

MEIBOMIAN GLANDS SEGMENTATION IN NEAR-INFRARED IMAGES WITH WEAKLY SUPERVISED DEEP LEARNING

Xiaoming Liu^{a, b}, Shuo Wang^{a, b}, Ying Zhang^c

^aSchool of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, 430065, China

^bHubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan, 430065, China

^cWuhan Aier Eye Hospital, Wuhan, China

ABSTRACT

Near-infrared imaging is currently the most effective clinical method for evaluating the morphology of the meibomian glands in patients. Meibomian gland dysfunction (MGD) is a chronic and diffuse disease of the meibomian glands, which is an important cause of eye diseases such as dry-eye and blepharitis. Therefore, it is important to monitor the gland-drop and gland morphology for MGD patients. In this paper, we proposed a new scribble-supervised deep learning method for segmenting the meibomian glands. The proposed segmentation network consists of two stages. The first stage uses the U-Net network to obtain the meibomian region segmentation map. The second stage focuses on the meibomian region, combining spatial attention, gradient map and label filtering to generate the meibomian gland segmentation results. Experimental results on a local meibomian gland dataset demonstrate the effectiveness of the proposed segmentation framework.

Index Terms—Near-infrared imaging, Meibomian gland dysfunction, meibomian gland segmentation, scribble-supervised, spatial attention

1. INTRODUCTION

The meibomian glands are a row of greatly enlarged sebaceous glands that run along the lid margin posterior to the lid lashes. They are essential for the maintenance of the structural and refractive integrity of the ocular surface [1, 2]. Meibomian gland dysfunction (MGD) is a diffuse disease of the meibomian glands and a major cause of eye diseases including dry eye and blepharitis [3, 4]. Epidemiological studies have shown that the prevalence of MGD has exceeded 35.9% globally, the prevalence of different races is between 21.2%-71.0% [5]. In this paper, we focus on meibomian glands segmentation of near-infrared imaging which is currently the most effective clinical method for evaluating the prognosis of the meibomian gland morphology [6]. Doctors use the meibomian gland images to observe the morphology and drop of the glands as important references for the diagnosis of MGD [7]. However, it is difficult for doctors to directly judge the patient's pathology through experience.

In response to the above problems, Koh et al. [8] used Gabor wavelets to extract width, length and curvature in the meibomian gland image. But the method was only applied to the analysis of upper eyelid images and was likely to cause

misdetetection. Arita et al. [9] used Wallis filter to enhance areas with low contrast in image processing, but the classic Wallis filter featured blocking effects. Prabhu et al. [10] used fine-tuning U-Net for the automatic segmentation of meibomian glands. Khan et al. [11] proposed an adversarial learning based method. However, there are not many methods for automatic segmentation of meibomian glands. A large number of glands and irregular boundaries result in a lot of labor and time costs for pixel-based labeling in the training dataset.

Recently, deep learning has been widely applied to medical image processing [12-16]. Weakly supervised and semi-supervised learning have gradually been studied in natural images and medical images segmentation tasks [17-20]. For samples with noisy labels, the model often predicts greater fluctuations during the training process. An idea is to average the model to remove fluctuations. Tarvainen et al. [21] used exponential moving average (EMA) of model parameters during training. Nguyen et al. [22] average the predictions of the teacher model on different epochs. The gradient and contour information of the image can assist segmentation [23, 24].

Inspired by the Scribble2Label [20], in this paper, we proposed two-stage scribble supervision of meibomian glands segmentation. Generally, 5-8 minutes are required to label glands with polygons for each image even for an expert, but it only takes 1-2 minutes to annotate an image with scribble annotations.

Considering that the glands are distributed in the region of the meibomian glands, we adopt a two-stage segmentation method. To make reasonable use of the gradient information of the original image, we adjusted the U-Net network. The curvature of the gland shape and irregular edges make the segmentation challenging. In order to better obtain the spatial information of the feature, we use the spatial attention block to optimize the network.

The contributions of the paper include:

(1) We proposed a novel two-stage scribble supervised segmentation approach for meibomian glands in near-infrared images. To the best of our knowledge, this is the first time that weak supervision has been used for meibomian gland segmentation.

(2) We gradually expand the pseudo-label according to the reliability of the predicted feature and optimize the network in combination with the spatial attention to obtain more accurate predictions.

