Retinal vascular segmentation based on depth-separable convolution and attention mechanisms

Xiaopeng Liu[1], Dongxu Gao[2],Congyi Zhang[3] ,Hongwei Gao[1] ,and Zhaojie Ju[4(✉)]

[1] School of Automation and Electrical Engineering, Shenyang Ligong University,
Shenyang 110159, China
ghw1978@sohu.com

[2] School of Computing, University of Portsmouth, Portsmouth PO13HE, UK
dongxu.gao@port.ac.uk

[3] School of Computing, University of Portsmouth, Portsmouth PO13HE, UK
Congyi.zhang@port.ac.uk

[4] School of Computing, University of Portsmouth, Portsmouth PO13HE, UK
Zhaojie.Ju@port.ac.uk

summary: Retinal vascular segmentation is an important research direction in the field of medical image processing, its main purpose is to automatically segment the vascular area from the fundus image, and provide doctors with more accurate diagnosis results and treatment plans. In recent years, with the continuous development of deep learning technology, retinal vascular segmentation algorithm based on deep learning has gradually become a research hotspot. In this paper, the retinal vascular segmentation algorithm based on deep learning is mainly improved, and the retinal vascular segmentation algorithm based on IPN-V2 is improved, in an attempt to make new explorations.

The retinal vascular segmentation algorithm based on IPN-V2 provides global information, but requires a large amount of image data and label information, the image size is different, and most importantly, the accuracy of the model for the segmentation of the original image is not enough.Therefore, this paper improves the retinal vascular segmentation algorithm based on IPN-V2, introduces the attention mechanism, and constructs a retinal vascular segmentation model based on ASR-IPN-V2, which enables the model to extract more image details from the original image through the depth-separable convolution and convolutional block attention mechanisms.

Experiments show that the retinal vascular segmentation model based on ASR-IPN-V2 greatly improves the efficiency of retinal vascular segmentation.

keyword: IPN-V2 model; Segmentation of retinal blood vessels; Attention mechanism

1 Introduction

Retinal vascular segmentation refers to the segmentation of retinal blood vessels in fundus images, which plays an important role in the early diagnosis, treatment and evaluation of ophthalmic diseases.Optical coherence tomography angiography (OCTA) is a non-invasive test that enables high-resolution imaging of ocular blood vessels.The combination of OCTA technology and retinal vascular segmentation technology can realize accurate and non-invasive imaging and segmentation of ocular blood vessels, further improving the effect of diagnosis and treatment.In the traditional retinal vascular segmentation method, image processing and machine learning technology are mainly used, but due to the noise in the image, the complex morphology of blood vessels and the differences in different pathological conditions, the segmentation effect of these methods is often unsatisfactory.

In recent years, with the rapid development of deep learning technology, retinal vascular segmentation methods based on deep learning have attracted more and more attention. Deep learning is a machine learning technology that can automatically learn feature representation, which has strong expression ability and generalization ability, and can effectively solve some problems in traditional methods. Among them, the Convolutional Neural Network (CNN) is a deep learning model with powerful feature extraction and classification capabilities. CNN-based retinal vascular segmentation methods mainly include FCN network, U-Net network, Seg Net network, etc., which can automatically learn feature representation and improve the segmentation effect.

Deep learning is a hot technology in the field of artificial intelligence, with a wide range of application prospects and trends.The following are the application prospects and trends of deep learning:

(1) Computer vision: Deep learning is widely used in computer vision, including image recognition, object detection, face recognition, image segmentation and other fields [1].Object detection and image classification have always been the focus of deep learning applications in the field of computer vision.Through the training and optimization of deep learning algorithms, the recognition and classification of objects can be achieved, which supports applications such as image retrieval and object tracking. Deep learning also plays an important role in video analytics.Through the training and optimization of deep learning algorithms, applications such as object tracking and action recognition in video can be realized, so as to provide support for video surveillance, video analysis and other fields [2]. Deep learning also has a wide range of applications in 3D reconstruction.Through the training and optimization of deep learning algorithms, the reconstruction and recognition of three-dimensional space can be realized, so as to provide support for architecture, engineering, geology and other fields.With the continuous optimization and performance improvement of deep learning algorithms, the application field of computer vision will be more extensive.

(2) Natural language processing: Deep learning also has a wide range of applications in natural language processing, including speech recognition, machine translation, sentiment analysis and other fields [3].Deep learning also has a wide range of applications in text classification and information extraction. Through the training and optimization of deep learning algorithms, the classification of text and information extraction can be realized, so as to provide support for text mining, intelligence analysis and other fields [4].Deep learning is also becoming more widely used in automated question answering and knowledge graphs.Through the training and optimization of deep learning algorithms, the understanding and answering of natural language problems can be realized, which provides support for the question answering system and knowledge graph construction.Deep learning is also becoming more widely used in emotion recognition and text generation. Through the training and optimization of deep learning algorithms, the recognition and expression of emotions can be realized, which can support applications such as sentiment analysis and text generation.With the continuous improvement and optimization of deep learning algorithms, the application of natural language processing will become more popular and mature [5].

