SRNet: Striped Pyramid Pooling and Relational Transformer for Retinal Vessel Segmentation

Wei Yan*
2021212122@nwnu.edu.cn
Yun Jiang
jiangyun@nwnu.edu.cn
Zequn Zhang
2021222178@nwnu.edu.cn
Yao Yan
2022222261@nwnu.edu.cn
Bingxi Liu
2022222292@nwnu.edu.cn

School of Computer Science and Engineering,
Northwest Normal University,
Lanzhou, Gansu Province, China

Abstract

The morphology of retinal vessels is crucial for diagnosing and screening retinal diseases such as age-related macular degeneration and diabetic retinopathy. Retinal vessels segmentation is an indispensable part of retinal disease screening and diagnosis. However, due to the inherent complex structural features of retinal vessels, it remains a challenging visual task. Based on the type of input, retinal vessels segmentation approaches can be roughly divided into both image-level and patches-level methods, which have their respective benefits and drawbacks. To better leverage both input methods, we design a Relational Transformer Module (RTM) to effectively combine local patches-level information with image-level global contextual information. Furthermore, retinal vessels exhibit varying lengths with tree-like branching patterns, making the classical rectangular pooling inefficient in capturing accurate vessels information because they are better suited for uniformly distributed objects. To better capture contextual information, we further developed a Striped Pyramid Pooling Module (SPPM) to adapt to the tree-like distribution of retinal vessels. Based upon these foundations, we propose a retinal vessels segmentation **Net**work with the **S**triped Pyramid Pooling Module and the **R**elational Transformer Module (**SRNet**). Experimental validation showed that our SRNet outperforms other advanced methods on the DRIVE and CHASE datasets.

1 Introduction

Retinal vessels are important structures that can be observed in fundus images. The width, tortuosity, trend, and branch changes of blood vessels are important means for diagnosing various diseases [6]. For example, diabetic retinopathy is a microvascular complication caused by elevated blood sugar levels that causes blood vessels in the retina to swell

[24]. Therefore, the automatic segmentation of retinal vessels is particularly important in the process of assisting medical diagnosis. With the development of deep learning [9, 12, 15], it has solved some problems that cannot be solved by traditional methods [2, 20, 35], and has become the mainstream of retinal vessel segmentation. Among them, U-Net [21] performed the best where it uses the encoder to extract the rich semantic information of the image, and uses the decoder to achieve accurate segmentation. Some improved versions of U-Net methods [1, 19, 27, 37] also achieved advanced results. Due to the inherent characteristics of retinal vessels: slender branches and indistinguishable ends, retinal vessels are difficult to segment. To further improve the segmentation accuracy, it is particularly important to obtain the global context information of the retinal image and the information of the subtle ends. Wang et al. [26] proposed a dual-path U-Net to capture semantic information on the context path with multi-scale convolutional blocks. SCS-Net [32] utilized an Adaptive Feature Fusion (AFF) module to guide efficient fusion between adjacent hierarchical features to capture more contextual semantic information. Cheng et al. [6] encoded context information by sampling mixed features from the orientation-invariant local context. Bridge-Net [36] incorporated the recurrent neural network (RNN) into the convolutional neural network (CNN) to provide the contextual information to generate a probabilistic map of retinal vessels.

The above methods can be roughly divided into two categories according to the type of input. One is to use the entire image as the input of the network [8, 17, 28, 36]. This type of method can preserve the remote background information of the retinal image to the greatest extent, and help the neural network perceive the overall structure and shape of the retinal vessels. However, it cannot effectively segment thin and low-contrast vessels. The other is to divide the image into multiple patches as the input of the network [21, 30, 33, 34]. This type of method can better pay attention to the details of retinal blood vessels, and can effectively solve the defect of insufficient training data in the former type of method. However, retinal vessels span multiple patches, which makes it impossible to establish long-distance dependencies on a single patch, ultimately cannot well display the geometric features and global context features of blood vessels. Therefore, the method of combining two types of input methods [29] seems to be able to make up for their respective shortcomings, and the problem lies in how to fuse the feature information generated by these two types of input methods.

