RBGNet: Reliable Boundary-Guided Segmentation of Choroidal Neovascularization

Tao Chen[1,2], Yitian Zhao[1(✉)], Lei Mou[1], Dan Zhang[3], Xiayu Xu[4], Mengting Liu[5], Huazhu Fu[6], and Jiong Zhang[1(✉)]

[1] Institute of Biomedical Engineering, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, China
{yitian.zhao,zhangjiong}@nimte.ac.cn
[2] Cixi Biomedical Research Institute, Wenzhou Medical University, Ningbo, China
[3] School of Cyber Science and Engineering, Ningbo University of Technology, Ningbo, China
[4] School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, China
[5] School of Biomedical Engineering, Sun Yat-sen University, Shenzhen, China
[6] Institute of High Performance Computing, A*STAR, Singapore, Singapore

Abstract. Choroidal neovascularization (CNV) is a leading cause of visual impairment in retinal diseases. Optical coherence tomography angiography (OCTA) enables non-invasive CNV visualization with micrometerscale resolution, aiding precise extraction and analysis. Nevertheless, the irregular shape patterns, variable scales, and blurred lesion boundaries of CNVs present challenges for their precise segmentation in OCTA images. In this study, we propose a **R**eliable **B**oundary-**G**uided choroidal neovascularization segmentation **Net**work (RBGNet) to address these issues. Specifically, our RBGNet comprises a dual-stream encoder and a multi-task decoder. The encoder consists of a convolutional neural network (CNN) stream and a transformer stream. The transformer captures global context and establishes long-range dependencies, compensating for the limitations of the CNN. The decoder is designed with multiple tasks to address specific challenges. Reliable boundary guidance is achieved by evaluating the uncertainty of each pixel label, By assigning it as a weight to regions with highly unstable boundaries, the network's ability to learn precise boundary locations can be improved, ultimately leading to more accurate segmentation results. The prediction results are also used to adaptively adjust the weighting factors between losses to guide the network's learning process. Our experimental results demonstrate that RBGNet outperforms existing methods, achieving a Dice score of 90.42% for CNV region segmentation and 90.25% for CNV vessel segmentation. https://github.com/iMED-Lab/RBGnet-Pytorch.git.

Keywords: CNV · OCTA · Transformer · Uncertainty · Multi-task

1 Introduction

Age-related macular degeneration (AMD) is a leading cause of blindness worldwide, primarily attributable to choroidal neovascularization (CNV) [1]. Optical

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14223, pp. 163–172, 2023.
https://doi.org/10.1007/978-3-031-43901-8_16

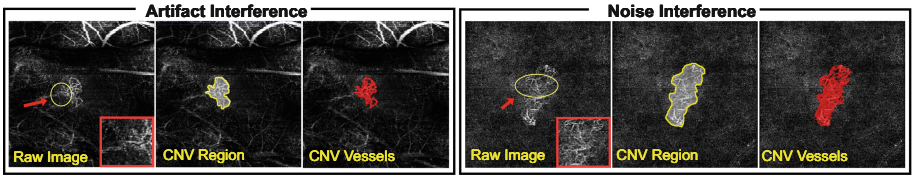

Fig. 1. Interference to CNV in OCTA images. Red arrows indicate artifacts and noise interference, respectively. (Color figure online)

coherence tomography angiography (OCTA), a non-invasive imaging technique, has gained popularity in recent years due to its ability to visualize blood flow in the retina and choroid with micrometer depth resolution [2]. Thus, automated CNV segmentation based on OCTA images can facilitate quantitative analysis and enhance the diagnosis performance of AMD [3]. However, the accurate segmentation of CNV from OCTA images poses a significant challenge due to the complex morphology of CNVs and the presence of imaging artifacts [4], as illustrated in Fig. 1. Hence, reliable CNV segmentation is promptly needed to assist ophthalmologists in making informed clinical decisions.