(3) Speech recognition: Speech recognition is a technology that converts human speech into computer-readable text, and it has a wide range of applications in artificial intelligence, smart home, intelligent customer service and other fields [6].The development of deep learning technology has greatly improved the accuracy of speech recognition, making the application of speech recognition in real life more popular [7].Deep learning technology can provide powerful support for the training of speech recognition models.Through the training and optimization of deep learning algorithms, the feature extraction of speech signals and the training of speech recognition models can be realized, so as to improve the accuracy and stability of speech recognition.Deep learning technology also has a wide range of applications in multilingual speech recognition.Through the training and optimization of deep learning algorithms, the recognition and conversion of multilingual speech signals can be realized, so as to provide support for multinational enterprises and international conferences.

At present, deep learning has shown excellent results in the fields of speech recognition, natural language processing, and computer vision.There are many kinds of deep learning algorithm models, including automatic coding machines, restricted Boltzmann machines, deep neural networks, convolutional neural networks, recurrent neural networks, and neural networks with multi-network model fusion, among which there are both traditional algorithms and the latest algorithm models [8].When applying deep learning technology, it is necessary to reasonably select algorithm models according to the characteristics of different fields, and cannot abandon the use because the models are old.Specific analysis of specific problems, reasonable selection of deep learning models is crucial.Although deep learning technology is still evolving, it will still be full of opportunities and challenges in the future.

The retinal vascular segmentation method based on deep learning has the advantages of high precision, high efficiency and automation, and has been widely used in the diagnosis and treatment of many ophthalmic diseases.However, there is still some room for development in OCTA reconstruction using deep learning methods.For example, methods and bases based on convolutional neural networks require a large amount of labeled data to train the model, and have high requirements for the quality and quantity of data. The instability of GAN model training and the possible distortion of the generated image may occur.With the continuous development and improvement of deep learning technology, retinal vascular segmentation method based on deep learning will have wider application prospects.

The IPN-V2 model successfully overcomes the "checkerboard effect" by introducing attention mechanisms and residual blocks, improving the accuracy and robustness of segmentation.Since the IPN-V2 model is a graph-to-graph method, it needs to be trained with complete image information, which requires a large amount of labeled data to train the model.In practice, obtaining sufficient labeled data is a very time-consuming and difficult task, especially for some rare cases or diseases that require a high degree of expertise. This study has conducted an in-depth study on retinal vascular segmentation based on deep learning, and proposed a new method to improve the IPN-V2 model for retinal vascular segmentation, which has important theoretical and practical significance.

In theory, although the IPN-V2 model performs well in the retinal vascular segmentation algorithm, it still has some limitations, such as insufficient capture of detailed information and insufficient processing of edge information.Therefore, this paper improves the IPN-V2 model to improve the performance and stability of the IPN-V2 model.This improvement provides a theoretical basis for the further development of IPN-V2 model, and enriches the research results of IPN-V2 model and retinal vascular segmentation algorithm.

In practice, the study of retinal vascular segmentation based on deep learning has become a hot area of current research.Through the improvement and application of IPN-V2 model in practice, this paper provides a practical reference for the retinal vascular segmentation method based on deep learning.At the same time, the improved IPN-V2 model proposed in this paper has achieved good results in practice, which verifies the effectiveness and practicability of the model.Therefore, the practical research in this paper provides an important reference and reference value for the application and promotion of retinal vascular segmentation algorithm based on deep learning.

The main contribution of this paper is the following improvements to the IPN-V2 model:

1. The amount of parameters in the network and the resources consumed during calculation are reduced with deep separable convolution.

2. The introduction of attention mechanism can effectively improve the accuracy of network segmentation.

The rest of this article is organized as follows. In the second part, we reviewed the work related to retinal vascular segmentation from the segmentation method. The third part introduces our methodology, and then the results and analysis of the experiment are presented in the fourth section. The final part is the summary.

2 Related work

2.1 Retinal vascular segmentation method based on machine learning

For retinal vascular segmentation tasks, the usual segmentation methods include Bayesian method, K-nearest neighbor method, support vector machine, random forest, Ada Boost, etc., all of which are typical supervised machine learning algorithms.For example, Wu Kui et al. (2016) consider that when the 2DGabor wavelet algorithm is used alone, the vascular morphology and structural information will be ignored, and the 2DGabor wavelet transform is no longer used alone, but a combined line detection operator is introduced into the algorithm, which first uses wavelet changes and detection operators to process the fundus image, and after processing, a six-dimensional pixel feature vector is obtained, and finally a Bayesian Gaussian mixture model is adopted to classify all pixels in the fundus image. The final segmentation result is superior to the 2DGabor wavelet algorithm alone [9].In general, retinal vascular segmentation methods based on traditional supervised machine learning are based on the premise of feature extraction, and the quality of the extracted features determines the performance of subsequent classifier classification.However, determining which features to extract requires a certain amount of accumulated experience, which leads to segmentation results that are largely influenced by personal experience.

