Proceedings of the 2nd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/76/20240578

Unet retinal vessel image segmentation algorithm based on coordinate attention mechanism and Min-ASPP module optimisation

Yiheng Ye[1,*], Qichao Zhao[2], Tiancheng Bao[3]

[1]School of Information Science, Guangdong University of Finance &Economics, Guangzhou, Guangdong, 510145, China
[2]College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[3]Electrical Engineering, and Computer, and Systems Engineering, Rensselaer Polytechnic Institute, New York, Troy, 12180, USA

[*]Corresponding author email: 2335901957@student.gdufe.edu.cn

Abstract. In this paper, the Unet image segmentation algorithm is optimised and improved by introducing the coordinate attention mechanism and integrating the Min-ASPP module, and compared with the traditional Unet model, targeting the segmentation of retinal blood vessel images, which is a key problem in the field of medical image processing. The experimental results show that the optimised Unet algorithm combining the coordinate attention mechanism and Min-ASPP module has a smoother performance in the loss curve, faster convergence speed, and lower final converged loss value, which has an obvious advantage compared with the traditional Unet algorithm. Meanwhile, the optimised Unet algorithm shows higher accuracy and stability in the prediction results, providing ophthalmologists with more reliable information about the retinal vascular structure. Therefore, the research results of this paper show that the Unet image segmentation algorithm optimised by combining the coordinate attention mechanism and Min-ASPP module has better application prospects and effects in the retinal blood vessel image segmentation task, which provides a useful exploration and practice for improving the level of medical image processing technology.

Keywords: Attention Mechanism, Min-ASPP, Image Segmentation.

1. Introduction

Retinal vascular image segmentation is an important research direction in the field of medical image processing, which aims to accurately extract the vascular structure from retinal imaging data to provide ophthalmologists with auxiliary information for the diagnosis and treatment of retinal diseases [1]. With the development of digital medical imaging technology, retinal vascular image segmentation plays an increasingly important role in ophthalmology clinical diagnosis, disease monitoring, and treatment effect assessment. Traditional image segmentation methods based on rule and feature engineering are often difficult to achieve satisfactory segmentation results in complex retinal vascular structures, so deep

DOI: 10.54254/2755-2721/76/20240578

learning techniques have been introduced into the field of retinal vascular image segmentation in recent years and have made significant progress [2,3].

Deep learning image segmentation algorithms play an important role in retinal blood vessel image segmentation. Compared with traditional methods, deep learning algorithms have more powerful feature learning capability and adaptability, and are able to automatically learn abstract expressions representing different classes of features from a large amount of labelled data, thus achieving more accurate and robust image segmentation results. Convolutional neural network (CNN) [4], as one of the most commonly used architectures in deep learning, is widely used in retinal blood vessel image segmentation tasks. By constructing end-to-end neural network models, CNNs can effectively capture visual features at different scales and levels to achieve accurate segmentation of complex retinal vascular structures.

In recent years, many deep learning-based methods for retinal blood vessel image segmentation have emerged. Among them, some methods adopt structures such as full convolutional network (FCN) [5] or U-Net to achieve fine segmentation of fine vascular structures while preserving spatial information; others introduce modules such as attention mechanism and residual connectivity to enhance the network's ability to perceive local features and edge information; and still others combine multimodal data or utilise adversarial generative networks (GANs) [6] for weakly supervised learning, which improves the generalisation ability of the model on incompletely labelled data.

Overall, deep learning image segmentation algorithms provide strong support for increasing automation, reducing physicians' workload, improving diagnostic accuracy, and accelerating clinical decision-making in retinal vessel image segmentation. In this paper, the Unet image segmentation algorithm is improved by combining the coordinate attention mechanism and integrating the Min-ASPP module, which compares the effect of the traditional Unet image segmentation model, and provides a certain research basis for the subsequent research. In the future, as the technology continues to progress and the quality of dataset annotation improves, the deep learning-based retinal blood vessel image segmentation method will be further improved and play a greater value in the field of ophthalmic medicine.

2. Data set sources and data analysis

The dataset used in this paper is an open source image dataset linked to (https://www.kaggle.com/datasets/abdallahwagih/retina-blood-vessel), which contains a comprehensive collection of retinal fundus images that are carefully annotated for vessel segmentation. Accurate segmentation of blood vessels is a critical task in ophthalmology as it helps in early detection and management of various retinal pathologies such as diabetic retinopathy and macular degeneration. Part of the image is shown in Fig. 1 and its corresponding image segmentationmask is shown in Fig. 2.

