Microaneurysms segmentation with a U-Net based on recurrent residual convolutional neural network

Caixia Kou
Wei Li
Wei Liang
Zekuan Yu
Jianchen Hao

Microaneurysms segmentation with a U-Net based on recurrent residual convolutional neural network

Caixia Kou,^{a,*} Wei Li,^a Wei Liang,^a Zekuan Yu,^{b,*} and Jianchen Hao^c

^aBeijing University of Posts and Telecommunications, Haidian District, Beijing, China

^bPeking University, Haidian District, Beijing, China

^cPeking University First Hospital, Xicheng District, Beijing, China

Abstract. Microaneurysms (MAs) play an important role in the diagnosis of clinical diabetic retinopathy at the early stage. Annotation of MAs manually by experts is laborious and so it is essential to develop automatic segmentation methods. Automatic MA segmentation remains a challenging task mainly due to the low local contrast of the image and the small size of MAs. A deep learning-based method called U-Net has become one of the most popular methods for the medical image segmentation task. We propose an architecture for U-Net, named deep recurrent U-Net (DRU-Net), obtained by combining the deep residual model and recurrent convolutional operations into U-Net. In the MA segmentation task, DRU-Net can accumulate effective features much better than the typical U-Net. The proposed method is evaluated on two publicly available datasets: E-Ophtha and IDRiD. Our results show that the proposed DRU-Net achieves the best performance with 0.9999 accuracy value and 0.9943 area under curve (AUC) value on the E-Ophtha dataset. And on the IDRiD dataset, it has achieved 0.987 AUC value (to our knowledge, this is the first result of segmenting MAs on this dataset). Compared with other methods, such as U-Net, FCNN, and ResU-Net, our architecture (DRU-Net) achieves state-of-the-art performance. © 2019 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.6.2.025008]

Keywords: U-Net; microaneurysms; segmentation; deep recurrent U-Net.

Paper 18240R received Nov. 18, 2018; accepted for publication May 31, 2019; published online Jun. 19, 2019.

1 Introduction

Diabetes is one of the most serious diseases all over the world since it can lead to severe complications. Among the complications, diabetic retinopathy (DR) is so common that it is the leading cause of blindness in working-age people.¹ DR can be generally classified into two stages: nonproliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR).² Patients may not have any symptoms of vision problems at the early stage of DR, but as the disease develops it can cause vision loss or blindness, so early detection of DR is very important to save patients' vision. Early detection of DR can prevent 90% of cases from resulting in blindness.³

Microaneurysms (MAs) are the earliest signs of DR, which are visible to ophthalmologists. MAs are small swellings in the retina's tiny blood vessels showing as tiny round and red spots on the retina, as shown in Fig. 1. However, it is a time-consuming work for ophthalmologists to identify MAs manually due to the limited number of ophthalmologists and the large number of people who require screening. Hence, developing an automatic MA segmentation method has become a necessity. Automatic MA segmentation is a challenging task because MAs are small in size, sometimes only a few pixels. Furthermore, the boundaries of MAs are not always well defined and the local contrast of the background is low, even in high-resolution images. MAs may be confused with visually similar fundus structures such as hemorrhages (HEs) and junctions in thin vessels.⁴

MAs are the earliest clinical signs of DR, so a considerable amount of research has been done on MAs.⁵ Many MA

segmentation algorithms are divided into unsupervised learning and supervised learning methods.

Unsupervised learning methods without a training process include morphological processing, wavelet transformation, and template matching. Spencer et al.⁶ have proposed a MA segmentation algorithm in fluorescein angiograms. They have employed a series of morphological operations that remove the vasculature and leave the other small structures representing the MAs. The ceiling of sensitivity reaches 0.82 and matches that of the clinicians' results. Quéllec et al.⁷ have proposed an adaptive wavelet method that can improve the performance of wavelet filtering and optimize the pattern-matching parameters of the frequency domain. Lazar and Hajdu⁸ have applied the analysis of directional cross-section profiles to detect MAs. A score value is assigned to each pixel in the image, based on the diameter and roundness, and then the actual MAs are obtained based on the threshold. The method is ranked at the fifth place among the ten best participants in the Retinopathy Online Challenge.

