

Sk-Conv and SPP-based UNet for lesion segmentation of coronary optical coherence tomography

Zhan Wang^a, Jiawei Zheng^b, Peilin Jiang^{a,*} and Dengfeng Gao^b

^a*School of Software, Xi'an Jiaotong University, Xi'an, Shaanxi, China*

^b*Department of Cardiovascular Medicine, Second Affiliated Hospital of Xi'an, Jiaotong University, Xi'an, Shaanxi, China*

Abstract.

BACKGROUND: Coronary artery disease (CAD) manifests with a blockage the coronary arteries, usually due to plaque buildup, and has a serious impact on the human life. Atherosclerotic plaques, including fibrous plaques, lipid plaques, and calcified plaques can lead to occurrence of CAD. Optical coherence tomography (OCT) is employed in the clinical practice as it clearly provides a detailed display of the lesion plaques, thereby assessing the patient's condition. Analyzing the OCT images manually is a very tedious and time-consuming task for the clinicians. Therefore, automatic segmentation of the coronary OCT images is necessary.

OBJECTIVE: In view of the good utility of Unet network in the segmentation of medical images, the present study proposed the development of a Unet network based on Sk-Conv and spatial pyramid pooling modules to segment the coronary OCT images.

METHODS: In order to extract multi-scale features, these two modules were added at the bottom of UNet. Meanwhile, ablation experiments are designed to verify each module is effective.

RESULTS: After testing, our model achieves 0.8935 on f1 score and 0.7497 on mIOU. Compared to the current advanced models, our model performs better.

CONCLUSION: Our model achieves good results on OCT sequences.

Keywords: UNet, OCT, SK-Net, lesion segmentation

1. Introduction

The use of optical coherence tomography (OCT) in coronary atherosclerosis can clearly show the site and level of the lesion to the clinicians. It takes a long time for professional doctors to analyze OCT images manually. Therefore, the automatic segmentation of coronary OCT images is necessary.

In fact, many patch segmentation methods have been proposed. Based on the Haralick texture signature and K-means clustering, Prakash [1] introduced a new patch segmentation method. In the study by Celi et al. [2], a novel method was employed that introduced the best global threshold of OSTU and morphological processing. Xu [3] proposed an OCT image system that adopted the support vector machine

*Corresponding author: Peilin Jiang, School of Software, Xi'an Jiaotong University, Xi'an, Shaanxi, China. E-mail: pljiang@xjtu.edu.cn.

(SVM) for automatic detection of atherosclerotic diseases. For achieving semi-automatic segmentation on OCT images, Huang et al. [4] adopted a method on the basis of image feature extraction and SVM. This method achieved a precision of 89% in the detection of fibrous plaques, 79.3% for calcified plaques, and 86.5% for lipid plaques. Ughi [5] and Athanasiou applied the machine learning in the form of random forest and K-means clustering model to detect the calcified plaques on coronary OCT. In an approach employing deep learning, Gessert [7] developed a method with deep learning models to discover and classify the calcifications and plaques between fibers/lipids. Li employed a fully automatic method on the basis of convolutional neural network for segmentation of calcified plaques [8], and the F1 score for pixel-level calcification classification reached 0.883 ± 0.008 .

Due to the good segmentation performance of UNet [9] on medical images, many related models have been proposed. The Att-UNet model [10] proposed a medical attention gate model that paid attention to the important local features from encoders. UNet++ [11] decreased the semantic gap between the shallow and deep feature maps by redesigning skip connection to fill up the hollow structure of UNet. With the dense atrous convolution module (DAC) and the residual multi-kernel pooling module (RMP), the CE-Net model proposed a context extractor to obtain deeper semantic feature maps. The DU-Net model [13] extends UNet and two other types of connections with dense blocks to obtain more information of the feature map. Cao [14] introduced the dilated convolution of 4 parallel branches before concatenating at UNet's each layer to decrease the semantic gap of features between the encoder and its corresponding decoder.

