Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Automatic Segmentation of Coronary Lumen and External Elastic Membrane in Intravascular Ultrasound Images Using 8-layer U-Net

Liang Dong
The Second Affiliated Hospital of Zhejiang University
Wei Lu
ArteryFlow Techology Co., Ltd.
Jun Jiang
Second Affiliated Hospital, School of Medicine, Zhejiang University
Ya Zhao
ArteryFlow Techology Co.,Ltd.
Xiangfen Song
ArteryFlow Techology Co., Ltd.
Xiaochang Leng
ArteryFlow Techology Co.,Ltd.
Hang Zhao
Arteryflow Techology Co., Ltd.
Jian'an Wang
The Second Affiliated Hosiptal of Zhejiang University
Changling Li
The Second Affiliated Hospital of Zhejiang University
Jianping Xiang (ygnjd2016@gmail.com)
ArteryFlow Techology Co., Ltd.

Research

Keywords: IVUS, Coronary, Segmentation, Lumen, EEM, MeshGrid, U-Net.

Posted Date: January 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-42239/v4

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published on February 6th, 2021. See the published version at https://doi.org/10.1186/s12938-021-00852-0.

Abstract

Background: Intravascular ultrasound (IVUS) is the golden standard in accessing the coronary lesions, stenosis, and atherosclerosis plaques. In this paper, a fully-automatic approach by an 8-layer U-Net is developed to segment the coronary artery lumen and the area bounded by external elastic membrane (EEM), i.e. cross section area (EEM-CSA). The database comprises of single-vendor and single-frequency IVUS data. Particularly, the proposed data augmentation of MeshGrid combined with flip and rotation operations is implemented, improving the model performance without pre- or post-processing of the raw IVUS images.

Results: The mean intersection of union (mIoU) of 0.937 and 0.804 for the lumen and EEM-CSA respectively were achieved, which exceeded the manual labeling accuracy of the clinician. **Conclusion:** The accuracy shown by the proposed method is sufficient for subsequent reconstruction of 3D IVUS images, which is essential for doctors' diagnosis in the tissue characterization of coronary artery walls and plaque compositions, qualitatively and quantitatively.

Background

Coronary heart disease has been the leading cause of death worldwide [1], and the coronary atherosclerosis is the dominant cause of coronary heart disease. In early atherosclerosis, coronary artery remodeling slows down the progression of vascular stenosis with the accumulation of coronary plaques. Intravascular ultrasound (IVUS) is one of the most effective real-time medical imaging techniques, which plays a critical role in the diagnosis and treatment of coronary heart disease.

2D IVUS images acquired serially by an IVUS catheter pulling back through the coronary artery can evaluate arterial distensibility caused by atherosclerotic plaque. The accurate segmentation of lumen and external elastic membrane cross section area (EEM-CSA) from 2D coronary IVUS images that contributes to assessing the atherosclerosis plaque and its vulnerability by measuring lumen diameter, plaque eccentricity, plaque burden, etc. has crucial clinical significance. However, it is time-consuming and experience-dependent for doctors to manually delineate the lumen and EEM contours on the 2D IVUS images. A typical IVUS pullback contains more than 3000 images, so an accurate, fast, and fully automatic segmentation of lumen and EEM-CSA is highly desirable but remains a challenging task due to the relative complexity of IVUS images.

Several segmentation techniques and methods in image processing and computer vision have been performed for coronary IVUS images [2, 3][2-4][1-3] . Traditional image processing methods, including graph search, active surfaces, and active contours, were applied to segment IVUS images based on local image properties or global gray level properties. 3D fast marching method [4, 5], incorporating the texture gradient and the gray level gradient, was applied to segment the walls of the coronary artery with an interactive initialization on EEM borders. In recent years, deep learning has been widely applied in the medical imaging analysis and achieved remarkable results [6, 7]. It has been utilized to detect the lumen

and media-adventitia borders in IVUS due to its capabilities in automatic feature extraction [8, 9]{Su, 2017 #47;Yang, #49;Su, 2017 #58}.