Supervised learning-based methods are usually used to extract the different kinds of features such as texture, color, geography, and size of the MA regions. Then the feature vectors are applied into different classifiers to train, which can distinguish MA regions and non-MA regions. Most supervised learning MA segmentation methods are based on pixel value. A double-loop filter is used to extract candidate regions of MAs,⁹ and then the candidate lesions are classified into MAs or false positives using a rule-based method and an artificial neural network. Akram et al.¹⁰ have proposed a hybrid classifier integrating a Gaussian mixture model, a support vector machine (SVM), and an extension of multimodel media to improve the

*Address all correspondence to Caixia Kou, E-mail: koucx@lsec.cc.ac.cn; Zekuan Yu, E-mail: yuzekuan518@163.com

Fig. 1 Fundus image and corresponding MA segmentation result.

accuracy of classification. In the 219 images of DIARETDB0 and DIARETDB1, this method has a sensitivity value of 0.9864, a specificity value of 0.9969, and an accuracy value of 0.9940. Srivastava et al.¹¹ have used filters based on the Frangi filter which is manually designed to distinguish between red lesions and blood vessels. Filters are applied to multiple-sized image patches to extract features and these features are classified by a SVM. Experiments on the datasets of DIARETDB1 and MESSIDOR show this method to have a 0.97 area under curve (AUC) value for MA detection.

Recently, deep learning methods, especially convolutional neural networks (CNNs) have achieved great success in many image tasks, such as classification, segmentation, and object detection.^{12–14} For MA segmentation, Haloi et al.¹⁵ have used a deep neural network with three convolutional layers and two fully connected layers, achieving 0.98 AUC value on ROC dataset. This method without manual feature extraction is a major improvement over other methods. Compared with Haloi method,¹⁵ Chudzik et al.¹⁶ have applied a fully convolutional neural network (FCNN) architecture with batch normalization (BN)¹⁷ layers and Dice coefficient loss function. This FCNN is used to divide pixels on small patches into MAs or non-MAs. Also a fine-tuning scheme is proposed to reduce training time. A 10-layer CNN has been proposed by Tan et al.¹⁸ that can automatically segment MAs based on the probabilities of output. Dai et al.¹⁹ have proposed a multisieving convolutional neural network (MS-CNN) which applies multiple classifiers to construct training data. The MS-CNN uses the original fundus images and the rough segmentation maps generated by a weak image-to-text mapping model as input to generate the final high-quality segmentation. The framework achieves 0.997 precision value and 0.878 recall value on DIARETDB dataset.

U-Net,²⁰ one of a FCNN architecture¹³ for image segmentation that accepts an image as an input and returns segmentation probability distribution as an output, has become a mainstream medical image segmentation method because of its huge success in blood vessel, optic disc, and exudates biomedical image segmentation. Two efficient U-Net-type networks have been presented recently. One proposed by Feng et al.²¹ is obtained by combining residual blocks into U-Net to segment optic disc and exudates, achieving the high F-score of 0.9093 and 0.8150, respectively. The other presented by Alom et al.²² is a R2U-Net which combines recurrent convolutional operations with U-Net to segment the blood vessel. The R2U-Net achieves 0.9914 AUC value on the STARE dataset, which is a state-of-the-art result. Inspired by these two modified U-Nets with the state-of-the-art performance, we apply recurrent residual

convolutional neural network (RRCNN) blocks into U-Net to further improve its performance in a MA segmentation task.

Specifically, in this paper, we use the RRCNN block, which includes a basic block and two recurrent convolutional layer (RCL) units instead of regular convolutional layers on U-Net. The obtained modified U-Net is used to segment the MAs. The rest of this paper is organized as follows: Sec. 2 proposes methodology, Sec. 3 deals with experiments, Sec. 4 demonstrates the discussion, and Sec. 5 concludes the work.

2 Proposed Methodology

In this section, first we give a brief review to the typical U-Net architecture. Then, we introduce our modified U-Net.

2.1 Typical U-Net Architecture

The U-Net²⁰ architecture is based on FCNN which does not use the full connected layer and allows images to be used as input and returns binary maps as output. As shown in Fig. 2, the U-Net consists of a contracting (encoding) path and a symmetric expanding (decoding) path. In the contracting path, successive convolution layers are followed by pooling operations. In the expanding path, pooling operators are replaced by upsampling operators. The upsampled output combined with high-resolution features from the contracting path can supplement the information lost in the pooling process. Because of its special structure, the U-Net shows satisfactory performance in biomedical image segmentation.

