DUF-Net: A Retinal Vessel Segmentation Method Integrating Global and Local Features with Skeleton Fitting Assistance

Xuelin Xu[1]*, Ren Lin[2], Jianwei Chen[3], Huabin He[4]

School of Computer Science and Mathematics, Fujian University of Technology, Fujian, China[1, 2, 3]
School of Electronic, Electrical Engineering and Physics, Fujian University of Technology, Fujian, China[4]
Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fujian University of Technology, Fujian, China[1, 2, 3]

Abstract—**Assisted evaluation through retinal vessel segmentation facilitates the early prevention and diagnosis of retinal lesions. To address the scarcity of medical samples, current research commonly employs image patching techniques to augment the training dataset. However, the vascular features in fundus images exhibit complex distribution, patch-based methods frequently encounter the challenge of isolated patches lacking contextual information, consequently resulting in issues such as vessel discontinuity and loss. Additionally, there are a higher number of samples with strong contrast vessels compared to those with weak contrast vessels in retinal images. Moreover, within individual patches, there are more pixels of strong contrast vessels compared to weak contrast vessels, leading to lower segmentation accuracy for small vessels. Hence, this study introduces a patch-based deep neural network method for retinal vessel segmentation to address the issues. Firstly, a novel architecture, termed Double U-Net with a Feature Fusion Module (DUF-Net), is proposed. This network structure effectively supplements missing contextual information and improves the problem of vessel discontinuity. Furthermore, an algorithm is introduced to classify vascular patches based on their contrast levels. Subsequently, conventional data augmentation methods were employed to achieve a balance in the number of samples with strong and weak contrast vessels. Additionally, method with skeleton fitting assistance is introduced to improve the segmentation of vessels with weak contrast. Finally, the proposed method is evaluated across four publicly available datasets: DRIVE, CHASE_DB1, STARE, and HRF. The results demonstrate that the proposed method effectively ensures the continuity of segmented blood vessels while maintaining accuracy.**

Keywords—Fundus image; vessel segmentation; skeleton fitting; data augmentation; patch classification

I. INTRODUCTION

The eye stands as one of the paramount sensory organs in the human body, serving as the primary conduit for external stimuli reception. Presently, propelled by societal advancements and the proliferation of electronic devices, there is a sustained escalation in the prevalence of ophthalmic diseases among patients. Today, with the development of society and the widespread use of electronic devices, the number of patients with ophthalmic diseases continues to increase. Ophthalmic diseases are closely associated with retinal lesions, with glaucoma, diabetic retinopathy, and diabetic macular degeneration being major causes of global blindness [1]. Retinal fundus images represent the singular non-invasive modality for observing the deep microvascular system, encompassing a diverse array of retinal structures such as the retinal vascular tree, optic disc, fovea centralis, and macula [2]. The early clinical characteristics of diabetic retinopathy encompass microaneurysms, dot and blot hemorrhages, cotton wool spots, and intraretinal microvascular abnormalities [3]. The characteristic sign of glaucoma is optic disc cupping, and the ratio of the cup to disc area in fundus images, referred to as the cup-to-disc ratio (CDR), serves as a vital structural metric for evaluating the presence and advancement of glaucoma [4]. The hallmark of age-related macular degeneration (AMD) is the infiltration of choroidal vasculature into the macular region, accompanied by heightened vascular permeability [4]. Variations in retinal structure are pivotal for diagnosing diabetic retinopathy, glaucoma, and AMD. Given that retinal vessel segmentation is essential for visualizing and quantifying retinal pathology, it is an indispensable component in the analysis of retinal diseases [5]. The conventional method for retinal vessel segmentation is characterized by its costly, intricate, and time-intensive nature. Furthermore, challenges such as uneven illumination, complex vascular structures, and low vessel-background contrast in images contribute to inconsistencies in vessel segmentation across different experts [6]. This has spurred the advancement of automated retinal vessel segmentation technology.

Traditional segmentation methods mainly include line detection [7], edge detection [8], matched filtering [9] [10], and shape-based methods [11] [12]. These methods perform vessel segmentation based on vessel edge pixels or shape features. The main reason is that, compared to the background, the edge and shape information of vessels with strong contrast are more prominent, making their features easier to learn. In recent years, deep learning methods have propelled the progress of retinal vessel segmentation techniques, surpassing traditional machine learning methods [13]. Deep learning methods do not require manual feature design and can effectively extract key features from data while achieving good accuracy and generalization capabilities, thus promoting the development of retinal vessel segmentation methods [14] [15]. Currently, there are generally two types of retinal vessel segmentation methods based on deep learning [16]. End-to-end methods are one of the types, sacrificing spatial resolution to ease memory constraints during

*Corresponding Author.

training. However, these methods lead to in the loss of spatial information in retinal vessel images. On the other hand, patch-based segmentation methods are another type. Although these methods enhance data samples, it may lead to a lack of contextual information between independent patches. Through observation, we notice that there are more samples of vessel patches with strong contrast compared to those with weak contrast, and within individual patches, there are more pixels of vessels with strong contrast compared to those with weak contrast. These issues may result in problems such as vessel discontinuity and poor segmentation of small vessels when restoring segmented blocks to the original image size.

Therefore, in this paper, we first compare the pixel values of vessels and background within vessel patches. If a patch contains a higher proportion of pixels with strong vessel contrast, it is classified as a Contrast Strong Vessel Patch (CSVP); otherwise, it is classified as a Contrast Weak Vessel Patch (CWVP). Next, we introduce a novel network structure named DUF-Net, which is capable of learning both global and local information, effectively supplementing missing contextual information between patches. Additionally, we design a patch classification algorithm to perform patch classification and quantity statistics. Data augmentation is utilized to augment deficient image patches, aiming to balance the number of samples for various types of retinal vessel pathology, including CSVP and CWVP. Additionally, a training method integrating skeleton fitting assistance is introduced to enhance the model's segmentation capability for CWVP within individual samples. In summary, the principal contributions of this paper can be outlined as follows:

- In order to compensate for the absence of contextual information between patches, We introduce a novel network structure named DUF-Net. It is capable of simultaneously learning both global and local features, guiding the model to capture contextual information surrounding the patches and thereby improving the completeness of vessel segmentation.

- To improve the segmentation capability of CWVP, a patch classification algorithm was designed to balance the quantity of CSVP and CWVP samples. Additionally, during training, prior knowledge of vessel skeletons and corresponding loss functions were introduced to guide the model in learning CWVP features and address the issue of pixel imbalance within individual patches.

- We assessed the proposed method across four publicly available datasets and conducted comparisons with six latest methods. Experimental findings affirm the efficacy and robustness of the proposed approach.

The remaining sections of this paper are organized as follows. Section II provides a review of traditional and deep learning methods in retinal vessel segmentation. Section III provides a detailed explanation of the proposed method. Section IV showcases the implementation details of the experiments. Section V presents the experimental results. Section VI and VII respectively discuss and summarize the proposed method.

II. RELATED WORK

Research on retinal vessel segmentation can be categorized into traditional methods and deep learning methods.

A. Traditional Methods

Traditional methods involve direct detection of features or edge pixels in retinal vessel images. Sheng et al. [17] combined geometric structures, texture, color, and spatial information in the image while simultaneously refining the segmentation results using a minimum spanning superpixel tree to refine segmentation results. Lam et al. [18] proposed a novel multi-concave modeling approach for handling bright lesions in the perceptual space and removing dark lesions that differ from the retinal vascular structure. To improve vessel segmentation efficiency, Rezaee and Haddadnia [19]employed a skeletonization and fuzzy entropy thresholding segmentation algorithm. They differentiated retinal main vessels from other tissue components through adaptive filtering and fuzzy entropy thresholding. In summary, although the aforementioned methods do not require training and have lower complexity, they heavily rely on filter design and often exhibit lower accuracy.

