

DAMMGA-Net: Direction-Aware Attention with Genetic Algorithm-Optimized Loss for Retinal Vessel Segmentation

Yintao Hong

City University of Macau

Yajun Xie

xyj@fzfu.edu.cn

Fuzhou University of International Studies and Trade

Yanyan Wu

City University of Macau

Jinwei Wang

City University of Macau

Article

Keywords: retinal vessel segmentation, direction-aware mixed attention, multi-scale fixed receptive-field pyramid, genetic algorithm, loss function search, class imbalance

Posted Date: April 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8934073/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

DAMMGA-NET: DIRECTION-AWARE ATTENTION WITH GENETIC ALGORITHM-OPTIMIZED LOSS FOR RETINAL VESSEL SEGMENTATION

Yintao Hong

Faculty of Data Science, City University of Macau.
School of Intelligent Manufacturing Engineering, Liming Vocational University.
Key Laboratory of Data Science and Intelligent Computing, Fuzhou University of International Studies and Trade .
D24092100431@cityu.edu.mo

Yajun Xie*

School of Big Data, Fuzhou University of International Studies and Trade.
Key Laboratory of Data Science and Intelligent Computing, Fuzhou University of International Studies and Trade.
xyj@fzfu.edu.cn

Yanyan Wu

Faculty of Data Science, City University of Macau.
Key Laboratory of Data Science and Intelligent Computing, Fuzhou University of International Studies and Trade.
D24092100054@cityu.edu.mo

JinWei Wang

Faculty of Data Science, City University of Macau.
School of Intelligent Manufacturing Engineering, Liming Vocational University.
D24092100091@cityu.edu.mo

2

ABSTRACT

Retinal vessel segmentation is fundamental for early screening and quantitative analysis of fundus diseases, yet frequent fragmentation of fine vessels, pronounced scale variation, and severe foreground-background imbalance still limit structural continuity and detail preservation. To address these challenges, we propose DAMMGA-Net, a lightweight retinal vessel segmentation framework built on an HFEM encoder and decoder-side Refinement Blocks. The proposed framework improves segmentation through three complementary components. First, the Direction-Aware Mixed Attention Module (DAMM) enhances directional continuity modeling for thin, tortuous, and branching vessels. Second, the Multi-scale Fixed Receptive-field Parallel Pyramid (MFRPP) aggregates cross-receptive-field information at the bottleneck to improve representation consistency across vessel scales. Third, GA-Loss adaptively optimizes the weights and internal parameters of a bounded composite-loss family to alleviate optimization bias caused by class imbalance. Experiments on DRIVE, CHASE_DB1, and STARE show that DAMMGA-Net achieves stable and competitive per-

*Corresponding author: xyj@fzfu.edu.cn

²This work was supported by the National Natural Science Foundation of China (NSFC) 12371378; Fujian Province University Industry-University Cooperation Project 2025H6016; Fujian Provincial Natural Science Foundation Project 2023J011127; Key Laboratory Foundation Project of Fuzhou University of International Studies and Trade SZ2025002.

ORCID(s): 0009-0006-0356-7995 (Y. Hong); 0000-0002-4796-5032 (Y. Xie); 0000-0001-6098-701X (Y. Wu); 0000-0003-1162-628X (J. Wang).

formance while better preserving fine vessels, reaching AUC 0.9945, F1 0.8721, and Acc 0.9804 on STARE, and AUC 0.9879, F1 0.8413, and Acc 0.9732 on DRIVE. Qualitative results further show improved connectivity and integrity in fine branches, tortuous vessels, and low-contrast regions. These findings indicate that the synergy of directional continuity modeling, stable cross-scale aggregation, and dataset-adaptive loss optimization is effective for retinal vessel segmentation under severe class imbalance.

Keywords retinal vessel segmentation, direction-aware mixed attention, multi-scale fixed receptive-field pyramid, genetic algorithm, loss function search, class imbalance

1 Introduction

Retinal vascular networks can be directly observed noninvasively, and precise analysis of their morphological characteristics is of great value for the early diagnosis and monitoring of fundus diseases such as diabetic retinopathy, hypertensive retinopathy, and glaucoma [1]. High-quality vessel segmentation maps are not only fundamental for downstream tasks such as vessel diameter measurement and arterio-venous classification, but also provide critical support for intelligent computer-aided diagnosis [2]. Therefore, achieving accurate and stable retinal vessel segmentation in complex backgrounds has become an important research problem in medical image analysis and computer vision.

Vessel structures exhibit pronounced variations in caliber, branching, multi-scale features and directionality [3]. Traditional segmentation methods often rely on filters, morphological operators, or vessel enhancement models to extract vessel structures [4]. These methods perform reasonably in simple scenarios, but due to the multi-scale characteristics, including uneven vessel thickness, branching, and curvature, they often lack robustness under noise interference or pathological background, leading to false positives or missed detections.

Despite the widespread adoption of convolution-based architectures such as U-Net for retinal vessel segmentation, their reliance on local receptive fields fundamentally limits the modeling of long-range structural continuity and directional vascular patterns [5]. Although U-Net achieves reliable multi-scale fusion through skip connections, its locality-constrained operations often fail to preserve fine-grained curvatures and branching continuity, leading to discontinuities and missed detections in small or tortuous vessels [6].

Recently, CNNs with large convolutional kernels (e.g., RepLKNet) have been proposed to approximate global receptive fields more efficiently while maintaining the advantages of convolution [7]. Although such architectures demonstrate performance comparable to or surpassing Transformers in natural image tasks, they still face inherent limitations in medical image segmentation. In particular, both Transformer-based architectures and large-kernel CNNs share a fundamental constraint: they lack robust mechanisms to capture the directional curvature and scale-varying vascular structures characteristic of retinal vessels. These limitations motivate the design of DAMMGA-Net, in which the Multi-scale Fixed Receptive-field Parallel Pyramid (MFRPP) is introduced for efficient and stable multi-scale aggregation, while the Direction-Aware Mixed Attention (DAMM) is designed to enhance directional continuity modeling of retinal vasculature.

To overcome these locality-induced limitations, recent works have explored two primary directions: enhancing global dependency modeling and enlarging effective receptive fields. Self-attention mechanisms (Transformers), owing to their ability to model long-range interactions, have become an important avenue to augment U-Net’s global feature representation. In recent years, Transformers have been widely applied in semantic segmentation and medical image segmentation, excelling at capturing complex vessel branches and curved structures. However, these methods commonly suffer from high computational complexity and difficult training, limiting their feasibility in clinical applications [8]. For example, the CSAT network [9] introduces cross-scale attention to improve global modeling but requires long training cycles; MTPA-UNet [10] enhances multi-scale representation by combining Transformer blocks with CNNs; Transformer-U-Net employs a dual-path decoder to strengthen local detail reconstruction [11]; and TCDDU-Net design integrates both Transformer and convolution streams for boundary refinement [12]. Despite their strong modeling capabilities, these architectures typically involve large parameter footprints and considerable computational overhead, restricting deployment in resource-limited scenarios.

While architectural innovations address spatial modeling challenges, the severe foreground-background imbalance in retinal images introduces an additional optimization difficulty that must also be carefully handled. Existing studies mainly follow two directions: (1) hyperparameter optimization of a single loss, for example setting class weights in weighted BCE, tuning α_{tv} and β_{tv} in the Tversky loss to trade off FP and FN, or adjusting γ and α in the Focal loss to emphasize hard examples; and (2) weighted composition of multiple losses, with common forms such as $\alpha \cdot \text{BCE} + \beta \cdot \text{Dice} + \gamma \cdot \text{Tversky}$ (optionally combined with Focal) to jointly account for region overlap and pixel-wise discrimination. However, both lines of work typically depend heavily on manual expertise and extensive tuning, and tend

to generalize poorly across datasets. In recent years, automated loss function search has gradually emerged, including reinforcement learning-based loss exploration (AM-LFS) [13], differentiable proxy optimization (AutoSeg-Loss) [14], genetic expression evolution (AutoLoss-Zero) [15], and Monte Carlo Tree Search (MCTS) [16]. While these methods can discover or even “create” new loss structures, they often incur high computational costs and may not generalize well in small-sample medical settings. Therefore, instead of conducting open-ended symbolic loss generation, we adopt a lightweight genetic algorithm optimization strategy to efficiently search for optimal configurations within an existing composite-loss space, thereby better matching the practical constraints of retinal vessel segmentation.

Beyond global dependency and receptive-field modeling, the directionality of retinal vasculature introduces an additional layer of structural complexity. Previous works attempted to enhance directional sensitivity using anisotropic convolution kernels (e.g., 1×3 , 3×1) [17], or axial attention mechanisms that decompose 2D attention into directional components. Likewise, Coordinate Attention [18] and Polarized Self-Attention [19] incorporate positional cues into channel attention to improve fine-grained localization. While these approaches provide useful directional cues, they remain limited in modeling multi-directional curvature and often treat direction-specific operators separately from channel or pixel attention, thus lacking a unified integration strategy. To address this gap, the proposed DAMM module fuses 1×7 and 7×1 directional kernels with pixel-wise and channel attention in a parallel mixed-attention design, enabling richer orientation modeling and improved continuity of fine vessels.

More importantly, DAMMGA-Net is designed as an integrated framework in which the components act synergistically rather than as isolated add-ons. **(1) Structural modeling level.** DAMM is inserted into the skip-connection pathway to enhance direction-aware continuity and pixel-wise saliency, making thin, tortuous, and low-contrast vessels more distinguishable, especially around bifurcations and curved segments. Built upon these direction-enhanced representations, MFRPP is placed at the bottleneck to aggregate multi-scale context with fixed receptive fields, enabling efficient modeling of vessels with varying calibers while avoiding the heavy overhead of Transformer-style global modeling. **(2) Loss-function optimization level.** Complementary to the structural design, the Genetic-Algorithm-optimized Loss (GA-Loss) mitigates severe foreground-background imbalance by adaptively tuning the weights and internal parameters of a composite-loss family under explicit constraints, using validation performance as the fitness signal within a finite and interpretable search space.

In this sense, DAMM improves local directional discrimination, MFRPP strengthens scale-aware semantic aggregation, and GA-Loss aligns the training objective with the imbalance-sensitive nature of retinal vessel segmentation. This structural-optimization co-design clearly distinguishes DAMMGA-Net from existing U-Net variants that mainly refine architecture alone, Transformer-based methods that pursue global modeling at high computational cost, and generic automatic loss-search frameworks that emphasize open-ended search rather than task-specific, resource-efficient optimization.

Overall, DAMMGA-Net provides a unified structural-optimization framework for retinal vessel segmentation by jointly addressing directional continuity modeling, multi-scale representation, and imbalance-aware objective adaptation.

2 Model Architecture

DAMMGA-Net follows an encoder-decoder paradigm similar to the classical U-Net, but introduces a set of structurally coherent modules that jointly enhance directional modeling, hierarchical feature representation, and multi-scale robustness. As illustrated in Figure 1, the network consists of four encoding stages and four decoding stages, where multi-level feature alignment and aggregation are employed to preserve spatial resolution and facilitate the faithful reconstruction of fine vascular structures. To maintain a lightweight backbone, the initial number of feature channels is set to 16 and doubles with depth (16/32/64/128/256), in contrast to the canonical U-Net that typically starts from 64 channels.

Instead of relying solely on convolutional operations with locality-constrained receptive fields, DAMMGA-Net incorporates three key components at strategic depths of the network: the Hierarchical Feature Encoding Module (HFEM), the Direction-Aware Mixed Attention Module (DAMM), and the Multi-scale Fixed Receptive-field Parallel Pyramid (MFRPP). In the encoder, HFEM progressively extracts low- to high-level semantic features while preserving fine structural cues. These enriched encoder features are further enhanced by DAMM, which is inserted along the skip pathways to model directional dependencies, pixel-wise discrimination, and channel-wise contextual relevance in a unified mixed-attention formulation, thereby improving the continuity modeling of thin, tortuous, and branching vessels.

Within the decoder, feature maps are gradually upsampled and fused with their encoder counterparts to recover vessel details at multiple resolutions. To mitigate artifacts and information bottlenecks commonly observed in naive upsam-

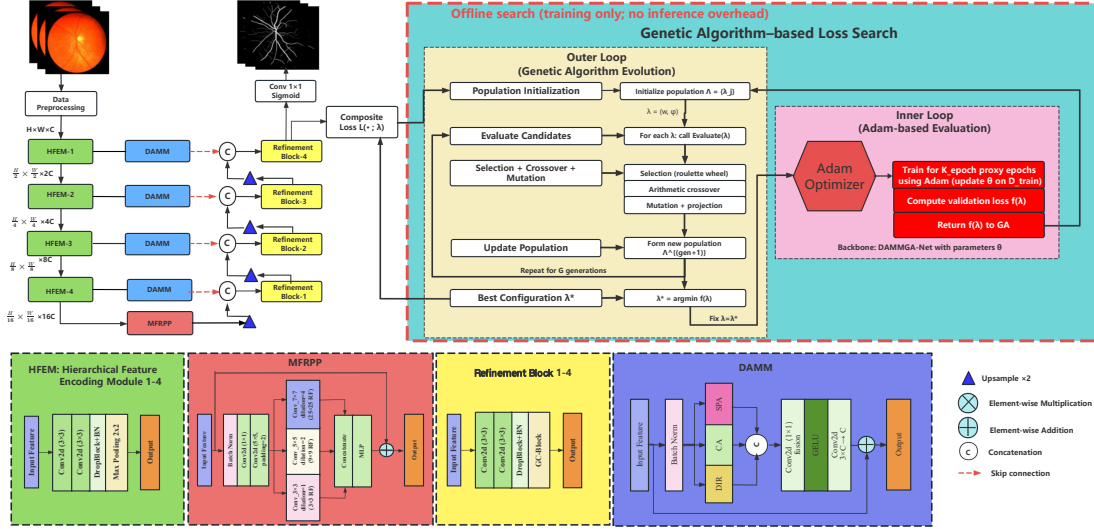


Figure 1: Overall architecture of the proposed DAMMGA-Net. The network is built on a lightweight encoder–decoder backbone consisting of HFEM-based encoding and decoder-side Refinement Blocks. On this backbone, DAMM and MFRPP are introduced as the two main architectural modules for direction-aware modeling and multi-scale aggregation, respectively. GA-Loss is an offline loss-search strategy used during training to adapt the optimization objective to each dataset.

pling, a dedicated Refinement Block is introduced after each fusion step for feature refinement. Meanwhile, unlike designs that repeatedly deploy multi-scale context modules at multiple depths, MFRPP is applied only once at the encoder output (i.e., the bottleneck), following prior end-stage multi-scale aggregation strategies such as ASPP- or DAC-style bottleneck modules [20, 21, 22]. Through three parallel atrous convolution branches with fixed dilation configurations, MFRPP enhances robustness to vessel caliber variations while introducing only limited computational overhead.

Built on a lightweight backbone with HFEM-based encoding and decoder-side Refinement, DAMMGA-Net integrates DAMM and MFRPP as its two architectural innovations to address directional continuity and multi-scale variability in retinal vessel segmentation. GA-Loss serves as the third main contribution of the proposed method by operating only on the training objective, without introducing any additional network parameters during inference. The effects of DAMM and MFRPP are empirically validated through the ablation and qualitative analyses in Section 4 (Tables 2–3 and Figures 10–11, 16–18).

2.1 Direction-Aware Mixed Attention Module (DAMM)

Retinal vessels exhibit highly anisotropic morphology, varying contrast, and fine-scale branching patterns. These properties require a representation mechanism capable of capturing direction-aware continuity, pixel-level discriminability, and channel-wise semantic relevance simultaneously. Traditional convolution layers or single-dimension attention mechanisms alone cannot satisfy these requirements, often leading to fragmented representations of curved or low-contrast vessels.

To address this challenge, we design the *Direction-Aware Mixed Attention Module (DAMM)*, which provides a unified formulation for directional pattern extraction, spatial saliency estimation, and channel-wise feature recalibration. DAMM achieves this through three lightweight and complementary branches: the *Directional Feature Branch (DIR)*, the *Saliency-Preserving Attention Branch (SPA)*, and the *Channel Attention Branch (CA)*. Each branch focuses on a distinct aspect of vascular characteristics, and their integration offers a more holistic and structurally aligned feature representation.

As shown in Figure 2, the three branches operate in parallel and generate feature representations with different inductive biases. Their outputs are concatenated along the channel dimension and subsequently fused using a point-wise MLP (PWConv-GELU-PWConv). A residual connection is finally applied to preserve the original feature semantics while enhancing vessel-relevant information. Formally, the DAMM output is expressed as:

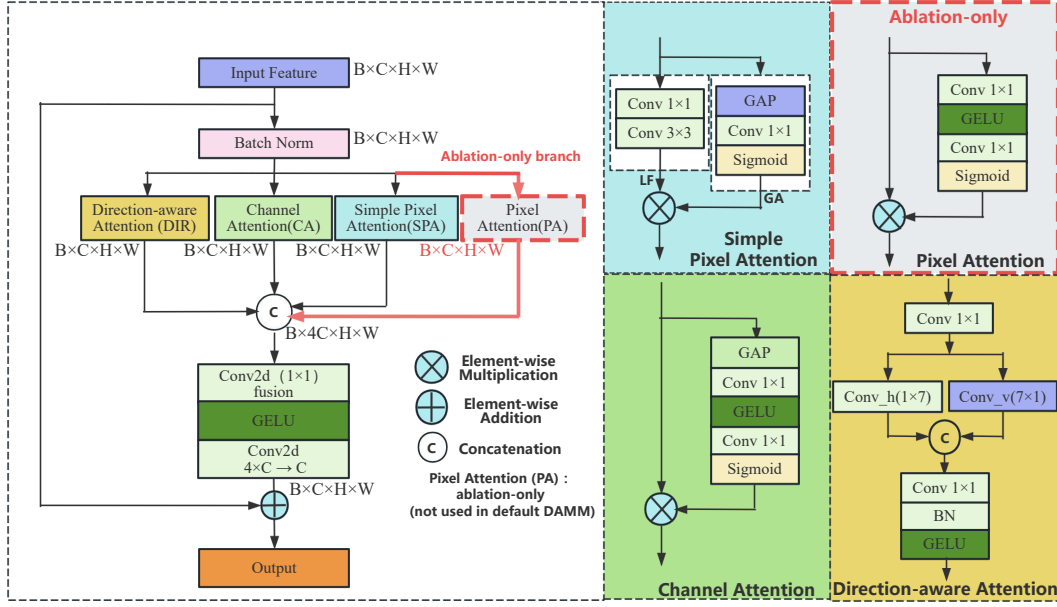


Figure 2: Structural diagram of the Direction-Aware Mixed Attention Module (DAMM). DAMM integrates three complementary branches: the DIR branch employs orthogonal convolutions to model vascular continuity and bifurcation; the CA branch enhances inter-channel feature selectivity with a lightweight GELU-based design; and the SPA branch introduces a simplified local-global gating mechanism for pixel-wise attention. Compared with CBAM [23] and SENet [24], these modifications highlight DAMM’s task-oriented innovations for vascular segmentation. The PA branch is shown only for ablation completeness and is not included in the default DAMM configuration (SPA+CA+DIR); quantitative results are reported in Section 4.3.2 (Table 3).

$$\text{DAMM}(x) = x + \text{MLP}(\text{Concat}[\text{DIR}(x), \text{SPA}(x), \text{CA}(x)]). \quad (1)$$

This mixed-attention formulation allows DAMM to simultaneously encode orientation cues, local saliency, and global channel dependencies. By design, it provides direction-sensitive responses essential for continuous vessel tracking, while remaining lightweight enough for integration into multi-stage feature processing pipelines.

