

# DUNet: A deformable network for retinal vessel segmentation<sup>☆</sup>

Qiangguo Jin <sup>a</sup>, Zhaopeng Meng <sup>a,d</sup>, Tuan D. Pham <sup>b</sup>, Qi Chen <sup>a</sup>, Leyi Wei <sup>c</sup>, Ran Su <sup>a,\*</sup>

<sup>a</sup> School of Computer Software, College of Intelligence and Computing, Tianjin University, Tianjin, China

<sup>b</sup> Department of Biomedical Engineering, Linköping University, Sweden

<sup>c</sup> School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University, Tianjin, China

<sup>d</sup> Tianjin University of Traditional Chinese Medicine, Tianjin, China



## HIGHLIGHTS

- A deep neural network (DUNet) for automatic segmentation of retinal vessel is built.
- DUNet exploits the retinal vessels' features with a U-shape architecture.
- DUNet captures retinal vessels adaptively, according to vessels' scales and shapes.
- DUNet extracts more detailed vessels than deformable neural network and U-Net.
- Comparisons between many methods show competitive performance and generalization.

## ARTICLE INFO

### Article history:

Received 25 December 2018

Received in revised form 26 April 2019

Accepted 27 April 2019

Available online 3 May 2019

### Keywords:

Retinal blood vessel

Segmentation

DUNet

U-Net

Deformable convolution

## ABSTRACT

Automatic segmentation of retinal vessels in fundus images plays an important role in the diagnosis of some diseases such as diabetes and hypertension. In this paper, we propose Deformable U-Net (DUNet), which exploits the retinal vessels' local features with a U-shape architecture, in an end to end manner for retinal vessel segmentation. Inspired by the recently introduced deformable convolutional networks, we integrate the deformable convolution into the proposed network. The DUNet, with upsampling operators to increase the output resolution, is designed to extract context information and enable precise localization by combining low-level features with high-level ones. Furthermore, DUNet captures the retinal vessels at various shapes and scales by adaptively adjusting the receptive fields according to vessels' scales and shapes. Public datasets: DRIVE, STARE, CHASE\_DB1 and HRF are used to test our models. Detailed comparisons between the proposed network and the deformable neural network, U-Net are provided in our study. Results show that more detailed vessels can be extracted by DUNet and it exhibits state-of-the-art performance for retinal vessel segmentation with a global accuracy of 0.9566/0.9641/0.9610/0.9651 and AUC of 0.9802/0.9832/0.9804/0.9831 on DRIVE, STARE, CHASE\_DB1 and HRF respectively. Moreover, to show the generalization ability of the DUNet, we use another two retinal vessel data sets, i.e., WIDE and SYNTHE, to qualitatively and quantitatively analyze and compare with other methods. Extensive cross-training evaluations are used to further assess the extensibility of DUNet. The proposed method has the potential to be applied to the early diagnosis of diseases.

© 2019 Elsevier B.V. All rights reserved.

## 1. Introduction

The morphological and topographical changes of retinal vessels may indicate some pathological diseases, such as diabetes and hypertension. Diabetic Retinopathy (DR) caused by elevated

blood sugar levels, is a complication of diabetes in which retinal blood vessels leak into the retina, accompanying with the swelling of the retinal vessels [1]. DR can cause the growth of new blood vessels [2]. It must be noticeable if a diabetic patient appears in a swelling of the retinal vessels. Hypertensive Retinopathy (HR) is another commonly seen retina disease caused by high blood pressure [3]. An increased vascular tortuosity or narrowing of vessels can be observed in a patient with high blood pressure [4]. New technology for surgeons to directly inspect the fundus of the retina is analyzing the retinal fundus images by computer-aided diagnosis [5]. The information on the shape of blood vessels and bifurcation acquired through retinal vascular

<sup>☆</sup> No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to <https://doi.org/10.1016/j.knosys.2019.04.025>.

\* Corresponding author.

E-mail address: [ran.su@tju.edu.cn](mailto:ran.su@tju.edu.cn) (R. Su).

can not only enhance the monitoring of DR/HR but also reduce the repeatability of inter- and intra-observer [6]. Moreover, blood vessels are also used as landmarks for registration of the same patient's retinal images, which gathered from different sources [7]. In all cases, retinal blood vessels extracted from fundus images is crucial and can be applied to the early diagnosis of some severe diseases. This inspires the proposal of more accurate retinal blood vessel detection algorithms in order to facilitate the early diagnosis of pathological diseases.