Common feature fusion methods include element-wise addition and dimension addition [13, 22] as well as feature map multiplication [3, 16]. Element-wise addition can be performed using simple mathematical operations such as scalar addition, which makes it faster than more complex operations like matrix multiplication. But since element-wise addition is linear, it lacks expressive power compared to more complex non-linear alternatives. Dimensional addition can build multiscale representations by building higher-order tensor products, which can encode finer nuances between objects. But it tends to create an overcomplete representation of the data, meaning that the resulting tensors contain many redundancies or unnecessary components. It doesn't actually help improve prediction accuracy, which may lead to an increased risk of overfitting.

Feature map multiplication also helps to reduce spatial dimensions by selectively combining features and discarding unimportant information, which enhances the ability of deep learning models to focus on relevant regions and suppress irrelevant regions, thereby improving object detection, recognition, and classification accuracy. Therefore, using attention [10, 31] or transformer [4, 11, 14] can focus on local features and establish long-distance dependencies between image patches. Based on the joint image-level and patches-level segmentation framework, we propose a novel fusion method to make better use of the advan-

tages brought by the two input methods. Specifically, we communicate the feature information obtained from the two input methods through a well-designed cross-transformer and self-attention. This module can effectively focus on the local information of the vessels end and the long-range context correlation of the blood vessels.

In addition, unlike other medical image segmentation tasks, retinal vessels are thin and long, distributed in an irregular tree. Previous pooling pyramid methods detect input feature maps within square or strip windows, which limits them to capture anisotropic contextual information in retinal vessels. To further solve this problem, we propose a novel striped pooling pyramid module to better adapt to the morphological features and distribution characteristics of retinal vessels, so as to better capture the contextual information of retinal vessels. Specifically, we use four long and narrow pooling kernels in different directions to capture the contextual information of retinal vessels, and fuse vessels feature information from different directions. The main contributions of our work are as follows:

- We propose a **Relational Transformer Module** for fusing image-level and patches-level information, which can combine the strengths of image-level and patches-level segmentation frameworks. The introduction of cross-transformer and self-attention effectively focuses on the local information of the vessels end and the long-distance context correlation of the vessels.
- According to the inherent characteristics of retinal vessels, we designed a novel **Striped Pooling Pyramid Module**, which can better capture the characteristic information of thin tree-like long blood vessels and further improve the segmentation accuracy.
- Based on the above innovations, we propose a retinal vessels segmentation network (**SRNet**). We conduct comprehensive experiments on DRIVE and CHASE datasets, all achieving state-of-the-art performance.

2 Proposed Method

In this section, we first introduce the framework of our proposed SRNet in Section 2.1 and then describe the Relational Transformer Module in Section 2.2. Finally describe the designed Striped Pooling Pyramid Module in Section 2.3.

2.1 Framework Overview

As shown in Figure 1, our SRNet consists of a shared encoder backbone, a Striped Pooling Pyramid Module (SPPM), a Relational Transformer Module (RTM), and an up-sampling block. Following the practice of Wang et al. [29], we crop the input image into patches $I^{(i)} \in R^{H\times W\times 3}$, $i \in N^2$, and down-sample the image to $\frac{1}{N}(h \times w)$ to get $I' \in R^{\frac{H}{N}\times\frac{W}{N}\times 3}$, and then input them into the shared encoder. Both branches use the same encoding backbone with shared weights, and both branches can operate in parallel by merging batches. Specifically, we employ an encoder similar to VGG [23] and ResNet [9] with 4 layers of convolution blocks to extract feature maps. The outputs of the two branches are subjected to pooling pyramid operation through our designed SPPM to obtain image-level and patches-level feature maps F' and $F^{(i)}$. The feature maps of the two branches communicate through RTM, which can effectively focus on the local information of the vessels end and the long-range contextual correlation of the vessels. Finally, the predicted segmentation results are obtained through an up-sampling decoder.