In machine learning, manual feature extraction is utilized for retinal fundus image analysis, followed by classification using common classifiers such as k-nearest neighbors (KNN) [20], support vector machine (SVM) [21], and Bayesian methods [22] [23]. Orlando et al. [24] utilized conditional random fields and support vector machines for retinal fundus image vessel segmentation. Zhu et al. [25] presented a supervised approach employing Extreme Learning Machine (ELM). This method involved constructing matrices by extracting features from each pixel of each retinal image and the manually labeled pixels, which were then input into ELM. These approaches heavily depend on prior knowledge or necessitate a series of complex operations for extracting discriminative features, thereby lacking generalization ability [26].

The segmentation performance of traditional methods needs improvement, especially when facing environments with retinal lesion features and low brightness. Traditional methods lack generalization capabilities in such scenarios. As a result, deep neural networks have found extensive applications in the field of retinal imaging.

B. Deep Learning Methods

Deep learning approaches equipped with automatic feature recognition capabilities exhibit superior performance in the field of retinal vessel segmentation compared to traditional methods. Overall, these methods can be categorized into two main types [16]. An end-to-end training approach is one of the types, because of its simple and stable performance has attracted the attention of many researchers. Hu et al. [27] incorporate a saliency mechanism to leverage features from one block as attention cues for the features of the subsequent block, effectively mitigating the problem of data imbalance. Moreover, o enhance the integrity and continuity of vessels after segmentation, Mou et al. [28] utilize a dense dilated model to integrate the newly proposed dense dilated feature extraction blocks, with the goal of extracting and accumulating

features across various scales. For improving the segmentation of fine vessels, Mishra et al. [29] employ the mean retinal vessel width and align it with the effective receptive field to identify the location of auxiliary supervision, thereby directing the model's attention towards fine vessels. The consecutive downsampling leads to the loss of a considerable amount of spatial feature information. To mitigate this challenge, Wang et al. [30] introduced two paths separately in the encoder and decoder, enhancing the model's capability for detailed representation and discrimination.

Patch-based segmentation methods are another type. In recent years, patch-based segmentation methods, which reduce computational pressure while minimizing the impact on vessel morphology, have been extensively investigated by many scholars. Dasgupta et al. [31] extracted 28×28 patches from preprocessed images and fed them into a CNN network for vessel segmentation. To ameliorate the problem of pixel imbalance between contrast strong vessels and contrast weak vessels in retinal images, Yan et al. [32] cropped retinal images into small 128×128 patches and proposed a loss function to balance contrast strong vessel and contrast weak vessel, enabling the model to learn more effective vessel segmentation features. Wu et al. [33] link two identical multi-scale backbone networks to facilitate the direct propagation of useful multi-scale features from shallow layers to deep layers. Wang et al. [34] designed three decoder networks: the first network dynamically localized image segmentation difficulty areas, while the other two networks are utilized for segmenting difficult and easy regions, respectively. Meanwhile, attention mechanisms were introduced in the network to enhance the focus on features of difficult areas in images. Yang et al. [35] designed an effective loss function to accommodate the two distinct vessel segmentation tasks, thereby improving the imbalance issue between thick and thin vessel segmentation ratios. Tan et al. [36] addressed the issue of having more pixels of strong contrast vessels than weak contrast vessels within individual patches by introducing skeleton priors and contrast losses. However, this method utilized maximum pooling on all extracted features before introducing skeleton-assisted vessel segmentation, leading to some contrast weak vessel missing and insufficient guidance for the model to learn contrast weak vessel. Therefore, we design a retinal vessel segmentation method that incorporates skeleton fitting assistance and fusion of global and local features to address vessel discontinuity and improve contrast weak vessel segmentation effectiveness.

III. METHODOLOGY

Firstly, we preprocess the original images by cropping them into several small patches. We balance the number of different class image patches through sample correction, and then perform image enhancement on CWVP. Secondly, the preprocessed patch samples are fed into the proposed Double U-Net with a Feature Fusion Module Networks (DUF-Net) to obtain corresponding segmentation maps. The DUF-Net architecture consists of green and blue Base-Nets, along with a Feature Fusion Module (FFM) designed to learn and fuse

global and local information from the patches. Additionally, the feature maps outputted by the two Base-Net models are fed into the Skeleton Fitting Assistance (SFA) section. Finally, the segmented patches are restored to the original image size. The specific workflow is illustrated in Fig. 1.
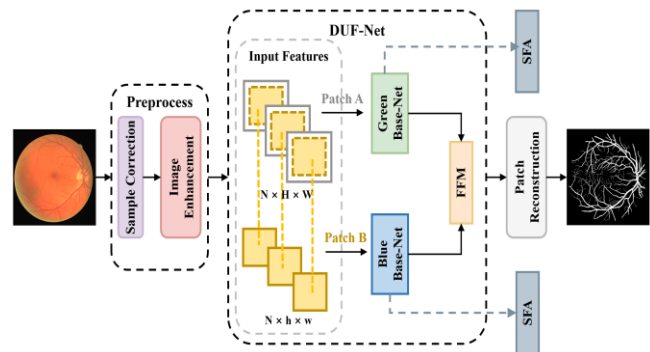

Fig. 1. Workflow of the proposed a retinal vessel segmentation method integrating global and local features with skeleton fitting assistance.

A. Preprocessing

Due to the limited quantity of existing retinal vessel datasets and the difficulty for models to learn contrast weak vessel features, we propose a preprocessing method consisting of two main stages: sample correction and image enhancement, as illustrated in Fig. 2.
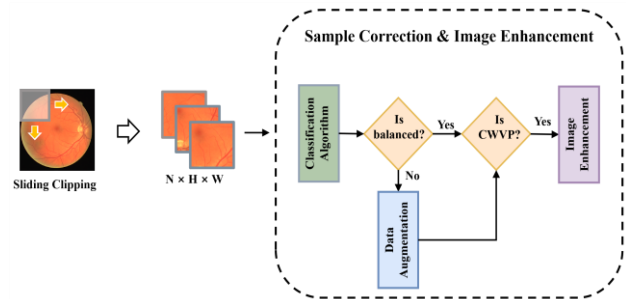

Fig. 2. The preprocessing steps.

1) Sample correction: This paper employs a patch-based approach for image segmentation. Firstly, patches are extracted from the original image using a sliding window approach to augment the model's training samples. Within these samples, patches are categorized into CSVP and CWVP. Typically, CSVP samples outnumber CWVP samples, resulting in suboptimal segmentation performance for CWVP samples. To address the issue of imbalanced CSVP and CWVP sample quantities, this paper designs a patch classification algorithm, as shown in Algorithm 1. This algorithm aims to achieve a balanced state with a nearly 1:1 ratio of samples for different classes in the training set. Data augmentation techniques are applied to CWVP; including random flipping, rotation, and Gaussian noise addition, to bring its quantity closer to that of CSVP.

Algorithm 1: patch classification Algorithm

Require: The Ground Truth *Gt* of the patch.

Ensure: The class of the patch.

1:      Set *Gt_temp = Gt* / 255.
2:      Set *first_eroded* = the erosion of *Gt_temp*.
3:      Set *dilated* = the dilation of *first_eroded*.
4:      Set T1 = Gt_temp - dilated_image.
5:      Set *second_eroded* = the erosion of *first_eroded*.
6:      Set T2 = first_eroded - second_eroded.
7:      Calculate the sum of pixels with a value of 1 in images *T1* and *T2* to obtain *S1* and *S2*, respectively.
8:      **If** *S1 > S2*
9:              *c* = CWVP
10:    else
11:            *c* = CSVP
12:    Output the class *c* of the patch

---

2) The details of Algorithm 1 are described as follows:

a) Step 1: Perform the first erosion operation on the ground truth, followed by a dilation operation, to obtain an image representing contrast strong vessel pixels.

b) Step 2: The ground truth is subtracted from the result of dilation in Step 1 to obtain the image T1, which represents contrast weak vessel pixels that are deleted after the first erosion.

c) Step 3: Perform second erosion on the result obtained after the first erosion in Step 1.

d) Step 4: Subtract the result obtained after the second erosion in Step 3 from the result obtained after the first erosion in Step 1 to obtain the image T2, representing contrast weak vessel pixels that are deleted after the second erosion.

e) Step 5: Calculate the pixel sum for both T1 and T2 images to obtain S1 and S2, respectively. If S1 is greater than S2, the patch is classified as CWVP; otherwise, it is classified as CSVP.

