# FRNet: A Full-Resolution Convolutional Neural Network for OCTA Vascular Segmentation


1st Liang Wang
*Shenyang Ligong University*
*School of Automation and Electrical Engineering*
Shenyang,China
wangliang@stu.sylu.edu.cn

2nd Dongxu Gao*
*University of Portsmouth*
*School of Computing*
England,Britain
dongxu.gao@port.ac.uk

3rd Hongwei Gao
*Shenyang Ligong University*
*School of Automation and Electrical Engineering*
Shenyang,China
30963915@qq.com

4th Guozhang Jiang
W*uhan University of Science and Technology*
*Key Laboratory of Metallurgical Equipment and Control Technology of Ministry of Education*
Wuhan,China
whjgz@wust.edu.cn

5th Zhaojie Ju
*University of Portsmouth*
*School of Computing*
England,Britain
juzhaojie@gmail.com



***Abstract:*** **Optical coherence tomography angiography (OCTA) is an advanced, non-invasive imaging technique. Automatic segmentation of vascular networks in retinal OCTA images is essential for the early diagnosis and progression assessment of various vision-related diseases. However, most existing methods for OCTA image vessel segmentation rely on encoder-decoder architectures, which typically involve a high number of parameters and result in slower inference speeds.In this paper, we introduce FRNet V2, an accurate and efficient neural network specifically designed for retinal vessel segmentation in OCTA images. This is achieved by integrating a modified recursive ConvNeXt V2 block and a recursive HAAM attention mechanism into a full-resolution convolutional network framework. We evaluate our proposed method on two large public datasets: ROSSA and OCTA-500. Experimental results demonstrate that our network achieves performance on par with other methods, while boasting significantly fewer parameters and faster inference speeds. This makes it highly suitable for practical industrial applications.**



***Index Terms:*** **Optical coherence tomography angiography; Vascular segmentation; Neural networks; ConvNeXt V2; Dataset**


## I. Introduction

The retina is a critical physiological tissue within the eye, and its health is closely linked to overall ocular well-being. The retinal vascular network is a vital component, and its morphological changes can not only aid in the identification and classification of systemic, metabolic, and hematological diseases[1], but also help to better evaluate disease progression and assess the effectiveness of treatments[2]. Consequently, the quantification of retinal vessels can assist ophthalmologists in clinical diagnosis and research, tracking disease progression, and monitoring treatment responses[3, 4]. The quantification of retinal vessels is closely tied to advancements in observational tools and improvements in algorithmic performance.

Optical Coherence Tomography Angiography (OCTA) is an advanced, non-invasive ophthalmic imaging technology that utilizes flowing red blood cells as a natural contrast agent. It leverages the relative movement of red blood cells and surrounding tissues as a labeling feature for blood flow. Combining OCT imaging technology, dynamic scattering signals are collected in three-dimensional space within biological tissues. By analyzing the dynamic characteristics of these scattering signals, dynamic blood flow is identified and static surrounding tissues are eliminated, achieving motion-contrast-based, unlabeled angiography. This represents a groundbreaking non-invasive angiography technology[5].

OCTA (Optical Coherence Tomography Angiography) has demonstrated significant value in the diagnosis, monitoring, and evaluation of treatment efficacy for a variety of retinal diseases, including but not limited to: age-related macular degeneration (AMD)[6], diabetic retinopathy (DR)[7], venous occlusion, glaucoma[8], and retinal vasculitis associated with autoimmune diseases.

The segmentation of retinal OCTA images involves the automatic or semi-automatic differentiation of different retinal layers and vascular structures within the acquired OCTA images. This is crucial for accurate analysis and quantification of retinal diseases. Figure 1 (a) shows OCTA images, and Figure 1(b) displays the corresponding segmented images.


* Corresponding Author

Currently, retinal OCTA segmentation is largely based on deep learning. Deep learning algorithms can automatically extract features, reducing the dependence on medical prior knowledge and significantly improving the segmentation outcome. However, most of the current deep learning-based methods[9, 10, 11] employ encoder-decoder architectures for segmenting OCTA retina images.