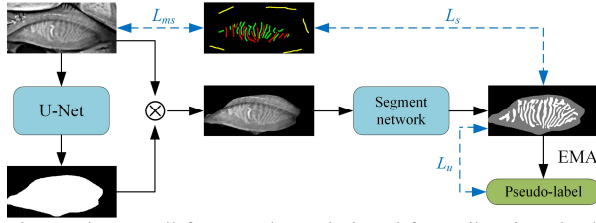


Fig. 1. The overall framework we designed for meibomian glands segmentation of near infrared images. The actual scribble thickness used in our experiment was 1 pixel, but it is widened to 5 pixels in this figure for better visualization. (Green: glands, red: meibomian region, yellow: background, black: unlabeled region)

(3) We utilize the gradient information of the original image at the decoder part of the segmentation network, which can coarsely locate the target contour and help segmentation.

2. THE PROPOSED APPROACH

In this section, we explain the proposed method for segmentation of the meibomian region and meibomian glands in detail. With the prior knowledge that the meibomian glands are in the meibomian region, we tackle the problem with a two-stage segmentation network (see Fig. 1).

In the first stage, we train the U-Net network to coarse segment the meibomian region with scribble annotations using only the scribbled pixel loss. In the second stage, we focus the segmentation task on the meibomian region. To optimize the segment network, spatial attention blocks are incorporated, which help the network pay more attention to the target region. For scribbled pixels, a cross-entropy loss is applied. For unscribbled pixels, our network automatically generates reliable labels using the exponential moving average of the predictions during training and filters the pseudo-label by uncertainty threshold.

2.1. Coarse segmentation of meibomian region

We use the U-Net network as our first stage segmentation network architecture, which aims to coarsely segment the meibomian region. The input sources for this stage are the original image and the user-given scribbles (see Fig. 1). Here, the given scribbles are labeled pixels (denoted as red and yellow for the meibomian region and background, respectively), and the rest of the pixels are unlabeled pixels (denoted as black). For scribbled pixels, a standard cross-entropy loss is applied as

$$L_{ms} = -\frac{1}{N} \sum_{k \in N} [s_k \log(p_k) + (1-s_k) \log(1-p_k)] \quad (1)$$

where N is a set of scribbled pixels, s_k is the label of the scribble pixel k and p_k is the prediction of the pixel k .

It is observed with only scribble annotations, the meibomian region can be successfully segmented.

2.2. Fine segmentation of meibomian glands

In this part, we will introduce the fine segmentation of the

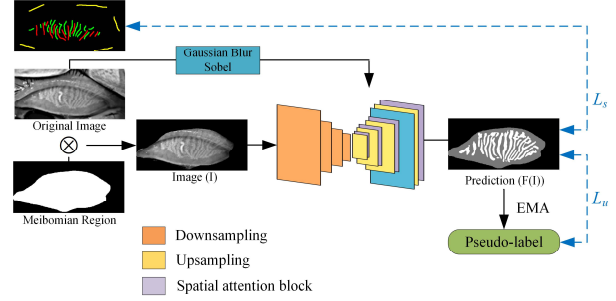


Fig. 2. The framework we designed for fine segmentation of meibomian glands. The purple block is the spatial attention block, the orange block is the down-sampling module, the yellow block is the up-sampling module, and the blue block is the gradient map.

meibomian glands in detail. First, we use the segmentation result of the first stage as a mask, do a dot multiplication with the original image to obtain the image I as the input of the second stage to concentrate the gland segmentation task on the meibomian region (see Fig. 2). We dilate the segmentation result in the first stage by two pixels to prevent the meibomian region from shrinking too much. Then we feed the image I into the segment network F to obtain the prediction ($F(I)$).

The spatial attention blocks are incorporated into the U-Net architecture to highlight salient features that are passed through the skip connections (see Fig. 2 purple blocks). The final decoder of the network will be superimposed with the gradient map of the original image (see Fig. 2 blue blocks). In the initial training stage, we will compare the prediction map with only labeled pixels, then compare the predictions with unlabeled reliable labels obtained using an uncertainty threshold. To make the predictions more reliable and not fluctuate greatly due to a less accurate prediction, the whole process uses the EMA method to ensemble the predictions.