3) Image enhancement: This paper employs a sliding window cropping approach to increase training samples; however, the background of fundus images exhibits relatively uniform brightness and weak lighting compared to blood vessels. In scenarios with low brightness and weak lighting, the model lacks generalization ability and struggles to achieve satisfactory segmentation results in complex environments. Moreover, In CWVP, the vessels exhibit lower contrast with the background, leading to increased learning difficulty and the introduction of noise during the segmentation process, affecting accuracy. The basic idea is to enhance the contrast of randomly selected CWVP samples, considering that different contrasts typically highlight different details.The main concept of this method is outlined in Eq. (1). Specifically, a probability factor p is randomly generated from the range [0, 1]. When p falls within the range [0, 0.5] and the patch corresponds to CWVP, contrast enhancement is applied. Otherwise, the image remains unaltered, without contrast enhancement.

$$X_{new} = \begin{cases} \partial \cdot T(X), & X \in CWVP \ and \ p \in [0, 0.5] \\ X, & otherwise \end{cases} \quad (1)$$

where, *X* represents the original fundus image, X_{new} denotes the obtained new image sample, *T* signifies contrast enhancement applied to the image, and ∂ represents a randomly generated contrast factor within the range [0.5, 1.5].

B. DUF-Net

The architecture of the DUF-Net is depicted in Fig. 3, consisting of green and blue Base-Nets and a Feature Fusion Module (FFM). Patch A corresponds to the output obtained in Fig. 2, with lengths and widths denoted as H and W, respectively. Due to the patch segmentation approach, there is a loss of contextual information around the patch, leading to the issue of vessel discontinuity after segmentation. Therefore, while feeding Patch A into the model, it is simultaneously center-cropped to obtain Patch B, with lengths and widths represented as h and w, respectively. The final segmentation map outputted by the model has the same dimensions as Patch B. Patch A is inputted into the green Base-Net for learning global features, while Patch B is inputted into the blue Base-Net to capture local features. Subsequently, the Feature Fusion Module (FFM) is employed to integrate global and local features, completing the supplementation of missing contextual information to prevent vessel discontinuity issues after segmentation.
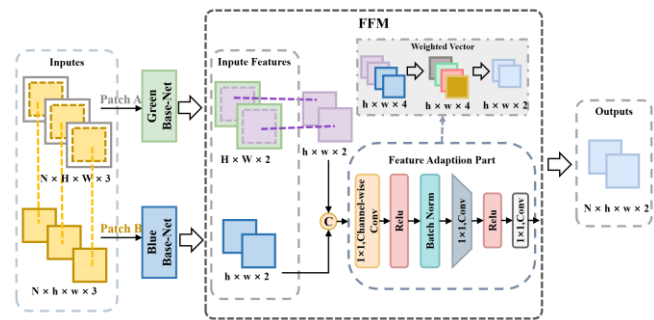

Fig. 3.   DUF-Net architecture.

In Fig. 3, the Base-Net utilizes the U-Net model, which typically employs a downsampling structure to achieve a larger receptive field for capturing more semantic information. It utilizes upsampling to restore the original image, and spatial information recovery is achieved through skip connections, resulting in improved accuracy [37]. In order to reduce model parameters and minimize the loss of spatial structural information during the model downsampling process, an enhanced U-Net is designed, as illustrated in Fig. 4. The number of downsampling steps is reduced from four to three, and the channel count is halved from [64, 128, 256, 512] to [32, 64, 128, 256].The DUF-Net model employs the Base-Net as a feature generator for producing global and local information features. Subsequently, Patch A and Patch B are separately input into the green and blue Base-Nets, resulting in corresponding features *Fa* and *Fb*. Fa represents features obtained from Patch A, incorporating contextual information around the patch, while *Fb* represents local information features derived from Patch B.
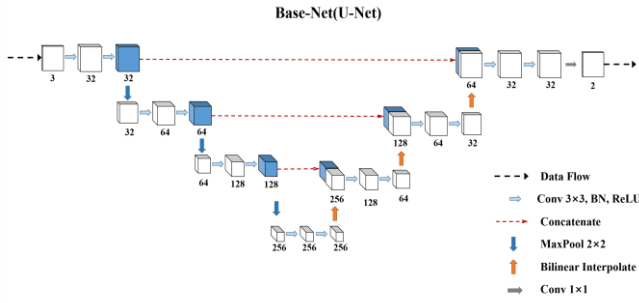
Fig. 4.   Base-Net architecture.

Furthermore, we introduce a Feature Fusion Module (FFM) designed to merge feature information of varying scales generated by different Base-Net. As depicted in Fig. 3, this module is primarily composed of three components: the dimension adaptation stage, the feature concatenation stage, and the feature adaptation stage. *Fa* and *Fb* represent the features generated from the Base-Net model. Due to the disparate patch sizes of Patch A and Patch B, the corresponding feature sizes differ. To facilitate effective fusion, *Fa*'s feature is centrally cropped, resizing it from H $\times$ W to h$\times$w, yielding the feature *Fac*. This cropping ensures consistency in size with *Fb* while retaining contextual features around the patch. The next step involves concatenating the two input features. As outputs from different-sized input images yield distinct features, a simple concatenation might not effectively integrate global and local features, potentially leading to the loss of edge information during learning. Thus, for effective feature fusion, the feature adaptation stage employs a 1$\times$1 channel-wise convolution to adjust the weights of the concatenated features. This convolution can be considered a weight vector that highlights important channel information. Subsequently, two 1$\times$1 convolutions are used to learn the adjusted features' weights, reducing the channel count by half. Y_{out} denotes the output of the fusion module and can be computed as given in Eq. (2):

$$Y_{out} = \gamma(\gamma([Fa^c, Fb] \otimes k_1) * k_2) * k_3 \quad (2)$$

where, *k1*, *k2*, and *k3* represent 1$\times$1$\times$4, 1$\times$1$\times$2, and 1$\times$1$\times$2 convolutional filter kernels, respectively. γ represents the ReLU activation function. $\otimes$ signifies channel-wise convolution and * signifies the regular convolution operation. To enhance the fusion of concatenated feature information and reduce model parameters, channel-wise convolution is utilized to redistribute the weights of the concatenated features. Finally, *k2* and *k3* are employed to downscale the concatenated features. This design ensures the effective integration of features at different scales, incorporating both global and local information.

C. Loss Mapping with Skeleton Fitting Assistance

On the DRIVE dataset [39], approximately 77% of the vessel pixels belong to contrast strong vessels, while only 23% are contrast weak vessels [38]. Therefore, this paper introduces SFA to enhance the segmentation capability of contrast weak vessels. Three feature maps are generated in this study: one from the green Base-Net, another from the blue Base-Net, and

the third from the Feature Fusion Module (FFM). Specifically, the feature maps outputted by the two Base-Nets undergo skeleton fitting assisted loss mapping, while the output segmentation image from the FFM is combined with the ground truth for loss calculation. The specific workflow is depicted in Fig. 5.
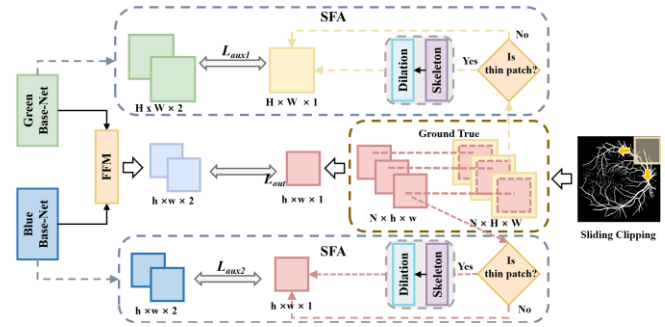

Fig. 5.   The process workflow of loss mapping with skeleton fitting assistance.