In this paper, we describe the development and evaluation of a U-Net [10] based pipeline that automatically segments the lumen and EEM-CSA from 2D IVUS images. The pipeline has two major steps: first, the data augmentation of MeshGrid combined with flip and rotation operations (MeshGrid+Flip+Rotate) is performed on raw IVUS images; second, an 8-layer deep U-Net is used for pixel-level prediction.

Results

Experiments were carried out for segmenting the lumen and EEM-CSA with four augmentation strategies of No Augmentation, Flip+Rotate, MeshGrid, MeshGrid+Flip+Rotate. The contours predicted by the method with MeshGrid+Flip+Rotate augmentation (3rd row of Figure 1) were in higher agreement with the ground truth (2nd row of Figure 1) for a range of morphologies, in comparison with those with No augmentation (4th row of Figure 1), MeshGrid (5th row of Figure 1), and Flip+Rotate (6th row of Figure 1). Particularly obvious difference is shown in the 2nd and 3rd column of Figure 1.

Table 1 quantifies the segmentation results from the four data augmentation strategies. Compared with the other three data augmentation strategies, the way of MeshGrid+Flip+Rotate presented better segmentation performance in both lumen and EEM-CSA, with mIoU of 0.937 and 0.804, respectively. The reason why 8-layer U-Net does not exceed EEM segmentation result of Ji Yang et al.[10] could be addressed from two perspectives. On the one hand, the training data is too small to train a powerful segmentation model; for another, 8-layer U-Net's architecture could be further optimization.

Table 1. The quantitative performance for the four different data augmentation strategies

Augmentation Strategies	MIoU (Lumen)	MIoU (EEM-CSA)
No Augmentation	0.872	0.703
Flip+Rotate	0.915	0.759
MeshGrid	0.894	0.747
MeshGrid+Flip+Rotate	**0.937**	0.804
Ji Yang et al.[10]	0.900	**0.860**

Discussion

The IVUS images varied significantly from the intensity gradient of edge of lumen to the contour curvature of plaques. The current dataset was limited to the single-vendor and single-brand. However, the proposed method provided acceptable segmentation results for both lumen and EEM-CSA from the frames in the dataset. On the visual comparison, in the case of the EEM-CSA segmentation, the performance was lower for complex frames while it was comparable good for simple frames; excellent results were seen for lumen segmentation for all cases. The segmentation of bifurcation images might be difficult due to the ambiguous vessel definition. The amount of calcified plaque could be possibly underestimated by IVUS and images of the vessel wall could be degraded due to geometric distortions, and shadows at the back of calcified plaques, all of these could make the model difficult to train. The case data including calcified plaque needs to be enriched in future to train a more robust model.

The data augmentation of MeshGrid+Flip+Rotate helped improve the segmentation performance. It was capable to generalize well to eliminate the outliers (3rd row of Figure 3). Neither pre-processing steps nor post-processing steps were necessary. The model was trained well on the current dataset, which provided mIoU of 0.937 for lumen predictions. However, when the testing set deviated far from the training set, such as serious artifacts, mixture plaques and branch vessels, the accuracy for EEM-CSA became relatively low (mIoU of 0.804). It can be improved largely when more coronary IVUS data of different categories are collected for training in the future.

From clinical perspective, a clinical threshold to assess the quality of the method should be provided by expert physicians to interpret the segmentation feature with complex or simple frames. For example, a fast pullback through the calcified lesion may result in loss of image features, increase of catheter artifacts and calcified shadows from the echogenicity of the lumen and plaque textures, which makes it tougher to be annotated even by experienced physicians. The cardiac cycle motion and coronary vessel pulsation due to the variability or arrhythmia of heart rate might push the catheter to touch the vessel and plaque boundaries, which increase the artifacts and motion jitters in the IVUS images.