The principle of unsupervised machine learning is to self-summarize and summarize the knowledge contained in unlabeled data, and use this knowledge to categorize new things.Commonly used unsupervised machine learning methods include K-means clustering algorithm, EM algorithm (GMM-expectation maximization), FCM clustering algorithm, etc.For example, Roychowdhury et al. (2014) implemented a three-step retinal vascular segmentation algorithm: first acquire a binary image; For the extracted binary images, the EM algorithm is selected as the classifier to determine the category of the pixels inside. Finally, by flattening the main part of the blood vessel and the divided blood vessel pixels, the algorithm can reduce the dependence on the training data and achieve consistent segmentation accuracy on healthy pictures and pathological pictures [10].Jia Hong et al. (2020) consider that the previous retinal vascular segmentation algorithm based on spatial FCM clustering only focuses on the clustering characteristics of pixels in feature space, but ignores the structural coherence information of the domain in pixel space, so the algorithm combines the local linear structure constraints of blood vessels in the pixel spatial domain to solve the above problems, compared with the feature-space FCM clustering algorithm, the retinal vascular segmentation results constructed in this way have better continuity and higher sensitivity to microscopic blood vessels [11].In summary, the unsupervised machine learning method achieves the goal of training classification algorithms using no labeled data or less labeled data, and solves the problem of insufficient labeled data.However, since the data of unsupervised learning is not labeled, the evaluation of the effectiveness of such algorithms is still a problem to be solved.

2.2 Retinal vascular segmentation method based on deep learning

With the continuous improvement of deep learning technology, its application in computer vision is becoming more and more extensive.In the field of medical image processing, new deep learning models and auxiliary modules have been introduced one after another, continuously improving and improving the continuity and accuracy of retinal vascular segmentation.

The U-NET model was originally proposed to be used to segment the cells in the electron microscope image, which is composed of two parts: encoder and decoder, the shallow semantic information of the image is obtained in the coding stage, the high-level semantic information of the image is extracted in the decoding stage, and the image size is restored, and the proposed cascade structure can realize the fusion of the shallow feature map and the deeper feature map obtained in the coding stage, so as to obtain more comprehensive context information.Lu Huihui (2018) analyzed the network structure of U-net, built a U-net network model, and realized fundus image vascular segmentation based on convolutional neural network.Gao Dongxu (2019) proposed a U-net-based fringe noise removal framework [12], which is able to remove fringe noise in OCTA images, leaving a clean image.Dali Chen (2020) proposes a semi-supervised learning method for vascular segmentation with limited labeled data.In this method, they use an improved U-Net deep learning network to segment the vascular tree. On this basis, the authors implement the training dataset update strategy based on U-Net network.In order to analyze the segmentation performance of the semi-supervised learning method proposed in this paper, a large number of experiments are carried out.Experimental results show that the proposed method can avoid the problem of insufficient manual labeling and achieve satisfactory performance [13].Sun Ying (2022) proposes a U-shaped network that fuses rough neurons and channel attention mechanisms. The network first introduces the concept of upper and lower approximation in rough set theory to design rough neurons. Then, based on rough neurons, the rough channel attention module is constructed, which adopts global maximum pooling and global average pooling to construct approximate neurons in the U-Net jump connection, and performs weighted summation between neurons to reasonably roughen the established channel dependency, which not only contains global information but also has local characteristics, which can effectively realize the accurate recalibration of extracted retinal vascular features. Then, residual connections are added to transfer features directly from the lower layer to the higher layer, which helps to solve the problem of network performance degradation and effectively extract richer retinal vascular features [14].

In summary, compared with traditional supervised machine learning methods, the deep learning-based method does not need to manually extract features, but the model independently learns the information contained in the data, avoiding the influence of human subjective factors. Compared with unsupervised machine learning methods, although deep learning requires more labeled data, it is more accurate than unsupervised machine learning methods.Although the emergence of deep learning technology has greatly helped to improve the segmentation of retinal blood vessels, the segmentation of tiny blood vessels is still a major challenge, and the continuity of segmented blood vessels is also an urgent problem to be solved. At the same time, there are relatively few studies on improving the sensitivity of vascular segmentation. All in all, there are still many areas for improvement in the task of automatic division of retinal blood vessels by deep learning methods.

2.3 Dataset

The retinal vascular image public dataset is a collection of public data integrated by some medical institutions or scientific research organizations after collecting and processing retinal vascular images [15]. These datasets consist of two parts, one of which is a color image of retinal blood vessels, and the other part is a corresponding image of retinal blood vessels manually labeled by experts [16]. In this paper, the OCTA-500 dataset is mainly used to train and test the proposed network model.

3 Retinal vascular segmentation based on attention mechanism

Retinal vascular segmentation is difficult to accurately segment because the ends of the vessels are small and easily confused with the background. As a result, the blood vessels at the image border may be misdivided into backgrounds or blood vessels may be broken. The image projection network (IPN-V2) enables micro-resolution to visualize the three-dimensional structure of retinal blood vessels, which can realize 3D-2D RV segmentation and FAZ segmentation, and solve the problem of weak recognition ability and "checkerboard effect" in the horizontal direction. In order to further improve the segmentation accuracy of retinal blood vessels, this paper proposes a retinal vascular segmentation algorithm based on improved IPN-V2, which optimizes the IPN-V2 network structure and improves the block processing effect of raw data images.

