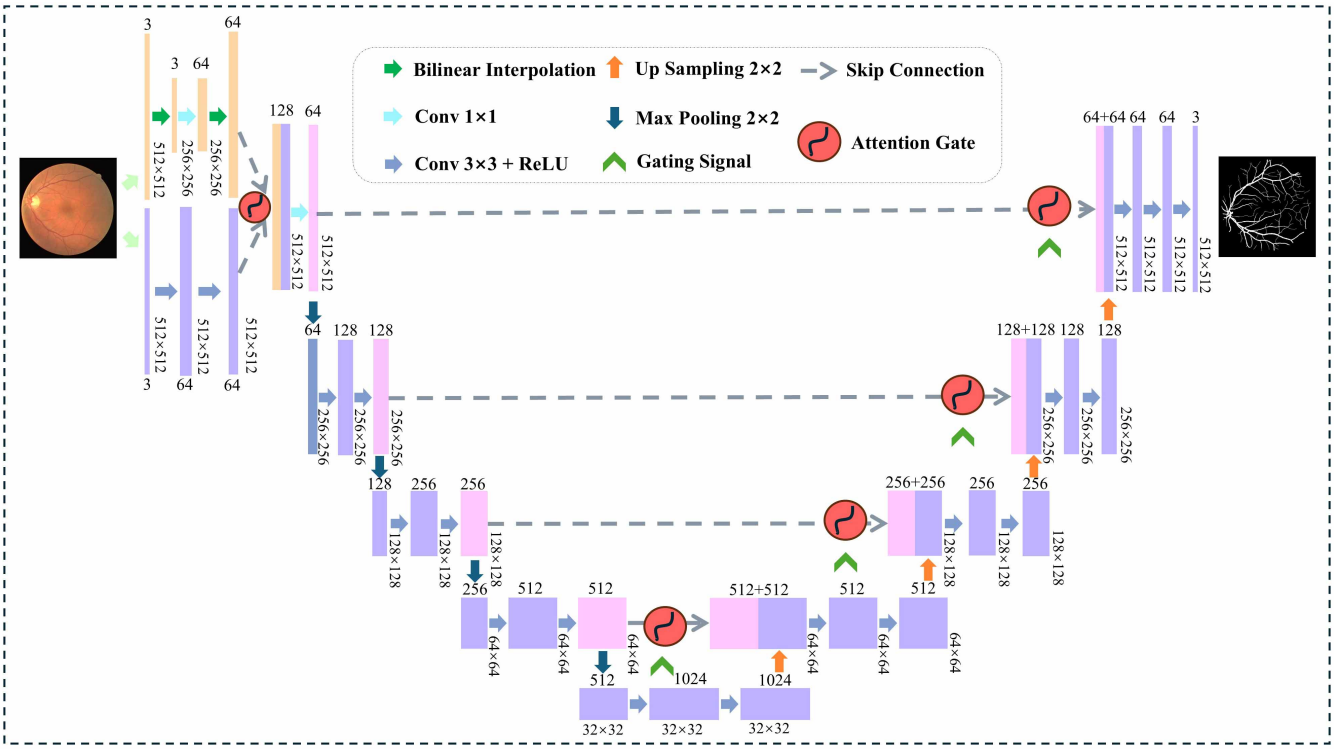


Fig. 3. Generator used in the proposed AA-WGAN.

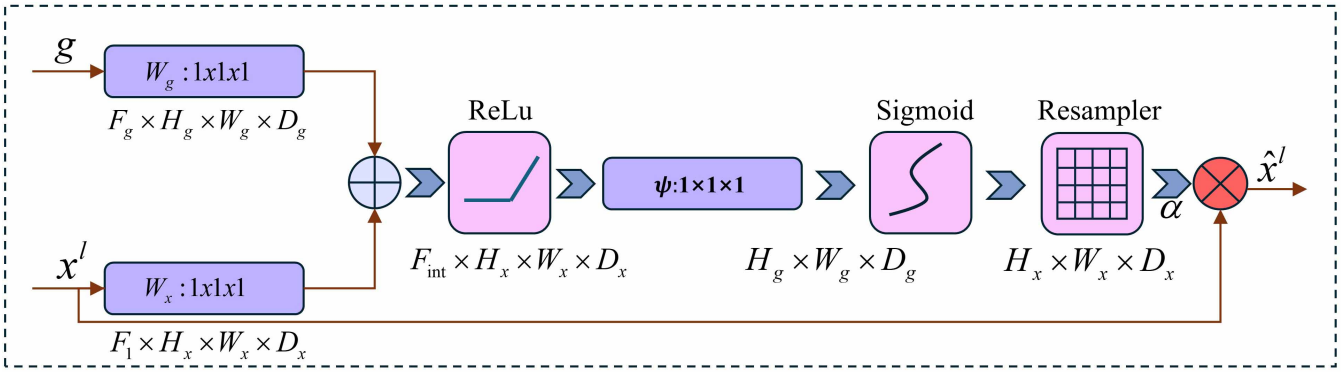


Fig. 4. Attention gate used in the proposed generator.

convolutional and downsampling layers. To augment the model's capacity to represent vessels at multiple scales, we introduce a multi-scale input fusion module. This module allows the generator to capture vascular features at both the original and a lower-resolution scale, thereby extracting complementary global and local information. After convolutional processing, the low-resolution features are upsampled using bilinear interpolation and concatenated with the high-resolution features from the initial encoder layer along the channel dimension. This architecture enables the model to effectively capture retinal vessel structures across different spatial scales and enhances its ability to delineate fine and complex vessel branches, leading to improved generalization across varying anatomical patterns. The decoder resembles the conventional U-Net structure but is augmented with attention gate modules preceding each encoder–decoder fusion step, as depicted in Figure 4. Overall, the generator design ensures the effective transmission of informative features and significantly enhances the model's ability to represent vessels at multiple scales, particularly microvasculature. In this study, the discriminator $D$ in FMA-WGAN is designed to differentiate between the generated segmentation masks and the corresponding ground truth masks. As shown in Figure 5, we employ a fully convolutional PatchGAN-style [32] discriminator and integrate the Wasserstein GAN with Gradient Penalty (WGAN-GP) framework to promote stable adversarial training. The discriminator assigns scores to each patch within the input and aggregates these scores by computing the mean across the patch map, effectively approximating the Wasserstein distance between the real and generated distributions. This mechanism provides a continuous and differentiable signal to guide the generator in producing segmentation masks that closely resemble the real ones, particularly in terms of structural detail and fine-grained accuracy.