2.1.1 Directional Feature Pathway (DIR)

Blood vessels exhibit prominent tubular morphology and directional distribution, which cannot be explicitly modeled using standard convolution operations alone. To address this, we introduce a directional feature branch (DIR) that employs orthogonal strip-shaped convolution kernels of sizes 1×7 and 7×1 , designed to capture horizontal and vertical vessel features, respectively.

The underlying rationale is that horizontal convolutions are sensitive to vessels oriented along the x -axis, while vertical convolutions respond to vessels aligned along the y -axis. The outputs of these two branches are concatenated and subsequently fused via an additional 1×1 convolution layer, followed by Batch Normalization (BN) and a nonlinear activation function to yield the final directional features.

This choice also follows the observation that overly large kernels may yield diminishing returns in large-kernel ablations (e.g., LitePose) [25], and 7×7 convolutions have been widely adopted as a practical compromise in attention or aggregation modules for retinal vessel segmentation [26]. We further justify this setting via strip-kernel ablations reported in Table S1.

The architecture of the DIR branch is illustrated in Figure 3, and the corresponding computation is formally expressed as:

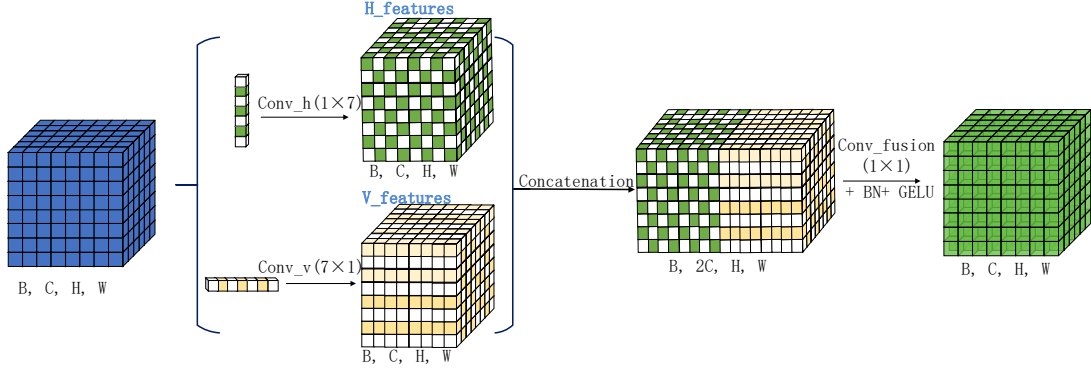


Figure 3: Directional Feature Pathway (DIR). Orthogonal strip-shaped convolutions (1×7 and 7×1) extract orientation-sensitive responses associated with transverse and longitudinal vascular trajectories. The outputs are fused through a 1×1 convolution followed by BN and GELU to produce the final direction-aware representation.

$$\text{DIR}(x) = \text{GELU}(\text{BN}(W_{\text{fusion}}(\text{Concat}[W_h(x), W_v(x)]))). \quad (2)$$

where W_h and W_v denote the 1×7 and 7×1 convolution operations, respectively, and W_{fusion} represents the 1×1 convolution used for feature fusion.

To provide an explicit empirical justification, we systematically evaluate strip-kernel sizes $k \in \{3, 5, 7, 9\}$ and multi-kernel combinations (5,7) and (3,5,7) under identical settings (Table S1). The results confirm that $k = 7$ provides a stable performance–efficiency trade-off, whereas larger kernels or multi-kernel combinations may bring diminishing returns and potential redundancy. Therefore, we adopt $1 \times 7/7 \times 1$ as the default choice, while treating multi-kernel variants as optional settings when higher sensitivity is preferred.

2.1.2 Saliency-Preserving Attention (SPA)

Small retinal vessels often exhibit low contrast and appear as weak signals easily dominated by surrounding background structures. To improve the preservation of these subtle vascular patterns, we introduce the *Saliency-Preserving Attention (SPA)* branch, a lightweight local-global gating mechanism designed to enhance pixel-level vascular saliency.

As shown in Figure 4, SPA combines two complementary components: (1) a local convolutional pathway composed of a 1×1 convolution followed by a 3×3 convolution to extract fine-grained texture responses, and (2) a global response pathway generated by applying global average pooling (GAP) followed by a 1×1 convolution. The two feature representations are fused via element-wise multiplication to yield a saliency-preserving attention map:

$$\text{SPA}(x) = W_v(x) \otimes W_g(x), \quad (3)$$

where $W_v(x)$ models local structural cues and $W_g(x)$ provides global context modulation, and $\otimes$ denotes element-wise multiplication.

Compared with the spatial attention mechanism in CBAM [23], SPA offers three key advantages tailored for retinal vessel segmentation:

- (1) **Saliency-oriented design:** By using GAP-based gating, SPA selectively amplifies weak vessel signals while suppressing background noise.
- (2) **Reduced computational overhead:** The global pathway compresses contextual information to a single scalar per channel, avoiding high-cost spatial convolutions.
- (3) **Better adaptability to fine vessels:** The local-global interaction effectively highlights tiny vascular structures that standard spatial attention mechanisms may overlook.

This design enables SPA to preserve subtle vascular details with minimal complexity, making it well suited for medical image segmentation scenarios characterized by fine structures and low foreground-background contrast.

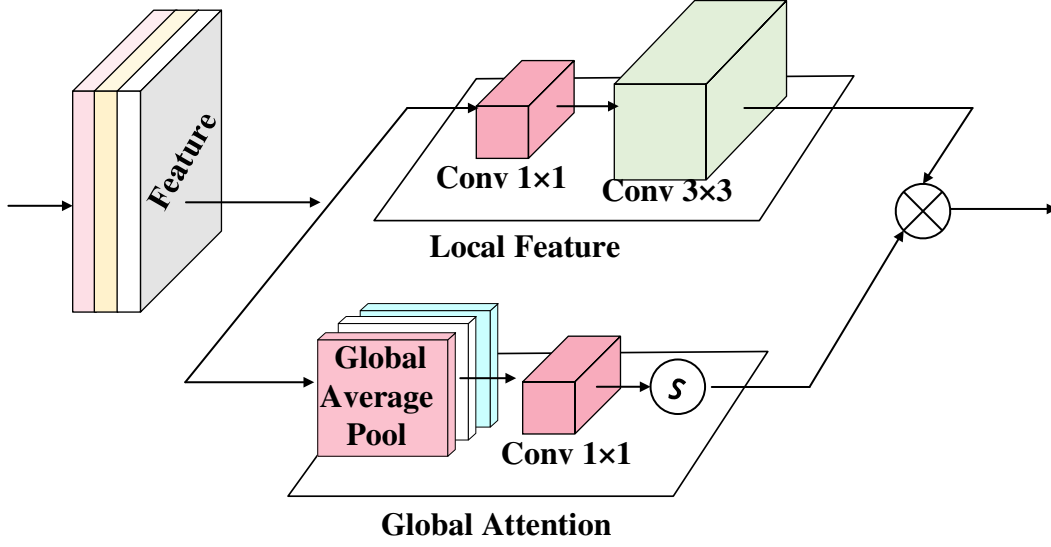


Figure 4: Architecture of the Saliency-Preserving Attention (SPA) branch. The SPA branch generates a pixel-level saliency map by combining local convolutional responses with a global average-guided gating mechanism, selectively enhancing vascular signal preservation while maintaining low computational cost.

2.1.3 Channel Recalibration Pathway (CA)

Vascular patterns distribute unevenly across channels, necessitating selective enhancement of task-relevant feature dimensions. The Channel Recalibration Pathway (CA) extracts global channel descriptors via GAP, then models inter-channel interactions through two 1×1 convolutions with GELU activation:

$$CA(x) = \sigma(W_2(\text{GELU}(W_1(\text{GAP}(x)))))) \otimes x, \quad (4)$$

where σ is the sigmoid gating function.

Compared with SENet [24], CA:

- replaces fully connected layers with 1×1 convolutions for efficiency;
- leverages GELU for smoother gradients and better modeling of subtle vascular patterns;
- provides a lightweight recalibration mechanism well-suited for medical segmentation.

2.1.4 Fusion Strategy

The outputs of the DIR, SPA, and CA pathways form complementary orientation-aware, saliency-enhanced, and channel-refined representations. These are concatenated and passed through a point-wise MLP (PWConv-GELU-PWConv) to yield a unified mixed-attention output:

$$\text{DAMM}(x) = x + W_2(\text{GELU}(W_1(\text{Concat}[\text{DIR}(x), \text{SPA}(x), \text{CA}(x)]))). \quad (5)$$

This fusion strategy preserves heterogeneous cues while avoiding the information bottlenecks found in serial attention mechanisms. Ablation studies confirm that the complementarity of the three pathways—not the number of pathways—is essential to DAMM’s effectiveness.

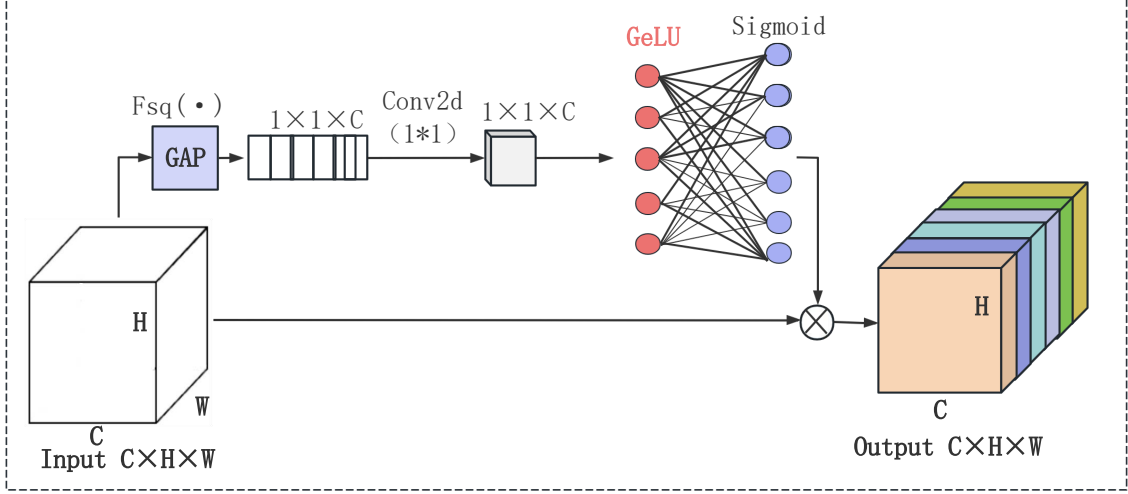


Figure 5: Architecture of the channel attention (CA) branch. Global channel descriptors are first obtained via global average pooling (GAP). They are then passed through two successive 1×1 convolutions with a GeLU non-linearity, instead of ReLU, to provide smoother channel-wise gating and stabilize optimization. Finally, a sigmoid activation produces channel-wise weights that recalibrate the input feature map.

2.2 Multi-scale Fixed Receptive-field Parallel Pyramid (MFRPP)

Retinal vessel segmentation requires the simultaneous modeling of vessels across a wide range of scales, from large trunks to fine capillaries. To enhance multi-scale feature representation while maintaining computational efficiency, we propose a lightweight Multi-scale Fixed Receptive-field Parallel Pyramid (MFRPP) module. The architecture is illustrated in Figure 6.

MFRPP consists of three parallel atrous convolution branches, each configured with a fixed dilation rate $d = \{1, 2, 4\}$, corresponding to different receptive field sizes. Specifically:

- (1) The small receptive field branch (3×3 , $d = 1$) captures fine-grained vessels and texture boundaries;
- (2) The medium receptive field branch (5×5 , $d = 2$) is suited for vessels of intermediate caliber;
- (3) The large receptive field branch (7×7 , $d = 4$) models the global structure of large trunk vessels.

The distribution of effective receptive fields across these branches is shown in Figure 7.

Empirical results show that while adaptive dilation strategies offer flexibility, they may lead to training instability, particularly on datasets such as CHASE_DB1. In contrast, the fixed dilation configuration ensures more stable convergence and is better aligned with the anatomical variability in vessel calibers inherent to retinal imaging tasks.

Each branch is composed of a convolutional layer followed by Batch Normalization (BN) and ReLU activation. The outputs of the three branches are concatenated along the channel dimension and then fused by a point-wise MLP, which consists of two 1×1 convolutions with a GELU activation in between, resulting in a multi-scale context-enhanced feature representation. The fusion process is defined as:

$$\text{MFRPP}(x) = W_2 \left(\text{GELU} \left(W_1 \left(\text{Concat} [\text{Conv}_{3 \times 3, d=1}(x), \text{Conv}_{5 \times 5, d=2}(x), \text{Conv}_{7 \times 7, d=4}(x)] \right) \right) \right). \quad (6)$$

Here, W_1 and W_2 denote the two point-wise 1×1 convolution layers in the MLP.

By leveraging this parallel architecture with fixed receptive fields, MFRPP effectively enhanced the network’s ability to represent vessel structures across varying spatial scales.

Placement rationale. Although the hierarchical encoder provides implicit multi-scale cues through downsampling, such “resolution-scale” variation may weaken fine capillaries when semantic abstraction increases. MFRPP is therefore designed to complement the encoder by explicitly constructing “receptive-field-scale” diversity at the same semantic level using fixed dilation rates ($d = 1, 2, 4$). We place MFRPP only at the encoder output (bottleneck), where features are semantically rich and channel capacity is sufficient for robust multi-scale fusion. Importantly, this design

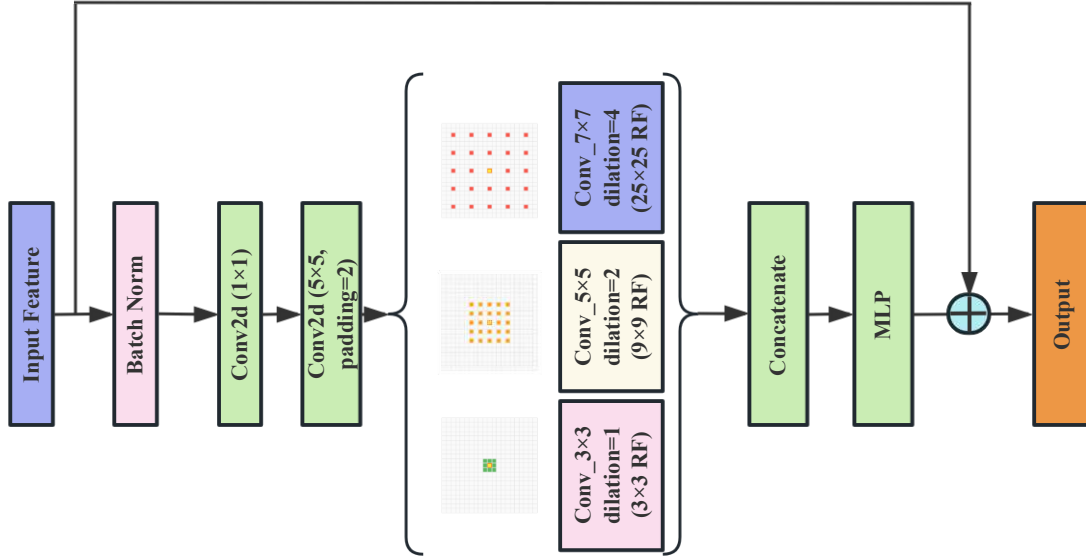


Figure 6: Schematic diagram of the Multi-scale Fixed Receptive-field Parallel Pyramid (MFRPP) structure. The module comprises three parallel branches, each employing convolutional kernels with fixed dilation rates of $d = 1, 2$, and 4 (corresponding to 3×3 , 5×5 , and 7×7 kernels) to capture vascular features across different spatial scales. Outputs from the three branches are concatenated along the channel dimension and then fused by a point-wise MLP, which consists of two 1×1 convolutions with a GELU activation in between, to produce multi-scale context-enhanced representations. This design effectively addresses scale variance in retinal vessel segmentation while maintaining computational efficiency.

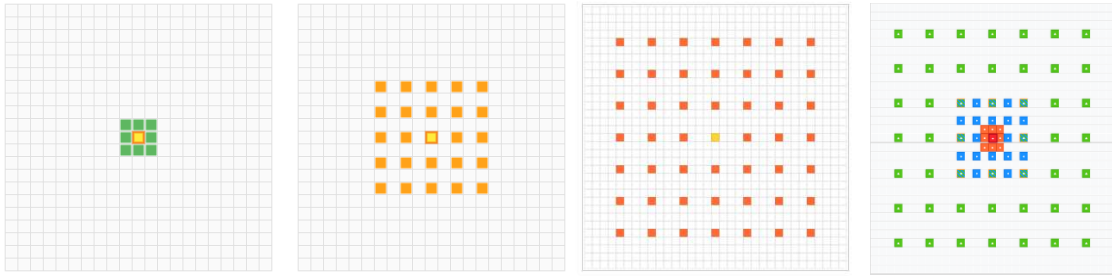


Figure 7: Receptive field visualization of the parallel branches in the MFRPP module. From left to right, the subfigures represent the small (3×3), medium (9×9), and large (25×25) receptive fields, as well as their composite multi-scale receptive field. The fixed dilation rate configuration ensures training stability and enhances the model's ability to represent vessels of varying scales.

follows a widely adopted paradigm in segmentation: DeepLabv3/DeepLabv3+ performs atrous operations mainly in later backbone stages and aggregates multi-scale context via an end-stage ASPP head rather than applying atrous convolutions uniformly at every stage [20]; similarly, U-Net variants such as CE-Net and DoubleU-Net introduce atrous-based context modules (e.g., DAC/ASPP) at the bottleneck to enhance multi-scale representation [21, 22]. In contrast, employing atrous convolutions at every encoder layer may accumulate computation and increase the risk of gridding artifacts. This is particularly detrimental for thin vessels, because their cross-sectional width spans only a few pixels and their connectivity relies on dense local sampling; gridding artifacts may create periodic blind spots in the receptive field, leading to fragmented responses and broken vessel segments.