Figure 1. Original image.

DOI: 10.54254/2755-2721/76/20240578

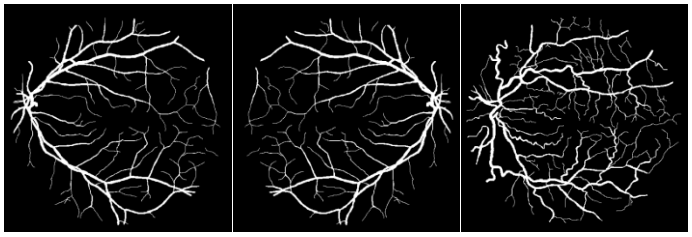

Figure 2. Image mask.

The dataset has the following advantages, the images in the dataset have different sizes, mimicking the diversity of retinal images in the real world. Secondly the dataset contains a range of retinal conditions including different vessel widths, branching patterns and the presence of anomalies making it suitable for assessing the robustness of segmentation models.

3. Method

3.1. Unet

Unet is a classical deep learning image segmentation model whose structural design is inspired by the encoder-decoder architecture [7].The compression path of Unet consists of alternating convolutional and pooling layers, which are used to progressively extract high-level feature representations of the input image. Through multiple convolution operations, the network can capture feature information at different scales and abstraction levels, while gradually reducing the size of the feature map through pooling operations to achieve effective compression and extraction of the input image information.The expansion path of Unet consists of a convolutional layer and an up-sampling (inverse convolution or transposed convolution) layer, which is used to progressively reduce and restore the high-level feature representations extracted in the compression path to the original input image size. In the expansion path, the feature maps are gradually enlarged in size by the up-sampling operation and spliced with the feature maps of the same size level in the corresponding compression path, thus fusing the information of different abstraction levels to help refine the segmentation results [8]. The model structure of Unet is shown in Fig. 3.

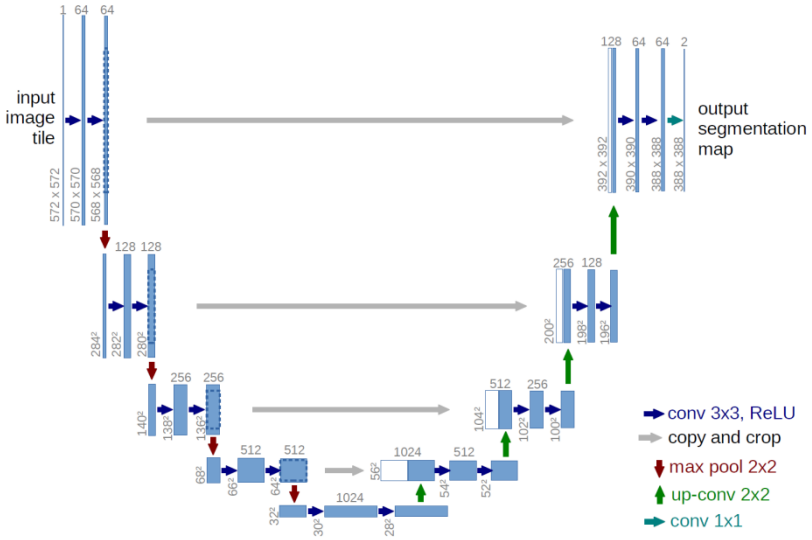

Figure 3. The model structure of Unet.

The Unet model is designed with a short connection structure, which allows the network to take advantage of the exchange of information between different layers, helping to alleviate the problem of

DOI: 10.54254/2755-2721/76/20240578

gradient vanishing, accelerate training convergence, and better preserve detailed information. In addition, Unet often employs appropriate loss functions (e.g., Dice Loss, Cross Entropy Loss, etc.) to guide network learning to generate accurate segmentation results.

3.2. Coordinate Attention Mechanism

Coordinate attention mechanism is an attention mechanism used to enhance the perception of spatial information by neural networks. In the traditional attention mechanism, the model usually learns the importance weights of different locations based on the global information of the input features, but this method may not be able to fully take into account the relative relationship between different locations and the spatial structure. The coordinate attention mechanism, on the other hand, enables the model to better understand the relative positional relationships between different locations by introducing location coordinate information, which improves the model's performance when dealing with spatial perception tasks [9].

Specifically, the coordinate attention mechanism makes the model simultaneously consider the relative position of each location in the whole space during the learning process by splicing or computing the position coordinate information with the input features. In this way, the model can better capture the spatial dependencies between different locations, which helps to improve the performance when processing data with obvious spatial structure features such as images and videos.