2.2 Modified U-Net Based on RRCNN Block

In this subsection, we introduce our proposed deep recurrent U-Net (DRU-Net), which is an efficient deeper U-Net model combining the classical features of the deep residual model²³ and the recurrent convolutional operation.²⁴ Fundus image pre-processing, an important step in the segmentation of MAs, is also presented in this subsection.

2.2.1 DRU-Net architecture

The architecture of DRU-Net is to extend the U-Net by replacing regular convolutional layers with the RRCNN blocks. It consists of an encoding path (left side) and a decoding path (right side), as shown in Fig. 3. Obviously, the DRU-Net allows each pixel in the input image to extract its own features using the convolutional layers. There are seven RRCNN blocks and two basic blocks in DRU-Net. Then, each RRCNN block is followed by 2×2 max-pooling in the encoding path. After the encoding process, the number of feature maps increases from 1 to 256. The decoding path applies 2×2 upsampling operation to RRCNN blocks, which doubles the size of the feature maps. Every upsampled output combines with high-resolution features in the encoding path. Finally, 1×1 convolution and sigmoid activation functions are used to get a segmentation probability distribution map.

In the whole framework process, a basic block, as shown in Fig. 4(a), includes a 3×3 convolution followed by a BN¹⁷ layer and a rectified linear layer. The typical residual block, as shown in Fig. 4(b), consists of a shortcut and a few stacked layers: convolutional layers and rectified linear unit (ReLU) layers. In the DRU-Net, we modify the typical residual block to the RRCNN block, as shown in Fig. 4(c). The RRCNN blocks, which implement recurrent convolutional operations, include two RCL units

Fig. 2 Typical U-Net architecture.

Fig. 3 The DRU-Net architecture based on typical U-Net architecture with RRCNN block.

and a basic block on the shortcut. We apply the RCL²⁴ units, instead of stacked layers, in order to incorporate recurrent connections into each convolutional layer. The unfolded RCL unit is shown in Fig. 4(c).

The unfolded RCL for t time steps is a feed-forward subnetwork of depth $t + 1$. And t is the number of the recurrent convolutional operations. For example, $t = 2$ leads to a feed-forward subnetwork with the largest depth of 3 and the smallest depth of 1. When $t = 0$, only feed-forward computation takes place. The RCL unit can be explained mathematically according to the recurrent CNN.²⁴ For a pixel at (i, j) on the k 'th feature map in RCL unit, its net input is defined as

$$z_{ijk}(t) = (w_k^f)^T * u^{f(i,j)}(t) + (w_k^r)^T * v^{r(i,j)}(t-1) + b_k, \quad (1)$$

where $u^{f(i,j)}(t)$ and $v^{r(i,j)}(t-1)$ are the feed-forward convolutional and the recurrent inputs, respectively. And w_k^f and w_k^r are the weights of the convolutional layer and the recurrent layer. In addition, b_k is the bias. When the recurrent input evolves over iterations, the feed-forward convolutional input remains the same in all iterations. Its net input is followed by ReLU activation function and a BN. The ReLU activation function is defined as

$$\phi(x) = \max(0, x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}. \quad (2)$$

Each RRCNN block has two RCL units and the last output of RCL unit passes through the residual block. Moreover, in order

Fig. 4 Proposed RRCNN block for retinal image segmentation. (a) Basic block, (b) typical residual block, and (c) RRCNN block and the unfolding RCL unit.

to make the input and output channels of the RRCNN block the same when adding elements, we add an extra basic block with 3×3 convolution to the shortcut. Consider x_t as the input of the RRCNN block. Denote the output of the RRCNN block as x_{t+1} and it can be expressed as

$$x_{t+1} = O(x_t) + F(x_t, w_t), \quad (3)$$

where x_t is the input of the RRCNN block, $O(x_t)$ represents the output on shortcut, and $F(x_t, w_t)$ is the output of the last RCL unit. Then, the x_{t+1} is used to feed into downsampling and upsampling layers in the encoding and decoding convolutional units of DRU-Net.

