Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

A saliency and Gaussian net model for retinal vessel segmentation*

Lan-yan XUE[†‡1,2], Jia-wen LIN[1], Xin-rong CAO[1], Shao-hua ZHENG[1], Lun YU[1]

[1]*College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China*

[2]*Institute of Computer and Information, Fujian Agriculture and Forestry University, Fuzhou 350002, China*

[†]E-mail: xuelanyan@126.com

Received June 19, 2017; Revision accepted Mar. 9, 2018; Crosschecked Aug. 15, 2019

Abstract: Retinal vessel segmentation is a significant problem in the analysis of fundus images. A novel deep learning structure called the Gaussian net (GNET) model combined with a saliency model is proposed for retinal vessel segmentation. A saliency image is used as the input of the GNET model replacing the original image. The GNET model adopts a bilaterally symmetrical structure. In the left structure, the first layer is upsampling and the other layers are max-pooling. In the right structure, the final layer is max-pooling and the other layers are upsampling. The proposed approach is evaluated using the DRIVE database. Experimental results indicate that the GNET model can obtain more precise features and subtle details than the UNET models. The proposed algorithm performs well in extracting vessel networks, and is more accurate than other deep learning methods. Retinal vessel segmentation can help extract vessel change characteristics and provide a basis for screening the cerebrovascular diseases.

Key words: Retinal vessel segmentation; Saliency model; Gaussian net (GNET); Feature learning
https://doi.org/10.1631/FITEE.1700404 **CLC number:** TP391

1 Introduction

Cerebrovascular disease is regarded as one of the three major causes of death. Stroke is the most common cerebrovascular disease, and hypertension is the most important and independent risk factor for stroke. Ikram et al. (2006) discovered that the risk of stroke in a patient with hypertension was related to the retinal artery diameter, vein diameter, and arteriovenous ratio. Information about the personal risk of potential cerebrovascular diseases can be obtained through the quantization parameter of the retinal vessel, which is widely used in clinical practice and may improve the prevention of strokes in hypertensive patients. Retinal vessel segmentation facilitates the quantification of characteristics. The advantage of this method is that it can overcome the randomness of boundary selection and the subjective error of quantization, and provide a convenient way for doctors to select vessels of interest. This can help in the early diagnosis of diseases and the monitoring of prognosis and treatment.

Many different approaches to vessel segmentation have been proposed, which can be divided into two categories (Zhu et al., 2015): unsupervised methods (including vessel tracking, matched filters, morphological processing, and a deformable model) and supervised methods (based on pixel classification systems such as neural networks).

1. Unsupervised methods

(1) Vessel tracking. Vessel tracking is based on the continuous structure of the retinal vessel. The initial seed should be chosen and followed in the direction of the vessel. The vessel between two points can be obtained when the termination condition is reached. Liu and Sun (1993) first proposed the vessel tracking method. Solouma et al. (2002) proposed a new real-time method, initiating a grid of seed contours over the whole image by splitting deformation

‡ Corresponding author

* Project supported by the Natural Science Foundation of Fujian Province, China (No. 2016J0129) and the Educational Commission of Fujian Province of China (No. JAT170180)

ORCID: Lan-yan XUE, http://orcid.org/0000-0003-2886-2983

and merging according to the preset criteria until the whole vessel tree was demarcated. A Gaussian filter was then used to filter the image to extract the vessels. Kumar et al. (2015) proposed a modified multiscale vessel method and reduced the processing by applying this vessel only within the connected blood vessel region. They proposed a novel method, combined with blood vessel segmentation, centerline extraction, and radius detection. Vessel tracking provides precise vessel connectivity information at branching and crossover points for early detection of many systemic diseases. However, vessel tracking may be confused by vessel crossings and bifurcations and may terminate when the contrast between the vessels and the background is weak.

(2) Matched filters. A matched filter is based on the gray distribution of the vessel section, which conforms to the Gaussian distribution. The vessel points can be determined by the maximum response value of the convolution between the Gaussian filter and the retinal image. Chaudhuri et al. (1989) designed a two-dimensional matched filter to detect a vessel in 12 directions, and obtained the maximum response as the output. Odstrcilik et al. (2013) combined a matched filter and a minimum error to segment a vessel. A matched filter can adopt the property of a vessel section sufficiently. However, the accuracy of this approach is low when the vessels and background have low contrast.

(3) Morphological processing. In morphological processing, dilation and erosion are used to process the image. The vessel edge can be obtained when the original image is subtracted from the processed image. Zana and Klein (2001) combined morphology and curvature estimation to extract a vessel. Based on this approach, Ayala et al. (2005) used median fuzzy set methods to extract vessels. Imani and Pourreza (2016) used morphological component analysis (MCA) to improve the detection of retinal blood vessels. First, an MCA algorithm with appropriate transforms was adopted to separate vessels and lesions from each other. Then Morlet wavelet transform was applied to enhance the retinal vessels. Finally, the vessel map was obtained by adaptive thresholding. This method can obtain a good noise immunity performance without relying on prior knowledge of vessels.

(4) Deformable model. An algorithm based on a deformable model is used to describe the boundary of the target adopting a continuous curve. Vese and Chan (2002) proposed a multiphase level set framework for image segmentation, using the Mumford and Shah model based on active contours without edges. Zhao et al. (2014) proposed retinal vessel segmentation based on level set and region growing.

2. Supervised methods

Franklin and Rajan (2014) used a back-propagation algorithm in a neural network for vessel segmentation. Zhu et al. (2016) proposed an effective method for retinal vessel segmentation based on supervised learning, in which a 39-dimensional feature vector was extracted for each pixel, consisting of local, morphological, and Gabor features. The sampled set was initially treated by the classification and regression tree as a weak classifier, and was then strengthened by a trained AdaBoost-based classifier as a strong classifier to classify pixels. Supervised methods are time consuming because they require training, which depends on hand-labeled vessel segmentation for references.

