Elongated Physiological Structure Segmentation via Spatial and Scale Uncertainty-Aware Network

Yinglin Zhang[1,2], Ruiling Xi[2], Huazhu Fu[4], Dave Towey[1], RuiBin Bai[1], Risa Higashita[2,3(✉)], and Jiang Liu[1,2,3(✉)]

[1] School of Computer Science, University of Nottingham Ningbo China, Ningbo 315100, China
[2] Research Institute of Trustworthy Autonomous Systems and Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China
rise@mail.sustech.edu.cn, liuj@sustech.edu.cn
[3] Tomey Corporation, Nagoya 451-0051, Japan
[4] Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore

Abstract. Robust and accurate segmentation for elongated physiological structures is challenging, especially in the ambiguous region, such as the corneal endothelium microscope image with uneven illumination or the fundus image with disease interference. In this paper, we present a spatial and scale uncertainty-aware network (SSU-Net) that fully uses both spatial and scale uncertainty to highlight ambiguous regions and integrate hierarchical structure contexts. First, we estimate epistemic and aleatoric spatial uncertainty maps using Monte Carlo dropout to approximate Bayesian networks. Based on these spatial uncertainty maps, we propose the gated soft uncertainty-aware (GSUA) module to guide the model to focus on ambiguous regions. Second, we extract the uncertainty under different scales and propose the multi-scale uncertainty-aware (MSUA) fusion module to integrate structure contexts from hierarchical predictions, strengthening the final prediction. Finally, we visualize the uncertainty map of final prediction, providing interpretability for segmentation results. Experiment results show that the SSU-Net performs best on cornea endothelial cell and retinal vessel segmentation tasks. Moreover, compared with counterpart uncertainty-based methods, SSU-Net is more accurate and robust.

Keywords: Uncertainty · Medical Image Segmentation · Elongated Physiological Structure · Deep Learning

1   Introduction

Robust and accurate elongated physiological structure segmentation is crucial for computer-aided diagnosis and quantification of clinical parameters [26,27]. Manual delineation is tedious and laborious. Recently, deep learning-based methods [14,16,19] have been proposed to delineate targets automatically. However, they are not able to outline correctly in ambiguous regions where exist uneven illumination, artifacts, or interference from the disease.

Many researchers have tried to use uncertainty information to concentrate on the ambiguous region, and to evaluate the reliability of model's prediction. According to the source of prediction errors [3], uncertainty is categorized into two types: epistemic and aleatoric. The main methods for uncertainty estimation are as follows. Bayesian neural networks [15] place a probability distribution over model weights, but are hard to optimize. Monte Carlo dropout [10] approximates the Gaussian process by embedding the dropout operation into the neural network layers and calculating the variance of several times inference. Deep Ensembles [8] combine the outputs from a group of independent models to estimate uncertainty. Softmax uncertainty [13,17,18] performs well in distinguishing examples that are easy or fallible to classify. Once the uncertainty information has been estimated, we are able to pay more attention to the ambiguous region. Xie et al. [24] used the cross-attention module to extract influential features for ambiguous regions based on pixel-level uncertainty. Yang et al. [25] achieved uncertainty awareness by training with a multi-confidence mask, and further used self-attention block with feature aware filter together to highlight uncertain areas. Wang et al. [23] annotated alpha matte for medical images and used it as a soft label to intuitively promote the network to focus on uncertain areas. Kohl et al. [7] proposed a generative model to produce multiple reasonable hypotheses for clinical experts to select from, which improved the diagnosis reliability. However, existing works applied the 'hard' attention to utilize uncertainty, which lacks the ability of adaptive adjustment and ignores neighboring uncertain regions. In addition, features at different scales contain rich structural and semantic contexts, which are essential for elongated physiological structure segmentation, such as cobweb corneal endothelial cells and retinal vessels.

