

# Precise segmentation of offshore farms in high-resolution SAR Images based on improved UNet++

Chuang Yu<sup>\*a,b,c,d</sup>, Yunpeng Liu<sup>#b</sup>, Xin Xia<sup>§b,d</sup>

<sup>a</sup>Key Laboratory of Opto-Electronic Information Processing, Chinese Academy of Sciences, Shenyang 110016, China; <sup>b</sup>Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China; <sup>c</sup>Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China; <sup>d</sup>School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China

## ABSTRACT

The segmentation of the offshore farm area in the high-resolution SAR image is of great significance for the statistics of the farming area and the analysis of the rationality of the farming layout. However, the SAR images have the characteristics of a lot of noise and inconspicuous features. It is difficult to achieve precise segmentation by directly using non-learning image segmentation methods. Therefore, we propose a precise segmentation scheme for offshore farms in high-resolution SAR images based on improved UNet++. We first adopt a simulated annealing strategy for the update of the learning rate during the network training. By initializing the learning rate multiple times, we avoid the network from falling into a local optimum. Secondly, for the data set studied, we verify that the segmentation performance of resizing the image to 256×256 pixels is better than that of 512×512 pixels. Finally, we propose an improved UNet++, which uses SE-Net as the feature extraction network to enhance the feature learning ability. Extensive experimental results show that, compared to some state-of-the-art methods, the proposed scheme achieves superior performance with a Frequency Weighted Intersection over Union (FWIoU) of 0.9853 on the high-resolution SAR offshore farm data set.

**Keywords:** Segmentation, offshore farms, sar image, improved UNet++

## 1. INTRODUCTION

In aquaculture, excessive aquaculture density and improper cage layout will cause the deterioration of water quality, which is not conducive to sustainable development<sup>1</sup>. At the same time, relying solely on manual measurement consumes a lot of manpower and material resources. Therefore, the use of remote sensing images to achieve accurate segmentation of the target farm area is of great significance<sup>2</sup>.

In the early stage, the image segmentation methods are mainly based on edge detection<sup>3</sup>, region growth<sup>4</sup>, threshold processing<sup>5</sup> and energy functional<sup>6</sup> to segment the target area. Although the above non-learning traditional segmentation methods can segment the specific target in the specific scene, the segmentation performance is easily affected by the environment and the imaging source. SAR images with a lot of noise and lack of texture and color information will be affected a lot. Using traditional image segmentation methods will result in poor segmentation effects<sup>7</sup>. With the emergence of deep learning networks, its powerful feature extraction capabilities have been verified in many fields. FCN<sup>8</sup> is one of the earliest methods of deep learning methods applied in the field of image segmentation. It has achieved excellent results in image segmentation tasks and promoted the application of deep learning methods in this field. Subsequently, many deep learning segmentation methods such as U-Net<sup>9</sup>, Linknet<sup>10</sup> and UNet++<sup>11</sup> have been proposed, which continuously improves the segmentation accuracy under different scenes and different image sources.

In order to improve the segmentation accuracy of deep learning methods, an effective feature extraction network structure is crucial<sup>12</sup>. Simonyan et al. propose VGG16 and VGG19<sup>13</sup>. They use several consecutive 3×3 convolution kernels to replace the larger convolution kernel in AlexNet. However, the networks use the fully connected layers, which require a large amount of calculation. He et al. propose the Resnet network<sup>14</sup>, which design the residual structure to solve the gradient disappearance or explosion caused by the superposition of network layers. In order to further improve the

\*yuchuang1205@163.com; #ypliu@sia.cn; §xiaxin\_krist@163.com

feature extraction ability of the network, Jie et al. introduce the attention mechanism into Resnet and propose SE-Net, which can help the network focus on learning important feature information<sup>15</sup>. In addition to the above-mentioned effective feature extraction networks, Xception<sup>16</sup>, inceptionv4<sup>17</sup>, etc. have also achieved good feature extraction results on some public data sets.

