

PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://spiedigitallibrary.org/conference-proceedings-of-spie)

Segmentation of RBCC disruption and myopic stretch line in retinal OCT images using an improved U-shape network

Diao, Yichao, Chen, Xinjian, Fan, Ying, Xie, Jiamin, Chen, Qiuying, et al.

Yichao Diao, Xinjian Chen, Ying Fan, Jiamin Xie, Qiuying Chen, Lingjiao Pan, Weifang Zhu, "Segmentation of RBCC disruption and myopic stretch line in retinal OCT images using an improved U-shape network," Proc. SPIE 11596, Medical Imaging 2021: Image Processing, 115962V (15 February 2021); doi: 10.1117/12.2579559

SPIE.

Event: SPIE Medical Imaging, 2021, Online Only

Segmentation of RBCC disruption and myopic stretch line in retinal OCT images using an improved U-shape network

Yichao Diao¹, Xinjian Chen^{1,2}, Ying Fan³, Jiamin Xie³, Qiuying Chen³, Lingjiao Pan⁴, Weifang Zhu^{1*}

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 215123, China
 3. Shanghai General Hospital, Shanghai 200080, China
 4. School of Electrical and Information Engineering, Jiangsu University of Technology, Changzhou, Jiangsu Province, 213000, China
- *indicates the corresponding author.

ABSTRACT

Pathologic myopia (PM) is a major cause of legal blindness in the world. Linear lesions are closely related to PM, which include two types of lesions in the posterior fundus of pathologic eyes in optical coherence tomography (OCT) images: retinal pigment epithelium-Bruch's membrane-choriocapillaris complex (RBCC) disruption and myopic stretch line (MSL). In this paper, a fully automated method based on U-shape network is proposed to segment RBCC disruption and MSL in retinal OCT images. Compared with the original U-Net, there are two main improvements in the proposed network: (1) We creatively propose a new downsampling module named as feature aggregation pooling module (FAPM), which aggregates context information and local information. (2) Deep supervision module (DSM) is adopted to help the network converge faster and improve the segmentation performance. The proposed method was evaluated via 3-fold cross-validation strategy on a dataset composed of 667 2D OCT B-scan images. The mean Dice similarity coefficient, Sensitivity and Jaccard of RBCC disruption and MSL are 0.626, 0.665, 0.491 and 0.739, 0.814, 0.626, respectively. The primary experimental results show the effectiveness of our proposed method.

KEYWORDS: U-Net, downsampling, deep supervision, myopic stretch lines, RBCC disruption

1. INTRODUCTION

Pathologic myopia (PM) is one of the major causes of visual impairment and even blindness worldwide with about 1% prevalence of the population^[1-3]. Linear lesions are closely related to PM, which include two types of lesions in the posterior fundus of pathologic eyes: retinal pigment epithelium-Bruch's membrane-choriocapillaris complex (RBCC) disruption and myopic stretch line (MSL). At present, there are few techniques focused on the automatic segmentation of linear lesions, except the segmentation of linear lesions based on improved c-GAN in indocyanine green angiography (ICGA) images^[4]. Although ICGA is the standard technology for linear lesion diagnosis in ophthalmic clinic, it is an invasive examination. OCT is a non-invasive imaging technology which is widely used in ophthalmic clinic, in which linear lesions can be divided into RBCC disruption and MSL^[5]. The research on RBCC disruption and MSL in OCT images is currently at an initial stage. So automatic segmentation of RBCC disruption and MSL in OCT images is of great importance to the large scale screening of linear lesions for pathologic myopic.

As shown in Fig.1(a), in OCT images, RBCC disruptions (the regions indicated by yellow arrows) are mainly shown as the discontinuities of the retinal pigment epithelium (RPE) with the corresponding increase of light penetrance into the deeper tissues^[6]. MSLs (the regions indicated by red arrows) are revealed as irregular clumping of RPE cells above the large choroidal vessel^[5]. Fig.1 (b) shows the corresponding ground truth of the RBCC disruptions and MSLs. The main challenges of RBCC disruption and MSL segmentation include the variant sizes, blurred borders and small targets, which are the common and difficult problems in medical image segmentation.