The original segmentation map is cropped into N patches of size H$\times$W using a sliding window approach. As the output segmentation map from the Feature Fusion Module (FFM) has dimensions h$\times$w, central cropping is performed to obtain h$\times$w patches. Subsequently, the cross-entropy loss function is applied for loss calculation, and it is given in Eq. (3):

$$L_{out} = -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \right] \quad (2)$$

where, *N* represents the total number of pixels in the patch, p_i denotes the model's predicted probability for pixel , y_i represents the true label of pixel *i*.

In Algorithm 1, we propose an algorithm for the classification of vessel patches. In the SFA, we adopt a similar approach for vessel patch classification. To enhance the model's learning capability for CWV, skeleton prior knowledge is employed to dilate vessel pixels within CWVP. Through the classification algorithm, H$\times$W patches of the ground truth, as well as the center-cropped h$\times$w patches, are classified. When a patch is identified as CSVP, no additional processing is performed, and the Dice loss is directly computed with the corresponding Base-Net output feature map. In the case of a patch classified as CWVP, the patch undergoes skeleton extraction and dilation operations to generate a new ground truth. The Dice loss is then calculated by comparing the outputs of the two Base-Nets with the new ground truth. The Dice loss function is given by Eq. (4):

$$L_{aux} = \begin{cases} 1 - \dfrac{2|P \cap D(S(Y))|}{|P| + |D(S(Y))|}, & Y \in CWVP \\[4mm] 1 - \dfrac{2|P \cap (Y)|}{|P| + |Y|}, & Y \in CSVP \end{cases} \quad (3)$$

where, *P* represents the model's prediction result, *Y* is the corresponding ground truth, *S*(·) denotes the skeleton extraction operation, *D*(·) represents the dilation operation.

The final loss function is given in Eq. (5):

$$L_{all} = \begin{cases} L_{out} + \dfrac{L_{out}}{L_{aux1}} L_{aux1} + \dfrac{L_{out}}{L_{aux2}} L_{aux2}, & Y \in CWVP \\ L_{out} + L_{aux1} + L_{aux2}, & Y \in CSVP \end{cases} \quad (4)$$

where, L_{aux1} and L_{aux2} are the corresponding loss functions for the two SFA components.

IV. EXPERIMENTS

A. Fundus Datasets

The existing widely used datasets include DRIVE [39], STARE [40], CHASE_DB1 [41], and HRF [42].

1) DRIVE dataset: The dataset comprises 40 retinal images, each measuring 565 × 584 pixels, with 20 images allocated for training. Additionally, the dataset has been pre-divided into training and testing sets. In the experiments, we utilize the first set of labels to evaluate the proposed method.

2) STARE dataset: The dataset consists of 20 images, each with dimensions of 700 × 605 pixels, with half of them representing pathological cases and the other half representing normal cases. We choose 10 images for training and another 10 images for testing.

3) CHASE_DB1 dataset: The dataset comprises 28 retinal images from various children, each with dimensions of 999 × 960 pixels for both left and right eyes. Annotations in CHASE_DB1 are provided by two different observers. In our experiments, we utilize first set of labels as the ground truth for evaluation. Specifically, we designate 8 images as test samples and allocate the remaining images for training samples.

4) HRF dataset: The dataset consists of 45 high-resolution color retinal images. These images are categorized into three classes: healthy, diabetic retinopathy, and glaucoma, each comprising 15 images. For training, we select 15 images from each category: healthy children, diabetic retinopathy patients, and glaucoma patients. The remaining 30 images are allocated for testing.

B. Evaluation Metrics

In retinal images, binary pixel values typically classify pixels into vessel (positive) and non-vessel (negative) categories. Based on this comparison, there are four fundamental pixel metrics: pixels labeled as vessels and correctly predicted as vessels are defined as True Positives (TP); pixels labeled as backgrounds but incorrectly predicted as vessels are False Positives (FP); similarly, False Negatives (FN) represent pixels labeled as vessels but incorrectly predicted as backgrounds, while True Negatives (TN) denote pixels labeled as backgrounds and correctly predicted as backgrounds. Furthermore, we employed eight metrics for evaluation, including Accuracy (Acc), Sensitivity (Sen), Specificity (Spe), Precision (Pre), F1 score (F1), G-mean (G),

Matthews Correlation Coefficient (MCC), and Area Under the ROC Curve (AUC). all metrics are defined as follows:

$$Acc = \frac{TP+TN}{TP+FN+TN+FP}, \quad (5)$$

$$Sen = \frac{TP}{TP+FN}, \quad (6)$$

$$Spe = \frac{TN}{TN+FP}, \quad (7)$$

$$Pre = \frac{TP}{TP+FP}, \quad (8)$$

$$F1 = \frac{2 \cdot Pre \cdot Sen}{Pre+Sen}, \quad (9)$$

$$G = \sqrt{Sen \cdot Spe}, \quad (10)$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}, \quad (11)$$

C. Implementation Details

The proposed model was implemented using PyTorch 1.12.1 on an RTX 3060 GPU with 12GB of memory. Stochastic gradient descent (SGD) was utilized for parameter optimization. Additionally, a "poly" learning rate policy was applied, where the learning rate was multiplied by $(1 - \frac{iter}{Max\_iter})^{power}$ and the power was set to 0.9, with a base learning rate of 0.01 in training. Furthermore, the batch size was set to 1 and each experiment was conducted for 100 epochs. During testing, the test images were partitioned into overlapping small patches. Subsequently, the segmented samples generated by the model were reconstructed into complete segmentation results.

V. RESULTS

A. Comparison with Other Methods

1) Comparison of composite metrics: We initially compared it with six methods on four datasets, including FANet[43], Convmixer[44], BFMD[45], GT-U-Net[46], SA-Unet[47], and CFPNetM[48], with quantitative metrics shown in Table I to Table IV.. Overall, the proposed method showed improvements in metrics,especially Acc and Pre. As illustrated in Table I, across the DRIVE dataset, Sen, G, and AUC metrics outperformed the six methods, reflecting superior segmentation results. F1 and MCC metrics were also close to the optimal values. Among them, G, AUC, F1, and MCC metrics are more comprehensive, indicating that the proposed method has better vascular segmentation capabilities.As depicted in Table III, on the STARE dataset, Acc, Spe, Pre, and F1 metrics achieved optimal results, and the other four metrics ranked second among the seven methods, showing results close to the optimal values.

Particularly, the F1 metric, reflecting the connectivity and accuracy of vascular segmentation, indicated the good segmentation demonstrated by the proposed method on this dataset indicate its excellent segmentation performance. As shown in Table IV, the HRF dataset includes high-resolution images of patients afflicted with diabetic retinopathy and glaucoma, posing challenges for vessel segmentation. Experimental results demonstrate that the proposed method for retinal vessel segmentation outperforms others, achieving the highest accuracy metric (Acc), while also securing the second or third position in the comparison of comprehensive metrics. These results suggest that the devised approach demonstrates promising segmentation capabilities for high-resolution and pathological images. The examples provided in 0 demonstrate the favorable segmentation performance of the proposed method. In the first and last rows of 0, DUF-Net exhibits more complete segmentation of thin blood vessels with fewer vessel interruptions. The features of DUF-Net contribute to preserving the continuity and accuracy of thin blood vessels in datasets like HRF, which include images with diabetic retinopathy and glaucoma. This capability is crucial for supporting the prevention and treatment of retinal pathologies.