A retinal vessel segmentation method based on deep learning has achieved a higher precision than other supervised algorithms (Zhu et al., 2015). Maji et al. (2015) proposed a hybrid architecture based on deep and ensemble learning to detect vessels, followed by the unsupervised learning of sparse denoising and auto-encoding training with a random forest used to detect vessels. Ronneberger et al. (2015) proposed a UNET model, which was established on a fully convolutional network (FCN). The UNET model was so named because of its "U"-like shape. The back-half operation of the UNET model was upsampling instead of max-pooling. The model has been used for biological image segmentation including retinal vessels, and has achieved good results. Liu et al. (2014) and Fu et al. (2016) proposed a novel method for retinal vessel segmentation based on deep learning and a random field, respectively.

Given the differentiation in color, brightness, and texture features between regions of retinal vessels and background in fundus images, a saliency model can be used to highlight the vessel region. The saliency characteristics of the underlying data can highlight the vessels in images.

The left half structure of the UNET model is max-pooling. However, some details are lost in spite of upsampling in the back-half structure. In this study,

we propose a novel deep learning model called Gaussian net (GNET), since the structure resembles a Gaussian distribution curve. The GNET model is an improved version of the UNET model. The GNET model adopts a bilaterally symmetrical structure. In the left structure, the first layer is upsampling and the other layers are max-pooling. In the right structure, the final layer is max-pooling and the other layers are upsampling. The original image is directly used as an input of the new algorithm, which can ensure the universality of the learning characteristics without the need to manually design features, based on prior knowledge. Compared with the UNET model, the improved GNET model is proposed to obtain more precise features and subtle details.

Consider the advantages of a saliency model and deep learning, both of which are combined with vessel segmentation. There are two important ideas in the proposed algorithm: (1) A low-level and bottom-up visual saliency model is adopted to detect the fundus image. By computing the distance between the mean pixel value and the Gaussian blurred version of the fundus image, the distance image is used in the saliency image. This method can highlight the saliency region of the retinal vessel and obtain clear edges with a complete resolution. (2) Given that the GNET model used to learn the characteristics of the fundus image may lose some details, an improved model is proposed, and the classifying results are obtained through a training classifier.

The flowchart of the proposed algorithm is shown in Fig. 1.

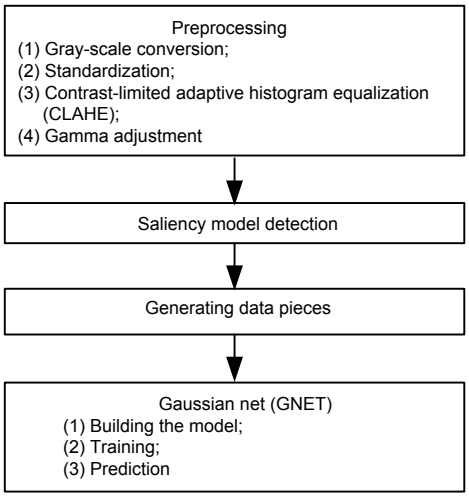


Fig. 1 Flowchart of the proposed algorithm

2 Related work

2.1 Deep learning

2.1.1 Fully convolutional network

An FCN (Shelhamer et al., 2017) has been transformed from a convolutional neural network (CNN). A CNN is a multilayer neural network, with each layer containing convolution and pooling transformations. The CNN structure contains convolution, sampling, and fully connected layers. Fixed-length feature vectors can be obtained in the fully connected layer and can be classified by the softmax classifier. In an FCN model, the fully connected layer is used instead of the convolution layer. The resolution becomes lower after pooling. The deconvolution layer is followed by the last convolution layer. The output image has the same size as the input image. Then pixels in the feature map can be classified by the softmax classifier. The structure of the FCN is shown in Fig. 2.

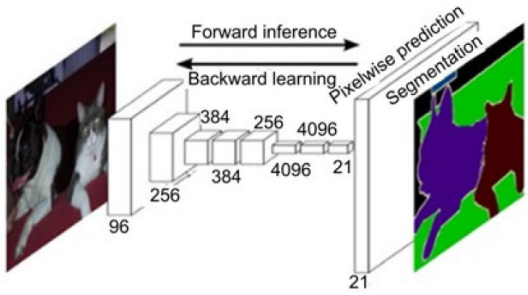


Fig. 2 Structure of the fully convolutional network

1. Convolution layer

The input image is convolved with a convolution kernel, which can generate a new feature map. The size of the output image will be smaller than that of the input image. For example, consider the size of an input image to be 32×32 pixels and the size of a convolution layer with four convolution kernels to be 5×5 pixels. Four feature maps will be obtained after convolution, and the size of the feature maps will be (32−5+1)×(32−5+1)=28×28 pixels. These feature maps comprise the input layer of the subsequent convolution layer.

2. Pooling layer

The pooling layer is the downsampling layer. The size of the feature maps will decrease after downsampling. The max-pooling operation is adopted extensively and the stride is two. The

maximum value of the region is 2×2 pixels, which is considered a characteristic of this region. After downsampling, the size of the feature maps will be reduced to a quarter.

3. Deconvolution layer

The deconvolution layer is the upsampling layer. The size of the feature maps will increase after upsampling. The size of the image becomes small after pooling, and can remain unchanged when the same stride is chosen for upsampling.

4. Dropout

Dropout can prevent overfitting in the case of few training samples. In overfitting, the network is good for fitting the training set because the loss is small and the accuracy is high. The accuracy decreases when the loss of the testing set increases. Some of the hidden layer nodes are lost, which prevents the network from fitting the training set when each sample is trained.

2.1.2 UNET model

The UNET model (Ronneberger et al., 2015) has been established based on the FCN. However, some changes and extensions were made to enable the network to obtain a better precision with fewer training images. The back-half operation of the UNET model is upsampling instead of pooling of the FCN. First, the pooling operation is implemented to decrease the resolution of the image. Then the upsampling operation is implemented to increase the image resolution. The output layer will be merged with the high-resolution features of the left structure. Thus, many accurate features can be obtained. The UNET model is symmetrical such that the resolutions of the input and output images are the same.

