RIP-AV: Joint Representative Instance Pre-training with Context Aware Network for Retinal Artery/Vein Segmentation

Wei Dai[1], Yinghao Yao[2], Hengte Kong[1,2], Zhen Ji Chen[1], Sheng Wang[1,2], Qingshi Bai[1,2], Haojun Sun[1,2], Yongxin Yang[3] and Jianzhong Su[1,2]

[1]National Engineering Research Center of Ophthalmology and Optometry, Eye Hospital, Wenzhou Medical University, Wenzhou, 325027, Zhejiang, China
[2]Oujiang Laboratory (Zhejiang Lab for Regenerative Medicine, Vision and Brain Health), Eye Hospital, Wenzhou Medical University, Wenzhou, 325011, Zhejiang, China
[3]School of Electronic Engineering and Computer Science, Queen Mary University of London, London, E1 4NS, UK
sujz@wmu.edu.cn

Abstract. Accurate deep learning-based segmentation of retinal arteries and veins (A/V) enables improved diagnosis, monitoring, and management of ocular fundus diseases and systemic diseases. However, existing resized and patch-based algorithms face challenges with redundancy, overlooking thin vessels, and underperforming in low-contrast edge areas of the retinal images, due to imbalanced background-to-A/V ratios and limited contexts. Here, we have developed a novel deep learning framework for retinal A/V segmentation, named RIP-AV, which integrates a Representative Instance Pre-training (RIP) task with a context-aware network for retinal A/V segmentation for the first time. Initially, we develop a direct yet effective algorithm for vascular patch-pair selection (PPS) and then introduce a RIP task, formulated as a multi-label problem, aiming at enhancing the network's capability to learn latent arteriovenous features from diverse spatial locations across vascular patches. Subsequently, in the training phase, we introduce two novel modules: Patch Context Fusion (PCF) module and Distance Aware (DA) module. They are designed to improve the discriminability and continuity of thin vessels, especially in low-contrast edge areas, by leveraging the relationship between vascular patches and their surrounding contexts cooperatively and complementarily. The effectiveness of RIP-AV has been validated on three publicly available retinal datasets: AV-DRIVE, LES-AV, and HRF, demonstrating remarkable accuracies of 0.970, 0.967, and 0.981, respectively, thereby outperforming existing state-of-the-art methods. Notably, our method achieves a significant 1.7% improvement in accuracy on the HRF dataset, particularly enhancing the segmentation of thin edge arteries and veins.

Keywords: Retinal Artery/Vein Segmentation, Pre-training, Context Aware

1 Introduction

The morphological changes of retinal vessels have been well-documented as significant biomarkers for various ocular diseases and systemic diseases in vivo [1-3]. Color fundus photography, a non-invasive imaging modality, affords direct inspection of the intricate retinal vasculature, encompassing both arteries and veins [4]. The identification of venous dilation and decreased arteriovenous width ratio associated with microaneurysms in the edge regions play a role in the early diagnosis of diabetic retinopathy [1]. Furthermore, the vascular alteration in hypertensive retinopathy (HR) is characterized by the narrowing of retinal arteries and increased tortuosity [2]. Hence, precise segmentation of retinal arteries and veins (A/V), followed by quantitative analysis, can serve as valuable clinical assistance in disease diagnosis and management, further preventing the risks of visual impairment and systemic complications in their early phase.

The recent development of deep learning (DL) has given rise to remarkable progress of automatic retinal A/V segmentation [5-14, 31-33]. Generally, current DL methods process high-resolution retinal images through encoder-decoder networks, such as U-Net [15], generative adversarial network (GAN [16]), and fully convolutional network (FCN [17]), to achieve pixel-level A/V segmentation. These methods can be categorized into resize-based method [5-8] and patch-based method [9-14], according to whether the image-level or patch-level input is processed or not. The resize-based method mainly scales the full image to a smaller dimension and then feeds it into network, however, overlooking details of thin vessels due to image compression. A patch-based method is thus introduced to preserve all details by primarily partitioning full images into multiple small patches and randomly selecting some for in-depth analysis, achieving notable A/V segmentation performance. Whereas, redundancy and context are the main limitations of this method, which can cause overfitting and arteriovenous confusion in edge regions.