The RRCNN block helps to develop a more efficient deeper model. The RCL unit with recurrent connections improves the effectiveness of feature accumulation in the segmentation algorithm based on CNN for medical imaging. Furthermore, the effectiveness of feature accumulation achieves precise segmentation and the DRU-Net has better performance than the typical U-Net model.²⁰

For training the network, we use a class-balancing loss function²⁵ proposed for contour detection in natural images, which is defined as

$$L(W, \tilde{y}_i) = -\beta \sum_{i \in Y_+} \log(\tilde{y}_i) - (1 - \beta) \sum_{i \in Y_-} \log(1 - \tilde{y}_i), \quad (4)$$

where Y_+ and Y_- represent the truth labels of Y corresponding to the background and foreground sets of the ground, respectively. In addition, we use $\beta = |Y_-|/|Y|$ and $1 - \beta = |Y_+|/|Y|$ for balancing the large number of MAs and non-MAs. And $\tilde{y}_i$ is the predicted probability value obtained by the last sigmoid layer.

2.2.2 Preprocessing

Original fundus images are preprocessed in order to get better image quality for training. Because the green channel has the best contrast, the green channel of RGB fundus image is used for further processing. When the contrast of the captured image is too low, it is difficult to detect and isolate objects of interest. Therefore, image enhancement is a very important step before the learning process. The histogram equalization and contrast-limited adaptive histogram equalization (CLAHE) are two classical approaches. To achieve the most obvious contrast images, we develop a weight adjustment method which

combines histogram equalization and CLAHE. In addition, a data normalization method is used to eliminate the influence of other transformation functions on image transformation. All the preprocessing results are shown in Fig. 5.

3 Experiments and Results

In this section, we first describe the details of the public retinal image dataset and the evaluation metrics are also introduced. To test the numerical performance of DRU-Net, three other MA segmentation methods, namely, FCNN, typical U-Net, and ResU-Net, are also implemented for comparison. In addition, we demonstrate the effect of the basic block and the different numbers of RCL units in DRU-Net architecture through several numerical experiments.

3.1 Dataset

E-Ophtha²⁶ dataset is a publicly accessible digital retinal images dataset, including 463 color fundus images, and is used for scientific research in lesion segmentation. The images of this dataset are captured within the OPHDIAT telemedical network for DR screening in the framework of the TeleOphta project. Lesions are carefully contoured by an ophthalmologist using software developed by ADCIS. These annotations are afterward checked by a second ophthalmologist, which is precise enough for segmentation evaluation. E-Ophtha dataset consists of two subsets: E-Ophtha-EX (exudates) and E-Ophtha-MA (microaneurysms). In our experiments, we use E-Ophtha-MA dataset for MA segmentation. It contains 233 images with no lesion (i.e., healthy) and 148 with MAs. The images have four different image sizes of 1440×960 to 2544×1696 pixels. To acquire consistent resolutions, all the images are resized into 1024×1024 pixels. Our training data consist of 80% of the MA dataset and 80% of the healthy dataset, and the remaining images are used for testing.

The Indian Diabetic Retinopathy Image Dataset (IDRiD)²⁷ is a retinal fundus image dataset, which is available for the segmentation and grading challenge of DR, organized by International Symposium on Biomedical Imaging (ISBI) Conference 2018. The fundus images in IDRiD are captured by a retinal specialist at an Eye Clinic located in Nanded, Maharashtra, India. The pixel-level annotation is done by a master's student using software developed by ADCIS and then is reviewed by two retinal specialists, which is similar to the E-Ophtha dataset. IDRiD dataset provides precise pixel-level

Fig. 5 (a) Original image, (b) green channel of RGB-colored image, (c) green channel image after CLAHE processing, (d) green channel image after histogram equalization processing, and (e) average mean of (c) and (d).

annotation of abnormalities associated with DR, such as MAs, soft exudates (SE), hard exudates (EX), and HEs, which is invaluable resource for performance evaluation of individual lesion segmentation techniques. The dataset consists of 81 color fundus images with signs of DR and pixel-level annotations of EX, HE, MA, and SE. The images have a resolution of 4288×2848 pixels with 50-deg field of view. The 81 color fundus images and annotations of MAs are used in our experiments. IDRiD provides a partition of training and testing sets, with 54 images for training and the remaining 27 images for testing.

3.2 Evaluation Metrics

To evaluate the segmentation task, the following numbers of samples are exactly calculated. True positive (TP) is the number of positive pixels correctly classified by the classifier. True negative (TN) represents the number of negative pixels correctly classified as negative. False positive (FP) and false negative (FN) are the number of negative pixels and positive pixels that are misclassified, respectively. To set a better representation, the accuracy (Acc), sensitivity (Se), and specificity (Sp) are defined as follows: $\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$, $\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$, $\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$.