The UNET model and the simplified UNET model are shown in Figs. 3 and 4, respectively. The UNET model can preserve the complete position information, which is important for pixel segmentation. However, complete feature details cannot be obtained.

The stair layer C_i (i=1, 2, …, 9) in Fig. 4 represents a series of operations in the original model, which involves convolution and dropout.

1. Interpolation operation

The size of the feature images is increased through deconvolution. However, the size is decreased after convolution. The sampling operation should be implemented in each convolution layer after deconvolution. Then the corresponding kernel is implemented to fill the details to restore the original image.

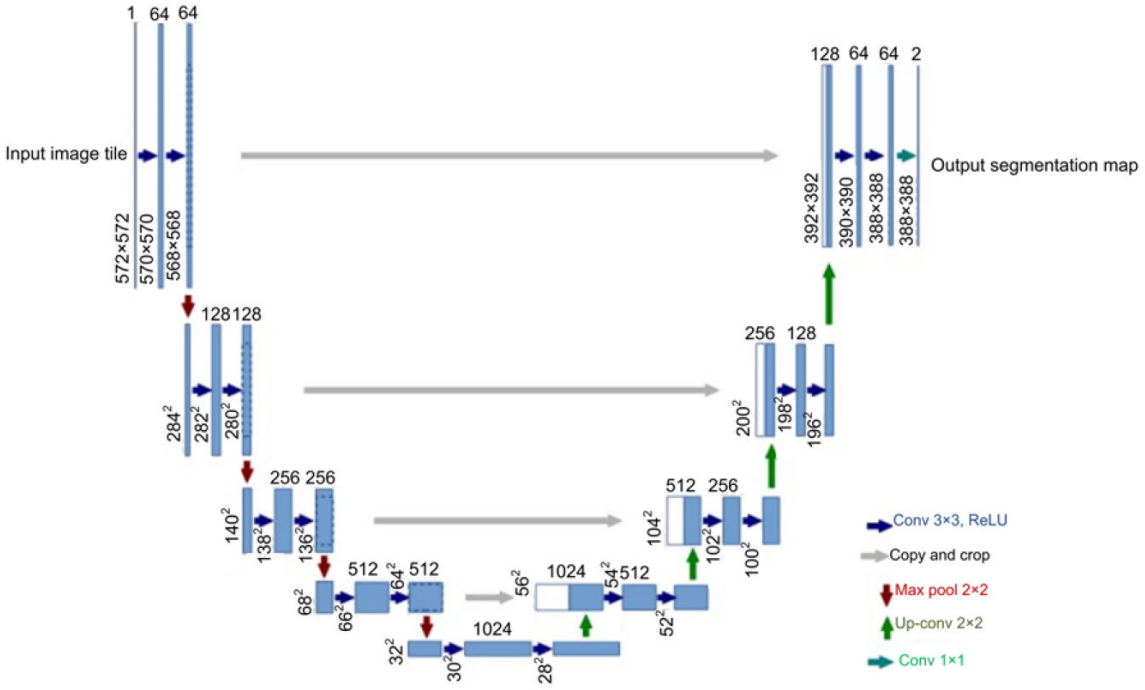


Fig. 3 Structure of the UNET model (References to color refer to the online version of this figure)

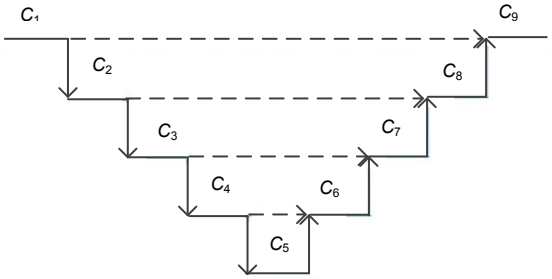


Fig. 4 Structure of the simplified UNET model
The upward arrow represents upsampling; the downward arrow represents max-pooling; the dotted line with an arrow represents merging

2. Merge operation

The output layer should be merged with the corresponding layer on the left structure, and the sampling should be completed.

2.1.3 GNET model

The left structure of the GNET model consists of pooling layers, and results in a low resolution. However, some details are lost in spite of upsampling in the back-half structure. The improved GNET model is proposed to obtain more precise features and subtle details. In the left structure, the first layer is upsampling and the other layers are max-pooling. In the right structure, the last layer is max-pooling and the other layers are upsampling. The novel model is also symmetrical. The simplified structure is shown in Fig. 5. The novel structure is called the GNET model, because it resembles the shape of the Gaussian distribution curve.

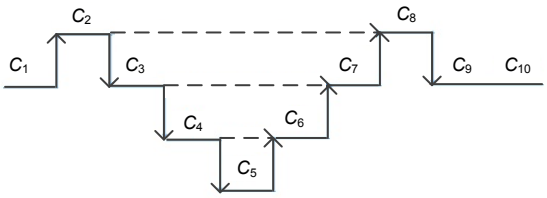


Fig. 5 Simplified structure of the GNET model
The upward arrow represents upsampling; the downward arrow represents max-pooling; the dotted line with an arrow represents merging

2.2 Saliency detection

Saliency detection is widely used in image retrieval, target recognition, and image segmentation. Fu et al. (2013) adopted comparison features, space features, and similarities to detect saliency after clustering all pixels of an image. Liu et al. (2014) proposed a saliency detection model based on hierarchical segmentation. The key of fine-grained image classification was to find the saliency of the image (Peng et al., 2018). Xiao et al. (2015) integrated two level attentions: an object-level attention that selects image patches relevant to the object and a part-level attention that selects discriminative and saliency parts. He et al. (2017) further exploited the object-part attention model (OPAM) for weakly supervised fine-grained image classification.