Considering the powerful learning ability of deep learning networks and the importance of feature extraction networks, we propose a precise segmentation scheme for offshore farms in high-resolution SAR images based on improved UNet++. The contributions of this research are as follows:

- (1) The simulated annealing strategy is used to update the learning rate during network training. It reduces the network weight updates into a local optimum, and promotes the positive update of the network weight.
- (2) For the data set studied, we verify that the segmentation performance of resizing an image to 256×256 pixels is better than that of 512×512 pixels. It can achieve better segmentation accuracy in a shorter time-consuming.
- (3) An improved UNet++ is proposed. We compare the segmentation effects of multiple feature extraction networks. SE-Net is selected as the feature extraction network to replace the encoding part of UNet++.

The rest of this research is arranged as follows: Section 2, the proposed scheme is introduced in detail. Section 3 elaborates on the experiments and experimental results. Section 4 summarizes the full text.

## 2. METHOD

### 2.1 The proposed scheme

In order to achieve precise segmentation of offshore farms in high-resolution SAR images, we propose a segmentation scheme based on improved UNet++. The overall segmentation flowchart is shown in Figure 1. From Figure 1, the scheme can be divided into model training and model application. In the model training stage, firstly, the training samples are resized to 256×256 pixels. Secondly, performing data augmentation operations on the samples to expand the sample set and reduce the overfitting of the training model<sup>18</sup>. Then, the enhanced data is input into the improved UNet++, and the simulated annealing strategy is used to constrain the learning rate of the network. Finally, the best trained model is chose by the validation samples. In the model application stage, the test set is resized to 256×256 pixels and input into the trained model for testing. Then, resizing the segmentation image to the size of the original image, which is the required binary segmentation image.

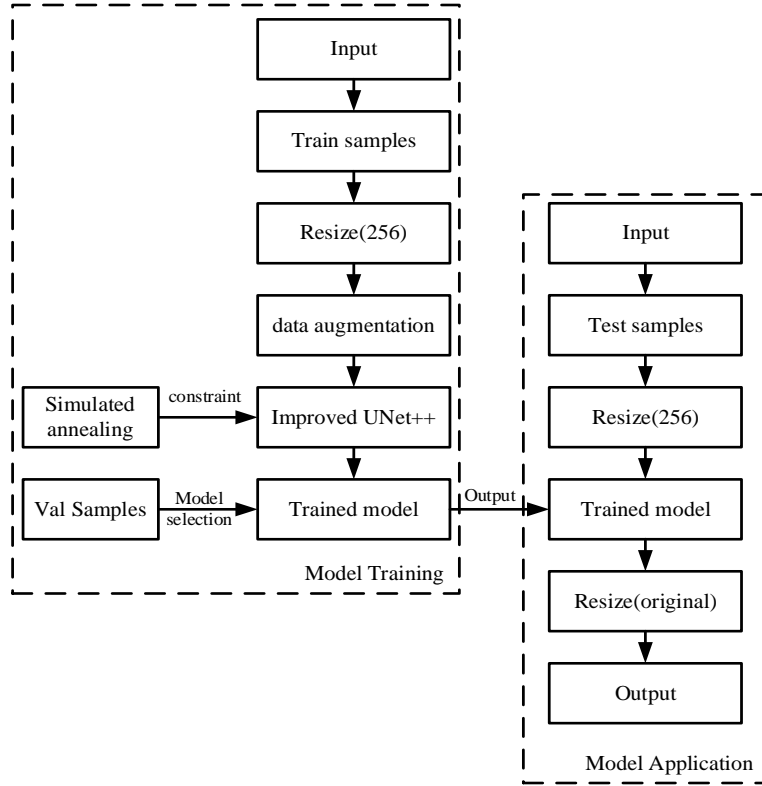


Figure 1. Overall scheme. “Resize(256)” denotes to resize the image to 256×256 pixels. “Resize(original)” denotes to resize the image to the original size.

