PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

IF-Net: information fusion network for meibomian gland area and atrophy area segmentation

Dengfeng Liu, Weifang Zhu, Xinyu Zhuang, Xiaofeng Zhang, Xinjian Chen, et al.

Dengfeng Liu, Weifang Zhu, Xinyu Zhuang, Xiaofeng Zhang, Xinjian Chen, Fei Shi, Dehui Xiang, "IF-Net: information fusion network for meibomian gland area and atrophy area segmentation," Proc. SPIE 12464, Medical Imaging 2023: Image Processing, 124643X (3 April 2023); doi: 10.1117/12.2654206

Event: SPIE Medical Imaging, 2023, San Diego, California, United States

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 13 Apr 2023 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

IF-Net: Information fusion network for meibomian gland area and atrophy area segmentation

Dengfeng Liu[1], Weifang Zhu[1], Dehui Xiang[1], Fei Shi[1], Xinyu Zhuang[3], Xiaofeng Zhang[4], Xinjian Chen[1,2,]*

[1]School of Electronics and Information Engineering, Soochow University, Suzhou, Jiangsu Province, 215006, China

[2]State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou, Jiangsu Province, 215123, China

[3]The First Affiliated Hospital of Soochow University, Suzhou, Jiangsu Province, 215006, China

[4]Department of Ophthalmology, Dushu Lake Hospital Affiliated to Soochow University, Suzhou, Jiangsu Province, 215000, China

*indicates the corresponding author, E-mail:xjchen@suda.edu.cn

ABSTRACT

Meibomian glands dysfunction (MGD) is the main cause of dry eyes. The degree of meibomian gland atrophy plays an important role in the clinical diagnosis of MGD. The automatic quantification of meibomian gland area (MGA) and meibomian gland atrophy area (MGAA) is challenging due to the blurred boundary and various shapes. A U-shaped information fusion network (IF-Net) is proposed for the segmentation of MGA and MGAA in this paper. The contributions of this paper are as follows: (1) An information fusion (IF) module is designed to fuse the context information from the spatial dimension and the channel dimension respectively, which effectively reduces the loss of information caused by continuous downsampling. (2) A parallel path connection (PPC) is proposed and inserted into skip connections. On one hand, it can suppress the noise of different levels of information. On the other hand, it can make up the lack of information via the original simple skip connection of U-Net. Our proposed IF-Net has been evaluated on 505 infrared MG images from 300 subjects and achieves the average Dice similarity coefficient (DSC) of 84.81% and the average intersection over union (IoU) of 74.44% on MGAA segmentation, which indicates the primary effectiveness of the proposed method.

Keywords: Dry eyes, meibomian gland dysfunction, spatial attention, channel attention, information fusion

1. INTRODUCTION

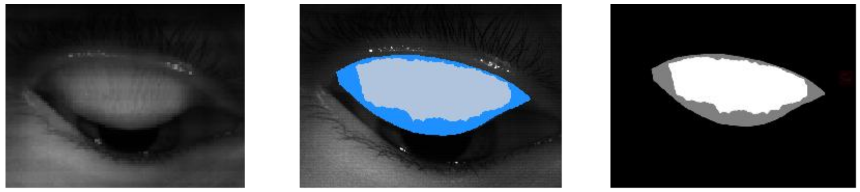
Meibomian gland dysfunction (MGD) is the main cause of dry eyes. Biological parameters of meibomian gland (MG) such as height, tortuosity and the degree of atrophy are closely related to its function[1]. In clinic, physicians need to manually draw the size of the meibomian area and the meibomian gland area (MGA) on the infrared MG image, and then subtract the two areas to obtain the meibomian gland atrophy area (MGAA), which is tedious and time-consuming. Doctors can estimate the degree of meibomian gland atrophy from images of infrared meibomian gland, but such a diagnosis is largely subject to subjective factors. With the rapid development of artificial intelligence, more and more intelligent technologies have been applied in medical image processing. Fig.1 shows an original infrared MG image and the ground truth.