TABLE I.   COMPREHENSIVELY COMPARING PERFORMANCE ON DRIVE DATASET (UNIT: %)

Dataset	Method	Accurary	Sensitivity	Specificity	Precision	F1-score	G-mean	MCC	AUC
DRIVE	FANet[43]	96.16	80.14	97.88	80.32	80.23	88.57	78.10	89.01
	Convmixer[44]	96.65	77.45	98.52	84.21	81.29	87.16	78.68	87.98
	BFMD[45]	96.65	72.27	99.11	89.18	79.71	84.60	78.50	85.69
	GT-U-Net[46]	96.80	76.35	98.85	87.23	81.26	86.83	79.83	87.60
	SA-Unet[47]	96.68	81.95	98.00	78.60	80.24	89.62	78.45	89.97
	CFPNetM[48]	95.77	73.39	98.09	81.28	73.58	83.08	73.07	85.73
	Ours	96.70	82.18	98.00	78.68	80.39	89.74	78.61	90.09

TABLE II.   COMPREHENSIVELY COMPARING PERFORMANCE ON CHASE_DB1 DATASET (UNIT: %)

Dataset	Method	Accurary	Sensitivity	Specificity	Precision	F1-score	G-mean	MCC	AUC
CHASE_DB1	FANet[43]	92.66	60.30	95.36	52.73	55.72	75.53	52.22	77.83
	Convmixer[44]	97.20	75.30	98.99	88.90	81.47	86.39	80.35	87.23
	BFMD[45]	97.10	71.37	99.26	88.99	79.19	84.16	78.21	85.31
	GT-U-Net[46]	97.23	72.22	99.33	90.08	80.14	84.69	79.24	85.77
	SA-Unet[47]	96.99	72.16	99.07	86.67	78.69	84.53	77.49	85.61
	CFPNetM[48]	96.87	80.27	98.34	80.99	80.62	88.84	78.92	89.30
	Ours	97.24	70.75	99.44	91.29	79.67	83.86	78.97	85.09

TABLE III.   COMPREHENSIVELY COMPARING PERFORMANCE ON STARE DATASET (UNIT: %)

Dataset	Method	Accurary	Sensitivity	Specificity	Precision	F1-score	G-mean	MCC	AUC
STARE	FANet[43]	87.33	74.36	88.29	38.52	48.45	80.27	47.30	81.32
	Convmixer[44]	96.61	79.71	97.58	78.66	77.67	76.72	87.88	88.83
	BFMD [45]	96.86	82.45	98.03	77.75	79.71	89.76	78.24	90.24
	GT-U-Net[46]	97.18	79.65	98.61	82.89	80.61	88.39	79.47	89.12
	SA-Unet[47]	96.74	78.97	98.16	79.35	77.60	87.50	76.72	88.56
	CFPNetM[48]	93.54	90.76	93.75	56.42	68.81	92.16	68.27	92.25
	Ours	97.31	84.95	98.80	85.01	83.36	91.36	81.98	91.66

TABLE IV.   COMPREHENSIVELY COMPARING PERFORMANCE ON HRF DATASET (UNIT: %)

| Dataset	Method	Accurary	Sensitivity	Specificity	Precision	F1-score	G-mean	MCC	AUC
HRF	FANet[43]	96.96	82.15	98.21	79.28	80.47	89.77	78.98	90.18
	Convmixer[44]	96.90	70.07	99.16	87.60	77.48	83.22	76.60	84.61
	BFMD[45]	94.33	78.40	95.68	60.94	68.40	86.59	66.08	87.03
	GT-U-Net[46]	96.64	68.37	99.04	86.00	75.55	82.10	74.68	83.70
	SA-Unet[47]	96.18	74.55	97.97	75.79	74.75	85.31	72.93	86.26
	CFPNetM[48]	96.81	82.38	98.03	77.97	79.90	89.83	78.33	90.20
	Ours	97.08	76.76	98.79	84.33	80.02	86.98	78.75	87.76
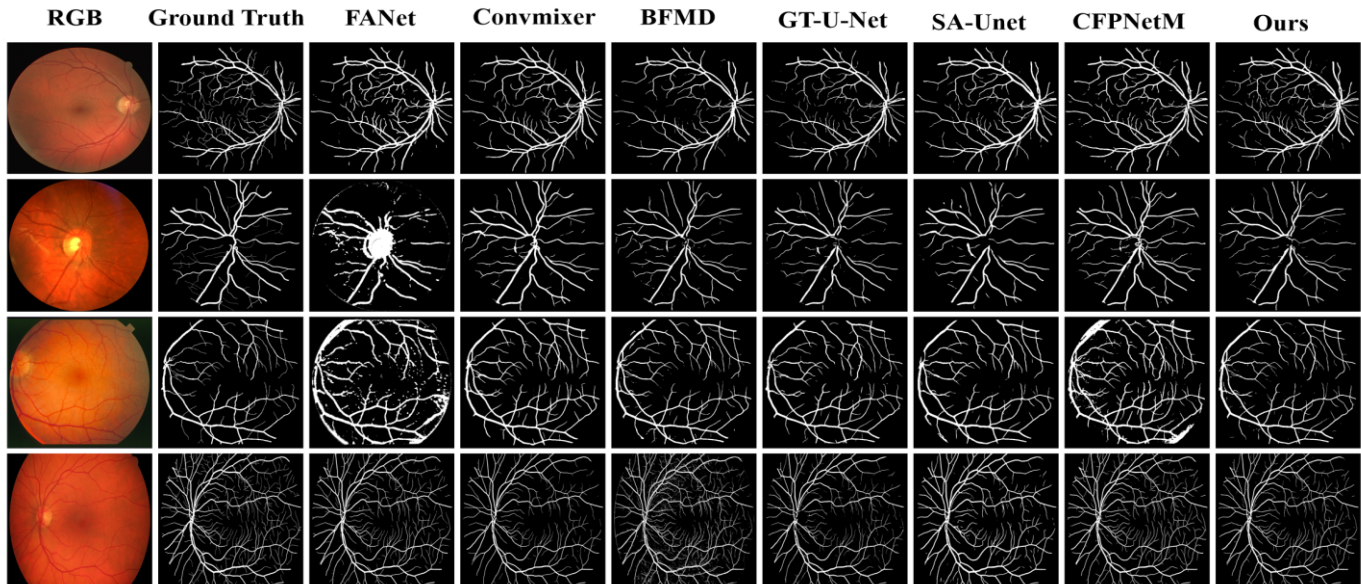
Fig. 6. The segmentation outcomes from the DRIVE, CHASE_DB1, STARE, and HRF datasets are juxtaposed in the figures below (from top to bottom).

