UAU-Net: United Attention U-Shaped Network for the Segmentation of Pigment Deposits in Fundus Images of Retinitis Pigmentosa

Jingcheng Xu[1], Zhuoshi Wang[2], Weifang Zhu[1], Yi Zhou[1], Yan Sun[3],
Zhuang Li[3], Ming Liu[1], Wenhao Tan[1], Ling Xu[3](✉), and Xinjian Chen[1,4](✉)

[1] School of Electronics and Information Engineering,
Soochow University, Suzhou, China
xjchen@suda.edu.cn

[2] Gene Center of Precision Medicine Innovation Institute,
He University, Shenyang, China

[3] Genetic Counseling Clinic, He Eye Specialist Hospital, Shenyang, China
xuling@hsyk.com.cn

[4] State Key Laboratory of Radiation Medicine and Protection,
Soochow University, Suzhou, China

Abstract. Retinitis Pigmentosa (RP) is a retinal disease with high rate of blindness. Retinal pigment deposits are a typical symptom of RP, whose automatic segmentation is crucial to the early diagnosis of RP. In fundus images, pigment deposits have various shapes and sizes and are scattered randomly, which makes the automatic segmentation very challenging. In this paper, we propose a United Attention U-shaped Network (UAU-Net) for segmentation of pigment deposits in fundus images, comprising four parts: encoder, decoder, Multi-scale Global Attention Module (MsGAM), and Spatial-enhanced Attention Module (SEAM). The MsGAM is proposed to extract multi-scale spatial and channel information in constructing associations between different locations of pigment deposits, and the SEAM is proposed to preserve detailed features and enhance the model's ability to segment small targets. Comprehensive experiments on 215 fundus images show that UAU-Net outperforms other state-of-the-art methods with Dice and Intersection-over-Union of 60.25% and 44.91%, respectively.

Keywords: Retinitis pigmentosa · Pigment deposits segmentation · United attention mechanism · Deep neural network

1 Introduction

Retinitis Pigmentosa (RP) is an inherited retinal dystrophy caused by loss of photoreceptors with a global prevalence of approximately 0.025% [8], which is irreversible and has high risk of blindness. Therefore, it is significant for RP

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
B. Antony et al. (Eds.): OMIA 2023, LNCS 14096, pp. 52–61, 2023.
https://doi.org/10.1007/978-3-031-44013-7_6

patients to be promptly diagnosed at an early stage. Retinal pigment deposits are a classic RP feature observable in fundus images, and a method that automatically and accurately segments retinal pigment deposits would be important for the early diagnosis of RP as well as for grading studies. However, pigment deposits in fundus images have various shapes and sizes and are scattered randomly, and blurring of the image also leads to reduced contrast of pigment deposits, all of which bring challenges to segmentation.

At this stage, there are many methods to segment small targets like pigment deposits, such as diabetic retinal exudates. Sambyal et al. [12] proposed a new upsampling technique to improve segmentation performance. Wang et al. [16] proposed a Contextual and local collaborative network (CLC-Net), using a collaborative architecture that comprises a contextual branch and a local branch. For segmenting pigment deposits, currently, most methods for segmenting pigmentation are based on traditional machine learning. Brancati et al. [3] proposed a supervised segmentation method for pigment signs detection in fundus images using an integrated classifier with Random Forests and Adaptive Boosting, and they also proposed another method [2] based on the relationship between adjacent region features. These methods rely on manually designed features and cannot achieve automatic end-to-end segmentation. With the development of deep learning (DL), neural network-based methods have also started to emerge. Brancati et al. [4] first applied a U-Net [11] based network to segment pigment deposits in retinal fundus images. Arsalan et al. [1] proposed an automatic RP Segmentation Network (RPS-Net) that is able to enhance the ability to segment pigment deposits through multiple dense connections between convolutional layers. However, these DL-based methods do not efficiently extract semantic information, which tend to negatively affect the final segmentation results.

To tackle these problems and improve the ability to segment small and dispersed targets, in this paper, we propose a novel end-to-end learning framework UAU-Net. Our main contributions include: 1) Proposing a Multi-scale Global Attention Module (MsGAM) to capture multi-scale global semantic features from encoder to enable model to ignore the problem of scattered randomly pigment deposits. 2) Introducing a Spatial-enhanced Attention Module (SEAM) to guide model in learning contextual information to preserve detailed features against small target segmentation. 3) Combining MsGAM and SEAM, a U-shaped structure-based model UAU-Net is proposed to achieve good pigment deposits segmentation performance on all three different fundus image data.

