BILATERAL-VIT FOR ROBUST FOVEA LOCALIZATION

anonymous

ABSTRACT

The fovea is an important anatomical landmark of the retina. Detecting the location of the fovea is essential for the analysis of many retinal diseases. However, robust fovea localization remains a challenging problem, as the fovea region often appears fuzzy, and retina diseases may further obscure its appearance. This paper proposes a novel vision transformer (ViT) approach that integrates information both inside and outside the fovea region to achieve robust fovea localization. Our proposed network named Bilateral-Vision-Transformer (Bilateral-ViT) consists of two network branches: a transformer-based main network branch for integrating global context across the entire fundus image and a vessel branch for explicitly incorporating the structure of blood vessels. The encoded features from both network branches are subsequently merged with a customized multi-scale feature fusion (MFF) module. Our comprehensive experiments demonstrate that the proposed approach is significantly more robust for diseased images and establishes the new state of the arts on both `Messidor` and `PALM` datasets.

***Index Terms*—** fovea localization, vision transformer, bilateral neural network, feature fusion

1. INTRODUCTION

The macula is the central region of the retina. The fovea is an important anatomical landmark located in the center of the macula, responsible for the most crucial part of a person's vision [1]. The severity of vision loss due to retinal diseases is usually related to the distance between the associated lesions and fovea. Therefore, detecting the location of fovea is essential for the analysis of many retinal diseases.

Despite its importance, robust fovea localization remains a challenging problem. The color contrast between the fovea region and its surrounding tissue is poor, leading to a fuzzy appearance. Furthermore, the fovea appearance may be obscured by lesions in the diseased retina; for example, geographic atrophy and hemorrhages significantly alter the fovea appearance. Such issues make it more difficult to perform localization based on the fovea appearance alone. Fortunately, anatomical structures outside the fovea region, such as blood vessels, are also helpful for localization [2, 3]. For this reason, we propose a novel vision transformer (ViT) approach that integrates information both inside and outside the fovea region to achieve robust fovea localization.

Our proposed network, named Bilateral-Vision-Transformer (Bilateral-ViT), consists of two network branches. We adopt a transformer-based U-net architecture [4] as the **main branch** for effectively integrating global context across the entire fundus image. In addition, we design a **vessel branch** that takes in a blood vessel segmentation map for explicitly incorporating the structure of blood vessels. Finally, the encoded features from both network branches are merged with a customized multi-scale feature fusion (MFF) module, leading to significantly improved performance. Thus, our key contributions are as follows:

- We propose a novel vision-transformer-based network architecture that explicitly incorporates global image context and structure of blood vessels for robust foveal localization.
- We demonstrate that the proposed approach is significantly more robust for challenging settings such as fovea localization in diseased retinas (over 9% improvements for specific evaluations). It also has a better generalization capability compared to the baseline methods, as shown in cross-dataset experiments.
- We establish the new state of the arts on both `Messidor` and `PALM` datasets.

2. RELATED WORK

Before convolutional neural networks (CNNs) have gained popularity in medical image analysis applications, researchers usually utilize hand-craft features for fovea localization. Most works use anatomical relationships among optic discs (OD), blood vessels, and fovea regions. Deka *et al.* [5] and Medhi *et al.* [6] generate the region of interest (ROI) using processed blood vessels for macula estimation. Certain methods utilize OD in the prediction of ROI and fovea center by selecting specific OD diameters [7], estimating OD orientations and minimum intensity values [8, 9]. Other applications use combined OD and blood vessels features to improve the performance of fovea localization [2, 3]. These methods generally perform less competitively than more recent deep-learning-based approaches.