To address these limitations, we propose a novel deep learning framework, RIP-AV, which joints a Representative Instance Pre-training (RIP) task with a context-aware network for the retinal A/V segmentation. This framework begins with a novel patch-pair selection (PPS) algorithm to pick vascular patches and their contextual counterparts from retinal images, thereby capturing representative instances while reducing redundancy. The RIP component of our framework is formulated as a multi-label classification problem, aimed at bolstering the network's proficiency in identifying latent arteriovenous features across various spatial regions within the picked vascular patches. To further refine the segmentation process, we introduce two innovative modules: the Patch Context Fusion (PCF) module and the Distance Aware (DA) module. They are integrated into the network, enhancing the contextual interplay between patches, and enriching the perception of vascular edges. The comparison experiment demonstrates a superior performance of RIP-AV when benchmarked against the state-of-the-art methods, across three publicly available retinal datasets—AV-DRIVE, LES-AV, and HRF.

2 Method

2.1 Framework Overview

The proposed RIP-AV architecture, as shown in Fig. 1, is structured into two main stages: RIP and subsequent model training. The RIP stage utilizes a PPS algorithm (see Sect. 2.2) to extract vascular patches at multiple scales (64, 96, 128, and 256) and their respective contexts from a given dataset $\mathbb{X} = \{x_1, ... x_n\}$ with their corresponding pixel-level ground truth $\mathbb{Y} = \{y_1 ... y_n\}$. This stage is dedicated to learning embedding arteriovenous features among the vascular patches $P = \{P_1, ... P_m\}$.

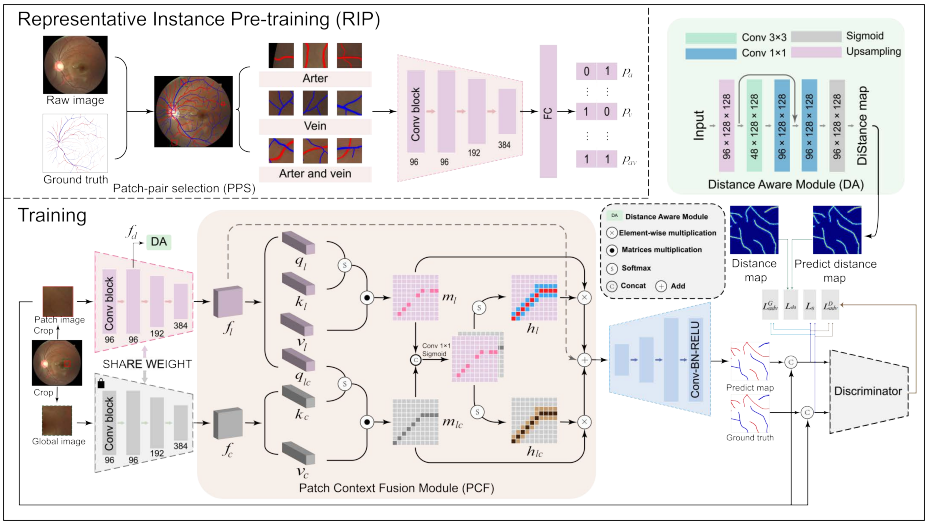

Fig. 1. Overview of our proposed framework.

Building on this pre-trained foundation, the final segmentation network is trained, which incorporates both the DA module (see Sect. 2.3) and the PCF module (see Sect. 2.4) to assimilate and apply the relationship between vascular patches and their contextual information.

In the training workflow, each vascular patch P_i, and its contexts C_i are resized to 256 and input into the dual-stream encoder E. The E, consisting of the first three convolutional blocks of ConvNeXt [18], processes these inputs to extract shallow and high-level feature maps, which are then sequentially processed by the DA and PCF modules. The decoder, employing standard convolutional-batch normalization-ReLU (Conv-BN-ReLU) structures, outputs a three-channel prediction map $G(P_i)$ for artery, vein, and overall vessel segmentation. This prediction map $G(P_i)$, along with y_i, and the patch P_i, are passed into the Discriminator D, which functions to evaluate the differences between the predicted and ground truth maps, ensuring quality of the segmentation. The discriminator is based on the standard PatchGAN [19] architecture, which provides a robust assessment of local image patches.

