Dynamic Statistical Attention-based lightweight model for Retinal Vessel Segmentation: DyStA-RetNet

Amit Bhati [a] , Samir Jain [a] , Neha Gour [b] , Pritee Khanna [a],*, Aparajita Ojha [a] ,
Naoufel Werghi [b]

[a] *PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur 482005, India*
[b] *Khalifa University, Abu Dhabi, United Arab Emirates*

ARTICLE INFO

ABSTRACT

Background and Objective: Accurate extraction of retinal vascular components is vital in diagnosing and treating retinal diseases. Achieving precise segmentation of retinal blood vessels is challenging due to their complex structure and overlapping vessels with other anatomical features. Existing deep neural networks often suffer from false positives at vessel branches or missing fragile vessel patterns. Also, deployment of the existing models in resource-constrained environments is challenging due to their computational complexity. An attention-based and computationally efficient architecture is proposed in this work to bridge this gap while enabling improved segmentation of retinal vascular structures.

Methods: The proposed dynamic statistical attention-based lightweight model for retinal vessel segmentation (DyStA-RetNet) employs a shallow CNN-based encoder–decoder architecture. One branch of the decoder utilizes a partial decoder connecting encoder layers with decoder layers to allow the transfer of high-level semantic information, whereas the other branch helps to incorporate low-level information. The multi-scale dynamic attention block empowers the network to accurately identify different-sized tree-shaped vessel patterns during the reconstruction phase in the decoder. The statistical spatial attention block improves the feature learning capability. By effectively integrating low-level and high-level semantic information, DYStA-RetNet significantly improves the performance of vessel segmentation.

Results: Experiments performed on four benchmark datasets (DRIVE, STARE, CHASEDB, and HRF) exhibit the adaptability of DYStA-RetNet for clinical applications with a significantly smaller number of trainable parameters (37.19K) and GFLOPS (0.75), and superior segmentation performance.

Conclusion: The proposed lightweight CNN-based DYStA-RetNet efficiently extracts complex retinal vascular components from fundus images. It is computationally efficient and deployable in resource-constrained environments.

1. Introduction

Retinal vessels are crucial components of the human blood circulation system, which offer a unique opportunity for immediate and non-invasive observation. These vessels possess key features like shape, tortuosity, and a distinctive structural tree-like pattern, which are crucial for early-stage diagnosis of ocular diseases [1]. Abnormalities such as artery thickening, retinal capillary increase, and hemorrhages in the vessel structure can be observed in patients with diabetic retinopathy. Furthermore, retinal vasculature can provide valuable predictive information about patient age, haemoglobin level, blood pressure, and other heart disease parameters, aiding in disease diagnosis and risk identification [2].

Manual extraction of retinal vessels is labor-intensive and time-consuming, emphasizing the need for efficient and rapid vascular segmentation algorithms that utilize both conventional and data-driven approaches [3]. Conventional methods employ hand-crafted filters such as Gabor and Gaussian to extract semantic information for vessel clustering and segmentation. Recently emerged Data-driven methods are leveraging U-Net based models and their variants to achieve remarkable performance compared to traditional approaches [4]. These deep learning methods focus on enhancing the design of U-Net structures to capture superior feature representations or decoupling the structure and textures of retinal images. However, available methods often

---

face challenges such as huge training parameters and complex network structures, making them infeasible to use in resource-constrained environments.

To enhance diagnostic capabilities for ocular diseases, this work presents a lightweight, attention-based shallow encoder–decoder model for efficient retinal blood vessel segmentation. The objective is to improve the accuracy and efficiency of vessel segmentation while reducing memory usage and computational complexity. The key contributions of this work are:

1. A lightweight shallow encoder–decoder architecture to extract retinal vessels from fundus images.
2. A self-attention mechanism-based multi-scale dynamic statistical attention module to efficiently identify the positions of pixels representing the complex tree-shaped vessel patterns during the reconstruction phase.
3. A modified version of a partial decoder to allow the proposed network to flow high-level features directly into the decoder part for improved learning.
4. Experimental analysis on four clinically acquired benchmark datasets.
5. A cross-dataset study to validate the performance of the proposed model on unseen data of different ethnicities.

2. Related work

Convolutional Neural Network (CNN), particularly fully convolutional networks (FCNs), achieved remarkable progress in enhancing medical image segmentation [5–7]. However, it is noticed that FCNs occasionally produced inaccurate boundaries due to de-convolution based upsampling operations on small-sized feature maps. Ronneberger et al. [8] resolved this issue by providing a skip connection between encoder–decoder blocks. This refined architecture allowed more information flow between encoder–decoder resulting in better segmentation results. Today, this U-Net architecture has been widely adopted for end-to-end retinal vessel segmentation [6,8–12].

Yan et al. [9] introduced a U-Net based architecture with a joint loss function that combines pixel-wise and segmentation-based losses. This approach successfully balanced the importance of thick and thin vessels during training and achieved good segmentation performance on DRIVE [13] and CHASE-DB 1 [14] datasets. However, large kernel sizes (5×5 and 7×7) used in the model increased memory and computational requirements. Mou et al. [15] trained a dilated attention-based U-Net model that achieved impressive vessel reconstruction but with a larger memory footprint due to its 56.03M trainable parameters. Zhang et al. [16] proposed a dual CNN-based BCU-Net combining U-Net and ConvNeXt. The U-Net branch captured local contextual information efficiently and the ConvNeXt branch refined global features. Bridging two models increases its size (102.3M parameters) and computations (460.6G FLOPS).

The need for lightweight models in medical image segmentation originates from the need to deploy these systems in resource-constrained situations such as mobile devices, edge computing platforms, and telemedicine applications. In such scenarios, traditional deep learning models with high memory and compute requirements can often be unsuitable for real-time processing. In order to fill this gap, lightweight models optimize their architectures to use fewer parameters and FLOPs without substantially sacrificing segmentation accuracy. In medical imaging, where timely and accurate analysis is crucial, these models are especially important. To address this issue, Laibacher et al. [6] introduced a lightweight MobileNet-based M2U-Net architecture with reduced trainable parameters but experienced a marginal decrease in performance. Galdran et al. [12] introduced a shallow W-Net architecture consisting of a pair of U-Net models, which achieved comparable performance with only 64K trainable parameters. Despite a smaller memory footprint, cascading of U-Net pairs resulted

in a higher FLOPs count, i.e., 1.5G FLOPS. Liu et al. [17] also proposed a lightweight model, WaveNet, which includes a detail enhancement and denoising block (DED) and a multi-scale feature fusion block (MFF). The DED block handles the information loss in micro-vessel structures and allows the model to extract more semantic information about thin vessel structures, whereas the MFF block utilizes multi-scale feature fusion for improved segmentation learning.

Li et al. [18] proposed a multi-scale residual similarity gathering and response cue erasing method that outperformed state-of-the-art methods on retinal vessel branching. All these methods suffered from a large number of trainable parameters, making them unsuitable for resource-constrained environments. Li et al. [19] proposed a multi-dimensional attention-based segmentation model for vessel information enhancement. They proposed deep guidance fusion (DGF) and a cross-pooling semantic enhancement (CPSE) block for multi-scale feature refinement. Although their model is shallow and computationally optimized, it suffers from incorrect segmentation due to patch-cropping operation.