To the best knowledge of us, there are no automatic methods for the segmentation of RBCC disruption and MSL in OCT images. In this paper, we propose an improved U-shape network for the segmentation of RBCC disruption and MSL and achieve good performances. In order to solve the challenge of small target segmentation, we propose a feature aggregation pooling module (FAPM) and embed it into the encoder path, which can help the network to aggregate context information and local information. In order to make the network converge faster, deep supervision modules are added in the decoder path. Finally, we also use data augmentation to improve the generalization of the model and use the joint cross entropy loss and Dice loss as the loss function to solve the data imbalance problem.

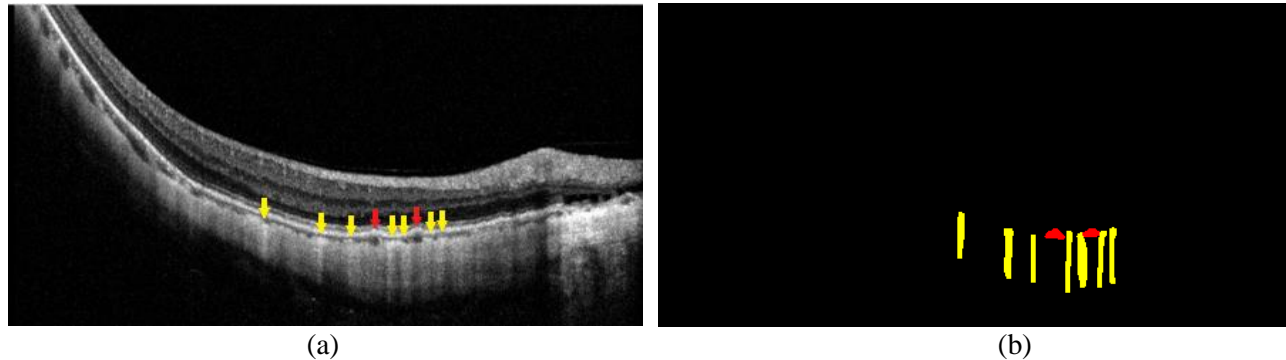


Fig.1. (a) RBCC disruption and MSL in the OCT image (yellow arrows for RBCC disruption, red for MSL) (b) The ground truth corresponding to the OCT image (yellow area is RBCC disruption and red is MSL)

2. METHODS

2.1 Overall structure of the network

U-Net^[7], whose structure contains an encoder path and a decoder path, merges features through skip connections, and is very popular in medical image segmentation. In this paper, we propose an improved U-shape network to segment RBCC disruption and MSL in retinal OCT images. Fig.2 (a) shows the structure of our proposed network. Because the MSLs are very small, it is easy to lose information using traditional maxpooling downsampling. Besides, maxpooling downsampling has a limited field of view and does not consider context information. To solve this problem, we propose a feature aggregation pooling module (FAPM), which aggregates context information and local information to enhance the network's segmentation of RBCC disruption and MSL. In addition, we add deep supervision modules in the decoder path to help the network converge faster.