2.2 Representative Instance Pre-training

The primary target of the RIP is the recognition of the embedding arteriovenous features within vascular patches, pivotal for subsequent segmentation tasks. The PPS algorithm, a critical pre-processing step for RIP, is designed to select patches representative of vascular structures. We observe that all randomly cropped patches fall into four categories: P_a (artery-only), P_v (vein-only), P_{av} (containing both arteries and veins), and P_b (none of vascular structures). The PPS is crafted to exclude P_b, thereby enhancing the specificity of the training set, as described in Algorithm 1. It ensures the inclusion of vascular features by leveraging the skeleton line — a representation of the vascular core structure — thus guaranteeing that selected patches are imbued with relevant vascular information.

Following that, the RIP phase then utilizes these categorized patches to update the neural network, facilitating the learning of a feature space that captures the essence of arteriovenous differentiation across diverse spatial areas. During this phase, the RIP is formulated as a multi-label classification task, with binary vectors representing the presence of arterial (P_a: [1,0]) and venous (P_v: [0,1]) feature, as well as their co-occurrence P_{av} ([1,1]). The binary encoding serves as a guidance for the network to distinguish between arterial and venous features, laying the groundwork for subsequent A/V segmentation tasks.

Algorithm 1: Patch-pair Selection (PPS)

Input: image: x, label: y, patch size: p, and type of local patch: P_t.
Output: local patch: P, and contextual patch: C
initialize $P, C \leftarrow 0$
$s \leftarrow y.getSkeleton$ /* generate skeleton line for y */
while not find proper P **do**
  random select a point $s^{(i,j)}$ from s
  $C \leftarrow x[i-p{:}\,i+p, j-p{:}\,j+p]$ /* get contextual patch C */
  random select a point $C^{(m,n)}$ from C
  $P \leftarrow C[m{:}\,m+p,\ n{:}\,n+p]$
  if $P == P_t$ and $P.countNonZero \geq \alpha$ and $P \subseteq C$ **then**
    break /* find proper P */
  end if
end while
return P, C

2.3 Distance Aware Module

The capability to accurately delineate edges is fundamental to the analysis of vascular structures [20]. Taking inspiration from the innate capacity of convolutional neural networks (CNNs) to capture edge details at shallow layers, here, we design the DA module, consisting of a sequence of convolutional layers, to enhance the perception of the edges of blood vessel, by refining shallow-level feature maps of E. As illustrated in the top right of Fig. 1, the feature map f_d , derived from the second layer of E, is first

processed by the upsampling layer to enlarge resolution, Subsequently, the 3 × 3 and the 1 × 1 convolutional layer are sequentially used to encode and refine spatial relationships essential to vascular structures. Finally, the refined feature map is ultimately transformed by a sigmoid activation function into a probabilistic distance map $DA(f_d)$. To optimize the parameters of the DA module, we employ a distance-aware loss function $\mathcal{L}_{da}$ like [13], aiming to minimize gaps between the predicted distance maps and the ground truth. $\mathcal{L}_{da}$ is depicted in the following equation:

$$\begin{aligned} \mathcal{L}_{da}(DA(f_d), S(y)) &= \frac{1}{\sum_{i=1}^{N} y} \sum_{j=1}^{N} \frac{1}{S(y_j)+\varepsilon} \cdot smooth_{L_1}(DA(f_{dj}), S(y_j)) \\ smooth_{L_1}(x, y) &= \begin{cases} 0.5(x-y)^2, & \text{if } |x-y| < 0 \\ |x-y| - 0.5, & \text{otherwise} \end{cases} \end{aligned} \tag{1}$$

where N denotes the batch number; $\mathcal{E}$ is a small value to avoid the division by zero; $S(y_j)$ denotes the jth distance map which is transferred from the ground truth y_j via Euclidean distance transformation [21]. $smooth_{L_1}$ serves as a soft version of conventional L1 distance.

2.4 Patch Context Fusion Module

Motivated by previous studies [22, 23, 34], which have shown that the significance of integrating context information for a more comprehensive understanding of spatial relationships within images, we propose a novel PCF module. It is designed to leverage the relationship between local patches and surrounding context, by applying self-attention and cross-attention jointly.