Most deep learning-based methods formulate the fovea localization as a regression task [10, 11, 12, 13]. Some methods utilize retinal structures, such as OD and blood vessels, as constraints for inferring the location of fovea. For example, Meyer *et al.* [11] adopt a pixel-wise distance regression approach for joint OD and fovea localization. Besides the regression-based approaches, Sedai *et al.* [14] propose a two-stage image segmentation framework for segmenting the image region around the fovea. Unlike all previous works,

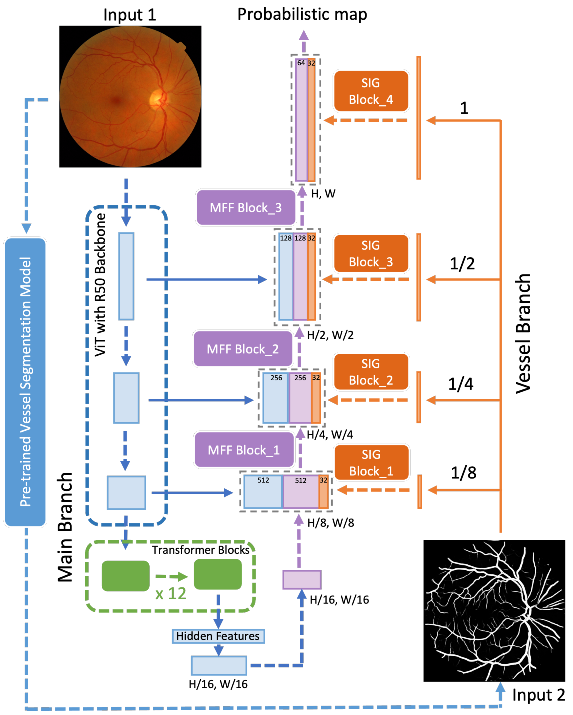

Fig. 1. The overall architecture of our proposed Bilateral-ViT network.

we customize the recent transformer-based segmentation network to incorporate blood vessel information and demonstrate its superior performance compared to the existing approaches.

3. METHODOLOGY

3.1. Network Architecture

The overall architecture of Bilateral-ViT is illustrated in Fig. 1. The proposed Bilateral-ViT is based on a U-shape architecture with a vision transformer-based encoder (**the main branch**) for exploiting long-range contexts. In addition, we design a **vessel branch** to encode structure information from blood vessel segmentation maps. Finally, Multi-scale Feature Fusion (MFF) blocks are designed to effectively fuse data from the main and vessel branches.

Main Branch. We adopt the TransUNet [4] as the main branch due to its superior performance on other medical image segmentation tasks. In the main branch, we utilize a CNN-Transformer hybrid structure as the encoder. The CNN part is used as the initial feature extractor. It provides features at different scales for the skip connections to compensate for the information loss in the downsampling operation. The extracted features are then processed by 12 consecutive transformer blocks at the bottleneck of the UNet architecture. The transformer encodes the long-range dependencies of the input fundus image due to the multi-head self-attention structure. The output features of the last transformer block are then resized for later decoding operations.

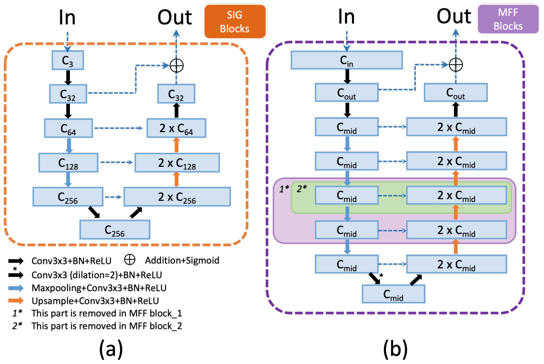
In
Out
SIG Blocks
MFF Blocks
Conv3x3+BN+ReLU
Addition+Sigmoid
Conv3x3 (dilation=2)+BN+ReLU
Maxpooling+Conv3x3+BN+ReLU
Upsample+Conv3x3+BN+ReLU
1* This part is removed in MFF block_1
2* This part is removed in MFF block_2
(a)
(b)

Fig. 2. The structures of SIG block and MFF block. The subscript of C denotes channel depths. C_{in}, C_{mid} and C_{out} represent channel depths of input, intermediate, and output feature maps for the MFF block, respectively. We set C_{mid} of three MFF blocks to small numbers, *i.e.*128, 64, 32, for improving the efficiency of multi-scale feature fusion.