Our future research is to extend the current dataset to enhance the robustness and generality of our method presented in this paper. The heterogeneous dataset of IVUS images shall cover different medical centers, different probe frequencies from different venders. More IVUS image categories from different artery pullback sections and different characteristics should be considered: plaque, bifurcations, branches, shadow artifact, stent, and catheter artifact, etc. Each frame shall be cross-labeled by three expert physicians according to the respective categories to assess the method, which will make it more convincing.

Conclusion

In this paper, an 8-layer U-Net is proposed with the data augmentation of MeshGrid+Flip+Rotate, which specifically fits for the coronary IVUS lumen and EEM-CSA segmentation task. The experimental results show its superiority in segmentation accuracy and efficiency. Furthermore, it provides a good start for the

image-based gating to implement 3D IVUS reconstruction when fused with X-ray projections, which enables fluid and dynamic analysis on plaques and vascular walls of coronary arteries.

Method

In this section, we first introduce the coronary IVUS dataset used for training and testing. Then, the 8-layer deep U-Net architecture that predicts the masks for the lumen and the EEM-CSA of IVUS images is presented. The training details are described, and the metric for evaluating the proposed method is illustrated.

Dataset and Augmentation

We use the coronary IVUS dataset from The Second Affiliated Hospital of Zhejiang University School of Medicine. It consists of *in vivo* pullback of coronary artery acquired by the iLab IVUS from Boston Scientific Corporation equipped with the 40-MHz OptiCross catheter. It contains IVUS frames from 30 patients, which are chosen at the end-diastolic cardiac phase in DICOM formats, with the resolution of 512×512. The dataset is divided into two parts, 567 frames of 24 patients for training and 108 frames of 6 patients for testing, respectively. The training set is used for building the deep learning model and the testing set is used to evaluate the model performance.

IVUS images contain catheter, lumen, endothelium, intima, media, external elastic membrane, adventitia, atherosclerosis plaque. The external elastic membrane is usually treated as the borders of media and adventitia. The media is gray or dark as it contains dense smooth muscle. The adventitia is similar to external tissues surrounding the vascular walls. The endothelium and intima are thinner than the lumen and media. Thus, the lumen and EEM-CSA can be manually annotated by experienced physicians as the ground truth for metric evaluation. Each IVUS frame has been manually annotated for the lumen and EEM-CSA in the short-axis view by three clinical experts, daily working with the specific IVUS brand from the Cardiology Department, shown in Figure 2. Each expert is blinded to the other two experts' annotations and each frame is repeatedly labeled by each of the three experts to ensure the correctness and blindness of the annotations. From the visual point of view of annotation, 92% of the annotated cases have high consistency.

The training set comprises of 567 frames, which is not large enough for training a CNN model from scratch. Data augmentation is essential for better performance. The augmentation is twofold and performed online. First, the coronary IVUS raw images and the corresponding ground truth are randomly (1) rotated at angles: 90°, 180° or 270°; (2) flipped up-down or left-right. Secondly, the MeshGrid is added to the raw image at pixel-level, providing the relative location information. Due to the relatively fixed position like intima and adventitia in IVUS images, MeshGrid could play a good guiding role in training process, which guides the neural network where to look.

Network Architectures

The U-Net is one type of the fully convolutional network (FCN) [11] and is the most common convolutional network architecture for biomedical image segmentation. It consists of encoder and decoder parts and predicts segmentation mask at pixel-level instead of image-level classification. The encoder part is used for down-sampling and extracts higher-level features. The decoder part is used for up-sampling the output from the encoder part and concatenates the feature maps of the corresponding layer by skip connection. The skip-connection is to relieve the gradient diffusion problem due to deep layers. The final decoder layer is activated by softmax to produce the class probability map to recover the segment predictions.