Dual-path networks are commonly used in retinal vessel segmentation for effectively capturing both local and global features, which help the model for precise segmentation of complicated vessel structures. Wang et al. [11] proposed a dual encoder-based segmentation model that fused multi-scale low-level and high-level features for improved network learning. This computationally intensive approach outperformed the method proposed by Yan et al. [9]. Li et al. [20] proposed a dual encoder-based edge enhancement graph convolution network. Their dual encoder detects vessel edges and preserves those during the downsampling process. The graph convolution module utilizes channel information by capturing topological relationship of pixels across the channels.

Li et al. [21] proposed a dual-path progressive fusion network for precise segmentation of retinal vessels. Its dual-path encoder comprises a CNN path for local feature detection and a recurrent convolutional path for contextual information extraction. The model utilizes interactive fusion, cross-layer fusion, and scale feature fusion blocks to aggregate detailed and contextual features for better learning. Huang et al. [22] proposed a double branch fusion U-Net with a hard example weighting strategy and a random channel attention mechanism to enhance segmentation and reduce overfitting. In order to enhance learning, one of the branches employs a hard example weighting technique, which prioritizes difficult cases. Although the approach is effective, its efficiency is limited by its reliance on computationally costly morphological processes. Jian et al. [23] proposed the dual-branch U-Net framework for medical image segmentation, enhancing the traditional U-Net with a parallel dual-branch encoder and attention module for better feature extraction and noise reduction. To enhance information fusion, the model substitutes pooling with convolution in the lower layers of the model. However, the attention mechanism and convolutions increase the amount of parameters. Furthermore, slight vessel fragmentation in the prediction results is also observed.

Recently, transformer-based models have become a promising method for medical image segmentation because of their capacity to capture long-range dependencies and integrate local–global features through self-attention mechanisms. Lin et al. [24] proposed a transformer-based model, known as Stimulus-Guided Adaptive Transformer Network (SGAT-Net), which focuses on precise retinal blood vessel segmentation. Central to their approach is the incorporation of a Stimulus-Guided Adaptive Module (SGA-Module), which effectively captures local–global features by leveraging inductive bias and a self-attention mechanism. Despite the promising capabilities of SGAT-Net, the model faces challenges in distinguishing fine retinal blood vessels from the background.

Despite progress in retinal vessel segmentation, most methods suffer from common limitations: (1) Fragmented vessel segmentation: Existing methods often produce discontinuous vessel segments, failing to connect thin and intricate vessel structures effectively. (2) High
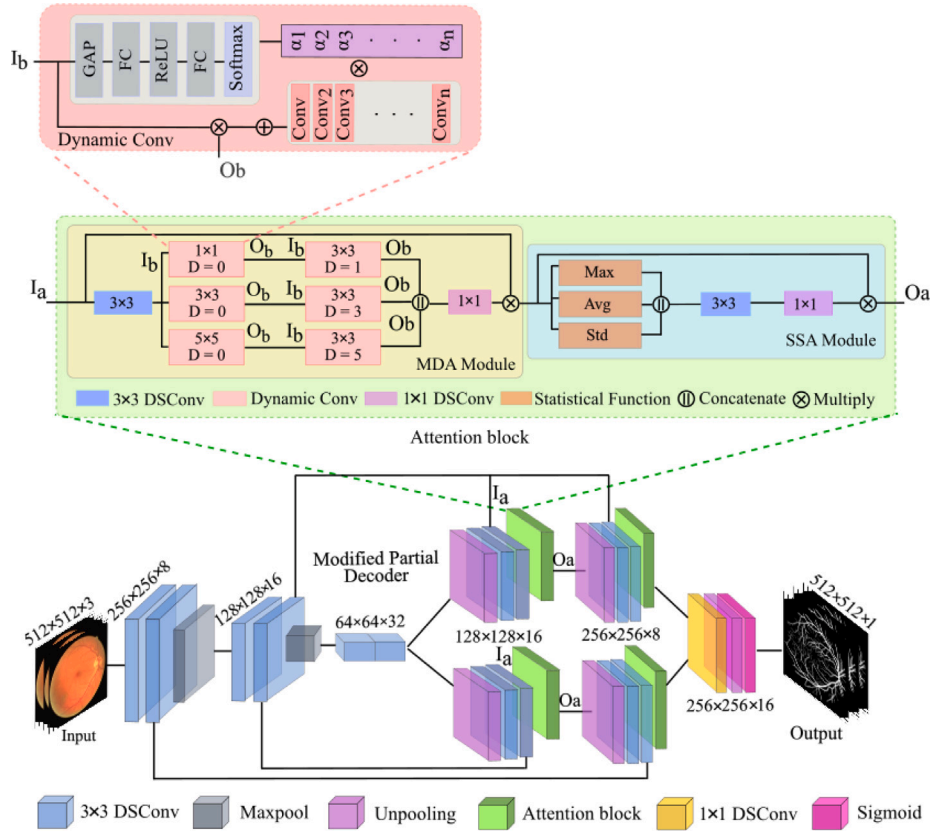
Fig. 1. Block diagram of the proposed DYStA-RetNet for retinal vessel segmentation.

computational costs: Many methods rely on large kernel sizes, deep networks, or complex fusion mechanisms, which increase the number of parameters and computational overhead, making them unsuitable for resource-constrained environments. (3) Imbalance in vessel and background pixels: The uneven distribution of vessel and background pixels in fundus images causes segmentation errors, particularly in thin vessel regions. The proposed Dynamic Statistical Attention-based Lightweight model (DYStA-RetNet) addresses these challenges by learning intricate vessel tree structures and focusing on the branching and break-point areas using a dynamic statistical attention mechanism. The proposed attention module learns complex vessel patterns and vessel breakpoints with minimal computational overhead. Simulation results show that the proposed model achieved better performance over state-of-the-art methods on benchmark datasets (DRIVE [13], CHASE-DB1 [14], HRF [25], and STARE [26]).

3. Method

3.1. Architecture of the proposed DYStA-RetNet

The proposed DYStA-RetNet shown in Fig. 1 is a computationally efficient shallow CNN architecture. It aims to enhance the extraction of vessel branching features and expand the network's capacity through the use of re-calibrated weighted attention maps for decoding vessel structures. Notably, the decoder section of DYStA-RetNet leverages the attention block, which employs statistical spatial attention (SSA) and multi-scale dynamic attention (MDA) operations to further enhance the encoded features. As shown in Fig. 1, the encoder part consists of three convolution layers, while the decoder comprises two convolution layers. Also, depth-wise separable convolution (DSConv) has proven its capabilities in reducing model size and computation [27]. DSConv factorizes a standard convolution operation into a depthwise convolution and a 1×1 pointwise convolution. This factorization significantly

reduces both computation and model size [28]. Inspired by this, the proposed model incorporates a series of residual 3×3 DSConv instead of standard 3×3 convolutions in both encoder and decoder modules. Batch normalization is commonly used to normalize feature maps by computing mean and variance on batches. While batch normalization is effective in preventing over-fitting, it may not perform well when the batch size is small [29,30]. To address this issue, group normalization can be used as an alternative. Group normalization divides channels into groups and normalizes their features within each group, making the outcome independent of the batch dimension.

In the proposed network, group normalization layers are incorporated in the encoder–decoder modules to mitigate the limitations of batch normalization. In the encoder module, max-pooling with a factor of 2 is used to downsample the input. The decoder module, on the other hand, utilizes transposed convolution operations to upsample the input and learns the pixel mapping corresponding to the extracted features from the encoder. The shallow network is configured with 8, 16, and 32 filters in the encoder convolution layers, while the decoder layers have 16 and 8 filters.