## 2.2 Simulated annealing strategy

It is very easy for the network training to fall into the local optimal during the training. Therefore, we introduce a simulated annealing strategy<sup>19</sup> for the update of the learning rate, as shown in equations (1) and (2).

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min})(1 + \cos(\frac{T_{cur}}{T_i} \pi)) \quad (1)$$

$$T_i = T_0 \beta^i \quad (2)$$

where  $\eta_t$  denotes the current learning rate.  $\eta_{\max}$  and  $\eta_{\min}$  respectively denote the set initial learning rate and minimum learning rate.  $T_0$  is the epoch of the first restart.  $\beta$  denotes the multiple of epochs compared to the previous restart interval, which is a constant.  $T_{cur}$  is the number of epochs since the last restart and  $T_i$  is the number of epochs between two restarts.

As shown in Figure 2, it shows the learning rate changes with epochs when  $\eta_{\max} = 0.0003$ ,  $\eta_{\min} = 0.00001$ ,  $T_0 = 3$ ,  $\beta = 2$ . The change in the learning rate is rippled, and the number of epochs between restarts every two times increases multiply. When the weights generated by epoch training fall into the local optimum, initializing the learning rate will help it jump out of the local optimum. And the epochs between the next two restarts are getting larger and larger, which provides a sufficient number of iterations for the weight to converge to the minimum.

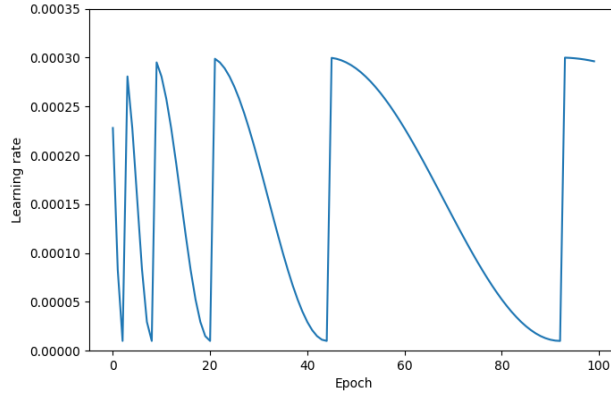


Figure 2. Using simulated annealing strategy, the change of learning rate during network training.

### 2.3 Improved UNet++

In order to enhance the feature extraction capability of the original UNet++ network, we integrate SE-Net as a feature extraction network structure into UNet++, and propose an improved UNet++. As shown in Figure 3, our improved UNet++ uses SE-Resnet101 (SE-Net) as the feature extraction network module, and replaces the encoding part of the original UNet++ with the feature map output from the five stages of SE-Resnet101. SE-Resnet101 is formed by adding an SE attention module that distributes weights to each channel on the Resnet101 network structure, which can pay more attention to the feature information of the target. At the same time, our proposed network is based on UNet++. UNet++ is an improvement based on U-Net. It has excellent performance in segmentation tasks and is widely used.

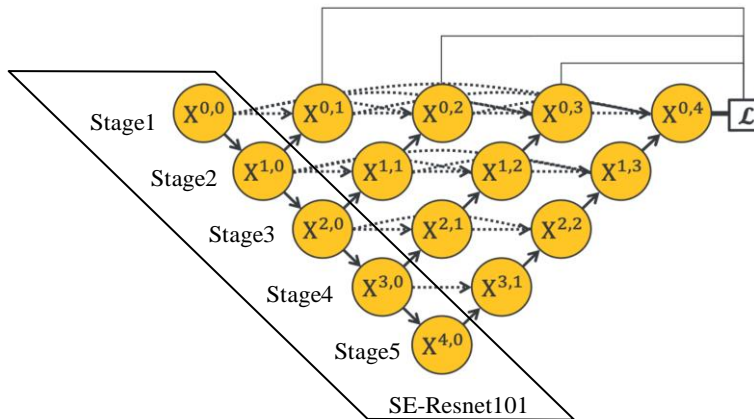


Figure 3. the network structure of Improve UNet++.