Inspired by the attention mechanism and the work of Cao et al., our study proposed two improved methods based on UNet. First, a four-branch Sk-Conv module [15] was added to the last layer of the encoder to synthesize messages from multiple convolution kernels. In addition, we introduced a spatial pyramid pooling (spp) module [16] to the original UNet network to get multi-scale features.

2. Dataset and methodology

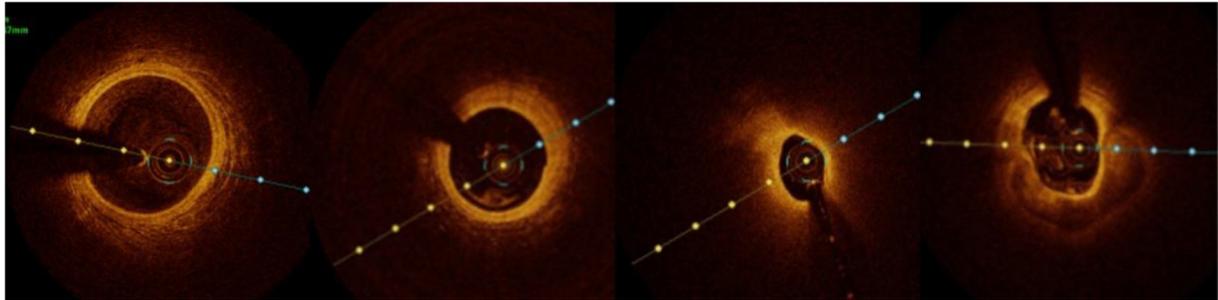
2.1. Dataset and preprocessing

In the present study, raw OCT images of the culprit vessels of each patient were acquired through the ILUMIEN OPTIS system built into the OCT Mobile Dragonfly. A total of 5624 frames of OCT original images of the culprit blood vessels of 15 patients were simultaneously annotated by 2 observers through ITK-Snap software, and the normal tissues and fibers in each frame of OCT original images of the culprit vessels of the patients were marked with four different colors. Plaques, lipid plaques, and calcified plaques (Fig. 1), and were reviewed by a clinical expert for training and testing of deep learning models.

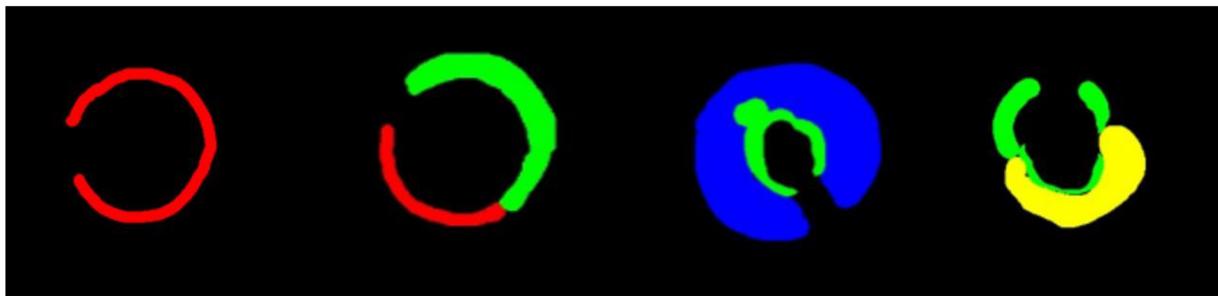
Since the original image or label with the size of $736 * 736$ was too large and consisted too much information, which was unnecessary for segmentation, the adjustment of the image size was necessary. After many prior experiments, the original image was resized to $370 * 370$ size for more accurate prediction. Among the 5624 images provided, we selected 1500 images to test and the others were augmented to train 49488 images. Enhancements include 90° , 180° and 270° rotations along with the horizontal and vertical flips.