As shown in Fig. 1, the PCF comprises two main paths that process local patch feature map $f_l \in \mathbb{R}^{c*h*w}$ and its contextual feature map $f_c \in \mathbb{R}^{c*h*w}$, which are then projected into queries (q_l, q_{lc}), keys (k_l, k_c) and values (v_l, v_c) representations. To underscore the features within patches, a standard self-attention mechanism [24] is adopted get the attention map $m_l \in \mathbb{R}^{c*h*w}$ with q_l, k_l, and v_l. Simultaneously, cross-attention mechanism [24] is introduced, using q_{lc}, k_c, and v_c, to generate contextual attention map $m_{lc} \in \mathbb{R}^{c*h*w}$ that reveals the relationship between patch and its context. Then, the weight map, which is significant in determining the influence of each attention map, is computed by applying a sequence of operations. Concretely, the weight maps in the PCF are generated by first concatenating the attention maps. This concatenated map is then refined through a 1 × 1 convolution, normalized by a sigmoid function, and finally, importance probabilities are calculated by a softmax function, parametrized by the weight maps h_l, h_{lc}. The final output of PCF is computed by merging attention maps with their corresponding weight maps and then adding f_l as a residual, aiming to integrate contextual information while preserving the essential features of the local feature map.

3 Experiments and Results

3.1 Datasets and Evaluation Metrics

We conduct our experiments on three public retinal datasets: AV-DRIVE [25], LES-AV [26], and HRF [27]. The DRIVE-AV dataset comprises 40 retinal images, each with a resolution of 565 × 584 pixels, accompanied by pixel-wise ground truth. 20 images are used for training and the remaining 20 for testing. The LES-AV dataset includes 21 images with the resolution of 1620 × 1444 and one image of 2196 × 1958, with a division of 11 images for training and the remaining 11 for testing. The HRF dataset contains 45 images at 3504 × 2336 pixels. 30 images are used for training and 15 for testing.

As for evaluation metrics, we assess the performance of proposed method using three principal metrics, consistent with [12, 13]: accuracy (ACC), sensitivity (SN), and specificity (SP). SN is defined as the proportion of true positives (TP) over the total number of actual positives (TP + false negatives (FN)), reflecting the model's capability to correctly identify arteries. Specificity (SP) measures the proportion of true negatives (TN) over the total number of actual negatives (TN + false positives (FP)), indicating the model's accuracy in recognizing veins. Accuracy (ACC) represents the overall correctness of the model, calculated as the sum of TP and TN over the total case count (TP + TN + FP + FN).

3.2 Implementation Details

In this work, all experiments are conducted using the PyTorch [28] framework on a NVIDIA RTX 4090 GPU with 24GB of memory. During the selection of multi-scale patches by PPS, it is ensured that the overlap between same scale patches does not exceed 80%. In the training phase, we utilize PPS at different scales (AV-DRIVE [64, 128, 256], LES-AV and HRF [96, 128, 256]) to extract multi-scale patches and their context. Then, data augmentations are performed using random flipping, affine transformation, rotation and shift to avoid overfitting. For testing, sliding windows with overlaps are used to extract patches, with each patch's context being the entire image. Results from these patches are integrated using strategies same as [11-14]. Adam [29] optimizer is adopted to refine the model parameters, configuring β_1 to 0.5 and β_2 to 0.999. Batch size is set to 8. The initial learning rate is 0.0002, reduced by 50% every 7,000 epochs, up to a limit of 30,000 epochs. The parameters of the segmentation network are optimized using a combination of loss functions: the distance-aware loss $\mathcal{L}_{da}$, the adversarial loss $\mathcal{L}_{adv}^{D}$, $\mathcal{L}_{adv}^{G}$, and the segmentation loss $\mathcal{L}_{s}$, which is the sum of combination of classic binary cross entropy loss and Dice loss between the prediction and ground truth. Further details and source code are available at https://github.com/wei-dai00/RIP-AV.

3.3 Quantitative and Qualitative Comparison with Advanced Methods

In this section, we compare the performance of our proposed RIP-AV with resize-based methods [5-7] and patch-based methods [11-14]. The comparison is conducted both qualitatively and quantitatively across three datasets: AV-DRIVE, LES-AV, and HRF. The quantitative comparison results, as presented in Table 1, demonstrate that our model has consistently outperformed resize-based and patch-based approaches across three datasets, showing the best overall performance. Specifically, the RIP-AV method achieves outstanding accuracies of 0.970 ($\pm$ 0.02, $P > 0.05$) on AV-DRIVE, 0.967 ($\pm$ 0.01, $P > 0.05$) on LES-AV, and 0.981 ($\pm$ 0.01, $P > 0.05$) on HRF. Notably, on the HRF dataset with high density of small vessels and lesions, RIP-AV has shown at least a 1.7% increase in term of accuracy over the other methods. In Fig. 2, the visualization results of qualitative comparison illustrate the superior performance of proposed method, particularly within the detailed regions marked by yellow and green dashed boxes. Here, RIP-AV demonstrates superior accuracy in capturing fine edge arteries and veins, as well as maintaining their continuity, compared to other state-of-the-art methods.