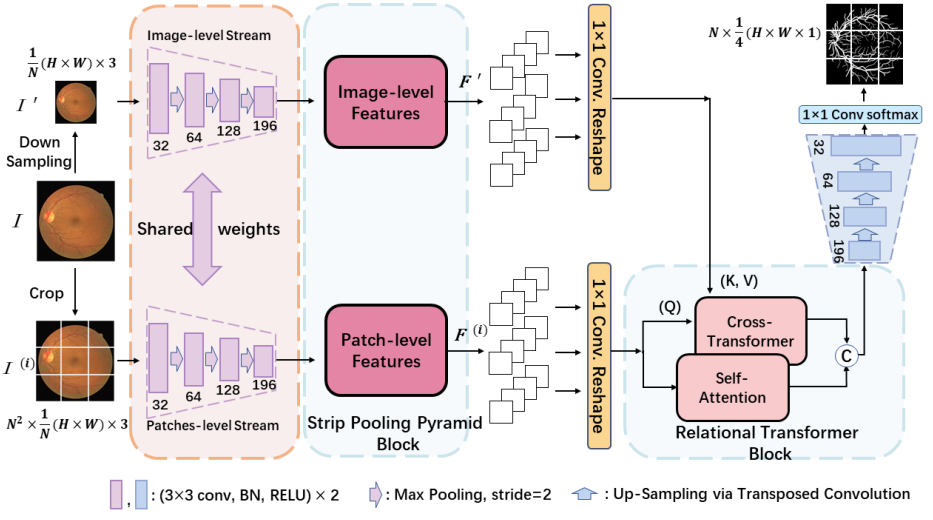

Figure 1: The overall architecture of our SRNet. The shared encoder consists of 4 layers of convolutional blocks, SPPM and RTM, and an up-sampling decoder.

2.2 Relational Transformer Module

Relational Transformer Module (RTM) consists of a self-attention head and a cross-attention head, which are used to capture the relationship between patches and the relationship between patches and image, as shown in Figure 2. In each attention head, three 1×1 convolutions and reshape are used to generate query, key, and value generators Q_i, K_i, V_i, $i \in s, c$. In the self-attention head and cross-attention head, queries and keys can be described as:

$$H_s(F_p) = K_s(F_p)^T Q_s(F_p) \tag{1}$$

$$H_c(F_p, F_i) = K_c(F_i)^T Q_c(F_p) \tag{2}$$

where s and c denote the self-attention head and cross-attention head, respectively. It should be emphasized that unlike the self-attention head, Q, K, and V all come from the image-level branch, the cross-attention head generates query from the image-level branch feature F_i to integrate vessels information. Next, the individual attention features of the two heads are calculated as:

$$G_s(F_p) = V_s(F_p)\, softmax(H_s(F_p)) \tag{3}$$

$$G_c(F_p, F_i) = V_c(F_i)\, softmax(H_c(F_p, F_i)) \tag{4}$$

We also use residual learning for each head to get the output:

$$F_i = W_i G_i(F_p, F_i) \oplus F_p \quad i \in \{s, c\} \tag{5}$$

Among them, W_i is a linear embedding of 1×1 convolution, and the $\oplus$ operation is performed through the residual connection of element-wise addition.

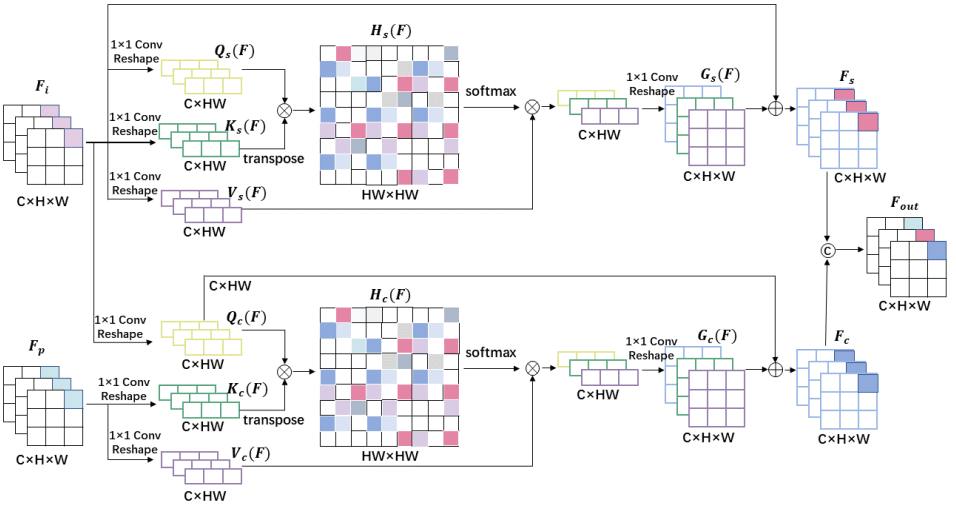

Figure 2: The details of the Relational Transformer Module(RTM).