2 Method

2.1 Proposed Model

Overall Architecture: The diagram of proposed UAU-Net is shown in Fig. 1, with an overall U-shaped network consisting of encoder, decoder, MsGAM, and SEAM. The MsGAM is inserted at the top of the encoder to capture multi-scale global semantic information. And the SEAM is placed between the encoder and the decoder at each layer as a skip connection to extract more contextual

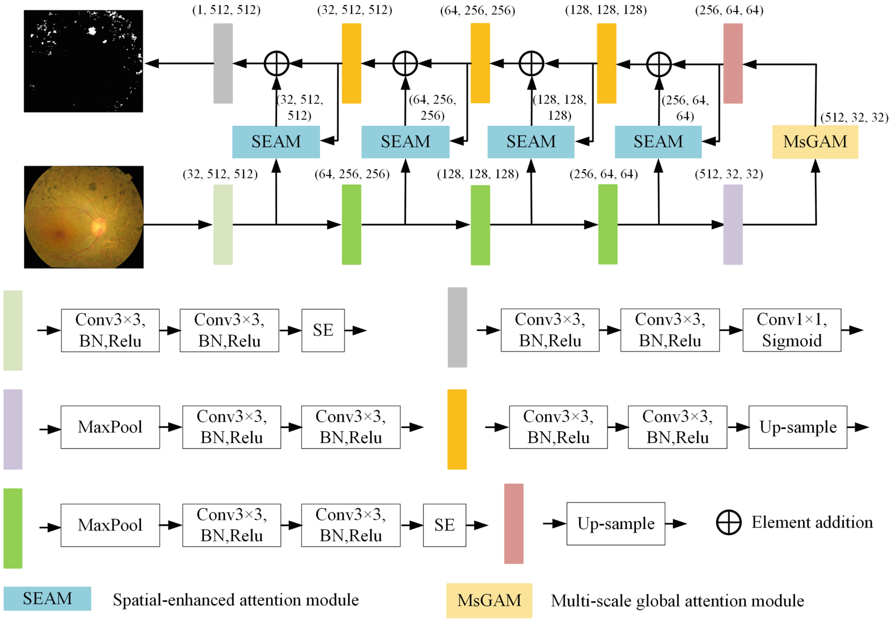

Fig. 1. Diagram of the proposed UAU-Net. Above each block is the number of output channels and the output feature map size. Fundus images are fed into the network, passed through encoder to extract high-level features, then through MsGAM to capture multi-scale global semantic information. Finally, these features are recovered by decoder and meanwhile the contextual semantic information flows are introduced by SEAM.

semantic information. Our goal is to obtain an end-to-end pigment deposits segmentation model by feeding fundus images into the network.

Backbone for Image Segmentation: The backbone of the proposed segmentation model is a U-Net [11] based network with Squeeze-and-Excitation (SE) block added to first four layers of encoder. Compared with U-Net encoder, adding the SE block enhances the ability of the encoder in extracting semantic features.

Multi-scale Global Attention Module: Due to the complexity of pigment deposits in fundus images, segmentation model needs to enhance the extraction of multi-scale global features. Inspired by convolutional block attention module [17] and Self-attention [14], a novel MsGAM is designed and inserted to top layer of encoder, which is illustrated in Fig. 2. Unlike the traditional self-attention, three inputs Key (K), Query (Q), and Value (V) each have different input information. The feature map X_{in} obtained from the encoder is fed into 1×1 convolution as input of Q. The feature map X_m is obtained by multi-scale information fusion on X_{in}, then X_m is passed through a global averag pooling

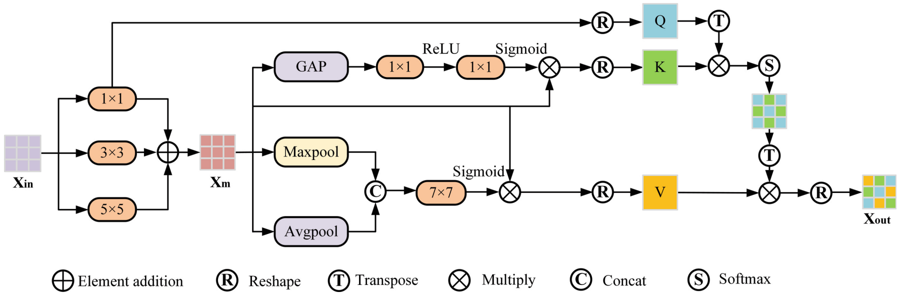

Fig. 2. The illustration of Multi-scale Global Attention Module (MsGAM).