### *C. Fourier Frequency Domain Loss*

The baseline loss function adopted in our model is the L1 loss, or mean absolute error, which enforces

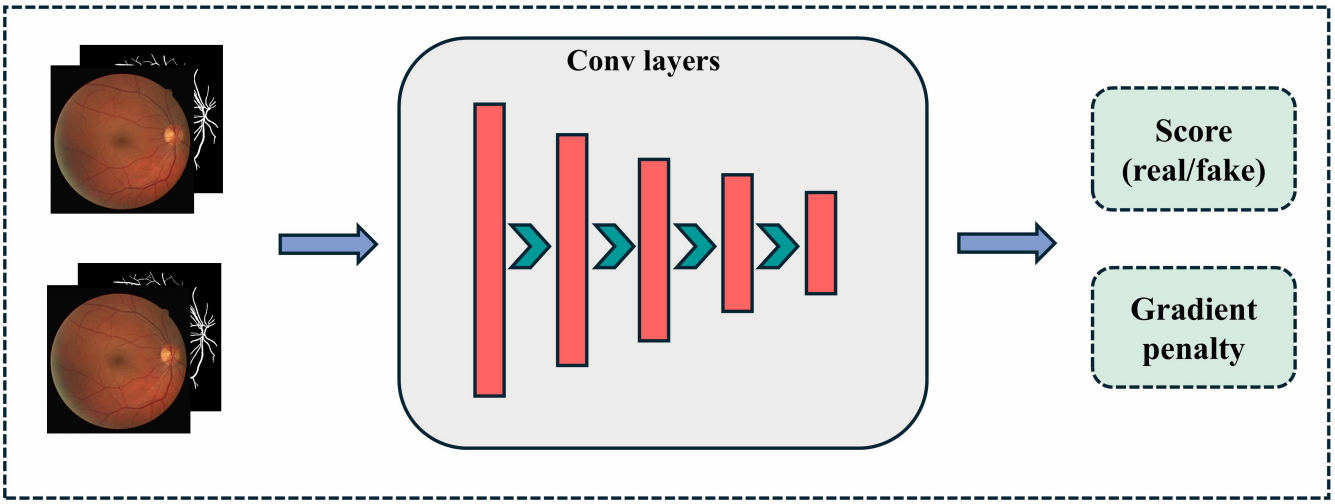


Fig. 5. Discriminator used in the proposed AA-WGAN.

pixel-wise similarity between the generator output and the corresponding ground truth image. However, the L1 loss is inherently limited in capturing high-frequency details such as edges and fine structures. To address this limitation, we incorporate a Fourier-based frequency-domain loss to enhance the model's sensitivity to high-frequency discrepancies. To better understand the frequency-domain characteristics of retinal vessel images, we analyzed the spectral distribution by computing the proportion of high-frequency regions in both spatial extent and energy contribution. Experimental analysis reveals that high-frequency regions cover approximately 60.7% of the spectral area, while contributing only 15.1% of the total spectral energy. Although the energy associated with high-frequency components is relatively low, these regions encode crucial visual information such as vessel edges, microvascular structures, and fine texture—elements that are vital for precise segmentation. This empirical evidence substantiates the theoretical rationale for introducing frequency-domain constraints in segmentation models.For more detailed information, please refer to the Appendix section. By applying the Fourier transform, we project both the generated and real images into the frequency domain and enforce structural alignment, particularly in the high-frequency bands. For the Fourier transform of a 2D image:

$$\mathcal{F}(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y)\, e^{-j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)} \tag{2}$$

After applying the Fourier transform to both the generated image and the ground truth image, their corresponding grayscale magnitude spectra are defined as follows:

$$F_{\text{gen}}(u,v), \quad F_{\text{real}}(u,v) \tag{3}$$

The magnitude difference between the predicted frequency spectrum and the target spectrum is computed as follows:

$$\Delta(u,v) = |F_{\text{gen}}(u,v) - F_{\text{real}}(u,v)| \tag{4}$$

To emphasize high-frequency error regions, the Focal Frequency Loss incorporates a dynamic weighting mechanism that assigns adaptive weights to each frequency component as follows:

$$W(u,v) = \log\left(1 + \Delta(u,v)^{\alpha}\right), \quad \alpha > 0 \tag{5}$$

In this formulation, the hyperparameter $\alpha$ controls the strength of the weighting, and $\log(1+\cdot)$ serves to smooth the gradients, thereby preventing gradient explosion. The final frequency-domain loss is computed by performing an element-wise multiplication between the weighting matrix and the spectral magnitude difference, followed by normalization:

$$\mathcal{L}_{\text{freq}} = \frac{\sum_{u,v} W(u,v) \cdot \Delta(u,v)}{\sum_{u,v} W(u,v) + \varepsilon} \tag{6}$$

In this context, the constant $\varepsilon$ is introduced to prevent division by zero. To comprehensively improve the performance of the model in retinal vessel segmentation, we combine time-domain and frequency-domain constraints by integrating the L1 loss and the Focal Frequency Loss. The L1 loss operates in the pixel domain, promoting global structural alignment and accuracy between the predicted and ground truth images. On the other hand, the Focal Frequency Loss functions in the Fourier domain, selectively focusing on regions exhibiting high spectral discrepancy—particularly in the high-frequency range—which includes critical anatomical features such as thin vessels, edges, and textures. This dual-domain design enhances the model's sensitivity to both global structure and local detail. Consequently, the final objective function of the generator in our framework is formulated as the sum of adversarial loss, L1 loss, and frequency-domain loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{adv}} + \lambda_{L1}\mathcal{L}_{L1} + \lambda_{\text{reg}}\mathcal{L}_{\text{FFL}} \tag{7}$$

In this formulation, $\mathcal{L}_{\text{adv}}$ denotes the adversarial loss term based on the WGAN-GP objective, while $\mathcal{L}_{L1}$ and $\mathcal{L}_{\text{FFL}}$ represent the L1 loss and the Fourier-domain loss, respectively. The weighting factors $\lambda_{L1}$ and $\lambda_{\text{reg}}$ are used to scale the contributions of the time-domain and frequency-domain losses. These coefficients are selected via cross-validation to balance the trade-off between global image fidelity and the preservation of fine structural details in the generated segmentation maps.

## IV. Results and Discussion

In this section, we present the evaluation of the proposed FMA-WGAN across three widely utilized open retinal fundus datasets: DRIVE, STARE, and CHASE_DB1. Extensive comparative experiments are carried out against a range of cutting-edge techniques to rigorously validate the effectiveness of our model in retinal vessel segmentation.