Several methods have been proposed to segment CNV regions from OCTA images, including handcraft feature descriptors [5,6] and deep learning-based techniques [7]. For example, a saliency-based method for automated segmentation of CNV regions in OCTA images was proposed by Liu *et al.* [5], which capitalizes on distinguishing features of CNV regions with higher intensity compared to background artifacts and noise. In [6], an unsupervised algorithm for CNV segmentation was proposed, which utilizes a density cell-like P system. However, their accuracy is restricted by weak saliency and ambiguous boundaries. With the recent advancements of deep learning, several methods have also been proposed for CNV segmentation in OCT images. U-shaped multiscale information fusion networks are proposed in [8] and [9] for segmenting CNVs with multiscale scenarios. Wang *et al.* [7] further proposed a two-stage CNN-based architecture based on OCTA images that is capable of extracting both CNV regions and vessel details. However, common issues including substantial scale variations of CNV regions and low-contrast microvascular boundaries were not fully deliberated in previous network designs. Thus, more dedicated modules with scale adaptivity and boundary refinement properties need to be explored to solve existing challenges.

Previously, Gal *et al.* [10] proposed a theory to effectively model uncertainty with dropout NNs. Afterward, Bragman *et al.* [11] applied this method to the field of medical image analysis. Nair *et al.* [12] showed the success of using dropout for the detection of three-dimensional multiple sclerosis lesions. Motivated by these findings, we also consider taking advantage of the pixel-level uncertainty estimation and making it adaptive to the segmentation of ambiguous CNV boundaries.

In this work, we propose a reliable boundary-guided network (RBGNet) to simultaneously segment both the CNV regions and vessels. Our proposed method is composed of a dual-branch encoder and a boundary uncertainty-guided multi-task decoder. The dual-branch encoder is designed to capture both of the global long-range dependencies and the local context of CNVs with significant scale variations, while the proposed uncertainty-guided multi-task decoder is designed to strengthen the model to segment ambiguous boundaries. The uncertainty is achieved by approximating a Bayesian network through Monte Carlo dropout.

The main contributions are summarized as follows:

(a) We propose a multi-task joint optimization method to interactively learn shape patterns and boundary contours for more accurate segmentation of CNV regions and vessels.
(b) We design a dual-stream encoder structure to take advantages of the CNN and transformer, which promote the network to effectively learn both local and global information.
(c) We propose an uncertainty estimation-guided weight optimization strategy to provide reliable guidance for multi-task network training.

2 Proposed Method

The proposed RBGNet comprises three primary components: a dual-stream encoder, a multi-task decoder, and an uncertainty-guided weight optimization strategy, as depicted in Fig. 2. The OCTA image is firstly processed using a dual-stream encoder that combines Convolutional Neural Networks (CNNs) and Vision Transformer (ViT) models [13] to produce high-dimensional semantic features. This approach enhances the local context and global dependencies of the image, thereby improving the CNV feature representation. These features are then fed into a multi-task decoder, which integrates information from multiple tasks to achieve better CNV segmentation. To further optimize the model's performance, we introduce a pixel-level uncertainty estimation approach that enhances the model's capacity to handle ambiguous region boundaries.

2.1 Dual-Stream Encoder

The CNVs in OCTA images are of various shapes and a wide range of scales, which pose challenges to the accurate segmentation of CNVs. To extract the long-range dependencies of cross-scale CNV information, we employ VIT [13] as an independent stream in the feature encoder of the proposed method. Moreover, the proposed dual-stream encoder is also embedded with a CNN-based multi-scale encoder to obtain the local context of CNV regions. Specifically, the VIT stream utilizes a stack of twelve transformer layers to extract features from flattened uniform non-overlapping patches, which seamlessly integrates with the CNN encoder stream through skip connections, similar to [14]. However, unlike [14], we divide the output representations of the transformer