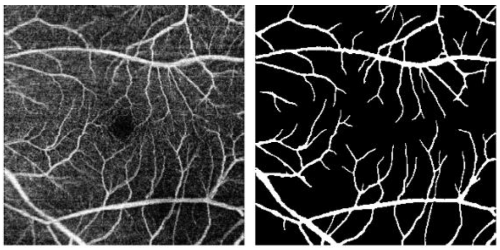


Fig. 1 OCTA image and its segmentation

At present, retinal OCTA segmentation is largely based on deep learning. Deep learning algorithms can automatically extract features, reducing the reliance on medical prior knowledge and significantly improving the segmentation outcome compared to traditional methods. However, most of the current deep learning-based methods[9, 10, 11] employ encoder-decoder architectures for segmenting OCTA retina images.

Although the aforementioned methods achieve high accuracy, they often involve a large number of parameters and slower inference speeds, which can be inconvenient for industrial applications. With the advancement of deep learning technologies, Transformers[12] have shown a trend towards replacing convolutional networks in the 2020s. However, the work of Liu et al. on ConvNeXt [13] demonstrates that rationally designed convolutional networks can outperform Transformers. Additionally, Sanghyun et al. proposed a complete convolutional mask self-encoder framework and a new Global Response Normalization (GRN) layer, which can be integrated into the ConvNeXt architecture, known as ConvNeXt V2[14]. Compared to pure ConvNets, ConvNeXt V2 shows significant performance improvements across various recognition benchmarks.

In order to address these issues and reduce the diagnosis time for ophthalmologists, it is necessary to design a lightweight and high-precision OCTA segmentation network. In this article, we make the following contributions:

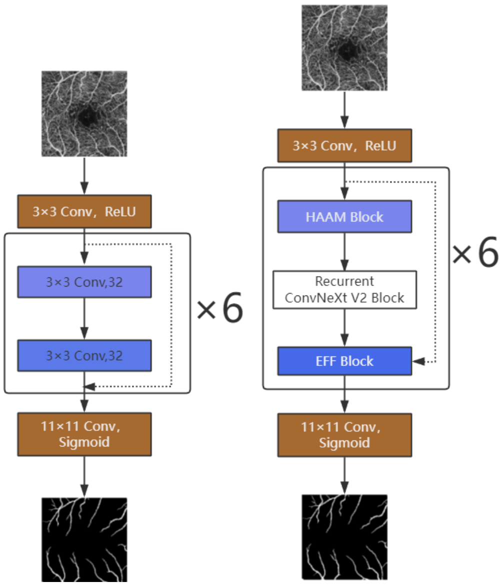
- A full-resolution convolutional network (FRNet V2) is proposed. By utilizing the modified ConvNeXt V2[14] as the core model, this network incorporates deep separable convolutions and recursive convolutions, avoiding up-sampling and down-sampling operations. This results in a model with fewer parameters and higher inference speed, making it highly suitable for medical ophthalmology diagnosis.
- By combining the deeply separable HAAM attention mechanism[20] with the modified ConvNeXt V2, the model can focus more effectively on the features that are most relevant to the current task by assigning different weights to different parts of the input. As a new hybrid adaptive attention module, the HAAM Attention Module can adaptively select receptive fields of different scales from both channel and spatial dimensions. We replace the ordinary convolution in this module with a multi-scale deeply separable recursive convolution to form channel and spatial attention, which reduces the computational complexity of the model and improves its segmentation accuracy.

The structure of this paper is as follows: In Section 2, the proposed approach is elaborated. In Section 3, we introduce experimental data and metrics, discuss datasets and metrics used to evaluate model performance, and compare our model with other models. Section 4 focuses on ablation studies. Section 5 gives the overall conclusion.

## II. Research Methodology

OCTA images contain many tiny blood vessels that may consist of just a few pixels. For the widely used encoder-decoder architecture, the image is first downsampled and then upsampled, which makes it difficult to fully recover the pixels of these small vessels and affects the segmentation accuracy. In addition, each down-sampling operation often results in doubling the number of channels in the convolution layer, leading to a model with a large number of parameters. This is why we believe that the encoder-decoder model is not the most suitable for OCTA vascular segmentation.To solve these problems, we take inspiration from [15], which proposes a full-resolution network (FRNet-base). As shown in Fig. 2 (a), FRNet-base consists of convolution modules without any downsampling or upsampling modules. These convolution blocks have the same structure as the BasicBlock in ResNet [16], where the residual connection is added after two convolutions. FRNet-base contains six such convolution blocks, each with a channel count of 32. The advantages of this design are:(1) Reduced Number of Parameters: The total number of convolution channels is greatly reduced, resulting in a significant reduction in the number of parameters.(2) Preservation of Small Vessel Information: It avoids the loss of small blood vessel information during the downsampling process, enabling higher accuracy in segmentation.