Low-level features require more computations due to their larger spatial resolutions. However, their contribution to performance may vary depending on the specific task. Conversely, high-level features obtained from deeper layers of the encoder capture more abstract and complex information, making them highly valuable for semantic understanding. By integrating the generated saliency map with the decoder layers, prediction performance and computational efficiency [31] of the network can be enhanced.

In the proposed DYStA-RetNet, as shown in Fig. 1, the decoder module splits into two parts. One part of the decoder allows only high-level semantic information to improve pixel placement for vessel reconstruction. This modified partial decoder (MPD) block enables faster and more accurate vessel branching pattern segmentation. In the other part, the features encoded from the encoder unit are passed to the

decoder for incorporating both low-level and high-level semantic features. In each decoder layer, the integrated features are passed through an attention block, which applies multi-scale dynamic attention and statistical spatial attention on the feature map at each stage. Finally, the output of both decoders concatenated and passed through a 1×1 convolution followed by unpooling and sigmoid activation function to generate segmentation masks. The detailed architecture of the attention block is presented in the next subsection.

Algorithm 1 Algorithm for Dynamic Statistical Attention

1: Compute 3×3 DSConv on input feature map $F_{H \times W}$
2: Apply a series $1 \times 1, 3 \times 3$ and 5×5 Dynamic convolution to obtained refined multi-scale features M_x, M_y, M_z
3: Concatenate M_x, M_y, M_z and apply 1×1 DSConv
4: Perform element-wise multiplication on input feature map and weighted matrix obtained from step 3 to get S
5: Initialize $S_m, S_a, S_s, i, j = 0$
6: **while** $i \leq H$ and $j \leq W$ **do**
7:     **for all** pixel $p \in S$ at loc(i,j) **do**
8:         Compute Max(p) and update S_m at (i, j)
9:         Compute Avg(p) and update S_a at (i, j)
10:        Compute Std(p) and update S_s at (i, j)
11: Concatenate S_m, S_a, S_s to obtain weighted matrix G
12: Apply 3×3 and 1×1 DSConv on G
13: Perform element-wise multiplication of re-calibrated weighted matrix obtained with step 12 with weighted matrix S

3.2. Proposed attention (MDA+SSA) module

Common factors that contribute to the discontinuity of retinal vessel segmentation include uneven brightness in thin vessel regions and background color similarity, where there is minimal contrast between fine vessels. These challenges make it difficult for the model to differentiate between the background and the vessels. Therefore, the network must learn patterns associated with the branch breakpoints of the vessels, enabling accurate identification. This can be achieved by directing the network's attention to these specific areas and extracting relevant information. As shown in Fig. 1, the proposed segmentation model incorporates the multi-scale dynamic attention (MDA) and statistical spatial attention (SSA) module for feature enhancement.

The MDA module utilizes the power of dynamic convolution [32] which uses squeeze-excitation block to generate attention weights for kernels. The generated convolution kernels are then aggregated together and so obtained final attention weights are multiplied with the input feature map. To capture multi-scale features along with channel redundant information reduction, the MDA module uses different scale (1, 3, 5) dilation rates in dynamic convolution to allow the model to learn different sized features efficiently as given in Algorithm 1. The attention weights for different receptive size kernel convolutions are aggregated to obtain MDA attention weight. These attention weights are then multiplied with the input feature map to get enhanced features. By aggregating multi-scale features and applying attention mechanisms, the MDA module ensures that redundant feature (highly similar) maps are down-weighted, while preserving diverse and significant features that enhance the model's focus on critical areas like vessel branching and break-points. Dynamic attention enables the MDA module to learn attention weights dynamically for emphasizing diverse and significant features. The MDA module enables the network to focus more on vessel branching and break-point areas. Feature maps generated by a convolutional layer inherently act as a feature detector.

By Analyzing the pixel value statistics across the feature maps, valuable insights into the similarities and variations in feature learning across channels at each pixel position can be obtained. The SSA module applies max, average and standard deviation statistical functions across

the channel on input feature maps to collect statistical information for pixel locations. The max operation in a pixel location across all channels signifies the dominant pixel values, whereas the average provides an average value for that location. Standard deviation (Std) is also applied to understand the distribution of intensity at a pixel position across various channels. Std operator allows the model to learn the contribution of each feature map uniquely to the learned representation. The learning capability of the CNN model gets better as it observes diverse data patterns. The obtained statistical attention weights are concatenated followed by 3×3 and 1×1 convolution to obtain SSA attention weights as shown in Fig. 1. These weight matrices are then multiplied with feature maps input in SSA to get enhanced feature maps providing support to the decoder layer for vessel pixel reconstruction.

4. Results

4.1. Datasets

The model is validated on four publicly available fundus image benchmark datasets. **DRIVE** [13] dataset has 40 fundus images including seven pathological cases with a resolution of 768×584 pixels. These fundus images were captured using a Canon CR5 non-dilated 3 CCD camera, with a field of view (FOV) spanning 45 degrees. Each image's FOV is circular, with a diameter measuring 540 pixels, and the dataset includes corresponding masked images representing the FOV. Following the standard protocols used in SOTA methods [17,19,22], the original train–test split is used in the experimentation for the DRIVE dataset. **HRF** [25] dataset contains 45 fundus images 3504×3504 pixels for normal, glaucoma, and diabetic retinopathy classes. Following the standard protocols used in SOTA methods [12], 15 out of 45 images are used for training. **CHASE-DB1** [14] dataset comprises 14 pairs of color retinal images of 999×960 pixels. Following the standard protocols used in SOTA methods [17,19,22], 20 and 8 images are used for training and testing respectively. **STARE** [26] dataset comprises 40 fundus images with a resolution of 605×700 pixels, accompanied by meticulously hand-labeled ground truth data, specifically for vessel segmentation. Following the standard protocols used in SOTA methods [17,19,22], 15 and 5 images are used for training and testing respectively. The images in these datasets have different illumination conditions and hazy vessel patterns, which makes vessel segmentation more complex.

4.2. Basic settings

The proposed DYStA-RetNet is implemented on NVidia DGX-A100 having AMD ROM 7742, 2.25 GHz 128 cores CPU with 40 GB memory size. It utilizes Adam optimizer with an initial learning rate of $1e-4$ and a decay rate of $1e-5$ for the vanishing gradient issue. The model is trained for 100 epochs with a batch size of 4 using the DICE loss function. The performance of the proposed model is evaluated on AUC, F1, and Dice Score. AUC evaluates the discriminating capability of the model ranging between 0 and 1. The F1 score is defined as the harmonic mean of precision and recall. Similar to AUC, F1 score also ranges from 0 to 1. In contrast, the dice score measures the similarity between the predicted label to the ground truth. A higher score of AUC, F1, and DICE score indicates a better performance.

4.3. Ablation study

An ablation study is performed to observe the role and impact of the proposed MPD block and attention module on different components of DYStA-RetNet. The first variant of DYStA-RetNet is performed without MPD and attention block. MPD is used in the second variant but without attention. The partial decoder effectively transfers the high-level feature information from the encoder to the decoder module, resulting

Table 1
Ablation study of the proposed DYStA-RetNet.