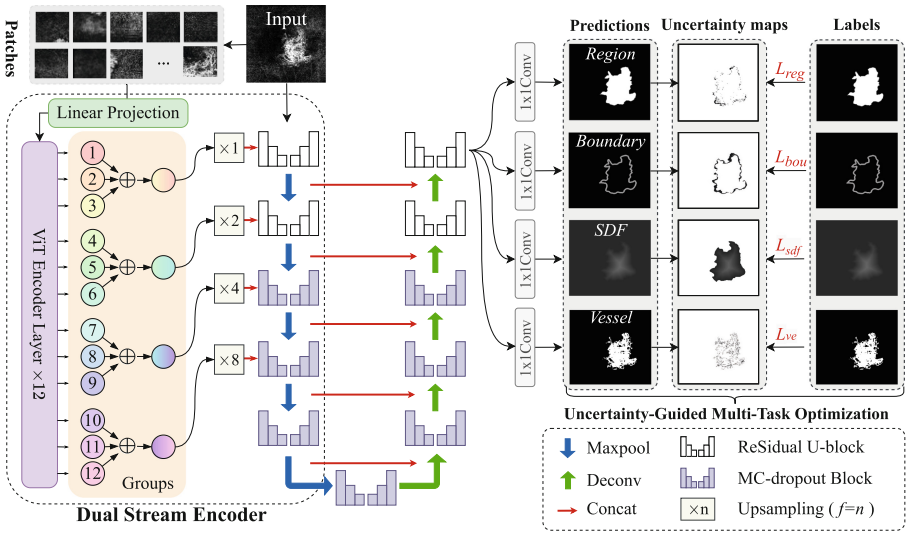

Fig. 2. Schematic diagram of the proposed RBGNet, which contains a dual-stream encoder and a multi-task decoder.

layers in VIT into four groups, each containing three feature representations. These groups correspond to the scales of the CNN stream, uniformly arranged from shallow to deep. Then, we perform element-wise summation for each group of representations, followed by reshaping them into non-overlapping patch sizes. The reshaped representation is further upsampled to the corresponding CNN feature resolution. The CNN branch is a U-shaped network [15] that extracts multi-scale features using ReSidual U-block (RSU) proposed by [16] to preserve high-resolution information locally. To integrate complementary information from the CNN and VIT features, we concatenate them in each of the first four feature extraction layers of the CNN branch. To enhance the features from the dual-branch encoder, we apply a bottleneck layer that consists of a feature extraction block, max-pooling, upsampling, and another feature extraction block.

2.2 Multi-task Decoder

The proposed multi-task decoder performs two main tasks: CNV region segmentation and vessel segmentation. The CNV region segmentation task contains two auxiliary subtasks including boundary prediction and shape regression. Each task is implemented at the end of the decoder using a 1×1 convolutional layer followed by a Sigmoid activation, as shown in Fig. 2.

Region Segmentation: Region segmentation of CNV allows accurate assessment of lesion size. This task aims to accurately segment the entire CNV regions in OCTA images via boundary prediction and shape regression. Region segmentation is typically accomplished by categorizing individual pixels as either

belonging to the CNV region or the background region. The purpose of region boundary prediction is to explicitly enhance the model's focus on ambiguous boundaries, allowing for more accurate region segmentation. The process of shape regression for region segmentation involves the transformation of boundary regression into a task of signed distance field regression. This is achieved by assigning a signed distance to each pixel, representing its distance from the boundary, with negative values inside the boundary, positive values outside the boundary, and zero values on the boundary. By converting the ground truth into a signed distance map (SDM), the network can learn CNV shape patterns from the rich shape pattern information contained in the SDMs.

Vessel Segmentation: To improve the vessel segmentation performance, we propose to guide the model to focus on low-contrast vessel details by estimating their pixel uncertainty. Simultaneously, the complementary information from the CNV region segmentation task is further utilized to eliminate the interference of vessel pixels outside the region, thus better refining the vessel segmentation results. The proposed multi-task decoder improves the segmentation accuracy of regions and vessels by explicitly or implicitly using the information between individual tasks and optimizing each task itself.

2.3 Uncertainty-Guided Multi-Task Optimization

Uncertainty Estimation: In contrast to traditional deep learning models that produce deterministic predictions, Bayesian neural networks [17] can provide not only predictions but also uncertainty. It treats the network weights as random variables with a priori distribution and infers the posterior distribution of the weights. In this paper, we employ Monte Carlo dropout (MC-dropout) [10] to approximate Bayesian networks and capture the uncertainty of the model. Bayesian inference offers a rigorous method for making decisions in the presence of uncertainty. However, the computational complexity of computing the posterior distribution often renders it infeasible. This issue is usually solved by finding the best approximation in a finite space.

To learn the weight distribution of the network, we minimize the Kullback-Leibler (KL) scatter between the true posterior distribution and its approximation. The probability distribution of each pixel is obtained based on Dropout to sample the posterior weight distribution M times. Then, the mean P_i of each pixel is used to generate the prediction, while the variance V_i is used to quantify the uncertainty of the pixel. This process can be described as follows.

$$P_i = \frac{1}{M}\sum_{m=1}^{M}\hat{P}_m, \quad \text{and} \quad V_i = \frac{1}{M}\sum_{m=1}^{M}\left(\hat{P}_m - P_i\right)^2. \tag{1}$$