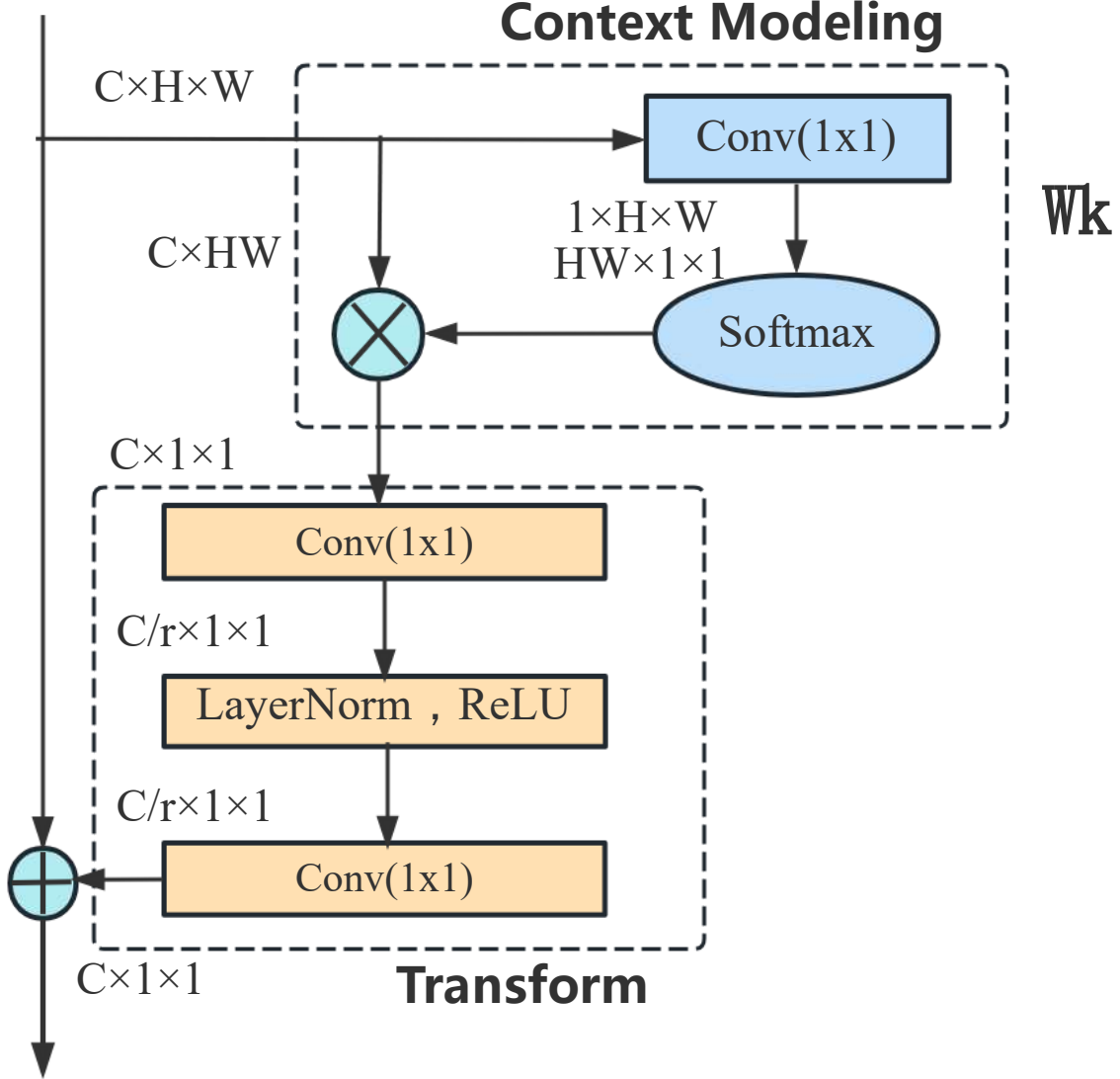


Figure 8: Schematic illustration of the Global Context (GC) Block. The module consists of two components: context modeling and feature transformation. The former applies a 1×1 convolution followed by softmax to compute spatial attention weights for global context aggregation, while the latter uses 1×1 convolution, normalization, and ReLU activation for nonlinear mapping and integrates the result with the original input. This design complements local receptive fields by capturing long-range dependencies, thereby improving continuity modeling for vessel bifurcations and elongated structures.

2.3 Decoder-Side Refinement Block

The Refinement Block serves as a lightweight feature enhancement module placed at the end of each decoding stage. As illustrated in Figure 1 (yellow region), the block consists of two 3×3 convolutions followed by DropBlock

regularization and Batch Normalization, and concludes with a Global Context (GC) Block for long-range dependency modeling.

The internal mechanism of the GC Block is shown in Figure 8. It aggregates global spatial context using a 1×1 projection and softmax-based attention weighting, and then transforms the aggregated descriptor through a bottleneck structure with LayerNorm, ReLU, and a final 1×1 convolution. By embedding this GC module inside the Refinement Block, our framework effectively complements local texture extraction with global contextual cues, improving continuity at bifurcations and elongated vascular structures.

Formally, given an input feature map x , the Refinement Block is expressed as:

$$\text{Refine}(x) = \text{GC}(\text{DropBlock}(\text{BN}(\text{Conv}_{3 \times 3}(\text{Conv}_{3 \times 3}(x)))). \quad (7)$$

2.4 Backbone Encoder (HFEM)

The Hierarchical Feature Encoding Module (HFEM), illustrated in Figure 1, is designed as a lightweight yet expressive feature extractor that progressively refines structural cues before they are processed by the subsequent attention and multi-scale modules. Unlike conventional encoder blocks that rely heavily on deep stacking or aggressive down-sampling, HFEM emphasizes *hierarchical refinement* through carefully arranged convolutional transformations and spatial regularization.

Specifically, HFEM applies two successive 3×3 convolutional layers to extract low-level edge patterns and intermediate structural textures. These layers work in tandem to stabilize early-stage representations and enhance the visibility of vascular boundaries. To prevent overfitting on limited retinal imaging datasets, DropBlock regularization is inserted after the convolutional operations, jointly with Batch Normalization to maintain stable feature statistics during training.

Following feature refinement, a 2×2 max-pooling operation reduces spatial resolution while preserving key morphological structures, enabling the encoded representation to retain the fundamental topological properties of the vessels—such as branching patterns and local curvature. The resulting hierarchical feature embeddings provide a powerful foundation for subsequent encoder-decoder processing, wherein DAMM is applied along the skip-connection feature paths for direction-aware enhancement, MFRPP is introduced at the bottleneck for multi-scale contextual aggregation, and decoder-side refinement is employed for progressive feature reconstruction.

Overall, HFEM acts as an effective front-end encoder that balances detail preservation, hierarchical abstraction, and regularization, providing stable and representationally rich features for the DAMMGA-Net architecture.

In summary, the architectural components introduced above collectively form the core representation stream of DAMMGA-Net, enabling enhanced modeling of multi-scale structures, directional vascular continuity, and global contextual interactions.

However, the segmentation performance of retinal vessels is not solely determined by architectural capacity. The inherent foreground-background imbalance and the fragility of fine-scale vessels make the training objective equally critical. To ensure stable optimization and improve the recall of tiny or low-contrast vessels, DAMMGA-Net incorporates a lightweight genetic-algorithm-driven loss optimization strategy (GA-Loss).

The following section formalizes this optimization component by defining the composite-loss search space, the evolutionary pipeline, and its integration with the proposed architecture.

3 Algorithm Design

While the lightweight backbone and the two architectural modules introduced in Section 2 define the representation pathway of DAMMGA-Net, retinal vessel segmentation performance is also strongly influenced by the optimization objective. In particular, severe foreground-background imbalance and the fragile continuity of fine vascular structures may still lead to unstable training dynamics, even when the network architecture is sufficiently expressive. Therefore, in addition to the two architectural contributions (DAMM and MFRPP), we introduce a lightweight genetic-algorithm-driven loss optimization strategy, termed GA-Loss, to complement the representational capability of DAMMGA-Net under class-imbalanced conditions. As the optimization-side contribution of the proposed framework, GA-Loss automatically searches for effective composite-loss configurations and their internal parameters, thereby reducing dependence on manual tuning and improving the preservation of small vessels and low-contrast structures.

The remainder of this section formalizes the GA-Loss formulation, describes the bi-level optimization process, and outlines the evolutionary search procedures employed in DAMMGA-Net.

3.1 Problem Formulation

We formulate loss function optimization as a bi-level problem under class imbalance. Let the base loss set be $\{L_i(\theta; \phi_i)\}_{i=1}^m$, where $m = 4$ in this work (Binary Cross-Entropy, Dice, Focal, and Tversky). Here, ϕ_i denotes the internal parameters of the i -th loss. Specifically, for Focal loss we use (α_f, γ_f) , and for Tversky loss we use $(\alpha_{tv}, \beta_{tv})$.

Let $\mathbf{w} = [w_1, \dots, w_m]^\top$ denote the weight vector subject to the constraints:

$$w_i \geq 0, \quad \sum_{i=1}^m w_i = 1. \quad (8)$$

Each candidate configuration is then defined as:

$$\lambda = (\mathbf{w}, \phi), \quad \phi = \{\phi_i\}_{i=1}^m, \quad \phi \in [\phi_{\min}, \phi_{\max}]. \quad (9)$$

The composite loss function becomes:

$$\mathcal{L}(\theta; \lambda) = \sum_{i=1}^m w_i \cdot L_i(\theta; \phi_i). \quad (10)$$

We adopt a two-level optimization strategy:

$$\theta^*(\lambda) \approx \text{Adam}_T(\theta_0; \lambda). \quad (11)$$

$$f(\lambda) = \frac{1}{|\mathcal{D}_{val}|} \sum_{(x,y) \in \mathcal{D}_{val}} \mathcal{L}(\theta^*(\lambda); \lambda). \quad (12)$$

$$\lambda^* = \arg \min_{\lambda} f(\lambda). \quad (13)$$

Here, θ_0 is the common initialization, K_{epoch} denotes the number of proxy-training epochs in each fitness evaluation, and T denotes the corresponding total number of mini-batch Adam updates used to implement one such evaluation in Algorithm 2. Thus, the search budget is specified at the epoch level via K_{epoch} , while the inner optimization is written at the update-step level via $\text{Adam}_T(\cdot)$. The fitness score $f(\lambda)$ is evaluated on the validation set.

Validation loss is used as the search fitness rather than threshold-dependent segmentation metrics, because each candidate is evaluated after only a short proxy training stage ($K_{\text{epoch}} = 15$ epochs), under which validation loss provides a denser and more stable ranking signal. In contrast, F1-score and sensitivity are reported only for the final evaluation after the selected optimal loss configuration has been fully retrained.

3.2 Search Framework Overview

The outer search loop follows a classical Genetic Algorithm (GA), where each individual corresponds to a candidate configuration $\lambda = (\mathbf{w}, \phi)$. The overall procedure is:

1. **Initialization:** A population $\Lambda_0 = \{\lambda^{(1)}, \dots, \lambda^{(N)}\}$ is randomly initialized, with $\mathbf{w} \in \Delta^3$ (non-negative and summing to 1) and ϕ_i sampled from feasible ranges $[\phi_{\min}, \phi_{\max}]$.
2. **Evaluation:** For each $\lambda^{(i)}$, we train a segmentation model using $\mathcal{L}(\cdot; \lambda^{(i)})$ for K_{epoch} proxy-training epochs, and compute $f(\lambda^{(i)}) = \text{ValLoss}(\lambda^{(i)})$.
3. **Selection and Variation:** - Elitism: top- E individuals are preserved. - Parent selection: performed via roulette wheel with smoothing $\delta > 0$:

$$p_i \propto \frac{1}{f(\lambda_i) + \delta}.$$

- Crossover: offspring are generated via arithmetic crossover:

$$\lambda_c = \beta \lambda_{p1} + (1 - \beta) \lambda_{p2}, \quad \beta \in [0, 1].$$

- Mutation: Gaussian mutation is applied to both weights and internal parameters:

$$\mathbf{w} \leftarrow \mathbf{w} + \mathcal{N}(0, \sigma_w^2). \quad (14)$$

$$\phi \leftarrow \phi + \mathcal{N}(0, \sigma_\phi^2). \quad (15)$$

with random toggling probability p_{toggle} for sparsity.

- Projection and clipping: after mutation,

$$\mathbf{w} \leftarrow \text{PROJECT-TO-SIMPLEX}(\max(0, \mathbf{w})). \quad (16)$$

$$\phi \leftarrow \text{CLIP}(\phi; \phi_{\min}, \phi_{\max}). \quad (17)$$

4. **Termination:** Repeat for G generations or until convergence (in practice we used a fixed maximum of $G=20$ generations). The final output is

$$\lambda^* = \arg \min_{\lambda \in \Lambda} f(\lambda).$$

This design decouples the loss function search from gradient-based optimization and enables flexible integration of domain-specific priors. Empirically, it enables stable and interpretable discovery of loss formulations tailored to class imbalance and topological preservation in vessel segmentation tasks.

Search protocol and counts: Each candidate is trained for $K_{\text{epoch}} = 15$ proxy-training epochs. The outer loop runs for $G = 20$ generations with population size $N = 5$, and all N candidates are evaluated once per generation. Therefore, the total number of fitness evaluations is $N_{\text{eval}} = G \times N = 100$ (for $G = 10$, $N_{\text{eval}} = 50$). Hence, the total number of proxy-training epochs is $K_{\text{epoch}} \times N_{\text{eval}} = 1,500$. The best configuration is then fully retrained for 70 epochs before the final evaluation.

Definition of a fitness evaluation: An evaluation (fitness evaluation) denotes one execution of Algorithm 2 for a single candidate loss configuration, that is, training the candidate for K_{epoch} proxy-training epochs under fixed initialization, data split, and training settings, and then computing its validation fitness for selection, crossover, and mutation in the outer loop. In Algorithm 2, this proxy-training process is written at the mini-batch level using the corresponding T gradient-update steps.

GA-Loss search settings. In the reported experiments, the genetic search adopts a population size of 5 for 20 generations, with mutation rate set to 0.30, crossover rate set to 0.50, elite size set to 1, and roulette-wheel selection used for parent sampling. The search space includes Focal loss parameters $\gamma \in [1, 5]$ and $\alpha \in [0.1, 0.9]$, as well as Tversky loss parameters $\alpha, \beta \in [0.1, 0.9]$, while the smoothing term is fixed to 1.0 for Dice and Tversky losses. The active loss weights are constrained to be non-negative and normalized such that their sum equals 1. Under the budgeted search setting, the fitness of each candidate is defined as the final validation loss after $K_{\text{epoch}} = 15$ epochs of quick training. To control the search cost, the GA search is performed once using a fixed random seed (42) to determine the loss configuration, and the selected configuration is then retrained and evaluated under multiple training seeds as described in Section 4.

3.3 Input Parameters and Definitions

Table 1 summarizes the input parameters and definitions used in GA-Loss.

3.4 Pseudocode of the Search Framework

To facilitate understanding of the composite loss search framework, Table 1 summarizes the core variables and symbols used throughout the evolutionary optimization process. These parameters are shared across the five modular procedures, namely Initialization, Evaluation, Selection, Crossover, and Mutation, as detailed in Algorithm 3 to Algorithm 7 (see Appendix A.1 for pseudocode implementations).

4 Experimental Results and Analysis

To validate the independent and joint contributions of DAMM and MFRPP (Section 2) and GA-Loss (Section 3), the evaluation was organized as follows: module-level ablations, branch-level ablations, comparisons with state-of-the-art methods, and qualitative visual analyses.

The proposed algorithm was implemented using the PyTorch framework and evaluated on a system equipped with an NVIDIA GeForce RTX 4060 GPU (8 GB VRAM) and CUDA version 11.8.

Training hyperparameters and preprocessing. We use the Adam optimizer with a learning rate of 5×10^{-4} and a weight decay of 1×10^{-4} . The batch size is set to 1, and the final training is performed for 70 epochs [27]. The backbone starts with 16 channels and doubles with depth, i.e., 16/32/64/128/256. DropBlock is applied with `keep_prob=0.87` and `block_size=7`, and deterministic settings of the NVIDIA CUDA Deep Neural Network library (cuDNN) were

Table 1: Input Parameters and Definitions in Composite Loss Search

Symbol	Name	Meaning	Range / Notes
$\mathcal{D}_{\text{train}}, \mathcal{D}_{\text{val}}, \mathcal{D}_{\text{test}}$	Datasets	Train/validation/test splits	Disjoint; no leakage
θ_0	Init params	Common initialization for each evaluation	Same seed/order across candidates
N	Population size	#candidates per generation	Integer; e.g., 5–20; $N > E$
E	Elites	#best individuals carried over	$1 \leq E < N$
G	Max generations	Outer-loop termination	Early stopping optional
N_{eval}	Total evaluations	Total number of fitness evaluations in the outer loop	$N_{\text{eval}} = G \times N$
K_{epoch}	Proxy epochs	Number of proxy-training epochs per fitness evaluation	Fixed across candidates; e.g., 15
T	Gradient steps	Total number of mini-batch Adam updates in Algorithm 2	Determined by K_{epoch} and the number of mini-batches per epoch
$\alpha, \beta_1, \beta_2, \varepsilon$	Adam hyper-params	LR, momentum coeffs, stabilizer	Typical $\beta_1=0.9, \beta_2=0.999$
δ	Selection smoothing	Stabilizer for roulette selection	$\delta > 0$
$[\phi_{\min}, \phi_{\max}]$	Bounds	Feasible domain of internal loss params	Per-loss; e.g., Focal (α_f, γ_f), Tversky (α_{tv}, β_{tv})
σ_w, σ_ϕ	Mutation scales	Std. dev. of Gaussian noise on $\mathbf{w}, \phi$	Tunable
p_{toggle}	Toggle prob.	Randomly set some w_i to zero (sparsify)	Small prob., e.g., ≤ 0.2 ; followed by projection
$\lambda = (\mathbf{w}, \phi)$	Configuration	Loss weights and internal parameters	$\mathbf{w} \in \Delta^3, \sum_i w_i = 1$
$f(\lambda)$	Fitness	Mean validation loss under λ	Minimize; forward only on $\mathcal{D}_{\text{val}}$

Algorithm 1 GA-Driven Composite Loss Search (Outer Loop)

Require: $\mathcal{D}_{\text{train}}, \mathcal{D}_{\text{val}}, \theta_0$; population size N , elites $E \geq 1$, max generations G ; proxy epochs K_{epoch} ; Adam hyper-parameters $(\alpha, \beta_1, \beta_2, \varepsilon)$; bounds $[\phi_{\min}, \phi_{\max}]$; selection smoothing $\delta > 0$; mutation scales σ_w, σ_ϕ ; toggle prob. p_{toggle} .

Ensure: Best configuration λ^* . $\triangleright$ sample $\lambda = (\mathbf{w}, \phi)$; $\mathbf{w} \geq 0, \sum \mathbf{w} = 1, \phi \in [\phi_{\min}, \phi_{\max}]$

```

1:  $\Lambda \leftarrow \text{INITIALIZE POPULATION}(N)$ 
2: for gen = 1 to  $G$  do
3:   for all  $\lambda \in \Lambda$  do
4:      $f(\lambda) \leftarrow \text{EVALUATE}(\lambda, \theta_0, K_{\text{epoch}}, \alpha, \beta_1, \beta_2, \varepsilon, \mathcal{D}_{\text{train}}, \mathcal{D}_{\text{val}})$   $\triangleright$  Algorithm 2
5:   end for
6:    $\Lambda_{\text{elite}} \leftarrow \text{TOP } E \text{ INDIVIDUALS}(\Lambda, f, E)$ 
7:   Define selection probabilities  $p_i \propto 1/(f(\lambda_i) + \delta)$  (roulette wheel)  $\triangleright$  Algorithm 7
8:    $\Lambda_{\text{next}} \leftarrow \emptyset$ 
9:   while  $|\Lambda_{\text{next}}| < N - E$  do
10:     $j_1 \leftarrow \text{ROULETTE WHEEL}(\{f(\lambda)\}, \delta)$   $\triangleright$  Algorithm 7
11:     $j_2 \leftarrow \text{ROULETTE WHEEL}(\{f(\lambda)\}, \delta)$   $\triangleright$  Algorithm 7
12:     $\lambda^{(p1)} \leftarrow \Lambda[j_1], \lambda^{(p2)} \leftarrow \Lambda[j_2]$ 
13:     $\lambda^{(c)} \leftarrow \text{ARITHMETIC CROSSOVER}(\lambda^{(p1)}, \lambda^{(p2)})$   $\triangleright$  Algorithm 3
14:     $\lambda^{(c)} \leftarrow \text{MUTATE}(\lambda^{(c)}; \sigma_w, \sigma_\phi, p_{\text{toggle}})$   $\triangleright$  Algorithm 4
15:     $\mathbf{w}^{(c)} \leftarrow \text{PROJECT TO SIMPLEX}(\max(0, \mathbf{w}^{(c)}))$   $\triangleright$  Algorithm 5
16:     $\phi^{(c)} \leftarrow \text{CLIP}(\phi^{(c)}; \phi_{\min}, \phi_{\max})$   $\triangleright$  Algorithm 6
17:     $\Lambda_{\text{next}} \leftarrow \Lambda_{\text{next}} \cup \{\lambda^{(c)}\}$ 
18:   end while
19:    $\Lambda \leftarrow \Lambda_{\text{elite}} \cup \Lambda_{\text{next}}$ 
20: end for
21: return  $\lambda^* = \arg \min_{\lambda \in \Lambda} f(\lambda)$ 

```

Algorithm 2 EVALUATE(λ): Adam Inner Loop and Validation Fitness

Require: Candidate $\lambda = (w, \phi)$; initialization θ_0 ; proxy epochs K_{epoch} ; Adam $(\alpha, \beta_1, \beta_2, \varepsilon)$; data $\mathcal{D}_{\text{train}}, \mathcal{D}_{\text{val}}$.