Method	DRIVE			HRF			CHASE-DB 1			STARE		
	F1	AUC	DICE	F1	AUC	DICE	F1	AUC	DICE	F1	AUC	DICE
Experiment[a]	81.93	97.18	80.86	77.66	98.17	80.43	81.80	96.49	81.01	77.37	96.05	75.23
Experiment[b]	82.41	97.79	81.55	78.35	98.98	81.41	82.04	97.18	81.89	78.08	97.52	76.88
Experiment[c]	82.78	98.05	83.89	80.23	98.67	83.11	84.08	98.05	82.02	78.61	97.86	76.92
Experiment[d]	84.81	98.86	84.68	81.92	99.14	83.89	85.25	99.31	84.01	86.18	99.12	82.27

[a] Variant 1 (Without MPD branch, without attention).
[b] Variant 2 (With MPD branch, without attention).
[c] Variant 3 (With MPD branch, with attention placed at Skip Connection).
[d] Variant 4 (With MPD branch, with attention placed at feature fusion layer) - Proposed.

Table 2
Performance of the proposed attention module (MDA+SSA) in comparison to state-of-the-art attention methods.

Dataset	CBAM [33]		SE [34]		SFA [35]		MDA+SSA	
	AUC	DICE	AUC	DICE	AUC	DICE	AUC	DICE
DRIVE	96.63	79.76	95.4	77.63	98.07	83.93	**98.86**	**84.68**
HRF	97.52	80.52	97.73	81.07	98.69	86.18	**99.14**	**83.89**
CHASE-DB1	96.12	79.89	95.16	77.05	98.86	82.95	**99.31**	**84.01**
STARE	95.82	76.37	94.65	74.13	98.61	81.21	**99.12**	**82.27**

in better reconstruction of vessel structure. In the third variant, the attention module is included in MPD but at the skip connection layer, which refines high-level features of the encoder and sends them to the decoder part for feature fusion. Finally, the fourth variant (proposed DYStA-RetNet shown in Fig. 1) incorporates the attention module placed after the feature fusion layer in MPD to refine fused features of the encoder–decoder unit. From Table 1, it can be observed that the attention module placed after the feature fusion layer in the proposed DYStA-RetNet, significantly improves the segmentation performance of the model on all four benchmark datasets.

4.4. Proposed attention (MDA+SSA) module in comparison with other attention methods

The efficacy of the proposed attention module for vessel segmentation is experimentally observed against three state-of-the-art attention modules. For this study, the attention module in DYStA-RetNet is replaced by Convolutional Block Attention Module (CBAM) [33], Squeeze and Excitation (SE) module [34], and Split-Fused Attention (SFA) module [35]. From Table 2, it can be observed that the proposed attention module (MDA+SSA) in DYStA-RetNet outperforms state-of-the-art attention modules for all four datasets. It has been observed that CBAM [33], SE [34], and SFA [35] modules fail to reconstruct some fine vessels and branching patterns as their focus is on global information and they overlook local information. Also, the pooling operation used in these modules has complex computations bringing more computational cost. The proposed MDA+SSA attention block efficiently identifies fine vessels and patterns by utilizing multi-scale features and the statistical relationships among vessel pixel spaces, with minimal computational overhead.

4.5. Performance comparison of the proposed DYStA-RetNet

The performance of the proposed DYStA-RetNet is compared with state-of-the-art models and summarized in Tables 3–5. It can be observed from the tabulated results that the proposed model is lightweight and requires a significantly lesser number of training parameters and FLOPS in comparison to other state-of-the-art methods. The proposed DYStA-RetNet showcases better performance compared to the best results from recent methods for the DRIVE dataset (Table 3), with an improvement of 1.61% and 1.65% respectively in terms of F1 score and DICE parameter as compared to Jian et al. 2023 [23]. For CHASE-DB1 also (Table 3), 1.22% and 2.46% improvements in F1 score and DICE parameter, respectively, are observed as compared to

Galdran et al. 2022 [12]. Similarly, DYStA-RetNet performs better than the state-of-the-art methods on HRF and STARE datasets (Table 4). Despite low brightness and less resolution images in the STARE dataset, DYStA-RetNet can better extract retinal vessels.

Also, there are significant differences in training parameters and FLOPS. From Table 5, it can be observed that the proposed DYStA-RetNet requires the least number of trainable parameters (only 37.19K as memory footprint) and computational units (0.75 GFLOPS) compared to state-of-the-art methods. It is observed that integration of the attention block with MPD requires an additional 3K parameters and 0.28G FLOPS, but it significantly improves the performance of the base model as shown in Table 1. The quantitative and quantitative results (presented below) demonstrate the effectiveness of the proposed method for enhancing vascular segmentation performance.

5. Discussion

5.1. Qualitative evaluation

Quantitative results demonstrate the effectiveness of the proposed method for enhancing vascular segmentation performance. The proposed attention module enables the network to focus on regions prone to vessel break points and learn branching patterns and structures of such cases to avoid pixel placement errors in the decoder part. It can be observed in Fig. 2, that the recent state-of-the-art methods, WNet (Galdran et al. [12]) and WaveNet (Liu et al. [17]) failed to capture fragile thin vessels comprising a very few pixels (cases highlighted with yellow color annotation marks). However, the proposed DYStA-RetNet is able to detect these vessels efficiently. In some cases, both state-of-the-art methods suffer from false detection of vessels (highlighted with red color marks). It can be observed that such cases are much less in the proposed DYStA-RetNet. Downsampling operation causes loss of information which may be important while reconstructing vessel patterns, especially when vessel thickness is a few pixels. The state-of-the-art methods suffer from this issue and fail to detect such fine vessels. To handle this issue, the proposed DYStA-RetNet uses two parallel branches of decoder, one allowing low-level important features to flow in each decoder layer and the other branch allowing both high and low-level features to flow in decoder layers. The combination of this arrangement with the attention module, allows DYStA-RetNet to identify fine vessels and complex branching patterns.

The characteristic response and visual analysis of vessel breaking areas are conducted through the results as shown in Fig. 3. The qualitative results depict the effectiveness of the attention module over other

Table 3

Results on DRIVE and CHASE-DB1 datasets.

Methods	DRIVE			CHASE-DB 1		
	F1	AUC	DICE	F1	AUC	DICE
Laibacher et al. 2019 [6]	80.22	97.14	80.91	79.56	97.03	80.06
Wang et al. 2019 [11]	82.93	97.72	82.7	81.91	98.12	80.37
Galdran et al. 2022 [12]	82.92	98.09	82.82	84.03	98.44	81.55
Li et al. 2022 [18]	81.27	98.43	82.02	81.84	98.35	80.86
Li et al. 2022 [20]	82.88	98.66	–	82.61	98.98	–
Huang et al. 2022 [22]	82.89	98.11	82.91	83.75	98.79	81.38
Liu et al. 2023 [17]	82.54	97.63	82.13	83.49	97.02	80.49
Li et al. 2023 [21]	83.03	98.24	–	83.02	98.68	–
Jian et al. 2023 [23]	83.20	98.65	83.03	81.90	98.52	80.91
Li et al. 2024 [19]	83.04	98.30	–	83.85	98.81	–
Proposed DYStA-RetNet	**84.81**	**98.86**	**84.68**	**85.25**	**99.31**	**84.01**

Table 4

Results on HRF and STARE datasets.

