Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/46/20241190

Comparison of deep learning-based models for retinal vessels segmentation

Zhuorui Bi

School of Electronic Information Engineering, China University of Petroleum Beijing, Beijing, 100000, China

2020011640@student.cup.edu.cn

Abstract. With the trend of social diversification and the rapid popularization of information, more and more people use mobile phones to deal with matters in life, study and work, as well as a place of entertainment in their leisure time. People use their eyes at high intensity for a long time. In addition, Low eye care attention greatly increases the chance of developing eye diseases. Such lesions are closely linked to eye diseases such as glaucoma and diabetes, and these lesions are the main cause of blindness. The retina has a complex structure, diverse shapes, and high curvature. It is time-consuming and difficult to diagnose diseases from retinal images. Doctors' subjective judgments on images account for a large proportion. In order to make segmentation performance more efficient and reduce manual segmentation errors, an intelligent system that can automatically extract retinal blood vessels from fundus images has become an urgent problem. Medical picture segmentation and the merging of machine learning have gained popularity as a study area with the growth and attention of deep learning. Various models are applied to accurate segmentation of the retina. Through computer feature extraction and sample training of images, it can be directly learned from many data samples. The development of this technology saves time and cost, and at the same time meets the medical needs in remote areas. Doctors can efficiently screen and diagnose through the vascular status of the fundus to determine diabetic retinopathy, glaucoma, microaneurysms and other eye disease fundus manifestations.

Keywords: Retinal Vessel Segmentation, Convolutional Neural Network, Deep Learning.

1. Introduction

Images of the human body's tiny vessels can be seen in the fundus retinal blood vessels. Clinical pathological characteristics such as disease and atherosclerosis are reflected by the curvature and diameter of blood vessels. Image processing is a crucial stage in the segmentation of retinal images. Subsequent image processing is directly impacted by segmentation. Particularly in relation to therapeutic applications, the study of retinal blood vessel segmentation is important practically.

Convolutional neural network is the most used structure in deep learning. Its deep structured convolutional neural network's core component is a feedforward neural network with a convolutional layer [1]. In the 1980s and 1990s, convolutional neural networks only had partial functional structures. Convolutional layer and pooling layer in the convolutional neural network are partially formed by the recognition model developed by researcher Kunihiko Fukushima. Alexander Waibel designed the first

© 2024 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/46/20241190

convolutional neural network in 1987. network and successfully applied it to handwritten digit recognition. In 1989, Yann LeCun constructed a convolutional neural network applied to computer vision problems, the initial version of LeNet. Yann LeCun built a more complete convolutional neural network LeNet-5 which based on LeNet, in 1998 and apply it to handwritten digit recognition and achieve success. The basic CNN consists of five parts. The image first passes through the input layer, which converts the information features of the image into a two-dimensional matrix. The two-dimensional matrix passes through the convolution layer, and the image features are obtained by multiplying with the convolution kernel, and then passes through the pooling layer. The model extracts feature from the feature map to effectively prevent over-fitting, and then goes through the fully connected layer to expand it into a one-dimensional vector. After another calculation, the recognition probability is obtained, and finally the result with the highest probability is output through the output layer. CNN can use the shared parameters of convolution and pooling operations as well as the characteristics of local connections to greatly reduce the number of parameters that need to be trained, thereby reducing the computational and storage overhead of the model. It may be utilized for more than only visual tasks like object identification and picture categorization, and picture segmentation, it is applicable to a variety of disciplines, including speech recognition and natural language processing [2].

As deep learning achieves results in retinal blood vessel segmentation, problems have gradually emerged, include the intricacy of the fundus picture and the absence of fundus imaging data. This has prompted people to continuously optimize and adjust the model to solve different imaging problems. Optimize the evaluation criteria of the model to achieve more accurate and clearer retinal blood vessel segmentation images, allowing ophthalmologists to make more accurate judgments on lesions. Based on the above appeal, this article evaluates and contrasts the most recent models for segmenting retinal blood vessels from images. Their respective advantages and characteristics provide doctors with a theoretical basis for further treatment, and assist in the early detection and diagnosis of clinical fundus lesions.

2. Models

2.1. U-Net

The U-Net network is simple and symmetrical in structure, showing a "U" shape, consisting of convolution and pooling units [3]. The left half is the down-sampling stage of the encoder, which is used to extract image features. The right half is the up-sampling stage of the decoder, which is used to reproduce the features and output feature results. Shallow features and deep features are spliced through the concatenate operation. Features are spliced. A normalizing layer comes after each module's convolutional layer,whose role is to adjust the data distribution in the network to prevent overfitting. The numbers of convolution kernels in the convolutional layer in each module are 32, 64, 128, 256, and 512 respectively. By cutting the size of the input features in half while maintaining the channels' original dimensions, the pooling layer is used to minimize the dimensionality of the features. After passing through the encoder, The input image's high-level feature vector is retrieved once it has gone through the encoder, and it is then given to the decoder. The encoder part uses the ReLU function.