This paper proposes a spatial and scale uncertainty-aware network (SSU-Net) for elongated physiological structure segmentation, which fully uses both spatial and scale uncertainty to highlight ambiguous regions and integrate hierarchical structure contexts. First, we use a gated soft uncertainty-aware (GSUA) module to adaptively highlight ambiguous areas based on spatial uncertainty maps. Second, we extract the uncertainty under different scales and propose the multi-scale uncertainty-aware (MSUA) fusion module to integrate hierarchical predictions for enhancing the final segmentation. Experiment results on segmentation tasks of the cornea endothelium and retinal vessel show the effectiveness of SSU-Net.
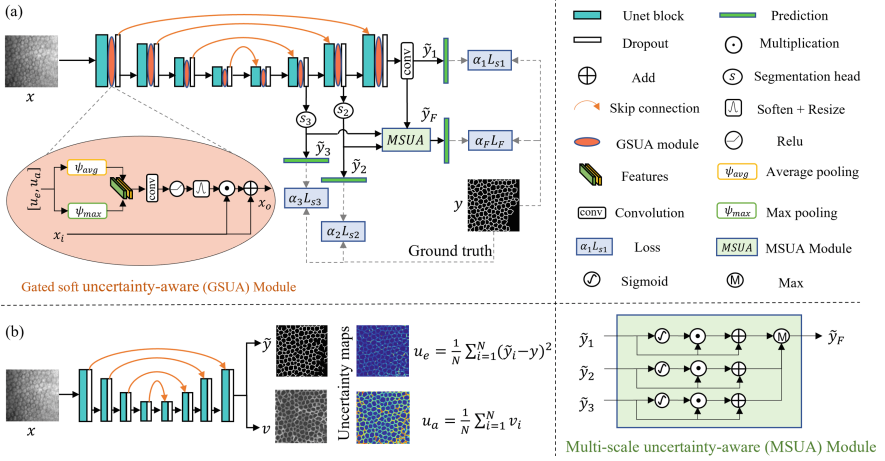
Fig. 1. The pipeline of the proposed algorithm. (a) The framework of spatial and scale uncertainty-aware network, SSU-Net. (b) We estimate the spatial uncertainty maps with Bayesian approximate network.

2   Method

Figure 1 (a) illustrates the framework of the proposed spatial and scale uncertainty-aware network, SSU-Net. The gated soft uncertainty-aware (GSUA) module enables the network to focus on the ambiguous region indicated by the spatial uncertainty maps $[u_e, u_a]$. Specifically, we construct a Bayesian approximate network to generate spatial uncertainty maps by introducing Monte Carlo dropout [2] into U-Net, as shown in Fig. 1 (b). The Bayesian approximate network has two outputs: segmentation prediction $\hat{y}$ and the estimation of aleatoric uncertainty v. We can calculate the epistemic and aleatoric uncertainty maps, u_e and u_a, after multiple inferences. Furthermore, we consider the sigmoid probabilities of predictions under different scales as the second uncertainty source, and fuse the predictions $\{\hat{y}_1, \hat{y}_2, \hat{y}_3\}$ from multiple scales using the multi-scale uncertainty-aware (MSUA) module. $\hat{y}_F$ is the final target output.

2.1   Spatial Uncertainty and Gated Soft Uncertainty-Aware Module

Spatial Uncertainty. Since the epistemic and aleatoric uncertainty maps are used to find the hard-to-classify spatial areas in this work, we regard them as spatial uncertainty. Referring to [6], we add dropout after each UNet block to approximate the Bayesian network, which learns the segmentation $\tilde{y}$ and aleatoric uncertainty v simultaneously. During inference, we sample a group of predictions $\{\tilde{y}_i\}_{i=1}^N$ and $\{v_i\}_{i=1}^N$ by N stochastic forward pass. In this work, we set $N = 16$. The epistemic u_e and aleatoric u_a uncertainty are formulated by

Eq. (1), where y is the ground truth.

$$u_e = \frac{1}{N}\sum_{i=1}^{N}(\tilde{y}_i - y)^2, u_a = \frac{1}{N}\sum_{i=1}^{N} v_i \qquad (1)$$