2.2. Architecture

We employed the Sk-Conv module to obtain more information of the different receptive fields on the basis of UNet by dynamically selecting the size of convolution kernels. Meanwhile, we introduced the SPP module to obtain multi-scale feature information of the images. The architecture of the method adopted in the study is shown in Fig. 2.



(a) OCT images with the size of 370 * 370



(b) Labels with the size of 370 * 370

Fig. 1. a) Four cropped OCT images with the size of 370 * 370; b) Corresponding ground truth labels: red – normal, green – fibrosis, blue – lipid, yellow – calcification and black – background (bg).

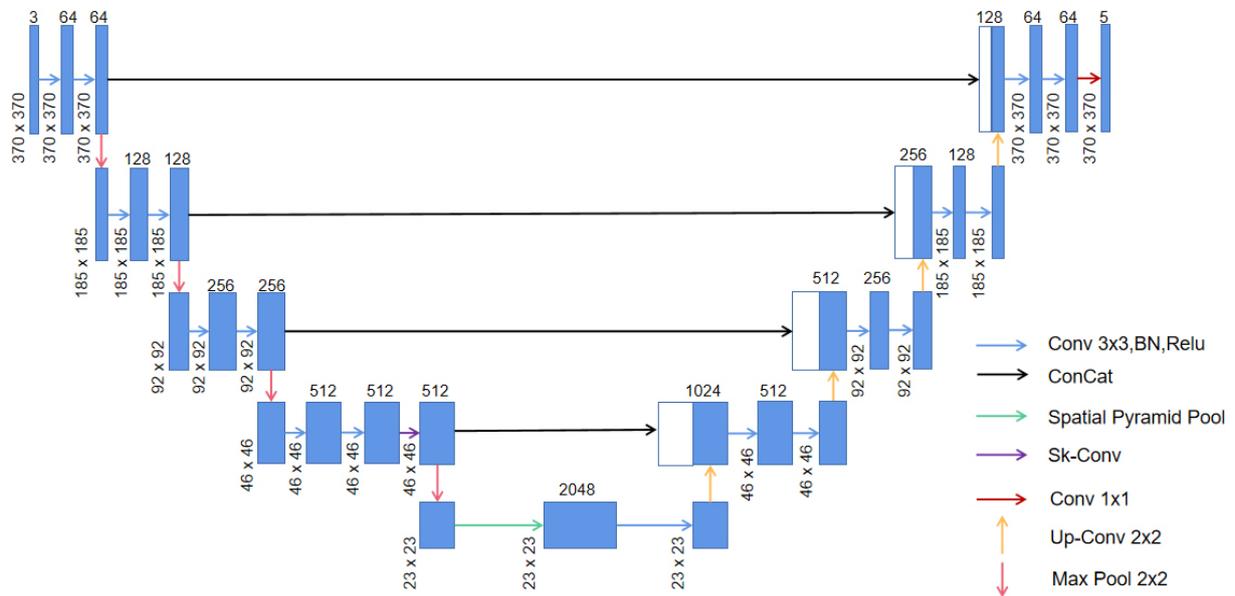


Fig. 2. Network architecture of the present study; green arrow indicates the spp module and purple arrow shows the Sk-Conv module.