The encoder part has 9 blocks and each incorporates two repeated operations of 3×3 convolution, batch normalization (BN) and LeakyReLU activation. The down-sampling operation of 3×3 convolution with stride 2×2 reduces feature maps by half. The size of the 8th block is 2×2 to capture the deeper abstract information. The decoder part has 8 blocks to restore the image dimension. Each up-sampling operation contains a 5×5 deconvolution with stride 2. The skip-connection concatenates the corresponding feature maps. The last convolution of 1co outputs the probability map of mask class prediction by softmax activation. The entire architecture is shown in Figure 3. The parameter initialization of all layers of the model uses the random initialization method.

Compared to other U-Net variations, our proposed U-Net was no major innovation in structure. We replaced the original 4-layer network with an 8-layer network, which been able to extract deeper image features.The actual results also confirmed this simple deepening design.

Implementation Details

The model was trained and evaluated on Dell PowerEdge T640 server with Xeon Silver 4114 processor, 128GB of RAM, and four Nvidia GTX 1080Ti graphics cards. It took less than 90 minutes for training and 10ms per image for inference.

We implement model training with TensorFlow framework. The less frames are not enough to train the CNN from scratch, in addition to data augmentation, we also employed transfer learning to initialize the encoder of U-Net's weights using VGG16[12]. The optimizer was Adam[13], which was fast and robust. The weights were initialized randomly and the batch size was set to 16. The initial learning rate was 0.001 with the decay of 0.1 every 2000 iterations. A total of 8000~10000 iterations were done for training. Lumen and EEM-CSA were trained and predicted at one shot with the softmax function as the output activation, which gave each pixel its class probability. The loss function was the sparse softmax cross entropy:

$$L(p_{\hat{y}}, p) = -\sum_{j=1}^{K} p_{\hat{y}_i} \log(p_j) \quad (1)$$

$$p_j = \text{softmax}(x_j) = \frac{e^{x_j}}{\sum_{k=1}^{K} e^{x_k}} \quad (2)$$

with *K* being the number of classes, p_j being the predicted probability belonging to class *j*, p_{y_i} being the true probability.

Evaluation Criteria

In semantic segmentation, the mean intersection over union (mIoU) is a widely used metric to evaluate the model, which is a common measure in semantic segmentation[14]. We compute the mIoU score between the ground truth and the predicted masks:

$$\text{mIoU} = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \quad (3)$$

with *k* being the number of classes excluding background, p_{ij} being the number of pixels of class *i* predicted to class *j*.

Abbreviations

3D: Three-dimensional

IVUS: Intravascular Ultrasound

EEM-CSA: External Elastic Membrane Cross Section Area

mIoU: mean Intersection of Union

FCN: Fully Convolutional Network

Declarations

Acknowledgements

The authors thanks the Second Affiliated Hospital of Zhejiang University for providing the IVUS datasets.

Funding

This study was supported by Zhejiang Provincial Public Welfare Technology Research Project (NO. LGF20H020012).

Ethics approval and consent to participate

This study protocol was approved by the Second Affiliated Hospital of Zhejiang University. We followed CONSORT guideline to perform this study.

Consent for publication

Not applicable.

Competing interests

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

Authors' contributions

Liang Dong:Project administration; Wei Lu:Original draft writing; Jun Jiang:Data curation; Ya Zhao:Original draft writing; Xiangfen Song:Formal analysis and paper revise; Xiaochang Leng:Data curation; Hang Zhao: Formal analysis and paper revise; Jian'an Wang:Project administration; Changling Li:Project administration; Jianping Xiang:Project administration.

Availability of data and materials

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

References

[1] S. Hay, "Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016," *The Lancet,* vol. 390, no. 10100, 2017.

[2] M. Faraji, I. Cheng, I. Naudin, and A. Basu, "Segmentation of arterial walls in intravascular ultrasound cross-sectional images using extremal region selection," *Ultrasonics,* vol. 84, pp. 356-365, 2018.

[3] J. H. Lee, Y. N. Hwang, G. Y. Kim, and K. S. Min, "Segmentation of the lumen and media-adventitial borders in intravascular ultrasound images using a geometric deformable model," *IET Image Processing,* vol. 12, no. 10, pp. 1881-1891, 2018.