Ensure: Fitness $f(\lambda)$.

- 1: Fix random seed; $\theta \leftarrow \theta_0$; reset Adam state $(m, v) \leftarrow (0, 0)$
 - 2: Let T denote the total number of mini-batch Adam updates induced by iterating over $\mathcal{D}_{\text{train}}$ for K_{epoch} proxy-training epochs
 - 3: **for** step = 1 to T **do**
 - 4: Sample mini-batch $\mathcal{B} \subset \mathcal{D}_{\text{train}}$
 - 5: $L_{\text{step}} \leftarrow \sum_{i=1}^4 w_i L_i(\theta; \phi_i | \mathcal{B})$
 - 6: $\theta \leftarrow \text{ADAM UPDATE}(\theta, \nabla_{\theta} L_{\text{step}}; \alpha, \beta_1, \beta_2, \varepsilon)$
 - 7: **end for**
 - 8: $f(\lambda) \leftarrow \text{MEAN LOSS OVER}(\mathcal{D}_{\text{val}}, \theta, \lambda)$
 - 9: **return** $f(\lambda)$
-

used to improve reproducibility. For CHASE_DB1, the original images (999×960) and the corresponding masks are first symmetrically zero-padded and then resized to a fixed resolution of 1008×1008 for both training and evaluation.

Random seeds and reporting protocol. Unless otherwise stated, all reproduced experiments are evaluated using three independent training seeds (11, 21, and 42). For readability, the main-text tables report averaged values, whereas variability (mean $\pm$ std) is explicitly provided in the supplementary ablation tables where relevant. For GA-Loss, the search stage is performed once with a fixed random seed (42) to obtain the loss configuration, and the selected configuration is then fully retrained and evaluated under the same three training seeds for final reporting.

Efficiency profiling. Params denote the number of trainable parameters. FLOPs were profiled under a fixed input size using the same profiling tool for all methods. Unless otherwise stated, FLOPs were profiled using the actual model input setting. For CHASE_DB1, the input tensor was set to $1 \times 1 \times 1008 \times 1008$, because the raw RGB fundus image is first converted into a preprocessed single-channel image through weighted channel fusion, normalization, and CLAHE, and the resulting single-channel image is used as the network input. FLOPs are reported as $2 \times \text{MACs}$. All models were trained from scratch without pretraining or distillation, and the same training protocol was used for fair comparison across settings.

GA-Loss searches within a bounded composite-loss family by optimizing only the loss weights and a small set of internal parameters under a fixed budget, which keeps the search cost controllable and reproducible. In contrast, AutoML loss-generation methods often rely on large, unconstrained search spaces and report cluster- and implementation-dependent costs; therefore, we summarize their resource usage and runtime as reported in the original papers (Table S2).

4.1 Dataset Processing

Experiments were conducted on three public datasets: DRIVE [28], CHASE_DB1 [29], and STARE [30]. Figure 9 illustrates the preprocessing pipeline from raw RGB fundus images to the final single-channel network input, and should be interpreted as a preprocessing workflow rather than a network-structure input schematic. Specifically, each original RGB fundus image was first decomposed into three channels and compressed into a single-channel image through weighted channel fusion. The fused image was subsequently normalized to reduce pixel-intensity inconsistencies and inter-image illumination variation, followed by CLAHE (Contrast Limited Adaptive Histogram Equalization) to enhance vessel contrast and suppress uneven illumination. Gamma correction was then applied to further refine the intensity distribution and improve the visibility of vascular structures. The resulting gamma-corrected single-channel image was used as the actual input to the segmentation network.

The CHASE_DB1 dataset consists of 28 fundus images with a resolution of 999×960 pixels, corresponding to the left and right eyes of 14 subjects. To enlarge the training set, rotation and flipping were applied, increasing the number of training images to 72. The STARE dataset comprises 20 fundus images with a resolution of 700×605 pixels, whereas the DRIVE dataset contains 40 fundus images with a resolution of 584×565 pixels. For all three datasets, a unified preprocessing pipeline was employed, in which each original RGB fundus image was compressed into a single-channel image through weighted channel fusion, followed by normalization, CLAHE enhancement, and gamma correction. The resulting gamma-corrected single-channel image was used as the actual input to the segmentation network. No additional augmentation was applied to the STARE or DRIVE training images.

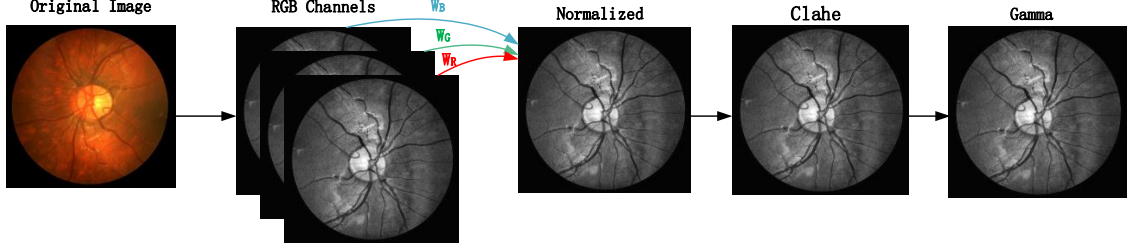


Figure 9: Preprocessing pipeline from raw RGB fundus images to the final single-channel network input [27]. From left to right: (1) the original RGB image is decomposed into three channels; (2) the channels are weighted-fused into a single-channel image, $I_g = w_R R + w_G G + w_B B$ with $w_R + w_G + w_B = 1$ (colored arrows denote the contributions from the red, green, and blue channels, respectively); (3) intensity normalization mitigates inter-image illumination variation; (4) CLAHE (contrast-limited adaptive histogram equalization) enhances vessel contrast and suppresses uneven illumination; (5) gamma correction is further applied to adjust intensity distribution and improve the visibility of vascular structures. The gamma-corrected image is used as the final network input. This figure depicts the preprocessing workflow rather than the network architecture or an input-interface schematic.

4.2 Datasets and Evaluation Metrics

To ensure a fair comparison across models, we used three different random seeds and reported the average performance over runs. Specifically, we evaluated the Area Under the ROC Curve (AUC), F1-score (F1), Sensitivity (Sen), Specificity (Sp), and Accuracy (Acc). F1 was computed using a fixed probability threshold to ensure consistency. We define these metrics as follows:

$$\left\{ \begin{array}{l} \text{Sen} = \frac{TP}{TP + FN}, \\ \text{Sp} = \frac{TN}{TN + FP}, \\ \text{Acc} = \frac{TP + TN}{TP + TN + FP + FN}, \\ F1 = \frac{2TP}{2TP + FP + FN}, \\ \text{AUC} = \int_0^1 TPR(FPR^{-1}(t)) dt. \end{array} \right. \quad (18)$$

In addition to conventional pixel-wise metrics, we also report topology-aware metrics (clDice, $|\Delta\beta_0|$, Rel.- β_0 Err., and $|\Delta\beta_1|$), since retinal vessel segmentation is highly sensitive to connectivity preservation and centerline continuity.

4.3 Ablation Study and Analysis

To assess the contribution of each architectural component and internal design choice, two types of ablations were performed: (1) **module-level ablations**, examining the influence of major components (MFRPP, DAMM, and the internal mechanisms DropBlock and GC Block); and (2) **branch-level ablations** within DAMM, evaluating the DIR, CA, SPA, and PA branches. All experiments were conducted using identical training settings and dataset partitions.

Clarification on Ablation Intent. Although DropBlock and GC Block appear as separate entries in the ablation table, they are *not* external plug-in modules added to the network. Instead, they represent internal alternatives for regularization (DropBlock) and context aggregation (GC Block) within HFEM and the Refinement Block. Their inclusion in the ablation study serves to validate internal architectural design choices, ensuring that the encoder and decoder employ the most effective regularization and global-context modeling mechanisms.

Module-level ablation results indicate that MFRPP improves multi-scale feature representation through stable receptive fields across dilated branches, which is beneficial for the detection of fine vessels while preserving global contextual consistency. DAMM enhances boundary discrimination and directional continuity by integrating orientation-aware, channel-wise, and pixel-level cues. The internal components (DropBlock and GC Block) further support fea-

ture regularization and contextual modeling within the encoder-decoder pathway. Overall, these results suggest that MFRPP and DAMM play complementary roles and serve as the primary task-oriented components of DAMMGA-Net.

Branch-level ablation results provide additional insights into DAMM’s parallel design: SPA improves boundary sharpness, DIR enhances orientation and bifurcation continuity, and CA contributes global semantic stability. Removing any branch leads to degradation in either boundary precision or fine-structure continuity, confirming the necessity of multi-perspective attention.

4.3.1 Module-Level Ablation

Table 2: DAMMGA-Net Module Ablation Study Results

Method	Module Components				Performance Metrics (%)					Topology Metrics			
	MFRPP	DAMM	DropBlock	GC Block	Acc	Sen	Sp	F1	AUC	clDice	$ \Delta\beta_0 $	Rel- β_0 Err	$ \Delta\beta_1 $
<i>Baseline Model</i>													
Baseline	–	–	–	–	97.26	79.08	98.49	78.45	97.82	0.7993	83.50	48.3125	4.00
<i>Single Module Addition</i>													
Baseline + MFRPP	✓	–	–	–	97.43	77.43	98.78	79.16	98.63	0.8184	89.75	52.9688	6.12
Baseline + DAMM	–	✓	–	–	97.42	82.15	98.44	80.02	98.56	0.8171	84.12	51.0000	4.75
Baseline + DropBlock	–	–	✓	–	97.29	79.11	98.51	78.63	97.85	0.8235	60.62	35.9062	6.12
Baseline + GC Block	–	–	–	✓	97.29	80.87	98.45	78.66	98.58	0.8193	67.88	40.0000	5.38
<i>Dual Module Combination</i>													
Baseline + MFRPP + DAMM	✓	✓	–	–	97.27	81.26	98.35	78.96	98.25	0.8195	73.75	43.6562	6.25
Baseline + MFRPP + DropBlock	✓	–	✓	–	97.18	82.50	98.17	78.65	98.23	0.8236	77.88	46.5625	5.12
Baseline + MFRPP + GC Block	✓	–	–	✓	97.32	81.17	98.04	79.23	98.56	0.8282	72.25	44.2500	5.00
Baseline + DAMM + DropBlock	–	✓	✓	–	97.22	85.56	97.98	79.48	98.64	0.8191	71.50	41.7500	5.75
Baseline + DAMM + GC Block	–	✓	–	✓	97.35	81.36	98.43	79.64	98.76	0.8217	63.88	37.5000	6.12
Baseline + DropBlock + GC Block	–	–	✓	✓	97.36	78.41	98.63	78.90	98.55	0.8193	67.88	40.0000	5.38
<i>Triple Module Combination</i>													
Baseline + MFRPP + DAMM + DropBlock	✓	✓	✓	–	97.23	82.30	98.23	78.91	98.38	0.8258	80.50	46.9062	6.25
Baseline + MFRPP + DAMM + GC Block	✓	✓	–	✓	97.36	81.50	98.59	79.61	98.71	0.8145	75.25	42.6250	5.88
Baseline + MFRPP + DropBlock + GC Block	✓	–	✓	✓	97.36	80.78	98.47	79.40	98.63	0.8041	60.62	52.0625	4.500
Baseline + DAMM + DropBlock + GC Block	–	✓	✓	✓	97.37	81.55	98.43	79.63	98.70	0.8251	76.62	46.9688	3.62
<i>Complete Model (All Modules)</i>													
Baseline + MFRPP + DAMM + DropBlock + GC Block	✓	✓	✓	✓	97.37	83.06	98.33	79.92	98.75	0.8238	64.38	37.4062	3.88

Note: 1. Baseline for module-level ablation (Vanilla Backbone): the DAMMGA-Net vanilla backbone, i.e., the plain U-Net backbone without DAMM, MFRPP, GC Block, or DropBlock. 2. Topology metrics include clDice, $|\Delta\beta_0|$, Rel- β_0 Err, and $|\Delta\beta_1|$. 3. “Rel- β_0 Err” denotes the relative error of the zero-dimensional Betti number.

To better align the empirical analysis with the design motivation of MFRPP, we further examine its effect in the module-level ablation study. As shown in Table 2, the effect of MFRPP is evaluated on CHASE_DB1. Relative to the module-level ablation baseline, adding MFRPP increases F1 from 78.45% to 79.16% and AUC from 97.82% to 98.63%. In addition, clDice improves from 0.7993 to 0.8184, indicating better centerline-level overlap and more coherent structural representation. These results suggest that MFRPP mainly benefits scale-aware feature aggregation and structural consistency under the lightweight backbone setting.

To provide module-level qualitative evidence, we further visualize the response distribution of the model with and without MFRPP using activation heatmaps. As shown in Fig. 10, the variant with MFRPP exhibits stronger and more spatially coherent responses along vessel regions within the focus mask, while the counterpart without MFRPP shows relatively weaker and more diffuse activations. The overlap visualization further suggests that MFRPP improves the alignment between high-response regions and vessel structures, especially around vessels with varying calibers. This observation is consistent with the role of MFRPP as a fixed-receptive-field multi-scale aggregation module.

By aggregating multi-scale features with fixed receptive fields, MFRPP helps the network capture both coarse vascular structures and subtle details. Compared with adaptive atrous strategies, the fixed dilation rates of 1, 2, and 4 provide a stable and lightweight way to model vessels of varying calibers, especially thin capillaries. MFRPP is applied only once at the bottleneck, which enhances multi-scale representation without introducing the accumulated cost and potential gridding artifacts of layer-wise atrous designs [20, 21, 22]. Overall, these design choices make MFRPP a lightweight yet effective multi-scale component for retinal vessel segmentation.

To improve the modeling of directional vascular features, we introduce the DAMM module, which integrates directional convolution, channel attention, and simple pixel attention in a parallel manner. In the module-level ablation study, introducing DAMM alone improves F1 from 78.45% to 80.02%, increases Sen from 79.08% to 82.15%, and achieves an AUC of 98.56%, consistently outperforming the plain baseline. By jointly modeling directional, spatial, and semantic cues, DAMM enhances the network’s ability to recognize fine vessels and branching structures. The

results further suggest that DAMM remains effective when combined with other regularization or context modules, highlighting its role as a core component for vascular morphology modeling.

Topology-aware evaluation in Table 2 further shows that the proposed design improves not only pixel-wise performance but also structural consistency. Compared with the plain U-Net baseline, the complete model achieves a higher cDice (0.8238 vs. 0.7993), a lower connectivity-related topology error $|\Delta\beta_0|$ (64.38 vs. 83.50), a lower Rel.- β_0 Err. (37.4062 vs. 48.3125), and a slightly lower $|\Delta\beta_1|$ (3.88 vs. 4.00), indicating better preservation of vessel continuity and centerline overlap. Although individual modules affect different topology aspects in different ways, their joint combination yields the most balanced overall topology-aware performance. These results suggest that the combined use of MFRPP, DAMM, DropBlock, and the GC Block contributes not only to region-level accuracy but also to topology-related vessel structure recovery.

Overall, the module-level ablation results support the complementary roles of MFRPP and DAMM: MFRPP primarily improves multi-scale robustness, whereas DAMM more strongly enhances directional modeling and fine-vessel sensitivity. Together with DropBlock and the GC Block, they form a balanced design that improves both segmentation accuracy and topology preservation.

Table 3: DAMM Branch Ablation Study Results

Method	DAMM Branches				Performance Metrics (%)					Topology Metrics				Efficiency			
	SPA	CA	PA	DIR	Acc	Sen	Sp	F1	AUC	cDice	$ \Delta\beta_0 $	Rel.- β_0 Err	$ \Delta\beta_1 $	Params (M)	FLOPs (G)	Latency (s)	FPS
<i>Baseline Model (w/o DAMM)</i>																	
Baseline	–	–	–	–	97.36	80.78	98.47	79.40	98.63	0.8041	88.1250	52.0625	4.5000	1.24	59.74	0.033	30.57
<i>Baseline + DAMM Shell (Framework w/o Branches)</i>																	
Baseline + DAMM Shell (all branches disabled)	–	–	–	–	97.35	81.35	98.43	79.46	98.57	0.7996	87.1250	49.8125	6.1250	1.28	66.10	0.049	20.31
<i>Baseline + Single Branch Activation</i>																	
Baseline + SPA	✓	–	–	–	97.42	81.33	98.50	79.90	98.70	0.8052	72.7500	43.4375	5.3750	1.29	67.16	0.056	17.96
Baseline + CA	–	✓	–	–	97.22	83.18	98.17	79.07	98.66	0.8077	69.6250	42.0000	5.8750	1.29	66.13	0.051	19.45
Baseline + PA	–	–	✓	–	97.45	79.78	98.64	79.79	98.43	0.8134	67.1250	40.6875	4.2500	1.28	66.20	0.052	19.30
Baseline + DIR	–	–	–	✓	97.39	81.14	98.48	79.65	98.72	0.8099	71.2500	41.5625	5.5000	1.37	78.70	0.060	16.57
<i>Baseline + Dual Branch Combination</i>																	
Baseline + SPA+CA	✓	✓	–	–	97.30	81.53	98.36	79.18	98.55	0.8081	66.2500	38.8750	5.0000	1.33	70.31	0.061	16.42
Baseline + SPA+PA	✓	–	✓	–	97.40	82.94	98.37	80.07	98.77	0.8139	64.8750	38.4688	5.6250	1.32	70.39	0.061	16.45
Baseline + SPA+DIR	✓	–	–	✓	97.29	82.32	98.30	79.30	98.55	0.8020	110.3750	64.6562	6.7500	1.40	82.88	0.069	14.47
Baseline + CA+PA	–	✓	✓	–	97.31	83.60	98.24	79.68	98.54	0.8067	73.0000	44.0000	3.3750	1.31	69.35	0.057	17.66
Baseline + CA+DIR	–	✓	–	✓	97.23	85.60	98.01	79.56	98.71	0.8137	73.1250	41.9062	6.1250	1.40	81.85	0.065	15.35
Baseline + PA+DIR	–	–	✓	✓	97.34	84.33	98.22	80.01	98.70	0.8133	57.6250	33.4375	5.8750	1.39	81.92	0.065	15.33
<i>Triple Branch Combination</i>																	
Baseline + SPA+CA+DIR(DAMM)	✓	✓	–	✓	97.38	83.56	98.31	80.08	98.42	0.8148	47.7500	27.7812	4.6250	1.43	86.03	0.071	14.05
Baseline + SPA+PA+DIR	✓	–	✓	✓	97.25	82.94	98.21	79.14	98.64	0.8029	73.8750	42.9375	6.1250	1.42	86.11	0.072	13.95
Baseline + CA+PA+DIR	–	✓	✓	✓	97.29	74.59	98.82	77.65	96.87	0.8117	62.8750	37.7812	5.6250	1.42	85.07	0.067	14.97
Baseline + SPA+CA+PA	✓	✓	✓	–	97.28	84.18	98.17	79.62	98.45	0.8076	69.8750	42.0000	3.7500	1.35	73.54	0.063	15.87
<i>Complete DAMM Model</i>																	
Baseline+SPA+CA+PA+DIR	✓	✓	✓	✓	97.39	83.45	98.33	80.15	98.71	0.8108	69.5000	40.4375	5.1250	1.46	89.26	0.075	13.31

Note: 1. Baseline for branch-level ablation (Branch-ablation Backbone): U-Net + MFRPP + DropBlock + GC Block, with the DAMM module entirely removed. This baseline is distinct from the Vanilla Backbone used in the module-level ablation table. 2. “DAMM Shell” denotes the insertion of the DAMM module wrapper with all internal branches disabled; therefore, it is not equivalent to the branch-ablation baseline in which DAMM is completely removed. 3. Symbol explanation: ✓ indicates branch activation; “–” indicates the branch is disabled. 4. Red values indicate improvement compared to the baseline; green values indicate degradation. 5. For topology metrics, cDice is better when larger, while $|\Delta\beta_0|$, Rel.- β_0 error, and $|\Delta\beta_1|$ are better when smaller. 6. Efficiency profiling: Params denote the number of trainable parameters. FLOPs were profiled using the actual model input setting and the same profiling tool. For CHASE_DB1, the input tensor was set to $1 \times 1 \times 1008 \times 1008$, because the raw RGB fundus image is first converted into a preprocessed single-channel image through weighted channel fusion, normalization, and CLAHE, and the resulting single-channel image is used as the network input. FLOPs are reported as $2 \times \text{MACs}$. 7. All models were trained from scratch without pretraining or distillation, and the same training protocol was used for fair comparison across settings. 8. Inference time and FPS were measured on the same GPU under the same input size and batch size, averaged over N forward passes after discarding the first M warm-up passes for runtime stabilization (GPU warm-up only, no weight warm-start). 9. Unless otherwise stated, the default DAMM configuration includes DIR, SPA, and CA, whereas PA is evaluated only as an auxiliary branch in the branch-ablation experiments.