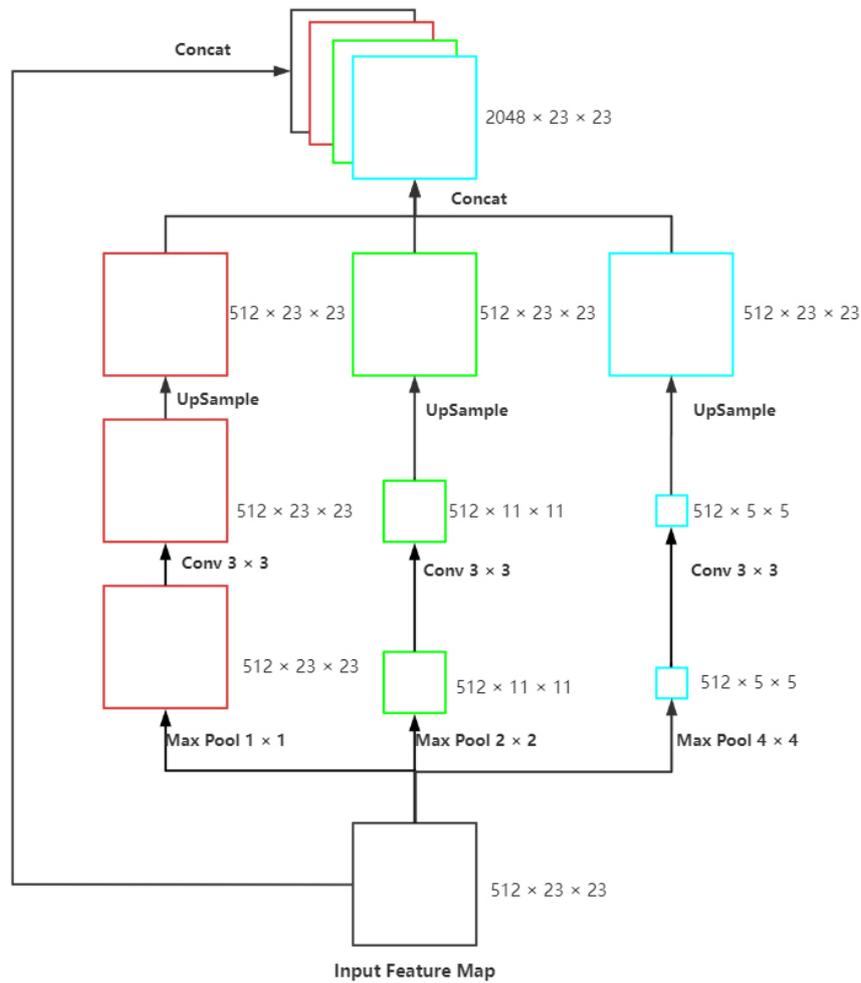


Fig. 3. Proposed SPP model: The max-pooling of three scales, convolution with the size of 3×3 and up-sampling were applied in turn to extract multi-scale features. Output was finally concatenated the with the input.

2.3. Spatial pyramid pooling

For achieving the multi-scale extraction of advanced features of the encoder, the module called SPP (Fig. 3) was applied at the last layer of UNet to perform pooling of three scales on the input feature map. Subsequently, we performed convolution to extract the features with different scales, upsample to the size of the input and concatenated with the original feature map. This module enhanced the robustness of this network to differentiate the spatial layout and resolution.

2.4. Sk-Conv

Since the receptive field of a fixed-size convolution kernel is fixed, we introduced the Sk-Conv (Fig. 4) to generate the information of convolution kernels with different sizes. As a result, the method could resize the receptive field adaptively, presented the importance of different convolution kernels, and raised the generalization ability of the net.