3 The proposed method

The original fundus images should be preprocessed by a series of transformations. The saliency image can be obtained from the gray image through the saliency model. The pieces of saliency images are the inputs of the CNN. A 48×48-pixel region is selected as a piece. The input of the convolution layer is the same as the output. The advantage of size consistency is that the front layer and the latter layer can be directly merged without shear. To minimize overfitting, the dropout layer is added between the convolution layers in each stair layer. The final convolution layer whose size is 1×1 pixel is transformed into two layers. Finally, the output with two labels can be obtained using the softmax activation function.

3.1 Preprocessing

The 20 training set images should be preprocessed with the following transformations before training: gray-scale conversion, standardization, contrast-limited adaptive histogram equalization, and Gamma adjustment. First, the gray image is normalized as

$$f_{\text{gout}} = \frac{f_{\text{gin}} - f_{\text{gin\_mean}}}{f_{\text{gin\_std}}}, \tag{1}$$

$$f_{\text{normalized}} = \frac{f_{\text{gout}} - f_{\text{gout\_min}}}{f_{\text{gout\_max}} - f_{\text{gout\_min}}} \times 255, \tag{2}$$

$$f_{\text{out}} = 255 \times \left(\frac{f_{\text{normalized}}}{255} \right)^{\frac{1}{\alpha}}, \tag{3}$$

where f_{gin} is the gray image, f_{gin_mean} the average of the image, f_{gin_std} the standard deviation of the image, $f_{normalized}$ the normalized image, f_{out} the image after adjustment using the Gamma curve, and α is set to 0.25. Then the image can be adjusted using the Gamma curve. The preprocessed images are shown in Fig. 6.

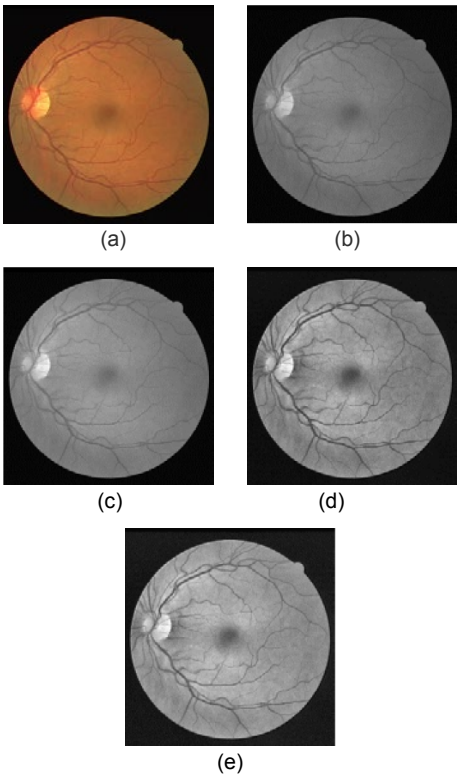


Fig. 6 Preprocessing processes: (a) original color image; (b) gray image; (c) standardized image; (d) equalization image; (e) Gamma correction image

3.2 Saliency model

Given the differentiation in color, brightness, and texture features between the regions of retinal vessels and the background in fundus images, saliency detection can highlight the retinal vessels. We adopt a frequency-tuned approach for computing saliency in images using low-level features of gray images, which is easy for quick implementation and can provide full-resolution saliency maps.

The pixel distance to describe the saliency for the gray space is defined as (Achanta et al., 2009)

$$D = | f_{out\_m} - f_{out\_G}(x, y) |, \quad (4)$$

where f_{out_m} is the mean pixel value of the preprocessed images and f_{out_G} the image after Gaussian filtering. The fine texture details, noises, and coding artifacts can be eliminated in the Gaussian blurred version. The retinal vessels can be highlighted and the lesion of the fundus images will be removed in this saliency model. The saliency image is shown in Fig. 7.

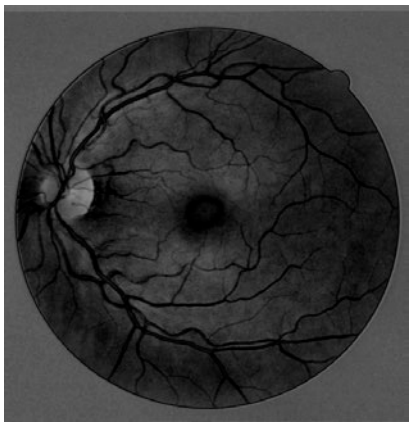


Fig. 7 Saliency image

3.3 Generating data pieces

The training network may not be effective when the sample number is small. Thus, the training sets are extended into more sub-images in the GNET model. Each image is randomly intercepted into 9500 pieces. The training set has a total of 190 000 pieces. An arbitrary pixel of the image is chosen as the center, and a 48×48-pixel region is selected as a piece. The stride is five. If the size does not match, then the piece should be filled in black to meet the size of the pieces. Note that the pieces may be located partially or entirely in the outer portion of the area of interest. The pieces will overlap. The network can learn to distinguish between the retinal vessels and the boundary region of interest. Sample pieces are shown in Fig. 8.

3.4 Building the model

After a series of preprocessing operations and the generation of additional training samples, the data pieces are sent to the network for feature learning. The GNET model has nine stair layers.

In the first stair layer (C_1), the input image is convoluted with 32 convolution kernels with the size of 3×3 pixels, and 20% of the connection of the input neurons is randomly disconnected. Then the output

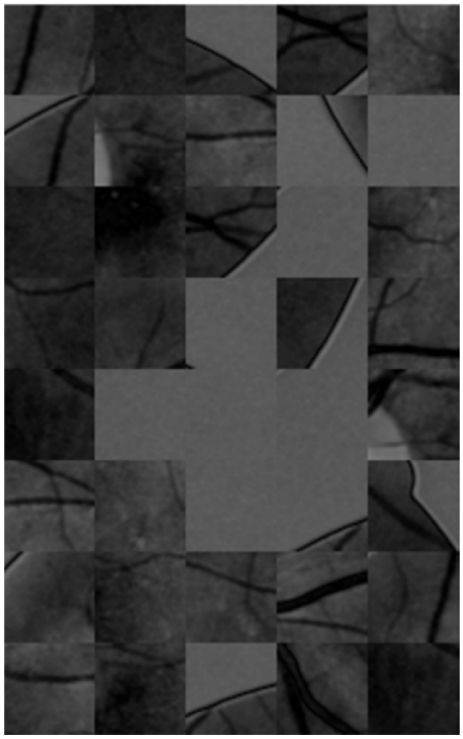


Fig. 8 A sample of image pieces

images are also convoluted with 32 convolution kernels with the size of 3×3 pixels. The output feature images are upsampling, with a stride of two. These feature maps are the input images of the subsequent stair layers.