Vessel Branch. In the vessel branch, we aim to exploit the structure information from the blood vessels. Unlike the main branch, where the input is a fundus image, we put in a vessel segmentation map generated by a pre-trained model. The pre-trained vessel segmentation model is built on the DRIVE dataset [15] with the TransUNet [4] architecture. Four identical spatial information guidance (SIG) blocks are utilized in the vessel branch to extract multi-scale vessel-based features. The rescaled vessel segmentation maps are fed into the SIG blocks, the details of which are illustrated in Fig. 2-a. The design of SIG blocks makes extensive use of customized ReSidual U-blocks (RSU). Qin *et al.* [16] indicate that the RSU block is superior in performance to other embedded structures (*e.g.*, plain convolution, residual-like, inception-like, and dense-like blocks), due to the enlarged receptive fields of the embedded U-shape architecture.

Multi-scale Feature Fusion (MFF) blocks. In contrast to the plain convolutional decoder blocks of the basic TransUNet, we use three Multi-scale Feature Fusion (MFF) blocks as the decoders for effective multi-scale feature fusion. The input to each MFF block is the concatenation of three types of features: (1) the multi-scale skip-connection features from the main branch, (ii) the hidden feature encoded by the last transformer block or the previous MFF block, (iii) the multi-scale SIG features from the vessel branch. The architecture of the MFF blocks is illustrated in Fig. 2-b, which is similar to one of the SIG blocks. From MFF block_1 to MFF block_3, we gradually increase the number of network layers in each MFF block. In this way, the later MFF blocks can capture more spatial context corresponding to larger feature maps. In the end, the concatenated feature maps of MFF block_3 and SIG block_4 are passed to two convolutional layers for outputting the fovea region score maps.

3.2. Implmentation Details

We first remove the uninformative black background from the original fundus image, then pad and resize the cropped image

Table 1. Comparison of performance on normal and diseased retinal images of both Messidor and PALM dataset. The best and second best results are highlighted in bold and italics respectively.

	1/8 R(%)		1/4 R(%)		1/2 R(%)		1R(%)		2R(%)	
Messidor	Normal	Diseased	Normal	Diseased	Normal	Diseased	Normal	Diseased	Normal	Diseased
UNet (2015) [17]	82.65	79.00	95.15	93.33	97.76	95.00	97.95	95.33	97.95	95.33
U2 Net (2020) [16]	86.19	81.33	**98.51**	97.33	*99.63*	99.50	*99.63*	99.50	*99.63*	99.50
TransUNet (2021) [4]	*87.31*	**84.33**	*98.32*	*97.67*	**100.00**	*99.83*	**100.00**	*99.83*	**100.00**	*99.83*
Bilateral-ViT (**Proposed**)	**87.50**	*84.00*	**98.51**	**98.67**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**

	1/8 R(%)		1/4 R(%)		1/2 R(%)		2/3 R(%)		1R(%)	
PALM	Normal	Diseased	Normal	Diseased	Normal	Diseased	Normal	Diseased	Normal	Diseased
UNet (2015) [17]	57.45	9.43	74.47	18.87	76.60	41.51	76.60	50.94	76.60	64.15
U2 Net (2020) [16]	*70.21*	*11.32*	*93.62*	*28.30*	*95.74*	*60.38*	95.74	*77.36*	*97.87*	*84.91*
TransUNet (2021) [4]	**82.98**	5.66	**95.74**	18.87	**97.87**	43.40	*97.87*	52.83	*97.87*	75.47
Bilateral-ViT (**Proposed**)	**82.98**	**13.21**	**95.74**	**37.74**	**97.87**	**69.81**	**100.00**	**81.13**	**100.00**	**92.45**

region to a spatial resolution of 512×512. We perform intensity normalization and data augmentation on the input images of the main branch and the vessel branch. To train our Bilateral-ViT network, we generate circular fovea segmentation masks from the ground-truth fovea coordinates. During the testing phase, we apply the sigmoid function to network prediction for the probabilistic map. We then collect all pixels with significant probabilistic scores and calculate their median coordinates as the final fovea location coordinates.

All experiments are coded by PyTorch and conducted on one NVIDIA GeForce GTX TITAN GPU. The weights of convolutional and linear layers are initialized by Kaiming initialization protocol [18]. The initial learning rate is $1e^{-3}$ and gradually decays to $1e^{-7}$ over 200 epochs by Cosine Annealing LR strategy. The optimizer is Adam [19] and the batch size is 2. We employ a combination of dice loss and binary cross-entropy as the loss function.