| Methods	HRF			STARE		
	F1	AUC	DICE	F1	AUC	DICE
Laibacher et al. 2019 [6]	79.85	97.43	78.14	78.01	97.42	72.98
Galdran et al. 2022 [12]	80.18	98.42	81.04	80.32	98.72	78.32
Li et al. 2022 [18]	77.62	98.24	79.63	78.15	97.31	75.66
Li et al. 2022 [20]	80.97	98.45	–	83.63	98.99	–
Huang et al. 2022 [22]	80.66	98.57	81.88	85.01	98.92	81.78
Liu et al. 2023 [17]	79.55	97.16	80.01	81.40	98.62	78.29
Li et al. 2023 [21]	-	-	–	83.66	98.98	–
Jian et al. 2023 [23]	79.83	97.76	78.80	84.69	98.71	81.09
Li et al. 2024 [19]	-	-	–	83.27	98.76	–
Proposed DYStA-RetNet	**81.92**	**99.14**	**83.89**	**86.18**	**99.12**	**82.27**
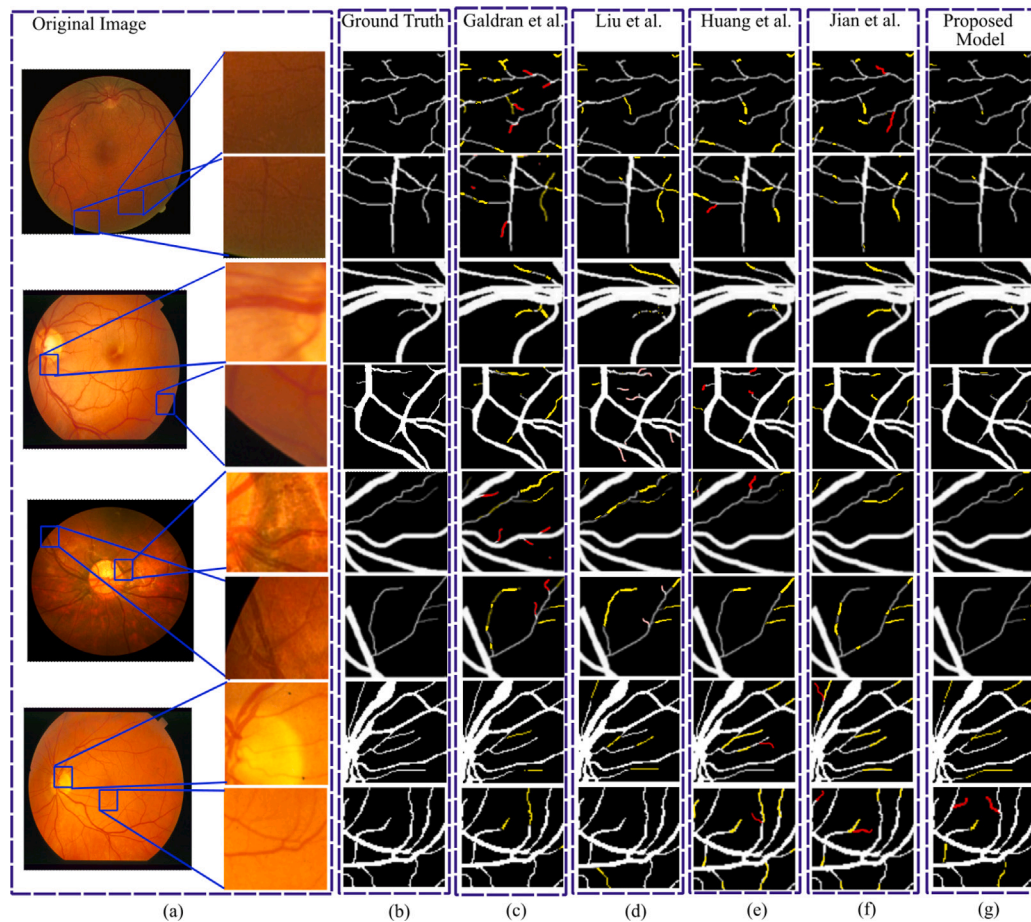
Fig. 2. Qualitative segmentation of fine retinal vessels having a few pixels (a) fundus image, (b) ground truth, (c) WNet [12] (d) segmented results WaveNet [17]. (e) Huang et al. [22] (f) Jian et al. [23]. (g) DYStA-RetNet. Yellow pixels denote false negative pixels, whereas red pixels denote false positive pixels.
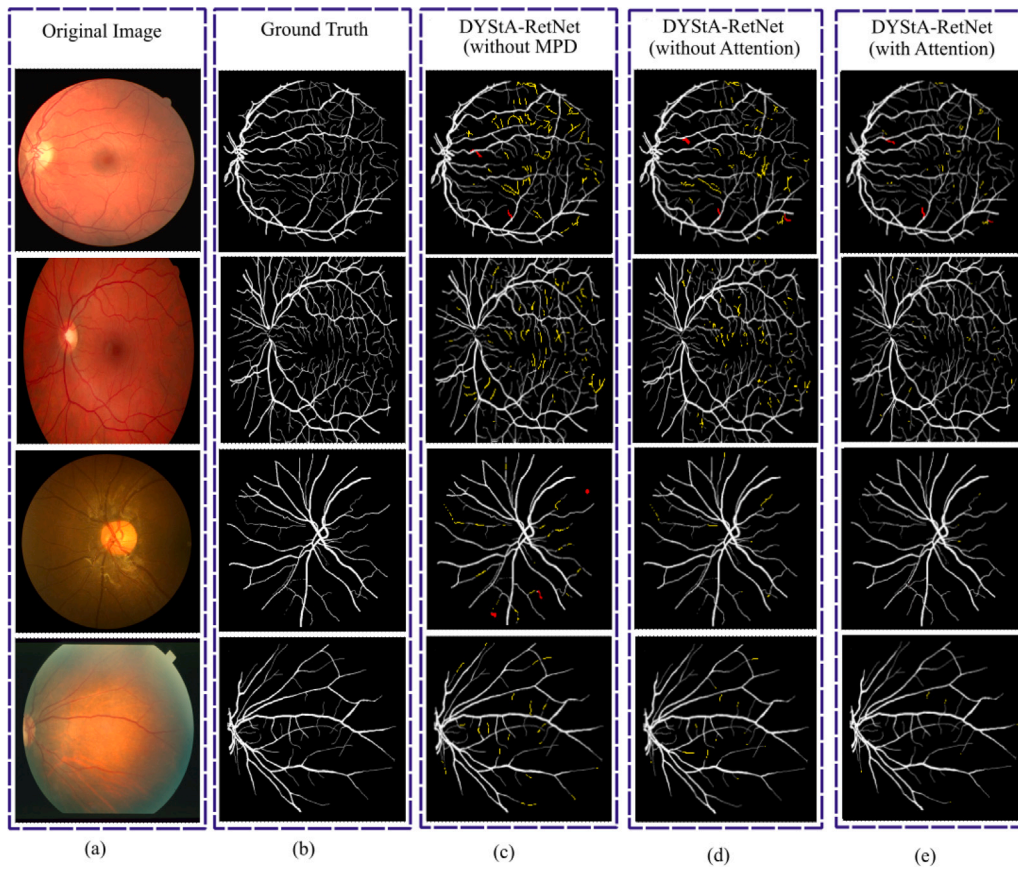
Fig. 3. Visual comparison of retinal vessel segmentation (a) original fundus image, (b) ground truth, (c) DYStA-RetNet without MPD branch, (d) DYStA-RetNet with MPD branch, but without attention module, and (e). DYStA-RetNet. Yellow pixels denote false negative pixels, whereas red pixels denote false positive pixels.

Table 5
Comparison on parameters and FLOPs count.