However, the retinal blood vessels present extremely complicated structures, together with high tortuosity and various shapes [8], which makes the blood vessel segmentation task quite challenging. Different approaches have been proposed for blood vessel detection. They are mainly divided into two categories: manual segmentation and algorithmic segmentation. The manual way is time-consuming and in high-demand of skilled technical staff. Therefore, automated segmentation of retinal vessels, which can release the intense burden of manual segmentation, is highly demanded. However, due to the uneven intensity distribution of the retinal vascular images, the subtle contrast between the target vessels and the background of the images, high complexity of the vessel structures, coupled with image noise pollution, it is quite challenging to segment the retinal blood vessels in an accurate and efficient way.

In this study, we propose a Fully Convolutional Neural Network (FCN) based network named Deformable U-Net (DUNet) that enhances deep neural networks' capability of segmenting vessels in an end-to-end and pixel-to-pixel manner. It is designed to have a U-shape similar to U-Net [9], where upsampling operators with a large number of feature channels are stacked symmetrically to the conventional Convolutional Neural Network (CNN), so context information is captured and propagated to higher resolution layers and thus a more precise segmentation is obtained. The U-Net not only locates the object with high-level features but also accurately segments the object with the low-level features, which has shown its excellence in medical imaging tasks. However, extensive researches have shown that U-Net had its limitations, as a result, different variants of U-Net models have been proposed [2,10]. In this study, we believe that the limitation of U-Net remains in the following points: first, feature accumulation reaches its limitation due to the depth of U-Net, which may fail in extracting some tiny vessels. Second, the high complexity of the vessel structures bring difficulties for U-Net. Based on all of the above and inspired by the recently proposed deformable convolutional networks (Deformable-ConvNet) [11], we integrate deformable convolution blocks both in the encoder and decoder to capture the geometric transformations. Therefore, the receptive fields are adaptively adjusted according to the objects' scales and shapes. Complicated vessel structures can be well detected after then. Deformable-ConvNet and U-Net are used for comparison in this study. The importance and originality of this study are that we demonstrate that DUNet outperforms the other methods on the listed datasets and becomes a new state-of-the-art method, which is a powerful computer-aided system in clinical surgery and early diagnosis. Secondly, it shows the competitive generalization of DUNet in dealing with retinal vessel segmentation tasks on different datasets. As far as we know, the DUNet is tested on the most wide-ranging retinal datasets to show its extensibility. Thirdly, the DUNet is inspired by the U-Net architecture and introduces deformable convolution blocks, which enlarges the U-Net family in medical image analysis.

The rest of this paper is structured as follows: in the next section, we give a brief literature review of related work. Section 3 explains the architecture of DUNet and systematic retinal blood vessel segmentation method. Deformable-ConvNet and U-Net are also introduced briefly in this section. Experimental results are presented in Section 4, where we evaluate the proposed method on different retinal blood vessel datasets. Conclusions and discussions are given in Section 5.

## 2. Related work

The objective of retinal blood vessel segmentation is to locate and identify retinal vessel structures in the fundus images. With the development of imaging technology, various intelligent algorithms have been applied to retinal vessel segmentation. According to the learning patterns, retinal vessel segmentation methods can be divided into supervised methods and unsupervised methods. A brief overview of vessel segmentation from these two aspects is given.