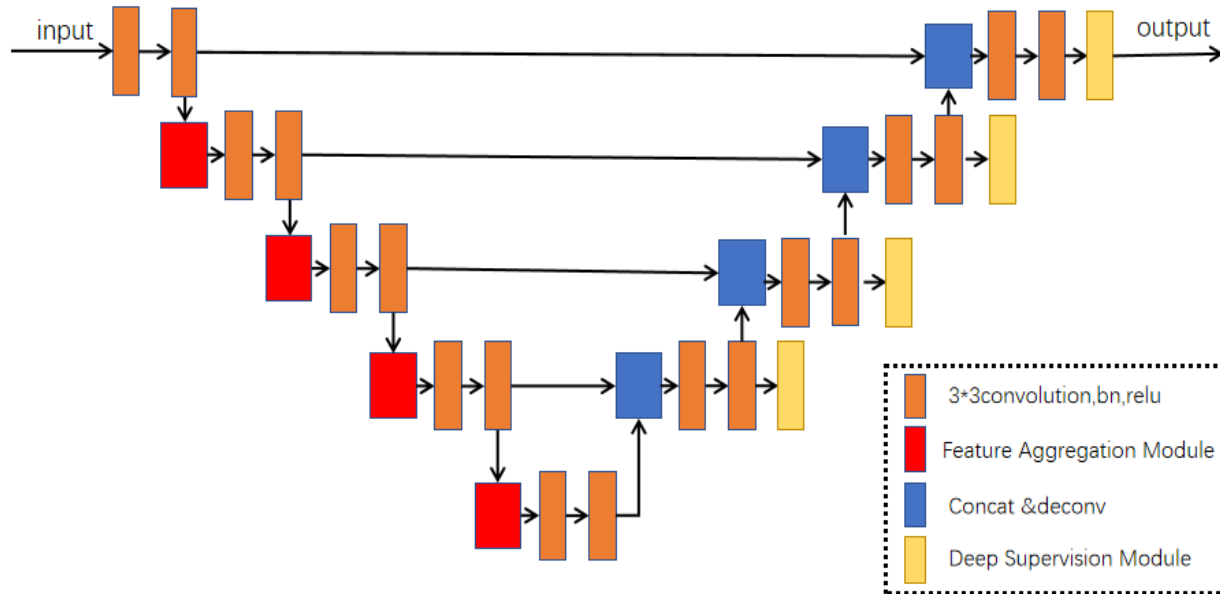
2.2 Feature aggregation pooling module

Considering that the shape of RBCC disruption is vertical stripe, we use stripe pooling^[8] to not only collect contextual information, but also make the network pay more attention to the vertical shape of RBCC disruption. There are two streams for the extraction of context information, which include vertical stripe pooling + 1*3 convolution and horizontal stripe pooling + 1*3 convolution to change the size of the input feature map from $H*W$ to $1*W$ and $H*1$, respectively. Then the sizes of these two feature maps will be changed to $H/2*W/2$ by bilinear interpolation. Finally, the two streams are fused by addition. Because the MSLs are very small with uniquely bugling shape, the local information is very important. First, a $7*7$ convolution is adopted for the input feature map to achieve larger field of view. Then through a sigmoid function, the weighted feature map is multiplied with the original input feature map. Finally resize the output feature map's size to $H/2*W/2$. This effectively considers the importance of each pixel, rather than simply taking the maximum value like maxpooling. Finally, we merge the above two modules to form a new downsampling module, which combines context information and local information. Fig. 2(b) shows details of the proposed FAPM.

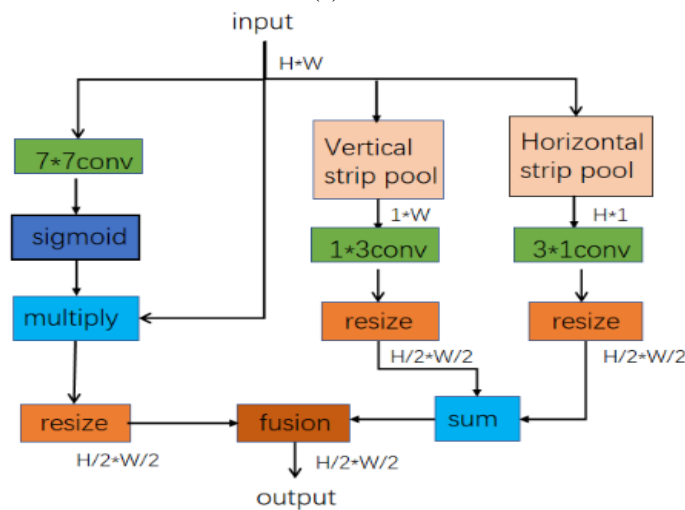
2.3 Deep supervision module

Deep supervision module was first proposed by Reference [9], but the network used in the [9] is not deep enough. In 2015, Wang et al. applied deep supervision to deep networks for the first time^[10], and proved the effectiveness of the method.

The key idea of deep supervision is to add some auxiliary classification modules in the middle layer of the deep neural network to solve the problems of gradient disappearance and slow convergence. We add four deep supervision modules into the decoder part of the segmentation network, whose structure is shown in Figure 2(c). After the input feature map, a 1×1 convolution is used to reduce the channel to 1, and then through bilinear upsampling, the map becomes the same size as the ground truth. In this way, the ground truth can supervise the output of feature map by each deep supervision module in the back propagation of the loss function. The experiments showed that the addition of the deep supervision modules to the decoder part can improve the segmentation performance of the network effectively.