Method	Params	FLOPS
Laibacher et al. 2019 [6]	549 K	1.89 G
Galdran et al. 2022 [12]	68.48 K	0.92 G
Li et al. 2022 [18]	2.01 M	15.31 G
Li et al. 2022 [20]	14.11 M	295.03 G
Huang et al. 2022 [22]	42.49 M	8.49 G
Liu et al. 2023 [17]	1.50 M	42.61 G
Li et al. 2023 [21]	56.99 M	3.63 G
Jian et al. 2023 [23]	58.36 M	12.65 G
Li et al. 2024 [19]	9.66 M	4.03 G
Proposed DYStA-RetNet	**37.19 K**	**0.75 G**

variants of the proposed model. It can be observed from Fig. 3 that DYStA-RetNet with attention module can identify the complex structures of vessels more accurately, even for branching patterns where thin branch patterns have pixel values very similar to the background pixels. The yellow annotations in Fig. 3(c) and (d) are showing vessel breakpoints and missing areas that DYStA-RetNet (without MPD and attention block) could not capture. It can be observed that these vessel breakpoints and missing areas are covered with MPD and attention block, as shown in Fig. 3(e).

An experiment is performed to evaluate the performance of the proposed model on low-quality fundus images. For this purpose, 23 low-quality images are taken from 4 datasets. Models proposed in [6, 12,18] and DYStA-RetNet are experimented with such low-quality images. The average dice scores are observed as 71.86 for Laibacher et al. [6], 76.39 for Li et al. [18], 77.64 Galdran et al. [12] and 78.93 for DYStA-RetNet.

It is observed that for the cases when images suffer from low contrast, the proposed model can detect the retinal vessels efficiently

as compared to state-of-the-art methods. Out of these 23 images, the proposed DYStA-RetNet underperforms for only two images. Some example images and their analysis are shown in Table 6, including one of the two images where DYStA-RetNet could not beat [18]. It can be seen that when image is too bright and fine vessels get mixed with the background color, the proposed shallow model fails to extract such fine vessels.

5.2. Cross-dataset evaluation of DYStA-RetNet

In the real-world scenario, the medical data come from different acquisition sources. The heterogeneity in the data sources leads to weaker generalization capability of the deep learning models. The models trained on one data source may exhibit low testing performance on data from unseen sources. An exhaustive cross-dataset experimentation study is performed for the proposed DYStA-RetNet to address the issue of model generalizability. The performance of the cross-dataset experimentation is reported in Table 7 considering methods proposed by Laibacher et al. [6], Li et al. [18], and Galdran et al. [12] as they have lesser number of parameters among available state-of-the-art methods and the results of models proposed in these works are compared with the proposed DYStA-RetNet (Table 7). DYStA-RetNet trained using fundus images of HRF, STARE, and CHASE-DB1 datasets respectively and tested on all other datasets performed better as compared to recent methods. Considering a smaller number of training parameters and flops combined with superior vessel segmentation performance, DYStA-RetNet can be considered well-suited for deployment in real-life clinical applications.

Table 6
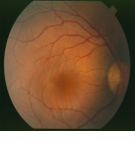Results on some challenging images with low contrast, brightness, and hemorrhage vessels.

Image	Metric	Laibacher et al. [6]	Li et al. [18]	Galdran et al. [12]	Huang et al. [22]	Jian et al. [23]	DYStA-RetNet
	F1	76.42	77.11	76.83	76.67	77.04	**77.49**
	AUC	94.25	95.15	94.64	94.17	94.7	**95.86**
	DICE	72.18	74.19	73.39	73.09	73.45	**74.31**
	F1	76.66	79.92	82.28	80.16	81.26	**82.75**
	AUC	94.03	96.58	95.64	95.18	96.31	**97.01**
	DICE	76.41	76.86	79.15	78.82	76.92	**81.69**
	F1	75.84	76.5	78.36	77.62	76.92	**79.09**
	AUC	94.29	95.68	96.12	95.82	95.1	**96.48**
	DICE	69.21	73.96	78.86	75.11	74.52	**79.11**
	F1	74.63	**76.86**	76.03	74.77	73.38	76.21
	AUC	94.62	**95.42**	94.55	94.61	93.85	94.76
	DICE	69.65	**80.58**	79.18	75.41	74.47	79.62

Table 7
Cross-dataset evaluation of DYStA-RetNet.

Method	Training	DRIVE			HRF			CHASE-DB 1			STARE		
		F1	AUC	DICE	F1	AUC	DICE	F1	AUC	DICE	F1	AUC	DICE
Laibacher et al. 2019 [6]	DRIVE	80.22	97.14	80.91	71.26	93.77	69.92	71.41	97.83	75.14	78.09	96.71	75.81
Li et al. 2022 [18]		81.27	98.43	82.02	71.79	94.42	70.52	72.85	98.35	76.78	78.11	97.19	76.16
Galdran et al. 2022 [12]		82.92	98.09	82.82	72.31	95.9	70.39	76.39	97.22	75.13	81.56	98.15	81.46
Huang et al. 2022 [22]		82.89	98.11	82.91	74.62	96.36	72.04	80.72	95.56	74.41	80.91	95.70	79.89
Jian et al. 2023 [23]		83.20	98.65	83.03	72.19	95.92	71.02	76.85	93.22	70.18	79.7	95.84	77.37
DYStA-RetNet		**84.81**	**98.86**	**84.68**	**77.42**	**97.57**	**76.28**	**81.42**	**98.19**	**77.48**	**81.93**	**98.86**	**82.08**
Laibacher et al. 2019 [6]	HRF	70.16	93.04	61.66	79.85	97.43	78.14	73.63	93.87	59.12	70.03	92.75	62.94
Li et al. 2022 [18]		70.49	93.77	63.98	77.62	98.24	79.63	73.82	94.28	59.4	70.51	92.04	62.47
Galdran et al. 2022 [12]		71.1	94.17	64.79	80.18	98.42	81.04	74.52	94.94	60.63	71.43	93.15	64.86
Huang et al. 2022 [22]		70.34	92.65	62.86	80.66	98.57	81.88	73.02	94.18	58.83	72.25	93.64	65.06
Jian et al. 2023 [23]		68.63	90.21	59.77	79.83	97.76	78.80	71.15	93.68	54.17	68.9	90.84	61.97
DYStA-RetNet		**72.72**	**95.36**	**66.46**	**81.92**	**99.14**	**83.89**	**78.41**	**95.04**	**63.52**	**72.55**	**95.18**	**66.84**
Laibacher et al. 2019 [6]	CHASE-DB 1	75.3	92.04	75.01	78.06	94.18	78.15	79.56	97.03	80.06	75.76	97.12	72.87
Li et al. 2022 [18]		77.72	92.93	75.14	78.99	95.62	78.18	81.84	98.35	80.86	76.37	97.27	73.58
Galdran et al. 2022 [12]		78.17	93.38	77.4	80.59	97.77	79.74	84.03	98.44	81.55	76.85	97.97	72.14
Huang et al. 2022 [22]		81.98	96.38	78.4	81.29	98.07	78.08	83.75	98.79	81.38	81.73	96.92	71.84
Jian et al. 2023 [23]		78.13	95.27	72.51	78.26	95.83	74.08	81.90	98.52	80.91	78.16	92.07	68.33
DYStA-RetNet		**82.03**	**96.59**	**78.74**	**82.75**	**98.66**	**81.64**	**85.25**	**99.31**	**84.01**	**81.58**	**98.16**	**75.48**
Laibacher et al. 2019 [6]	STARE	78.02	94.26	69.83	75.87	93.22	68.66	78.43	94.22	70.1	78.01	93.42	68.98
Li et al. 2022 [18]		78.82	94.79	70.25	76.24	96.54	67.07	79.52	95.12	71.34	79.87	94.31	70.66
Galdran et al. 2022 [12]		79.32	96.39	72.09	77.43	97.03	69.23	80.83	97.12	72.31	80.32	96.72	72.32
Huang et al. 2022 [22]		81.21	97.01	72.64	77.97	97.67	70.12	80.17	96.91	71.45	85.01	98.92	81.78
Jian et al. 2023 [23]		81.23	97.25	72.69	75.19	97.04	67.68	77.37	94.27	69.83	84.69	98.71	81.09
DYStA-RetNet		**81.53**	**97.42**	**73.13**	**79.61**	**98.12**	**72.83**	**82.46**	**98.56**	**77.65**	**86.18**	**99.12**	**82.27**