### A. Datasets

- **DRIVE** : This dataset contains 40 retinal scans with dimensions of 564 × 584 pixels. Of these, 33 are without lesions, whereas 7 exhibit indications of diabetic retinopathy.
- **STARE** : This collection consists of 20 fundus photographs, covering both healthy and diseased cases, each with a size of 700 × 605 pixels.
- **CHASE_DB1** : The dataset is made up of 28 color fundus images captured from both eyes of 14 children, each image presented at a resolution of 999 × 960 pixels.

### B. Data preprocessing

We designed a multi-step preprocessing pipeline. Initially, the original RGB images are transformed into the YCbCr space, enabling the decomposition of luminance and chrominance information. Thereafter, Contrast Limited Adaptive Histogram Equalization (CLAHE) is individually conducted on the Y, Cb, and Cr components to refine local contrast and facilitate the visualization of delicate vascular structures. Finally, gamma correction is performed on the enhanced image to further optimize brightness levels and emphasize image details.

### C. Evaluation metrics

To quantitatively assess the segmentation ability of our model, four standard metrics are employed, namely sensitivity, specificity, accuracy, and the area under the ROC curve (AUC). The mathematical definitions of SE, SP, and ACC are provided in Eqs. (8)–(10)

$$SE = \frac{TP}{TP + FN} \tag{8}$$

$$SP = \frac{TN}{TN + FP} \tag{9}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

The manually annotated vessel segmentation maps provided by clinical experts serve as the reference standard for evaluation. In this framework, $TP$, $TN$, $FP$, and $FN$ represent true positives, true negatives, false positives, and false negatives, respectively. The model's performance is evaluated based on four key indices: sensitivity, specificity, accuracy , and the area under the ROC curve. Larger values of SE, SP, ACC, and AUC—approaching 1—indicate superior segmentation accuracy. Furthermore, the ROC curve together with its AUC score are employed for overall model assessment, where a larger AUC reflects a more effective and reliable segmentation model. where a larger AUC reflects a more effective and reliable segmentation model.

### D. Implementation details

Due to the limited number of images and the high resolution of the fundus datasets, data augmentation was employed to enhance training diversity and mitigate overfitting. Specifically, the augmented images were randomly cropped to a predefined size and subjected to random horizontal flipping, thereby increasing the variability and robustness of the training samples.

The experiments were conducted on a workstation equipped with an NVIDIA GeForce RTX 4090 GPU. The network architecture was implemented, trained, and evaluated using the PyTorch deep learning framework. During training, the batch size was set to 1, and the initial learning rate for both the generator and the discriminator was fixed at 0.0001. The Adam optimizer was employed, and training was carried out for a total of 80 epochs. Following the WGAN-GP protocol, the discriminator was updated five times for every generator update in each iteration. The total loss function used in training was a weighted combination of three components: adversarial loss, L1 loss (spatial domain), and Focal Frequency Loss (frequency domain). The corresponding weights for these loss terms were set to $\lambda_{L1} = 100$ for the L1 loss, $\lambda_{reg} = 10.0$ for the frequency-domain loss, and $\lambda_{gp} = 10.0$ for the WGAN-GP gradient penalty.

### E. Ablation experiments

To validate the effectiveness of the proposed model, a series of ablation studies were conducted to assess the individual and combined contributions of key architectural components. The Attention Gate U-Net (AG-Unet) was employed as the baseline model. Subsequently, three modules—namely the Multi-scale input fusion, WGAN-GP adversarial optimization, and Frequency-aware loss—were incrementally integrated into the baseline. The performance of each configuration was then systematically evaluated across three publicly available retinal image datasets: DRIVE, CHASE_DB1, and STARE.

TABLE I
Results of ablation experiments on the DRIVE dataset.

| Model | ACC | SPE | SEN | AUC |
|---|---|---|---|---|
| Baseline | 0.9480 | 0.9646 | 0.7826 | 0.9534 |
| Baseline + WGAN | 0.9513 | 0.9632 | 0.8010 | 0.9657 |
| Baseline + WGAN + ML | 0.9590 | **0.9730** | 0.8562 | 0.9752 |
| Baseline + WGAN + FA | 0.9623 | 0.9693 | 0.8620 | 0.9788 |
| Baseline + ML + FA | 0.9627 | 0.9680 | 0.8572 | 0.9789 |
| FMA-WGAN (Ours) | **0.9652** | 0.9723 | **0.8733** | **0.9845** |

TABLE II
Results of ablation experiments on the STARE dataset.

| Model | ACC | SPE | SEN | AUC |
|---|---|---|---|---|
| Baseline | 0.9531 | 0.9624 | 0.8242 | 0.9563 |
| Baseline + WGAN | 0.9564 | 0.9647 | 0.8127 | 0.9596 |
| Baseline + WGAN + ML | 0.9662 | **0.9767** | 0.8667 | 0.9771 |
| Baseline + WGAN + FA | **0.9746** | 0.9724 | 0.8740 | 0.9801 |
| Baseline + ML + FA | 0.9710 | 0.9734 | 0.8781 | 0.9778 |
| FMA-WGAN (Ours) | 0.9743 | 0.9760 | **0.8839** | **0.9892** |

TABLE III
RESULTS OF ABLATION EXPERIMENTS ON THE CHASE_DB1 DATASET.

| Model | ACC | SPE | SEN | AUC |
|---|---|---|---|---|
| Baseline | 0.9534 | 0.9598 | 0.8065 | 0.9554 |
| Baseline + WGAN | 0.9576 | 0.9651 | 0.7938 | 0.9661 |
| Baseline + WGAN + ML | 0.9673 | 0.9793 | 0.8425 | 0.9792 |
| Baseline + WGAN + FA | 0.9712 | 0.9747 | 0.8564 | **0.9884** |
| Baseline + ML + FA | 0.9681 | 0.9762 | 0.8512 | 0.9821 |
| FMA-WGAN (Ours) | **0.9738** | **0.9825** | **0.8606** | 0.9879 |

The quantitative results of the ablation study, including accuracy (ACC), sensitivity (SE), specificity (SP), and AUC, are presented in Table I–III. As observed, the inclusion of the WGAN-GP module on top of the baseline AG-Unet model led to consistent performance gains across the DRIVE, CHASE_DB1, and STARE datasets. On average, ACC increased by 0.0036 and AUC by 0.0088, indicating that the WGAN-GP module enhances segmentation quality by stabilizing adversarial training and improving the delineation of fine structures and vessel boundaries. Further improvements were achieved by sequentially incorporating the Multi-scale and Frequency-aware modules. Notably, sensitivity exhibited significant enhancement, with average gains of 0.0526 and 0.0616, respectively, across the three datasets. These results suggest that the Multi-scale module effectively captures vascular features across different spatial scales, thereby improving recognition of complex vessel architectures. Simultaneously, the Frequency-aware module contributes by extracting high-frequency structural details, ultimately boosting the model's generalization ability in diverse retinal imaging scenarios.