Spatial attention block. With the reduction of the spatial resolution of feature maps, small objects that show large shape variability are difficult to discriminate and the features may be lost in the high-level features [25]. Therefore, it is difficult to learn discriminative features for glands by using traditional networks, resulting in the performance degradation. We use spatial attention to optimize the network, make the network pay more attention to the objects, assist in identifying the glands, and generate higher quality pseudo labels.

We use x^l to represent the low-level feature map at the scale m in the encoder, and x^h to represent the high-level feature map with lower spatial resolution upsampled at the $m+1$ scale from the end of the decoder. As shown in Fig. 3, x^h and x^l are compressed by a $1 \times 1 \times C_h$ and $1 \times 1 \times C_l$ convolution respectively. Then their sum is normalized by the softmax activation function. The obtained feature map is fed into another $1 \times 1 \times 1$ convolution followed by a Sigmoid function to obtain a pixel-wise attention coefficient $\alpha \in [0, 1]^{H \times W}$. Finally, x^l is multiplied with α .

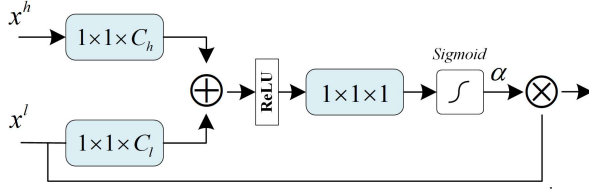


Fig. 3. Details of the spatial attention block. Input features (x^l) are scaled with attention coefficients (α) which are calculated in spatial attention block.

Gradient map. It can be seen in Fig. 4 that the gradient map of the original image can coarsely locate the target contour and help segmentation. Hence, we use the Gaussian Blur and Sobel operator to get the gradient map and superimpose it with the final decoder in the U-Net network.

Scribble Supervision. Considering that the gland is a long strip, we use scribble annotations. In detail, the labels are manual long lines, that is we use scribble-supervised learning. The given scribbles are labeled pixels (denoted as green and red for the glands and meibomian region, respectively, as shown in Fig. 2), and the rest of the pixels are unlabeled pixels (denoted as black). For scribbled pixels, the proposed network is trained with cross entropy loss as

$$L_s = -\frac{1}{N} \sum_{j \in N} \sum_{c=1}^C s_j^c \log(p_i^c) \quad (2)$$

where N is a set of scribbled pixels, C is the number of categories of object (in our case, $C=2$), s_j^c is an indicator variable to determine whether the category of the scribbled pixel j is the same as the category c , and p_i^c is the prediction of category c at iteration i .

EMA. For unscribbled pixels, in order to automatically generate more reliable labels, our network uses exponential moving averages (EMA) for predictions during the training process. To achieve the valuable ensemble and reduce computational costs, we regularly calculate the EMA of the predictions over the training process every γ (ensemble momentum) epoch: $y_n = \alpha p_i + (1-\alpha)y_{n-1}$, where y is the average of predictions, $y_0 = p_1$, α is the EMA weight, and n is how many times the predictions are averaged.

Filtered pseudo-label. After E iterations, we learn with the filtered pseudo-label which is generated by EMA. For the generated pseudo-labels, we estimate the uncertainty [26] of the predictions for each pixel, $u = -\sum_{c=1}^C p_i^c \log p_i^c$, then use an uncertainty threshold H to filter the labels, the unscribbled pixel loss is defined as

$$L_u = -\frac{1}{G} \sum_{x \in G} [I(u_x < H) y_n' \log(p_i)] \quad (3)$$

where G is a set of generated pseudo label pixels, I is a indicated function, u_x is the estimated uncertainty u of the

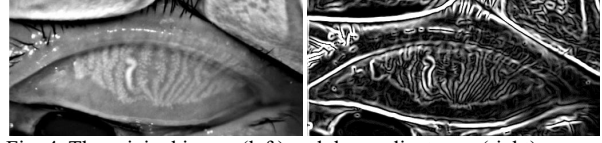


Fig. 4. The original image (left) and the gradient map (right).

unscribbled pixel x and $y_n' = \arg\max(y_n)$. It is noted that with the training process, the uncertainty of initial unlabeled pixels will decrease, and more pixels will be added into the pseudo-labeled training set.

2.3. Training

In this paper, we adopt a two-stage training scheme for meibomian glands segmentation, and the overall training procedure is presented below. In each stage, we use the model with the highest IoU on the validation set for testing.

First Stage. We firstly train a U-Net using scribble annotated meibomian glands datasets to minimize the loss L_{ms} in (1).