Table 1. The performance comparison with state-of-art methods on the AV-DRIVE, LES-AV and HRF datasets. Best performances are highlighted in **bold**.

	AV-DRIVE			LES-AV			HRF		
Method	SN	SP	ACC	SN	SP	ACC	SN	SP	ACC
Resize-based									
Galdran [5]	0.940	0.930	0.900	-	-	-	0.850	0.910	0.910
Galdran [6]	0.902	0.963	0.936	0.852	0.959	0.914	0.942	0.928	0.935
Li [7]	0.956	0.905	0.928	**0.982**	0.940	0.959	0.972	0.895	0.930
Patch-based									
Chen [11]	0.945	0.963	0.955	-	-	-	0.968	0.945	0.955
Hu [12]	0.936	0.974	0.955	0.943	0.947	0.945	0.959	0.970	0.964
Chen [13]	0.952	0.963	0.958	-	-	-	0.97	0.961	0.965
Luo [14]	**0.968**	0.972	**0.970**	0.961	0.941	0.951	-	-	-
Patch-pair selection (our)									
Our	0.962	**0.976**	**0.970**	0.963	**0.971**	**0.967**	**0.980**	**0.982**	**0.981**

Table 2. Ablation study on AV-DRIVE dataset. Best performances are highlighted in **bold**.

w/o DA	**w/o RIP**	**w/o PCF**	**SN**	**SP**	**ACC**
×	×	×	0.937	0.958	0.948
√	×	×	0.945	0.957	0.951
√	√	×	0.953	0.967	0.961
√	√	√	**0.962**	**0.976**	**0.970**

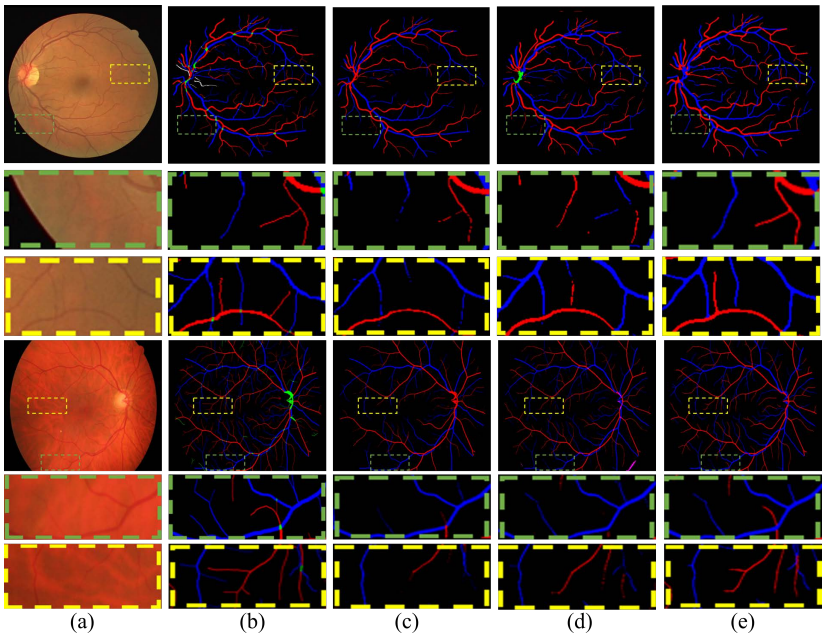

Fig. 2. Visualization of the results of retinal A/V segmentation. From left to right: (a) Raw image, (b) Ground truth, (c) Resize-based method [7], (d) Patch-based method [13] (e) our proposed RIP-AV.