We concatenate F_i from the self-attention head and F_p from the cross-attention head to get the final RTM output:

$$F_{out} = [F_p, F_i] \tag{6}$$

where $[\cdot,\cdot]$ denotes the channel connection.

2.3 Striped Pyramid Pooling Module

The inherent morphology of retinal vessels makes segmentation very difficult: the retinal vessel branches are slender, the boundaries are difficult to distinguish, and the relationship between vessels is complex [29]. In this case, contextual information around retinal vessels is extremely important for vessel segmentation. Previously, to obtain global image-level features, spatial pyramid pooling was widely used. However, commonly used pooling pyramid modules usually use $N \times N$ square pooling kernels (as shown in Figure 3 (a)), $1 \times N$ or $N \times 1$ strip pooling kernels (as shown in Figure 3 (b))), it cannot capture the dendritic curve features such as retinal vessels in the fundus well, and it will inevitably introduce irrelevant information from adjacent pixels. In order to better collect context information around retinal vessels, we propose a Strip Pooling Pyramid Module (SPPM) suitable for the elongated tree-like distribution of retinal vessels (as shown in (c) in Figure 3).

Specifically, we devise a novel method named Striped Pooling Pyramid Module (SPPM), it utilizes strip pooling operations in horizontal and vertical as well as two diagonal directions to help the network capture long-range contextual information in different spatial dimensions, such a pooling kernel design is more in line with the tree-like distribution of vessel shapes. Figure 4 depicts our proposed SPPM, let $X \in R^{C \times H \times W}$ be the input tensor, where C denotes the number of channels. We first feed X into four parallel paths, which consist of a horizontal, vertical, left-diagonal, and right-diagonal strip pooling layer, followed by a 1D convolutional layer with a kernel size of 1, used to modulate the current position and its

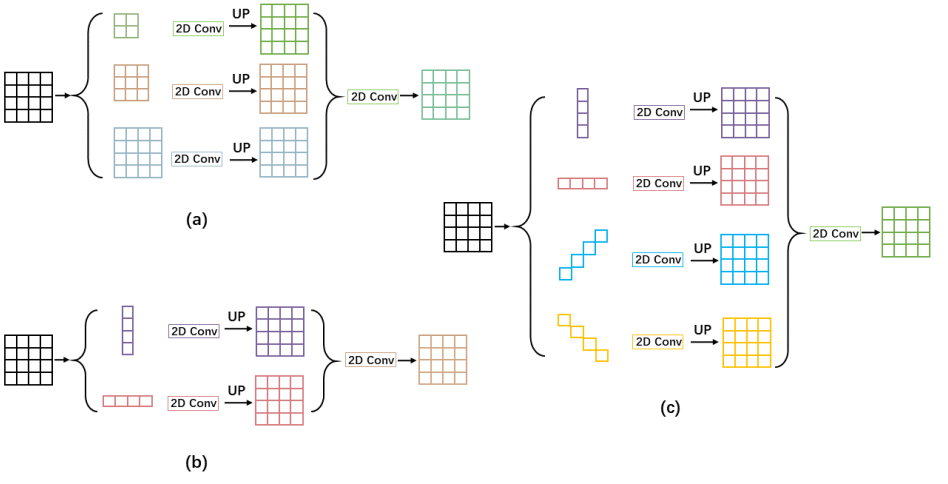

Figure 3: (a) Regular square pooling pyramid. (b) Regular strip pooling pyramid. (c) Our proposed pooling pyramid with a diagonal pooling kernel.

neighbor features. This gives $y^{horizontal} \in R^{C\times H}$, $y^{vertical} \in R^{C\times H}$, $y^{leftdiagonal} \in R^{C\times\sqrt{H^2+W^2}}$, $y^{rightdiagonal} \in R^{C\times\sqrt{H^2+W^2}}$. To get an output $z \in R^{C\times H\times W}$ that contains more useful global priors, we first combine y^h, y^w, y^l and y^r as follows to get $y \in R^{C\times H\times W}$:

$$y_c = y_c^h + y_c^w + y_c^l + y_c^r \tag{7}$$

Then, calculate the output z as:

$$z = Scale\,(x,\ \delta\,(f\,(y))) \tag{8}$$

where Scale$(\cdot,\cdot)$ refers to element-wise multiplication, δ is the sigmoid function and f is a 1×1 convolution.