4. EXPERIMENTS

We perform experiments on Messidor [20] and PALM [21] datasets. The Messidor dataset is for diabetic retinopathy analysis. It consists of 540 normal and 660 diseased retinas. We utilize 1136 images from this dataset with fovea locations provided by [22]. The PALM dataset was released by the Pathologic Myopia Challenge (PALM) 2019. It consists of 400 images annotated with fovea locations, in which 213 images are pathologic myopia, and the remaining 187 images are normal retinas. For the fairness of comparisons, we keep our data split identical to [13].

To evaluate the performance of fovea localization, we adopt the following evaluation protocol [22]: the fovea localization is considered successful when the Euclidean distance between the ground-truth and predicted fovea coordinates is no larger than a predefined threshold value, such as the optic disc radius R. For a comprehensive evaluation, accuracy corresponding to different evaluation threshold (for example, $2R$ indicating the predefined threshold values are set to twice the optic disc radius R) is usually reported.

4.1. Fovea Localization on Normal and Diseased Images

In Table 1, we evaluate the performance of normal and diseased cases separately. We reimplement several widely used segmentation networks as comparison baselines, such as UNet [17], U2 Net [16], and TransUNet [4]. Bilateral-ViT obtains 100% accuracy from $1/2R$ to $1R$ on all the images of Messidor, and 100% accuracy from $2/3R$ to $1R$ on normal images of PALM. It demonstrates that the performance of Bilateral-ViT is highly reliable for normal fundus images.

Table 2. Comparison with existing studies on the Messidor and PALM datasets based on the R rule. The best and second best results are highlighted in bold and italics respectively.

Messidor	1/8 R (%)	1/4 R (%)	1/2 R (%)	1R (%)	2R (%)
Gegundez-Arias *et al.*(2013) [22]	-	76.32	93.84	98.24	99.30
Aquino (2014) [3]	-	83.01	91.28	98.24	99.56
Dashtbozorg *et al.*(2016) [23]	-	66.50	93.75	98.87	-
Girard *et al.*(2016) [24]	-	-	94.00	98.00	-
Molina-Casado *et al.*(2017) [25]	-	-	96.08	98.58	99.50
Al-Bander *et al.*(2018) [10]	-	66.80	91.40	96.60	99.50
Meyer *et al.*(2018) [11]	70.33	94.01	97.71	99.74	-
GeethaRamani *et al.*(2018) [26]	-	85.00	94.08	99.33	-
Zheng *et al.*(2019) [27]	60.39	91.36	98.32	99.03	-
Huang *et al.*(2020) [12]	-	70.10	89.20	99.25	-
Xie *et al.*(2020) [13]	*83.81*	*98.15*	*99.74*	*99.82*	**100.00**
Bilateral-ViT (**Proposed**)	**85.65**	**98.59**	**100.00**	**100.00**	**100.00**

PALM	1/8 R (%)	1/4 R (%)	1/2 R (%)	2/3 R (%)	1R (%)
Xie *et al.*(2020) [13]	-	-	-	*87*	*94*
Bilateral-ViT (**Proposed**)	**46**	**65**	**83**	**90**	**96**

For the diseased cases on the PALM dataset, Bilateral-ViT reaches 92.45% foveal localization accuracy for the threshold of $1R$ and significantly outperforms the second-best results by a large margin (7.54%). Fig. 3 provides some visual results of fovea localization on diseases images from the PALM dataset. Our Bilateral-ViT generates the most accurate predictions for the severely diseased image images with large atrophic regions (see Fig. 3-a and Fig. 3-b), or the heavily blurred image (see Fig. 3-c). In Fig. 3-d where the fovea is close to the image border, the predicted fovea locations from baseline networks (UNet and U2 Net) appear on the wrong side of the optic disc. However, TransUNet [4] and our method still perform well potentially due to their long-range modeling capability. Such results highlight that our proposed Bilateral-ViT has a significant advantage for diseased cases.