3.1 IPN-V2-based retinal vascular segmentation

Retinal vascular density is used to diagnose the health of the retinal vascular system, which provides a basis for clinical diagnosis of retinal diseases and significantly improves the decision-making efficiency of doctors. Both OCT and OCTA can provide 3D data, but most retinal metrics, such as blood vessel density and FAZ area, are quantified on projection maps rather than in 3D space. Moreover, the failure of layer segmentation makes the index difficult to quantify, which has become a bottleneck in the field of retinal disease analysis.To this end, Ming chao Li (2020) proposes an image projection network (IPN), a novel end-to-end architecture that enables 3D to 2D image segmentation in optical coherence tomography angiography (OCTA) images. IPN uses the original 3D volume as input instead of a projection map, avoiding the use of retinal layer segmentation, thus overcoming the negative effects of false layer segmentation. In addition, using volume information can help the network better distinguish similar areas in the projection map, such as non-perfusion areas and FAZ regions. However, IPN still has some limitations. For example, the lack of downsampling in the horizontal direction makes the IPN lack a wide range of receptive fields in the horizontal plane, which leads to weak recognition in the horizontal direction. In addition, because 3D networks are computationally intensive, raw data needs to be cut into small pieces when computing resources are limited. In the stitching process of the segmentation results, the segmentation results have a "checkerboard effect".

To this end, Mingchao Li (2020) further proposed image projection network-V2 (IPN-V2) [17]. IPN-V2 introduces planar perceptrons to enhance perceptron capabilities in the horizontal direction. At the same time, in order to overcome the "checkerboard effect" during splicing, a global retraining process, called IPN-V2+, is added to IPN-V2, which stitches the 2D feature map output by the network and then trains a global network. This process not only overcomes the "checkerboard effect", but also provides global information that compensates for the limitations of patch training. The IPN-V2 network architecture is shown in Figure 1.

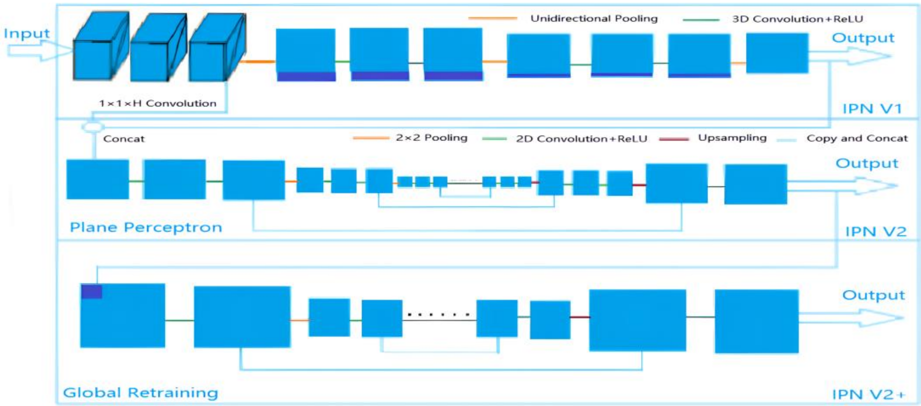


Figure 1 IPN-V2 network architecture

IPN-V2 expands the IPN and adds a flat fluoroscope, which strengthens the perspective capability in the lateral direction. At the same time, a global retraining process was introduced, overcoming the "checkerboard effect". Experimental results show that IPN-V2 has better performance than IPN and other deep learning methods in segmentation tasks. However, although IPN-V2 provides global information and contains complete image information every time training, this graph-to-graph method requires a large amount of image data and label information, and the dataset of medical images, especially the dataset of fundus images, is often relatively small, and at the same time, fundus images often contain a large number of vascular modalities, if the complete image is used every time training, it is not conducive to the convergence of the network, but also occupies a lot of memory space.For the image, the original image data is chunked, and the problem is to select different tiling schemes considering the specific characteristics of the vascular map. The segmentation method based on image blocks can not only enhance the data by chunking, increase the data set, but also reduce the memory occupation by reducing the number of vascular modalities. Therefore, in the task of retinal vascular segmentation, the segmentation method based on image blocks is adopted to improve the retinal vascular segmentation algorithm based on IPN-V2.

3.2 Attention mechanism

The Attention Mechanism has been widely used in various research fields of deep learning. The attention mechanism allows neural networks to reallocate computing resources and focus on key areas to highlight important local information. The attention mechanism often used in the field of computer vision mainly includes channel attention mechanism and spatial attention mechanism.

Channel attention mechanism

Convolutional neural networks can rely on convolution operations and use local receptive field ideas to fuse spatial domain information with channel domain information for feature extraction [18]. The Channel Attention (CA) mechanism, as the name implies, strengthens the network model from the channel dimension to improve its performance.In 2017, SENet[19], which used the channel attention mechanism, won first place in the ILSVRC 2017 classification task. For convolutional neural networks, the most important core computation is the convolution operation, through which new features are learned from the input feature map. Therefore, for convolutional neural networks, the process of learning features is to fuse the features of local regions, which can be divided into spatial features and channel features.

The innovation of SENet is to focus on the relationship between different channels, hoping to improve the branching structure of the model, so that the network model can learn the importance of different channels during the training process.SENet presents a channel attention module SE, which hopes to enhance the performance of the network model from the channel dimension. In order to obtain the importance of the feature map, the SE module uses the Squeeze operation to process the original multi-channel feature map, compress it into a one-dimensional vector according to each channel, and also learn the weight features of each channel, and then strengthen the original feature map with the learned feature vector.The SE structure is essentially similar to the Inception structure, except that the Inception structure adopted no longer focuses on the integration of features at different scales, but more on the importance between different channels. Therefore, the SE module mainly considers the dependence of different channels of neural network on feature selection in feature selection, and improves the segmentation accuracy through the strengthening of channel features.