Full convolutional network (FCN)[2] is widely used in image segmentation due to its excellent feature extraction capability. U-Net[3] greatly improves the performance of medical image segmentation by using jump connection and encoder-decoder structure. Some other U-Net based segmentation networks replace the convolution in the deep encoder layers with dilated convolution, which can maintain a large feature map while having a large receptive field. CE-Net[4] uses convolution branches with several different receptive fields to improve the ability of the model to obtain multi-scale information. The multi-scale weighted shared attention module proposed by GDCSeg-Net[5] can focus on information at different scales, and can integrate feature information with channel and spatial attention mechanism at multiple scales simultaneously, which can obtain the target feature information effectively. CS^2-Net[6] introduces a mixed attention mechanism to increase the focus on the shape of the target.

Medical Imaging 2023: Image Processing, edited by Olivier Colliot, Ivana Išgum,
Proc. of SPIE Vol. 12464, 124643X · © 2023 SPIE
1605-7422 · doi: 10.1117/12.2654206

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 13 Apr 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

The irregular shape of the meibomian gland regions increases the challenge of segmentation. In this paper, we propose a U-shaped information fusion network (IF-Net) for the segmentation of MGA and MGAA, in which the information fusion (IF) module is designed to overcome the loss of feature information caused by multiple downsampling and the parallel path connection (PPC) module is proposed and inserted into skip connections to alleviate the neglection of semantic information in the original skip connection of U-Net.



(a) (b) (c)

Figure 1: Original infrared MG image and ground truth. (a) Original infrared MG image. (b) Original infrared image with MGAA and MGA: MGA (gray), MGAA (blue). (c) Ground truth: MGA (white), MGAA (gray).

2. METHODS

In this section, we will introduce the proposed method in three parts: the overall structure of the network, IF module and PPC module.

2.1 Overall structure of the proposed network

In recent years, U-Net has greatly improved the performance of medical image segmentation by using skip connection and encoder-decoder structure. Although U-Net obtains good segmentation effect with its special encoder-decoder structure and novel skip connection mode, it still has the following deficiencies: (1) Although downsampling operation of feature makes the convolutional network have large receptive field, continuous downsampling will lead to the loss of contextual semantic information. (2) Skip connection can make up for the information loss caused by downsampling operation in the encoding process, and assist the decoder to recover the feature information. However, a simple skip connection will fuse the features extracted by the encoder with the adjacent feature information, which will introduce the background noise or useless information into the shallow features and affect the quality of the final decoder output feature.