4.2. Comparison to the State-of-the-art methods

From Table 2, the Bilateral-ViT achieves state-of-the-art performance for all the evaluation settings. In particular, on the Messidor dataset, at $1/8R$, our network reaches the best accuracy of 85.65% with a gain of 1.84% compared to the second-best score (83.81%) [13]. It also reaches an accuracy of 100% at evaluation thresholds of $1/2R$, $1R$, and $2R$; in other words, the localization errors are at most $1/2R$ (approx-

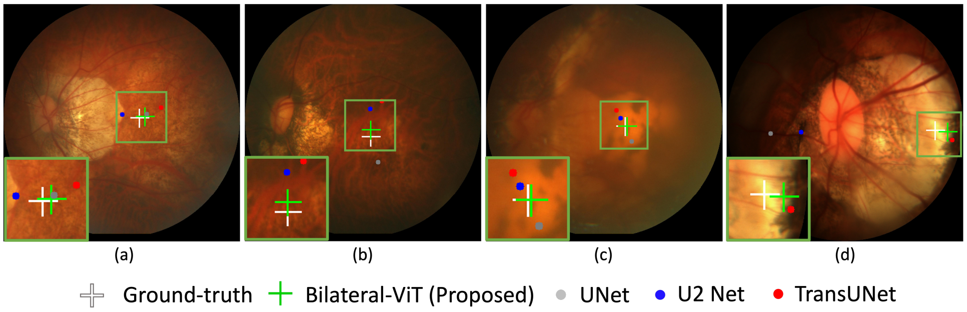

Fig. 3. Visual results of fovea localization predicted by different methods.

Table 3. **Top** and **Bottom**: Performance of the ablation study on the Messidor and PALM datasets respectively. VB refers to the vessel branch. The best and second best results are highlighted in bold and italics.

Messidor	1/8 R (%)	1/4 R (%)	1/2 R (%)	1R (%)	2R (%)
ViT+plain decoder (TransUNet [4])	**85.74**	97.98	*99.91*	*99.91*	*99.91*
ViT+VB+plain decoder	85.56	*98.33*	99.74	*99.91*	*99.91*
ViT+VB+MFF (**Proposed**)	*85.65*	**98.59**	**100.00**	**100.00**	**100.00**
ViT+VB (fundu as the input)+MFF	*85.65*	97.89	*99.91*	**100.00**	**100.00**

PALM	1/8 R (%)	1/4 R (%)	1/2 R (%)	2/3 R (%)	1R (%)
ViT+plain decoder (TransUNet [4])	42	55	69	74	86
ViT+VB+plain decoder	*45*	52	72	77	85
ViT+VB+MFF (**Proposed**)	**46**	**65**	**83**	**90**	**96**
ViT+VB (fundu as the input)+MFF	43	*58*	*82*	*89*	**96**

imately 19 pixels for an input image size of 512×512). PALM is a considerably more challenging dataset due to fewer images and complex diseased patterns. Our method achieves the accuracy of 90% and 96% at $2/3R$ and $1R$, which is 3% and 2% better than the previous work [13], respectively.

4.3. Ablation Study and Cross-Dataset Experiments

We conduct a comprehensive set of ablation experiments to evaluate the effectiveness of different components (see Table 3):

- ViT+plain decoder: the TransUNet architecture [4] comprised of a vision transformer-based encoder and a plain decoder is used as the comparison baseline.
- ViT+VB+plain decoder: we add the vessel branch (vessel segmentation mask as the input) to the baseline network.
- ViT+VB+MFF (**the proposed Bilateral-ViT**): we add the vessel branch (vessel segmentation mask as the input) and MFF blocks to the baseline network.
- ViT+VB (fundus as the input)+MFF: we add the vessel branch (fundu image as the input) and MFF blocks to the baseline network. This configuration compares the performance differences between fundus images and vessel segmentation maps as inputs to the vessel branch.