In summary, the complete model—FMA-WGAN (Ours)—which incorporates the Multi-scale input strategy, WGAN-GP-based discriminator, and the Frequency-aware loss module, achieved superior performance on all three benchmark datasets. These results conclusively demonstrate that the proposed combination of architectural and loss-function enhancements substantially improves the model's capability in retinal vessel segmentation, particularly in capturing fine structures and complex vascular patterns.

TABLE IV
EXPERIMENTAL RESULTS ON THE DRIVE DATASET.

| Model | Year | ACC | SPE | SEN | AUC |
|---|---|---|---|---|---|
| U-Net++ [33] | 2018 | 0.9573 | 0.9761 | 0.7762 | 0.9796 |
| IterNet [34] | 2020 | 0.9574 | 0.9831 | 0.7791 | 0.9813 |
| Wave-Net [35] | 2023 | 0.9561 | 0.9764 | 0.8164 | – |
| AFNet [36] | 2023 | 0.9580 | 0.9818 | 0.8139 | 0.9820 |
| ResDO-Unet [37] | 2023 | 0.9561 | 0.9791 | 0.7985 | – |
| WS-DMF [38] | 2024 | 0.9494 | **0.9841** | 0.8504 | – |
| SMDAA [39] | 2024 | 0.9615 | 0.9703 | 0.8609 | 0.9806 |
| (DA-U)$^{+}$Net [40] | 2025 | 0.9546 | 0.9720 | 0.8300 | – |
| PA-Net [21] | 2025 | 0.9582 | 0.9807 | 0.8284 | 0.9833 |
| **FMA-WGAN (Ours)** | – | **0.9652** | 0.9723 | **0.8733** | **0.9845** |

### *F. Comparison with the state-of-the-art methods*

A comparative study was carried out between the proposed FMA-WGAN model and a set of advanced approaches for retinal vessel segmentation. The experiments were implemented on three benchmark datasets: DRIVE, STARE, and CHASE_DB1. Model evaluation employed four key metrics: accuracy (ACC), sensitivity (SEN), specificity (SPE), and the area under the ROC curve (AUC). A summary of the detailed results is presented in Table IV–VI.

TABLE V
EXPERIMENTAL RESULTS ON THE STARE DATASET.

| Model | Year | ACC | SPE | SEN | AUC |
|---|---|---|---|---|---|
| U-Net++ [33] | 2018 | 0.9612 | 0.9801 | 0.8122 | 0.9841 |
| IterNet [34] | 2020 | 0.9701 | **0.9886** | 0.7715 | 0.9881 |
| Wave-Net [35] | 2023 | 0.9641 | 0.9836 | 0.7902 | – |
| AFNet [36] | 2023 | 0.9712 | 0.9779 | 0.7972 | 0.9902 |
| ResDO-Unet [37] | 2023 | 0.9567 | 0.9792 | 0.7963 | – |
| WS-DMF [38] | 2024 | 0.9613 | 0.9854 | 0.8448 | – |
| SMDAA [39] | 2024 | 0.9631 | 0.9711 | 0.8654 | 0.9819 |
| (DA-U)$^{2}$Net [40] | 2025 | 0.9620 | 0.9734 | 0.8210 | – |
| PA-Net [21] | 2025 | 0.9709 | 0.9805 | 0.8813 | **0.9908** |
| **FMA-WGAN (Ours)** | – | **0.9743** | 0.9760 | **0.8839** | 0.9892 |

TABLE VI
EXPERIMENTAL RESULTS ON THE CHASE_DB1 DATASET.

| Model | Year | ACC | SPE | SEN | AUC |
|---|---|---|---|---|---|
| U-Net++ [33] | 2018 | 0.9657 | 0.9801 | 0.8012 | 0.9869 |
| IterNet [34] | 2020 | 0.9655 | 0.9823 | 0.7970 | 0.9851 |
| Wave-Net [35] | 2023 | 0.9664 | 0.9821 | 0.8284 | – |
| AFNet [36] | 2023 | 0.9669 | 0.9817 | 0.8194 | 0.9867 |
| ResDO-Unet [37] | 2023 | 0.9672 | 0.9794 | 0.8020 | – |
| WS-DMF [38] | 2024 | 0.9566 | 0.9797 | 0.7841 | – |
| SMDAA [39] | 2024 | 0.9689 | 0.9749 | **0.8939** | 0.9832 |
| (DA-U)$^{2}$Net [40] | 2025 | 0.9581 | 0.9720 | 0.8160 | – |
| PA-Net [21] | 2025 | 0.9677 | 0.9779 | 0.8570 | 0.9875 |
| **FMA-WGAN (Ours)** | – | **0.9738** | **0.9825** | 0.8606 | **0.9879** |

Several key observations can be made based on the results presented in Table IV through Table VI: (1) On the DRIVE dataset, FMA-WGAN achieved the highest values in terms of accuracy (ACC), sensitivity (SE), and area under the ROC curve (AUC). On the STARE dataset, it attained the highest ACC and SE scores, while on the CHASE_DB1 dataset, it recorded the best ACC, specificity (SP), and AUC values. These results collectively demonstrate that FMA-WGAN offers the most consistent and robust performance across all three datasets, highlighting its superior capability for accurate retinal vessel segmentation. (2) Sensitivity (SE) quantifies the proportion of correctly detected vessel pixels relative to the total number of annotated vessel pixels, and serves as a key indicator of the model's effectiveness in identifying vascular structures. Due to the highly imbalanced nature of retinal images—where background pixels significantly outnumber vessel pixels—SE is particularly critical for evaluating segmentation performance. Our proposed FMA-WGAN consistently achieves higher SE values compared to other models, demonstrating its superior capability in recognizing vessel pixels. As shown in Figure 6, a bar chart comparison of SE across various models further confirms that FMA-WGAN outperforms competing approaches with a notable margin in sensitivity.