Uncertainty-Weighted Loss: In CNV region segmentation, the importance of each pixel may vary, especially for ambiguous boundaries, while assigning equal weights to all samples may not be optimal. To address this issue, uncertainty

maps are utilized to assign increased weights to pixels with higher levels of uncertainty. This, in turn, results in a more substantial impact on the update of the model parameters. Moreover, the incorporation of multiple tasks can generate different uncertainty weights for a single image, enabling a more comprehensive exploration of CNV boundary features via joint optimization. We employ a combination of loss functions, including binary cross-entropy (BCE) loss, mean squared error (MSE) loss, and Dice loss, to optimize the model parameters across all tasks. However, for the region shape regression task, we restricted the loss functions to only BCE and MSE. We incorporate uncertainty weights into the BCE loss by weighting each pixel to guide uncertainty on model training, i.e.,

$$\mathcal{L}_{UBCE} = -\frac{1}{N}\sum_{i=1}^{N}(1+V_i)\cdot\left[y_i \log\left(\hat{y}_i\right) + \left(1-y_i\right)\log\left(1-\hat{y}_i\right)\right], \tag{2}$$

where y_i and $\hat{y}_i$ are respective ground truth and prediction for pixel i. The total loss function can be expressed as $\mathcal{L} = \sum_t^T \lambda_t \mathcal{L}_t$, where $\mathcal{L}_t$ denotes the loss function for t^{th} task. λ_t denotes the loss weight, obtained by averaging the uncertainty map of the corresponding task and normalizing them to sum to 1.

3 Experimental Results

Dataset: The proposed RBGNet was evaluated on a dataset consisting of 74 OCTA images obtained using the Heidelberg OCT2 system (Heidelberg, Germany). All images were from AMD patients with CNV progression, captured in a 3×3 mm^2 area centered at the fovea. The enface projected OCTA images of the avascular complex were used for our experiments. All the images were resized into a resolution of 384×384 for experiments. The CNV areas and vessels were manually annotated by one senior ophthalmologist, and then reviewed and refined by another senior ophthalmologist. All images were acquired with regulatory approvals and patient consents as appropriate, following the Declaration of Helsinki.

Implementation Details: Our method is implemented based on the PyTorch framework with NVIDIA GeForce GTX 1080Ti. We train the model using an Adam optimizer with an initial learning rate of 0.0001 and a batch size of 4 for 300 epochs, without implementing a learning rate decay strategy. During training, the model inputs were subject to standard data augmentation pipelines, including random horizontal, vertical flips, random rotation, and random cropping. A 5-fold cross-validation approach is adopted to evaluate the performance.

Comparison with State-of-the-Arts: To benchmark our model's performance, we compared it with several state-of-the-art methods in the medical image segmentation field, including U-Net [15], CE-Net [18], CS-Net [19], TransUNet [20], and the backbone method U^{2}Net [16]. We use the Dice coefficient (Dice), intersection over union (IoU), false discovery rate (FDR), and area under the ROC curve (AUC) to evaluate the segmentation performance. The quantitative results are demonstrated in Table 1. Our results demonstrate that the

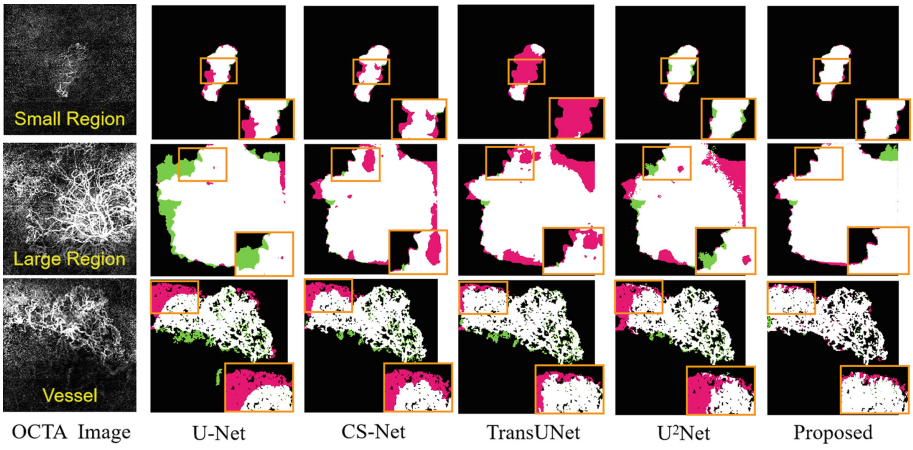

Fig. 3. Performance comparisons of different methods. Under-segmentation is shown in red, and over-segmentation is shown in green. (Color figure online)

Table 1. Performance comparisons for CNV segmentation.