### 2.1. Unsupervised method

The unsupervised method has no training samples in advance, and it constructs models directly by using filter responses or model-based techniques in most cases. There are many previous works on segmenting blood vessels in retinal images. Zana et al. presented an algorithm based on mathematical morphology and curvature evaluation for the detection of vessel-like patterns in a noisy environment and they obtained an accuracy of 0.9377 [12]. Chanwimaluang et al. proposed a free human intervention method to locate and extract blood vessels [13]. The proposed algorithm was composed of four steps, matched filtering, entropy-based thresholding, length filtering, and vascular intersection detection. However, this method is quite tediously long. Fraz et al. combined vessel centerlines detection and morphological bit plane slicing to extract vessel from retinal images [14]. Martinez-Perez et al. proposed a method to automatically segment retinal blood vessels based on multiscale feature extraction [15]. Niemeijer et al. compared a number of vessel segmentation algorithms [16]. According to this study, the highest accuracy of those compared algorithms reached 0.9416. Zhang et al. presented a retinal vessel segmentation algorithm using an unsupervised texton dictionary, where vessel textons were derived from responses of a multi-scale Gabor filter bank [17]. Better performance would be obtained if a proper pre-processing was carried out. Raja et al. proposed a three-stage method which automatically segmented the retinal blood vessels by eliminating the optic disk region from the retinal image [18]. Hassan et al. proposed a method which combined mathematical morphology and k-means clustering to segment blood vessels [19]. However, this method was not good at dealing with vessels of various widths. Tiny structures might be lost using this method. Oliveira et al. used a combined matched filter, Frangi's filter, and Gabor Wavelet filter to enhance the vessels [20]. They took the average of a few performance metrics to enhance the contrast between vessels and background. Jouandeau et al. presented an algorithm which was based on an adaptive random sampling algorithm [21]. Garg et al. proposed a segmentation approach which modeled the vessels as trenches [22]. They corrected the illumination, detected trenches by high curvature, and oriented the trenches in a particular direction first. Then they used a modified region growing method to extract the complete vessel structure. A threshold of mean illumination level that was set empirically might bring bias in this method. Zardadi et al. presented a faster unsupervised method for automatic detection of blood vessels in fundus images [23]. They enhanced the blood vessels in various directions; Then they presented an activation function on cellular responses; Next, they classified each pixel via an adaptive thresholding algorithm; Finally, morphological post-processing was carried out. However, several spots were falsely segmented into vessels which affected the final performance of the algorithm.

These traditional unsupervised methods do not utilize manual annotation, but a limitation of those methods is that they mainly rely on handcrafted features for vessel representation and segmentation [24]. Moreover, the performance of vessel extraction is not particularly good. In addition, the parameters used in the algorithm usually need to be carefully designed, due to the image noises and some pathological patterns.

## 2.2. Supervised method

Different from unsupervised learning, supervised learning requires hand-labeled data in order to build an optimally predictive model. All the inputs are mapped to the corresponding outputs using the built model. It has been widely applied to segmentation tasks. In order to achieve the purpose of this investigation, two processors are needed: one is an extractor to extract the feature vectors of pixels; The other one is a classifier to map extracted vectors to the corresponding labels. A number of feature extractors have been proposed, for instance, the Gabor filter [25], the Gaussian filter [26], etc. Various classifiers such as k-NN classifier [27], support vector machine (SVM) [28,29], artificial neural networks (ANN) [30], AdaBoost [31], etc., have been proposed to deal with different tasks.

Supervised methods were used widely in retinal vessel segmentation. Aslani et al. proposed a new segmentation method which characterized pixels with a vector of hybrid features calculated via a different extractor [32]. They trained a Random Forest (RF) classifier with the hybrid feature vector to classify vessel/non-vessel pixels. In order to simplify the model and increase efficiency, the number of Gabor features should be reduced as small as possible. Marín et al. used the Neural Network (NN) scheme for pixel classification and they computed a 7-D vector composed of gray-level and moment invariants-based features for pixel representation [33]. Yet the calculation cost was high and needed to be optimized.

For these traditional supervised methods, what features are used for classification greatly influence the final results of the prediction. However, they are often defined empirically, which requires human intervention and may cause bias. Therefore, an automated and effective feature extractor is highly demanded to achieve higher efficiency.

Deep learning is an architecture referring to an algorithm set which can solve the image, text and other tasks based on backpropagation and multi-layer neural network. One of the most significant contributions of deep learning is that it can replace handcrafted features with features automatically learned from deep hierarchical feature extraction method [34].

In a number of fields such as image processing, bioinformatics, and natural language processing, various deep learning architectures such as Deep Neural Networks, Convolutional Neural Networks, Deep Belief Networks and Recurrent Neural Networks have been widely used and have shown that they could produce state-of-the-art results on various tasks. Recently, there are some studies that investigated the vessel segmentation problems based on deep learning. Wang et al. preprocessed the retinal vessel images and then combined two superior classifiers, Convolutional Neural Network (CNN) and Random Forest (RF) together to carry out the segmentation [35]. Fu et al. used the deep learning architecture, formulated the vessel segmentation to a holistically-nested edge detection (HED) problem, and utilized the fully convolutional neural networks to generate vessel probability map [36]. Maji et al. used a ConvNet-ensemble based framework to process color fundus images and detect blood vessels [37]. Jiang et al. proposed a method which defined and computed pixels as primary features for segmentation, then a Neural Network (NN) classifier was trained using selected training data [38]. In this method, each pixel was represented by an 8-D vector. Then the unlabeled pixels were classified based on the vector. Azemin et al. estimated the impact of aging based on the results of the supervised vessel segmentation using artificial neural network [39]. It showed that different age groups affected different aspects of segmentation results. Liskowski et al. proposed a supervised segmentation architecture that used a Deep Neural Network with a large training dataset which was preprocessed via global contrast normalization, zero-phase whitening,