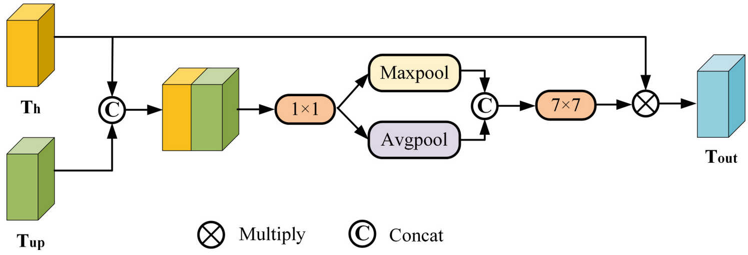

Fig. 3. The illustration of Spatial-enhanced Attention Module (SEAM).

and two 1×1 convolutions as input of K. After X_m undergoes a Spatial Attention Module (SAM) [17], it serves as input of V. The proposed MsGAM allows model to capture local spatial and channel information at multiple scales by compressing the channel dimension and spatial dimension, respectively, while obtaining global long-range information. This results in a higher discriminative power of model, which is suitable for segmenting widely distributed lesion regions like pigment deposits.

Spatial-Enhanced Attention Module: The simple skip connection in U-Net combines local information with different levels indiscriminately, ignoring semantic information and may be disturbed by unrelated features [13]. To solve this problem, we design a SEAM to take the place of the skip connection in U-Net, whose structure is shown in Fig. 3. The feature map T_h with high-resolution weak semantic features from encoder is combined with the up-sampled T_{up} with low resolution strong semantic features from decoder [15]. Subsequently, the obtained features are normalized by a 1×1 convolution and a SAM to get the weight information, which is then multiplied with T_h and finally the final feature map T_{out} is obtained. This allows the model to fully account for the contextual information to retain more semantic information, achieving a good segmentation of smaller targets.

Loss Function: We utilize a joint loss function of Dice loss and binary cross entropy (BCE) loss. Dice loss can effectively alleviate the data imbalance problem, which facilitates the segmentation of pigment deposits with varying size and random scatter, and the addition of BCE loss could stabilize the training of the model. The joint loss can be expressed as follows:

$$\begin{aligned} L_{joint} &= L_{Dice} + L_{BCE} \\ &= 1 - \frac{2\sum_{i=1}^{C} g_i \times p_i}{\sum_{i=1}^{C} g_i + p_i} - \frac{1}{C}\sum_{i=1}^{C} g_i \log p_i + (1-g_i)\log(1-p_i), \end{aligned} \tag{1}$$

where $0 \leq g \leq 1$ and $0 \leq p \leq 1$ are segmentation ground truth and predicted probability, respectively. C is the sum of output results in pixels, i is per pixel.

2.2 Dataset

An in-house dataset consisting of 215 fundus images with RP are used to evaluate the proposed UAU-Net, which was approved by IRB of the University and informed consent was obtained from all subjects. The corresponding segmentation ground truth was manually annotated by professional ophthalmologists. The dataset contains seven different sizes of fundus images, which are 3100 × 2848, 2592 × 1944, 2528 × 2036, 2048 × 1536, 1600 × 1216, 884 × 818 and 874 × 957, acquired by two types of cameras. For each original image, a crop of the smallest outer rectangle of the retinal region was used to remove useless background and mitigate its negative impact on segmentation performance. We roughly group the dataset randomly by 6:2:2 for training, validation, and testing, specifically 132, 42, and 41.

2.3 Evaluation Metrics

The evaluation metrics used in this paper include Dice, Intersection-over-Union (IoU), Accuracy (Acc), and Specificity (Spec), where Dice is regarded as the most important metric. These metrics are defined as:

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN}, IoU = \frac{TP}{TP + FP + FN}, \tag{2}$$

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}, Spec = \frac{TN}{TN + FN}, \tag{3}$$

where TP, FN, FP, and TN are the pixel number of true positive, false negative, false positive, and true negative, respectively.