We drew inspiration from ConvNeXt V2 [14] to further enhance FRNet-base. Based on the ConvNeXt V2 block, we set the convolution channel to 32. To increase the receptive field, we replaced the $1 \times 1$ pixel convolution with $3 \times 3$ convolution, and added the HAAM attention module before the ConvNeXt V2 block. Finally, we used the EFF feature fusion module to perform a feature fusion operation. The complete FRNet V2 model is shown in Figure 2 (b).In addition, we also applied the idea of recursive convolution in this module. Recursive convolution means that the input is looped through the convolution layer R times. In [17], it was shown to be very effective for vascular segmentation. In our model, we set R = 2. The structure of the cyclic ConvNeXt V2 block is shown in Figure 3.




（a） （b）

Fig .2 (a) FRNet-base model [15](b) FRNet V2 model

In this paper, we added a modified version of the HAAM attention module in front of the ConvNeXt V2 module. HAAM (Hybrid Adaptive Attention Module) is a hybrid adaptive attention mechanism. Instead of directly adding the HAAM module, we modified it by replacing the 3 × 3 and 5 × 5 convolutions with deep separable convolutions, and then combining them with recursive convolutions. Here, we set R to 2. The modified HAAM attention mechanism is shown in Figure 4.After replacing it with deeply separable recursive convolution, the number of parameters and computational requirements of the model are greatly reduced, accelerating the model's running speed, reducing hardware resource consumption, and decreasing the risk of overfitting. This modification aligns well with our goal of creating a lightweight and efficient model.

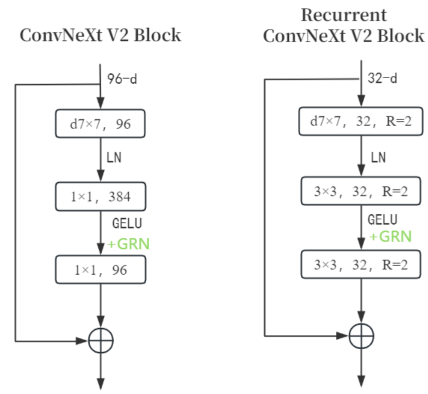



Fig. 3 Original ConvNeXt V2 block and modified recursive ConvNeXt V2 block

## III. Experimental Result

### A. Experimental data

To evaluate the approach in this paper, it was validated on public datasets of two OCTAs, namely, the OCTA-500 dataset [18] and the ROSSA dataset [19].

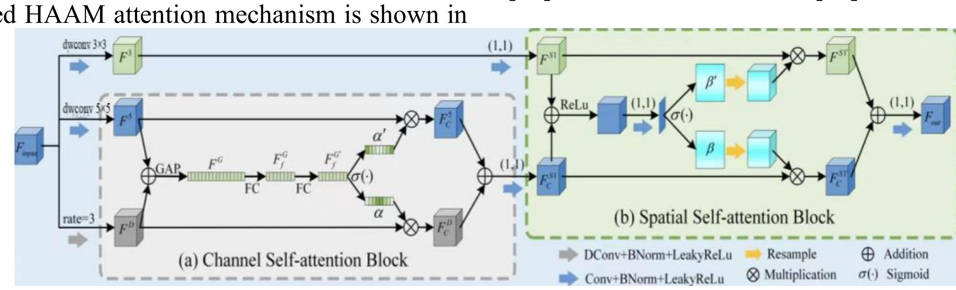



Fig. 4 Modified HAAM model

The OCTA-500 dataset contains two subsets: OCTA_6M and OCTA_3M. OCTA_6M consists of 300 subjects with images of size 400 × 400, of which 180 images are used for training, 100 images for testing, and 20 images for validation. OCTA_3M consists of 200 subjects with images of size 304 × 304, of which 140 images are used for training, 50 images for testing, and 10 images for validation.The ROSSA dataset contains 918 OCTA images and their corresponding vascular annotations, with a size of 320 × 320. In this paper, we divide ROSSA into training set (No. 1-No. 100 and No. 301-No. 918), validation set (No. 101-No. 200) and test set (No. 201-No. 300). To augment the data, we employed random rotation, added Gaussian noise, random sharpness adjustment, and horizontal, vertical, and diagonal flips for both datasets. Four data augmentation methods, each of which generates four new images from one original image, and during training and testing, we normalized the images in both datasets.