### 2.4 Loss function and loss optimization

For the research task, simply using the binary cross-entropy loss function will cause the loss value to be low but the segmentation effect is not accurate enough. Therefore, this research uses a combination of binary cross-entropy loss and dice loss<sup>20</sup>, and its expression is as shown in equation (3).

$$\varepsilon(Y, P) = -\frac{1}{N} \sum_{c=1}^C \sum_{n=1}^N (y_{n,c} \log p_{n,c} + \frac{2y_{n,c}p_{n,c}}{y_{n,c}^2 + p_{n,c}^2}) \quad (3)$$

where  $p_{n,c} \in P$  and  $y_{n,c} \in Y$  denote the target labels and predicted probabilities of the  $c$ -th classes and  $n$ -th pixels in the batch processing, respectively.  $Y$  and  $P$  denote the true value and prediction result of the test image, respectively.  $C$  and  $N$  denote the number of classes and pixels, respectively.

Using cross-entropy loss and dice loss integrated loss function can promote the rapid convergence of the model.  $\varepsilon(Y, P)$  will be continuously updated and gradually approach 0 during the network training.

### 3. METHOD

#### 3.1 Data set

The high-resolution SAR offshore farm data set is derived from the SAR data of Haisi No. 1 and Gaofen No. 3 with a resolution of 1-3m. The entire data set has a total of 4000 images with sizes ranging from 512 to 2048 pixels, and most of the samples are 512×512 pixels. The scene covers the coastal areas of southeastern China, and the target includes large-scale seafood farms commonly found in the coastal waters of China. Figure 4 shows part of the sample images and annotated images. We can find that the SAR image has a lot of noise and the features are not obvious.

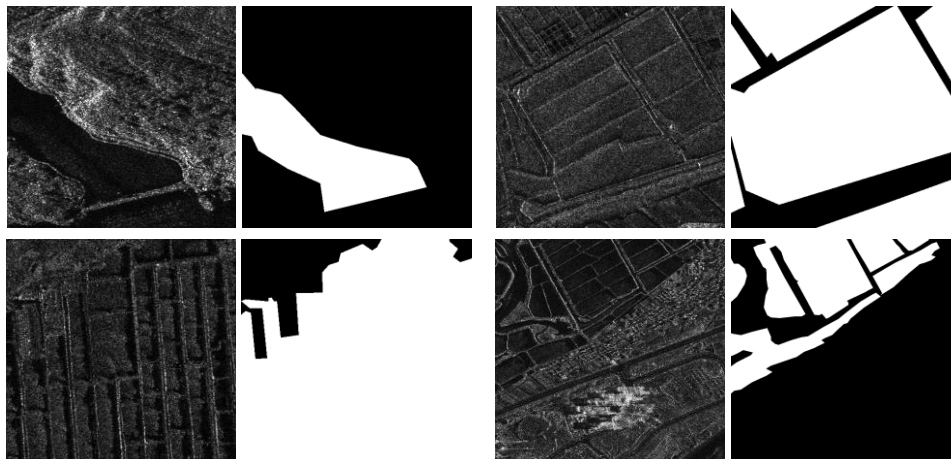


Figure 4. Some sample samples from the high-resolution SAR offshore farm data set.

#### 3.2 Experimental environment and parameter settings

The experimental environment of this research is ubuntu18.04 operating system. GPU is RTX 2080Ti 11G. The epochs are 100, the batch sizes are 8. The learning rate is 0.0003, the momentum is 0.9, and the weight decay is 0.0005. The initial number of rounds of simulated annealing is 3. The number of epochs between two restarts increases by multiples.

The optimization of the weights in the deep learning network requires a large number of training samples. In the experiment, we randomly use rotation transformation and flip transformation to enhance the data of the training samples to improve the robustness of the generated model.