(a)



(b)



(c)

Fig.2 (a) Overall structure of the network. (b) FAPM: Feature aggregation pooling module. (c) Deep supervision module.

2.4 Data augmentation strategy

The number of our dataset is not very large, which is a common problem and in medical image processing and may affect the training of the network. In order to increase the generalization and the robustness of the model, each training epoch adopts some of the online augmentation strategies including random flipping, random rotation and adding Gaussian noise to the image.

2.5 Loss function

Due to the small proportion of RBCC disruption and MSL in the image, this can cause data imbalance problems. The Dice coefficient loss^[11] can make small targets better noticed by the network. Therefore, we choose the sum of cross-entropy loss and Dice loss as the total loss function, which can effectively solve the problem of unbalanced data in the training process. The loss function can be defined as follows:

$$loss_{bce} = -\frac{1}{n} \sum_i^n [x_i \ln y_i + (1 - x_i) \ln(1 - y_i)] \quad (1)$$

$$loss_{dice} = 1 - \frac{1}{n} \sum_{i=1}^n \frac{2x_i y_i}{x_i^2 + y_i^2} \quad (2)$$

$$loss = loss_{bce} + loss_{dice} \quad (3)$$

Where x denotes the pixel value in the ground truth, y denotes the value predicted by the network, n represents the total number of pixels in the image, and i denotes the i th pixel.

3. RESULTS

3.1 Datasets

The 2D retinal OCT B-scan images used in our experiments are from 3D retinal OCT scans acquired by Heidelberg SPECTRALIS OCT (Heidelberg Engineering, Germany). The dataset includes 667 retinal OCT slice images with ground truth from 20 different patients (labeled as RBCC disruption, MSL, and normal), which are labeled under the supervision of two senior ophthalmologists.

3.2 Implementation Details

The size of the original 2D OCT B-scan image is 769*392, which is too large for the training of the proposed network. In order to facilitate the network training, the images are resized to 512*256 and used as the input of the network. The proposed network is realized on the Pytorch framework. In the training process, the SGD^[12] algorithm with an initial learning rate of 0.001, momentum of 0.9 and weight decay of 0.0001 is used to optimize the network. The batchsize is set to 2 and the number of epochs is 120. We use three-fold cross validation to evaluate the performance of the proposed method. On average, each fold includes about 220 images from different patients.

3.3 Evaluation metrics

Due to the limitation of the number of dataset, 3-fold cross evaluation strategy is adopted in our experiments. In order to evaluate the proposed network, three evaluation metrics including Dice similarity coefficient (DSC), Jaccard index and sensitivity are adopted as evaluation metrics. Sensitivity represents the proportion of pixels in the original foreground that are correctly marked as foreground. DSC and Jaccard are the measurements of similarity between two regions. The calculations of DSC, Jaccard and sensitivity are shown as follows:

$$DSC = 2 \frac{|LA_{pred} \cap LA_{gt}|}{|LA_{pred}| + |LA_{gt}|} \quad (4)$$

$$Jaccard = \frac{|LA_{pred} \cap LA_{gt}|}{|LA_{pred} \cup LA_{gt}|} \quad (5)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (6)$$

where TP and FN represent the number of true positive and false negative predictions, respectively. LA_{pred} and LA_{gt} represent the predicted result and the lesion area in the ground truth respectively. $|\cdot|$ means the size of the pixel set.

3.4 Results

We use U-net as the baseline. To evaluate the performance of our method, comparison experiments between SegNet^[13], PSPNet^[14] and our proposed improved U-shape network have been performed. We also do ablation experiments to prove the effectiveness of our proposed FAPM and deep supervision modules. Fig.3 shows some examples of segmentation results of the baseline and our proposed method. It can be seen that our proposed method can segment RBCC disruptions and MSLs better than the baseline with more true positive and less false positive.