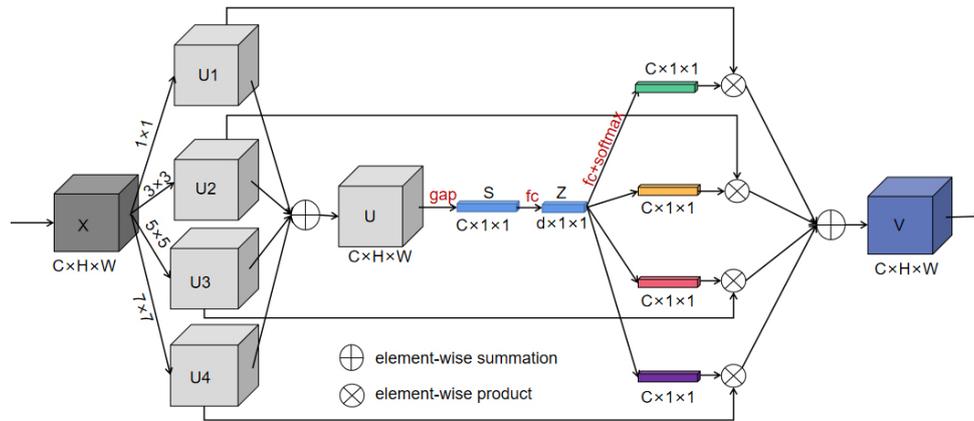


Fig. 4. SK-Conv module employed in our model.

In Fig. 4, SK-Net introduced a branch attention mechanism while paying attention to the channel attention, which could resize the receptive field adaptively and exhibited a better performance. The core in Sk-Net is the SK-Conv module, which contained three parts. In the first stage of the SK-Conv module which was called split in our network, convolution kernels with sizes of 1, 3, 5, and 7 were set up to perform convolution on the feature map X to obtain U_1 , U_2 , U_3 , and U_4 . In the second stage named fuse, the four outputs were added element-wise to generate U ; subsequently the global information S was obtained through global average pooling (GAP), and then the full connect layer (FC) was performed on S to find each channel. Then Z was restored to the size of S through a full connect layer again. In the last stage called select, softmax was used to obtain the weight matrices of kernels of different sizes and weighted operations with U_1 , U_2 , U_3 , and U_4 respectively, and finally the feature map V containing the attention of 4 branches was obtained.

2.5. Training

We used the stochastic gradient descend with a momentum of 0.9 during the model training. In the Sk-Conv module, there are three important hyper-parameters: M determines the number of kernels to be chosen, the group number G that shows the cardinality of each branch, and r indicates the number of output channels of the first fully connected layer in the fuse stage. We set $G = 1$, $M = 4$ and $r = 16$.

Inspired by Cao's study [14], category imbalance is very common for medical image segmentation, so we needed to specifically calculate the number of pixels of each target. Based on our statistics, the number of pixels of these five types of targets was 6244003560 (background), 36524472 (normal tissue), 222035352 (fibrous plaque), 158842704 (lipid plaque) and 113501112 (calcified plaque). Obviously, there was a big difference in the number of background pixels and other pixels. Therefore, we chose the Focal Loss as the focal loss and set $\alpha = 0.5$, $\gamma = 2$.

In addition, all the programs were implemented by Pytorch 1.10.1 toolkit and run on Red Hat 4.8.5 system with three NVIDIA Tesla P100 GPUs and an Inter(R) Xeon(R) Gold 5118 CPU @ 2.30GHz.

3. Results

3.1. Ablation studies

The ablation experiments (Table 1) were necessary in our research. According to our experiment, the

Table 1
Ablation study

Model	F1 score	mIOU
U-Net	0.8761	0.7082
U-Net + SPP	0.8914	0.7422
U-Net + SK-Conv	0.8833	0.7311
U-Net + SPP + SK-Conv	0.8935	0.7497