Spatial attention mechanisms

The Spatial Attention mechanism focuses attention on an important part of the spatial dimension, so that the network model can simultaneously suppress the expression of other location information, so as to achieve the purpose of improving the segmentation results. Due to the complex features of retinal vascular images and the existence of noise problems, false positive predictions are difficult to reduce. Different from channel attention, the spatial attention mechanism can enhance the local feature extraction ability of the network model to a certain extent.

The CBAM module[20] is proposed by Woo et al., similar to the channel attention SE module proposed by Hu et al., and is a simple channel spatial attention module suitable for convolutional neural networks. Although the SE module is effective in enhancing model performance, the location information in the feature map is generally ignored, which is critical for generating spatially selective feature maps. CBAM extracts and combines important information in space and channel, and proposes that given intermediate feature mapping, two independent dimensions of attention mapping are deduced along the channel and space, and then the attention map is multiplied by the input feature map to achieve adaptive feature refinement. CBAM module is also suitable for segmentation tasks, the module can adaptively refine the mapping of feature maps, pay attention to important feature information and suppress unimportant feature information, improve segmentation performance, its structure is shown in Figure 2:

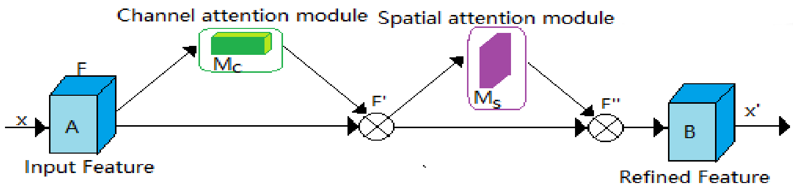


Figure 2 CBAM block structure diagram

The CBAM module is a lightweight general-purpose module that combines two attention mechanisms and can be added behind the convolutional layer of any network [21]. The specific calculation process of this module is shown in the following formula:

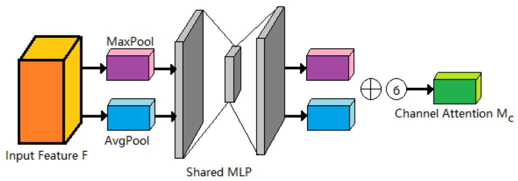
$$F' = M_c(F) \otimes F \tag{1}$$

$$F'' = M_s\left(F'\right) \otimes F' \tag{2}$$

$$\mathbf{M}_c(\mathbf{F}) = \sigma(\mathrm{MLP}(\mathrm{AvgPool}(\mathbf{F})) + \mathrm{MLP}(\mathrm{MaxPool}(\mathrm{F}))) \tag{3}$$

$$\mathbf{M}_s(\mathrm{F}) = \sigma\left(f^{7\times 7}([\mathrm{AvgPool}(\mathrm{F}); \mathrm{MaxPool}(\mathrm{F})])\right) \tag{4}$$

where F represents the input of feature Figure $(C\times H\times W)$, M_c is a channel attention map of one-dimensional $(C\times 1\times 1)$, M_s is a spatial attention map of 2D $(1\times H\times W)$, $\otimes$ indicates multiplication operation, F' is the intermediate output $(C\times H\times W)$, F'' is the final output $(C\times H\times W)$, MLP is a multilayer perceptron, AvgPool is an average pooling operation, and MaxPool is a maximum pooling operation, σ is the sigmoid activation function, $f^{7\times 7}$ is a convolution operation with a convolution kernel size of 7×7, [;] is the splicing operation of the channel dimension.



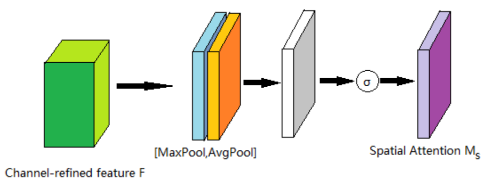


Figure 3 Channel attention block Figure 4 Spatial attention block

Figure 3 is the operation process of the channel attention submodule, which has one more global maximum pooling than the SE module, and the pooling operation itself is to extract high-level features, and different pooling means that the extracted high-level features are richer. Firstly, the average pooling and maximum pooling operations are used to aggregate the spatial information of a feature map to produce two spatial context descriptors: AvgPool(F) and MaxPool(F), which represent the features after average pooling and the features after maximum pooling, respectively. Then, both descriptors are fed forward into a network shared by both to produce a channel attention map. There are hidden layers of multi-layer sensing mechanisms into that shared network.After each spatial context descriptor is processed by the shared network, Mc(F) is obtained by using bitwise addition fusion output feature vector activated by sigmoid, and it is multiplied by the original input feature map to generate channel attention feature F' as the output of the channel attention submodule.

Figure 4 shows the operation process of the spatial attention submodule. The submodule takes the output feature map of the channel attention submodule as input, first completes the channel-based global maximum pooling and global average pooling operations, and then realizes the splicing of the two results according to the channel, merges into the feature map of channel number 2, and then becomes 1 channel through the convolution kernel size 7×7 standard convolutional layer. Then, the spatial attention feature is generated by sigmoid activation, and finally the feature is multiplied by the input feature of the spatial attention submodule, and the generated feature Figure F'' is used as the output of the spatial attention submodule, which is also the final output of the CBAM module.