Features: The training data of Patch is much larger than the number of training images, this addresses the issue of the lack of training data for medical images; the U-Net structure can combine the bottom and top information of the network to obtain low-resolution target segmentation information after multiple downsampling. This This kind of target information is the relationship feature between image information and image environment. This feature can reduce the difficulty of judging the segmentation target and background. As far as retinal blood vessel segmentation is concerned, blood vessel features can be more effectively extracted from the tissue fluid background of fundus photos. As for the shallow information of the higher layer, the input terminal is transferred to the output terminal of the same layer through the concatenate operation, which generates more accurate feature information, improves the accuracy, and solves the problem of certain noise in the blurred parts of the medical image.

DOI: 10.54254/2755-2721/46/20241190

2.2. DUNet

The DUNet model adopts a "U"-shaped structure similar to the U-Net network. It is also called a deformable U-Net network [4]. It also has a bilaterally symmetric encoder and decoder, but its decoding stage has a large number of feature channels. The upsampling operator is stacked symmetrically with the traditional CNN, which may more effectively gather contextual data and transmit it to higher resolution layers for more precise segmentation. The encoder and decoder sections stack deformable convolution blocks to record geometric alterations, which reflects the deformability of the object, achieving adaptive receptive field modification in accordance with the size and shape of the object, that is, different sizes and shapes can pass through Better identification of intricate vascular structures is made possible by deformed receptive field capture. The deformable convolution module has an offset at each sampling position to generate a new sampling position. The obtained conversion offset is batch normalized according to conditions, and then passes through the activation layer (ReLU). Compared with U-Net model, DUNet achieves more accurate segmentation and detection of more complex vascular structures through more feature channels and added deformable convolution modules.

2.3. FR-Net

FR-Net stands for "Full Resolution Network". Images of blood vessels are distinct from those of the heart and of cells [5]. Particularly the capillaries of the fundus blood vessels, blood vessels are particularly thin, have a low pixel ratio, and have low contrast.Based on the U-Net network, after ordinary downsampling, the valuable spatial information is reduced. It is difficult to learn information about small-scale blood vessels. FR-Net consists of upsampling, downsampling, feature aggregation module, and residual block. This model combines the idea that each layer in U-Net++ is closely connected with the subsequent layer, and performs horizontal summation through upsampling, downsampling, and convolution. Vertically expand, remove the tight connections of each layer, propose an interaction mechanism for multiple resolutions of convolution, Fusing the upsampling and downsampling feature maps of subsequent stages after adding a feature aggregation module, mapping the features of the preceding residual module. The data are then combined to achieve full resolution, expanding the receptive field and completely completing the context information. The FR-Net model achieves dense detection by introducing a residual module and a feature aggregation module, especially for microscopic targets, such as pixel-thin blood veins, that lose spatial information after downsampling. At the same time, accurate pixel segmentation is accomplished.

2.4. CE-Net

CE-Net is also called context encoder network. For many U-Net models and many other optimization models, convolution and pooling processes will be carried out continuously [6]. This operation will cause the loss of some geographic context and may be false during retinal diagnosis. Context extractor module, feature encoder module, and positive misjudgment make up the three primary components of CE-Net. Feature decoder module. The innovation of CE-Net is to replace the original U-Net encoder module with ResNet-34 blocks trained from ImageNet. The step of processing images is to first input the image into a feature encoder module, through a residual multi-kernel convolution (RMP) block and a dense convolution-free (DAC) block generates higher-level semantic feature maps, and finally, the extracted feature information is processed in the feature decoder. Compared with the original module,it added a shortcut mechanism, which uses pre-trained ResNet blocks, avoids the disappearance of gradients, and accelerates the network convergence speed. Payback for information loss brought on by strung-together convolution procedures and continuous pooling, the DAC module bypasses the data acquired in the connection and decoder. The CE-Net model integrates dense convolution-free modules and residual multi-core pooling into the ResetNet-improved U-Net network,make it feasible to store more spatial information and high-level feature information.