3 Experiments and Results

3.1 Implementation Details

We use Adam optimizer with an initial learning rate of 1.0×10^{-4} and a momentum of 0.9. The batch size is set to 4 and 350 epochs are trained. All images

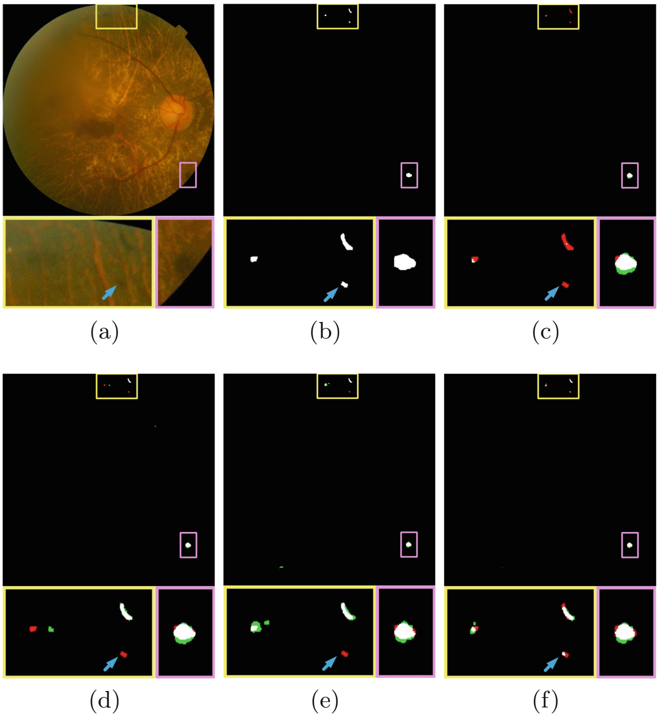

Fig. 4. Examples results of ablation study. (a) Original Fundus image (b) Ground truth (c) Baseline (d) Baseline+SEAM (e) Baseline+MsGAM (f) Baseline+MsGAM+SEAM (Proposed UAU-Net). Red, green, and white color represent FN, FP, and TP, respectively. The blue arrows point to the hard-to-segment pigment deposits. Best view in color and zoom in. (Color figure online)

are first resized to 1024 × 1024, then divided into 4 patches, i.e. 512 × 512 each, and then they are enhanced by flipping horizontally and vertically respectively, after which they are served as input to the model. The proposed UAU-Net is implemented using Pytorch and is trained with one NVIDIA RTX 3060 GPU of 12 GB memory.

3.2 Ablation Study

To investigate the contribution of each proposed module to segmentation performance, we conduct ablation experiments in four cases: 1) the first four layers of U-Net encoder join the SE module as Baseline, 2) combination of Baseline and SEAM, 3) combination of Baseline and MsGAM, and 4) the Baseline combines both SEAM and MsGAM, i.e., the segmentation model UAU-Net proposed in this paper. The experimental results are shown in Table 1.

The visualization results of the ablation study are shown in Fig. 4, with yellow and pink rectangular boxes representing two manually selected regions of

Table 1. Results of ablation study (mean ± standard deviation).

Method	Dice (%)	IoU (%)	Acc (%)	Spec (%)
Baseline	58.87±17.07	43.64±16.14	99.07±1.24	99.51±0.69
Baseline+SEAM	59.60±17.47	44.46±16.37	98.90±1.64	99.26±1.39
Baseline+MsGAM	59.96±16.10	44.62±15.87	98.92±1.79	99.31±1.48
UAU-Net (Proposed)	**60.25±16.11**	**44.91±15.73**	**99.09±1.27**	**99.53±0.73**

interest for better comparison. It can be clearly seen that the pigment deposits in the yellow box are smaller and the edges are more blurred, and they are more difficult to be segmented compared to those in the pink box. The addition of SEAM (Fig. 4(d)) and MsGAM (Fig. 4(e)) makes the pigment deposits segmentation more comprehensive compared to the Baseline (Fig. 4(c)). When both modules are used (Fig. 4(f)), the model accurately segmented the pigment deposits pointed by blue arrows, demonstrating that the proposed UAU-Net is able to successfully segment difficult small targets regardless of the variation in size and random scatter of pigment deposits. This is also demonstrated by the results listed in Table 1. Due to the fusion of contextual information, the addition of SEAM alone more efficiently enhances the feature selection ability of the model, with Dice increasing from 58.87% to 59.60% and IoU increasing from 43.64% to 44.46%. The multi-scale information fusion and global attention brought by the MsGAM allows the model to more accurately capture important features, with Dice and IoU rising further to 59.96% and 44.62%, respectively. When the two modules are combined with the Baseline, Acc, IoU, Dice, and Spec all achieve the best. The proposed segmentation model UAU-Net improves Dice by 1.38% and IoU by 1.27% compared to the Baseline.