3.3. Min-ASPP

Min-ASPP (Minimum Atrous Spatial Pyramid Pooling) module is an attentional mechanism for semantic segmentation tasks, aiming to improve the perception of neural networks for different scales of information and the performance of semantic segmentation.The Min-ASPP module combines the ideas of null convolution and spatial pyramid pooling through multi-scale feature extraction and minimisation operations to effectively capture semantic information at different scales in an image [10].

Specifically, the Min-ASPP module first adopts hollow convolution with different dilation rates for feature extraction, which can increase the receptive field while maintaining the feature map size, thus obtaining richer contextual information. Then, after obtaining multiple feature representations at different scales, the Min-ASPP module introduces a minimisation operation, i.e., using global average pooling or global minimum pooling, etc., to aggregate these multi-scale features and retain the most salient information.

Through the minimisation operation, the Min-ASPP module can filter the most important and representative features at each scale and integrate them into a unified feature representation. This design enables the network to better understand the image content at different scales and improves the ability to grasp the semantic information. At the same time, the Min-ASPP module also helps to reduce the number of parameters and the amount of computation, which improves the efficiency of the model and the speed of inference.

4. Result

4.1. Experimental setup

The CPU used in this paper is Intel Core i9, GPU is NQuadro RTX 8000, RAM is 64GB, storage is using SSD with 1TB SSD, operating system is Windows 10, based on PyTorch deep learning framework, python version is 3.10.

Before the experiment, the images are first preprocessed and all images are processed into 512×512 size, then the data are randomly divided into training and test sets according to the ratio of 7:3, 70% of the data are used for training and 30% of the data are used for testing.The Epoch is set to 50, the maximal learning rate of the model is set to 1e-4, and the batch size is set to 32 The model is adjusted adaptively according to the The model adaptively adjusts the learning rate according to the current batch size. The output loss value and the loss curve are set, five images are used for prediction, and the segmentation results of the five images are automatically output for comparison and analysis.

DOI: 10.54254/2755-2721/76/20240578

Experimental results

The Unet and Unet image segmentation algorithms optimised by combining the coordinate attention mechanism and incorporating the Min-ASPP module are trained respectively, and the changes in the loss curves for 50 epochs are shown in Fig. 4.

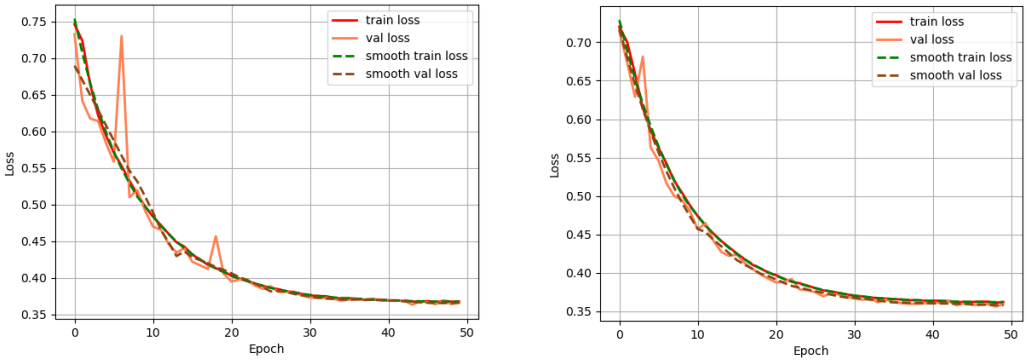

Figure 4. The changes in the loss curves.

As can be seen from the figure, the loss curve of the Unet image segmentation algorithm optimised by combining the coordinate attention mechanism and incorporating the Min-ASPP module is relatively smooth, and is faster than Unet in terms of convergence speed, while the final converged loss value is smaller than that of Unet.

The outputs of the Unet image algorithm and the results predicted by the Unet image segmentation algorithm optimised by combining the coordinate attention mechanism and incorporating the Min-ASPP module with the Unet image algorithm are shown in Fig. 5.