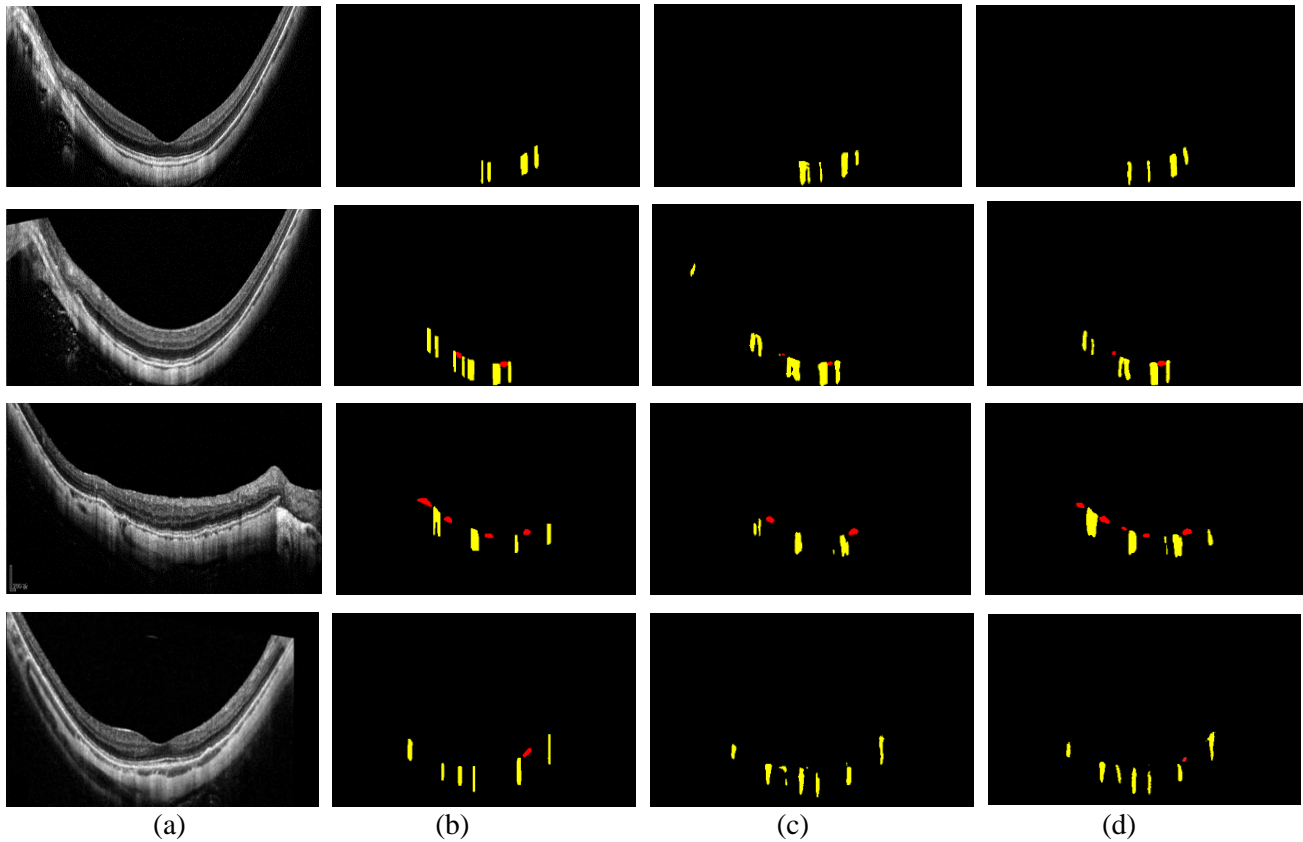


Fig.3 Segmentation results. (a) Original B-scan images. (b)Ground truth. (c) Baseline. (d) Our proposed method.

As vividly shown in Table 1, the overall average DSC of our baseline for RBCC disruptions and MSL segmentation is higher than SegNet and PSPNet. Compared with the above three networks, our improved U-shape network achieves significant improvements in DSC, sensitivity and Jaccard index. The results of the ablation experiment show that the

proposed FAPM and deep supervision modules are effective for the improvement of RBCC disruptions and MSL segmentation.