4.3.2 Branch-Level Ablation

To analyze the contribution of each branch within the DAMM module, we conducted branch-level ablation experiments on the CHASE_DB1 dataset. The experimental results demonstrate the following:

(1) **SPA Branch:** Removing the SPA branch led to a significant performance drop, Sen decreased from 83.45% to 74.59%, F1 from 80.15% to 77.65%, and AUC from 98.71% to 96.87%, marking the lowest performance among all combinations (see Table 3). This indicates that the SPA branch plays a critical role in model performance. Visual comparisons in Figure 11 show notable differences in boundary responses before and after SPA removal, particularly in fine vessel discontinuities and peripheral vessel edges. The red-highlighted areas illustrate regions where vessel structures are no longer enhanced after removing SPA, confirming its effectiveness in boundary preservation and fine vessel modeling. The SPA branch thus improves structural continuity by enhancing edge and microvascular features.

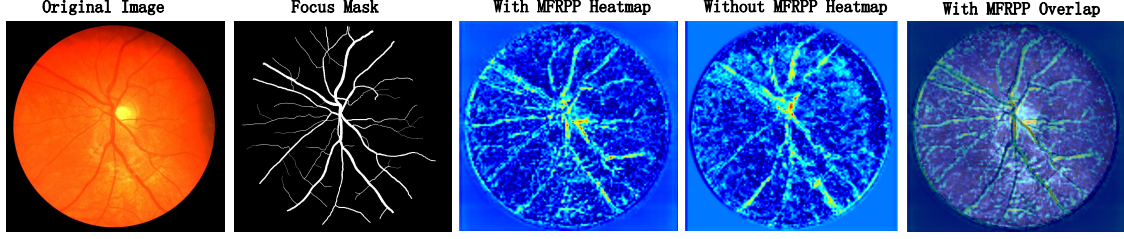


Figure 10: Module-level qualitative analysis of MFRPP using activation heatmaps. From left to right: original retinal image, focus mask, heatmap with MFRPP, heatmap without MFRPP, and the overlap visualization with MFRPP. Compared with the counterpart without MFRPP, the model with MFRPP exhibits stronger and more spatially coherent responses along vessel regions, supporting its role in scale-aware feature aggregation.

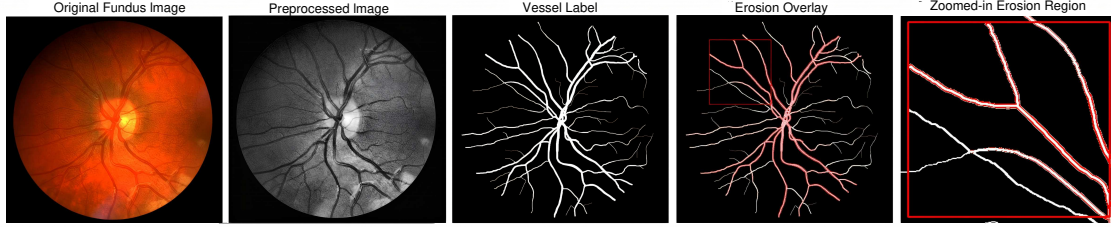


Figure 11: Enhancement effect of the SPA branch on vessel boundary modeling. From left to right: (a) original fundus image, (b) preprocessed result, (c) ground truth, (d) overlay comparison of Baseline+SPA+CA+DIR predictions before and after removing SPA, and (e) magnified region of interest. The **red regions** indicate the differences between predictions with and without SPA (mainly located at fine discontinuities and vessel boundaries), while the **white regions** represent the common major vessels.

(2) DIR Branch: When the DIR branch is activated independently, the AUC reaches 98.72%, the highest among all single-branch configurations, and it consistently maintains strong performance in multi-branch combinations. Removing DIR leads to a drop in F1 to 79.62% and AUC to 98.45%, indicating its importance in modeling the primary directions, bifurcations, and curved trajectories of vessels. In the Baseline+SPA+CA+DIR model, excluding DIR resulted in a 2% decrease in Sen and 0.9% drop in F1. As shown in the enhanced regions of Figure 12, DIR contributes to the modeling of fine directional vessels and enhances responses along the main vessel trunks and crossing regions.

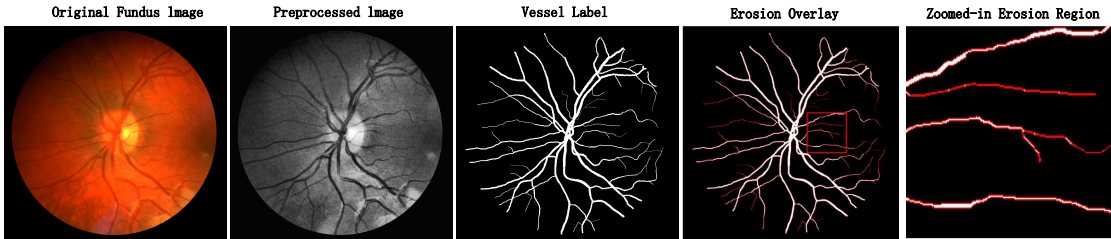


Figure 12: Enhancement effect of the DIR branch on directional vessel modeling. From left to right: (a) original fundus image, (b) preprocessed result, (c) ground truth, (d) overlay comparison of Baseline+SPA+CA+DIR predictions before and after removing DIR, and (e) magnified region of interest. The **red regions** indicate the differences between predictions with and without DIR (mainly located along vessel trunks, bifurcations, and curved paths), while the **white regions** represent the common vessel structures.

Kernel-size ablation in DIR: We analyze the impact of strip-kernel size in the DIR branch on CHASE_DB1 using three random seeds, evaluating single-k settings $k \in \{3, 5, 7, 9\}$ and multi-kernel combinations (Table S1). The shortest kernel ($k = 3$) performs the worst (F1: $78.95\% \pm 0.30\%$), indicating that a limited directional receptive field may be insufficient for preserving vessel continuity. As shown in Figure 13, increasing the strip-kernel size from $k = 3$ to $k = 7$ consistently improves the mean F1, and the $1 \times 7/7 \times 1$ configuration achieves the best overall performance, with the highest mean F1 ($79.89\% \pm 0.20\%$). When the kernel size is further enlarged to $k = 9$, the gain is not sustained and Precision decreases (Prec: $77.88\% \pm 0.16\%$), resulting in a slightly lower F1 ($79.62\% \pm 0.30\%$), which

suggests diminishing returns for overly large directional kernels. The multi-kernel setting (5,7) improves Sensitivity and AUC (Sen: $81.94\% \pm 0.41\%$, AUC: $98.75\% \pm 0.05\%$) but slightly reduces Specificity (Sp: $98.43\% \pm 0.03\%$). Considering this trade-off, we adopt the strip-kernel setting $k = 7$ (i.e., 1×7 and 7×1) as the default configuration in the main experiments, while the multi-kernel setting (5,7) may be considered in scenarios where sensitivity is prioritized. Figure 13 visually summarizes this trend, highlighting the advantage of the $1 \times 7/7 \times 1$ setting among all evaluated DIR kernel configurations.

Diagonal-kernel ablation in DIR: We further evaluate multi-directional kernels by enabling diagonal ($45^\circ/135^\circ$) filters in the DIR branch. As also illustrated in Figure 13, the diagonal-enabled variants do not surpass the best strip-only configuration in terms of mean F1. Table S4 further confirms that their effects are marginal and seed-sensitive in paired comparisons matched by `dir_ks` and random seed, yielding no consistent F1 improvement (4/12 improved vs. 8/12 decreased; mean $\Delta F1$ (diag-strip) = -0.0002). Although slight gains in Sensitivity and AUC are observed in a few cases, these improvements are not stable across kernel sizes or seeds. Therefore, we retain the horizontal/vertical strip-kernel design as the default DIR choice and report the diagonal-kernel results in the Supplementary Material for completeness, rather than adopting them in the main configuration.

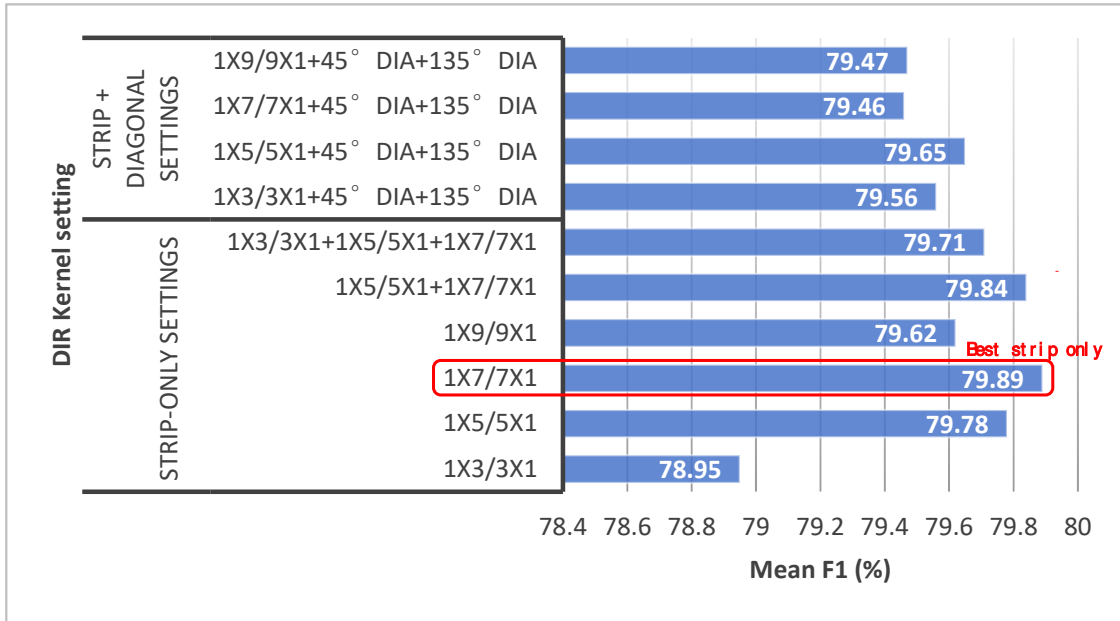


Figure 13: Comparison of mean F1 under different DIR kernel settings on CHASE_DB1. Among strip-only settings, the $1 \times 7/7 \times 1$ configuration achieves the best mean F1 (79.89). In contrast, diagonal-enabled variants do not provide a consistent improvement over the strip-only baseline.

Overall, the ablation results indicate that moderately large strip kernels are sufficient and most effective for directional vessel modeling in DIR, while introducing diagonal kernels does not provide stable additional gains; therefore, the $1 \times 7/7 \times 1$ strip-only design is adopted as the default setting.

(3) CA Branch: In the full three-branch configuration (SPA+CA+DIR), the model achieved a Sen of 83.56% and an AUC of 98.42%, outperforming the two-branch SPA+DIR variant (Sen: 82.94%, AUC: 98.23%). These results are also comparable to the full DAMM module (Sen: 83.45%, AUC: 98.71%), confirming that the CA branch contributes to improved detection by complementing spatial and directional features with semantic channel information.

(4) Topology and Performance-Efficiency Trade-off: Table 3 further complements the branch-level analysis by reporting both topology-aware and efficiency-related metrics. In terms of topology preservation, compared with the baseline model without DAMM, the adopted three-branch configuration (SPA+CA+DIR) improves $clDice$ from 0.8041 to 0.8148 and substantially reduces the connectivity-related topology errors $|\Delta\beta_0|$ (88.1250 to 47.7500) and $Rel.-\beta_0$ Err. (52.0625% to 27.7812%), indicating markedly better preservation of vessel continuity and centerline overlap. Meanwhile, $|\Delta\beta_1|$ shows a more branch-dependent behavior, suggesting that different attention combinations may affect different aspects of topology preservation in different ways.

From the efficiency perspective, SPA+CA+DIR combination achieves a favorable balance between accuracy and efficiency. Compared to the Four-Branch configuration, this setup reduces parameter count by 2.1% (from 1.46M to

1.43M), FLOPs by 3.6% (from 89.26G to 86.03G), and improves inference speed by 5.6% (from 13.31 FPS to 14.05 FPS), with minimal performance degradation (F1 drops by only 0.07%, AUC by 0.29%). These results suggest that the PA branch, which contributes relatively little, can be excluded during deployment to boost computational efficiency without significant loss in performance. The results validate the multi-dimensional parallel attention mechanism within the DAMM module: the consistent improvements in Sensitivity and F1, together with the reduced fragmentation in thin and crossing vessels observed in Figures 10-11, demonstrate its effectiveness in modeling fine vessels, bifurcations, and directional structures while maintaining a favorable trade-off between segmentation accuracy and generalization capability. These findings are consistent with the design goal of DAMM, namely enhancing directional dependency modeling for retinal vessel segmentation.

Overall, the branch-level ablation results highlight the distinct yet complementary roles of each branch: SPA mainly improves boundary and detail preservation, DIR strengthens directional modeling, and CA provides additional semantic calibration, whereas PA contributes comparatively limited gains. Taken together, these observations indicate that SPA, DIR, and CA capture complementary rather than redundant cues within DAMM. The consistent improvements in Sensitivity and F1, together with the reduced fragmentation in thin and crossing vessels observed in Figures 11–12, further validate the effectiveness of the adopted three-branch design for retinal vessel segmentation.

4.4 Analysis of Loss Function Optimization

This subsection reports two complementary experiments on GA-Loss. First, Section 4.4.1 presents a search-budget analysis of the genetic algorithm on CHASE_DB1 using the Vanilla Backbone, in order to examine the convergence behavior and the effect of search depth. Second, Section 4.4.2 validates the effectiveness and cross-dataset consistency of the GA-selected loss under the Core Backbone setting of DAMMGA-Net, by comparing it with BCE on DRIVE, CHASE_DB1, and STARE under the same architecture. This separation is intended to distinguish the behavior of the search procedure itself from the dataset-level performance gains brought by the selected loss configuration.

4.4.1 Hyperparameter Optimization

Loss-function optimization forms a high-dimensional hyperparameter search problem, requiring simultaneous optimization of the weight vector $[w_1, w_2, w_3, w_4]$ and the loss-specific parameter vector $[\alpha_{focal}, \gamma_{focal}, \beta_{tversky}, \alpha_{tversky}]$. Traditional gradient-based methods are often inadequate for this task because the mapping from hyperparameters to validation loss is noisy and non-smooth under stochastic training dynamics, producing unreliable gradient signals and leading to convergence toward suboptimal local minima.

To address these challenges, we adopt a gradient-free optimization strategy based on a Genetic Algorithm (GA). The GA evaluates each candidate configuration using its validation loss (fitness), allowing effective global exploration of the search space while satisfying constraints such as weight normalization. The GA jointly optimizes both the composition of the composite loss and the internal parameters of its constituent losses, tailoring the loss design to the structural properties of retinal vessels. Although the search is conducted in a composite-loss space, the best-performing solution after 20 generations converged to a degenerate one-hot configuration, i.e., a single Focal Loss with parameters $\gamma = 4.94$ and $\alpha = 0.201$.

We further analyze the search outcomes of the optimal loss configuration under 10- and 20-generation budgets (see Figures 14–16 and Table 4, based on the Vanilla Backbone on CHASE_DB1). In both cases, the best-performing solution converged to a degenerate one-hot configuration, i.e., a single Focal Loss, while the selected internal parameters differed slightly across search budgets. Specifically, α remained close to 0.20 ($\alpha = 0.199$ at 10 generations; $\alpha = 0.201$ at 20 generations), indicating a stable class-balance factor, whereas γ increased from 4.058 to 4.936 (+21.6%), suggesting that a stronger focus on hard examples was favored as the search budget increased. Correspondingly, the evaluation metrics improved as follows:

- Fitness decreased from 0.00182 to 0.00105 (-42.3%),
- F1 rose from 69.01% to 78.02%,
- Sensitivity increased from 55.45% to 87.67% (the missed-diagnosis rate dropped from 44.55% to 12.33%),
- Specificity slightly declined from 99.65% to 97.51%.

These results indicate that, under the Vanilla Backbone setting, increasing the search budget leads to a better loss configuration and a more recall-oriented trade-off. Therefore, Table 4 should be interpreted as a search-budget and convergence analysis of the GA procedure, rather than as a separate hyperparameter-tuning experiment for Focal Loss.

Table 4: Search-budget analysis of GA-Loss on CHASE_DB1 using the Vanilla Backbone.

Comparison Dimension	10-Gen	20-Gen	Change
<i>Search Configuration</i>			
Duration	1.3 hours	8.0 hours	+6.7h
Generations	10	20	+10
Evaluations	50	100	+50
Total Epochs	750	1500	+750
<i>Optimal Parameters</i>			
γ_{focal}	4.058	4.936	+0.878
α_{focal}	0.199	0.201	+0.002
Fitness Score	0.00182	0.00105	-42.3%
Convergence Gen	8	13	+5
<i>Performance Metrics (%)</i>			
F1 Score	69.01	78.02	+9.01
Sensitivity	55.45	87.67	+32.22
Specificity	99.65	97.51	-2.14
Miss. Diag. Rate	44.55	12.33	-32.22

Notes: “Vanilla Backbone” denotes the plain U-Net backbone used in the module-level ablation, without DAMM, MFRPP, GC Block, or DropBlock. This table evaluates the effect of GA search budget (10 vs. 20 generations) under the same simplified backbone setting on CHASE_DB1.