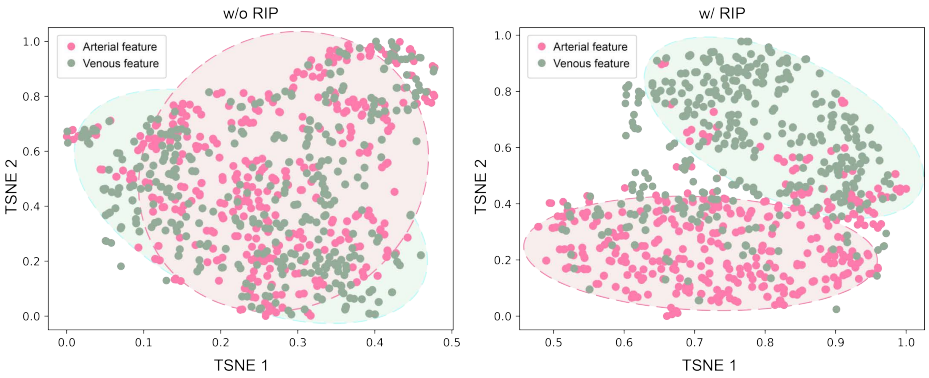

Fig. 3. Visualization of embedding features for w/o RIP in the ablation study. Pink and green dots denote arterial and venous embedding features, respectively.

3.4 Ablation Analysis

To investigate the contribution of each component, we perform an ablation study on AV-DRIVE dataset. Starting with a baseline U-Net-like GAN, we sequentially integrated a DA module, followed by the RIP and the PCF module, with results detailed in Table 2. The DA module improved accuracy to 0.951 by enhancing the perception of

vascular edge, while RIP provided a further boost to 0.961 by pre-training the network on arterial and venous features, as visually supported by t-SNE [30] in Fig. 3. The final incorporation of PCF achieved the top accuracy of 0.970, underscoring its effectiveness in integrating contextual information for superior A/V segmentation.

4 Conclusion

In this paper, we propose a novel deep learning framework, named RIP-AV, that is the first deep learning framework to combine the concepts of Representative Instance Pre-training (RIP) with the contextual analysis of local patches for retinal A/V segmentation. The comparison experiment on three public datasets shows the effectiveness of our method over current state-of-the-art methods. Furthermore, a comprehensive ablation study indicates the effectiveness and logic behind the proposed framework. In the future work, we will introduce more customed metrics to evaluate the performance of models on condition that vessels with varying widths. We believe that our contributions will foster further investigations into the correlations between retinal A/V metrics and ocular as well as cardiovascular health conditions, potentially leading to breakthroughs in predictive diagnostics and treatment strategies.

Acknowledgments. This study was funded by the National Natural Science Foundation of China (82172882) and supported by Biomedical Big Data Intelligent Computing Center of Oujiang Lab.

Disclosure of Interests. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the study.

References

1. Safi, H., Safi, S., Hafezi-Moghadam, A., Ahmadieh, H.: Early detection of diabetic retinopathy. Surv. Ophthalmol. **63**(5), 601–608 (2018)
2. Tapp, R.J., et al.: Associations of Retinal Microvascular Diameters and Tortuosity With Blood Pressure and Arterial Stiffness: United Kingdom Biobank. Hypertension **74**(6), 1383–1390 (2019)
3. Zekavat, S.M., et al.: Deep Learning of the Retina Enables Phenome- and Genome-Wide Analyses of the Microvasculature. Circulation **145**(2), 134–150 (2022)
4. Fogel-Levin, M., et al.: Advanced retinal imaging and applications for clinical practice: A consensus review. Surv. Ophthalmol. **67**(5), 1373–1390 (2022)
5. Galdran, A., Meyer, M., Costa, P., Campilho, A.: Uncertainty-aware artery/vein classification on retinal images. In: ISBI, pp. 556–560 (2019)
6. Galdran, A., Anjos, A., Dolz, J., Chakor, H., Lombaert, H., Ben Ayed, I.: The Little W-Net That Could: State-of-the-Art Retinal Vessel Segmentation with Minimalistic Models. arXiv preprint arXiv:2009.01907 (2020)
7. Li, L., Verma, M., Nakashima, Y., Kawasaki, R., Nagahara, H.: Joint learning of vessel segmentation and artery/vein classification with post-processing. In: Medical Imaging with Deep Learning, pp. 440–453 (2020)