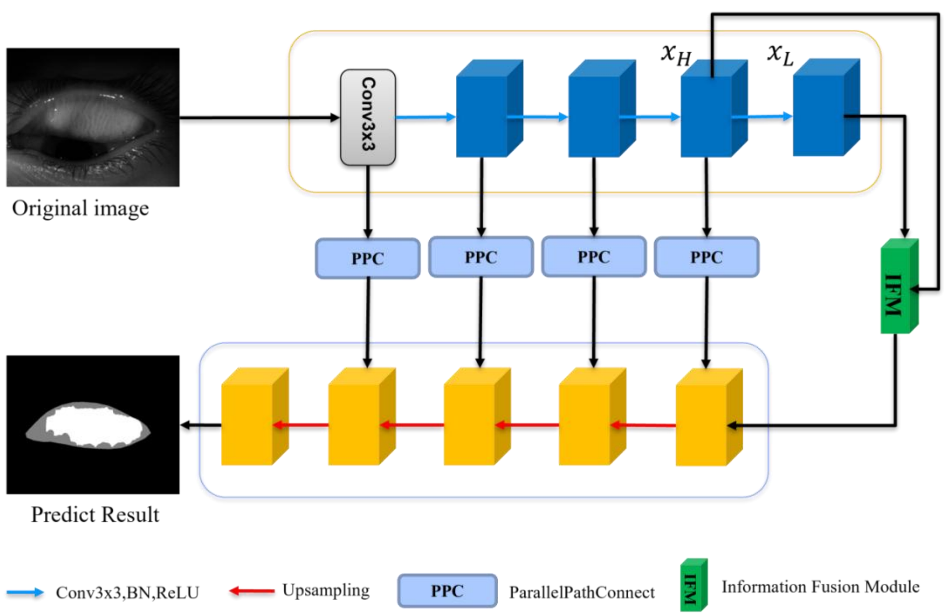


Figure 2. The overall structure of the IF-Net

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 13 Apr 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Fig. 2 illustrates the overall structure of the proposed IF-Net, which consists of four parts: encoder module, IF module, PPC module, and decoder module. The encoder module consists of four layers. In order to reduce the number of model parameters, the number of channels in each encoder layer is reduced to 1/2 of the ones in the original U-Net. Two 3x3 convolutions, a batch normalization and a Relu activation are performed after each downsampling operation. After each downsampling, the number of channels is doubled. The channels in the encoder path are 32, 64, 128, 256 and 512 respectively. The size of the output feature of the encoder path is 1/8 the size of the original input image. Between the last two layers of the encoder, the information fusion (IF) module is imbedded to fuse the contextual information to make up for the lack of information caused by continuous downsampling. In the skip connection, the parallel path connection (PPC) modules are designed and inserted to make up the lack of information via the original simple skip connection of U-Net.

2.2 Information Fusion (IF) Module

The structure of the IF module is as is shown in Fig. 3. The deep feature x_L and shallow feature x_H of the last two layers of the encoder are taken as the inputs of the IF module. We reduce the channel numbers of the two inputs to 1/4 of the original input by 1x1 convolution. The x_L is recovered to the same resolution of the x_H by upsampling and then added to x_H to obtain the fusion feature F_M, which has the richer semantic feature information. In order to obtain the global dependence relationship and establish the correlation between channels, we do not use the channel compression operation like SE-Net[7]. Although the operation can get the correlation between channels, it will also destroy the dependence relationship between the channels. A 1x1 convolution is adopted to realize the interaction between local channels and extract the dependencies. Finally, through the Sigmoid activation function, the result of channel fusion c_y can be obtained. Therefore, the channel fusion can be summarized as follows:

$$c_y = \sigma(Conv_{1x1}(Avg(Conv_{1x1}(x_H) + Up(Conv_{1x1}(x_L))))) \quad (1)$$

where σ represents the Sigmoid activation function, Up represents the upsampling and Avg represents the global average pooling.

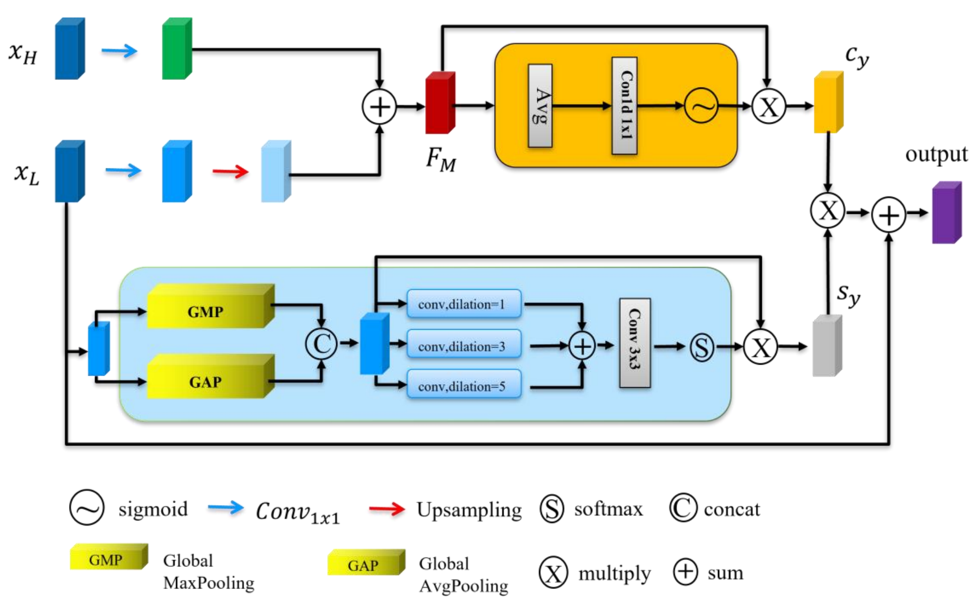


Figure 3. Structure of information fusion (IF) module

In order to enhance the spatial information of the features map, we make x_L pass through global average pooling (GAP) and global maximum pooling (GMP), which can obtain the global and local information respectively. Concatenate them to obtain the spatial information fusion. Then a multi-scale module consisting of three dilation convolutions with different dilation rates (1, 3 and 5) is added to capture the feature multi-scale information. In terms of spatial dimension, Softmax activation function is normalized to obtain the importance of each pixel relative to the region. We can obtain the