Table 2
Comparison among different models on four cases

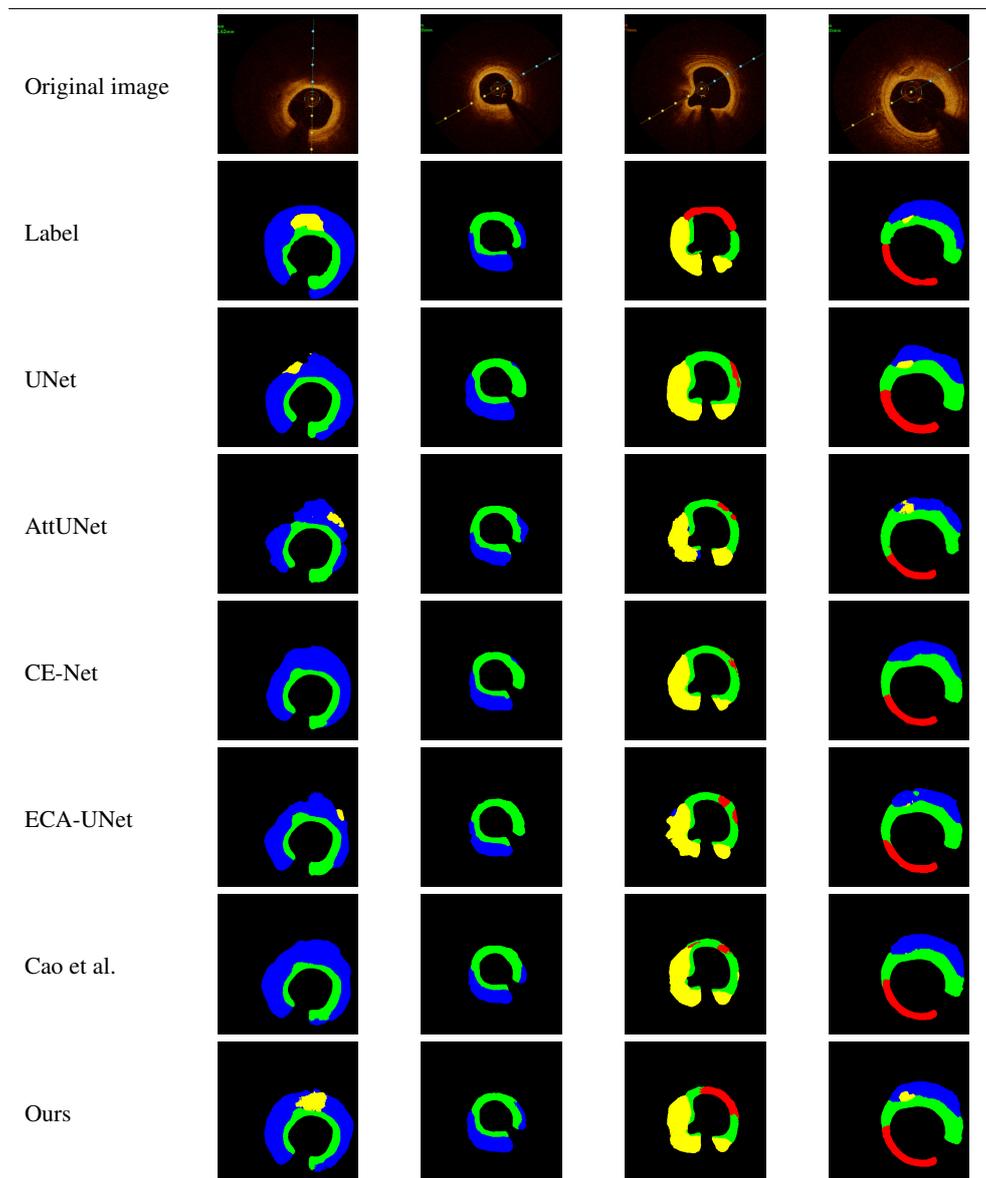


Table 3
Comparison of different methods indexes

Model and classes		Precision	Recall	F1 score	mIOU
UNet	Background	0.99	0.99	0.99	0.7082
	Normal tissue	0.76	0.74	0.75	
	Fibrous plaque	0.82	0.80	0.81	
	Lipid plaque	0.79	0.74	0.76	
	Calcified plaque	0.79	0.81	0.80	
Att-UNet	Background	0.99	0.99	0.99	0.6689
	Normal tissue	0.76	0.68	0.72	
	Fibrous plaque	0.80	0.79	0.80	
	Lipid plaque	0.74	0.68	0.71	
	Calcified plaque	0.77	0.72	0.74	
CE-Net	Background	0.99	0.99	0.99	0.7405
	Normal tissue	0.81	0.76	0.79	
	Fibrous plaque	0.83	0.83	0.83	
	Lipid plaque	0.83	0.78	0.80	
	Calcified plaque	0.86	0.79	0.82	
ECAUNet	Background	0.99	0.99	0.99	0.7165
	Normal tissue	0.77	0.77	0.77	
	Fibrous plaque	0.82	0.81	0.81	
	Lipid plaque	0.79	0.74	0.77	
	Calcified plaque	0.80	0.80	0.80	
Cao et al.	Background	0.99	0.99	0.99	0.7453
	Normal tissue	0.82	0.76	0.79	
	Fibrous plaque	0.82	0.84	0.83	
	Lipid plaque	0.82	0.80	0.81	
	Calcified plaque	0.83	0.83	0.83	
Ours	Background	0.99	0.99	0.99	0.7497
	Normal tissue	0.78	0.81	0.80	
	Fibrous plaque	0.84	0.82	0.83	
	Lipid plaque	0.82	0.80	0.81	
	Calcified plaque	0.83	0.83	0.83	

U-Net integrating the spp module or the Sk-Conv module performed better in terms of both the F1 score and mIOU. Finally, our model with both SPP module and Sk-Conv improved the F1 score by 1% and mIOU by 4%.

3.2. Experimental results

Our study systematically compared the differences between the UNet, Attention-UNet (Att-UNet), CE-Net, UNet with ECA-Module (ECA-U-Net) [18] and Cao's method in plaque segmentation, and presented the comparison of sensitivity, precision, F1 score and mIOU in the test set.