Figure 7 illustrates the Receiver Operating Characteristic (ROC) curves of seven different models evaluated on the DRIVE, CHASE_DB1, and STARE datasets. In general, ROC curves that are closer to the upper-left corner indicate superior classification performance. As shown in the figure, FMA-WGAN exhibits the most favorable ROC

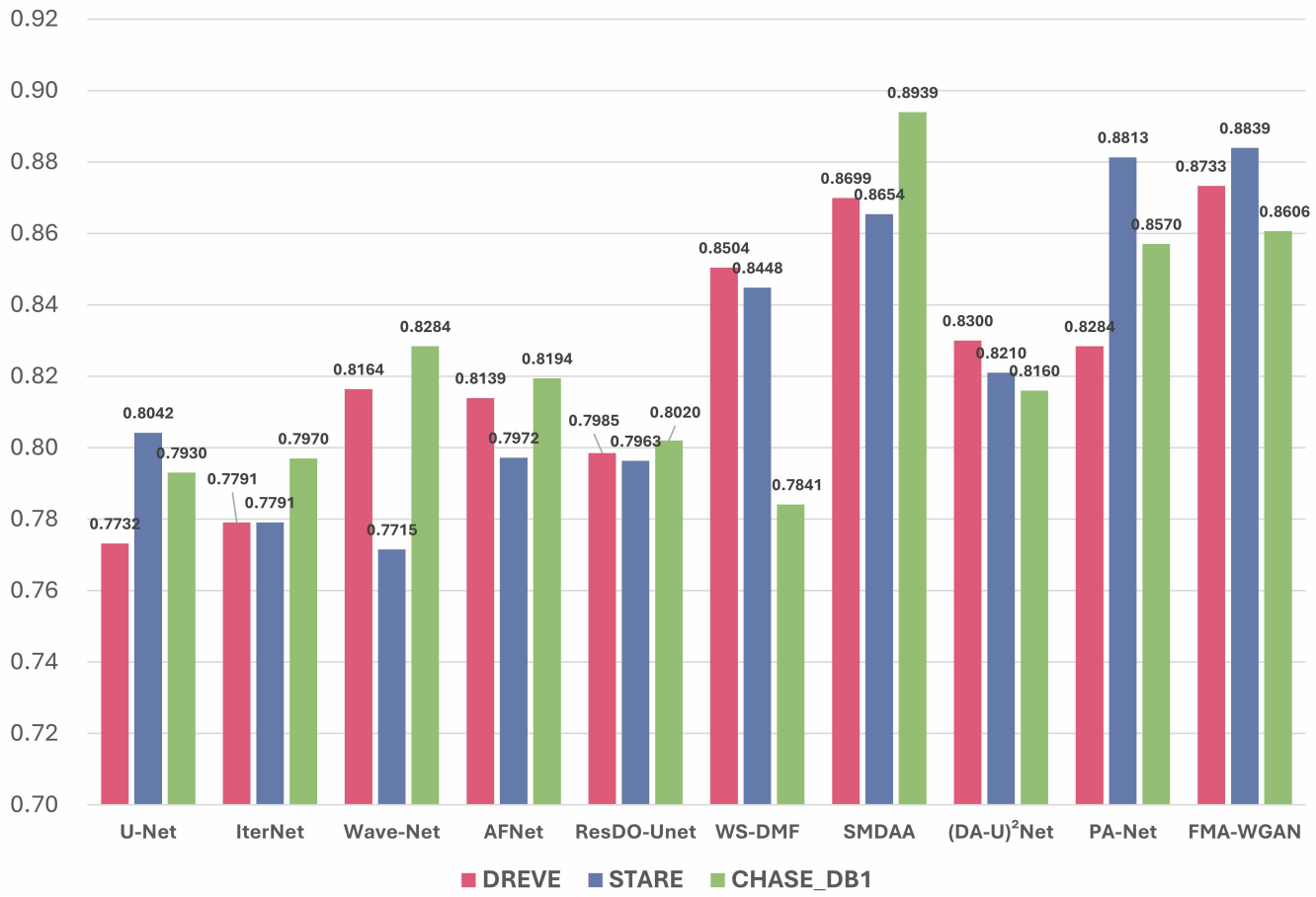


Fig. 6. Sensitivity comparison of different models.

positioning on both the DRIVE and CHASE_DB1 datasets, with curves approaching the ideal top-left corner. On the STARE dataset, its performance is similarly competitive, closely aligning with the best-performing models. These observations collectively demonstrate that FMA-WGAN achieves the most consistent and robust results in terms of ROC-based evaluation.

### *G. Qualitative analysis*

Figure 8 displays the qualitative segmentation results produced by FMA-WGAN across the DRIVE, STARE, and CHASE_DB1 datasets, alongside comparative results from U-Net++, SMADD, and PA-Net. To facilitate detailed visual analysis, magnified insets of the green-boxed regions are also provided. Several key observations emerge from this comparison: (1) FMA-WGAN demonstrates markedly improved accuracy in the segmentation of elongated vessels, exhibiting substantially fewer discontinuities compared to competing methods. This superior performance highlights the efficacy of the Multi-scale module in modeling long-range contextual dependencies by integrating features across multiple spatial resolutions. Such capability not only mitigates vessel fragmentation but also enhances the model's robustness and adaptability to complex retinal anatomical structures. (2) The model also excels at identifying finer microvascular details and preserving the continuity of the retinal vasculature. This advantage stems from the Frequency-aware module, which addresses the common issue of high-frequency detail loss in traditional segmentation methods. By emphasizing spectral-domain learning, it enables the model to recover subtle structural boundaries and minute vessel features more effectively.

In summary, FMA-WGAN delivers qualitatively superior segmentation results across all three datasets, clearly outperforming the benchmark and competing models in visual fidelity and structural preservation.

### *H. Discussion*

This section introduces an extensive set of experiments performed on three open datasets—DRIVE, CHASE_DB1, and STARE—to assess the performance of the proposed FMA-WGAN network. The results indicate that FMA-WGAN consistently surpasses a range of advanced methods with respect to accuracy (ACC), while also achieving high sensitivity (SEN), specificity (SP), and AUC across the majority of datasets. These findings verify the model's strong segmentation capability and generalization ability. The findings show that incorporating each module yields notable gains in segmentation accuracy over the datasets, validating the architectural design and underscoring the synergistic benefits of the integrated modules.