5.3. Generalizability of DYStA-RetNet

In order to evaluate generalizability of the proposed DYStA-RetNet, experiments were conducted on two distinct datasets: IDRiD [36] (same domain dataset) and KVASIR [37] (different domain dataset). IDRiD targets the extraction of soft-exudate (SE) lesions from fundus images for diabetic retinopathy, while KVASIR focuses on polyp segmentation in wireless capsule endoscopy (WCE). The diversity of these datasets, IDRiD addressing retinal imaging and KVASIR emphasizing gastrointestinal imaging, provides a robust framework to validate the model's adaptability to different medical imaging tasks and modalities.

For the KVASIR dataset, an 80:20 train–test split was employed to simulate real-world performance, while for the IDRID dataset, the original training and testing splits were used to ensure consistency with prior research. Comparative evaluations are performed against state-of-the-art methods for which code was available. As shown in Table 8, the quantitative analysis revealed that the proposed DYStA-RetNet performs better than state-of-the-art methods in important measures including F1-score, AUC and DICE score, demonstrating its superiority in segmenting complex patterns. These results highlight the model's ability to generalize across datasets with varying challenges, including variations in texture, scale, and boundary irregularities.

Also, qualitative assessment of DYStA-RetNet performed on both the datasets demonstrated better performance in detecting complex structures while maintaining significantly low false positives compared

Table 8
Model evaluation on different disease datasets.

Methods	IDRID (SE) [36][a]			KVASIR [37][b]		
	F1	AUC	DICE	F1	AUC	DICE
Laibacher et al. 2019 [6]	75.85	90.17	72.94	88.17	97.21	85.76
Galdran et al. 2022 [12]	80.64	95.86	77.64	91.63	98.92	89.15
Li et al. 2022 [18]	76.31	94.47	74.35	89.04	98.02	86.15
Huang et al. 2022 [22]	77.49	94.73	76.85	88.45	97.58	86.75
Jian et al. 2023 [23]	78.16	95.48	77.18	90.32	98.65	88.54
DYStA-RetNet	**82.15**	**97.03**	**79.42**	**92.52**	**99.17**	**91.11**

[a] Same domain dataset: different disease - Soft Exudates segmentation.
[b] Different domain dataset: WCE images for polyp segmentation.
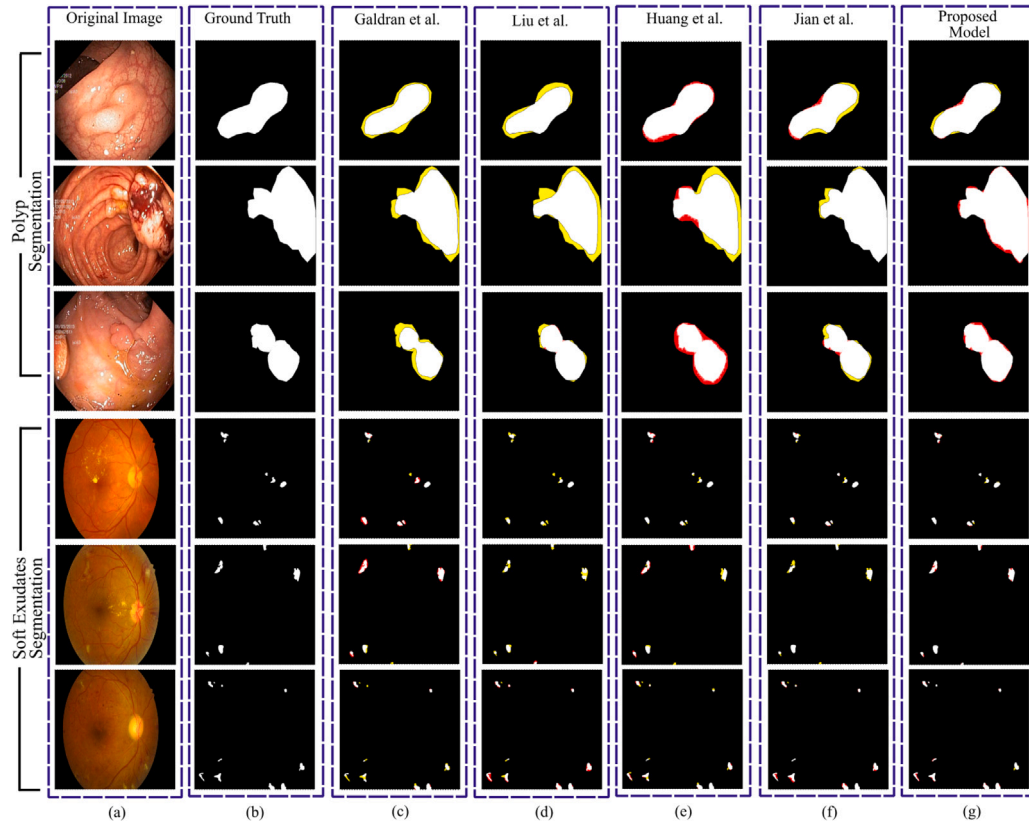
Fig. 4. Visual comparison of polyp and diabetic retinopathy lesion (soft-exudates) segmentation results (a) original fundus image, (b) ground truth, (c) Galdran et al. [12] (d) Liu et al. [17]. (e) Huang et al. [22] (f) Jian et al. [23]. (g) proposed DYStA-RetNet. Yellow pixels denote false negative pixels, whereas red pixels denote false positive pixels.

to state-of-the-art methods. As shown in Fig. 4, the proposed DYStA-RetNet effectively segmented polyps even in the presence of unclear or blended boundaries in the input image of KVASIR polyp segmentation dataset. This showcases its robustness in handling low-contrast regions and challenging gastrointestinal patterns. Similarly, for the IDRID-SE segmentation dataset, DYStA-RetNet accurately detected soft-exudate lesions, maintaining continuity and details in the presence of noise and complex retinal structures. These outcomes highlight the capacity of the proposed DYStA-RetNet to respond to a variety of issues specific to medical imaging, including differences in data quality, imaging modalities, and disease contexts. The performance of the proposed model on two completely distinct datasets highlights both, its potential utility across a wide range of diagnostic applications and its generalizability.

6. Conclusion

Retinal vessel segmentation is the primary step for early detection of retinal diseases. It is found to be challenging due to the presence of fragile thin vessels and complex breakpoints. This work presents

lightweight CNN-based DYStA-RetNet architecture for efficient retinal vessel segmentation. The key idea is to utilize a self-attention module in the proposed model that exploits spatial and channel information for efficient vessel pattern extraction. As a result, the proposed DYStA-RetNet effectively learns vessel breakpoints and branching patterns. Experimentation shows that the proposed model can efficiently detect thicker as well as micro-level vessels. The method outperforms state-of-the-art methods for all four benchmark datasets. Future work will focus on reducing false positives of micro-vessels.