Gated Soft Uncertainty-Aware Module. To endow the uncertainty-aware module with adaptive adjustment ability, we propose the gated soft uncertainty-aware (GSUA) module, as illustrated in Fig. 1. We extract salient descriptions from uncertainty maps by two parallel pooling $[\psi_{avg}, \psi_{max}]$ and a 1×1 convolution $f(\cdot)$ operation. The *relu* operation is set as a switch to filter out areas with small uncertainty values, further strengthening our attention on areas with high uncertainty. Since it is also usually difficult to classify the area adjacent to the high-uncertainty regions, we use the Gaussian kernel to soften the boundary in such regions. The GSUA module is formulated by:

$$x_o = x_i \odot g_s(\sigma(relu(f([\psi_{avg}(\mathbf{u}), \psi_{max}(\mathbf{u})])))) + x_i \qquad (2)$$

where $x_i, x_o \in R^{N \times c \times h \times w}$ are the input and output features respectively; $\mathbf{u} = [u_a, u_e] \in R^{N \times 2 \times H \times W}$ is a tensor of uncertainty maps; ψ_{avg} and ψ_{max} represent average and max pooling; σ is the sigmoid function; g_s denotes a convolution operation with Gaussian kernel and resizes the attention maps to the size of input features; $\odot$ is element-wise multiplication.

2.2 Scale Uncertainty and Multi-scale Uncertainty-Aware Module

Scale Uncertainty. To integrate the predictions from hierarchical layers during model training, we capture the uncertainty under multiple scales. The sigmoid function is a simple and effective way to estimate uncertainty for the binary classification task. We extract the multi-scale uncertainty by Equation (3), where u_s is the uncertainty map of prediction $\tilde{y}_s$ under scale $s \in \{1, 2, 3\}$.

$$u_s = \frac{1}{1 + e^{-\tilde{y}_s}} \qquad (3)$$

Multi-scale Uncertainty-Aware Module. With the uncertainty maps from different scales, all the hierarchical predictions $\{\tilde{y}_1, \tilde{y}_2, \tilde{y}_3\}$ are fused by the MSUA module to generate the enhanced prediction $\tilde{y}_F$, as illustrate in Fig. 1. The uncertainty map u_s provides the classification confidence for each pixel. Therefore, we use u_s to highlight the confident region of $\tilde{y}_s$ and further extract the max value across the different scales. The process is formulated by:

$$\tilde{y}_F(i, j) = \max_{s \in \{1,2,3\}} (y_s(i, j) \odot \sigma(y_s(i, j)) + y_s(i, j)) \qquad (4)$$

where $\tilde{y}_F(i, j)$ denotes the pixel value at location (i, j) of enhanced prediction; $y_s(i, j)$ is the value of prediction under scale $s \in \{1, 2, 3\}$; and σ denotes the sigmoid operation of Eq. (3).

2.3   Objective Function

As shown in Fig. 1, we optimize the model with supervision on four segmentation branches simultaneously, including supervision for predicting three scales and the final enhanced output. The loss function is summarized as follows:

$$\mathcal{L}_{total} = \alpha_1 \mathcal{L}_{s1} + \alpha_2 \mathcal{L}_{s2} + \alpha_3 \mathcal{L}_{s3} + \alpha_F \mathcal{L}_F \qquad (5)$$

where $\alpha_1, \alpha_2, \alpha_3$, and α_F are the weight parameters for sub loss $\mathcal{L}_{s1}, \mathcal{L}_{s2}, \mathcal{L}_{s3}$, and $\mathcal{L}_F$. In this experiment, we set all the weight parameters as 1. For these sub-losses, we adopt binary cross-entropy loss, as shown in Eq. (6).

$$\mathcal{L} = -[ylog\tilde{y} + (1-y)log(1-\tilde{y})] \qquad (6)$$

where y is the ground truth; the positive class value of each pixel is 1; the negative class value is 0; and $\tilde{y} \in (0,1)$ is the predicted probability value.

3   Experiment

3.1   Dataset

Two cornea endothelium microscope image datasets, TM-EM3000 and Rodrep, and one retinal fundus image dataset, FIVES, are used in this work. The private dataset **TM-EM3000** contains 183 images measured by EM3000 specular microscope (Tomey Corporation, Japan). Following Ruggeri et al. [20], we cropped a 192×192 pixels sub-region from its 260×480 pixels whole image. We used 155 images for model training, ten images for validation, and 18 images for testing. **Rodrep** [21] contains 52 in-vivo confocal corneal microscope images, from 23 Fuchs patients with endothelial corneal dystrophy. We used 40 for training, five images for validation, and seven for testing. **FIVES** [5] is the largest known high-resolution fundus image dataset: It covers normal eyes and three different eye diseases with a balanced distribution. There are 800 high-resolution images and the corresponding manual annotations, with 550 for training, 50 for validation, and 200 for testing.