4.4.2 Comparative Performance of Loss Functions

Loss function design critically affects segmentation performance, particularly under severe class imbalance. To further validate the practical utility of the selected loss configuration, we compare Binary Cross-Entropy (BCE) and GA-Loss under the Core Backbone setting of DAMMGA-Net on three public retinal vessel datasets, namely DRIVE, CHASE_DB1, and STARE. Here, the “Core Backbone” denotes the intrinsic encoder–decoder representation stream consisting of HFEM-based encoding and decoder-side Refinement Blocks, while excluding DAMM and MFRPP. Under this setting, DropBlock remains part of HFEM and the Refinement Blocks, and the GC Block remains embedded in the Refinement Blocks. The purpose of this experiment is not to perform strict cross-dataset generalization testing, but to examine whether the GA-selected loss yields consistent gains across different dataset conditions under the same architectural setting.

All experiments in this comparison were conducted using identical data splits, identical training settings, and the same Core Backbone architecture. Two loss functions were evaluated: (1) Binary Cross-Entropy (BCE) and (2) the GA-selected loss (GA-Loss). F1 was computed using a fixed probability threshold of 0.5. The results are summarized in Table 5.

The main observations from Table 5 are as follows:

- GA-Loss improves F1 on all three datasets, from 77.83% to 80.45% on DRIVE, from 75.45% to 78.15% on CHASE_DB1, and from 78.12% to 85.15% on STARE.
- Sensitivity also improves consistently, from 78.91% to 80.12% on DRIVE, from 81.93% to 83.06% on CHASE_DB1, and from 79.82% to 86.36% on STARE.
- AUC increases on all three datasets, from 93.53% to 95.30% on DRIVE, from 94.32% to 95.75% on CHASE_DB1, and from 93.51% to 97.46% on STARE, indicating more favorable threshold-free discrimination under the GA-selected loss.
- Specificity is not merely stable; it also increases in all three datasets, from 93.12% to 98.31% on DRIVE, from 98.23% to 98.63% on CHASE_DB1, and from 93.12% to 98.70% on STARE.
- Accuracy improves on DRIVE and STARE, while a slight decrease is observed on CHASE_DB1 (97.83% to 97.68%), suggesting that the gain brought by GA-Loss is mainly reflected in recall-sensitive and overlap-related metrics rather than a uniform increase in every metric.

Overall, these results indicate that the GA-selected loss improves recall-sensitive and overlap-related performance more consistently than a manually fixed BCE objective under the same Core Backbone setting. In particular, the gains

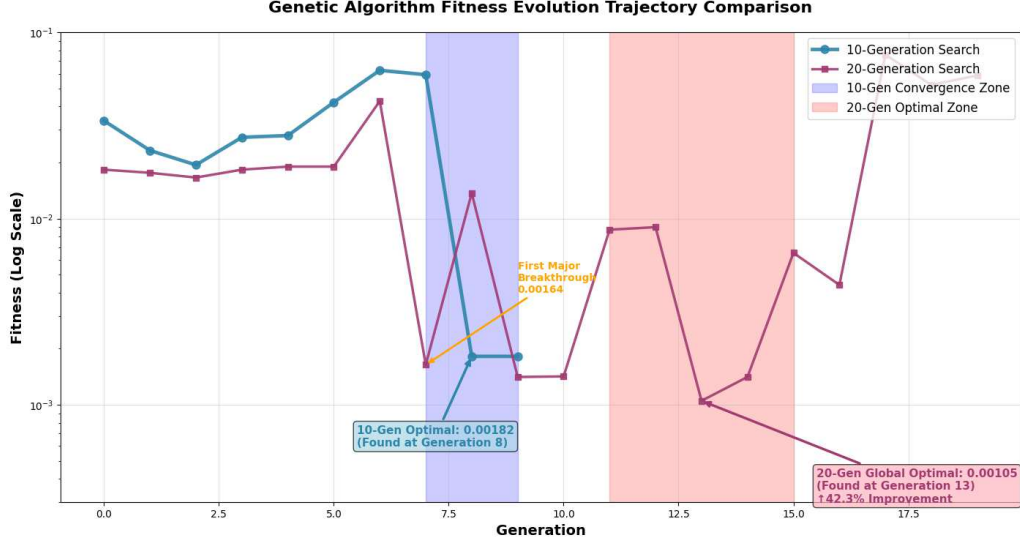


Figure 14: Fitness evolution trajectory.

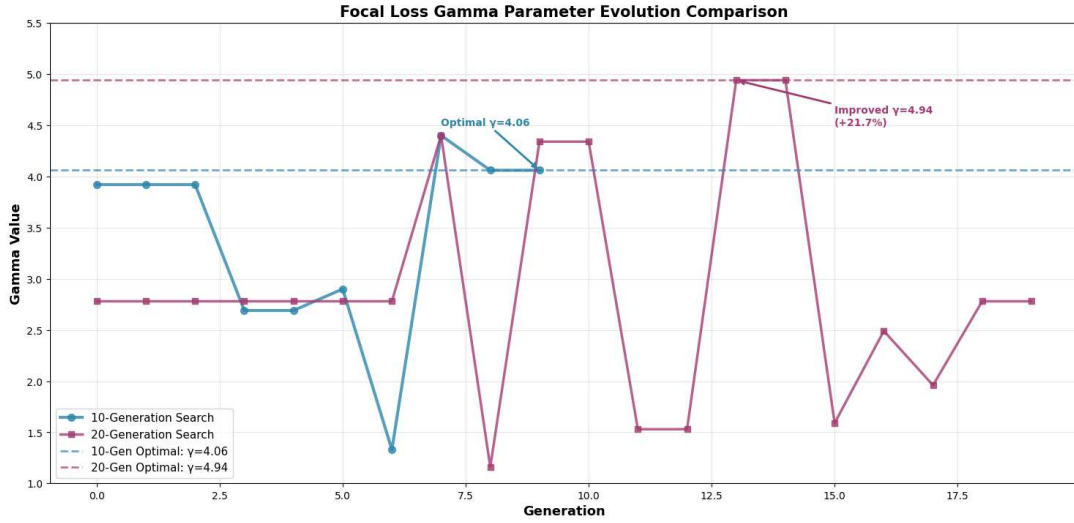


Figure 15: Evolution of the Gamma parameter in Focal Loss.

in F1, sensitivity, and AUC across all three datasets suggest that the selected loss better matches the class-imbalanced and thin-structure characteristics of retinal vessel segmentation.

4.5 Comparison with Existing Methods

We evaluate the proposed method on three public retinal vessel datasets (DRIVE, STARE, and CHASE_DB1) through both quantitative and qualitative comparisons. Unless otherwise specified, we use the three-branch DAMM setting (SPA+CA+DIR). For baselines without official implementations, the reported numbers are taken from the original papers, as noted in Table 6. DAMMGA-Net is an end-to-end framework composed of an HFEM encoder, the direction-aware mixed attention module (DAMM), a multi-scale fixed receptive-field parallel pyramid (MFRPP), and decoder-side progressive refinement (Refinement Blocks). In addition, GA-Loss searches a constrained composite-loss space to select dataset-adaptive training objectives, aiming to improve thin-vessel separability, cross-scale coherence, and structural continuity.

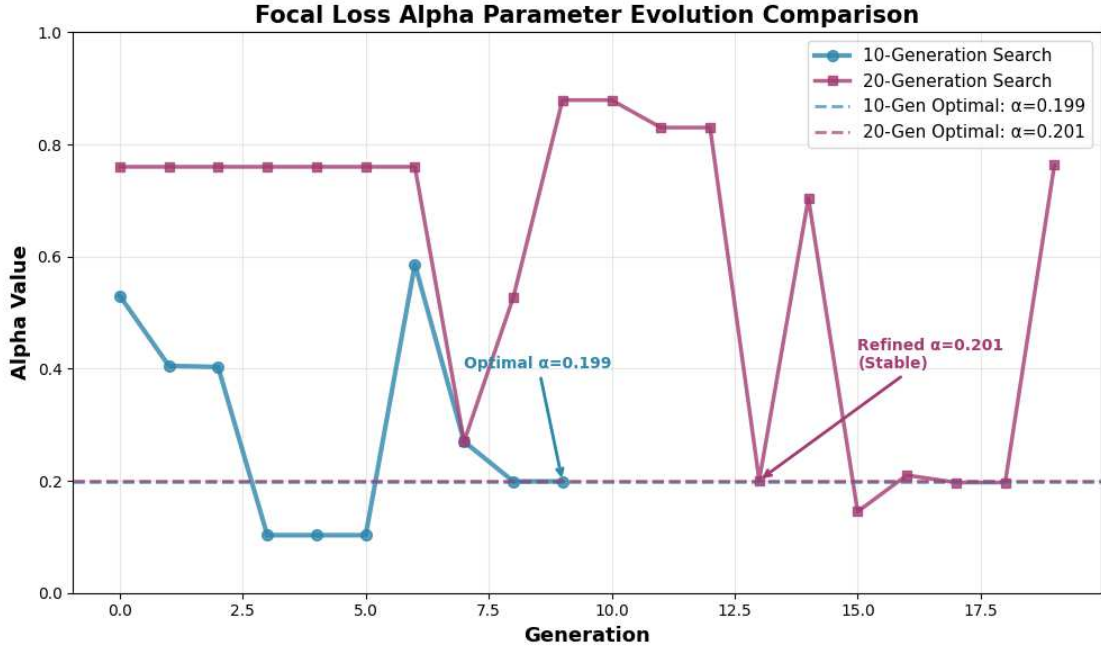


Figure 16: Evolution of the Alpha parameter in Focal Loss.

Table 5: Comparison of BCE and GA-Loss under the Core Backbone setting (HFEM + Refinement, w/o DAMM and w/o MFRPP).

Dataset	Loss	Acc	Sen	Sp	F1	AUC
DRIVE	BCE	94.96	78.91	93.12	77.83	93.53
	GA-Loss	95.37	80.12	98.31	80.45	95.30
CHASE_DB1	BCE	97.83	81.93	98.23	75.45	94.32
	GA-Loss	97.68	83.06	98.63	78.15	95.75
STARE	BCE	90.12	79.82	93.12	78.12	93.51
	GA-Loss	97.98	86.36	98.70	85.15	97.46

Notes: “Core Backbone” denotes the intrinsic encoder–decoder representation stream of DAMMGA-Net, consisting of HFEM-based encoding and decoder-side Refinement Blocks, while excluding the two main architectural contribution modules, namely DAMM and MFRPP. Under this setting, DropBlock remains part of HFEM and the Refinement Blocks, and the GC Block remains embedded in the Refinement Blocks. All metrics are reported as percentages (%). GA-Loss denotes the composite loss configuration obtained through genetic algorithm optimization. A fixed probability threshold of 0.5 was used for F1 computation. Results are averaged over three runs with identical training settings and different random seeds.

Table 6: Segmentation performance comparison of different methods on CHASE_DB1, STARE, and DRIVE datasets.

Model	Params (M)↓	FLOPs (G)↓	CHASE_DB1 (%)					STARE (%)					DRIVE (%)					Year
			Acc	Sen	Sp	F1	AUC	Acc	Sen	Sp	F1	AUC	Acc	Sen	Sp	F1	AUC	
U-Net(base=64) [31]	31.038	1592.919	95.78	82.88	97.01	77.83	97.72	97.56	81.87	98.70	82.73	98.94	96.90	80.91	98.46	81.91	98.63	2015
DUNet [32]	0.88	83.77	96.74	81.75	98.51	83.78	98.82	96.78	81.02	98.74	83.89	99.10	95.76	80.63	98.10	83.37	98.12	2019
RNA-Net [33]	–	–	96.66	80.78	98.28	81.31	98.49	96.51	79.18	99.13	85.21	99.23	97.02	81.27	98.62	83.06	98.60	2020
HAN [34]	–	–	96.80	83.39	98.23	82.91	98.81	97.45	80.51	98.93	83.16	98.66	95.91	80.91	98.23	83.93	98.33	2020
LACNN [35]	–	–	96.35	78.18	98.19	–	98.10	96.78	83.52	98.23	–	98.75	95.68	79.21	98.10	–	98.06	2020
MAN-RVS [27]	2.03	1905.51	96.73	84.02	98.01	82.48	98.74	96.65	79.14	98.70	82.76	98.64	95.81	80.46	98.05	83.03	98.27	2021
Bridge-Net [36]	–	–	96.83	85.02	98.11	83.48	98.84	96.68	80.02	98.64	82.89	99.01	95.68	81.15	97.80	82.72	98.10	2022
DCU-net [37]	–	–	96.64	80.75	98.41	82.78	98.72	–	–	–	–	–	95.75	79.53	98.28	83.03	98.44	2022
ResDO-UNet [38]	–	–	96.72	80.20	97.94	82.63	–	95.67	79.63	97.92	81.72	–	95.61	79.85	97.91	82.29	–	2023
Wave-Net [39]	1.50	42.61	96.64	82.84	98.21	83.49	–	96.41	79.03	98.36	81.40	–	95.61	81.64	97.64	82.54	–	2023
RCAR-UNet [40]	–	–	95.66	74.75	97.98	74.70	–	95.94	69.79	99.05	78.50	–	95.37	74.87	98.36	80.47	–	2024
RVS-FDSC [41]	–	12.05	97.43	83.56	98.56	80.50	98.67	97.18	80.04	98.82	78.75	98.08	96.92	82.21	98.78	82.71	98.56	2024
TCDDU-Net [12]	–	–	97.23	81.87	98.23	78.38	98.50	97.40	79.20	98.84	81.63	98.56	96.98	82.58	98.38	82.65	98.68	2024
LMFR-Net [42]	0.07	63.95	97.40	83.85	98.44	82.91	–	96.99	86.85	98.76	82.57	–	96.58	84.79	98.30	82.37	–	2025
PA-Net [43]	2.02	3.47	96.77	85.70	97.79	83.08	98.75	97.09	88.13	98.05	85.61	99.08	95.82	82.84	98.07	83.93	98.33	2025
MSTP-Net [44]	1.82	169.83	97.45	84.85	98.30	80.74	–	97.61	86.03	98.58	84.68	–	96.91	83.68	98.18	82.58	–	2025
Ours	1.43	86.03	97.68	83.06	98.33	80.15	98.75	98.04	87.93	98.87	87.21	99.45	97.32	83.96	98.56	84.13	98.79	2025

Note: 1. "–" indicates that the corresponding metric is not reported in the original paper; for Params/FLOPs, "–" indicates that the official implementation/details are unavailable or incomplete, thus Params/FLOPs cannot be fairly profiled under the same setting. 2. Bold numbers indicate the best results for each metric. 3. Params and FLOPs are computed under the actual model input setting using the same profiling tool. For CHASE_DB1, the input tensor is set to $1 \times 1 \times 1008 \times 1008$, because the raw RGB fundus image is first converted into a preprocessed single-channel image through weighted channel fusion, normalization, and CLAHE, and the resulting single-channel image is used as the network input. FLOPs are reported as $2 \times \text{MACs}$. 4. All models in this paper were trained from scratch without using pre-training, distillation, or any warm-start weights.

4.5.1 Quantitative comparison and per-metric best-method attribution

As summarized in Table 6, the top method varies across datasets and metrics. This is common in thin-structure segmentation under heavy class imbalance: sensitivity (Sen) is closely tied to false negatives (FN), specificity (Sp) to false positives (FP), and F1 reflects how precision and recall are negotiated. To avoid over-interpreting a single peak metric, we report dataset-level patterns and attribute each metric to its best method, highlighting both gains and associated costs.

(1) DRIVE: best Acc/F1/AUC, top-tier Sen with stronger FP control. On DRIVE, our method achieves the best **Acc, F1, and AUC** (Acc=97.32%, F1=84.13%, AUC=98.79%), indicating an advantage in overall correctness, overlap quality, and threshold-free discrimination. The highest Sen is reported by LMFR-Net (84.79%), but its **Acc and F1** (96.58% and 82.37%) are lower than ours, and AUC is not reported, suggesting that the recall gain does not translate into a better overall agreement under the reported setting. Notably, our Sen reaches 83.96%, ranking **2nd** among all compared methods (only behind LMFR-Net), while also achieving a higher Sp (ours 98.56% vs. LMFR-Net 98.30%). This indicates that the improved detection of thin vessels is not obtained by a marked increase in FP. In contrast, RNA-Net attains the highest Sp (98.62%) but yields a lower Sen (81.27%) than ours (83.96%), consistent with a more conservative decision tendency that may miss weak vessels. Overall, the Sen/Sp pairing and the leading F1/AUC on DRIVE favor DAMMGA-Net.

(2) STARE: best Acc/F1/AUC, near-peak Sen with clearer precision advantage. On STARE, our method again obtains the best **Acc, F1, and AUC** (Acc=98.04%, F1=87.21%, AUC=99.45%). PA-Net reports the highest Sen (88.13%), while ours is 87.97% (only 0.16 points lower, ranked **2nd**); however, PA-Net shows lower Sp and F1 (98.05% and 85.61%) than ours (Sp=98.87%, F1=87.21%). This pattern suggests that a small recall edge may come with more boundary expansion or local FP, limiting overall overlap quality. Similarly, RNA-Net yields the highest Sp (99.13%) but with a much lower Sen (79.18%) and a lower F1 (85.21%) than ours. Taken together, STARE results support that DAMMGA-Net maintains high recall while preserving precision, leading to the best F1/AUC in Table 6.

(3) CHASE_DB1: best Acc, non-peak Sen/F1, with stronger metric tension. On CHASE_DB1, our method achieves the best **Acc** (97.68%) and keeps Sp and AUC at a high level (Sp=98.33%, AUC=98.75%). Meanwhile, peak Sen and F1 are achieved by other methods (PA-Net: Sen=85.70%; Wave-Net: F1=83.49%; Bridge-Net: AUC=98.84%; LMFR-Net: Sp=98.44%). This indicates a stronger metric tension on CHASE_DB1, where dense capillaries and larger scale variation can make F1 more sensitive to boundary errors and thresholding. In this setting, our results correspond to an operating point that favors overall correctness and FP suppression while keeping Sen competitive (83.06%), whereas further F1 gains may require finer boundary control.

4.5.2 Visual comparison and error-mode interpretation

Figs. 17–19 provide overlay maps and error maps that help contextualize Table 6. In the overlays, all methods recover major vessel trunks; differences concentrate on thin branches, continuity at bifurcations/crossings, and low-contrast regions. In the error maps, blue FN gaps correspond to missed thin vessels and broken traces, whereas red FP marks often appear near textured background or along vessel borders.

Across several zoomed regions, our method exhibits fewer FN-induced breaks on weak branches, leading to more continuous trajectories; this agrees with the stronger Sen and the leading F1/AUC observed on DRIVE and STARE in Table 6. At the same time, occasional localized FP traces can be seen near high-contrast edges or strong background texture, appearing as mild boundary expansion or small false activations. This visual pattern is consistent with the fact that Sp/F1 does not peak simultaneously on every dataset and that boundary-related errors can affect F1 more noticeably on CHASE_DB1. Overall, the qualitative evidence supports a relatively coherent error profile: DAMMGA-Net tends to reduce discontinuity-causing FN errors while keeping FP largely localized, rather than trading one error type for large-scale artifacts.

5 Discussion

This section interprets the results in Table 6 together with the error distributions in Figs. 17–19 to explain where DAMMGA-Net gains accuracy and where errors remain. On DRIVE and STARE, the leading Acc/F1/AUC in Table 6, along with fewer FN gaps on thin branches in the error maps, suggests that the method is particularly effective at reducing discontinuities on weak vessels and improving overall agreement. Meanwhile, the remaining errors are often boundary-adjacent FP traces, which are more likely to affect F1 on higher-resolution and capillary-dense imagery such as CHASE_DB1, where small boundary deviations can shift the precision–recall balance more noticeably.