The performance of "ViT+plain decoder (TransUNet)" and "ViT+VB+plain decoder" are similar on both datasets; a possible reason is that the plain decoder does not have adequate capacity to fuse features from the vessel branch and transformer blocks. By further adding MFF blocks, the proposed Bilateral-ViT (ViT+VB+MFF) shows superior performance, suggesting the significance of the customized MFF blocks. The performance of "ViT+VB+MFF' is much better than "ViT+VB (fundus as the input)+MFF", demonstrating the usefulness of the vessel segmentation map. On the other hand, we note that "ViT+VB (fundus as the input)+MFF" outperforms all the existing works, implying our network can achieve the state-of-the-art performance even without the input of vessel segmentation map.

Table 4. Performance of cross-dataset experiments. The models used hare are exactly those selected in **Bottom** of Table 3. They are constructed on PALM only and generate the following results on Messidor. The higher results based on the R rule are better. The lower results based on distance errors are better. VB refers to the vessel branch. The best and second best results are highlighted in bold and italics respectively.

Cross-Dataset	1/8 R(%)	1/4 R(%)	1/2 R(%)	1R(%)	2R(%)	Errors
Xie *et al.* [13]	-	-	-	95.26	-	22.84
ViT+plain decoder (TransUNet)	77.82	*95.95*	**98.59**	*99.03*	99.30	10.76
ViT+VB+plain decoder	*78.17*	95.69	98.24	98.77	99.12	11.38
ViT+VB+MFF (**Proposed**)	**81.78**	**96.48**	*98.42*	**99.38**	**100.00**	**8.57**
ViT+VB (fundu as the input)+MFF	77.02	94.28	97.62	98.68	*99.47*	*10.69*

We conduct cross-dataset experiments to assess the generalization capability of the proposed Bilateral-ViT. The models are trained on PALM dataset and test on Messidor dataset. From Table 4, the accuracy is 99.38% at $1R$, which is a 4.12% improvement over the best-reported result (95.26%). The average localization error for the original image resolution is 8.57 pixels compared to the previous best result of 22.84 pixels. In addition, the proposed Bilateral-ViT outperforms the baselines by a significant margin, especially for $1/8R$, demonstrating its robustness for the cross-dataset setting.

5. CONCLUSIONS

This paper proposes a novel vision transformer (ViT) approach for robust fovea localization. It consists of a transformer-based main network branch for integrating global context and a vessel branch for explicitly incorporating the structure of blood vessels. The encoded features are subsequently merged with a customized multi-scale feature fusion (MFF) module. Our experiments demonstrate that the proposed approach has a significant advantage in handling diseased images. It also has excellent generalization capability, as shown in the cross-dataset experiments. Thanks to the transformer-based feature encoder, the incorporation of blood vessel structure, and the carefully designed MFF module, our approach establishes the new state of the arts on both Messidor and PALM datasets.

6. COMPLIANCE WITH ETHICAL STANDARDS

Human retinal images made publicly available through Messidor and PALM datasets are used in this study. As confirmed by the license accompanying the open-access data, no ethical approval is required.

7. REFERENCES

[1] JJ Weiter, GL Wing, CL Trempe, and MA Mainster, "Visual acuity related to retinal distance from the fovea in macular disease.," *Annals of ophthalmology*, 1984.

[2] Huiqi Li and Opas Chutatape, "Automated feature extraction in color retinal images by a model based approach," *IEEE Transactions on biomedical engineering*, 2004.

[3] Arturo Aquino, "Establishing the macular grading grid by means of fovea centre detection using anatomical-based and visual-based features," *Computers in biology and medicine*, 2014.

[4] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[5] Dharitri Deka, Jyoti Prakash Medhi, and SR Nirmala, "Detection of macula and fovea for disease analysis in color fundus images," in *2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS)*. IEEE, 2015.

[6] Jyoti Prakash Medhi and Samarendra Dandapat, "An effective fovea detection and automatic assessment of diabetic maculopathy in color fundus images," *Computers in biology and medicine*, 2016.

[7] Harihar Narasimha-Iyer, Ali Can, Badrinath Roysam, V Stewart, Howard L Tanenbaum, Anna Majerovics, and Hanumant Singh, "Robust detection and classification of longitudinal changes in color retinal fundus images for monitoring diabetic retinopathy," *IEEE transactions on biomedical engineering*, 2006.