Despite these promising outcomes, some limitations were observed under challenging conditions. Specifically, the model showed slightly decreased sensitivity on images with low contrast or artifact interference, revealing opportunities for further improvement in fine-detail recovery. Moreover, while FMA-WGAN introduces additional components

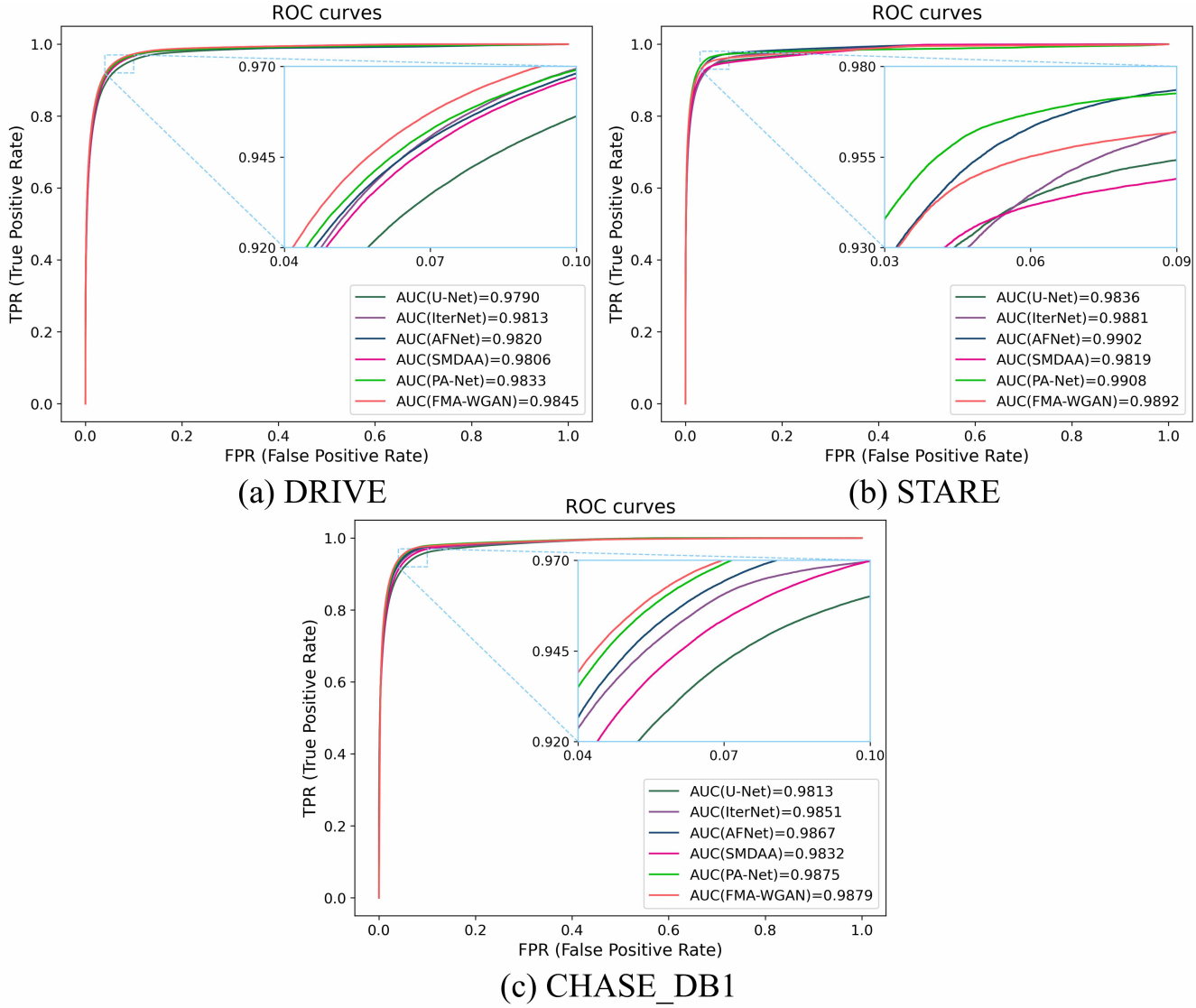


Fig. 7. ROC curves on the DRIVE, CHASE_DB1, and STARE datasets.

compared to the traditional U-Net architecture, and maintains a relatively moderate parameter scale, further optimization is still warranted in terms of computational efficiency and deployment feasibility.

## V. Conclusions

In this work, we propose a novel framework for retinal vessel segmentation, termed FMA-WGAN, which extends the classical Attention U-Net architecture by incorporating three key innovations: multi-scale feature learning, frequency-aware constraints, and generative adversarial training. Specifically, the WGAN-GP-based discriminator enhances the model's ability to distinguish fine vessel boundaries by enabling more stable adversarial learning. The multi-scale module enables the network to capture vascular structures across different spatial resolutions, thereby improving its robustness to scale variation. Furthermore, the frequency-aware loss guides the generator to emphasize high-frequency components within the image, substantially improving the modeling of microvascular details and subtle anatomical features.

Experiments conducted on the publicly accessible DRIVE, CHASE_DB1, and STARE datasets show that the proposed FMA-WGAN surpasses the baseline and other comparative methods across various evaluation indices, such as accuracy, sensitivity, specificity, and the area under the ROC curve (AUC). These findings confirm the individual effectiveness of the integrated components—namely multi-scale feature learning, WGAN-GP adversarial training, and frequency-aware loss—and highlight their synergistic contribution to enhancing overall segmentation performance.

Although FMA-WGAN has achieved substantial progress in retinal vessel segmentation, it may still face limitations when dealing with images characterized by extremely low contrast or the presence of noise-induced artifacts. In future work, we intend to investigate more robust preprocessing techniques that offer enhanced noise resilience. Furthermore, we plan to incorporate lightweight Transformer-based modules to improve both segmentation accuracy and computational efficiency. Additionally, we aim to explore self-supervised learning paradigms to alleviate the reliance on manually annotated datasets, thereby promoting the broader applicability of the proposed approach in real-world clinical scenarios.