[4] H. Cui *et al.*, "Fast marching and Runge–Kutta based method for centreline extraction of right coronary artery in human patients," *Cardiovascular engineering and technology,* vol. 7, no. 2, pp. 159-169, 2016.

[5] H. Cui, Y. Xia, Y. Zhang, and L. Zhong, "Validation of right coronary artery lumen area from cardiac computed tomography against intravascular ultrasound," *Machine Vision and Applications,* vol. 29, no. 8, pp. 1287-1298, 2018.

[6] D. Shen, G. Wu, and H.-I. Suk, "Deep Learning in Medical Image Analysis," *Annual Review of Biomedical Engineering,* vol. 19, no. 1, pp. annurev-bioeng-071516-044442.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature,* vol. 521, no. 7553, pp. 436-444, 2015.

[8] S. Su, Z. Gao, H. Zhang, L. Qiang, and S. Li, "Detection of lumen and media-adventitia borders in IVUS images using sparse auto-encoder neural network," in *International Symposium on Biomedical Imaging (ISBI)*, 2017.

[9] S. Su, Z. Hu, Q. Lin, W. K. Hau, Z. Gao, and H. Zhang, "An artificial neural network method for lumen and media-adventitia border detection in IVUS," *Computerized Medical Imaging and Graphics,* vol. 57, pp. 29-39, 2017.

[10] 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: Springer, pp. 234-241.

[11] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE transactions on pattern analysis and machine intelligence,* vol. 39, no. 4, pp. 640-651, 2017.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.

[13] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.

[14] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *arXiv preprint arXiv:2001.05566,* 2020.