Methods	Leison segmentation				Vessel segmentation			
	DICE	IoU	FDR	AUC	DICE	IoU	FDR	AUC
MF-Net [8]	0.8166	0.7142	0.1574	0.9084	0.7582	0.7191	0.1734	0.9062
U-Net [15]	0.8228	0.7290	0.1694	0.9148	0.8193	0.7280	0.1685	0.9175
CE-Net [18]	0.8690	0.7806	0.1173	0.9331	0.8372	0.7308	0.1459	0.9156
CS-Net [19]	0.8567	0.7689	0.1423	0.9365	0.8518	0.7622	0.1519	0.9378
TransUNet [20]	0.8414	0.7438	0.1254	0.9197	0.8301	0.7283	0.1423	0.9171
U^2Net [16]	0.8770	0.7930	0.1072	0.9383	0.8561	0.7662	0.0982	0.9214
Proposed	**0.9042**	**0.8328**	**0.0797**	**0.9464**	**0.9025**	**0.8229**	**0.0688**	**0.9455**

proposed method surpasses the existing state-of-the-art methods in both tasks. Specifically, our method achieved outstanding results on the test set for region segmentation, with a Dice of 90.42%, an IOU of 83.28%, and an AUC of 94.64%. The results in Table 1 indicate that superior vessel segmentation is positively associated with the model's ability to segment CNV regions.

The proposed method exhibits superior performance in precisely segmenting ambiguous boundaries of CNV regions, as demonstrated in the first two rows of Fig. 3. In contrast, existing state-of-the-art methods such as U-Net [15], and CS-Net [19] exhibit limitations in accurately segmenting complex and variable structures, leading to the misidentification of background structures as CNV regions. The illustrated quantitative results and performance comparisons serve as evidence of the proposed method's ability to simultaneously segment CNV regions and vessels with state-of-the-art performance. Accurate segmentation of

Table 2. Ablation results for CNV region segmentation and vessel segmentation.

Methods	Leison segmentation				Vessel segmentation			
	DICE	IoU	FDR	AUC	DICE	IoU	FDR	AUC
Backbone	0.8869	0.8051	0.1149	0.9473	0.8934	0.8160	0.0918	0.9401
Backbone + M1	0.8957	0.8194	0.1119	**0.9534**	0.8947	0.8177	0.0932	**0.9530**
Backbone + M2	0.9001	0.8260	**0.0856**	0.9448	0.8985	0.8183	0.0725	0.9322
Backbone + M3	**0.9042**	**0.8328**	0.0797	0.9464	**0.9025**	**0.8229**	**0.0688**	0.9455

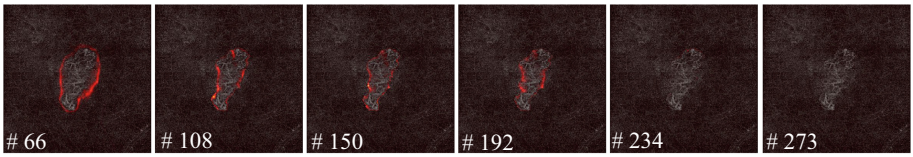

Fig. 4. Uncertainty maps during model training. #n represents the n^{th} epoch.

CNV region boundaries leads to improved vessel segmentation accuracy, which in turn refines the boundary of CNV regions.

Ablation Study: To validate the effectiveness of multi-task processing in enhancing CNV segmentation, we systematically integrated individual tasks into the multi-task training framework, by taking CNV region and vessel segmentation as the backbone. We successively incorporate M1: boundary prediction, M2: M1 and shape regression, and M3: M2 and uncertainty prediction, into the Backbone. The results for ablation are summarized in Table 2. We also illustrate in Fig. 4 the uncertainty probability maps generated by the proposed method during training. Note that the notation #n indicates the n^{th} training epoch, where the total number of training epochs is 300, and the model has specified to output the result in every 3 epochs. The best result is achieved at the $\#273^{th}$ epoch. Through Fig. 4, we can observe that the degree of uncertainty surrounding the segmented CNV regions diminishes gradually with the increased training epoch. These observations show the effectiveness of the uncertainty-weighted loss.