Starting with the second stair layer, two paths are divided. Downsampling is completed in one path. In the second stair layer, the output image of the previous layer is convoluted with 16 convolution kernels with the size of 3×3 pixels, and 20% of the connection of the input neurons is randomly disconnected. Then the output images are also convoluted with 16 convolution kernels with the size of 3×3 pixels. The output feature images are max-pooling, with a stride of two. The image will be sheared in the other path, which is merged with the output of the corresponding stair layer in the right structure. The information of the front and back layers is crossed. Positioning will be better.

Operations in other layers are the same as those in the second stair layer. However, the convolution parameters in the 10th layer are different from those in the previous layers. Two convolution kernels with the size of 1×1 pixel are used in the 10th layer to change the output into two channels. This layer is called the "activation layer," which is implemented by the activation function. In this study, we adopt softmax to achieve pixel segmentation.

The pieces of the intermediate layer are assembled to obtain the complete output feature map of the middle layer. The output feature maps in the first convolution layer of the 8th stair layer are shown in Fig. 9.

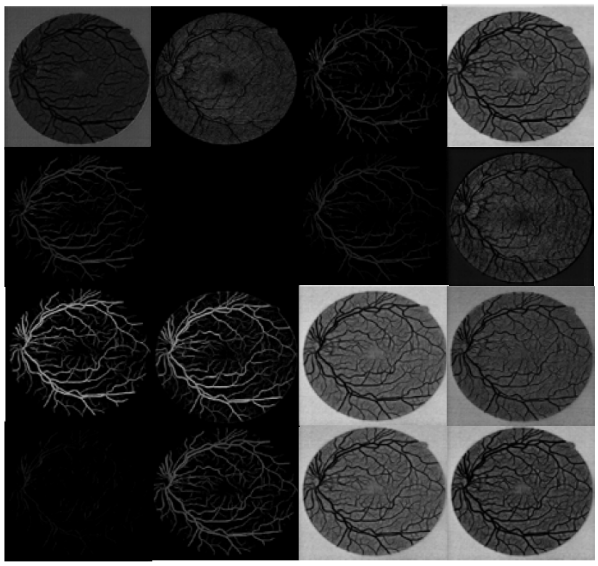


Fig. 9 Output feature maps of the 8th stair layer

3.5 Training process of the network

The feature map layer of the convolution layer is convoluted between the last layer and the convolution kernels. Then a bias b_j is added. The output feature images can be obtained using a rectified linear unit. Each output of the convolution layer can be a combination of many convolutions, expressed as

$$x_j^l = f\left[\sum_{i\in M_j} x_i^{l-1} \otimes k_{ij}^l\right] + b_j^l, \tag{5}$$

where x_j^l is the j^{th} output feature in the l^{th} layer, M_j the input map, k_{ij}^l the kernel in the l^{th} layer, and $\otimes$ stands for the convolution operation.

A rectified linear unit (ReLU) (Schmidhuber, 2015) is constant in the interval of $(-\infty, 0)$ and linear in the interval of $[0, +\infty)$. The function can be expressed as

$$f(z) = \max(0, z). \tag{6}$$

The derivative is expressed as

$$f'(z)=\begin{cases}0, & z<0,\\ 1, & z>0,\\ \text{undefined}, & z=0.\end{cases} \tag{7}$$

The process of GNET model training involves learning the weight kernels and the bias. The stochastic gradient descent method is adopted to solve the model parameters. The loss can be reduced through the iteration steps, and the network converges through constantly updating the weights.

Softmax is used in the last feature map to achieve pixel segmentation, expressed as

$$\text{softmax}(x)_i=\frac{\exp(x_i)}{\sum_j \exp(x_j)}. \tag{8}$$

The cross entropy is used to optimize the classification problem. If the probabilities of the final sample are assumed to be p and q, then

$$H(p,q)=\sum p(i)\cdot\log\frac{1}{q(i)}. \tag{9}$$

3.6 Prediction of the network

The fundus image can be predicted after the training of the network. The original image is cut into pieces every 5 pixels after preprocessing. The size of each piece is 48×48 pixels. Then the image is sent to the model and the corresponding weights are used to perform the prediction of the network. The input and output of the network are all pieces. These pieces can be assembled according to the size of the pieces and the stride. The black part caused by size mismatch in the fragmentation process is removed. The prediction result is shown in Fig. 10.

Fig. 10 Prediction result of the retinal vessels

4 Experimental results and analysis

4.1 Data collection

The proposed algorithm was evaluated using the DRIVE database. The database was established from a diabetic retinopathy screening program in the Netherlands. The ages of the 400 diabetic subjects were between 25 and 90. The resolution of the retinal images was 568×584, and the images were captured using a Cannon CR5 non-mydriatic 3CDD camera with a 45° field of view. The site provided hand-labeled data from two graders, which could be used to evaluate an algorithm's performance.

The proposed algorithm was implemented in the operating system of Microsoft Windows 10, with an Intel E3-1231 v3 CPU, ASUS B85-PRO GAMER motherboard, NVDIA GeForce GTX 1080 graphics, and 32 GB memory. The development environment was Visual Studio 2013+CUDA 8.0.27+cuDNN 5105, which used the Keras deep learning framework and Python language.