To convert the last probability map into a binary image, a threshold is defined by seeking the maximum value of $\text{Se} + \text{Sp} - 1$, and then this threshold is used to get the final binary results. Moreover, we also use the receiver operating characteristics (ROC) curve and the value of the AUC, which are the common evaluation measures for medical image segmentation tasks.

3.3 Implementation Details

All the experiments were implemented on the publicly available Keras framework with TensorFlow in Python 2.7 version and were performed on a single GPU machine with 256G of RAM

and an NVIDIA GeForce GTX 1080Ti. All the MA segmentation methods were trained using the Adam algorithm (momentum = 0.9, weight decay = 0.0005, initial learning rate = 0.001), which was considered to be suitable for the gradient descent method. Each model was trained with 150 epochs and when the loss function tended to a stable value, each epoch took ~ 20 min. To reduce the memory shortage caused by the excessive input image resolution, all the training images were divided into 381,600 patches with 48×48 pixels for training, and 75,600 patches for validation. Dividing into smaller patches also solved the imbalance problem in MA and non-MA pixels.

3.4 Performance of the Proposed Method

3.4.1 Qualitative results

The MA segmentation result is a probability map in which each pixel is a probability value. The threshold scheme is used to obtain a binary MA segmentation map which represents whether a pixel belongs to the MA class or not. The visual results based on E-Ophtha and IDRiD produced by our proposed model are shown in Figs. 6 and 7. When the number of MAs is very small, almost all of them can be segmented accurately. In the two datasets, the vast majority of MAs can be distinguished.

Compared to other segmentation tasks, the accurate segmentation of MAs using the proposed method in the case of different challenges on E-Ophtha dataset is shown in Fig. 8. In the case of the low local contrast of the background, as shown in Fig. 8(a), it is observed that our method can classify the MAs with high accuracy. All the MAs are segmented and the probability value of MAs is close to one using our method. When there exist some structures similar to MAs, as shown in Fig. 8(b), the similar structure can be accurately distinguished from MAs. In Fig. 8(c), the boundary of the MAs may be close to the blood vessels which makes the task of segmenting the MAs more challenging. The morphology of MAs and small blood vessels are

Fig. 6 MA segmentation results on E-Ophtha dataset. (a) Retinal image, (b) manual segmentation result as ground truth, (c) probability map of segmentation result, and (d) final segmentation results (white and black pixels denote the MA and non-MA regions, respectively).

very close to the background, as a result the MAs can be identified accurately based on the DRU-Net. In all, the above results show that our method of MA segmentation is robust with high segmentation accuracy.

3.4.2 Comparison with other methods

We also compare the performance of the segmentation using DRU-Net with other algorithms that have achieved state-of-the-art results on various segmentation tasks. We train the model of typical U-Net,²⁰ FCNN,¹³ and ResU-Net²¹ to segment the MA dataset which has good performance on the vessel image segmentation task. The visual results of the three models and DRU-Net on E-Ophtha are shown in Fig. 9.

It is can be observed that DRU-Net outperforms the other three methods in Fig. 9. By comparing Figs. 9(a) and 9(e), the results of FCNN segmentation can be observed to be undersegmented, and in some regions MAs are not segmented at all or only a few pixels of the MAs can be segmented. From Figs. 9(b) and 9(e), we can know that the results of typical U-Net are

oversegmented and the typical U-Net segments non-MAs into MAs which makes the results of segmentation inaccurate. In addition, as seen in Figs. 9(c) and 9(e), the ResU-Net segmentation is not stable and has no robustness. However, the DRU-Net method can segment accurately and all the MAs can be segmented, showing stronger performance than the other three methods.

In the following tables, we compared the proposed DRU-Net with other MA segmentation methods on various evaluation metrics. Among the comparisons, the best results are denoted in bold. In Table 1, the performance of four MA segmentation results on E-Ophtha dataset is listed. Different from Se, Sp, and Acc, AUC is independent of the threshold, and so we compare the AUC value of different models. The testing results show that the proposed DRU-Net model provides a higher AUC value of 0.9943. The performance of four MA segmentation results on IDRid is listed in Table 2. The results illustrate that the proposed RCL unit indeed contributes to the U-Net architecture and outperforms other U-Net models for MA segmentation. Furthermore, compared with other methods,^{28,29} which

Fig. 7 MA segmentation results on IDRiD dataset. (a) Retinal image, (b) manual segmentation result as ground truth, (c) probability map of segmentation result, and (d) final segmentation results (white and black pixels denote the MA and non-MA regions, respectively).