Table 2 shows some cases of the labels and segmentation results of different models. Our study offered more accurate segmentation than others. Furthermore, only few missed and wrong prediction on segmentation of normal tissues, fibrous plaque, lipid plaque and calcified plaque were noted when compared to the other methods.

Table 3 indicates the specific metrics of different methods on test images. The F1 scores and the mIOU of other models were no better than that of our method.

4. Conclusion

In the present study, we designed an improved UNet by adding a spatial pyramid pooling and Sk-Conv

to achieve better multi-scale feature extraction for accurately detecting and quantitatively calculating the degree of coronary atherosclerosis tissue through original OCT images. The results indicated that with the improvements of spp and Sk-Conv, our model achieved good results on the OCT sequences and performed better than the current advanced models.

The clinical significance of our results is that it did not only ensure accuracy, but also improved the efficiency of the interpretation and analysis of coronary and OCT images. In addition, our model could accurately segment the normal tissues and diseased plaques at the same time, which can assist the doctors to understand the degree of lesion of the OCT sequence, thereby deciding the course of treatment for the patients. However, in the present study, the pixels representing the normal tissue accounted for the least proportion of the training samples, and exhibited a poor metric. In this regard, the improvement in the loss function or adding weights to focal loss can be considered in future studies.

Acknowledgments

None to report.

Conflict of interest

None to report.