3.3 Retinal vascular segmentation model based on ASR-IPN-V2

IPN and IPN-V2 are used for effective feature selection and dimensionality reduction to achieve OCTA vascular segmentation of end-to-end structures with 3D output to 2D. However, IPN and its improvements increase the complexity of the model while also adding additional computation, which consumes higher computer resources and relatively long segmentation time. In view of the above problems, this paper introduces the attention mechanism, improves the retinal vascular segmentation algorithm based on IPN-V2, and proposes a retinal vascular segmentation algorithm based on ASR-IPN-V2.On the one hand, in each sampling layer, the deep separable convolution is used instead of the conventional convolution for feature learning while reducing the amount of parameters. On the other hand, a lightweight and efficient convolutional block attention-weighted module is added to obtain the importance of channel information and spatial information in the feature map by learning, and the weight of each feature map is adaptively adjusted according to the learned importance to complete the recalibration of the feature map. The retinal vascular segmentation algorithm based on ASR-IPN-V2 pays more attention to subtle segmentation, improves the accuracy of segmentation, and occupies less computing resources.

Depth separable convolution

In IPN and IPN-V2, PLM consists of three 3D convolutional layers and a unidirectional pooling layer. The convolutional layer is used to extract image features, and the unidirectional pooling layer is used to select valid features along the projection direction. At the end of the network, convolutional layers are used to reduce the number of channels to aggregate the 2D planar information obtained by the PLM module. The main role of convolution is feature extraction. A regular convolution operation is a joint mapping that implements channel and spatial dependencies.Depthwise separable convolutions[22] are a method of convolution proposed by Chollet et al. Deep separable convolution operation is the process of decomposing a conventional convolution operation into a deep convolution plus a point-by-point convolution, so that it performs spatial convolution while keeping channels separated, and then convolutes in the direction of depth. Compared with conventional convolution operations, deep separable convolution operations have the advantages of reducing computational complexity and reducing the number of parameters. The deeply separable convolutional structure is shown in Figure 5:

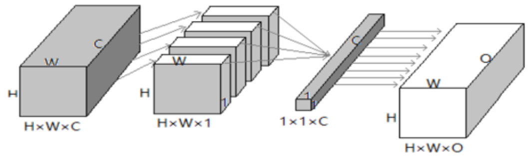


Figure 5 Depth separable convolutional structure diagram

Compared with conventional convolution operations, deep separable convolution broadens the width of the network, while greatly reducing the number of parameters in the network and reducing the resources consumed during calculation. Existing experiments have verified that deep separable convolution can reduce the number of parameters while ensuring task accuracy. Since the OCTA-500 segmentation dataset has the problem of small sample size, the deep separable convolution operation is compared with the conventional convolution operation, which will inevitably reduce the number of network parameters, reduce the computational complexity, and achieve better learning results. Therefore, this paper proposes to replace the conventional convolution operation in each sampling layer of IPN-V2 with a deep separable convolution operation, and replace the two conventional convolution operations used for feature extraction in each layer with a deep separable convolutional block.

Convolutional block attention mechanism

The CBAM module can achieve the parameter weights that are conducive to model training on the enlarged feature map at the minimum computational cost, reduce the weights of parameters that are unfavorable to the model, extract spatial and channel important information and combine them [23]. Due to the widespread phenomenon of small vascular ends and easy to be confused by the background in the images of OCTA-500 retinal blood vessels, some areas have missing vascular ends and small vascular parts are mistakenly divided into backgrounds when segmentation is used with IPN-V2. Therefore, this paper adds a CBAM module after each depth separable convolutional block (Previousconvblock) and before the next convolution block (Nextconvblock) to continue to learn the spatial information and channel information of the feature map and improve the accuracy of network segmentation. A schematic diagram of inserting a CBAM module into two adjacent convolutional blocks is shown in Figure 6 below:

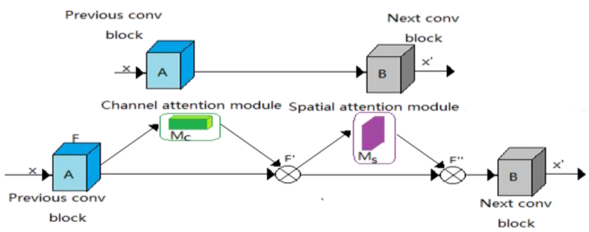


Figure 6 CBAM inserted into two adjacent convolutional blocks

The model uses deep separable convolution instead of regular convolution to reduce the number of parameters and computational complexity. At the same time, a lightweight and efficient convolutional block attention-weighted module is added, which can learn the importance of channel information and spatial information in the feature map, so as to better calibrate the feature map. Eventually, the model can complete the task of segmenting retinal blood vessels.

The final retinal vascular segmentation model of ASR-IPN-V2 is shown in Figure 7:

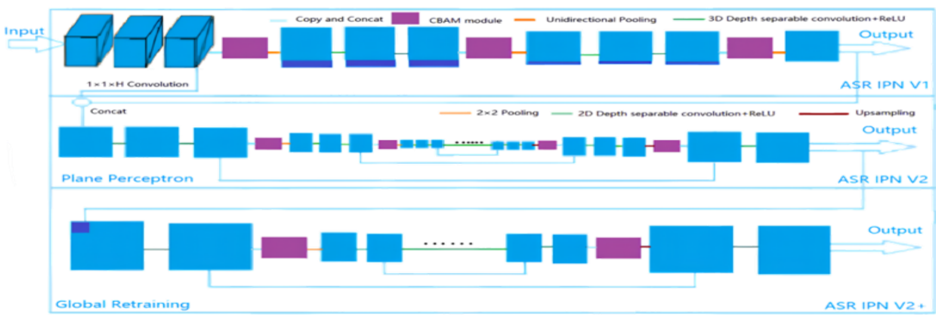


Figure 7 ASR-IPN-V2 network architecture

4 Experiments and discussion

The proposed methods and baselines are all implemented with PyTorch on 1 NVIDIA GeForce RTX 3090 GPU.The batch size is set as 4, the learning rate is 5×10−4 , epochs is 300, Adam is used as the optimizer.