### B. Evaluation index

In order to evaluate the performance of this model, this paper uses accuracy (*Acc*) and Dice coefficient (*Dice*) as evaluation metrics. *Acc* represents the ratio of correctly classified pixels by the network to the total number of pixels in the image. The higher the accuracy, the better the segmentation performance of the network. The *Dice* coefficient is a function used to measure the similarity of sets, and is often used to measure the similarity between two samples. The formula definitions of the two evaluation indicators in this paper are as follows:

$$Acc=\frac{TP+TN}{TP+FN+FP+TN} \tag{1}$$

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

In this article, we use the *Dice* loss as the loss function of our model, and the definition of the *Dice* loss function is as follows:

$$DiceLoss = 1 - \frac{2\left|X \cap Y\right| + \varepsilon}{\left|X\right| + \left|Y\right| + 2\varepsilon} \tag{3}$$

Where $X$ is the predicted image mask and $Y$ is the artificially identified real image mask, $\varepsilon$ is a small positive number (such as 10-5), used to smooth the loss function to ensure its derivability in all cases.

### C. Segmentation result

We performed vascular segmentation experiments on several popular networks, including U-Net[21], UNet++[22], ResUNet[23], FARGO[11], and FRNet-base[15]. The segmentation results are presented in Table 1. As can be seen from the table, our proposed approach achieves higher *Dice* and *Acc* scores on all datasets. Although our model has slightly more parameters and a longer inference time compared to FRNet-base[34], this is due to the additional convolution operation we incorporated. Despite this, the accuracy achieved is higher than FRNet-base, and the training is stable.

We can see that the number of parameters for FRNet V2 is more than two orders of magnitude lower than for other models (e.g., 0.29 M for FRNet V2 versus 17.52 M for FARGO), and its inference speed is also faster than that of its competitors. This demonstrates that our proposed methods are more efficient, and we anticipate that they will be more suitable for industrial applications.

TABLE 1 COMPARISON BETWEEN DIFFERENT APPROACHES

| Method | Dice（↑） | Acc（↑） | Param（↓） | Time（↓） |
|---|---|---|---|---|
| | | **OCTA_6M** | | |
| U-Net | 85.03 | 95.21 | 14.32M | 20.2ms |
| U-Net++ | 85.67 | 95.73 | 15.96M | 25.7ms |
| ResUNet | 88.10 | 96.03 | 32.52M | 32.4ms |
| FARGO | 89.01 | 98.12 | 17.52M | 29.6ms |
| FRNet-base | 88.85 | 98.02 | **0.12M** | **15.3ms** |
| FRNet V2 | **89.14** | **98.28** | 0.29M | 26.4ms |
| | | **OCTA_3M** | | |
| U-Net | 88.35 | 95.45 | 14.32M | 17.4ms |
| U-Net++ | 88.64 | 95.98 | 15.96M | 21.2ms |
| ResUNet | 90.03 | 96.18 | 32.52M | 26.3ms |
| FRRGO | 91.21 | 98.12 | 17.52M | 24.5ms |
| FRNet-base | 91.15 | 98.84 | **0.12M** | **12.1ms** |
| FRNet V2 | **91.57** | **98.99** | 0.29M | 19.5ms |
| | | **ROSSA** | | |
| U-Net | 89.53 | 96.64 | 14.32M | 18.9ms |
| U-Net++ | 90.16 | 96.97 | 15.96M | 23.9ms |
| ResUNet | 91.32 | 97.81 | 32.52M | 28.7ms |
| FRRGO | 91.23 | 98.03 | 17.52M | 27.5ms |
| FRNet-base | 92.12 | 98.24 | **0.12M** | **13.8ms** |
| FRNet V2 | **92.58** | **98.42** | 0.29M | 24.5ms |