In view of the evaluation metric for the local test set, we choose FWIoU<sup>21</sup> as the evaluation metric, as shown in equation (4).

$$FWIoU = \frac{1}{\sum_{i=0}^k \sum_{j=0}^k P_{ij}} \frac{\sum_{j=0}^k P_{ij} P_{ii}}{\sum_{j=0}^k P_{ij} + \sum_{j=0}^k P_{ji} - P_{ii}} \quad (4)$$

where  $P_{ii}$  denotes it was originally the  $i$ -th class, and it is predicted to be the  $i$ -th class;  $P_{ij}$  denotes it is originally the  $i$ -th class, but it is predicted to be the  $j$ -th class;  $P_{ji}$  denotes it is originally the  $j$ -th class, but it is predicted to be the  $i$ -th class.  $k$  denotes the total number of classes.

#### 3.3 Effect verification of simulated annealing strategy

In order to better verify the effectiveness of the simulated annealing strategy, both U-Net and UNet++ have been verified in the experiment. The experimental results are shown in Table 1. The images in the experiment are all resized to 256×256 pixels for experiment.

Table 1. Comparison of segmentation accuracy of whether to use simulated annealing strategy under U-Net and UNet++ networks.

Methods	Without simulated annealing	With simulated annealing
U-Net	0.9569	<b>0.9793</b>
UNet++	0.9628	<b>0.9819</b>

It can be seen from Table 1 that using the simulated annealing strategy, the segmentation performance of U-Net and UNet++ has been significantly improved. The U-Net and UNet++ respectively increased by 0.0224 and 0.0191 in FWIoU. At the same time, by observing the segmentation performance of U-Net and UNet++, it is found that the segmentation performance of the UNet++ network is better than that of U-Net.

### 3.4 Effect verification of resize strategy

For the high-resolution SAR offshore farm data set, we consider that the segmentation performance of uniformly resizing the image to 256×256 pixels for training is better than resizing the image to 512×512 pixels. In order to better verify this idea, we use different resize strategies on U-Net and UNet++, and compare the final segmentation performance. The experimental results are shown in Table 2. The simulated annealing strategy is used in the experiments.

Table 2. The impact of different resize strategies on segmentation accuracy under U-Net and UNet++ networks.

Methods	512	256
U-Net	0.9759	<b>0.9793</b>
UNet++	0.9788	<b>0.9819</b>

Note: “512” denotes to resize the image to 512×512 pixels. “256” denotes to resize the image to 256×256 pixels.

From Table 2, the segmentation accuracy of resizing to 256×256 pixels is better than that of 512×512 pixels under U-Net and UNet++. The U-Net and UNet++ effect respectively increase by 0.0034 and 0.0031 in FWIoU. On the one hand, by resizing the image to 256×256 pixels, the receptive field of convolution becomes relatively larger. On the other hand, the target area of data set is all in blocks, and the edges are relatively flat. Using a smaller size image input can increase the robustness of the input image to some small disturbances.

In addition to the comparative experiments of segmentation accuracy, we also compare the test time. As shown in Table 3, it will take less time to resize the image to 256×256 pixels. The speed of U-Net has increased by 20% (from 0.02 to 0.016), and the speed of UNet++ has increased by 33% (from 0.027 to 0.018).

Table 3. The impact of different resize strategies on test time under U-Net and UNet++ networks.

Methods	Time on GPU/s
U-Net (256)	<b>0.016</b>
U-Net (512)	0.020
U-Net++(256)	<b>0.018</b>
U-Net++(512)	0.027

Note: “512” denotes to resize the image to 512×512 pixels. “256” denotes to resize the image to 256×256 pixels.

### 3.5 Effect verification of improved U-Net++

In order to verify the segmentation performance of the proposed improved U-Net++, we compare it with multiple methods. The experimental results are shown in Table 4.

Table 4. Comparison of different segmentation methods.