5.1 Mechanism analysis: continuity and cross-scale modeling

Retinal vessels are elongated and curvilinear; bifurcations and crossings introduce abrupt orientation changes, and vessel calibers span a wide range. Consequently, segmentation quality depends not only on detection but also on whether predictions remain connected along trajectories and remain coherent across scales. DAMMGA-Net addresses these two aspects through the joint use of DAMM and MFRPP in the encoder.

DAMM couples spatial saliency, channel recalibration, and explicit directional modeling, encouraging orientation-consistent responses along vessel paths and thereby reducing broken traces on faint capillaries. This behavior is consistent with the error maps in Figs. 17–19, where fewer blue FN gaps are observed in several zoomed regions, indicating fewer continuity-breaking misses. In parallel, MFRPP employs fixed-dilation atrous branches to provide explicit multi-scale context coverage. Compared with highly dynamic receptive-field designs, this fixed decomposition offers a clearer scale assignment that can be easier to optimize in small-sample settings, and it complements the directional cue: multi-scale coverage mitigates scale-induced misses, while direction cues mitigate breaks around rapid orientation changes. Together, they provide decoder-side refinement with more informative structural cues for reconstructing both thick trunks and thin branches.

5.2 Decoder refinement and boundary-related errors

Progressive upsampling and cross-level fusion often introduce (i) boundary noise and texture-driven spurious activations and (ii) semantic misalignment across stages that can distort thin branches. DAMMGA-Net inserts Refinement Blocks after each fusion step and embeds a GC Block within them, allowing high-level semantics to participate in correction at multiple decoder stages. This design helps align features across stages and improves reconstruction around bifurcations and crossings.

However, the error maps also show occasional red FP traces concentrated near high-contrast borders or strongly textured regions, appearing as mild boundary expansion or small false activations. In conjunction with Table 6, these patterns suggest that when the model favors preserving continuity and retaining thin structures, some sharp edges can be locally interpreted as vessel extensions, which may limit Sp or F1 in certain cases. Importantly, these FP errors are typically localized rather than large-scale over-segmentation. For applications that require stricter precision, incorporating more explicit FP control (e.g., precision/Sp-aware terms or post-threshold calibration) may further reduce boundary-related false positives.

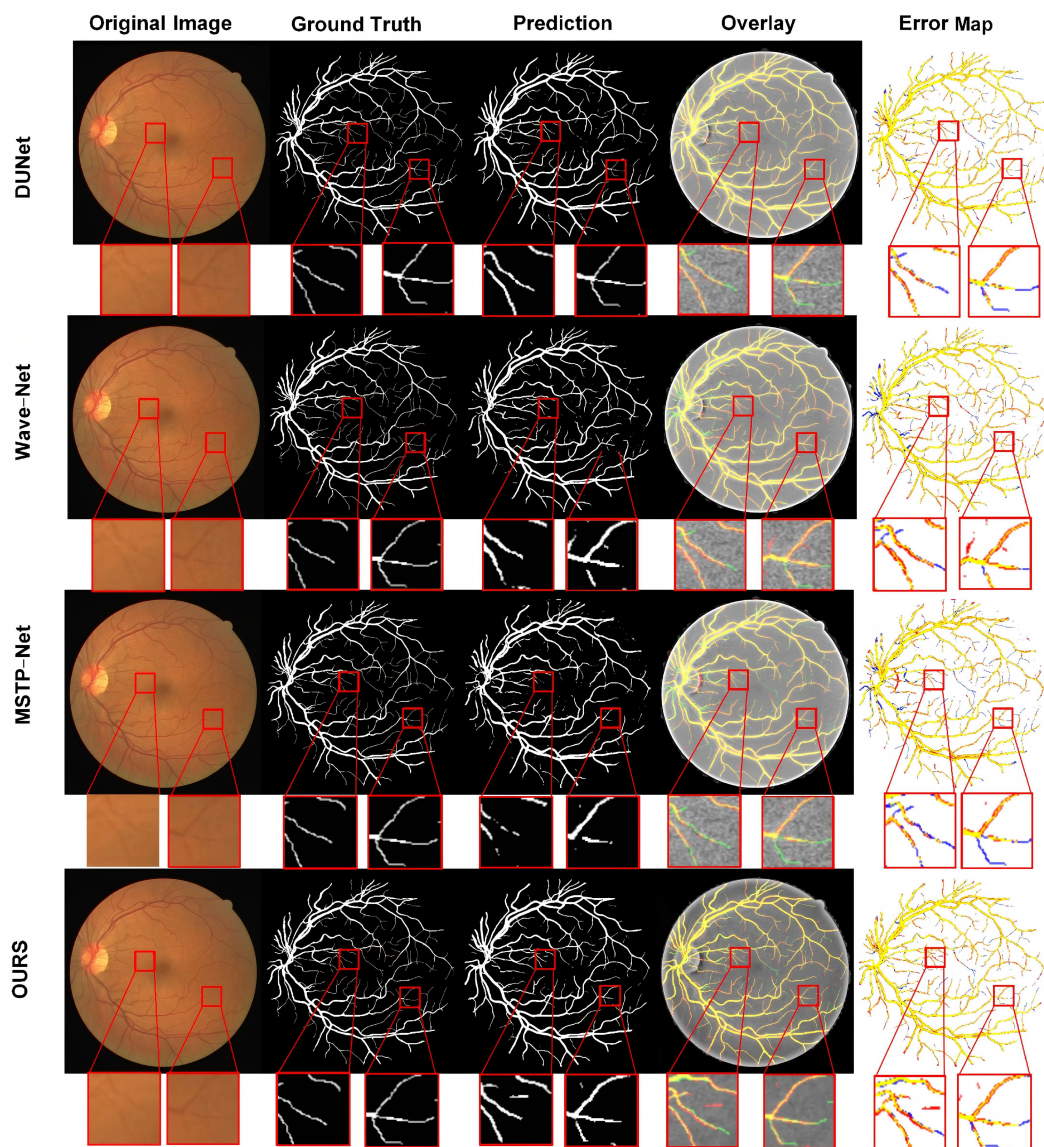


Figure 17: Segmentation results comparison on the DRIVE dataset (ROI enlarged $\times 2.5$). **Overlay Map:** Green = Ground Truth, Red = Prediction, Yellow = True Positives (correct predictions), Grey = Background of the original image. **Error Map:** White = True Negative (correct background), Blue = False Negative (missed vessels), Yellow = True Positive (correct vessels), Red = False Positive (incorrect vessel prediction in background).

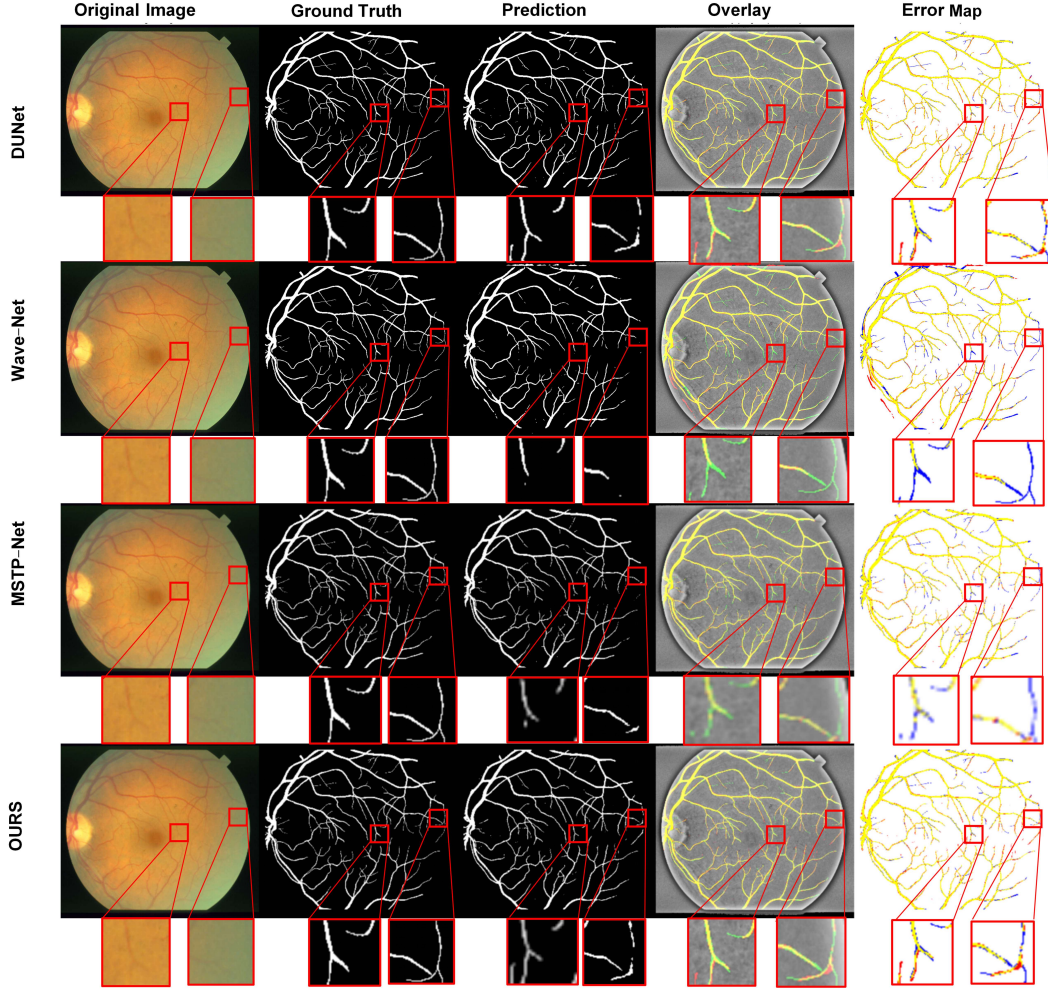


Figure 18: Segmentation results comparison on the STARE dataset (right: ROI enlarged $\times 2.5$). **Overlay Map:** Green = Ground Truth, Red = Prediction, Yellow = True Positives (correct predictions), Grey = Background of the original image. **Error Map:** White = True Negative (correct background), Blue = False Negative (missed vessels), Yellow = True Positive (correct vessels), Red = False Positive (incorrect vessel prediction in background).

5.3 Objective adaptation and dataset dependence

Datasets differ in noise, annotation tendencies, and imbalance profiles, so a fixed, manually tuned loss configuration may not transfer cleanly across benchmarks. GA-Loss treats the weights and internal parameters of a composite loss as search variables and selects objectives within a constrained space using validation feedback, reducing reliance on ad-hoc trial-and-error. Objective preference, however, also shapes how FN and FP are distributed: configurations that emphasize FN reduction can improve thin-vessel completeness but may slightly increase boundary FP. This is consistent with the joint observation that FN discontinuities are reduced in the error maps while localized FP traces can remain.

The stronger metric tension on CHASE_DB1 can be interpreted through dataset characteristics. With denser thin vessels and larger scale variation, small boundary expansions or local misses can more strongly influence the precision–recall balance, making F1 more sensitive to thresholding and boundary quality. Therefore, it is not unexpected that, on CHASE_DB1, Table 6 shows DAMMGA-Net achieving the best Acc while peak Sen/F1 is obtained by other methods. The evidence supports reading this behavior as an operating point that prioritizes overall correctness and FP suppression while keeping sensitivity competitive, rather than optimizing a single scalar metric to its maximum in a more demanding setting.

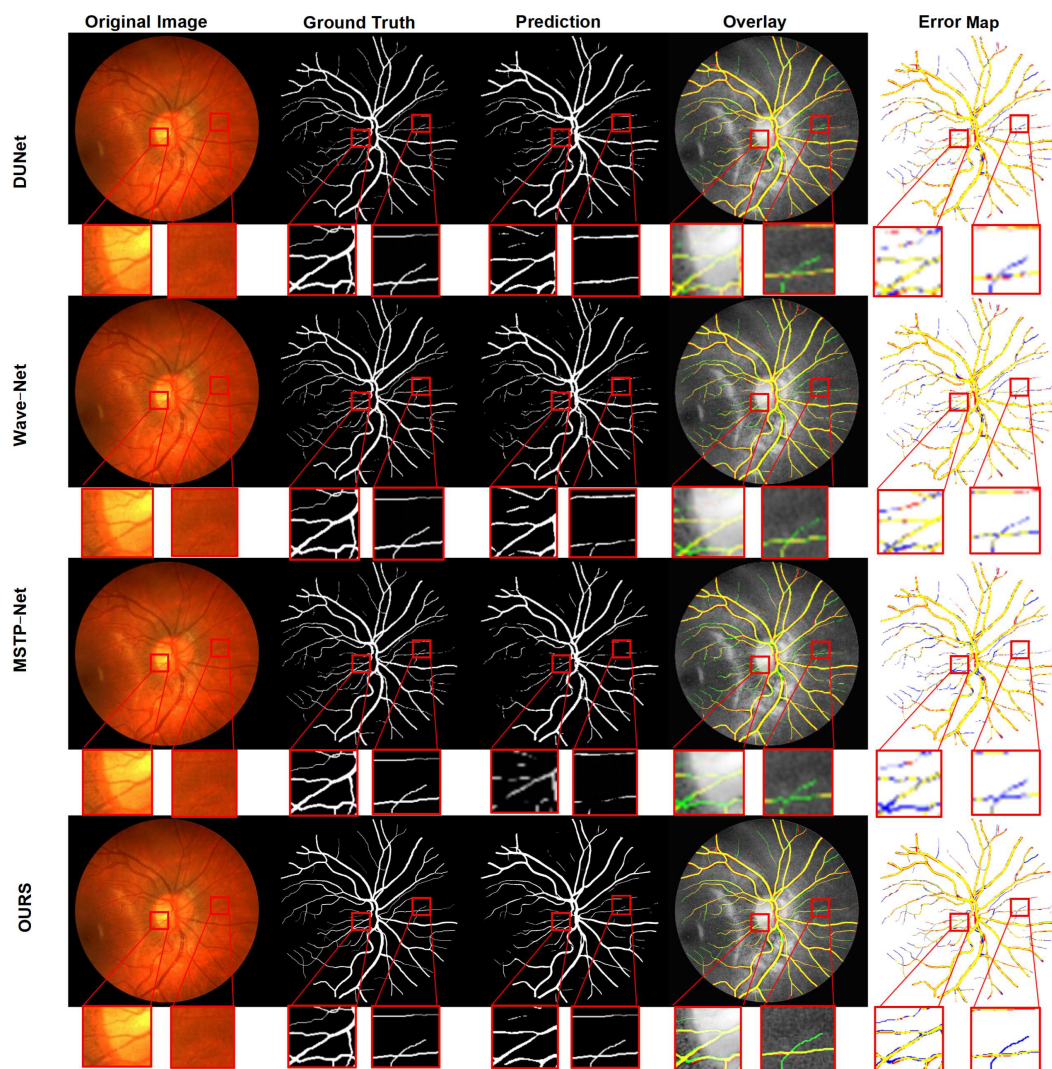


Figure 19: Segmentation results comparison on the CHASE_DB1 dataset (right: ROI enlarged $\times 2.5$). **Overlay Map:** Green = Ground Truth, Red = Prediction, Yellow = True Positives (correct predictions), Grey = Background of the original image. **Error Map:** White = True Negative (correct background), Blue = False Negative (missed vessels), Yellow = True Positive (correct vessels), Red = False Positive (incorrect vessel prediction in background).

5.4 Limitations and future directions

Despite these achievements, several limitations remain and will be addressed in future work:

(1) Cross-dataset generalization. Cross-dataset generalization (e.g., training on one dataset and directly testing on another without target-domain fine-tuning) is important for clinical translation. Although our method shows stable performance across three heterogeneous public datasets under their standard splits, such in-domain multi-dataset evaluation provides only indirect evidence of robustness and does not replace strict cross-dataset transfer testing. In particular, DRIVE, CHASE_DB1, and STARE differ in imaging devices and acquisition protocols, field of view (FOV) and resolution, and annotation criteria/inter-grader variability, which may cause distribution mismatch and performance degradation under direct cross-dataset transfer. Future work will systematically evaluate settings such as train-on-DRIVE and test-on-STARE/CHASE_DB1 (without target-domain fine-tuning), and explore domain generalization/adaptation strategies to improve robustness in multi-device and multi-center scenarios.

(2) Search overhead of GA-Loss. Although GA-Loss searches within a finite and interpretable composite-loss space and uses elitism to stabilize the search, it still introduces additional optimization overhead compared with using a fixed loss. We will further improve the efficiency of the search procedure (e.g., early stopping, surrogate screening, or reduced evaluation schedules) while maintaining reproducibility under limited computational resources.

(3) Fitness signal and objective alignment. A limitation of the current GA-Loss framework is that the outer-loop fitness is based on validation loss rather than task-level segmentation metrics. Future work will investigate multi-objective or metric-aware fitness designs, such as incorporating F1-score, sensitivity, or their smooth surrogates. We will also explore lightweight global-context modeling (e.g., Transformer-inspired designs) as well as weakly supervised strategies to reduce annotation reliance and enhance clinical applicability.

(4) Dilation flexibility and stability of MFRPP. In this work, MFRPP adopts fixed dilation rates $d = \{1, 2, 4\}$ to ensure stable convergence under high-resolution training. More flexible receptive-field allocation (e.g., learnable or image-adaptive dilation) may further benefit vessels with diverse calibers and curvatures, but we observed occasional non-convergence when attempting adaptive dilation under CHASE_DB1-like settings. Future work will investigate stability-aware adaptive dilation mechanisms (e.g., constrained dilation candidates and regularization) and conduct budget-controlled ablations of alternative dilation sets to better characterize the accuracy–efficiency trade-off.

(5) Orientation modeling beyond fixed directions. Although the current DIR branch mainly uses horizontal/vertical strip kernels for efficiency and stability, learnable orientation modules may further improve the modeling of curved and arbitrary vessel orientations, which will be explored in future work.

6 Conclusions

This study presented DAMMGA-Net for retinal vessel segmentation under severe class imbalance, thin-structure fragility, and pronounced cross-scale variation. The proposed framework combines direction-aware mixed attention (DAMM), bottleneck-level fixed receptive-field multi-scale aggregation (MFRPP), and an offline genetic-algorithm-based loss search (GA-Loss) within a lightweight encoder–decoder architecture.

Experimental results on DRIVE, STARE, and CHASE_DB1 show that DAMMGA-Net achieves competitive overall performance and provides consistent advantages in continuity-sensitive vessel segmentation, particularly in reducing fragmentation-related false negatives and improving the recovery of thin vascular structures. The ablation and qualitative analyses further support the roles of DAMM and MFRPP as the main task-oriented architectural components, while GA-Loss contributes to better objective adaptation under class-imbalanced conditions.

At the same time, the current results also indicate that boundary-related false positives and precision–recall balancing remain the main challenges, especially on more demanding high-resolution data. Moreover, the present conclusions are supported by in-domain evaluations on three public datasets and should not yet be interpreted as evidence of strict cross-dataset or clinical generalization.

Future work will therefore focus on stronger false-positive control, more efficient and metric-aware loss search, and broader cross-device/cross-center validation to further improve robustness and practical applicability.