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 13 Apr 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

weights between different channels. Finally, the spatial fusion result s_y is obtained by feature multiplication. In summary, the information fusion module can be summarized as:

$$output = x_L + s_y * c_y \tag{2}$$

The IF module is proposed to improve the channel and spatial information of features. On one hand, the global dependence between channels is obtained by the channel attention module[8]. On the other hand, multi-scale pooling module is added to capture the multi-scale information. Then a spatial attention module is introduced to capture global and local feature information respectively. In general, the IF module can make up for the loss of detailed target information caused by multiple downsampling to a certain extent.

2.3 Parallel Path Connection (PPC) Module

A simple skip connection can fuse the features extracted by the encoder with the adjacent feature information, which will introduce the background noise or useless information into the shallow features and affect the quality of the final output feature of the decoder. nnUNet[9] designs effective preprocessing operations including CLAHE and Gamma to suppress the noise of background. Attention-UNet[10] introduces gated attention in the skip connection to enhance foreground region and suppress background region.

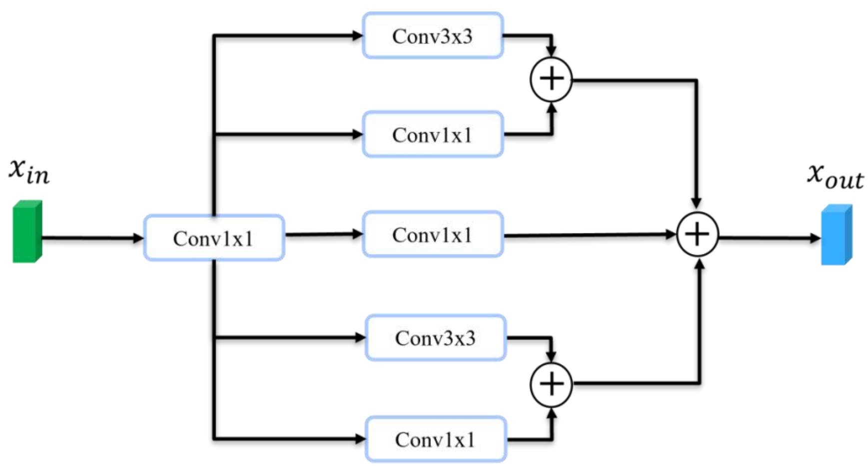


Figure 4. Structure of the parallel path connection (PPC)

The PPC module is proposed to replace with the simple skip connection. Fig. 4 illustrates the structure of the proposed PPC module. First, we first use 1x1 convolution to reduce the channel numbers to 1/4 of the original feature of the input feature x_{in} . Then we use concatenated 3x3 convolutions and a 1x1 convolution to do residual summation in parallel to fuse as much feature information as possible. Finally, we use a 1x1 convolution to compensate for the information and recover the channel numbers. The final output feature x_{out} is obtained via the summation of features from the three parallel paths.

After stacking 1x1 convolution for two times, the refinement of the features extracted by the encoder can effectively suppress the noise of background and alleviate the neglection of semantic information. In addition, the computing cost is greatly reduced and the inference speed of the network is accelerated.

3. RESULTS

3.1 Datasets

The datasets is from the First Affiliated Hospital of Soochow University, including 505 infrared MG images from 300 subjects with size 1024×1280. All the images are from 300 subjects with different eyes. In the paper, we randomly divide the 505 images into training set (303), validation set (101) and testing set (101). For simplicity, the images are resized to

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 13 Apr 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

512×512 and fed into the networks. In order to increase the diversity of the data, data augmentation strategies such as horizontal and vertical flip and rotation are adopted.