4.2 Evaluation methodology

To evaluate the algorithm, we calculated the sensitivity (Sen), specificity (Spec), and accuracy (ACC). Assume that true positive (TP) and true negative (TN) show the correct vessel pixels and background pixels respectively, consistent with the ophthalmologist's judgement. False positive (FP) and false negative (FN) show the wrong vessel pixels and background pixels respectively, according to the ophthalmologist's judgement. The evaluation formulae are expressed as

$$\text{Sen}=\frac{\text{TP}}{\text{TP}+\text{FN}}, \tag{10}$$

$$\text{Spec}=\frac{\text{TN}}{\text{FP}+\text{TN}}, \tag{11}$$

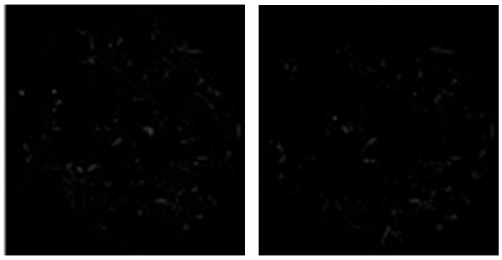
$$\text{ACC}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}}. \tag{12}$$

4.3 Results and analysis

All images were processed by the algorithm. The vessel segmentation results are shown in Fig. 11.

There were 331 712 pixels in one image. There was no big difference in the segmentation accuracy between the UNET model and the GNET model. The

segmentation results were not so different with the unaided eye. Thus, the two segmentation results and the labeling results can be subtracted separately, yielding the difference between the two segmentation results. If the segmentation results were close to the labeling results, the subtracted image would have few white pixels. The difference image between the segmentation results from the GNET model and the labeling results had fewer white pixels than that from the UNET model (Fig. 11). It was closer to the labeling and was more accurate.



(a) (b)

Fig. 11 Difference images between the segmentation results and the labeling results using the UNET model (a) and the GNET model (b)

The proposed algorithm was compared with the retinal vessel segmentation algorithms based on deep learning and manual labeling in the DRIVE database. Fig. 12 shows that the vessel network is continuous and clear, and that small vessels can be extracted by the proposed algorithm. The effect is good for those images with poor contrast. The segmentation results are close to those of manual labeling.

Table 1 shows the performance comparison of our method with the methods used by other researchers in terms of Spec, Sen, and ACC. The quantitative of both enhancement and segmentation steps shows that our method effectively detects the blood vessels with an accuracy of above 95%.

The performance indices of vessel segmentation in the DRIVE database are given in Table 2. The difference between the best and the worst segmentation results is due to the contrast, which results in the loss of some small vessels. The novel algorithm has a high accuracy, but is time consuming. The structure will be improved in future to solve this problem.

Table 1 Performance comparison between our method and the methods used by other researchers

Method	Spec	Sen	ACC
Wang et al. (2015)'s	–	0.7527	0.9457
Maji et al. (2015)'s	–	–	0.9327
Fu et al. (2016)'s	–	0.7603	0.9523
Ronneberger et al. (2015)'s	0.9835	0.7671	0.9559
Ours	0.9861	0.7967	0.9629

Spec: specificity; Sen: sensitivity; ACC: accuracy

Table 2 Performance indices of vessel segmentation from images in the DRIVE dataset

Image	Spec	Sen	ACC
01_test	0.9742	0.8656	0.9603
02_test	0.9871	0.8302	0.9639
03_test	0.9824	0.8012	0.9589
04_test	0.9898	0.7948	0.9650
05_test	0.9887	0.8043	0.9670
06_test	0.9914	0.7324	0.9564
07_test	0.9836	0.8273	0.9674
08_test	0.9822	0.7931	0.9638
09_test	0.9933	0.6889	0.9576
10_test	0.9830	0.8422	0.9683
11_test	0.9841	0.8331	0.9657
12_test	0.9848	0.8193	0.9656
13_test	0.9916	0.7351	0.9539
14_test	0.9799	0.8567	0.9664
15_test	0.9847	0.8015	0.9648
16_test	0.9853	0.8260	0.9656
17_test	0.9845	0.8013	0.9647
18_test	0.9910	0.7970	0.9652
19_test	0.9899	0.8024	0.9629
20_test	0.9914	0.7188	0.9547
Average	0.9861	0.7986	0.9629
Maximum	0.9933	0.8656	0.9683
Minimum	0.9742	0.6889	0.9539

Spec: specificity; Sen: sensitivity; ACC: accuracy

The receiver operating characteristic (ROC) curves of the UNET and GNET models are shown in Fig. 13. The ROC curves can be used to quantify the performance of a classifier. A good classifier has an area under curve (AUC) of about one. The true positive rate is on the y axis, whereas the false positive rate is on the x axis. The AUC of the UNET model is close to 0.9835 and that of the model combining saliency and GNET is close to 0.9843.