Table 1: The performances of segmentation with different evaluation metrics

Architecture	RBCC disruption			MSL		
	DSC	Jaccard	Sensitivity	DSC	Jaccard	Sensitivity
SegNet	0.578	0.444	0.628	0.617	0.504	0.742
PSPNet	0.560	0.422	0.587	0.630	0.505	0.757
Baseline	0.571	0.434	0.607	0.646	0.526	0.770
Baseline +FAPM	0.558	0.422	0.606	0.672	0.548	0.796
Baseline +DSM	0.601	0.465	0.624	0.717	0.601	0.816
Proposed method	0.626	0.491	0.665	0.739	0.626	0.814

4. CONCLUSIONS

In this paper, we propose an improved U-shaped network with FAPM and deep supervision modules for the automatic segmentation of RBCC disruption and MSL in retinal OCT images. The experimental results indicate that these two modules can improve the segmentation performance, which can provide great help for the further research on linear lesion in pathological myopia. We think the main reasons for the limitation of segmentation performance include: (1) The limitation of the number of the dataset causes the insufficient training of the proposed network. (2) Due to the small target characteristic of RBCC disruptions and MSL, the segmentation of RBCC disruption and MSL is a classical data imbalance problem, which is always a difficult problem in medical image segmentation. In order to improve the segmentation performance of the network, we are currently trying to combine carafe^[15], DUpsampling^[16] and other up-sampling methods with deep supervision.

5. 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 61622114 and 62001196, and in part by the National Basic Research Program of China under Grant 2014CB748600.

6. REFERECENS

- [1] S. Resnikoff, D. Pascolini, D.Etya'Ale, et al. "Global data on visual impairment in the year 2002. " Bulletin of the world health organization. 82(11), pp. 844–851, 2004.
- [2] C.K. Leung. "Diagnosing glaucoma progression with optical coherence tomography. " Current opinion in ophthalmology. 25(2), pp. 104–111.2014.
- [3] M. Banitt. "The choroid in glaucoma." Current opinion in ophthalmology. 24(2), pp.125-129, 2013.
- [4] H. Jiang, X. Chen, F. Shi, et al. "Improved cGAN based linear lesion segmentation in high myopia ICGA images. " Biomedical optics express, 10(5), pp.2355-2366, 2019.
- [5] K. Shinohara, M. Moriyama, N. Shimada, et al. "Myopic stretch lines: linear lesions in fundus of eyes with pathologic myopia that differ from lacquer cracks. " Retina, 34(3), pp. 461-469, 2014.
- [6] N.K. Wang, C.C. Lai, C.L. Chou, et al. "Choroidal thickness and biometric markers for the screening of lacquer cracks in patients with high myopia." PLoS One, 8(1), e53660, 2013.
- [7] O. Ronneberger, P. Fischer and T. Brox. "U-net: Convolutional networks for biomedical image segmentation. " Medical Image Computing and Computer Assisted Intervention, pp. 234-241,2015.
- [8] Q. Hou, L. Zhang, M.M Cheng, et al. "Strip Pooling: Rethinking Spatial Pooling for Scene Parsing." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4003-4012. 2020.
- [9] C.Y. Lee, S. Xie, P. Gallagher, et al. "Deeply-supervised nets." Artificial intelligence and statistics, pp. 562-570, 2015.
- [10] L. Wang, C.Y Lee, Z. Tu, et al. "Training deeper convolutional networks with deep supervision." arXiv preprint arXiv:1505.02496 ,2015.

- [11] F. Milletari, N. Navab and S.A. Ahmadi. "V-Net: Fully convolutional neural networks for volumetric medical image segmentation. " 2016 fourth international conference on 3D vision (3DV). IEEE, pp. 565-571, 2016.
- [12] L. Bottou. "Stochastic gradient descent tricks." Neural networks: Tricks of the trade. Springer, Berlin, Heidelberg, pp. 421-436. 2012.
- [13] V. Badrinarayanan, A. Kendall, R. Cipolla. "SegNet: A deep convolutional encoder-decoder architecture for scene segmentation. " IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), pp.2481-2495. 2017.
- [14] H. Zhao, J. Shi, X Qi, et al. "Pyramid scene parsing network. " IEEE Conference on Computer Vision and Pattern Recognition, pp.2881-2890, 2017.
- [15] J. Wang, K. Chen, R. Xu , et al. "Carafe: Content-aware reassembly of features." Proceedings of the IEEE International Conference on Computer Vision, pp.3007-3016, 2019.
- [16] Z. Tian, T. He, C. Shen, et al. "Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.3126-3135, 2019.