3.1 Conclusion

In summary, this study proposes a novel method to address the challenges of CNV segmentation in OCTA images. It incorporates a dual-branch encoder, multi-task optimization, and uncertainty-weighted loss to accurately segment CNV regions and vessels. The findings indicate that the utilization of cross-scale information, multi-task optimization, and uncertainty maps improve CNV segmentations. The proposed method exhibits superior performance compared to state-of-the-art methods, which suggests potential clinical implications for the diagnosis of CNV-related diseases. Nevertheless, further research is needed to validate the effectiveness of the proposed approach in large-scale clinical studies.

Acknowledgment. This work was supported in part by the National Science Foundation Program of China (62103398, 62272444), Zhejiang Provincial Natural Science Foundation of China (LZ23F010002, LR22F020008, LQ23F010002), in part by the Ningbo Natural Science Foundation (2022J143), and A*STAR AME Programmatic Fund (A20H4b0141) and Central Research Fund.

References

1. Friedman, D.S., et al.: Prevalence of age-related macular degeneration in the United States. Arch. Ophthalmol. **122**(4), 564–572 (2004)
2. Spaide, R.F., Klancnik, J.M., Cooney, M.J.: Retinal vascular layers imaged by fluorescein angiography and optical coherence tomography angiography. JAMA Ophthalmol. **133**(1), 45–50 (2015)
3. Jia, Y., et al.: Quantitative optical coherence tomography angiography of choroidal neovascularization in age-related macular degeneration. Ophthalmology **121**(7), 1435–1444 (2014)
4. Falavarjani, K.G., Al-Sheikh, M., Akil, H., Sadda, S.R.: Image artefacts in swept-source optical coherence tomography angiography. Br. J. Ophthalmol. **101**(5), 564–568 (2017)
5. Liu, L., Gao, S.S., Bailey, S.T., Huang, D., Li, D., Jia, Y.: Automated choroidal neovascularization detection algorithm for optical coherence tomography angiography. Biomed. Opt. Express **6**(9), 3564–3576 (2015)
6. Xue, J., Camino, A., Bailey, S.T., Liu, X., Li, D., Jia, Y.: Automatic quantification of choroidal neovascularization lesion area on OCT angiography based on density cell-like p systems with active membranes. Biomed. Opt. Express **9**(7), 3208–3219 (2018)
7. Wang, J., et al.: Automated diagnosis and segmentation of choroidal neovascularization in OCT angiography using deep learning. Biomed. Opt. Express **11**(2), 927–944 (2020)
8. Meng, Q., et al.: MF-Net: multi-scale information fusion network for CNV segmentation in retinal OCT images. Front. Neurosci. **15**, 743769 (2021)
9. Su, J., Chen, X., Ma, Y., Zhu, W., Shi, F.: Segmentation of choroid neovascularization in OCT images based on convolutional neural network with differential amplification blocks. In: Medical Imaging 2020: Image Processing, vol. 11313, pp. 491–497. SPIE (2020)
10. Gal, Y., Ghahramani, Z.: Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: International Conference on Machine Learning, pp. 1050–1059. PMLR (2016)
11. Bragman, F.J.S., et al.: Uncertainty in multitask learning: joint representations for probabilistic MR-only radiotherapy planning. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11073, pp. 3–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00937-3_1
12. Nair, T., Precup, D., Arnold, D.L., Arbel, T.: Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation. Med. Image Anal. **59**, 101557 (2020)
13. Dosovitskiy, A., et al.: An image is worth 16×16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)

14. Hatamizadeh, A., et al.: UNETR: transformers for 3D medical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 574–584 (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). https://doi.org/10.1007/978-3-319-24574-4_28
16. Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O.R., Jagersand, M.: U2-Net: going deeper with nested U-structure for salient object detection. Pattern Recogn. **106**, 107404 (2020)
17. Sedai, S., et al.: Uncertainty guided semi-supervised segmentation of retinal layers in OCT images. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 282–290. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_32
18. Gu, Z., et al.: CE-Net: context encoder network for 2D medical image segmentation. IEEE Trans. Med. Imaging **38**(10), 2281–2292 (2019)
19. Mou, L., et al.: CS-Net: channel and spatial attention network for curvilinear structure segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 721–730. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_80
20. Chen, J., et al.: TransUNet: transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)