## IV. Ablation Experiment

In this section, we conducted ablation experiments as shown in Figure 3. FRNet V2 can be seen as a network composed of full-resolution convolutions. To verify the effectiveness of the components, we performed ablation studies to evaluate how each component affects the results, using the ROSSA dataset as an example.We first experimented with each component in the ConvNeXt V2 block,, and the results are summarized in Table 2. The first line indicates the use of the same modules as in FRNet-base. The second line indicates the replacement of the previous module with the ConvNeXt V2 block, and the reduction in the number of parameters is due to the use of deeply separable convolutions in ConvNeXt V2. The third row indicates the use of a 3x3 convolution to replace the 1x1 convolution in the ConvNeXt V2 block. As the number of parameters increases, accuracy also increases. The fourth line represents the application of recursive convolution, which achieves optimal precision. Recursive convolution does not increase the number of parameters, but it does increase the inference time.

TABLE 2 ABLATION EXPERIMENTS OF THE CONVNEXT V2 MODULE IN FRNET V2

| Component | Dice (↑) | ACC (↑) | Param (↓) | Time (↓) |
|---|---|---|---|---|
| Residual Block | 92.12 | 98.24 | **0.12M** | **13.8ms** |
| ConvNeXtV2 Block | 91.53 | 97.94 | 0.08M | 12.3ms |
| 1×1⇒3×3 | 92.18 | 98.32 | 0.13M | 14.5ms |
| +Recurrent | **92.37** | **98.38** | 0.13M | 21.6ms |

We also performed experiments on the HAAM attention module, and the results are shown in Table 3. The first row indicates no addition of the HAAM attention module, the second row indicates the addition of the HAAM module, and the third row indicates the replacement of ordinary convolution with a deeply separable convolution. We can see that after adding the HAAM module, the accuracy improves, but the number of parameters and inference time also increase. When the convolution is replaced with a deeply separable convolution, you can see that the accuracy is improved while reducing the number of parameters and inference time. The fourth row represents the combination with recursive convolution. As mentioned earlier, it achieves the best accuracy without increasing the number of parameters, but the inference time does increase.

TABLE 4 ABLATION EXPERIMENTS OF THE HAAM MODULE IN FRNET V2

| Component | Dice (↑) | ACC (↑) | Param (↓) | Time (↓) |
|---|---|---|---|---|
| Without HAAM | 92.37 | 98.38 | **0.13M** | **21.6ms** |
| With HAAM | 92.44 | 98.39 | 0.61M | 22.4ms |
| Conv⇒DSConv | 92.45 | 98.40 | 0.29M | 21.8ms |
| +Recurrent | **92.58** | **98.42** | 0.29M | 24.5ms |

## V. Conclusion

This paper presents an accurate and efficient neural network (FRNet V2), which is a model for OCTA retinal vessel segmentation. In this model, we fuse an optimized recursive ConvNeXt V2 module with a self-attention module (HAAM). This unique combination enables FRNet V2 to perform convolutional operations at full resolution, which not only significantly improves the processing efficiency of the model but also ensures the accuracy of segmentation.Experimental results show that this model is 49 times lighter than U-Net, 55 times lighter than U-Net++, and 110 times lighter than ResUNet in OCTA retinal vessel segmentation. Additionally, FRNet V2 is 1.1 times faster than U-Net++ and 1.3 times faster than ResUNet. The architectural design of FRNet V2, though simple, achieves high-performance computation with very low model complexity. This means that the model can accomplish tasks quickly and accurately in resource-limited environments and is well-suited for users with real-time performance requirements and application scenarios with stringent requirements for computational efficiency.In the future, we will continue to explore this work and extend it to the field of 3D vessel segmentation, such as magnetic resonance angiography and computed tomography angiography.

## References


[1] MOU L, ZHAO Y, CHEN L, et al. Cs-net: channel and spatial attention network for curvilinear structure segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2019: 721-730.

[2] ZHAO Y, ZHENG Y, LIU Y, et al. Automatic 2-d/3-d vessel enhancement in multiple modality images using a weighted symmetry filter[J]. *IEEE* transactions on medical imaging, 2017, 37(2):438-450.

[3] Xia Ying, Shi Mei, Zhang Ming. Early screening value of optical coherence tomography angiography for retinal microcirculation disorders in patients with diabetic retinopathy [J]. China Pharmaceutical Herald, 2022, 19 (28): 4.