Second Stage. Once the segmentation result is output in the first stage, it will be combined with the original image as the input of the second stage, focusing the segmentation task on the meibomian region. We feed it into the second stage segmentation network and train it by minimize the scribbled loss L_s during the first few epoch E . Then train it by minimize the total loss, which is the combination of the scribbled loss L_s and the unscribbled loss L_u with the relative weight of μ , defined as

$$L = L_s + \mu L_u \quad (4)$$

3. EXPERIMENTATION AND RESULTS

3.1. Dataset

The proposed approach is evaluated on a local dataset which is collected from the Wuhan Aier Eye Hospital. The images come from 100 MGD adult subjects, and we selected 300 meibomian gland images (including upper and lower eyelids) with a size of 240×512 as the dataset. All images have scribbles and polygons labeled by clinical experts using ITK-SNAP [27].

3.2. Implementation details

The proposed approach is implemented on pytorch framework, the model is trained on the GTX750 GPU. The hyper-parameters used for our model are as follows:

uncertainty threshold $H = \frac{1}{4} u_{\max}$ (u_{\max} is the maximum uncertainty value, i.e., $\ln 3$ in our experiments); EMA Alpha $\alpha = 0.2$; Ensemble Momentum $\gamma = 5$; first few epoch $E = 50$; L_u 's weight $\mu = 0.5$. For the evaluation of our model generalization and performance, we randomly divide the dataset into 5 equal subsets at our datasets. Note that four of them are used for training and validating, the remaining 1 for testing. The final evaluation result is the average of the 5 folds.

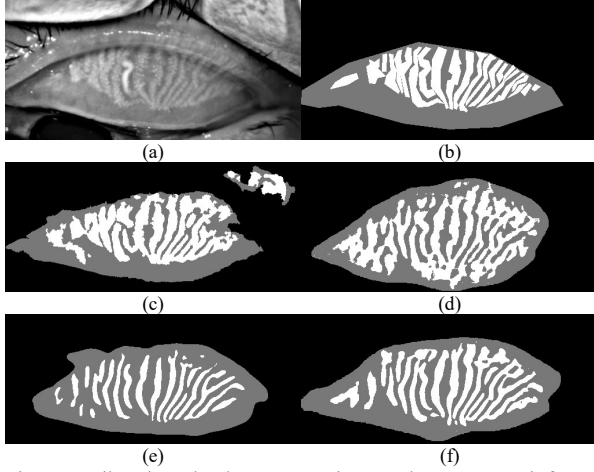


Fig. 5. Meibomian glands segmentation results on a near-infrared image with MGD. (a) the original image, (b) expert polygon annotation, (c) U-Net (Full supervision) [28], (d) CE only [17], (e) Scribble2Label [20], (f) the proposed method.

3.3. Comparative methods and metric

We compared the proposed method with three state-of-the-art methods which are U-Net [28], CE only [17], Scribble2Label [20].

For fair comparison, we implement the U-Net in PyTorch. Since the purpose of our first and second stages is to segment the meibomian glands, in the comparison method, we directly perform the three-class segmentation task on the original image. We utilize the IoU (Intersection-over-Union), mDice (mean Dice coefficient), AC (accuracy), SE (sensitivity), SP (specificity), AJI (Aggregated Jaccard Index) [29], which has been reported on medical image segmentation researches, as the comparative metric of these methods.

3.4. Results

Fig. 5 shows the segmentation results of the proposed method and the comparison method on a typical near-infrared image of MGD patients. Since the Meibomian gland image has many irregular gland shapes, it is one of the most difficult to segment in near-infrared images. It can be seen from Fig. 5 that when using the U-Net (full labeled) segmentation (Fig. 5(c)), although the segmentation effect is well in the meibomian region and the glands region, the glove part outside the meibomian area is also mistakenly segmented as the meibomian glands, which confirms to a certain extent the necessity of two-stage segmentation. Almost all the glands segmented by CE are fused together (Fig.5(d)), and the right part of the meibomian gland is missing. Scribble2Label is better than CE which may benefit from EMA and filtered pseudo-label approach, but it is insufficiently segmented in the gland part of the image (Fig.5(e)). When using the proposed method (Fig.5(f)), most single gland regions can be identified, and the gland segmentation can be concentrated in the meibomian gland region, it might be due to the spatial attention blocks and the gradient map.