References

- [1] Prakash A, Hewko M, Sowa M, Sherif S. Texture based segmentation method to detect atherosclerotic plaque from optical tomography images. In: Bouma R, Leitgeb B. (eds.) Optical Coherence Tomography and Coherence Techniques VI. SPIE Proceedings, vol. 8802, p. 88020S. Optical Society of America, Munich, 2013.
- [2] Celi S, Berti S. *In-vivo* segmentation and quantification of coronary lesions by optical coherence tomography images for a lesion type definition and stenosis grading. *Med. Image Anal.* 2014; 18(7): 1157-1168. doi: 10.1016/j.media.2014.06.011.
- [3] Xu M, Cheng J, Wong DWK, Taruya A, Liu J. "Automatic atherosclerotic heart disease detection in intracoronary optical coherence tomography images," 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2014. pp. 174-177. doi: 10.1109/EMBC2014.6943557.
- [4] Huang Y, He C, Wang J, Miao Y and Zhu T. Intravascular optical coherence tomography image segmentation based on support vector machine algorithm. *Molecular & Cellular Biomechanics.* 2018; 15(2): 117.
- [5] Ughi GJ, Adriaenssens T, Sinnaeve P, Desmet W, Hooge JD. Automated tissue characterization of *in vivo* atherosclerotic plaques by intravascular optical coherence tomography images. *Biomedical Optics Express.* 2013; 4(7): 1014-1030. doi: 10.1364/BOE.4.001014.
- [6] Athanasiou LS, Bourantas CV, Rigas G, Sakellarios AI, Fotiadis DI. Methodology for fully automated segmentation and plaque characterization in intracoronary optical coherence tomography images. *Journal of Biomedical Optics.* 2014; 19(2): 26009. doi: 10.1117/1.JBO.19.2.026009.
- [7] Gessert N, Lutz M, Heyder M, Latus S, Leistner DM, Abdelwahed YS, Schlaefer A. Automatic plaque detection in IV-OCT pullbacks using convolutional neural networks. *IEEE Trans. Med. Imag.* 2018; 38(2): 426-434. doi: 10.1109/TMI.2018.2865659.
- [8] Li C, Jia H, Tian J, He C, Lu F, Li K, Gong Y, Hu S, Yu B, Wang Z. Comprehensive Assessment of Coronary Calcification in Intravascular OCT Using a Spatial-Temporal Encoder-Decoder Network. *IEEE Trans Med Imaging.* 2022; 41(4): 857-868. doi: 10.1109/TMI.2021.3125061.
- [9] Ronneberger O, Fischer P, Brox T. "U-Net: Convolutional networks for biomedical image segmentation". In *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer, 2015. pp. 234-241.
- [10] Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B. Attention U-Net: Learning Where to Look for the Pancreas. 2018.

- [11] Zhou ZW, Rahman Siddiquee MM, Tajbakhsh N, Liang JM. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *Deep Learn Med Image Anal Multimodal Learn Clin Decis Support* (2018). 2018 Sep; 11045: 3-11. doi: 10.1007/978-3-030-00889-5_1.
- [12] Gu Z, Cheng J, Fu H, Zhou K, Hao H, Zhao Y, Zhang T, Gao S, Liu J. CE-Net: Context Encoder Network for 2D Medical Image Segmentation[J]. *IEEE Transactions on Medical Imaging*. 2019; 38(10): 2281-2292. doi: 10.1109/TMI.2019.2903562.
- [13] Ghamdi MA, Abdel-Mottaleb M, Collado-Mesa F. DU-Net: Convolutional Network for the Detection of Arterial Calcifications in Mammograms. *IEEE Transactions on Medical Imaging*. 2020; 39(10): 3240-3249. doi: 10.1109/TMI.2020.2989737.
- [14] Cao XY, Zheng JW, Liu Z, Jiang PL, Gao DF, Ma R. Improved U-Net for Plaque Segmentation of Intracoronary Optical Coherence Tomography Images. In: Farkas I, Masulli P, Otte S, Wermter S, eds. *ARTIFICIAL NEURAL NETWORKS AND MACHINE LEARNING – ICANN 2021, PT III*. Vol 12893. 30th International Conference on Artificial Neural Networks (ICANN); 2021. pp. 598-609. doi: 10.1007/978-3-030-86365-4_48.
- [15] Li X, Wang W, Hu X, Yang J. "Selective Kernel Networks," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. pp. 510-519. doi: 10.1109/CVPR.2019.00060.
- [16] He KM, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2015; 37(9): 1904-1916. doi: 10.1109/TPAMI.2015.2389824.
- [17] Lin TY, Goyal P, Girshick R, He KM, Dollár P. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. 2017; PP(99): 2999-3007. doi: 10.1109/TPAMI.2018.2858826.
- [18] Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q. "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020. pp. 11531-11539. doi: 10.1109/CVPR426002020.01155.