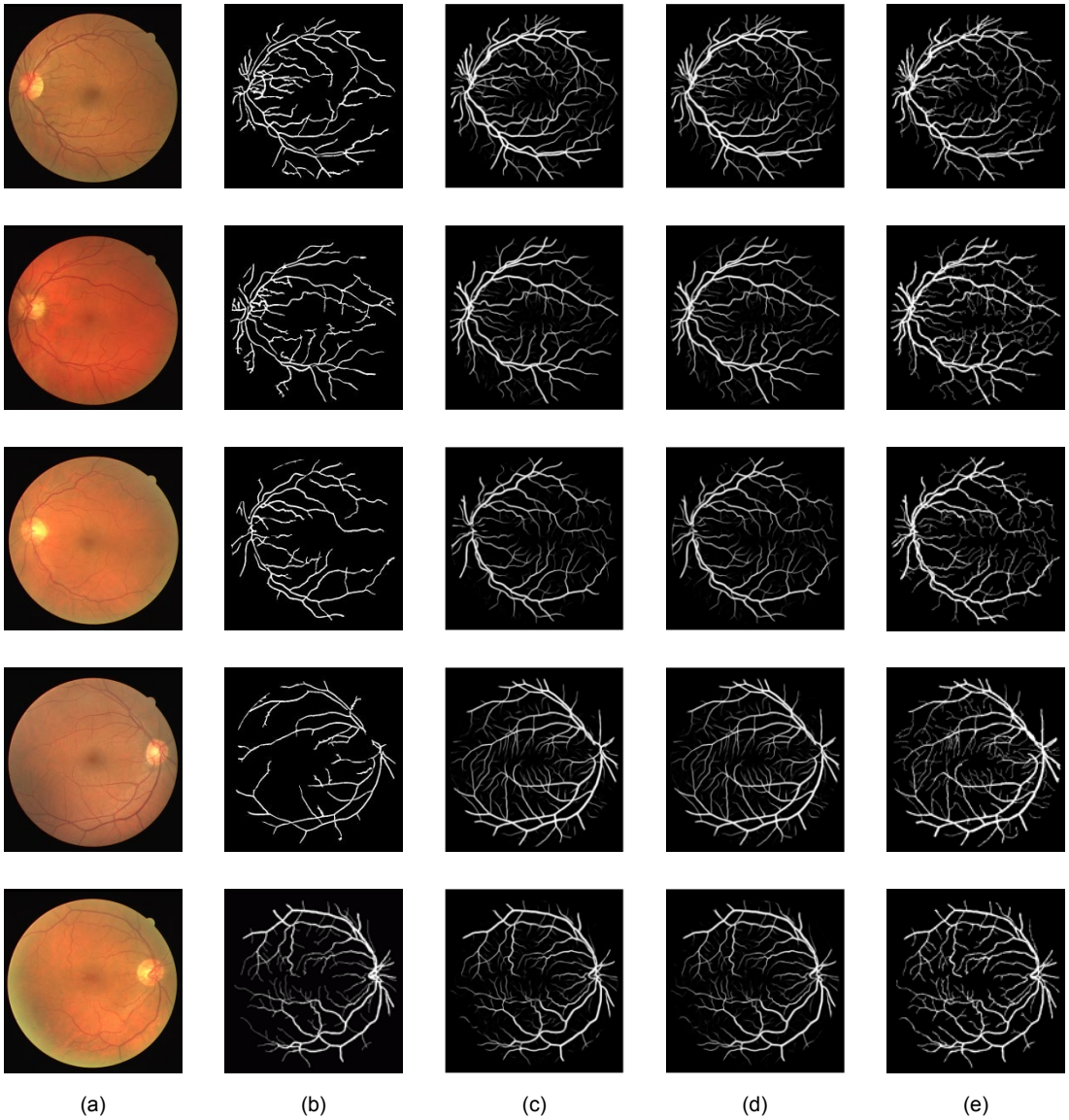


Fig. 12 Retinal vessel segmentation using images from the DRIVE database: (a) original images; (b) results using a CNN model; (c) results using a UNET model; (d) results using the proposed algorithm; (e) manual labeling images

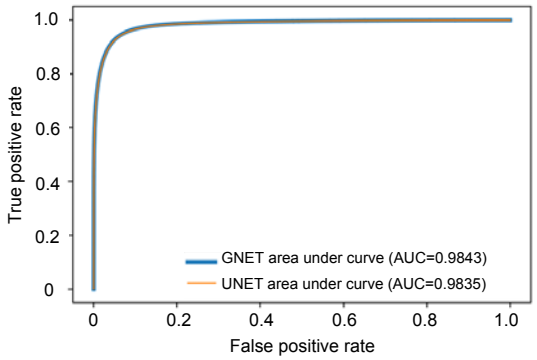


Fig. 13 ROC curves of the GNET and UNET models

5 Conclusions and future work

To obtain clear edges with a complete resolution of retinal vessels, a saliency image has been used as the input image of the deep learning network. A novel GNET model has been proposed to train the features and classify the pixels using classifiers. Upsampling has been operated before max-pooling and the opposite operation was implemented in the right layer to reduce the loss of several features caused by the size after convolution and to obtain details. The proposed algorithm has been evaluated using images from the

DRIVE database. Compared with other deep learning algorithms, the proposed algorithm had higher accuracy, sensitivity, and specificity. The retinal vessels can be accurately segmented, and vessel change characteristics can be extracted to provide a basis for the screening of cerebrovascular diseases.

Our future work will focus on three aspects: (1) Given that image super-resolution can increase the size of a small image and prevent the degradation of image quality, incorporate a deep convolutional layer (Hu et al., 2016); (2) Use an anchored neighborhood index algorithm (Wang et al., 2018) to generate more patches; (3) Apply Bayesian learning (Wang et al., 2017) to deep learning to produce a more accurate model from only a few training samples. These developments will be employed to further improve the performance of retinal vessel segmentation.

Compliance with ethics guidelines

Lan-yan XUE, Jia-wen LIN, Xin-rong CAO, Shao-hua ZHENG, and Lun YU declare that they have no conflict of interest.