Table 1. Quantitative comparison (in percentage) between the proposed approach and the comparative method by IoU, mDice, AC, SE, SP, AJI, the best results are bold.

Label	Method	IoU	mDice	AC	SE	SP	AJI
Full	U-Net [28]	76.7±0.1	78.5±0.1	84.3±0.7	93.1±2.2	82.6±1.6	27.5±0.8
	CE only [17]	63.1±0.2	67.4±0.4	68.5±1.0	87.2±3.1	67.9±2.2	18.8±1.8
	Scribble2Label [20]	68.7±0.2	71.6±0.2	73.5±0.7	90.0±2.6	71.7±1.6	56.3±0.9
	The proposed	70.2±0.1	72.9±0.2	74.2±0.5	90.2±2.5	69.6±1.5	41.9±0.5

The quantitative comparison of IoU, mDice, AC, SE, SP, AJI is shown in Table 1. The average value of each category is listed in the table. As shown in Table 1, compared with weakly-supervised methods, the fully-supervised U-Net performs better except for the AJI indicator. Compared with CE, Scribble2Label and our method have improved all indicators, which verifies the effectiveness of EMA method and pseudo-label filtering. The IoU and mDice values of the proposed method are 70.2% and 72.9%, respectively, which are better than other comparison weakly-supervised methods. The AJI of the proposed method is 41.9%, which is a 14.4% increase compared to the U-Net. Comparing our method with Scribble2Label, we can see that only the SP and AJI values are slightly weakened, which may be due to the more complete segmentation of the glands which will lead to certain connectivity. In summary, the proposed method can outperform other state-of-the-art weakly-supervised methods on meibomian gland segmentation.

4. CONCLUSIONS

In this paper, we proposed a weakly-supervised framework to segment the meibomian glands on the near-infrared image in two stages through strategies such as spatial attention, pseudo-label filtering, and effective use of gradient maps. As far as we know, this is the first time that weakly supervised deep learning has been used to segment the meibomian glands on near-infrared images. The framework is trained in two stages. The segmentation network is trained by using the scribble label data and the filtered pseudo-label to optimize the joint loss function. Experimental results on a local dataset show that the proposed method is superior to the latest weakly supervised segmentation method and is relatively close to the full supervision method of the meibomian glands.

REFERENCES

- [1] P. J. Driver and M. A. Lemp, "MAJOR REVIEW Meibomian Gland Dysfunction," vol. 40, no. 5, pp. 0-367, 1996.
- [2] T. H. Ko *et al.*, "Comparison of Ultrahigh- and Standard-Resolution Optical Coherence Tomography for Imaging Macular Pathology," *Ophthalmology*, vol. 111, no. 11, pp. 2033-2043, 2005.
- [3] N. J. Daniel *et al.*, "The International Workshop on Meibomian Gland Dysfunction: Report of the Definition and Classification Subcommittee," *Invest Ophthalmol Vis Sci*, vol. 52, no. 4, pp. 1930-1937, 2011.
- [4] K. Erich, K. Nadja, M. Thomas, O. Hiroto, and D. A. Sullivan, "The international workshop on meibomian gland dysfunction: report of the subcommittee on anatomy, physiology, and pathophysiology of the meibomian gland," *Investigative Ophthalmology & Visual Science*, vol. 52, no. 4, pp. 1938-78, 2011.
- [5] M. Varmaghani, "Global Prevalence of Meibomian Gland Dysfunction:

- A Systematic Review and Meta-Analysis," *Ocular Immunology and Inflammation*.
- [6] R. Arita, K. Itoh, K. Inoue, and S. Amano, "Noncontact Infrared Meibography to Document Age-Related Changes of the Meibomian Glands in a Normal Population," *Ophthalmology*, vol. 115, no. 5, pp. 911-915, 2008.
 - [7] G. N. Foulks and A. J. Bron, "Meibomian Gland Dysfunction: A Clinical Scheme for Description, Diagnosis, Classification, and Grading," *Ocular Surface*, vol. 1, no. 3, pp. 107-126, 2003.
 - [8] Y. W. Koh, T. Celik, H. K. Lee, A. Petznick, and L. H. Tong, "Detection of meibomian glands and classification of meibography images," *Journal of biomedical optics*, vol. 17, no. 8, p. 086008, 2012.
 - [9] R. Arita, J. Suehiro, T. Haraguchi, R. Shirakawa, H. Tokoro, and S. Amano, "Objective image analysis of the meibomian gland area," *British Journal of Ophthalmology*, vol. 98, no. 6, pp. 746-755, 2014.
 - [10] S. M. Prabhu, A. Chakiat, S. Shashank, K. P. Vunnavu, and R. Shetty, "Deep learning segmentation and quantification of Meibomian glands," *Biomedical Signal Processing and Control*, vol. 57, p. 101776, 2020.
 - [11] Z. K. Khan, A. I. Umar, S. H. Shirazi, A. Rasheed, and S. Gul, "Image based analysis of meibomian gland dysfunction using conditional generative adversarial neural network," *BMJ Open Ophthalmology*, vol. 6, no. 1, p. e000436, 2021.
 - [12] X. Liu, J. Cao, S. Wang, Y. Zhang, and M. Wang, "Confidence-Guided Topology-Preserving Layer Segmentation for Optical Coherence Tomography Images With Focus-Column Module," *IEEE Transactions on Instrumentation and Measurement*, vol. PP, no. 99, pp. 1-1, 2020.
 - [13] J. Tang, S. Guo, Q. Sun, Y. Deng, and D. Zhou, "Speckle reducing bilateral filter for cattle follicle segmentation," *Bmc Genomics*, 2010.
 - [14] X. Wang *et al.*, "A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT," *IEEE transactions on medical imaging*, vol. 39, no. 8, pp. 2615-2625, 2020.
 - [15] X. Liu *et al.*, "Automated Layer Segmentation of Retinal Optical Coherence Tomography Images Using a Deep Feature Enhanced Structured Random Forests Classifier," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, pp. 1404-1416, 2019.
 - [16] X. Liu, S. Wang, Y. Zhang, D. Liu, and W. Hu, "Automatic fluid segmentation in retinal optical coherence tomography images using attention based deep learning," *Neurocomputing*, 2021.
 - [17] M. Tang, F. Perazzi, A. Djelouah, I. Ben Ayed, C. Schroers, and Y. Boykov, "On regularized losses for weakly-supervised cnn segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 507-522.
 - [18] H. Kervadec, J. Dolz, S. Wang, E. Granger, and I. B. Ayed, "Bounding boxes for weakly supervised segmentation: Global constraints get close to full supervision," in *Medical Imaging with Deep Learning*, 2020, pp. 365-381: PMLR.
 - [19] X. Liu *et al.*, "Semi-supervised automatic segmentation of layer and fluid region in retinal optical coherence tomography images using adversarial learning," *IEEE Access*, vol. 7, pp. 3046-3061, 2018.
 - [20] H. Lee and W.-K. Jeong, "Scribble2Label: Scribble-Supervised Cell Segmentation via Self-generating Pseudo-Labels with Consistency," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2020, pp. 14-23: Springer.
 - [21] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Advances in neural information processing systems*, 2017, pp. 1195-1204.
 - [22] D. T. Nguyen, C. K. Mummadi, T. P. N. Ngo, T. H. P. Nguyen, L. Beggel, and T. Brox, "Self: Learning to filter noisy labels with self-ensembling," *arXiv preprint arXiv:1910.01842*, 2019.
 - [23] X. Liu, A. Yu, X. Wei, Z. Pan, and J. Tang, "Multimodal MR Image Synthesis Using Gradient Prior and Adversarial Learning," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1-1, 2020.
 - [24] X. Liu, Z. Guo, J. Cao, and J. Tang, "MDC-Net: A New Convolutional Neural Network for Nucleus Segmentation in Histopathology Images with Distance Maps and Contour Information," *Computers in Biology and Medicine*, p. 104543, 2021.
 - [25] X. Xi, X. Meng, Z. Qin, X. Nie, and X. Chen, "IA-net: Informative attention convolutional neural network for choroidal neovascularization segmentation in OCT images," *Biomedical Optics Express*, vol. 11, no. 11, p. 6122, 2020.
 - [26] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 605-613: Springer.
 - [27] P. A. Yushkevich, Y. Gao, and G. Gerig, "ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 3342-3345: IEEE.
 - [28] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234-241: Springer.
 - [29] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A Dataset and a Technique for Generalized Nuclear Segmentation for Computational Pathology," *IEEE Transactions on Medical Imaging*, pp. 1-1, 2017.