DOI: 10.54254/2755-2721/46/20241190

2.5. SA-UNet

A simple network for spatial attention is called SA-UNet. The introduced spatial attention module multiplies the input feature map and the attention map to perform adaptive feature refinement, thereby inferring along the spatial dimension [7]. The enhanced and optimized U-Net network in retinal blood vessel segmentation will unavoidably result in a network that is more complicated and challenging to understand. SA-UNet uses a structured dropout convolution block combined with DropBlock. The specific structure includes a U-shaped symmetrical decoder-encoder structure. A structured dropout convolution block and a maximum pooling operation are included in each encoder phase. A batch reduction and a DropBlock block come after each convolutional layer. The encoder consists of a series of modules with similar structures. Each module contains a DropBlock block, a batch normalization layer and a linear rectification unit. After the encoder encodes the input, it is decoded by the decoder, and the two are connected by a spatial attention module, which includes maximum pooling and average pooling operations. The decoder comprises an upsampling transposed convolution at each stage, half more feature channels are added before being linked to the encoder's feature map, followed by a dropout convolution block fused with Dropout, which aims to discard adjacent areas in the feature map to regularize function to prevent overfitting of the network. A lightweight network proposed by SA-UNet combined with DropBlock can effectively prevent network overfitting and data enhancement. Even small sample data can be well trained, especially for small blood vessel areas and edges in fundus vascular imaging. Vascular area.

2.6. RV-GAN Model

Encoder and decoder are used by the U-Net network structure to extract feature pictures, however the inability to recover the lost blood vessel feature images in the decoding stage and the loss of resolution in the encoding stage are problems that must be solved [8]. Autoencoding segmentation methods have limited ability to extract retinal vascular structures. In order to better realize the segmentation of microvessels and the localization of specific lesions, the RV-GAN architecture proposes using two multi-scale autoencoding discriminators Df, Dc and their corresponding two generators Gf, Gc, each with its own upsampling block and downsampling block. The generator Gc is responsible for extracting features and the generator Gf is responsible for learning features. For adversarial training, the two generators are paired with two corresponding discriminators Df and Dc.The generators are coarse generators and fine generators respectively, it has the ability to retrieve local data such micro branches, linkages, and obstructions. Synthesize fine-grained blood vessel segmentation images, then extract features from the discriminator, and combine loss due to feature matching in semantic segmentation. RV-GAN also proposes a new weighted feature matching loss, which gives more weight to the decoder feature map. Adversarial training of architectures. RV-GAN achieves high-fidelity and high-precision segmentation of retinal macro and microvascular structures, avoiding false positive segmentation in subspace.

2.7. ET-Net

ET-Net is called edge-preserving guidance network. Previous models generally carry out iterative enhancement for target segmentation, but ET-Net embeds edge detection methods to make up for the shortcomings of target segmentation [9]. Commonly used segmentation methods for medical images include edge detection and target segmentation. Although the edge detection method mainly focuses on local edge information, resulting in discontinuous segmentation boundaries, it can obtain important information about the shape of the object and reduce the amount of information to be processed. Protecting the boundary structure of microvessels, ET-Net integrates edge detection and target segmentation. The segmentation process relies heavily on the guidance of edge detection. ET-Net contains the U-Net infrastructure of encoder-decoder, and towards the conclusion, a weighted aggregation module and an edge guidance module are introduced. Edge detection is used by the edge segmentation module to keep an eye on the first convolutional layers while the edge information is used to avoid excessive divergence when extracting feature information. The introduction of a weighted

DOI: 10.54254/2755-2721/46/20241190

aggregation module enables all high-level characteristics, which facilitates object recovery and facilitates representation, guiding initial segmentation. ET-Net gives full play to the respective advantages of edge detection and object detection, highlights valuable features of higher-level layers, using multi-scale evaluation of edge information and contextual features to obtain information, improving segmentation accuracy, especially for the retina. Edge information in imaging.

3. Result

The fundus retinal image data sets in this article all come from DRIVE. Currently, most relevant studies are experimentally verified from this data set. The photos came from the Netherlands' Diabetic Retinopathy Screening Program. 400 diabetic participants in the study, who ranged in age from 25 to 90, were included, and it consisted of a total of 40 JPEG color fundus images; single-channel images have better vascular contrast than RGB images, so all color fundus photos were converted These are single-channel photos, including 7 cases of pathological abnormalities. Each of the 40 images includes 20 training sets and 20 test sets. The test models are all from the above segmented images [10].

Retinal blood vessel segmentation evaluation indicators: Use four commonly used indicators to objectively analyze retinal segmentation performance: ROC curve area (AUC), analyze the quality of the model, binary classification model accuracy (F1-score), accuracy (ACC), and sensitivity (Sensitive).

Table 1. Comparison of evaluation indicators of each model.