Methods	FWIoU
U-Net	0.9793
UNet++	0.9819
UNet++(Resnet101)	0.9813
UNet++(xception)	0.9836
UNet++(inceptionv4)	0.9841
Improved UNet++	<b>0.9853</b>

From Table 4, the proposed improved UNet++ has achieved relatively optimal segmentation accuracy. And compared to the original UNet++, the proposed network has an increase of 0.0034 in FWIoU. In order to show the segmentation effect of the proposed network more intuitively, the mask generated after the experimental segmentation is drawn to the original image as shown in Figure 5.

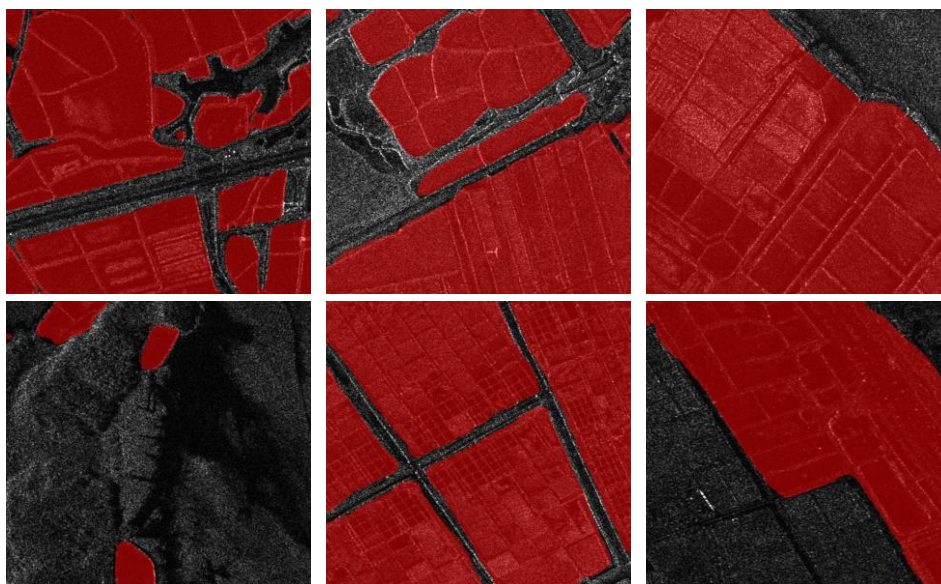


Figure 5. The segmentation effect of the test set under the improved UNet++.

From Figure 5, although there is a small part of the mis-segmentation area between two adjacent farm areas in some images, the segmentation boundaries of most images are relatively accurate. It shows that the overall segmentation effect of the proposed scheme is very good.

#### 4. CONCLUSION

Using high-resolution SAR images to accurately segment the area of offshore farms can assist in the statistics of farming density and the planning of the farming layout. A precise segmentation scheme for offshore farms in high-resolution SAR images based on improved UNet++ is proposed. We firstly adopt the simulated annealing strategy to update the learning rate during network training. It can reduce the network falling into a local optimum and promote the positive update of the network weights. Secondly, we verify that, for the data set studied, the segmentation effect of resizing image to  $256 \times 256$  pixels is better than that of  $512 \times 512$  pixels. This operation relatively increases the receptive field of each layer of the network, and increases the model's robustness to some disturbances such as noise. Finally, we compare the segmentation effect of using multiple different feature extraction networks to replace the encoding part of UNet++, and construct an improved UNet++, which uses SE-Net as the feature extraction network. The experimental results show

that on the high-resolution SAR offshore farm data set, the proposed improved UNet++ scheme is as high as 0.9853 in FWIoU.

## ACKNOWLEDGMENTS

This research is supported by the Innovation Project of Equipment Development Department--Information Perception Technology under Grant no. E01Z040601.