4.1 Experimental datasets

The dataset used in this experiment is the 20th version of the retinal vascular dataset OCTA-500 published by Professor Chen Qiang and his team on IEEE-Data Port, which is currently the largest OCTA image dataset.The OCTA-500 dataset contains optical coherence tomography angiography data from 500 individuals. Each eye contains one OCTA image, a total of 1,000 OCTA images, including information such as age, gender, and ophthalmic history, which can be used to study the impact of these factors on ophthalmic diseases. In addition to being divided into two subsets, OCTA_6M and OCTA_3M by field of view, the OCTA-500 dataset can also be classified based on factors such as data source, image quality, and more. Among them, OCTA_6M dataset includes OCTA images with a 6x6mm field of view, which is suitable for studying changes in blood flow in large areas of retina, such as diabetic retinopathy, venous obstructive retinopathy and other diseases. The OCTA_3M dataset includes OCTA images with a 3x3mm field of view, which is suitable for studying small local retinal blood flow changes, such as retinal aneurysm, prominent vitreoretinopathy and other diseases.In addition, the OCTA-500 dataset is categorized according to the source of the data, including multi-center datasets and single-center datasets. The multi-center dataset includes data from different medical institutions, covering different eye disease types and different patient populations, with high representativeness and universality. The single-center dataset, on the other hand, only includes data from a single medical institution, and the data source is relatively single. In this experiment, only the dataset used for segmentation in the OCTA_6M subset was selected, in which the original image selected was the OCTA image of the whole eye, and the selected label was the retinal vascular segmentation tag, and the original image and the real label were 300 pieces each.

4.2 Data preprocessing

In this experiment, the dataset was randomly sampled and divided into training set and test set according to a 7:3 ratio, including a total of 210 training images and 90 test images. In order to unify the input size of different networks in subsequent comparative experiments, the pixel size was uniformly adjusted before entering the network, so that the pixel size reached 512×512. The distribution of the dataset is shown in Table 1 below:

Table 1 Distribution of datasets

data set	Training set	Test set
Number of samples	210	90
Sample dimension	512*512*3	512*512*3

In the training process, the small amount of data is one of the main reasons for model underfitting. The relative scarcity of OCTA images results in a small amount of data in the OCTA-500 retinal vascular segmentation dataset. Therefore, in order to reduce the impact of underfitting, this paper adopts the method of random cropping, flipping, scaling, and panning of the input image to enhance the online data of the input image, making the data richer and more diverse.

4.3 Analysis of experimental results

In order to verify the rationality of various improvement measures of the algorithm in this paper, the method of ablation experiment is adopted to verify the OCTA-500 dataset. The results are shown in Table 2.

Table 2 Ablation test results

model	Acc(%)	Sn(%)	Sp(%)	F1(%)
U-Net	94.36	72.4	94.3	76.2
IPN	95.3	74.8	**95.1**	80.3
IPN-V2	96.5	78.3	93.6	82.0
IPN-V2+	96.7	78.4	93.5	82.1

Ours	**96.9**	**80.1**	93.2	**82.5**

4.4 Discussion

It can be seen from Table 2 that IPN-V2 has a certain improvement in the comprehensive performance of IPN, ACC has increased to 0.965, Sn and F1 values have also increased relatively little, and Sp values have decreased, because the global retraining process has improved the performance of the network, but it will also increase a large number of parameters to a certain extent, resulting in overfitting. Compared with IPN and IPN-V2, the ASR-IPN-V2 algorithm has better performance in various indicators, although the performance in Sp is slightly worse, but compared with IPN and IPN-V2, the proposed algorithm has a great improvement in other indicators. Overall, the ASR-IPN-V2 algorithm proposed in this paper has excellent performance on the OCTA-500 dataset.

5 CONCLUSION

Retinal vascular segmentation based on deep learning has great practical significance in the field of medical image processing, which can be used for disease monitoring and prediction, through the analysis and processing of fundus images, it can timely detect and monitor changes in diseases, provide more accurate data support for the treatment and management of diseases, help doctors better analyze and identify pathological features in fundus images, and provide more accurate reference for disease diagnosis and treatment. Combined with the current research status and development trend of retinal vascular segmentation based on deep learning, this paper improves the retinal vascular segmentation based on IPN-V2 according to the IPN-V2 network structure, and the main conclusions are as follows:

Retinal vascular segmentation algorithm based on ASR-IPN-V2. In the IPN-V2-based retinal vascular segmentation model, the attention mechanism is introduced to improve the IPN-V2-based retinal vascular segmentation algorithm. The retinal vascular segmentation algorithm based on ASR-IPN-V2 uses deep separable convolution instead of conventional convolution, which can greatly reduce the amount of network parameters, improve the computational efficiency of the network, and maintain good segmentation effect. Adding the CBAM module after each deep separable convolutional block can further improve the segmentation accuracy of the network, because the CBAM module can learn the spatial information and channel information of the feature map, so as to better capture the semantic information of the image.