3.2 Implementation Details

In the training process, we use the stochastic gradient descent (SGD) algorithm with an initial learning rate of 0.005 and a momentum of 0.9 to optimize the proposed IF-Net. The batch size is set to 4, and the number of epochs is set to 180. The combination of the Dice loss and the cross-entropy (CE) loss is adopted as the joint loss function. The coefficient of CE loss function α is set to 1.

$$L_{Total} = L_{Dice} + \alpha L_{CE} \tag{3}$$

3.3 Evaluation Metrics

To quantitatively evaluate the MGA and MGAA segmentation performance of the proposed IF-Net, three segmentation evaluation metrics including root-mean-square deviation (RMSD)[11], Dice similarity coefficient (DSC) and intersection over union (IoU) are used in the paper. RMSD is used to measure the error in predicting the percentage of gland atrophy (GA) [12]. The definitions of these metrics are shown as follows:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{4}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{5}$$

$$GA = \frac{S_1}{S_1 + S_2} \tag{6}$$

$$RMSD(GA) = \sqrt{\frac{1}{N}\sum\nolimits_{i=1}^{N}(GA_i - \overline{GA_i})^2} \tag{7}$$

where TP denotes true positive, FP denotes false positive, FN denotes false negative, TN denotes the true negative, S_1 denotes the area of MGAA, S_2 denotes the area of MGA, GA_i denotes the atrophy rate of the i_{th} image predicted by the neural network, $\overline{GA_i}$ denotes the atrophy rate of the i_{th} image calculated by Eq.(6).

3.4 Meibomian gland area and atrophy area Segmentation

The proposed method is compared with four state-of-art networks including U-Net, GDCSeg-Net, CS^2-Net and CE-Net. As shown in Table 1, our method outperforms other methods in all evaluation metrics. CE-Net uses the branches of multiple sensitive fields to obtain multi-scale information. However, it introduces a large number of parameters and increases the calculation cost. CS^2-Net introduces a strip pooling module, which can effectively recognize the strip target, but it has poor performance for the MGA and MGAA. GDCSeg-Net uses the fusion of multi-scale feature information with channel and spatial attention mechanism to increase the target locking, which also brings a large amount of parameters.

The ablation experiments illustrate the effectiveness of our proposed IF module and PPC module, which are shown in the Table 1. As can be seen from the results of the Baseline + IFM which separately adds the IF module, it is superior to the baseline in all evaluation metrics. The IF module integrates rich semantic features of high-level features and rich spatial information of low-level features, so that the network can focus more on the focal area and improve the recognition ability of small targets. The average of Dice and the average of IoU of Baseline + PPC that adds PPC module alone on the segmentation of MGA and MGAA increases by 0.55% and 0.73%, respectively. PPC module can retain more detailed information and alleviate the neglection of semantic information. The Baseline + PPC + IFM achieves the highest result. Compared with the baseline, the Dice coefficients of the proposed network improves by nearly 1.21% for MGA segmentation and 1.90% for MGAA segmentation, respectively and the IoU improves by nearly 1.40% for MGA

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 13 Apr 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

segmentation and 2.80% for MGAA segmentation, respectively. It shows that the IF module and PPC module can both improve the segmentation performance of the network.

Fig. 5 shows the segmentation results of different methods, which indicates the results of our proposed method are closer to the ground truth.

Table 1. The performance of segmentation with different evaluation metrics.

Methods		IoU(%)	Dice(%)	RMSD(%)	Parameter(M)
Baseline	MGAA	66.71	79.58	6.3	13.39
	MGA	79.19	87.47		
	AVG	72.95	83.53		
U-Net	MGAA	66.72	79.60	6.2	14.23
	MGA	79.12	87.43		
	AVG	72.92	83.52		
GDCSeg-Net	MGAA	68.48	80.97	6.4	25.10
	MGA	79.11	87.57		
	AVG	73.80	84.27		
CS^2-Net	MGAA	65.88	78.94	5.8	14.34
	MGA	79.27	87.84		
	AVG	72.58	83.34		
CE-Net	MGAA	64.21	77.75	5.9	29.00
	MGA	77.22	86.16		
	AVG	70.71	81.96		
Baseline + IFM	MGAA	68.07	80.32	4.4	14.79
	MGA	80.06	88.18		
	AVG	74.07	84.25		
Baseline + PPC	MGAA	67.64	80.23	4.7	20.01
	MGA	79.32	87.83		
	AVG	73.48	84.04		
Baseline + IFM + PPC	MGAA	**68.58**	**81.09**	**3.9**	16.83
	MGA	**80.31**	**88.53**		
	AVG	**74.44**	**84.81**		