3 Experiments

3.1 Datasets

The datasets we use are two commonly used benchmark datasets in the field of retinal vessels segmentation: DRIVE[25] and CHASE[7]. Detailed data descriptions are shown in Table 1.

3.2 Evaluation Metrics

We employ several wide-use testing metrics for quantitative evaluation: Accuracy (Acc), Sensitivity(Sen), and Area Under Curve (AUC). The calculation definition is: $Sen = \frac{TP}{TP+FN}$, $Acc = \frac{TP+TN}{TP+TN+FP+FN}$. Where TP for true positive; FP for false positive; TN for true negative; FN for false negative.

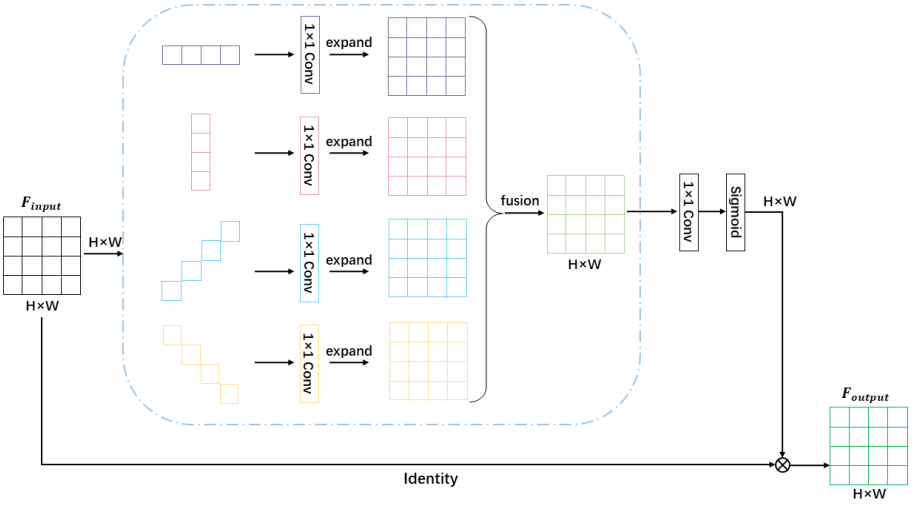

Figure 4: Schematic illustration of the Striped Pooling module(SPPM).

3.3 Implementation Details

Network Structure: The number of channels of each convolutional block in the encoder is set to 32, 64, 128, and 196, respectively, and the number of channels of convolutional blocks in the up-sampling decoder is set to 196, 128, 64, and 32, as shown in Figure 1. Finally, the final segmentation mask is obtained by 1×1 convolution and softmax.

Loss function: We use the same loss function as DA-Net [29]: binary cross-entropy loss L_{bce} and Dice loss [18] L_{dice} to construct the total loss L_{total} as follows:

$$L_{total} = L_{bce} + L_{dice} \tag{9}$$

Training: Our model is implemented based on the PyTorch framework and trained for 300 epochs on a single RTX 6000 GPU. All images are resized to 640×640 and then cropped into 160×160 ($1/4\ H \times W$) patches to feed the patches-level branch of the encoder. Random horizontal flipping and random rotation data augmentation are used to avoid overfitting, and the random probability is set to 0.5. Additionally, we use the Adam optimizer to train our model with a momentum size of 10^{-8}. The initial learning rate is set to 0.001, and the linear decay strategy is used to adjust the learning rate, the decay factor is 0.01, and the weight decay is 0.0005.