	AUC	F1-score	Accuracy	Sensitive
U-Net	0.9755	0.8142	-	-
DUNet	0.9802	0.8237	-	-
CE-Net	0.9779	-	0.9545	-
FR-Net	0.9889	0.8316	0.9705	0.8356
ET-Net	-	-	0.956	0.7744
SA-UNet	0.9864	0.8263	0.9698	-
RV-GAN	0.9887	0.8690	0.9790	0.7929

From the analysis of the above table, FR-Net has the best AUC value and Sensitive value, and the best authenticity and sensitivity of the detection method. This is related to its increased receptive field and the introduction of residual convolution module and feature convolution module to achieve dense features. For detection, feature maps fully complement relevant contextual information. RV-GAN has the best F1-score value and Accurary value, indicating that the model has higher accuracy and detection accuracy. Replacing the encoder and decoder with a generator and a discriminator ensures high-precision segmentation of microvascular structures. Thus, false positive segmentation is avoided to a large extent. At the same time, RV-GAN has good detection method authenticity, which is only 0.0002 different from FR-Net, and FR-Net also has good accuracy. For other networks, the performance index of SA-UNet is second only to FR-Net, followed by DU-Net and CE-Net, and finally ET-Net and U-Net.

4. Conclusion

In this article, the basic architecture of the U-Net network is first introduced and discussed, combined with the different shortcomings of the U-Net network, introduced other optimized networks, including their basic architecture, innovatively introduced modules, and specific solutions. Specific problems in retinal blood vessel segmentation are identified and each model is applied to the DRIVE data set test. This work comprehensively compared four main data indicators and found that the FR-Net and RV-GAN models have better indicators, while the most basic All indicators of the U-Net structure rank last. The U-Net network's encoder, decoder, and concatenate structure continue to have several issues. Focusing on the FR-Net network and RV-GAN network, which have two outstanding performance indicators, and the SA-UNet network with excellent comprehensive performance, the network has the following characteristics: it introduces a multi-resolution convolution interaction mechanism to realize

DOI: 10.54254/2755-2721/46/20241190

information interaction in adjacent stages and supplements advanced contextual information; it introduces a generative adversarial network structure to give the decoder feature map better weight, weighted target loss and adversarial property, using loss to obtain more realistic microvessels, arteries, and blood vessel structures; introducing a spatial attention module and a structured dropout layer for adaptive feature refinement. The three add specific modules to make full use of the very small amount of data, make up for all the shortcomings that may cause information loss before output, and maximize the display of retinal blood vessel characteristics.

Medical image analysis is largely based on medical image segmentation results. It is suggested that the U-Net network structure advance the process of medical picture segmentation.The iterative optimization of various models based on the basic ideas of U-Net is for better application in medical treatment. middle. At present, the accuracy and precision of image segmentation have been greatly improved, which has avoided false positive misdiagnosis by ophthalmologists to a certain extent. It is hoped that with the advancement of science and technology, lightweight networks with higher accuracy and precision can be realized. There are better models in 3D retinal blood vessel segmentation to achieve better segmentation results.

References

[1] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., et al. (2018). Recent advances in convolutional neural networks. Pattern recognition, 77, 354-377.

[2] Macukow, B. (2016). Neural networks–state of art, brief history, basic models and architecture. In Computer Information Systems and Industrial Management: 15th IFIP TC8 International Conference, CISIM 2016, 15, 3-14.

[3] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention, 3(18), 234-241.

[4] Jin, Q., Meng, Z., Pham, T. D., Chen, Q., Wei, L., & Su, R. (2019). DUNet: A deformable network for retinal vessel segmentation. Knowledge-Based Systems, 178, 149-162.

[5] Liu, W., Yang, H., Tian, T., Cao, Z., Pan, X., Xu, W., et al. (2022). Full-resolution network and dual-threshold iteration for retinal vessel and coronary angiograph segmentation. IEEE Journal of Biomedical and Health Informatics, 26(9), 4623-4634.

[6] Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., et al. (2019). Ce-net: Context encoder network for 2d medical image segmentation. IEEE transactions on medical imaging, 38(10), 2281-2292.

[7] Guo, C., Szemenyei, M., Yi, Y., Wang, W., Chen, B., & Fan, C. (2021). Sa-unet: Spatial attention u-net for retinal vessel segmentation. In 2020 25th international conference on pattern recognition, 1236-1242.

[8] Kamran, S. A., Hossain, K. F., Tavakkoli, A., Zuckerbrod, S. L., Sanders, K. M., & Baker, S. A. (2021). RV-GAN: Segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial network. In Medical Image Computing and Computer Assisted Intervention, 8(24), 34-44.

[9] Zhang, Z., Fu, H., Dai, H., Shen, J., Pang, Y., & Shao, L. (2019). Et-net: A generic edge-attention guidance network for medical image segmentation. In Medical Image Computing and Computer Assisted Intervention, 1(22), 442-450.

[10] DRIVE: Digital Retinal Images for Vessel Extraction, URL: https://drive.grand-challenge.org/, Last Accessed 2023/09/15.