References

Achanta R, Hemami S, Estrada F, et al., 2009. Frequency-tuned salient region detection. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.1597-1604. https://doi.org/10.1109/CVPR.2009.5206596

Ayala G, Leon T, Zapater V, 2005. Different averages of a fuzzy set with an application to vessel segmentation. *IEEE Trans Fuzzy Syst*, 13(3):384-393. https://doi.org/10.1109/TFUZZ.2004.839667

Chaudhuri S, Chatterjee S, Katz N, et al., 1989. Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Trans Med Imag*, 8(3):263-269. https://doi.org/10.1109/42.34715

Franklin SW, Rajan SE, 2014. Retinal vessel segmentation employing ANN technique by Gabor and moment invariants-based features. *Appl Soft Comput*, 22:94-100. https://doi.org/10.1016/j.asoc.2014.04.024

Fu HZ, Cao XC, Tu ZW, 2013. Cluster-based co-saliency detection. *IEEE Trans Imag Process*, 22(10):3766-3778. https://doi.org/10.1109/TIP.2013.2260166

Fu HZ, Xu YW, Kee DW, et al., 2016. Retinal vessel segmentation via deep learning network and fully-connected conditional random fields. Proc IEEE 13th Int Symp on Biomedical Imaging, p.698-701. https://doi.org/10.1109/ISBI.2016.7493362

He XT, Peng YX, Zhao JJ, 2017. Fine-grained discriminative localization via saliency-guided faster R-CNN. Proc 25th ACM Int Conf on Multimedia, p.627-635. https://doi.org/10.1145/3123266.3123319

Hu YT, Wang NN, Tao DC, et al., 2016. SERF: a simple, effective, robust, and fast image super-resolver from cascaded linear regression. *IEEE Trans Imag Process*, 25(9):4091-4102. https://doi.org/10.1109/TIP.2016.2580942

Ikram MK, de Jong FJ, Bos MJ, et al., 2006. Retinal vessel diameters and risk of stroke: the Rotterdam study. *Neurology*, 66(9):1339-1343. https://doi.org/10.1212/01.wnl.0000210533.24338.ea

Imani E, Pourreza HR, 2016. A novel method for retinal exudate segmentation using signal separation algorithm. *Comput Method Program Biomed*, 133:195-205. https://doi.org/10.1016/j.cmpb.2016.05.016

Kumar RP, Albregtsen F, Reimers M, et al., 2015. Blood vessel segmentation and centerline tracking using local structure analysis. Proc 6th European Conf of the Int Federation for Medical and Biological Engineering, p.122-125. https://doi.org/10.1007/978-3-319-11128-5_31

Liu I, Sun Y, 1993. Recursive tracking of vascular networks in angiograms based on the detection-deletion scheme. *IEEE Trans Med Imag*, 12(2):334-341. https://doi.org/10.1109/42.232264

Liu Z, Zou WB, Li LN, et al., 2014. Co-saliency detection based on hierarchical segmentation. *IEEE Signal Process Lett*, 21(1):88-92. https://doi.org/10.1109/LSP.2013.2292873

Maji D, Santara A, Ghosh S, et al., 2015. Deep neural network and random forest hybrid architecture for learning to detect retinal vessels in fundus images. Proc 37th Annual Int Conf of the IEEE Engineering in Medicine and Biology Society, p.3029-3032. https://doi.org/10.1109/EMBC.2015.7319030

Odstrcilik J, Radim K, Attila B, et al., 2013. Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database. *IET Image Process*, 7(4):373-383. https://doi.org/10.1049/iet-ipr.2012.0455

Peng YX, He XT, Zhao JJ, 2018. Object-part attention model for fine-grained image classification. *IEEE Trans Imag Process*, 27(3):1487-1500. https://doi.org/10.1109/TIP.2017.2774041

Ronneberger O, Fischer P, Brox T, 2015. U-Net: convolutional networks for biomedical image segmentation. Proc 18th Int Conf on Medical Image Computing and Computer-Assisted Intervention, p.234-241. https://doi.org/10.1007/978-3-319-24574-4_28

Schmidhuber J, 2015. Deep learning in neural networks: an overview. *Neur Netw*, 61:85-117. https://doi.org/10.1016/j.neunet.2014.09.003

Shelhamer E, Long J, Darrell T, 2017. Fully convolutional networks for semantic segmentation. *IEEE Trans Patt Anal Mach Intell*, 39(4):640-651. https://doi.org/10.1109/TPAMI.2016.2572683

Solouma NH, Youssef ABM, Badr YA, et al., 2002. A new real-time retinal tracking system for image-guided laser

treatment. *IEEE Trans Biomed Eng*, 49(9):1059-1067. https://doi.org/10.1109/TBME.2002.802059

Vese LA, Chan TF, 2002. A multiphase level set framework for image segmentation using the Mumford and Shah model. *Int J Comput Vis*, 50(3):271-293. https://doi.org/10.1023/A:1020874308076

Wang NN, Gao XB, Sun LY, et al., 2017. Bayesian face sketch synthesis. *IEEE Trans Imag Process*, 26(3):1264-1274. https://doi.org/10.1109/TIP.2017.2651375

Wang NN, Gao XB, Sun LY, et al., 2018. Anchored neighborhood index for face sketch synthesis. *IEEE Trans Circ Syst Video Technol*, 28(9):2154-2163. https://doi.org/10.1109/TCSVT.2017.2709465

Wang XH, Zhao YQ, Liao M, et al., 2015. Automatic segmentation for retinal vessel based on multi-scale 2D Gabor wavelet. *Acta Automat Sin*, 41(5):970-980 (in Chinese). https://doi.org/10.16383/j.aas.2015.c140185

Xiao TJ, Xu YC, Yang KY, et al., 2015. The application of two-level attention models in deep convolutional neural network for fine-grained image classification. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.842-850. https://doi.org/10.1109/CVPR.2015.7298685

Zana F, Klein JC, 2001. Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. *IEEE Trans Imag Process*, 10(7):1010-1019. https://doi.org/10.1109/83.931095

Zhao YQ, Wang XH, Wang XF, et al., 2014. Retinal vessels segmentation based on level set and region growing. *Patt Recog*, 47(7):2437-2446. https://doi.org/10.1016/j.patcog.2014.01.006

Zhu CZ, Zou BJ, Xiang Y, et al., 2015. A survey of retinal vessel segmentation in fundus images. *J Comput Aided Des Comput Graph*, 27(11):2046-2057 (in Chinese).

Zhu CZ, Zou BJ, Xiang Y, et al., 2016. An ensemble retinal vessel segmentation based on supervised learning in fundus images. *Chin J Electron*, 25(3):503-511. https://doi.org/10.1049/cje.2016.05.016