3.4 Comparison with Advanced Methods

Table 2 and 3 presents a quantitative comparison of our SRNet with state-of-the-art methods on DRIVE and CHASE datasets. From the table, we can see that our method SRNet has achieved the best *Acc*, *Sen*, and *AUC* on both benchmark datasets. Figure 5 shows the visualization results of some segmentations of our network on the two datasets. It can be clearly seen that our network can obtain better segmentation visual effects.

Table 1: Datasets details.

Dataset	Iamges	Size	Training	Testing
DRIVE	40	565×584	20	20
CHASE	28	999×960	20	8

Table 2: Comparison with other state-of-the-art methods on the DRIVE dataset. The best result is marked in **bold** and the second best result is underlined.

Method	year	Input Type	Acc	Sen	AUC
JL-UNet[34]	2018	Patches	95.56	77.92	97.84
MS-NFN[33]	2018	Patches	95.67	78.44	98.07
CE-Net[8]	2019	Image	95.45	83.09	97.79
CTF-Net[30]	2020	Patches	95.67	78.49	97.88
CGA-Net[28]	2021	Image	96.47	83.05	98.65
SCS-Net[32]	2021	Image	96.97	82.89	98.37
DA-Net[29]	2022	Joint	97.07	85.57	99.03
SRNet(our)	2023	Joint	**97.09**	**85.68**	**99.13**

Table 3: Comparison with other state-of-the-art methods on the CAHSE dataset. The best result is marked in **bold** and the second best result is underlined.

Method	year	Input Type	Acc	Sen	AUC
JL-UNet[34]	2018	Patches	96.10	76.33	97.81
MS-NFN[33]	2018	Patches	96.37	75.38	98.25
CE-Net[8]	2019	Image	96.89	81.52	98.30
CTF-Net[30]	2020	Patches	96.48	79.48	98.47
CGA-Net[28]	2021	Image	97.06	86.78	98.12
SCS-Net[32]	2021	Image	97.44	83.65	98.67
DA-Net[29]	2022	Joint	97.66	87.04	99.08
SRNet(our)	2023	Joint	**97.82**	**87.06**	**99.17**

Table 4: Ablation study on DRIVE dataset. RTM means Relational Transformer Module proposed in Sect. 2.2. SPPM means Striped Pooling Pyramid Module proposed in Sect. 2.3.

Methods	Acc	AUC	Flops	Parameters
Baseline w/image-level input	95.68	97.60	**21.5G**	**8.2M**
Baseline w/patches-level input	96.01	97.51	**21.5G**	**8.2M**
Baseline + **RTM**	96.73	98.69	21.9G	8.9M
Baseline + **SPPM**	96.65	98.58	22.3G	9.6M
Baseline + **all(our SRNet)**	**97.09**	**99.13**	23.7G	10.6M

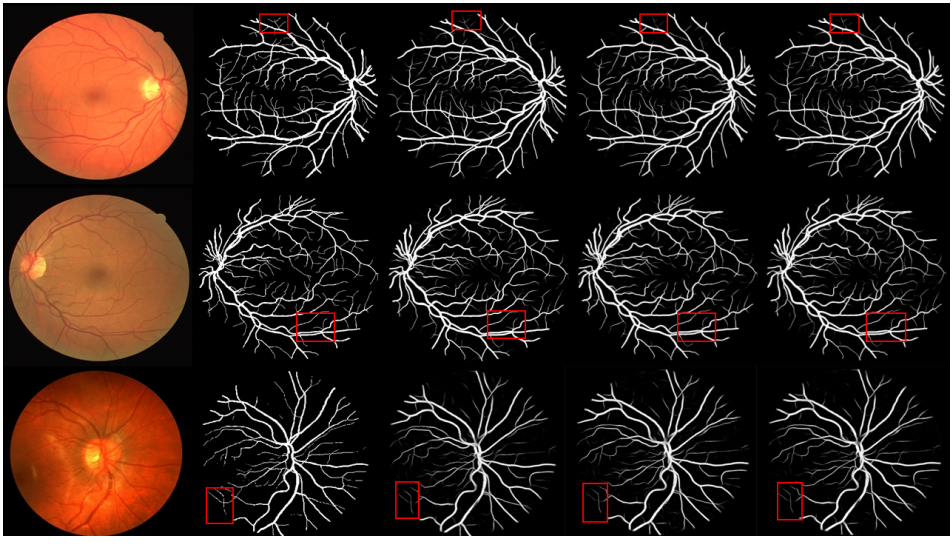

Figure 5: Visualization of segmentation results on DRIVE and CHASE datasets. The first and second lines are the DRIVE dataset, and the third line is the CHASE dataset. From Left to Right: Retina images, Ground truths, proposed **SRNet**, DA-Net[29], and CGANet [28] outputs.