2) Segmentation of optic disc and pathological vessel regions: Detailed comparisons of the optic disc and pathological vessel regions are illustrated in Fig. 7 demonstrating the magnified views of these areas. The first and second rows of Fig. 7 compare the optic disc regions with state-of-the-art methods from recent years. Due to challenges such as low contrast and uneven brightness, blood vessels around the optic disc region are often overlooked by other methods. In contrast, the proposed approach effectively identifies the vessels around the optic disc. Furthermore, the presence of bleeding around the lesions, which closely resembles vascular features, significantly increases the difficulty of vessel segmentation in lesion images. As evident in the third rows of Fig. 7, the proposed method accurately segments vessels in the pathological regions. This is primarily attributed to the integration of skeletal prior knowledge to augment the model's comprehension of fine vessels and enhance its discrimination between vessel and background pixels .In the fourth row of Fig. 7, a substantial amount of noise is evident in the segmentation image due to BFMD misclassifying numerous lesions and surrounding hemorrhagic areas as vessels. This could potentially compromise subsequent auxiliary diagnostics.
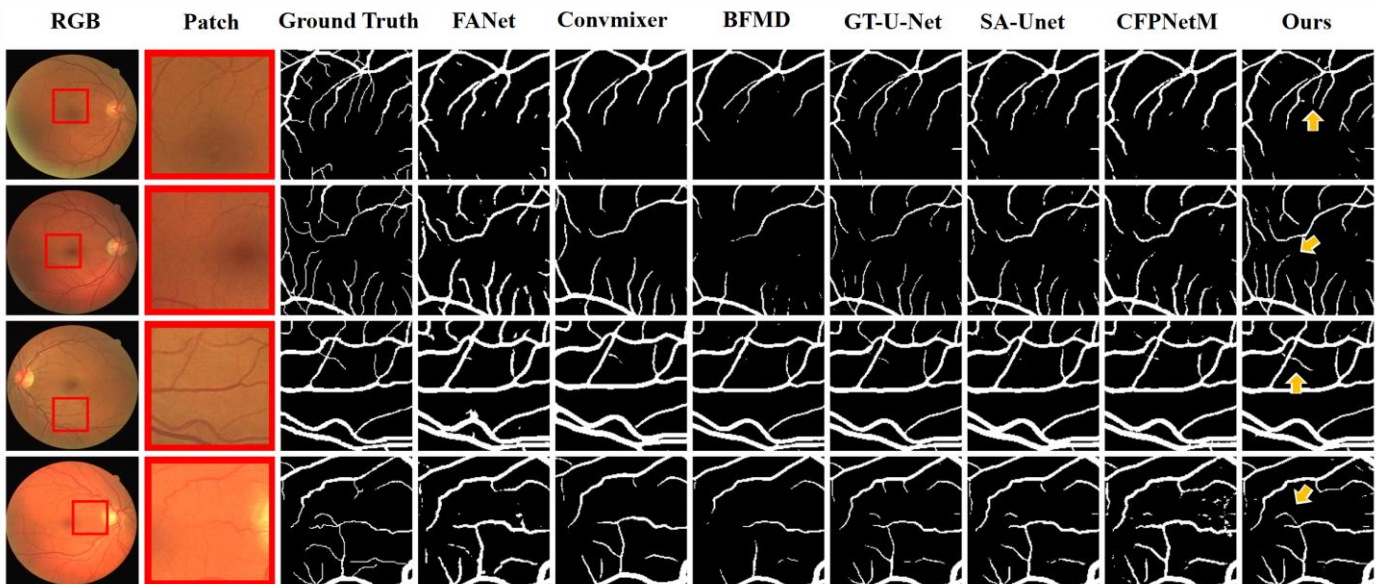

Fig. 7. Comparison of vessel segmentation results in the optic disc region (first two rows) and the lesion region (last two rows).

3) Segmentation of contrast weak vessels: The proposed method accurately segments contrast weak vessels, as illustrated in 0. When compared with other methods, it is evident that DUF-Net extracts more complete contrast weak vessels. This improvement can be attributed to the model's ability to learn both local and global features, effectively addressing the missing contextual information. This results in superior performance in detecting thin vessels, reflected in a higher accuracy (Acc) score.

B. Experiment for Algorithm Involves Categorizing CSVP and CWVP

In Algorithm 1, introduces a patch classification algorithm designed to categorize CSVP and CWVP. 0 illustrates the classification results of the patch classification algorithm on the DRIVE dataset. The first and second rows show the patches classified as CWVP by the algorithm and their corresponding ground truth. In this classification result, all vessels are characterized by low contrast. The third and fourth rows display the patches classified as CSVP by the algorithm and their corresponding ground truth. Here, it is evident that vessels with strong contrast are prominent.
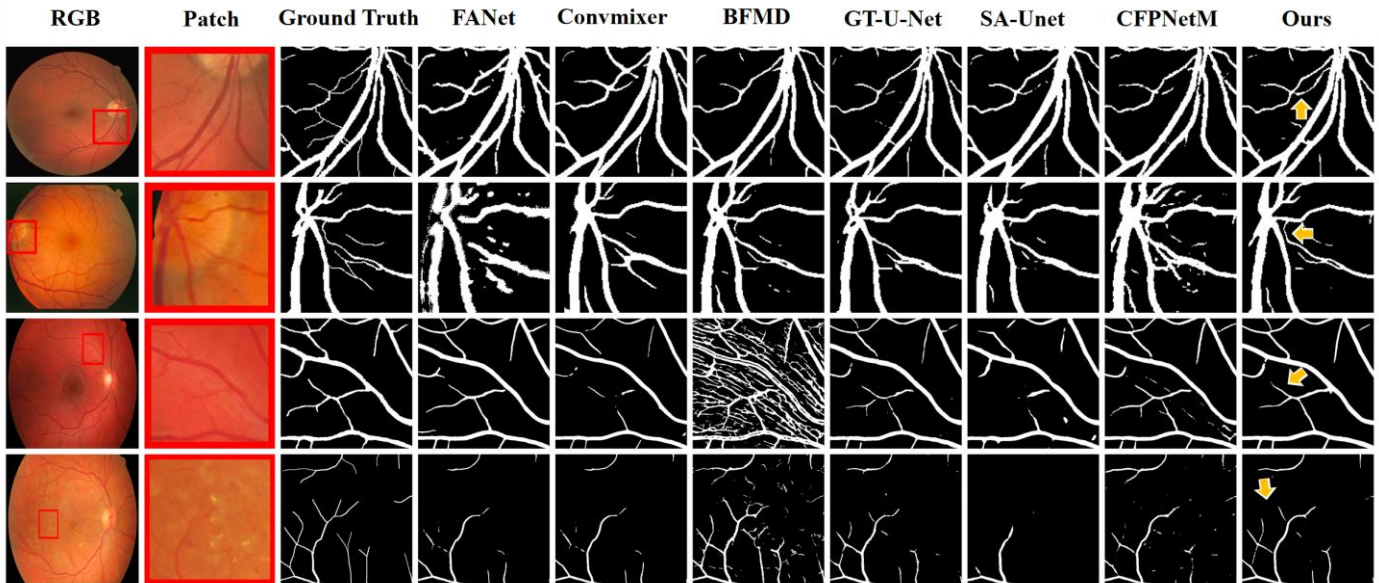

Fig. 8. Comparison of contrast weak vessel segmentation results.
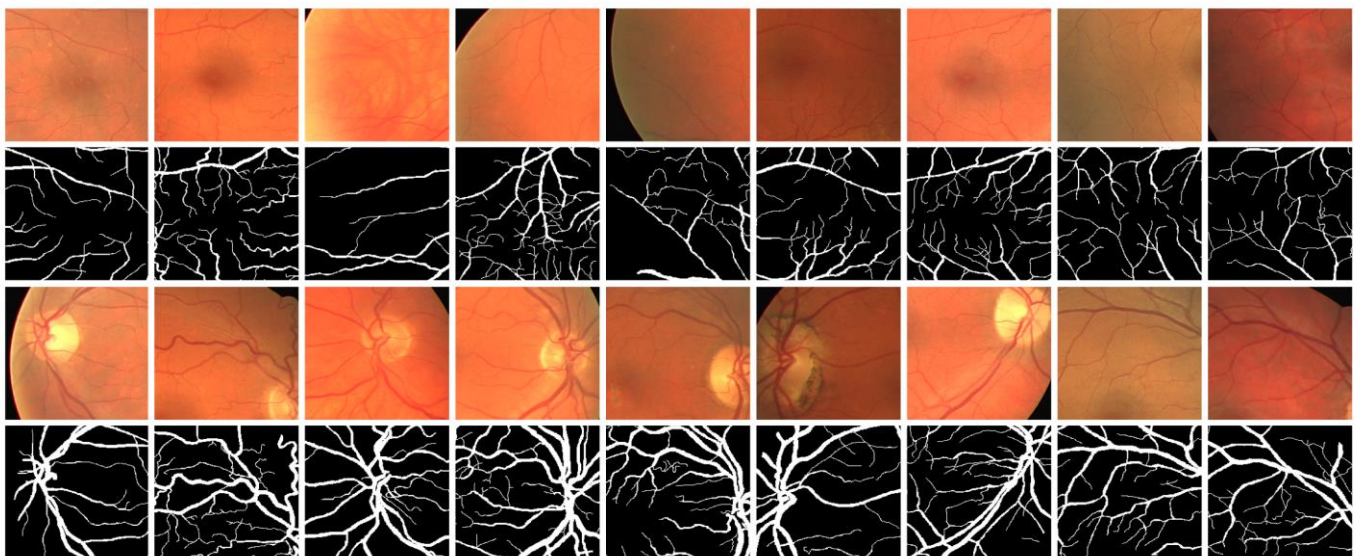

Fig. 9. The outcomes of patch classification into CSVP (first two rows) and CWVP (last two rows) on the DRIVE dataset are juxtaposed with their respective ground truth.