## REFERENCES

- [1] Zhou, W. Z., Zhang, Y. Y., Tu, W. L., Wang, H. Y., Cao, J. G., Wu, H. L., Lv, W. W. and Tan, Y. S., "Effect of the density of cage culture on the growth performance of *Whitmania pigra*," *Jiangsu Agricultural Science*, 46(24), 194-19 (2018). (in Chinese)
- [2] Yu, C., Hu, Z. H., Li, R. Q., Xia, X., Zhao, Y. C., Fan, X. and Bai, Y., "Segmentation and density statistics of mariculture cages from remote sensing images using mask R-CNN," *Information Processing in Agriculture* (In press), (2021).
- [3] Arbelaez, P., Maire, M., Fowlkes, C. and Malik, J., "Contour detection and hierarchical image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5), 898-916 (2011).
- [4] Wang, Y., Wang, L. C., Zheng, Y. F. and Lei, T., "Aurora image segmentation method based on region growing," *Computer Engineering and Applications*, 52(23), 190-195 (2016).
- [5] Wang, H. W., Liang, Y. Y. and Wang, Z. H., "Otsu image threshold segmentation method based on new genetic algorithm," *Laser Technology*, 38(3), 364-367 (2014).
- [6] Cui, W. C., Gong, G. Q., Lu, K., Sun, S. F. and Dong, F. M., "Convex-relaxed active contour model based on localised kernel mapping," *IET Image Processing*, 11(11), 976-985 (2017).
- [7] An, Y., Long, J. W. and Mabu, S., "A segmentation network with multiattention and its application to SAR image analysis," *IEEE Transactions on Electrical and Electronic Engineering*, 15(4), 570-576 (2020).
- [8] Shelhamer, E., Long, J. and Darrell, T., "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), 640-651 (2017).
- [9] Olaf, R., Philipp, F. and Thomas, B., "U-Net: Convolutional networks for biomedical image segmentation," *18th Inter. Conf. on Medical Image Computing and Computer-Assisted Intervention*, 234-241 (2015).
- [10] Abhishek, C. and Eugenio, C., "LinkNet: Exploiting encoder representations for efficient semantic segmentation," *2017 IEEE Visual Communications and Image Processing*, 1-4 (2017).
- [11] Zhou, Z. W., Siddiquee, M. M. R., Tajbakhsh, M. and Liang, J. M., "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, 39(6), 1856-1867 (2019).
- [12] Krizhevsky, A., Sutskever, I. and Hinton, G., "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, 25(2), (2012).
- [13] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," *Computer Science*, (2014).
- [14] He, K. M., Zhang, X. Y., Ren, S. Q. and Sun, J., "Deep residual learning for image recognition," *29th IEEE Con. on Computer Vision and Pattern Recognition*, 770-778, (2015).
- [15] Jie, H., Li, S. and Gang, S., "Squeeze-and-excitation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99, (2017).
- [16] Chollet, F., "Xception: Deep learning with depthwise separable convolutions," *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) IEEE*, 1800-1807 (2017).
- [17] Szegedy, C., Loffe, S., Vanhoucke, V. and Alemi, A., "Inception-v4, inception-ResNet and the impact of residual connections on learning," *31st AAAI Conf. on Artificial Intelligence*, 4278-4284 (2016).
- [18] Zoph, B., Cubuk, E. D., Ghiasi, G., Lin, T. Y., Shlens, J. and Le, Q. V., "Learning data augmentation strategies for object detection," *Lecture Notes in Computer Science*, 566-583 (2020).
- [19] Loshchilov, I. and Hutter, F., "SGDR: Stochastic gradient descent with warm restarts," *5th Inter. Conf. on Learning Representations*, (2016).



- [20] Milletari, F., Navab, N. and Ahmadi, S. A., "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," Proc. of 4th Inter. Conf. on 3D Vision, 565-571 (2016).
- [21] Alberto, G. G., Sergio, O. E., Sergiu, O., Victor, V. M. and Jose, G. R., "A review on deep learning techniques applied to semantic segmentation," Comput. Vis. Pattern Recognit., 48(5), 644-654 (2017).