3.2   Evaluation Metrics and Implementation Details

There is a class imbalance between foreground and background pixels. To better evaluate the segmentation performance, we choose $Dice$ score [22], $mIoU$, and $mAcc$ as evaluation metrics. We optimized the models using the RMSprop strategy with momentum $= 0.9$ and weight decay $= 1e-8$ for 100 epochs. The initial learning rate was 2e-4, and the input size of all networks was uniformly set to 256×256. Random shift and rescaling within a range of $[-0.3, +0.3]$ were used for data augmentation. We set the batch size to 1 based on our empirical observations. For uncertainty-based models, we set the dropout rate as 0.5 and no data augmentation. During testing, we inferred $N = 16$ times and obtained the final prediction $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} \tilde{y}_F^i$ and the epistemic uncertainty $u_e' = \frac{1}{N}\sum_{i=1}^{N}(\tilde{y}_F^i - \bar{y})^2$.

3.3   Ablation Study

We investigated the influence of GSUA and MSUA modules on TM-EM3000, as shown in Table 1. The MSUA increased performance by 0.69% on the Dice score, and GSUA increased by 0.19%. The MSUA module brought more improvement than GSUA, which indicated that multi-scale context is crucial for cornea endothelium cell segmentation. When using both GSUA and MSUA modules simultaneously, we achieved the best performance.

Table 1. Ablation study on TM-EM3000. GSUA denotes the gated soft uncertainty-aware module. MSUA means the multi-scale uncertainty-aware fusion module.

Models	GSUA	MSUA	Dice(%) ↑	mIoU(%) ↑	mAcc(%) ↑
Variant1	✗	✗	76.04	76.32	85.52
Variant2	✗	✓	76.73	76.68	**87.38**
Variant3	✓	✗	76.23	76.42	86.08
Variant4	✓	✓	**77.16**	**77.02**	87.34

3.4   Comparison with State-of-the-Art Methods

To study the effectiveness of the proposed SSU-Net, we compared it with a series of state-of-the-art methods. On TM-EM3000 and Rodrep, we implemented several popular networks for comparison: UNet [19], D-LinkNet [28], AttentionUNet [16], TransUNet [1], and uncertainty-based counterparts, Monte Carlo (MC) BayesianNet [2], Lee's method [9]. On the fundus image dataset FIVES, we additionally implemented several recent retinal vessel segmentation algorithms: FR-UNet [12], SA-UNet [4], and IterNet [11].

As shown in Table 2, the proposed SSU-Net achieved the best performance. On the TM-EM3000 dataset, the uncertainty-based methods outperformed the typical convolution and attention methods, which proves that introducing uncertainty is beneficial. On the Rodrep dataset, SSU-Net performed considerably better than Lee's uncertainty method, improving the Dice score by 4.98%, mIoU by 3.48%, and mAcc by 3.14%. The results further suggest that the multi-scale predictions fusion module is crucial to elevate the robustness. According to the indication of uncertainty map u'_e, we cropped and zoomed in two ambiguous regions of each image, as shown in Fig. 2. The visualization results suggested that the proposed SSU-Net effectively improved the segmentation performance in ambiguous regions. On the FIVES dataset, the performance of the specialized network for retinal blood vessel segmentation in the fundus was similar to that of UNet, TransUNet, and AttettionUNet. The uncertainty-based methods are uniformly significantly superior to the above methods. The proposed SSU-Net achieved the best performance, increasing the Dice score by 10.08%, mIoU by 7.29%, and mAcc by 7.51% compared with UNet. Qualitative analysis is shown in Fig. 3, further supporting the conclusions of quantitative analysis.