3.5 Ablation Study

Table 4 shows the results of our network ablation experiments on the DRIVE dataset. We use the classic U-Net [21] as a baseline and train the models separately with the two input forms mentioned above. The results show that our RTM and SPPM can significantly improve the scores of *Acc* and *AUC*. Finally, we combine all the proposed components to achieve the best segmentation performance. To investigate the additional cost brought by the proposed components, we also report the Flops and Parameters of each variant in the ablation study. Our SRNet achieves better results with only marginal increases in memory and computation consumption.

4 Conclusion

We propose a retinal vessel segmentation approach SRNet, in which image-level contextual information is introduced to local patches via a well-designed Relational Transformer Module. In addition, we design a Striped Pooling Pyramid Module with diagonal lines according to the vessel distribution features at the bottom of the encoder, which can effectively capture the contextual information that fits the tree-like morphological distribution of retinal vessels. Our SRNet significantly outperforms other state-of-the-art retinal vessel segmentation methods on the DRIVE and CHASE datasets, and can potentially be applied to other high-resolution or strip object and lesion segmentation tasks.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (No.61962054), the Cultivation plan of major Scientific Research Projects of Northwest Normal University (No.NWNU-LKZD2021-06). the National Natural Science Foundation of China (No.61163036).

References

[1] Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M Taha, and Vijayan K Asari. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*, 2018.

[2] Peter Bankhead, C Norman Scholfield, J Graham McGeown, and Tim M Curtis. Fast retinal vessel detection and measurement using wavelets and edge location refinement. *PloS one*, 7(3):e32435, 2012.

[3] Qi Bi, Shuang Yu, Wei Ji, Cheng Bian, Lijun Gong, Hanruo Liu, Kai Ma, and Yefeng Zheng. Local-global dual perception based deep multiple instance learning for retinal disease classification. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*, pages 55–64. Springer, 2021.

[4] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III*, pages 205–218. Springer, 2023.

[5] Erkang Cheng, Liang Du, Yi Wu, Ying J Zhu, Vasileios Megalooikonomou, and Haibin Ling. Discriminative vessel segmentation in retinal images by fusing context-aware hybrid features. *Machine vision and applications*, 25:1779–1792, 2014.

[6] Donald S Fong, Lloyd Aiello, Thomas W Gardner, George L King, George Blankenship, Jerry D Cavallerano, Fredrick L Ferris III, Ronald Klein, and American Diabetes Association. Retinopathy in diabetes. *Diabetes care*, 27(suppl_1):s84–s87, 2004.

[7] Muhammad Moazam Fraz, Paolo Remagnino, Andreas Hoppe, Bunyarit Uyyanonvara, Alicja R Rudnicka, Christopher G Owen, and Sarah A Barman. An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Transactions on Biomedical Engineering*, 59(9):2538–2548, 2012.

[8] Zaiwang Gu, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, and Jiang Liu. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE transactions on medical imaging*, 38(10):2281–2292, 2019.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[10] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[11] Shiqi Huang, Jianan Li, Yuze Xiao, Ning Shen, and Tingfa Xu. Rtnet: relation transformer network for diabetic retinopathy multi-lesion segmentation. *IEEE Transactions on Medical Imaging*, 41(6):1596–1607, 2022.

[12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[13] Xiaomeng Li, Xiaowei Hu, Lequan Yu, Lei Zhu, Chi-Wing Fu, and Pheng-Ann Heng. Canet: cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading. *IEEE transactions on medical imaging*, 39(5):1483–1493, 2019.