Fig. 8 (a) Segmentation results with low contrast, (b) a similar structure can accurately distinguish whether it belongs to MAs, and (c) segmentation results when the boundary of the MAs is close to the blood vessels.

Fig. 9 MA segmentation results on E-Ophtha dataset. (a) FCNN segmentation results, (b) typical U-Net segmentation results, (c) ResU-Net segmentation results, (d) DRU-Net segmentation results, and (e) ground truth.

Table 1 Performance of various models on E-Ophtha dataset.

Method	Se	Sp	Acc	AUC
FCNN	0.628	0.856	0.913	0.704
U-Net	0.717	0.943	0.997	0.838
ResU-Net	0.943	0.967	0.999	0.990
DRU-Net	0.964	0.963	0.999	0.994

Table 2 Performance of various models on IDRiD dataset.

Method	Se	Sp	Acc	AUC
FCNN	0.635	0.786	0.945	0.790
U-Net	0.732	0.643	0.924	0.761
ResU-Net	0.919	0.913	0.999	0.974
DRU-Net	0.930	0.936	0.999	0.982

have achieved good AUC values for MA segmentation on E-Ophtha dataset, our method achieves a better result, as shown in Table 3. Tables 1 and 3 show that the proposed DRU-Net model has better performance in the terms of the value of AUC and accuracy.

Table 3 Performance comparison of MA segmentation methods on E-Ophtha images.

Method	Se	Sp	Acc	AUC
Lam et al. ²⁸	—	—	—	0.940
Costa et al. ²⁹	—	—	—	0.971
DRU-Net	0.964	0.963	0.999	0.994

Table 4 The computational time for testing phase on E-Ophtha dataset.

Model	Time (s)/image
U-Net	3.82
ResU-Net	4.84
DRU-Net	5.48

Furthermore, we calculate the time for testing per image, as shown in Table 4. The processing times during the testing phase for the U-Net, ResU-Net, and DRU-Net models on the E-Ophtha dataset are 3.82, 5.48, and 4.84 s per image, respectively. The computational time of DRU-Net for testing per image is much less than ResU-Net, but the AUC value of DRU-Net is higher

Fig. 10 ROC curve for the best performance achieved with DRU-Net on E-Ophtha and IDRiD datasets.

than ResU-Net. The time taken by DRU-Net for testing per image on IDRiD dataset is 32.16 s, due to the high resolution.

We employ the ROC curve on E-Ophtha and IDRiD in Figs. 10(a) and 10(b), which is also a common evaluation measure for medical image segmentation tasks to evaluate the performance of the proposed approaches. Furthermore, it has achieved 0.987 AUC value on IDRiD, as far as we know, which is the first result of segmenting MAs on this dataset. Experimental results show that our model can segment MAs accurately on different datasets.

4 Discussion

U-Net architecture, which has various variants such as ResU-Net²¹ and M-Net,³⁰ is the mainstream biological image segmentation method and achieves better performance than other methods. Feng et al.²¹ have proposed a modified residual U-Net, which combines the residual blocks with typical U-Net. The modified residual U-Net, which is composed of residual blocks and basic blocks, is applied to segment both optic disc and exudates. Compared to the typical U-Net and other segmentation methods, this model achieves better results with 0.9093 and 0.8150 F-score value on optic disc and exudates segmentation, respectively. Alom et al.²² have proposed R2U-Net, which includes RCL units but not basic blocks, to segment retina blood vessel. They use R2U-Net to segment blood vessel on STARE, DRIVE, and CHASE_DB1 datasets, and the results of AUC value are 0.9914, 0.9784, 0.9816, respectively. All prior research results show that the U-Net with RCL unit has better segmentation results even if on different datasets. In R2U-Net,²² RRCNN blocks are with two RCLs and without basic block. To illustrate the effect of basic block on the shortcut for the MA segmentation, experiments on E-Ophtha dataset with different RRCNN blocks are compared in this paper. As shown in Table 5, the U-Net architecture with the RRCNN block, which includes two RCL units and a basic block on the shortcut, achieves higher AUC value than that without basic block. Our model has better performance than R2U-Net, as shown in Table 5. Adding an extra basic block on shortcut can make the same number of channels to add in RRCNN block in order

Table 5 Performance of the MA segmentation with and without basic block in RRCNN block.