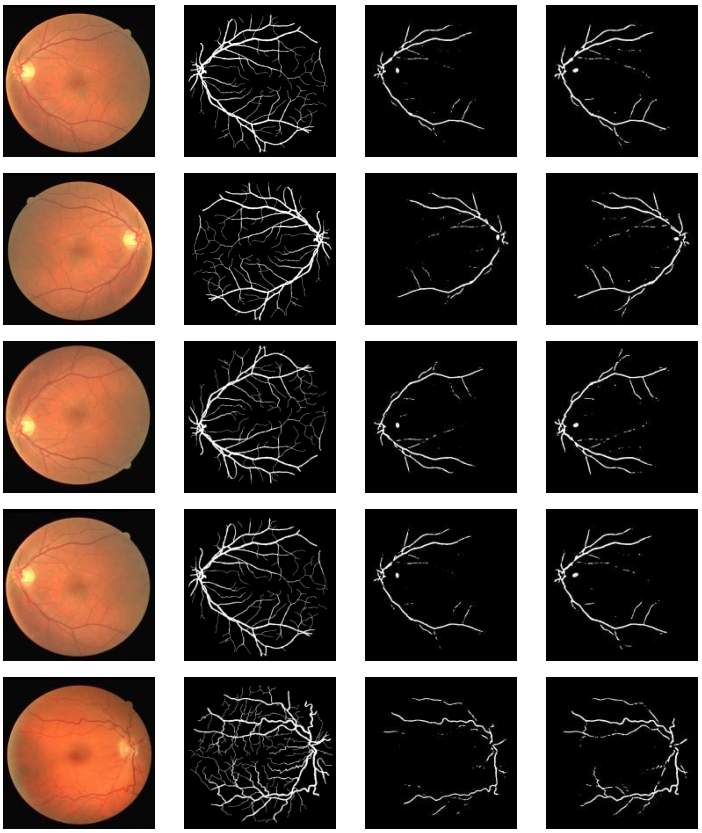

Figure 5. Projected results.

DOI: 10.54254/2755-2721/76/20240578

As can be seen from Figure 5, the first column is the original image, the second column is the result of manually correct segmented image, the third column is the result predicted by Unet image, and the fourth column is the result predicted by Unet image segmentation algorithm optimised by combining coordinate attention mechanism and incorporating Min-ASPP module. It can be clearly seen that the Unet image segmentation algorithm optimised by combining coordinate attention mechanism and incorporating Min-ASPP module is better than Unet in predicting the results.

5. Conclusion

In this paper, the Unet image segmentation algorithm is improved by introducing the coordinate attention mechanism and fusing the Min-ASPP module, and compared with the traditional Unet model. The experimental results show that the optimised Unet algorithm performs smoother on the loss curve, converges faster, and the final converged loss value is smaller, which has obvious advantages compared with the traditional Unet algorithm. In addition, the optimised Unet algorithm combined with the coordinate attention mechanism and Min-ASPP module performs better in predicting the results, and has higher accuracy and robustness compared to the traditional Unet algorithm. Therefore, it is concluded that the optimised Unet image segmentation algorithm significantly outperforms the traditional Unet model in terms of effectiveness and provides an effective and better performing solution for image segmentation tasks.

References

[1] Ma, Danqing, Bo, Dang, Shaojie, Li, Hengyi, Zang, Xinqi, Dong. "Implementation of computer vision technology based on artificial intelligence for medical image analysis". International Journal of Computer Science and Information Technology 1. 1(2023): 69–76.

[2] Ao Xiang, , Zongqing Qi, Han Wang, Qin Yang, Danqing Ma. "A Multimodal Fusion Network For Student Emotion Recognition Based on Transformer and Tensor Product." (2024).

[3] Patton, Niall, et al. "Retinal image analysis: concepts, applications and potential." Progress in retinal and eye research 25.1 (2006): 99-127.

[4] Maninis, Kevis-Kokitsi, et al. "Deep retinal image understanding." Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19. Springer International Publishing, 2016.

[5] Badar, Maryam, Muhammad Haris, and Anam Fatima. "Application of deep learning for retinal image analysis: A review." Computer Science Review 35 (2020): 100203.

[6] Chanwimaluang, Thitiporn, Guoliang Fan, and Stephen R. Fransen. "Hybrid retinal image registration." IEEE transactions on information technology in biomedicine 10.1 (2006): 129-142.

[7] Lee, Samuel C., and Yiming Wang. "Automatic retinal image quality assessment and enhancement." Medical imaging 1999: image processing. Vol. 3661. SPIE, 1999.

[8] Tobin, Kenneth W., et al. "Detection of anatomic structures in human retinal imagery." IEEE transactions on medical imaging 26.12 (2007): 1729-1739.

[9] Goh, James Kang Hao, et al. "Retinal imaging techniques for diabetic retinopathy screening." Journal of diabetes science and technology 10.2 (2016): 282-294.

[10] Sánchez, Clara I., et al. "Retinal image analysis to detect and quantify lesions associated with diabetic retinopathy." The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol. 1. IEEE, 2004.