C. Experiment for Loss Mapping with Skeleton Fitting Assistance

In Fig. 5, we proposed a loss mapping with skeleton fitting assistance composed of three loss functions: L_{aux1}, L_{aux2}, and L_{out}. The evolution of these three loss functions during training for 100 epochs on the DRIVE dataset is illustrated in 0 . The values of L_{aux1} and L_{aux2} are greater than the value of L_{out}. When the patch represents CSVP, the coefficient weights of the three loss functions are configured as 1, and the values of the three loss functions show a synchronous decreasing trend in 0(a). However, when the patch represents CWVP, the skeleton fitting assistance part involves computing losses using the feature map and the newly generated ground truth after skeleton extraction. This may bias the model towards learning the new ground truth, leading to suboptimal segmentation results. Therefore, an adaptive weight assignment method is designed in this paper. It uses L_{aux1}/L_{out} as the weight for L_{aux1}

and L_{aux2}/L_{out} as the weight for L_{aux2} to balance the loss functions. In 0(b), it can be observed that with the adaptive weight assignment, L_{out} shows a decreasing trend, while L_{aux1} and L_{aux2} remains within a certain range of variation.

D. Ablation Analysis

The ablation analysis presents the performance of each proposed method, as demonstrated in TABLE V. . For comparison, we refer to DUF-Net without contextual information as DU-Net. DU-Net does not undergo central cropping when fed into the model, employs blocks of the same size, and omits the use of the Feature Fusion Module (FFM). Additionally, DU-Net does not undergo the proposed preprocessing methods and the loss calculation of Skeleton Fitting Assist (SFA). By incorporating preprocessing methods, enhancing contextual information supplementation, and integrating FFM and SFA loss calculation, all metrics show corresponding improvements, as depicted in TABLE V. .
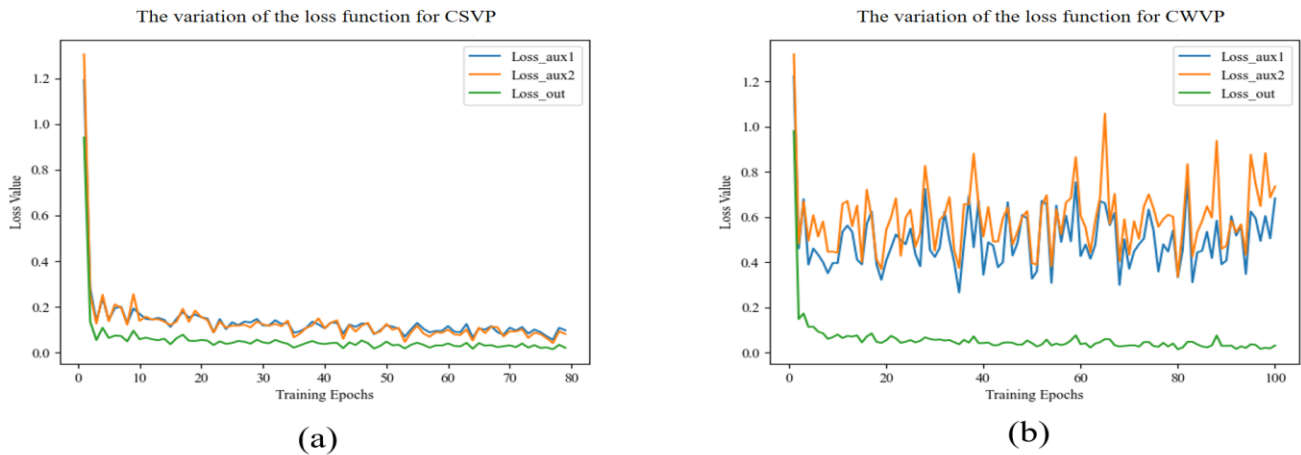

Fig. 10. The variation of the loss functions on the DRIVE dataset (a) The variation of the loss function for CSVP (b) The variation of the loss function for CWV.

TABLE V. THE RESULTS OF THE ABLATION ANALYSIS ON THE DRIVE DATASET (UNIT: %)

Method	Accurary	Sensitivity	Specificity	Precision	F1-score	G-mean	MCC	AUC
DU-Net	96.62	77.19	98.52	83.55	79.76	87.05	78.28	87.85
+Preprocess	96.65	81.97	97.94	79.47	80.22	87.39	77.98	88.27
+FFM	96.68	81.68	97.98	78.28	79.98	88.50	78.51	89.18
+SFA	96.70	82.18	98.00	78.68	80.39	89.74	78.61	90.09

VI. DISCUSSION

Compared to end-to-end segmentation, patch-based segmentation methods can significantly alleviate computational hardware pressure while increasing the training samples in datasets, thereby enhancing vessel segmentation capability, which has gradually attracted attention from many scholars [49]. However, when restoring the segmented patch results to the original size, issues such as vessel discontinuity and poor segmentation of small vessels may arise. Therefore, this paper designs a preprocessing method tailored for fundus images, a novel network architecture, and a training approach with skeleton fitting assistance to address the aforementioned issues.

Due to the difficulty in segmenting small vessels, datasets with a higher density of small vessels often exhibit lower

performance metrics compared to datasets with fewer small vessels, such as DRIVE [39] and HRF [42] datasets containing more small vessels than STARE [40] and CHASE_DB1 [41] datasets. Consequently, the performance of six vessel segmentation methods is significantly reduced. Additionally, existing fundus segmentation methods are primarily evaluated based on pixel measurements, where the contribution of small vessel pixels is relatively minor, limiting the potential improvement in performance metrics. The proposed method shows limited improvement on the HRF dataset. This is partly due to the high resolution of this dataset leads our classification algorithm to misclassify some CWVP samples as CSVP, resulting in incomplete feature capture of CWVP. Thus, there remains considerable room for improvement in our method.

VII. Conclusion

This paper proposes a novel segmentation method by addressing the limitations of patch-based retinal vessel segmentation. Firstly, the patch-based segmentation approach often overlooks the issue of imbalanced contrast between strong and weak vessel samples in the dataset. Therefore, this paper designs a vessel patch classification algorithm to balance the training data samples by augmenting contrast weak vessel samples based on the computed quantities. To address the missing contextual information around the patches, new network architecture, DUF-Net, is proposed. By feeding features of different scales into the network model, both global and local information are learned separately and then fused to complement the missing features. Lastly, skeleton prior knowledge is introduced to alleviate vessel discontinuity issues after segmentation, and the adaptive weight allocation mechanism is employed to adjust the imbalanced pixel distribution within blocks, thereby enhancing the model's segmentation capability for contrast weak vessels.

In the experimental results, the proposed method demonstrates promising performance. As retinal lesions significantly affect vessel segmentation results, future work will focus on enhancing lesioned retinal images using image enhancement techniques and loss functions to improve the performance of segmentation. Moreover, extending this method to other segmentation tasks in medicine, such as neuron segmentation and cell segmentation, will also be considered.

References

[1] Li T, Bo W, Hu C, et al. Applications of deep learning in fundus images: A review[J]. Medical Image Analysis, 2021, 69: 101971.

[2] Chen C, Chuah J H, Ali R, et al. Retinal vessel segmentation using deep learning: a review[J]. IEEE Access, 2021, 9: 111985-112004.

[3] Lechner J, O'Leary O E, Stitt A W. The pathology associated with diabetic retinopathy[J]. Vision research, 2017, 139: 7-14.

[4] Abràmoff M D, Garvin M K, Sonka M. Retinal imaging and image analysis[J]. IEEE reviews in biomedical engineering, 2010, 3: 169-208.

[5] L Srinidhi C, Aparna P, Rajan J. Recent advancements in retinal vessel segmentation[J]. Journal of medical systems, 2017, 41: 1-22.