[4] Ai Shibei, Zheng Zong, Cao Mingzhe, et al. Differences in biochemical indicators and optical coherence tomography angiography in patients with different degrees of diabetic retinopathy [J]. Acta Ophthalmologica Sinica, 2022, 37 (12): 8.

[5] SPAIDE R F, FUJIMOTO J G, WAHEED N K, et al. Optical coherence tomography angiography [J]. Progress in retinal and eye research, 2018, 64: 1- 55.

[6] DONNELLY R, EMSLIE-SMITH A M, GARDNER I D, et al. Vascular complications of diabetes [J]. Bmj, 2000, 320(7241): 1062-6

[7] YANNUZZI L A, NEGRãO S, TOMOHIRO I, et al. Retinal angiomatous proliferation in age–related macular degeneration [J]. Retina, 2012, 32: 416-34.

[8] RAO H L, PRADHAN Z S, WEINREB R N, et al. Regional comparisons of optical coherence tomography angiography vessel density in primary open-angle glaucoma[J]. American journal of ophthalmology, 2016, 171:75-83.

[9] LI M, CHEN Y, JI Z, et al. Image projection network: 3d to 2d image segmentation in octa images[J]. *IEEE* Transactions on Medical Imaging, 2020, 39(11):3343- 3354.
[10] MA Y, HAO H, XIE J, et al. Rose: a retinal oct-angiography vessel segmentation dataset and new model[J]. *IEEE* transactions on medical imaging, 2020, 40(3): 928-939.
[11] Linkai Peng, Li Lin, Pujin Cheng, Zhonghua Wang, and Xiaoying Tang. "FARGO - A Joint Framework for FAZ and RV Segmentation from OCTA Images.", Medical Image Computing and Computer-Assisted Intervention (2021): 42-51.
[12] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," ArXiv, vol. abs/2010.11929, 2020.
[13] Zhuang Liu, Hanzi Mao, Chaozheng Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie, "A convnet for the 2020s," 2022 *IEEE/CVF* Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11966–11976, 2022.
[14] Sanghyun Woo, Shoubhik Debnath, Ronghang Hu, Xinlei Chen, Zhuang Liu, In So Kweon, and Saining Xie. "ConvNeXt V2: Co-Designing and Scaling ConvNets With Masked Autoencoders", Computer Vision and Pattern Recognition abs/2301.00808 (2023): 16133-16142.
[15] Haojian Ning, Chengliang Wang, Xinrun Chen, and Shiying Li. "An Accurate and Efficient Neural Network for OCTA Vessel Segmentation and a New Dataset", *IEEE* International Conference on Acoustics, Speech, and Signal Processing abs/2309.09483 (2023): 1966-1970.
[16] Kaiming He, X. Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," 2016 *IEEE* Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.
[17] Md. Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M. Taha, and Vijayan K. Asari, "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation," ArXiv, vol. abs/1802.06955, 2018.
[18] Mingchao Li, Yuhan Zhang, Zexuan Ji, Keren Xie, Songtao Yuan, Qinghuai Liu, and Qiang Chen, "Ipn-v2 and octa-500: Methodology and dataset for retinal image segmentation," ArXiv, vol. abs/2012.07261, 2020.
[19] Haojian Ning, Chengliang Wang, Xinrun Chen, and Shiying Li. "An Accurate and Efficient Neural Network for OCTA Vessel Segmentation and a New Dataset", *IEEE* International Conference on Acoustics, Speech, and Signal Processing abs/2309.09483 (2023): 1966-1970.
[20] Gongping C, Lei L, Yu D, Jianxun Z, Moi H Y, et al. AAU-Net: An Adaptive Attention U-Net for Breast Lesions Segmentation in Ultrasound Images.[J], *IEEE* Transactions on Medical Imaging, 2023, 42(5): 1289-1300.
[21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," ArXiv, vol. abs/1505.04597, 2015.
[22] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang, "Unet++: A nested unet architecture for medical image segmentation," Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, S..., vol. 11045, pp. 311, 2018.
[23] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang, "Road extraction by deep residual u-net," *IEEE* Geoscience and Remote Sensing Letters, vol. 15, pp. 749753, 2018.