Method	Se	Sp	Acc	AUC
2 RCL (R2U-Net²²)	0.9576	0.9615	0.9999	0.9929
2 RCL + basic block	0.9590	0.9688	0.9999	0.9940
2 RCL + basic block + loss (DRU-Net)	0.9640	0.9632	0.9999	0.9943
3 RCL	0.9425	0.9853	0.9999	0.9935
3 RCL + basic block	0.9434	0.9715	0.9999	0.9923
3 RCL + loss	0.9428	0.9633	0.9999	0.9899
3 RCL + basic block + loss	0.9445	0.9721	0.9999	0.9888

Table 6 Different value of t on E-Ophtha dataset.

Method	Se	Sp	Acc	AUC
$t = 1$	0.9567	0.9629	0.9999	0.9930
$t = 2$	0.9640	0.9632	0.9999	0.9943
$t = 3$	0.9254	0.9582	0.9999	0.9853

to improve the accuracy of the segmentation which can make segmentation more precise.

In addition, we test different value of t on E-Ophtha dataset to get the best DRU-Net model. The experimental results are shown in Table 6. When $t = 2$, the segmentation of MAs is more accurate. By increasing or decreasing its value, the numerical segmentation results of MAs fall.

Based on RCNN and recurrent connections in the convolutional layers, we have designed the DRU-Net. Furthermore, it is for the first time that we used similar U-Net architecture to

segment MAs on E-Ophtha dataset and IDRiD dataset. DRU-Net architecture based on recurrent residual convolutional performs better than other network models (e.g., R2U-Net), which consists of the similar RRCNN block structure. Details of residual blocks, RCL unit, and loss functions are discussed and compared in Table 5. U-Net architecture with the RRCNN block, including two RCL units and a basic block, which uses the class-balancing loss function instead of the commonly used cross-entropy loss function, has strong performance with 0.9943 AUC value for MA segmentation. The RRCNN block containing different numbers of RCL units also has influence on the U-Net model. Experimental results show that the best architecture with three RCL units achieves an AUC value of 0.9935. Based on the above experiments, all changes of architectures are very close and have not significantly improved. DRU-Net achieves the best results with the class-balancing loss function and RRCNN block composed of two RCL units and a basic block.

5 Conclusion

In this paper, we propose the DRU-Net method, a U-Net architecture which consists of RRCNN blocks with basic block and RCL units, and which achieved good results in the MA segmentation task. We also demonstrate that RCL unit in the RRCNN block and basic block contributes to the U-Net. Different numbers of RCL units in the RRCNN block affect the performance of the uniform network; we find the best DRU-Net with two RCL units, which is combined with basic block to set up a good DRU-Net model. All the experiments show that our DRU-Net approach achieves better performance than existing methods, including U-Net and other residual U-Net-based improvement architectures in MA segmentation. In summary, our DRU-Net approach achieves state-of-the-art performance on the E-Ophtha dataset and the IDRiD dataset.

Disclosures

There are no conflicts of interest.

Acknowledgments

This work was funded by Chinese NSF Grant (No. 11871115).