[6] Dai P, Luo H, Sheng H, et al. A new approach to segment both main and peripheral retinal vessels based on gray-voting and gaussian mixture model[J]. PloS one, 2015, 10(6): e0127748.

[7] Nguyen U T V, Bhuiyan A, Park L A F, et al. An effective retinal blood vessel segmentation method using multi-scale line detection[J]. Pattern recognition, 2013, 46(3): 703-715.

[8] Lee Y, Hara T, Fujita H, et al. Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique[J]. IEEE Transactions on medical imaging, 2001, 20(7): 595-604.

[9] Hoover A D, Kouznetsova V, Goldbaum M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response[J]. IEEE Transactions on Medical imaging, 2000, 19(3): 203-210.

[10] Nyemeesha V, Ismail B M. Implementation of noise and hair removals from dermoscopy images using hybrid Gaussian filter[J]. Network Modeling Analysis in Health Informatics and Bioinformatics, 2021, 10: 1-10.

[11] Taşcı E, Uğur A. Shape and texture based novel features for automated juxtapleural nodule detection in lung CTs[J]. Journal of medical systems, 2015, 39: 1-13.

[12] Zakeri F S, Behnam H, Ahmadinejad N. Classification of benign and malignant breast masses based on shape and texture features in sonography images[J]. Journal of medical systems, 2012, 36: 1621-1627.

[13] Orlando J I, Prokofyeva E, Blaschko M B. A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images[J]. IEEE transactions on Biomedical Engineering, 2016, 64(1): 16-27.

[14] Li Q, Feng B, Xie L P, et al. A cross-modality learning approach for vessel segmentation in retinal images[J]. IEEE transactions on medical imaging, 2015, 35(1): 109-118.

[15] Liskowski P, Krawiec K. Segmenting retinal blood vessels with deep neural networks[J]. IEEE transactions on medical imaging, 2016, 35(11): 2369-2380.

[16] Zhang Y, He M, Chen Z, et al. Bridge-Net: Context-involved U-net with patch-based loss weight mapping for retinal blood vessel segmentation[J]. Expert Systems with Applications, 2022, 195: 116526.

[17] Sheng B, Li P, Mo S, et al. Retinal vessel segmentation using minimum spanning superpixel tree detector[J]. IEEE transactions on cybernetics, 2018, 49(7): 2707-2719.

[18] Lam B S Y, Gao Y, Liew A W C. General retinal vessel segmentation using regularization-based multiconcavity modeling[J]. IEEE Transactions on medical imaging, 2010, 29(7): 1369-1381.

[19] Rezaee K, Haddadnia J, Tashk A. Optimized clinical segmentation of retinal blood vessels by using combination of adaptive filtering, fuzzy entropy and skeletonization[J]. Applied Soft Computing, 2017, 52: 937-951.

[20] Staal J, Abràmoff M D, Niemeijer M, et al. Ridge-based vessel segmentation in color images of the retina[J]. IEEE transactions on medical imaging, 2004, 23(4): 501-509.

[21] You X, Peng Q, Yuan Y, et al. Segmentation of retinal blood vessels using the radial projection and semi-supervised approach[J]. Pattern recognition, 2011, 44(10-11): 2314-2324.

[22] Soares J V B, Leandro J J G, Cesar R M, et al. Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification[J]. IEEE Transactions on medical Imaging, 2006, 25(9): 1214-1222.

[23] Kumar Agarwal A, Angeline Ranjithamani D, Pavithra M, et al. Machine learning technique for the assembly-based image classification system[J]. J Nucl Ene Sci Power Generat Techno, 2021, 10(9).

[24] Orlando J I, Prokofyeva E, Blaschko M B. A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images[J]. IEEE transactions on Biomedical Engineering, 2016, 64(1): 16-27.

[25] Zhu C, Zou B, Zhao R, et al. Retinal vessel segmentation in colour fundus images using extreme learning machine[J]. Computerized Medical Imaging and Graphics, 2017, 55: 68-77.

[26] Hu K, Zhang Z, Niu X, et al. Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function[J]. Neurocomputing, 2018, 309: 179-191.

[27] Hu J, Wang H, Gao S, et al. S-unet: A bridge-style u-net framework with a saliency mechanism for retinal vessel segmentation[J]. IEEE Access, 2019, 7: 174167-174177.

[28] Mou L, Chen L, Cheng J, et al. Dense dilated network with probability regularized walk for vessel detection[J]. IEEE transactions on medical imaging, 2019, 39(5): 1392-1403.

[29] Mishra S, Chen D Z, Hu X S. A data-aware deep supervised method for retinal vessel segmentation[C]//2020 IEEE 17th international symposium on biomedical imaging (ISBI). IEEE, 2020: 1254-1257.

[30] Wang D, Hu G, Lyu C. Frnet: an end-to-end feature refinement neural network for medical image segmentation[J]. The Visual Computer, 2021, 37: 1101-1112.

[31] Dasgupta A, Singh S. A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation[C]//2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017). IEEE, 2017: 248-251.

[32] Yan Z, Yang X, Cheng K T. Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation[J]. IEEE Transactions on Biomedical Engineering, 2018, 65(9): 1912-1923.

[33] Wu Y, Xia Y, Song Y, et al. NFN+: A novel network followed network for retinal vessel segmentation[J]. Neural Networks, 2020, 126: 153-162.

[34] Wang D, Haytham A, Pottenburgh J, et al. Hard attention net for automatic retinal vessel segmentation[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 24(12): 3384-3396.

[35] Yang L, Wang H, Zeng Q, et al. A hybrid deep segmentation network for fundus vessels via deep-learning framework[J]. Neurocomputing, 2021, 448: 168-178.

[36] Tan, Yubo, et al. "Retinal vessel segmentation with skeletal prior and contrastive loss." IEEE Transactions on Medical Imaging 41.9 (2022): 2238-2251.

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

[38] Yan Z, Yang X, Cheng K T. A skeletal similarity metric for quality evaluation of retinal vessel segmentation[J]. IEEE transactions on medical imaging, 2017, 37(4): 1045-1057.

[39] Staal J, Abràmoff M D, Niemeijer M, et al. Ridge-based vessel segmentation in color images of the retina[J]. IEEE transactions on medical imaging, 2004, 23(4): 501-509.

[40] Hoover A D, Kouznetsova V, Goldbaum M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response[J]. IEEE Transactions on Medical imaging, 2000, 19(3): 203-210.

[41] Fraz M M, Remagnino P, Hoppe A, et al. An ensemble classification-based approach applied to retinal blood vessel segmentation[J]. IEEE Transactions on Biomedical Engineering, 2012, 59(9): 2538-2548.

[42] Odstrcilik J, Kolar R, Budai A, et al. Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database[J]. IET Image Processing, 2013, 7(4): 373-383.

[43] Tomar N K, Jha D, Riegler M A, et al. Fanet: A feedback attention network for improved biomedical image segmentation[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022.

[44] Solano A, Dietrich K N, Martínez-Sober M, et al. Deep Learning Architectures for Diagnosis of Diabetic Retinopathy[J]. Applied Sciences, 2023, 13(7): 4445.

[45] Deari S, Oksuz I, Ulukaya S. Block Attention and Switchable Normalization based Deep Learning Framework for Segmentation of Retinal Vessels[J]. IEEE Access, 2023.

[46] Li Y, Wang S, Wang J, et al. Gt u-net: A u-net like group transformer network for tooth root segmentation[C]//Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 12. Springer International Publishing, 2021: 386-395.

[47] Sun J, Darbehani F, Zaidi M, et al. Saunet: Shape attentive u-net for interpretable medical image segmentation[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23. Springer International Publishing, 2020: 797-806.

[48] Lou A, Guan S, Loew M. Cfpnet-m: A light-weight encoder-decoder based network for multimodal biomedical image real-time segmentation[J]. Computers in Biology and Medicine, 2023, 154: 106579.

[49] Xia H, Jiang F, Deng S, et al. Mapping functions driven robust retinal vessel segmentation via training patches[J]. IEEE access, 2018, 6: 61973-61982.