CRediT authorship contribution statement

Amit Bhati: Writing – original draft, Methodology. **Samir Jain:** Formal analysis. **Neha Gour:** Formal analysis, Data curation. **Pritee Khanna:** Writing – review & editing, Formal analysis, Conceptualization. **Aparajita Ojha:** Validation, Formal analysis, Conceptualization. **Naoufel Werghi:** Validation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] C.L. Srinidhi, P. Aparna, J. Rajan, Recent advancements in retinal vessel segmentation, J. Med. Syst. 41 (2017) 1–22.

[2] R. Poplin, A. Varadarajan, K. Blumer, Y. Liu, M. Mcconnell, G. Corrado, L. Peng, D. Webster, Predicting cardiovascular risk factors from retinal fundus photographs using deep learning. arxiv 2017, 2017, arXiv preprint arXiv:1708.09843.

[3] K. Sun, Y. Chen, Y. Chao, J. Geng, Y. Chen, A retinal vessel segmentation method based improved U-net model, Biomed. Signal Process. Control 82 (2023) 104574.

[4] Q. Qin, Y. Chen, A review of retinal vessel segmentation for fundus image analysis, Eng. Appl. Artif. Intell. 128 (2024) 107454.

[5] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[6] T. Laibacher, T. Weyde, S. Jalali, M2u-net: Effective and efficient retinal vessel segmentation for real-world applications, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.

[7] Z. Zhuo, J. Huang, K. Lu, D. Pan, S. Feng, A size-invariant convolutional network with dense connectivity applied to retinal vessel segmentation measured by a unique index, Comput. Methods Programs Biomed. 196 (2020) 105508.

[8] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer, 2015, pp. 234–241.

[9] Z. Yan, X. Yang, K.-T. Cheng, Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation, IEEE Trans. Biomed. Eng. 65 (9) (2018) 1912–1923.

[10] H. Zhao, H. Li, S. Maurer-Stroh, Y. Guo, Q. Deng, L. Cheng, Supervised segmentation of un-annotated retinal fundus images by synthesis, IEEE Trans. Med. Imaging 38 (1) (2018) 46–56.

[11] X. Wang, X. Jiang, J. Ren, Blood vessel segmentation from fundus image by a cascade classification framework, Pattern Recognit. 88 (2019) 331–341.

[12] A. Galdran, A. Anjos, J. Dolz, H. Chakor, H. Lombaert, I.B. Ayed, State-of-the-art retinal vessel segmentation with minimalistic models, Sci. Rep. 12 (1) (2022) 6174.

[13] A. Asad, A.T. Azar, N. El-Bendary, A.E. Hassaanien, et al., Ant colony based feature selection heuristics for retinal vessel segmentation, 2014, arXiv preprint arXiv:1403.1735.

[14] H.Y. Henry, X. Feng, Z. Wang, H. Sun, MixModule: Mixed CNN kernel module for medical image segmentation, in: 2020 IEEE 17th International Symposium on Biomedical Imaging, ISBI, IEEE, 2020, pp. 1508–1512.

[15] L. Mou, Y. Zhao, L. Chen, J. Cheng, Z. Gu, H. Hao, H. Qi, Y. Zheng, A. Frangi, J. Liu, CS-net: channel and spatial attention network for curvilinear structure segmentation, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22, 2019, pp. 721–730.

[16] H. Zhang, X. Zhong, G. Li, W. Liu, J. Liu, D. Ji, X. Li, J. Wu, BCU-Net: Bridging ConvNeXt and U-net for medical image segmentation, Comput. Biol. Med. 159 (2023) 106960.

[17] Y. Liu, J. Shen, L. Yang, H. Yu, G. Bian, Wave-net: A lightweight deep network for retinal vessel segmentation from fundus images, Comput. Biol. Med. 152 (2023) 106341.

[18] M. Li, S. Zhou, C. Chen, Y. Zhang, D. Liu, Z. Xiong, Retinal vessel segmentation with pixel-wise adaptive filters, in: 2022 IEEE 19th International Symposium on Biomedical Imaging, ISBI, IEEE, 2022, pp. 1–5.

[19] J. Li, G. Gao, L. Yang, Y. Liu, A retinal vessel segmentation network with multiple-dimension attention and adaptive feature fusion, Comput. Biol. Med. 172 (2024) 108315.

[20] Y. Li, Y. Zhang, W. Cui, B. Lei, X. Kuang, T. Zhang, Dual encoder-based dynamic-channel graph convolutional network with edge enhancement for retinal vessel segmentation, IEEE Trans. Med. Imaging 41 (8) (2022) 1975–1989.

[21] J. Li, G. Gao, L. Yang, G. Bian, Y. Liu, DPF-Net: A dual-path progressive fusion network for retinal vessel segmentation, IEEE Trans. Instrum. Meas. (2023).

[22] J. Huang, Z. Lin, Y. Chen, X. Zhang, W. Zhao, J. Zhang, Y. Li, X. He, M. Zhan, L. Lu, et al., DBFU-net: Double branch fusion U-net with hard example weighting train strategy to segment retinal vessel, PeerJ Comput. Sci. 8 (2022) e871.

[23] M. Jian, R. Wu, L. Fu, C. Yang, et al., Dual-branch-unet: A dual-branch convolutional neural network for medical image segmentation, CMES - Comput. Model. Eng. Sci. 137 (1) (2023) 705–716.

[24] J. Lin, X. Huang, H. Zhou, Y. Wang, Q. Zhang, Stimulus-guided adaptive transformer network for retinal blood vessel segmentation in fundus images, Med. Image Anal. 89 (2023) 102929.

[25] A. Budai, R. Bock, A. Maier, J. Hornegger, G. Michelson, Robust vessel segmentation in fundus images, Int. J. Biomed. Imaging 2013 (2013).

[26] A. Hoover, V. Kouznetsova, M. Goldbaum, Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response, IEEE Trans. Med. Imaging 19 (3) (2000) 203–210.

[27] F.N. Iandola, S. Han, M.W. Moskewicz, K. Ashraf, W.J. Dally, K. Keutzer, SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5mb model size, 2016, arXiv:1602.07360.

[28] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, MobileNets: Efficient convolutional neural networks for mobile vision applications, 2017, arXiv:1704.04861.

[29] S. Ioffe, Batch renormalization: Towards reducing minibatch dependence in batch-normalized models, in: I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), in: Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc., 2017.

[30] X. Lian, Revisit batch normalization: New understanding and refinement via composition optimization., in: I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), The 22nd International Conference on Artificial Intelligence and Statistics, PMLR, 2019.

[31] Z. Wu, L. Su, Q. Huang, Cascaded partial decoder for fast and accurate salient object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2019.

[32] Y. Chen, X. Dai, M. Liu, D. Chen, L. Yuan, Z. Liu, Dynamic convolution: Attention over convolution kernels, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11030–11039.

[33] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, Cbam: Convolutional block attention module, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 3–19.

[34] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.

[35] A. Bhati, S. Jain, N. Gour, P. Khanna, A. Ojha, N. Werghi, A shallow U-net with split-fused attention mechanism for retinal vessel segmentation, in: 2023 IEEE International Conference on Image Processing, ICIP, IEEE, 2023, pp. 3205–3209.

[36] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, F. Meriaudeau, Indian diabetic retinopathy image dataset (idrid): a database for diabetic retinopathy screening research, Data 3 (3) (2018) 25.

[37] D. Jha, P.H. Smedsrud, M.A. Riegler, P. Halvorsen, T. De Lange, D. Johansen, H.D. Johansen, Kvasir-seg: A segmented polyp dataset, in: MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, Proceedings, Part II 26, Springer, 2020, pp. 451–462.