[14] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[15] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[16] Yuhui Ma, Huaying Hao, Jianyang Xie, Huazhu Fu, Jiong Zhang, Jianlong Yang, Zhen Wang, Jiang Liu, Yalin Zheng, and Yitian Zhao. Rose: a retinal oct-angiography vessel segmentation dataset and new model. *IEEE transactions on medical imaging*, 40(3): 928–939, 2020.

[17] Kevis-Kokitsi Maninis, Jordi Pont-Tuset, Pablo Arbeláez, and Luc Van Gool. Deep retinal image understanding. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 140–148. Springer, 2016.

[18] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. Ieee, 2016.

[19] José Ignacio Orlando, Elena Prokofyeva, and Matthew B Blaschko. A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. *IEEE transactions on Biomedical Engineering*, 64(1):16–27, 2016.

[20] Elisa Ricci and Renzo Perfetti. Retinal blood vessel segmentation using line operators and support vector classification. *IEEE transactions on medical imaging*, 26(10):1357–1365, 2007.

[21] Olaf Ronneberger, Philipp Fischer, and Thomas 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*, pages 234–241. Springer, 2015.

[22] Mehdi Sadeghibakhi, Hamidreza Pourreza, and Hamidreza Mahyar. Multiple sclerosis lesions segmentation using attention-based cnns in flair images. *IEEE Journal of Translational Engineering in Health and Medicine*, 10:1–11, 2022.

[23] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[24] Thomas J Smart, Christopher J Richards, Rhythm Bhatnagar, Carlos Pavesio, Rupesh Agrawal, and Philip H Jones. A study of red blood cell deformability in diabetic retinopathy using optical tweezers. In *Optical trapping and optical micromanipulation XII*, volume 9548, pages 342–348. SPIE, 2015.

[25] Joes Staal, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram Van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging*, 23(4):501–509, 2004.

[26] Bo Wang, Shuang Qiu, and Huiguang He. Dual encoding u-net for retinal vessel segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22*, pages 84–92. Springer, 2019.

[27] Changwei Wang, Rongtao Xu, Shibiao Xu, Weiliang Meng, Jun Xiao, and Xiaopeng Zhang. Accurate lung nodules segmentation with detailed representation transfer and soft mask supervision. *arXiv preprint arXiv:2007.14556*, 2020.

[28] Changwei Wang, Rongtao Xu, Yuyang Zhang, Shibiao Xu, and Xiaopeng Zhang. Retinal vessel segmentation via context guide attention net with joint hard sample mining strategy. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1319–1323. IEEE, 2021.

[29] Changwei Wang, Rongtao Xu, Shibiao Xu, Weiliang Meng, and Xiaopeng Zhang. Danet: Dual branch transformer and adaptive strip upsampling for retinal vessels segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part II*, pages 528–538. Springer, 2022.

[30] Kun Wang, Xiaohong Zhang, Sheng Huang, Qiuli Wang, and Feiyu Chen. Ctf-net: Retinal vessel segmentation via deep coarse-to-fine supervision network. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1237–1241. IEEE, 2020.

[31] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

[32] Huisi Wu, Wei Wang, Jiafu Zhong, Baiying Lei, Zhenkun Wen, and Jing Qin. Scs-net: A scale and context sensitive network for retinal vessel segmentation. *Medical Image Analysis*, 70:102025, 2021.

[33] Yicheng Wu, Yong Xia, Yang Song, Yanning Zhang, and Weidong Cai. Multiscale network followed network model for retinal vessel segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11*, pages 119–126. Springer, 2018.

[34] Zengqiang Yan, Xin Yang, and Kwang-Ting Cheng. Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation. *IEEE Transactions on Biomedical Engineering*, 65(9):1912–1923, 2018.

[35] Bob Zhang, Lin Zhang, Lei Zhang, and Fakhri Karray. Retinal vessel extraction by matched filter with first-order derivative of gaussian. *Computers in biology and medicine*, 40(4):438–445, 2010.

[36] Yuan Zhang, Miao He, Zhineng Chen, Kai Hu, Xuanya Li, and Xieping Gao. Bridge-net: Context-involved u-net with patch-based loss weight mapping for retinal blood vessel segmentation. *Expert Systems with Applications*, 195:116526, 2022.

[37] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pages 3–11. Springer, 2018.