Through experimental analysis, it is found that compared with the IPN model and IPN-V2 model, the ASR-IPN-V2 model proposed in this paper has significantly improved the efficiency of retinal vascular segmentation, which has important practical significance for the early diagnosis, treatment and evaluation of ophthalmic diseases. At the same time, the research ideas and methods of this paper also provide reference for other studies of retinal vascular segmentation based on deep learning.

REFERENCES

1. Xu Weijun. Application of deep learning in the field of computer vision[J].Electronic Technology,2021,(05):20-21.)
2. SUN Yumeng. Application analysis of deep learning in computer vision analysis[J].China New Communications,2018,(23):169-171.)
3. WANG Jiangpeng. Development analysis of natural language processing technology based on deep learning[J].China Security,2022,(12):40-43.)
4. YE Fuming. Research on the application of deep learning in natural language processing NLP[J].Information Recording Materials,2021,(11):148-149.)
5. Yoav Goldberg,Che Wanxiang,Guo Jiang,Zhang Weinan,Liu Ming. Natural language processing based on deep learning[J].Journal of Chinese Information Technology,2021,(08):145.)
6. GUO Baobao,YANG Zhangyong,ZHANG Zixiang,ZHANG Jianing. Research on the application of deep learning technology in assisted decision-making[J].Science and Technology Innovation and Application,2020,(22):175-176.)
7. FANG Yun,MA Linzi. Deep learning technology and its application[J].Computer Knowledge and Technology,2020,(05):190-193.)
8. WANG Ying,ZHU Jiaming,SONG Qi. Retinal vascular image segmentation based on improved R2U-Net network[J].Radio Engineering,2022,(05):814-823.)

9. WU Kui, CAI Dongmei, JIA Peng, et al. Retinal vascular segmentation based on 2DGabor wavelet and combination line detection operator[J].Science Technology and Engineering,2016,16(12):106-112.)
10. Roychowdhury S，Koozekanani D D，Parhi K K. Blood vessel segmentation of fundus images by major vessel extraction and subimage classification[J]. IEEE Journal of Biomedical and Health Informatics，2014，19（3）：1118-1128.
11. JIA Hong, ZHENG Chujun, LI Canbiao, et al. FCM clustering retinal vascular segmentation based on local line structure constraints[J].Acta Optica Sinica,2020,40(9):40-49.)
12. Gao, D., Celik, N., Wu, X., Williams, B.M., Stylianides, A., Zheng, Y. (2020). A Novel Deep Learning Based OCTA De-striping Method. In: Zheng, Y., Williams, B., Chen, K. (eds) Medical Image Understanding and Analysis. MIUA 2019. Communications in Computer and Information Science, vol 1065. Springer, Cham. https://doi.org/10.1007/978-3-030-39343-4_16
13. Dali Chen，Yingying Ao，Shixin Liu.Semi-Supervised Learning Method of U-Net Deep Learning Network for Blood Vessel Segmentation in Retinal Images[J].Symmetry，2020，12（7）.
14. SUN Ying,DING Weiping,HUANG Jiashuang,JU Hengrong,LI Ming,GENG Yu. RCAR-UNet:Retinal vascular segmentation network based on rough attention mechanism[J].Computer Research and Development,2022:1-15.)
15. WANG Caiyun,GUAN Zhiyu,WU Xuanda,YAO Chen. Retinal vascular image segmentation algorithm based on multi-directional filtering[J].Progress in Laser and Optoelectronics,2022,(08):463-469.)
16. ZHANG Lijuan,MEI Chang,LI Chaoran,ZHANG Run. Retinal vascular image segmentation based on RAU-net[J].Infrared Technology,2021,(12):1222-1227+1233.)
17. Li M，Zhang Y ，Ji Z ，et al. IPN-V2 and OCTA-500：Methodology and Dataset for Retinal Image Segmentation[J]. 2020.
18. ZHANG Jia,ZHANG Lihong. Text Generation Image Method Based on Conditional Enhancement and Attention Mechanism[J].Journal of Testing Technology,2023,(02):112-119.)
19. Jie Hu and Li Shen and Samuel Albanie and Gang Sun and Enhua Wu,“Squeeze-and-Excitation Networks,”arXiv preprint arXiv:1709.01507,2017.
20. Sanghyun Woo and Jongchan Park and Joon-Young Lee and In So Kweon,“CBAM: Convolutional Block Attention Module,”arXiv preprint arXiv:1807.06521,2018.
21. JIANG Yun,LIU Wenhuan,LIANG Jing. Retinal vascular segmentation network of joint attention and Transformer[J].Computer Engineering and Science,2022,(11):2037-2047.)
22. Nithin Rao Koluguri and Taejin Park and Boris Ginsburg,“TitaNet: Neural Model for speaker representation with 1D Depth-wise separable convolutions and global context,”arXiv preprint arXiv:2110.04410,2021
23. JIANG Lulu,LI Sicong,CAO Jiawang,SUN Siqi,FENG Rui,ZOU Haidong. Retinal vascular segmentation based on multi-scale high-order attention mechanism[J].Computer Systems and Applications,2022,(10):368-374.)