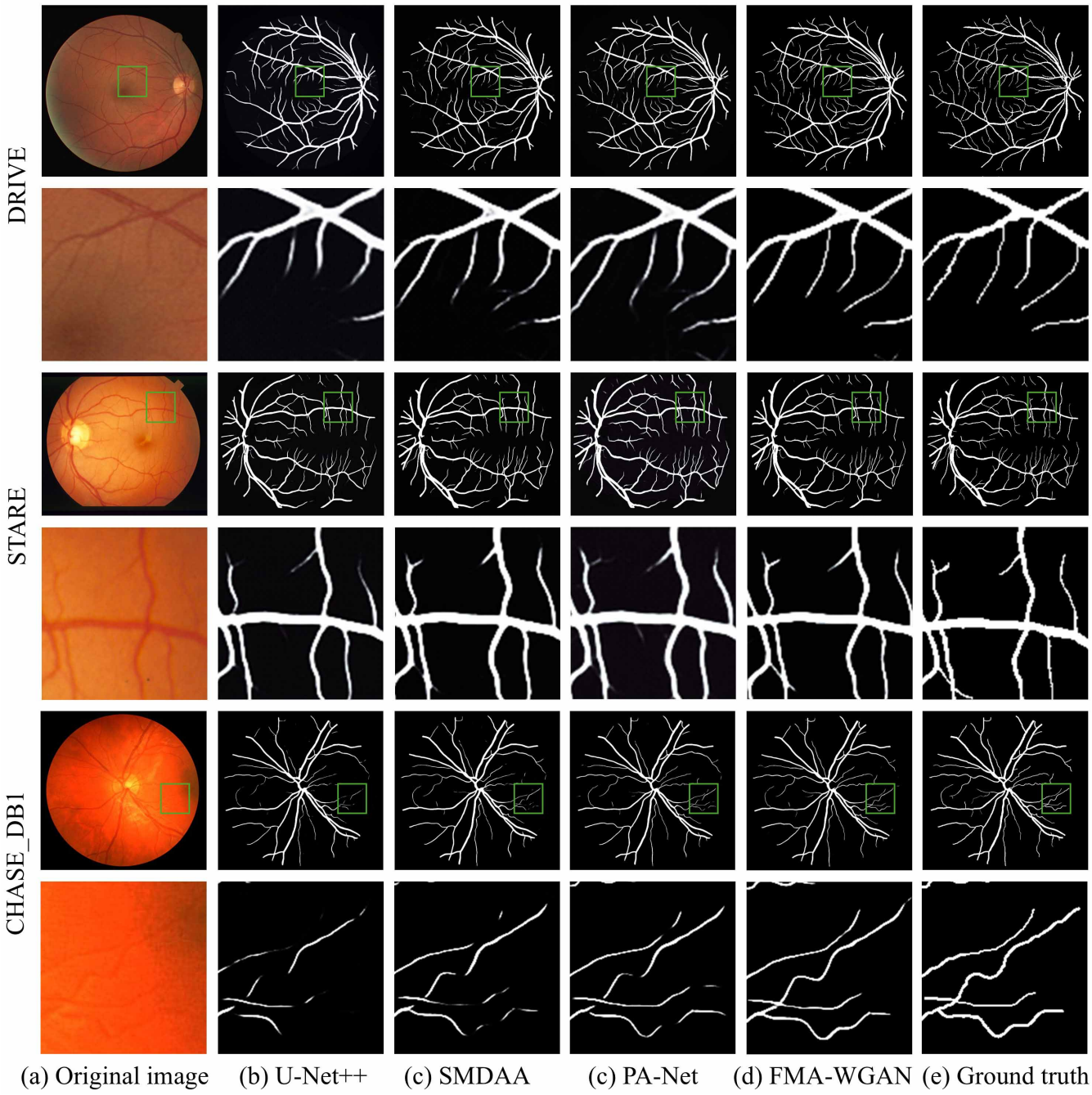


Fig. 8. Segmentation results on the DRIVE, STARE, and CHASE_DB1 datasets.

## APPENDIX

### *Experimental Verification of High-Frequency Characteristics*

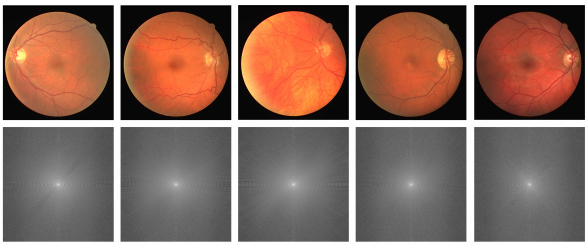

Fig. A1. Representative retinal images and their corresponding Fourier spectra from the DRIVE dataset.

To examine the energy distribution characteristics of retinal images in the frequency domain—particularly the relationship between vascular structures and high-frequency components—we conducted a spectral analysis on the DRIVE dataset using a two-dimensional Fourier transform (2D FFT). Each image was first converted to grayscale, and its frequency magnitude spectrum was computed via Fourier transformation. To facilitate clearer observation of frequency properties, the spectrum was shifted (FFT shift) such that the low-frequency components were repositioned at the center. Given an image size of $H \times W$, the center of the frequency spectrum is located at $(H/2, W/2)$. A radial frequency mask was constructed based on the Euclidean distance in the frequency domain, and the boundary between low and high frequencies was defined as half of the maximum possible

radial distance, calculated as follows:

$$r = 0.5 \cdot \sqrt{\left(\frac{H}{2}\right)^2 + \left(\frac{W}{2}\right)^2} \tag{11}$$

All frequency components with a radial distance greater than rrr from the center of the frequency domain are classified as belonging to the high-frequency region, while those within the radius rrr are considered low-frequency components. The spectral energy contribution of the high-frequency region is subsequently computed to quantify its proportion relative to the total energy of the image, as defined below:

$$\text{HighFreqEnergyRatio} = \frac{\sum_{(u,v)\in\mathcal{S}_{HF}} |F(u,v)|}{\sum_{(u,v)} |F(u,v)|} \tag{12}$$

Figure A1 illustrates the frequency spectra of representative retinal images from the DRIVE dataset after applying the two-dimensional Fourier transform. As shown, the high-frequency components are broadly distributed across the spectrum. The proportion of high-frequency energy present in the dataset was quantitatively analyzed, and experimental results indicate that high-frequency components account for approximately 15% of the total spectral energy in the images. In the context of retinal vessel segmentation, fine vascular structures and boundary regions are predominantly encoded in the high-frequency domain. Although these regions contribute relatively little energy, they are densely packed with structural details that are crucial for precise segmentation. These findings validate the theoretical motivation for introducing frequency-domain loss functions, which enhance the model's ability to reconstruct high-frequency features and improve segmentation accuracy, particularly for microvascular structures.

## REFERENCES


[1] X. Zhang, G. Thibault, E. Decencière, B. Marcotegui, B. Laÿ, R. Danno, G. Cazuguel, G. Quellec, M. Lamard, P. Massin, A. Chabouis, Z. Victor, and A. Erginay, "Exudate detection in color retinal images for mass screening of diabetic retinopathy," *Medical Image Analysis*, vol. 18, no. 7, pp. 1026–1043, 2014.