8. Zhou, Y., et al.: Learning to Address Intra-segment Misclassification in Retinal Imaging. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (eds.) MICCAI 2021. LNCS, vol 12901, pp. 482–492. Springer, Cham (2021)
9. Girard, F., Kavalec, C., Cheriet, F.:Joint segmentation and classification of retinal arteries/veins from fundus images. Artif. Intell. Med. **94**, 96–109 (2019)
10. Ma, W., Yu, S., Ma, K., Wang, J., Ding, X., Zheng, Y.: Multi-task Neural Networks with Spatial Activation for Retinal Vessel Segmentation and Artery/Vein Classification. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A. (eds.) MICCAI 2019. LNCS, vol 11764, pp. 769–778. Springer, Cham (2019)
11. Chen, W., et al.: TR-GAN: Topology Ranking GAN with Triplet Loss for Retinal Artery/Vein Classification. In: Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., Joskowicz, L. (eds.) MICCAI 2020, LNCS, vol 12265, pp. 616–625. Springer, Cham (2020)
12. Hu, J., et al.: Automatic Artery/Vein Classification Using a Vessel-Constraint Network for Multicenter Fundus Images. Front. Cell Dev. Biol. **9**, 659941 (2021)
13. Chen, W., et al.: TW-GAN: Topology and width aware GAN for retinal artery/vein classification. Med. Image Anal. **77**, 102340 (2022)
14. Luo, S., Heng, Z., Pagnucco, M., Song, Y.: Two-Stage Topological Refinement Network for Retinal Artery/Vein Classification. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), pp. 1–4. (2022)
15. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015, LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015)
16. Goodfellow, I., et al.: Generative adversarial nets. In: NeurIPS, vol. **27** (2014)
17. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: CVPR, pp. 3431-3440 (2015)
18. Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T., Xie, S.: A ConvNet for the 2020s. In: CVPR, pp. 11966–11976 (2022)
19. Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: CVPR, pp. 5967–5976 (2017)
20. Qu, Z., Zhuo, L., Cao, J., Li, X., Yin, H., Wang, Z.: TP-Net: Two-Path Network for Retinal Vessel Segmentation. JBHI **27**(4), 1979–1990 (2023)
21. Maurer J., C.R., Qi, R., Raghavan, V.: A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions. TPAMI **25**(2), 265–270 (2003)
22. Zhang, S., Song, L., Gao, C., Sang, N.: GLNet: Global Local Network for Weakly Supervised Action Localization. IEEE Trans. on Mult. **22**(10), 2610–2622 (2020)
23. Li, Q., Yang, W., Liu, W., Yu, Y., He, S.: From Contexts to Locality: Ultra-high Resolution Image Segmentation via Locality-aware Contextual Correlation. In: ICCV, pp. 7232–7241 (2021)
24. Vaswani, A., et al.: Attention is all you need. In: NeurIPS, pp. 5998–6008 (2017)
25. Hu, Q., Abràmoff, M.D., Garvin, M.K.: Automated Separation of Binary Overlapping Trees in Low-Contrast Color Retinal Images. In: MICCAI 2013, LNCS, vol. 8150, pp. 436–443. Springer, Berlin, Heidelberg (2013)
26. Odstrcilik, J., et al.: Retinal vessel segmentation by improved matched filtering: Evaluation on a new high-resolution fundus image database. IET Image Pro. **7**(4), 373–383 (2013)
27. Orlando, J.I., Barbosa Breda, J., van Keer, K., Blaschko, M.B., Blanco, P.J., Bulant, C.A.: Towards a Glaucoma Risk Index Based on Simulated Hemodynamics from Fundus Images. In: MICCAI 2018, LNCS, vol.11071, pp. 65–73. Springer, Cham (2018)

28. Paszke, A., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32** (2019)
29. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
30. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. (2008)
31. Morano, José, et al.: Simultaneous segmentation and classification of the retinal arteries and veins from color fundus images. Artificial Intelligence in Medicine **118** (2021)
32. Karlsson, R.A., and Sveinn H.H.: Artery vein classification in fundus images using serially connected U-Nets. Computer Methods and Programs in Biomedicine **216** (2022)
33. Zhao, A.D., et al.: Optimization of retinal artery/vein classification based on vascular topology. Biomedical Signal Processing and Control **88** (2024)
34. Wang, C., Xu, R., Xu, S., Meng, W., Zhang, X.: DA-Net: Dual Branch Transformer and Adaptive Strip Upsampling for Retinal Vessels Segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol 13432, pp. 528–538. Springer, Cham (2022).