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 13 Apr 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

(a) (b) (c) (d) (e) (f) (g)

Figure 5. Examples of meibomian gland area and atrophy area segmentation results, White: MGA, gray: MGAA. (a) original image; (b) ground truth; (c) U-Net; (d) CS^2-Net; (e) GDCSeg-Net; (f) CE-Net; (g) the proposed IF-Net

4. CONCLUSIONS

In this paper, we propose a U-shape information fusion network for MGA and MGAA segmentation. An IF module is designed and inserted after the last two layers of the encoder. By optimizing the channel information and spatial information of the extracted features, the information loss caused by continuous downsampling can be effectively reduced. In addition, the PPC module is designed and inserted at the skip connection, which can effectively suppress the background noise by stacking several 1x1 convolution for residual connection to refine the features extracted by the encoder and alleviate the neglection of semantic information in the original skip connection of U-Net. Our proposed IF-Net has been evaluated on 505 infrared MG images from 300 subjects and the results indicate the primary effectiveness of the proposed method.

ACKNOWLEDGEMENTS

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFA0701700, in part by the National Nature Science Foundation of China under Grant U20A20170, and in part by the Natural Science Research of Jiangsu Higher Education Institutions of China under Grant 21KJB510021.

REFERENCES

[1] Dai Q, Liu X, Lin X, et al. "A novel meibomian gland morphology analytic system based on a convolutional neural network," IEEE Access, 2021, 9: 23083-23094.
[2] Long J, Shelhamer E, Darrell T. "Fully convolutional networks for semantic segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4): 640-651.
[3] Ronneberger O, Fischer P, Brox T. "U-net: Convolutional networks for biomedical image segmentation," in Proceedings of International Conference on Medical image Computing and Computer-Assisted Intervention, 2015, pp. 234-241.
[4] Gu Z, Cheng J, Fu H, et al. "Ce-net: Context encoder network for 2d medical image segmentation," IEEE Transactions on Medical Imaging, 2019, 38(10): 2281-2292.
[5] Zhu Q, Chen X, Meng Q, et al. "GDCSeg-Net: General optic disc and cup segmentation network for multi-device fundus images," Biomedical Optics Express, 2021, 12(10): 6529-6544.
[6] Mou L, Zhao Y, Fu H, et al. "CS2-Net: Deep learning segmentation of curvilinear structures in medical imaging," Medical Image Analysis, 2021, 67: 101874.
[7] Hu J, Shen L, Sun G. "Squeeze-and-excitation networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132-7141.
[8] Vaswani A, Shazeer N, Parmar N, et al. "Attention is all you need," Advances in Neural Information Processing Systems, 2017, 30: 1706.03762.
[9] Isensee F, Jaeger P F, Kohl S A A, et al. "nnU-Net: A self-configuring method for deep learning-based biomedical

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 13 Apr 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

image segmentation," Nature methods, 2021, 18(2): 203-211.
[10] Oktay O, Schlemper J, Folgoc L L, et al. " Attention U-Net: Learning where to look for the pancreas," arXiv preprint, 2018, arXiv:1804.03999.
[11] Wang J, Yeh T N, Chakraborty R, et al. "A deep learning approach for meibomian gland atrophy evaluation in meibography images," Translational Vision Science & Technology, 2019, 8(6): 37-37.
[12] Arita R, Itoh K, Inoue K, et al. "Noncontact infrared meibography to document age-related changes of the meibomian glands in a normal population," Ophthalmology, 2008, 115(5): 911-915.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 13 Apr 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use