[2] Y. Zhang, Y. Xu, and J. Zhai, "Diabetic retinopathy image segmentation method based on fusion densenet and u-net network," *Engineering Letters*, vol. 33, no. 2, pp. 418–428, 2025.

[3] B. Zhang, L. Zhang, L. Zhang, and F. Karray, "Retinal vessel extraction by matched filter with first-order derivative of gaussian," *Computers in Biology and Medicine*, vol. 40, no. 4, pp. 438–445, 2010.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[5] D. Li, L. Peng, S. Peng, H. Xiao, and Y. Zhang, "Retinal vessel segmentation by using afnet," *The Visual Computer*, vol. 39, no. 5, pp. 1929–1941, 2023.

[6] J. Cao, J. Chen, Y. Gu, and J. Liu, "Mfa-unet: a vessel segmentation method based on multi-scale feature fusion and attention module," *Frontiers in Neuroscience*, vol. 17, pp. 1249331, 2023, eCollection 2023.

[7] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein gans," 2017.

[8] 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*, vol. 8, no. 3, pp. 263–269, 1989.

[9] D. Kumar, A. Pramanik, S. S. Kar, and S. P. Maity, "Retinal blood vessel segmentation using matched filter and laplacian of gaussian," in *2016 International Conference on Signal Processing and Communications (SPCOM)*, 2016, pp. 1–5.

[10] Q. Li, J. You, and D. Zhang, "Vessel segmentation and width estimation in retinal images using multiscale production of matched filter responses," *Expert Systems with Applications*, vol. 39, no. 9, pp. 7600–7610, 2012.

[11] Z. Jiang, J. Yepez, S. An, and S. Ko, "Fast, accurate and robust retinal vessel segmentation system," *Biocybernetics and Biomedical Engineering*, vol. 37, no. 3, pp. 412–421, 2017.

[12] Y. Zhao, L. Rada, K. Chen, S. P. Harding, and Y. Zheng, "Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images," *IEEE Transactions on Medical Imaging*, vol. 34, no. 9, pp. 1797–1807, Sep. 2015.

[13] U. T. Nguyen, A. Bhuiyan, L. A. Park, and K. Ramamohanarao, "An effective retinal blood vessel segmentation method using multi-scale line detection," *Pattern Recognition*, vol. 46, no. 3, pp. 703–715, 2013.

[14] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 9, pp. 2538–2548, 2012.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[16] Y. Chai and Y. Zhang, "Dim-unet: Enhancing medical image segmentation with dual-input hybrid attention and state space modeling," *Engineering Letters*, vol. 33, no. 6, pp. 2186–2194, 2025.

[17] C. Wang, R. Xu, Y. Zhang, S. Xu, and X. Zhang, "Retinal vessel segmentation via context guide attention net with joint hard sample mining strategy," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 2021, pp. 1319–1323.

[18] M. A. Al-masni and D.-H. Kim, "Cmm-net: Contextual multi-scale multi-level network for efficient biomedical image segmentation," *Scientific Reports*, vol. 11, no. 1, pp. 10191, 2021.

[19] C. Beeche, J. P. Singh, J. K. Leader, N. S. Gezer, A. P. Oruwari, K. K. Dansingani, J. Chhablani, and J. Pu, "Super u-net: A modularized generalizable architecture," *Pattern Recognition*, vol. 128, pp. 108669, 2022.

[20] L. Zhang, C. Xu, Y. Li, T. Liu, and J. Sun, "Mcse-u-net: multi-convolution blocks and squeeze and excitation blocks for vessel segmentation," *Quantitative Imaging in Medicine and Surgery*, vol. 14, no. 3, pp. Fill-in appropriate pages if needed, 2024.

[21] X. Luo, L. Peng, Z. Ke, J. Lin, and Z. Yu, "Pa-net: A hybrid architecture for retinal vessel segmentation," *Pattern Recognition*, vol. 161, pp. 111254, 2025.

[22] Y. Xie, J. Shang, Q. Yang, X. Qian, H. Zhang, and X. Tang, "Arsa-unet: Atrous residual network based on structure-adaptive model for retinal vessel segmentation," *Biomedical Signal Processing and Control*, vol. 96, pp. 106595, 2024.

[23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2023.

[24] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2021.

[25] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," 2021.

[26] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," 2021.

[27] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," 2021.

[28] N. Lv, L. Xu, Y. Chen, W. Sun, J. Tian, and S. Zhang, "Tcddu-net: combining transformer and convolutional dual-path decoding u-net for retinal vessel segmentation," *Scientific Reports*, vol. 14, no. 1, pp. 25978, 2024.

[29] J. Li, Q. Cheng, and C. Wu, "Gvit-rsnet: A retinal vessel segmentation network using graph convolutional attention and multi-scale vision transformer," *Pattern Recognition Letters*, vol. 189, pp. 182–187, 2025.

[30] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.

[31] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014.

[32] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," 2018.
[33] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham: Springer International Publishing, 2018, pp. 3–11.
[34] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, "Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks," in *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020, pp. 3645–3654.
[35] Y. Liu, J. Shen, L. Yang, H. Yu, and G. Bian, "Wave-net: A lightweight deep network for retinal vessel segmentation from fundus images," *Computers in Biology and Medicine*, vol. 152, pp. 106341, 2023.
[36] D. Li, L. Peng, S. Peng, H. Xiao, and Y. Zhang, "Retinal vessel segmentation by using afnet," *The Visual Computer*, vol. 39, no. 5, pp. 1929–1941, 2023.
[37] Y. Liu, J. Shen, L. Yang, G. Bian, and H. Yu, "Resdo-unet: A deep residual network for accurate retinal vessel segmentation from fundus images," *Biomedical Signal Processing and Control*, vol. 79, pp. 104087, 2023.
[38] 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*, vol. 283, pp. 111185, 2024.
[39] W. Zhou, X. Wang, X. Yang, Y. Hu, and Y. Yi, "Skeleton-guided multi-scale dual-coordinate attention aggregation network for retinal blood vessel segmentation," *Computers in Biology and Medicine*, vol. 181, pp. 109027, 2024.
[40] B. Chu, J. Zhao, W. Zheng, and Z. Xu, "(da-u)$^2$net: double attention u2net for retinal vessel segmentation," *BMC Ophthalmology*, vol. 25, no. 1, pp. 86, 2025.