Figures

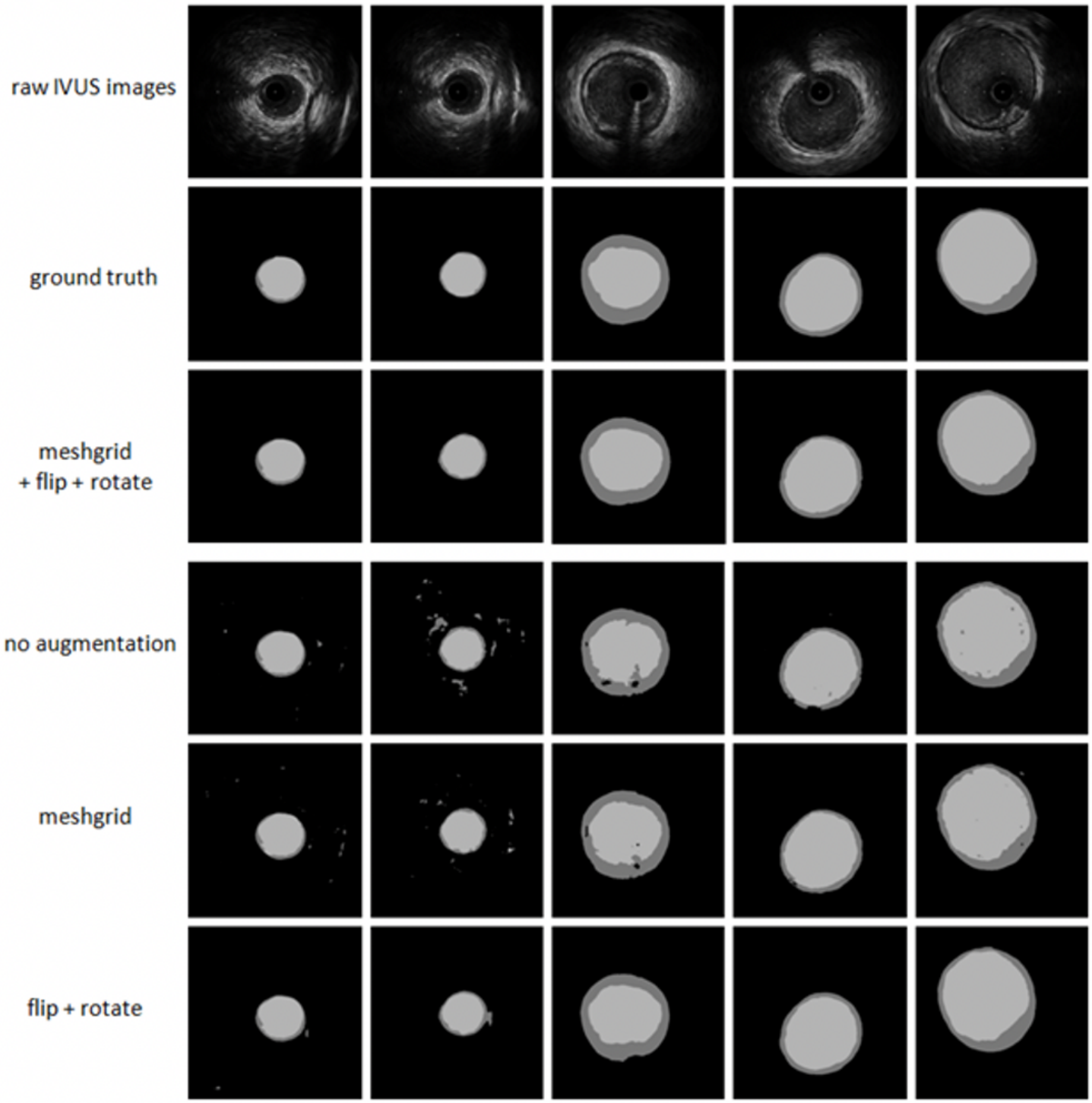

Figure 1

Result comparison for four data augmentation strategies.