3.3 Comparison Study

For a more comprehensive demonstration of the superiority of UAU-Net, we compare seven state-of-the-art methods, including U-Net [11], U-Net++ [18], Context Encoder Network (CE-Net) [7], Context Pyramid Fusion Network (CPFNet) [6], Attention U-Net (Att-UNet) [10], Curvilinear Structure Segmentation Network (CS^2-Net) [9], Unet-like pure Transformer for medical image segmentation (Swin-Unet) [5]. The hyperparameters of all competing methods are set the same as the proposed UAU-Net to ensure fairness.

The qualitative results in Fig. 5 show that the proposed UAU-Net achieves the best segmentation performance (Fig. 5(j)). It can be seen from Fig. 5(a) that the pigment deposits to be segmented are widely distributed and vary greatly in size, making it extremely difficult to identify and segment them. When comparing all segmentation results, we can see that the sum of the red and green regions of the UAU-Net (Fig. 5(j)) proposed in this paper is the smallest, which demonstrates the best segmentation performance.

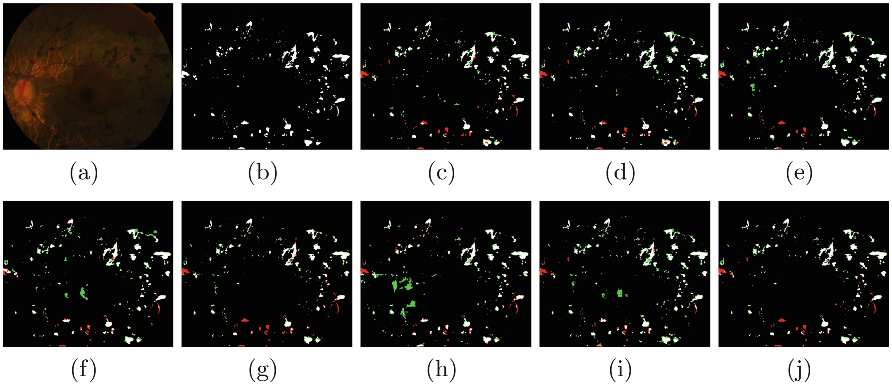

Fig. 5. Examples results of comparison experiments. (a) Original Fundus image (b) Ground truth (c) U-Net [11] (d) U-Net++ [18] (e) CE-Net [7] (f) CPFNet [6] (g) Att-UNet [10] (h) CS^2-Net [9] (i) Swin-Unet [5] (j)UAU-Net (Proposed). Red, green, and white color represent FN, FP, and TP, respectively. (Color figure online)

The comparison of quantitative performance is presented in Table 2, and the results demonstrate that our model achieves optimal results on Dice, IoU, and Acc, and also ranks second on Spec. Comparing to the best U-Net++ among competing methods, which has 58.51% and 43.62% on Dice and IoU, respectively, our UAU-Net still has a 1.74% and 1.29% boost on these two metrics. What's more, compared to Att-UNet, which has the highest computational cost (see the GFLOPs in Table 2), Dice and IoU improve 2.14% and 1.54%, respectively. With the exception of CS^2-Net and Swin-Unet, none of other methods focus on global attention, i.e., they ignore the long-distance dependence of features. For pigment deposit, which is a more widely distributed and variable-sized target, global attention allows long-distance features to be correlated with each other, making it easier to determine whether features at different distances are the same target. Although CS^2-Net and Swin-Unet also use a Self-attention related module and their computational cost is comparable to that of UAU-Net, they ignore the information of small targets and thus fail to segment the pigment deposits of small size. This leads to a Dice metric that are 8.17% and 6.28% respectively lower than the proposed model. It is worth noting that our method obtains the smallest standard deviation in both Dice and IoU, indicating that it obtains the most stable segmentation performance.

Table 2. Results of comparison experiments (mean ± standard deviation).