Table 2. Comparison of SSU-Net with some SOTA methods on cornea endothelial cell and retinal vessel segmentation tasks.

| Dataset	Models	Dice (%) ↑	mIoU (%) ↑	mAcc (%) ↑
TM-EM3000	UNet	71.28	72.67	81.34
	D-LinkNet	72.71	73.60	82.67
	AttentionUNet	72.29	73.46	82.06
	TransUNet	71.65	72.78	82.20
	BayesianNet	75.42	75.71	85.66
	Lee's	76.40	76.53	85.80
	SSU-Net	**77.16**	**77.02**	**87.34**
Rodrep	UNet	64.89	67.89	78.50
	D-LinkNet	65.40	68.22	79.21
	AttentionUNet	60.55	65.68	74.45
	TransUNet	66.56	68.87	80.43
	BayesianNet	65.62	68.15	80.24
	Lee's	63.51	66.62	79.44
	SSU-Net	**68.49**	**70.10**	**82.58**
FIVES	UNet	78.99	82.59	86.05
	D-LinkNet	73.45	78.03	82.89
	AttentionUNet	78.09	81.82	85.26
	TransUNet	80.81	83.19	88.18
	FR-UNet	78.47	82.09	85.51
	SA-UNet	79.15	82.05	88.27
	IterNet	79.32	82.54	85.56
	BayesianNet	88.70	89.65	93.01
	Lee's	88.79	89.70	93.37
	SSU-Net	**89.07**	**89.88**	**93.56**
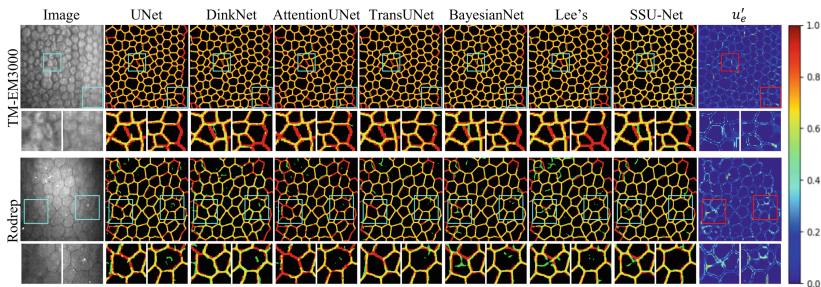
Fig. 2. Visualization for cornea endothelial cell segmentation. Red, green, and yellow line presents manual label, predicted segmentation result, and their overlap region. Indicated by the uncertainty map u'_e of the SSU-Net, we zoomed in two ambiguous local regions for clear observation. (Color figure online)
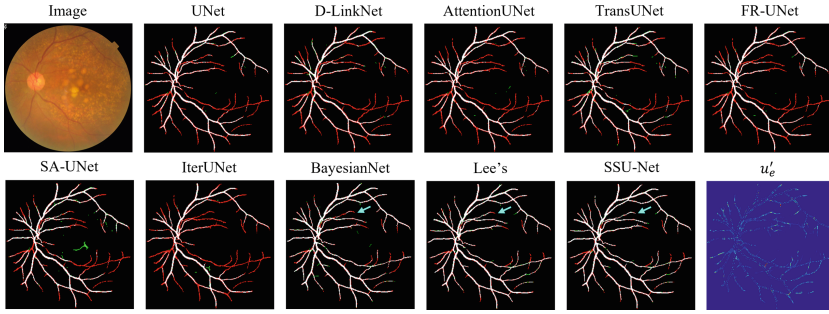
Fig. 3. Visualization results for retinal vessel segmentation on FIVES. Red, green, and yellow line presents manual label, predicted segmentation result, and their overlap region. The cyan arrow points to a sub-region with differences among uncertainty-based methods. We also visualize the epistemic uncertainty map u'_e of our SSU-Net. (Color figure online)

4   Conclusion

This paper proposes a spatial and scale uncertainty-aware network (SSU-Net) for elongated physiological structure segmentation. The ablation study shows the effectiveness of core components: the soft gated uncertainty-aware (GSUA) and the multi-scale uncertainty-aware (MSUA) fusion modules. Compared with some SOTA methods on cornea endothelial cell and retinal vessel image segmentation tasks, the proposed SSU-Net achieved the best segmentation performance and is more robust than other uncertainty-based methods. It is noteworthy that the SSU-Net performed considerably better than specialized retinal vessel segmentation networks. In the future, we plan to conduct experiments on various challenging situations to further explore the characteristics of SSU-Net.