References

- G. L. Ong et al., "Screening for sight-threatening diabetic retinopathy: comparison of fundus photography with automated color contrast threshold test," *Am. J. Ophthalmol.* **137**(3), 445–452 (2004).
- B. Zhang et al., "Detection of microaneurysms using multi-scale correlation coefficients," *Pattern Recognit.* **43**(6), 2237–2248 (2010).
- R. J. Tapp et al., "The prevalence of and factors associated with diabetic retinopathy in the Australian population," *Diabetes Care* **26**(6), 1731–1737 (2003).
- P. Chudzik et al., "Microaneurysm detection using deep learning and interleaved freezing," *Proc. SPIE* **10574**, 105741I (2018).
- N. Salamat, M. M. S. Missen, and A. Rashid, "Diabetic retinopathy techniques in retinal images: a review," *Artif. Intell. Med.* (2018).
- T. Spencer et al., "An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus," *Comput. Biomed. Res.* **29**(4), 284–302 (1996).
- G. Quellec et al., "Optimal wavelet transform for the detection of microaneurysms in retina photographs," *IEEE Trans. Med. Imaging* **27**(9), 1230–1241 (2008).
- I. Lazar and A. Hajdu, "Microaneurysm detection in retinal images using a rotating cross-section based model," in *IEEE Int. Symp. Biomed. Imaging: From Nano to Macro*, pp. 1405–1409 (2011).
- A. Mizutani et al., "Automated microaneurysm detection method based on double ring filter in retinal fundus images," *Proc. SPIE* **7260**, 72601N (2009).
- M. U. Akram, S. Khalid, and S. A. Khan, "Identification and classification of microaneurysms for early detection of diabetic retinopathy," *Pattern Recognit.* **46**(1), 107–116 (2013).
- R. Srivastava et al., "Detecting retinal microaneurysms and hemorrhages with robustness to the presence of blood vessels," *Comput. Meth. Programs Biomed.* **138**, 83–91 (2017).
- A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Adv. Neural Inf. Process. Syst.*, pp. 1097–1105 (2012).
- J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 3431–3440 (2015).
- S. Ren et al., "Faster R-CNN: towards real-time object detection with region proposal networks," in *Int. Conf. Neural Inf. Process. Syst.*, pp. 91–99 (2015).
- M. Haloi, "Improved microaneurysm detection using deep neural networks," arXiv:1505.04424 (2015).
- P. Chudzik et al., "Microaneurysm detection using fully convolutional neural networks," *Comput. Meth. Programs Biomed.* **158**, 185–192 (2018).
- S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *32nd Int. Conf. Machine Learning (ICML) 2015*, Vol. 1, pp. 448–456 (2015).
- J. H. Tan et al., "Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network," *Inf. Sci.* **420**, 66–76 (2017).
- L. Dai et al., "Clinical report guided retinal microaneurysm detection with multi-sieving deep learning," *IEEE Trans. Med. Imaging* **37**, 1149–1161 (2018).
- O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
- Z. Feng et al., "Deep retinal image segmentation: a FCN-based architecture with short and long skip connections for retinal image segmentation," *Lect. Notes Comput. Sci.* **10637**, 713–722 (2017).
- M. Z. Alom et al., "Recurrent residual convolutional neural network based on u-net (R2U-net) for medical image segmentation," arXiv:1802.06955 (2018).
- K. He et al., "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 770–778 (2016).
- M. Liang and X. Hu, "Recurrent convolutional neural network for object recognition," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 3367–3375 (2015).
- S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vision*, pp. 1395–1403 (2015).
- E. Decencière et al., "TeleOphta: machine learning and image processing methods for teleophthalmology," *IRBM* **34**(2), 196–203 (2013).
- P. Porwal et al., "Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research," *Data* **3**(3), 25 (2018).
- C. Lam et al., "Retinal lesion detection with deep learning using image patches," *Invest. Ophthalmol. Visual Sci.* **59**(1), 590–596 (2018).
- P. Costa et al., "EyeWes: weakly supervised pre-trained convolutional neural networks for diabetic retinopathy detection" (2018).
- H. Fu et al., "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation," *IEEE Trans. Med. Imaging* **37**, 1597–1605 (2018).

Caixia Kou is working in the School of Sciences, Beijing University of Posts and Telecommunications, China. She received her PhD from the State Key Laboratory of Scientific and Engineering Computing (LSEC), Chinese Academy of Sciences (CAS), Beijing, China, in 2011. Her major field is optimization algorithms and applications. She has proposed several efficient practical optimization algorithms and released the corresponding software packages. Her current interests include large-scale optimization methods and machine learning.

Wei Li is currently studying for an MS degree at Beijing University of Posts and Telecommunications. She received her BS degree in

mathematics from Hebei Normal University, Hebei, China, in 2017. Her current research interests include optimization theory and machine learning.

Wei Liang is currently studying for an MS degree at Beijing University of Posts and Telecommunications. He received his BS degree in mathematics from Hebei University of Technology, Tianjin, China, in 2016. His current research interests include optimization theory and machine learning.

Zekuan Yu is currently a PhD student in the Department of Biomedical Engineering, College of Engineering, Peking University, China. He received his BS degree in electrical engineering and

automation from China University of Mining and Technology in 2014. His current research interests include medical image analysis and computer-aided diagnosis, deep learning, and medical robot applications.

Jianchen Hao is working as an attending ophthalmologist in the Department of Ophthalmology, Peking University First Hospital, China. He received his MD degree from the Peking University Health Science Center in 2013. He has been involved with ophthalmology clinical work for many years and specializes in the diagnosis and treatment of fundus diseases such as diabetic retinopathy. Furthermore, he has engaged in clinical research relating to fundus diseases for many years.