Method	Dice (%)	IoU (%)	Acc (%)	Spec (%)	GFLOPs
U-Net [11]	57.75±18.69	42.83±17.20	99.03±1.33	99.48±0.88	124.47
U-Net++ [18]	58.51±18.98	43.62±17.16	98.99±1.30	99.33±0.97	139.61
CE-Net [7]	58.11±18.48	43.10±16.77	99.01±1.27	99.37±0.94	35.60
CPFNet [6]	56.91±19.43	42.07±17.18	99.00±1.27	99.41±0.83	**32.28**
Att-UNet [10]	58.11±19.81	43.37±17.61	99.02±1.44	99.53±0.82	333.37
CS^2-Net [9]	52.08±20.19	37.57±17.53	98.87±1.59	99.47±0.97	56.02
Swin-Unet [5]	53.97±18.93	39.13±16.98	98.88±1.44	99.29±1.00	45.43
UAU-Net (Proposed)	**60.25±16.11**	**44.91±15.73**	**99.09±1.27**	**99.53±0.73**	83.02

4 Conclusion

In this paper, we propose a novel end-to-end segmentation network, UAU-Net, which can significantly improve the retinal pigment deposits segmentation performance in fundus images. The model proposes two modules, MsGAM and SEAM, for solving the problems of pigment deposits with random scatter and large size variation, respectively. The MsGAM is able to fuse multi-scale long-range spatial and channel information to construct associations of the same segmentation target at different locations. The SEAM enables combining contextual information to retain detailed features well and improve the segmentation accuracy of small targets. The experimental results show that the proposed model has good performance and great potential in segmenting retinal pigment deposits.

References

1. Arsalan, M., Baek, N.R., Owais, M., Mahmood, T., Park, K.R.: Deep learning-based detection of pigment signs for analysis and diagnosis of retinitis pigmentosa. Sensors **20**(12), 3454 (2020)
2. Brancati, N., Frucci, M., Gragnaniello, D., Riccio, D., Di Iorio, V., Di Perna, L.: Automatic segmentation of pigment deposits in retinal fundus images of retinitis pigmentosa. Comput. Med. Imag. Graph. **66**, 73–81 (2018)
3. Brancati, N., et al.: Learning-based approach to segment pigment signs in fundus images for retinitis pigmentosa analysis. Neurocomputing **308**, 159–171 (2018)
4. Brancati, N., Frucci, M., Riccio, D., Di Perna, L., Simonelli, F.: Segmentation of pigment signs in fundus images for retinitis pigmentosa analysis by using deep learning. In: Ricci, E., Rota Bulò, S., Snoek, C., Lanz, O., Messelodi, S., Sebe, N. (eds.) ICIAP 2019. LNCS, vol. 11752, pp. 437–445. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30645-8_40
5. Cao, H., et al.: Swin-unet: unet-like pure transformer for medical image segmentation. In: European Conference on Computer Vision, pp. 205–218. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-25066-8_9
6. Feng, S., et al.: Cpfnet: context pyramid fusion network for medical image segmentation. IEEE Trans. Med. Imag. **39**(10), 3008–3018 (2020)

7. Gu, Z., et al.: Ce-net: context encoder network for 2d medical image segmentation. IEEE Trans. Med. Imag. **38**(10), 2281–2292 (2019)
8. Hamel, C.: Retinitis pigmentosa. Orphanet J. Rare Dis. **1**(1), 1–12 (2006)
9. Mou, L., et al.: Cs2-net: deep learning segmentation of curvilinear structures in medical imaging. Med. Image Anal. **67**, 101874 (2021)
10. Oktay, O., et al.: Attention u-net: learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
11. 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). https://doi.org/10.1007/978-3-319-24574-4_28
12. Sambyal, N., Saini, P., Syal, R., Gupta, V.: Modified u-net architecture for semantic segmentation of diabetic retinopathy images. Biocybernet. Biomed. Eng. **40**(3), 1094–1109 (2020)
13. Song, J., et al.: Global and local feature reconstruction for medical image segmentation. IEEE Trans. Med. Imag. **41**(9), 2273–2284 (2022)
14. Vaswani, A., et al.: Attention is all you need. Adv. Neural Inf. Process. Syst. **30** (2017)
15. Wang, M., et al.: Mstganet: automatic drusen segmentation from retinal oct images. IEEE Trans. Med. Imag. **41**(2), 394–406 (2021)
16. Wang, X., et al.: Clc-net: contextual and local collaborative network for lesion segmentation in diabetic retinopathy images. Neurocomputing **527**, 100–109 (2023)
17. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19 (2018)
18. Zhou, Z., Rahman Siddiquee, Md.M., Tajbakhsh, N., Liang, J.: UNet++: a nested U-net architecture for medical image segmentation. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 3–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_1