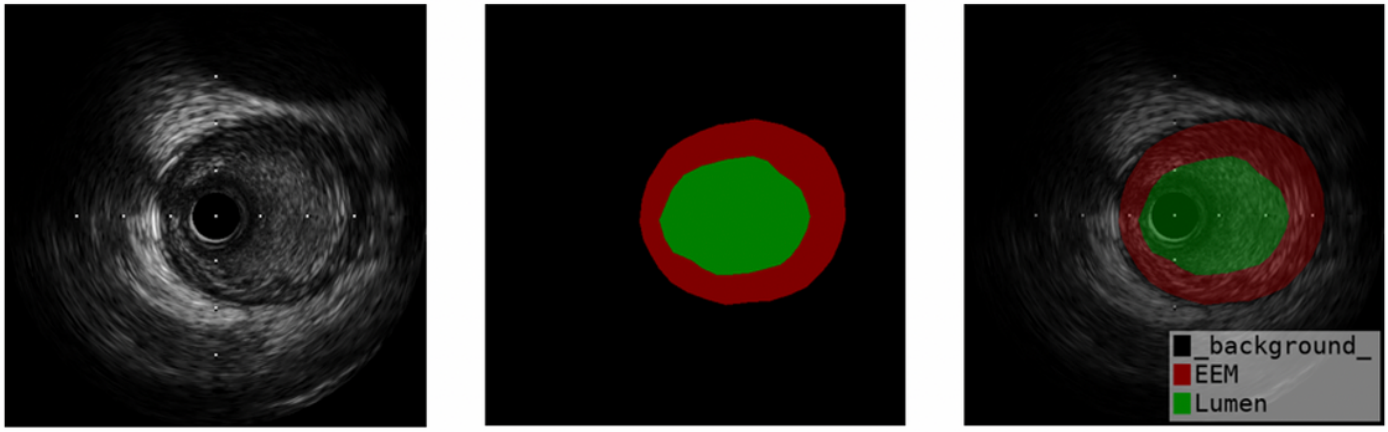

Figure 2

Ground truth labeling for lumen and EEM-CSA

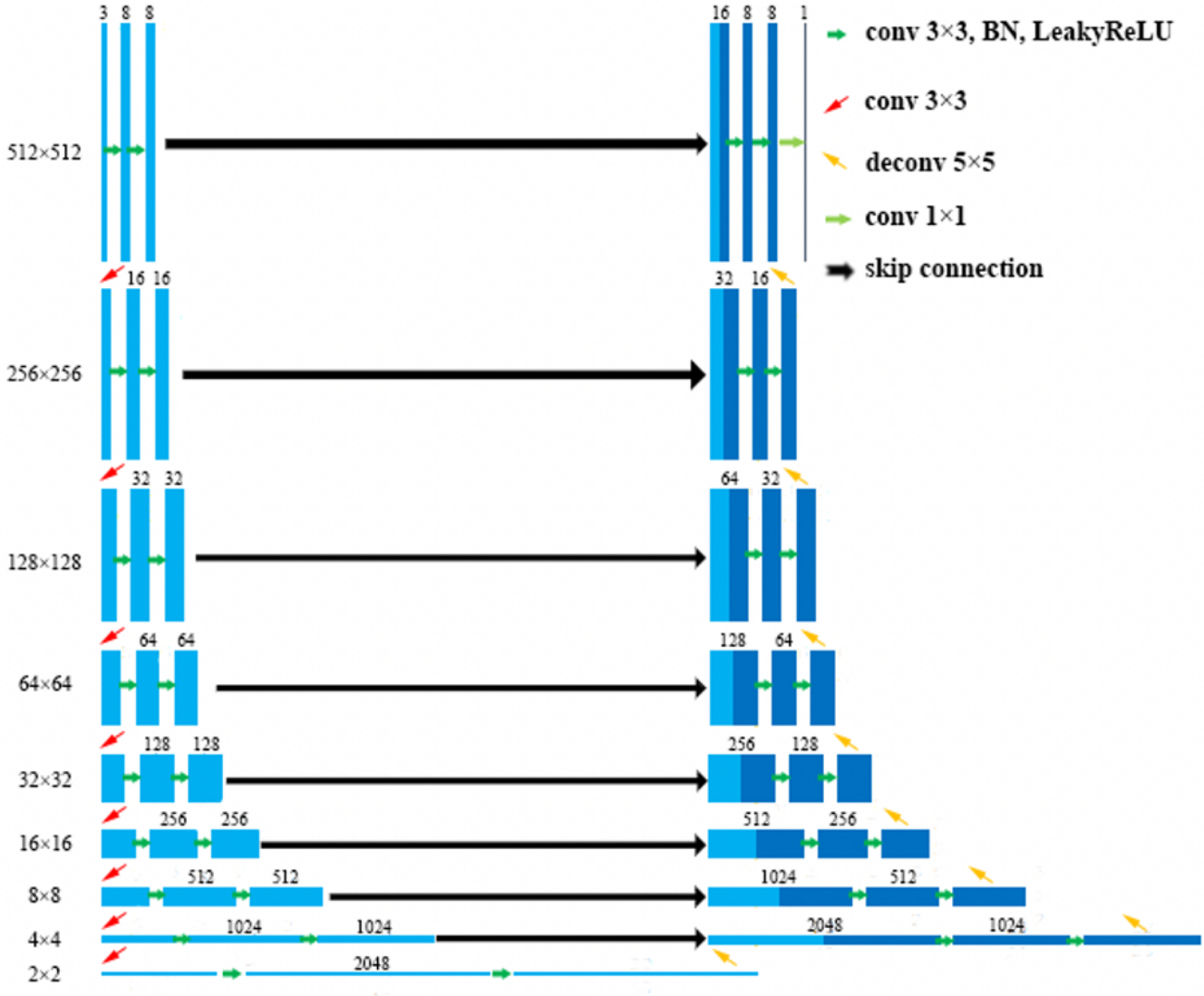

Figure 3

The proposed U-Net Architecture with eight layers