References

1. Chen, J., et al.: Transunet: transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
2. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: Balcan, M.F., Weinberger, K.Q. (eds.) Proceedings of ICML. Proceedings of Machine Learning Research, vol. 48, pp. 1050–1059. PMLR, New York, New York, USA (20–22 Jun 2016)

3. Gawlikowski, J., et al.: A survey of uncertainty in deep neural networks. arXiv preprint arXiv:2107.03342 (2021)
4. Guo, Changlu, et al.: Sa-unet: spatial attention u-net for retinal vessel segmentation. In: Proceedings of ICPR, pp. 1236–1242 (2021)
5. Jin, K., et al.: Fives: a fundus image dataset for artificial intelligence based vessel segmentation. Sci. Data **9**(1), 475 (2022)
6. Kendall, A., Gal, Y.: What uncertainties do we need in Bayesian deep learning for computer vision? Proc. NeurIPS **30** (2017)
7. Kohl, S., et al.: A probabilistic u-net for segmentation of ambiguous images. Proc. of NeurIPS **31** (2018)
8. Lakshminarayanan, B., et al.: Simple and scalable predictive uncertainty estimation using deep ensembles. Proc. of NeurIPS **30** (2017)
9. Lee, J., et al.: Method to minimize the errors of AI: quantifying and exploiting uncertainty of deep learning in brain tumor segmentation. Sensors **22**(6), 2406 (2022)
10. Leibig, C., et al.: Leveraging uncertainty information from deep neural networks for disease detection. Sci. Rep. **7**(1), 1–14 (2017)
11. Li, L., et al.: Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3656–3665 (2020)
12. Liu, W., et al.: Full-resolution network and dual-threshold iteration for retinal vessel and coronary angiograph segmentation. IEEE J. Biomed. Health Inform. **26**(9), 4623–4634 (2022)
13. Mehrtash, A., et al.: Confidence calibration and predictive uncertainty estimation for deep medical image segmentation. IEEE Trans. Med. Imaging **39**(12), 3868–3878 (2020)
14. Mou, L., et al.: Cs2-net: deep learning segmentation of curvilinear structures in medical imaging. Med. Image Anal. **67**, 101874 (2021)
15. Neal, R.M.: Bayesian learning for neural networks. IEEE Trans. Neural Netw. (1994)
16. Oktay, O., et al.: Attention u-net: learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
17. Pearce, T., et al.: Understanding softmax confidence and uncertainty. arXiv preprint arXiv:2106.04972 (2021)
18. Pidaparthy, H., et al.: Automatic play segmentation of hockey videos. In: Proceedings of CVPR, pp. 4585–4593 (2021)
19. 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
20. Ruggeri, A., et al.: A system for the automatic estimation of morphometric parameters of corneal endothelium in Alizarine red-stained images. Br. J. Ophthalmol. **94**(5), 643–647 (2010)
21. Selig, B., et al.: Fully automatic evaluation of the corneal endothelium from in vivo confocal microscopy. BMC Med. Imaging **15**(1), 1–15 (2015)
22. Taha, A.A., Hanbury, A.: Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med. Imaging **15**(1), 1–28 (2015)
23. Wang, L., et al.: Medical matting: a new perspective on medical segmentation with uncertainty. In: Proceedings of MICCAI, pp. 573–583 (2021)
24. Xie, Y., et al.: Uncertainty-aware cascade network for ultrasound image segmentation with ambiguous boundary. In: Proceedings of MICCAI, pp. 268–278 (2022)

25. Yang, H., et al.: Uncertainty-guided lung nodule segmentation with feature-aware attention. In: Proceedings of MICCAI, pp. 44–54 (2022)
26. Zhang, Y., et al.: A multi-branch hybrid transformer network for corneal endothelial cell segmentation. In: Proceedings of MICCAI, pp. 99–108 (2021)
27. Zhao, Y., et al.: Automated tortuosity analysis of nerve fibers in corneal confocal microscopy. IEEE Trans. Med. Imaging **39**(9), 2725–2737 (2020)
28. Zhou, L., et al.: D-linknet: linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In: Proceedings of CVPR, pp. 182–186 (June 2018)