7 Data availability

All datasets used in this study are publicly available:

- (1) **DRIVE:** <https://drive.grand-challenge.org/>

- (2) **CHASE_DB1**: https://github.com/LvNianZu/CHASE_DB1/tree/main
- (3) **STARE**: <https://cecas.clemson.edu/~ahoover/stare/probing/index.html>

8 Code availability

The source code and implementation details of the DAMM-MFRPP framework and GA-Loss will be made available upon publication. During the peer-review stage, reasonable reproduction support can be provided by the corresponding author, Yajun Xie, upon request.

References

- [1] Chunhui Chen, Joon Huang Chuah, Raza Ali, and Yizhou Wang. Retinal vessel segmentation using deep learning: a review. *IEEE Access*, 9:111985–112004, 2021.
- [2] Arunakranthi Godishala, Veena Raj, Daphne Teck Ching Lai, and Hayati Yassin. Survey on retinal vessel segmentation. *Multimedia Tools and Applications*, 84(11):8361–8411, 2025.
- [3] Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, and Max A Viergever. Multiscale vessel enhancement filtering. In *International conference on medical image computing and computer-assisted intervention*, pages 130–137. Springer, 1998.
- [4] David Lesage, Elsa D Angelini, Isabelle Bloch, and Gareth Funka-Lea. A review of 3d vessel lumen segmentation techniques: Models, features and extraction schemes. *Medical image analysis*, 13(6):819–845, 2009.
- [5] Adrian Galdran, André Anjos, José Dolz, Hadi Chakor, Hervé Lombaert, and Ismail Ben Ayed. State-of-the-art retinal vessel segmentation with minimalistic models. *Scientific Reports*, 12(1):6174, 2022.
- [6] X Yin, J Ma, WK Zhang, et al. From u-net to transformer: progress in the application of hybrid models in medical image segmentation. *Laser & Optoelectronics Progress*, 62(2):0200001, 2025.
- [7] Xiaohan Ding, Xiangyu Zhang, Jungong Han, and Guiguang Ding. Scaling up your kernels to 31x31: Revisiting large kernel design in cnns. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11963–11975, 2022.
- [8] Feilong Tang, Zhongxing Xu, Qiming Huang, Jinfeng Wang, Xianxu Hou, Jionglong Su, and Jingxin Liu. Duat: Dual-aggregation transformer network for medical image segmentation. In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, pages 343–356. Springer, 2023.
- [9] Jia Guo, Ning Jia, and Jun Bai. Transformer based on channel-spatial attention for accurate classification of scenes in remote sensing image. *Scientific Reports*, 12(1):15473, 2022.
- [10] Yun Jiang, Jing Liang, Tongtong Cheng, Xin Lin, Yuan Zhang, and Jinkun Dong. Mtpa_unet: Multi-scale transformer-position attention retinal vessel segmentation network joint transformer and cnn. *Sensors*, 22(12):4592, 2022.
- [11] Yishuo Zhang and Albert CS Chung. Retinal vessel segmentation by a transformer-u-net hybrid model with dual-path decoder. *IEEE Journal of Biomedical and Health Informatics*, 28(9):5347–5359, 2024.
- [12] Nianzu Lv, Li Xu, Yuling Chen, Wei Sun, Jiya Tian, and Shuping Zhang. Tcddu-net: combining transformer and convolutional dual-path decoding u-net for retinal vessel segmentation. *Scientific Reports*, 14(1):25978, 2024.
- [13] Chuming Li, Xin Yuan, Chen Lin, Minghao Guo, Wei Wu, Junjie Yan, and Wanli Ouyang. Am-lfs: Automl for loss function search. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8410–8419, 2019.
- [14] Hao Li, Chenxin Tao, Xizhou Zhu, Xiaogang Wang, Gao Huang, and Jifeng Dai. Auto seg-loss: Searching metric surrogates for semantic segmentation. In *International Conference on Learning Representations (ICLR)*, 2021.
- [15] Hao Li, Tianwen Fu, Jifeng Dai, Hongsheng Li, Gao Huang, and Xizhou Zhu. Autoloss-zero: Searching loss functions from scratch for generic tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1009–1018, 2022.
- [16] Xinyu Guo, Kai Wu, Xiaoyu Zhang, and Jing Liu. Automated loss function search for class-imbalanced node classification. In *International Conference on Machine Learning (ICML 2024)*, volume 235 of *Proceedings of Machine Learning Research*, Vienna, Austria, 2024.

- [17] Lei Mou, Yitian Zhao, Huazhu Fu, Yonghuai Liu, Jun Cheng, Yalin Zheng, Pan Su, Jianlong Yang, Li Chen, Alejandro F Frangi, et al. Cs2-net: Deep learning segmentation of curvilinear structures in medical imaging. *Medical image analysis*, 67:101874, 2021.
- [18] Qibin Hou, Daquan Zhou, and Jiashi Feng. Coordinate attention for efficient mobile network design. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13713–13722, 2021.
- [19] Huajun Liu, Fuqiang Liu, Xinyi Fan, and Dong Huang. Polarized self-attention: Towards high-quality pixel-wise mapping. *Neurocomputing*, 506:158–167, 2022.
- [20] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.
- [21] Zaiwang Gu, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, and Jiang Liu. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE Transactions on Medical Imaging*, 38(10):2281–2292, 2019.
- [22] Debesh Jha, Michael A. Riegler, Dag Johansen, Pål Halvorsen, and Håvard D. Johansen. Doubleu-net: A deep convolutional neural network for medical image segmentation. In *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, 2020.
- [23] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.
- [24] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.
- [25] Yihan Wang, Muyang Li, Han Cai, Wei-Ming Chen, and Song Han. Lite pose: Efficient architecture design for 2d human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13116–13126, 2022.
- [26] Xiaolong Hu, Liejun Wang, Shuli Cheng, and Yongming Li. Hdc-net: A hierarchical dilation convolutional network for retinal vessel segmentation. *PLOS ONE*, 16(9):e0257013, 2021.
- [27] Yuchen Yuan, Lei Zhang, Lituan Wang, and Haiying Huang. Multi-level attention network for retinal vessel segmentation. *IEEE Journal of Biomedical and Health Informatics*, 26(1):312–323, 2021.
- [28] Joes Staal, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram Van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging*, 23(4):501–509, 2004.
- [29] Muhammad Moazam Fraz, Paolo Remagnino, Andreas Hoppe, Bunyarit Uyyanonvara, Alicja R Rudnicka, Christopher G Owen, and Sarah A Barman. An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Transactions on Biomedical Engineering*, 59(9):2538–2548, 2012.
- [30] AD Hoover, Valentina Kouznetsova, and Michael Goldbaum. Locating blood vessels in retinal images by piece-wise threshold probing of a matched filter response. *IEEE Transactions on Medical imaging*, 19(3):203–210, 2000.
- [31] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015.
- [32] Qiangguo Jin, Zhaopeng Meng, Tuan D Pham, Qi Chen, Leyi Wei, and Ran Su. Dunet: A deformable network for retinal vessel segmentation. *Knowledge-Based Systems*, 178:149–162, 2019.
- [33] Yixuan Chen, Yuhan Dong, Yi Zhang, and Kai Zhang. Rna-net: Residual nonlocal attention network for retinal vessel segmentation. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1560–1565. IEEE, 2020.
- [34] Dongyi Wang, Ayman Haytham, Jessica Pottenburgh, Osamah Saeedi, and Yang Tao. Hard attention net for automatic retinal vessel segmentation. *IEEE Journal of Biomedical and Health Informatics*, 24(12):3384–3396, 2020.
- [35] Xiang Li, Yuchen Jiang, Minglei Li, and Shen Yin. Lightweight attention convolutional neural network for retinal vessel image segmentation. *IEEE Transactions on Industrial Informatics*, 17(3):1958–1967, 2020.
- [36] Yuan Zhang, Miao He, Zhineng Chen, Kai Hu, Xuanya Li, and Xieping Gao. Bridge-net: Context-involved u-net with patch-based loss weight mapping for retinal blood vessel segmentation. *Expert Systems with Applications*, 195:116526, 2022.
- [37] Xin Yang, Zhiqiang Li, Yingqing Guo, and Dake Zhou. Dcu-net: a deformable convolutional neural network based on cascade u-net for retinal vessel segmentation. *Multimedia Tools and Applications*, 81(11):15593–15607, 2022.

- [38] Yanhong Liu, Ji Shen, Lei Yang, Guibin Bian, and Hongnian Yu. Resdo-unet: A deep residual network for accurate retinal vessel segmentation from fundus images. *Biomedical Signal Processing and Control*, 79:104087, 2023.
- [39] Yanhong Liu, Ji Shen, Lei Yang, Hongnian Yu, and Guibin Bian. Wave-net: A lightweight deep network for retinal vessel segmentation from fundus images. *Computers in biology and medicine*, 152:106341, 2023.
- [40] Weiping Ding, Ying Sun, Jiashuang Huang, Hengrong Ju, Chongsheng Zhang, Guang Yang, and Chin-Teng Lin. Rcar-unet: Retinal vessel segmentation network algorithm via novel rough attention mechanism. *Information Sciences*, 657:120007, 2024.
- [41] Linfeng Kong and Yun Wu. Rvs-fdsc: A retinal vessel segmentation method with four-directional strip convolution to enhance feature extraction. *Biomedical Signal Processing and Control*, 95:106296, 2024.
- [42] WenHao Zhang, ShaoJun Qu, and YueWen Feng. Lmfr-net: lightweight multi-scale feature refinement network for retinal vessel segmentation. *Pattern Analysis and Applications*, 28(2):44, 2025.
- [43] Xuebing Luo, Lingxi Peng, Ziyang Ke, Jinhui Lin, and Zhiwen Yu. Pa-net: A hybrid architecture for retinal vessel segmentation. *Pattern Recognition*, 161:111254, 2025.
- [44] Jiahao Wang, Xiaobo Li, and Zhendi Ma. Multi-scale three-path network (mstp-net): A new architecture for retinal vessel segmentation. *Measurement*, 250:117100, 2025.

A Composite Loss Search: Inputs and Primitive Algorithms

A.1 Primitive Algorithms (used by GA outer loop)

Algorithm 3 ARITHMETIC-CROSSOVER

Require: Parents $\lambda^{(p1)} = (\mathbf{w}^{(p1)}, \phi^{(p1)})$, $\lambda^{(p2)} = (\mathbf{w}^{(p2)}, \phi^{(p2)})$

Ensure: Child $\lambda^{(c)} = (\mathbf{w}^{(c)}, \phi^{(c)})$

- 1: Sample $\alpha_c \sim \mathcal{U}(0, 1)$
 - 2: $\mathbf{w}^{(c)} \leftarrow \alpha_c \mathbf{w}^{(p1)} + (1 - \alpha_c) \mathbf{w}^{(p2)}$
 - 3: $\phi^{(c)} \leftarrow \alpha_c \phi^{(p1)} + (1 - \alpha_c) \phi^{(p2)}$
 - 4: **return** $\lambda^{(c)}$
-

Algorithm 4 MUTATE

Require: Individual $\lambda = (\mathbf{w}, \phi)$; mutation scales σ_w, σ_ϕ ; toggle prob. p_{toggle} ; optional max toggles k_{max}

Ensure: Mutated $\lambda = (\mathbf{w}, \phi)$

- 1: $\mathbf{w} \leftarrow \mathbf{w} + \mathcal{N}(0, \sigma_w^2 I)$ ▷ Gaussian perturbation on weights
 - 2: $\phi \leftarrow \phi + \mathcal{N}(0, \sigma_\phi^2 I)$ ▷ Gaussian perturbation on internal params
 - 3: With prob. p_{toggle} , pick up to k_{max} indices $\mathcal{S}$ and set $\mathbf{w}_i \leftarrow 0$ for $i \in \mathcal{S}$
 - 4: **return** $\lambda = (\mathbf{w}, \phi)$
-

Algorithm 5 PROJECT-TO-SIMPLEX

Require: $u \in \mathbb{R}^m$

Ensure: $\mathbf{w} \in \Delta^{m-1} = \{\mathbf{w} \geq 0, \sum_{i=1}^m \mathbf{w}_i = 1\}$

- 1: $\mu \leftarrow \text{SORT}(u)$ in descending order; $s_j \leftarrow \sum_{r=1}^j \mu_r$
 - 2: $\rho \leftarrow \max \{j \in \{1, \dots, m\} : \mu_j - \frac{1}{j}(s_j - 1) > 0\}$
 - 3: $\theta \leftarrow \frac{1}{\rho}(s_\rho - 1)$
 - 4: **for** $i = 1$ to m **do**
 - 5: $\mathbf{w}_i \leftarrow \max(u_i - \theta, 0)$
 - 6: **end for**
 - 7: **return** $\mathbf{w}$
-

Algorithm 6 CLIP

Require: $\phi \in \mathbb{R}^q$; lower bound $\phi_{\min}$; upper bound $\phi_{\max}$ (component-wise allowed)

Ensure: Clipped ϕ

- 1: **for** $i = 1$ to q **do**
 - 2: $\phi_i \leftarrow \min\{\max\{\phi_i, (\phi_{\min})_i\}, (\phi_{\max})_i\}$
 - 3: **end for**
 - 4: **return** ϕ
-

Algorithm 7 ROULETTE WHEEL

Require: Fitness values $\{f_i\}_{i=1}^N$ (smaller is better); smoothing $\delta > 0$

Ensure: One sampled index $j \in \{1, \dots, N\}$ (repeat to obtain a parent pair)

- 1: $s_i \leftarrow 1/(f_i + \delta)$; $p_i \leftarrow s_i / \sum_{r=1}^N s_r$
 - 2: Cumulative distribution $c_j \leftarrow \sum_{i=1}^j p_i, j = 1, \dots, N$
 - 3: Sample $u \sim \mathcal{U}(0, 1)$ and return the smallest j such that $c_j \geq u$
-

Table S1: Ablation on strip-kernel sizes in the directional branch (DIR) on CHASE_DB1. Results are reported as mean $\pm$ std over three random seeds (11/21/42). All settings are identical except the strip-kernel set; diagonal kernels are disabled in this table.

dir_ks	Sen	F1	AUC	Prec	Sp
3	0.8082 $\pm$ 0.0115	0.7895 $\pm$ 0.0030	0.9849 $\pm$ 0.0013	0.7718 $\pm$ 0.0130	0.9839 $\pm$ 0.0014
5	0.8114 $\pm$ 0.0004	0.7978 $\pm$ 0.0041	0.9874 $\pm$ 0.0011	0.7848 $\pm$ 0.0077	0.9850 $\pm$ 0.0007
7	0.8141 $\pm$ 0.0013	0.7989 $\pm$ 0.0020	0.9855 $\pm$ 0.0017	0.7844 $\pm$ 0.0036	0.9849 $\pm$ 0.0004
9	0.8144 $\pm$ 0.0057	0.7962 $\pm$ 0.0030	0.9866 $\pm$ 0.0008	0.7788 $\pm$ 0.0016	0.9845 $\pm$ 0.0002
5,7	0.8194 $\pm$ 0.0041	0.7984 $\pm$ 0.0012	0.9875 $\pm$ 0.0005	0.7784 $\pm$ 0.0025	0.9843 $\pm$ 0.0003
3,5,7	0.8124 $\pm$ 0.0110	0.7971 $\pm$ 0.0034	0.9870 $\pm$ 0.0010	0.7825 $\pm$ 0.0077	0.9848 $\pm$ 0.0009

Table S2: Resource and reproducibility comparison among loss-optimization strategies.

Dimension	Manual tuning	AutoML loss generation [†]			GA-Loss (ours)
		AM-LFS	AutoSeg-Loss	AutoLoss-Zero	
Representative examples	BCE+Dice (fixed)	AM-LFS	AutoSeg-Loss	AutoLoss-Zero	Constrained composite-loss search
Search space	Fixed / trial-and-error	Large / unconstrained	Large / unconstrained	Large / unconstrained	Bounded (weights + internal params)
Compute requirement	Single GPU (many reruns)	64 $\times$ GTX 1080Ti	8 $\times$ Tesla V100	4 $\times$ V100	Single GPU (RTX 4060, offline)
Reported / measured search time	N/A	Not separately reported Online bilevel search 32 sampled instances/round	$\sim$ 8.0 h	$\sim$ 48.0 h 300 proxy-task evals	$\sim$ 8.0 h 20 gen, 100 eval [‡]
Additional training cost	Multiple full trainings	Search coupled with training	Many trainings in search	Many trainings in search	Bounded evals + one final training
Inference overhead	None	None	None	None	None
Interpretability	High	Low-Medium	Low-Medium	Low-Medium	High
Reproducibility	User-dependent	Implementation-dependent	Implementation-dependent	Implementation-dependent	High (fixed budget)
Suitability (limited compute)	Moderate	Limited	Limited	Limited	High

Note: [†] Resource usage and search time for AM-LFS, AutoSeg-Loss, and AutoLoss-Zero are reported from their original papers. Their practical cost depends on implementation details, search-space size, proxy setting, and cluster scale. In particular, AM-LFS performs online bilevel search during training and does not report an isolated wall-clock search time.

[‡] GA-Loss cost is measured in our environment (single RTX 4060 GPU). The search is performed offline with a fixed budget of 20 generations and 100 evaluations, and does not increase inference-time complexity.

Table S3: Ablation on diagonal kernels (45 $^\circ$ /135 $^\circ$) in the directional branch on CHASE_DB1. Results are mean $\pm$ std over three random seeds (11/21/42). Here, diagonal kernels are enabled with diag_k matched to dir_ks (i.e., 3/5/7/9).

dir_ks	Sen	F1	AUC	Prec	Sp
3	0.8157 $\pm$ 0.0014	0.7956 $\pm$ 0.0010	0.9872 $\pm$ 0.0004	0.7765 $\pm$ 0.0010	0.9842 $\pm$ 0.0001
5	0.8134 $\pm$ 0.0081	0.7965 $\pm$ 0.0033	0.9861 $\pm$ 0.0011	0.7803 $\pm$ 0.0074	0.9846 $\pm$ 0.0008
7	0.8125 $\pm$ 0.0079	0.7946 $\pm$ 0.0038	0.9868 $\pm$ 0.0006	0.7776 $\pm$ 0.0046	0.9844 $\pm$ 0.0005
9	0.8112 $\pm$ 0.0114	0.7947 $\pm$ 0.0034	0.9870 $\pm$ 0.0011	0.7790 $\pm$ 0.0066	0.9845 $\pm$ 0.0008

Table S4: Paired differences between diagonal-enabled and strip-only models on CHASE_DB1. Δ is computed as (diag - strip) under identical settings, paired by dir_ks and seed (11/21/42). Values are reported as mean $\pm$ std across 3 seeds for each dir_ks. The Overall row is computed over all 12 paired runs (4 dir_ks $\times$ 3 seeds).

dir_ks	Δ F1 (mean $\pm$ std)	Δ Sen (mean $\pm$ std)	Δ AUC (mean $\pm$ std)	F1 Improved	F1 Decreased
3	+0.0062 $\pm$ 0.0037	+0.0075 $\pm$ 0.0101	+0.0024 $\pm$ 0.0016	3/3	0/3
5	-0.0014 $\pm$ 0.0008	+0.0020 $\pm$ 0.0084	-0.0013 $\pm$ 0.0007	0/3	3/3
7	-0.0043 $\pm$ 0.0029	-0.0016 $\pm$ 0.0072	+0.0012 $\pm$ 0.0020	0/3	3/3
9	-0.0015 $\pm$ 0.0063	-0.0032 $\pm$ 0.0165	+0.0003 $\pm$ 0.0017	1/3	2/3
Overall	-0.0002 $\pm$ 0.0053	+0.0012 $\pm$ 0.0104	+0.0007 $\pm$ 0.0019	4/12	8/12