[8] S Sekhar, Waleed Al-Nuaimy, and Asoke K Nandi, "Automated localisation of optic disk and fovea in retinal fundus images," in *2008 16th European Signal Processing Conference*. IEEE, 2008.

[9] Khawaja Muhammad Asim, A Basit, and Abdul Jalil, "Detection and localization of fovea in human retinal fundus images," in *2012 International Conference on Emerging Technologies*. IEEE, 2012.

[10] Baidaa Al-Bander, Waleed Al-Nuaimy, Bryan M Williams, and Yalin Zheng, "Multiscale sequential convolutional neural networks for simultaneous detection of fovea and optic disc," *Biomedical Signal Processing and Control*, 2018.

[11] Maria Ines Meyer, Adrian Galdran, Ana Maria Mendonça, and Aurélio Campilho, "A pixel-wise distance regression approach for joint retinal optical disc and fovea detection," in *MICCAI*. Springer, 2018.

[12] Yijin Huang, Zhiquan Zhong, Jin Yuan, and Xiaoying Tang, "Efficient and robust optic disc detection and fovea localization using region proposal network and cascaded network," *Biomedical Signal Processing and Control*, 2020.

[13] Ruitao Xie, Jingxin Liu, Rui Cao, Connor S Qiu, Jiang Duan, Jon Garibaldi, and Guoping Qiu, "End-to-end fovea localisation in colour fundus images with a hierarchical deep regression network," *IEEE Transactions on Medical Imaging*, 2020.

[14] Suman Sedai, Ruwan Tennakoon, Pallab Roy, Khoa Cao, and Rahil Garnavi, "Multi-stage segmentation of the fovea in retinal fundus images using fully convolutional neural networks," in *ISBI*. IEEE, 2017.

[15] 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*, 2004.

[16] Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R Zaiane, and Martin Jagersand, "U2-net: Going deeper with nested u-structure for salient object detection," *Pattern Recognition*, vol. 106, 2020.

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015.

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Identity mappings in deep residual networks," in *ECCV*. Springer, 2016.

[19] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[20] Etienne Decencière, Xiwei Zhang, Guy Cazuguel, Bruno Lay, Béatrice Cochener, Caroline Trone, Philippe Gain, Richard Ordonez, Pascale Massin, Ali Erginay, et al., "Feedback on a publicly distributed image database: the messidor database," *Image Analysis & Stereology*, 2014.

[21] Huazhu Fu, Fei Li, José Ignacio Orlando, Hrvoje Bogunović, Xu Sun, Jingan Liao, Yanwu Xu, Shaochong Zhang, and Xiulan Zhang, "Palm: Pathologic myopia challenge," 2019.

[22] Manuel E Gegundez-Arias, Diego Marin, Jose M Bravo, and Angel Suero, "Locating the fovea center position in digital fundus images using thresholding and feature extraction techniques," *Computerized Medical Imaging and Graphics*, 2013.

[23] Behdad Dashtbozorg, Jiong Zhang, Fan Huang, and Bart M ter Haar Romeny, "Automatic optic disc and fovea detection in retinal images using super-elliptical convergence index filters," in *International Conference on Image Analysis and Recognition*. Springer, 2016.

[24] Fantin Girard, Conrad Kavalec, Sébastien Grenier, Houssem Ben Tahar, and Farida Cheriet, "Simultaneous macula detection and optic disc boundary segmentation in retinal fundus images," in *Medical Imaging 2016: Image Processing*. International Society for Optics and Photonics, 2016.

[25] José M Molina-Casado, Enrique J Carmona, and Julián García-Feijoó, "Fast detection of the main anatomical structures in digital retinal images based on intra-and inter-structure relational knowledge," *Computer methods and programs in biomedicine*, 2017.

[26] R GeethaRamani and Lakshmi Balasubramanian, "Macula segmentation and fovea localization employing image processing and heuristic based clustering for automated retinal screening," *Computer methods and programs in biomedicine*, 2018.

[27] Shaohua Zheng, Youxing Zhu, Lin Pan, and Ting Zhou, "New simplified fovea and optic disc localization method for retinal images," *Journal of Medical Imaging and Health Informatics*, 2019.