geometric transformations and gamma corrections [40]. And the network classified multiple pixels simultaneously using a variant structured prediction method. Fu et al. regarded the segmentation as a boundary detection problem and they combined the Convolution Neural Networks (CNN) and Conditional Random Field (CRF) layers into an integrated deep network to achieve their goal [41]. Tan et al. proposed a 10-layer CNN to automatically, simultaneously segment and discriminate exudates, hemorrhages and micro-aneurysms [42]. They used a single CNN to segment those pathological features with a desirable accuracy and showed the power of deep learning's feature extraction ability. What is more, Tan et al. also used a 7-layer CNN to automatically and simultaneously segment optic disk, fovea and vasculature [43]. Xu et al. proposed an automatic architecture which detected the DR [44]. In order to improve the algorithm's performance, they introduced a data augmentation as a preprocessing strategy. This method outperformed the results obtained by using classical approaches on Kaggle dataset.

Overall, it is expected that deep learning approaches can overcome the difficulties existed in the traditional unsupervised and supervised methods. Although those researches have been carried out on retinal vessel segmentation, there is still room to gain a precise vessel map, which is meaningful for early detection of diseases. It is even more challenging to propose a simpler but effective model for this complex task. In our study, we developed a systematic framework using the fully convolutional based methods to finish the effective and automatic segmentation task of retinal blood vessels.

## 3. Methodology

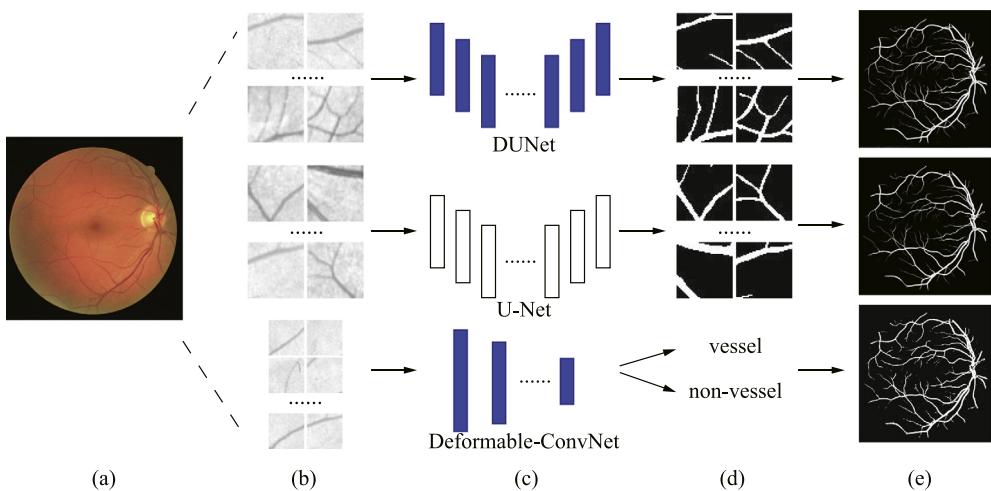
The aim of this study is to build deep learning models to segment retinal vessels in fundus images. Inspired by U-Net [9] and deformable convolutional network (Deformable-ConvNet) [11], we propose a new network named Deformable U-Net (DUNet<sup>1</sup>) for retinal vessel segmentation task. The proposed approach is designed to integrate the advantages of the deformable unit and U-Net architecture.

Fig. 1 shows an overview of the proposed DUNet, U-Net and Deformable-ConvNet. The raw images are preprocessed and cropped into small patches to establish training and validation dataset. In the contrast experiments, different models will be set with corresponding patch size. Since DUNet and U-Net are both pixel-to-pixel deep learning frameworks for segmentation, a 48 × 48 patch size is empirically used to trade off between computing complexity and efficiency, which is the same patch size referred to [10]. Meanwhile, Deformable-ConvNet is a model for vessel classification, a 29 × 29 patch size is chosen for training compared with [37,45]. After the inference of an image from the test dataset, all outputs from different models are re-composed to form a complete vessel probability map respectively.

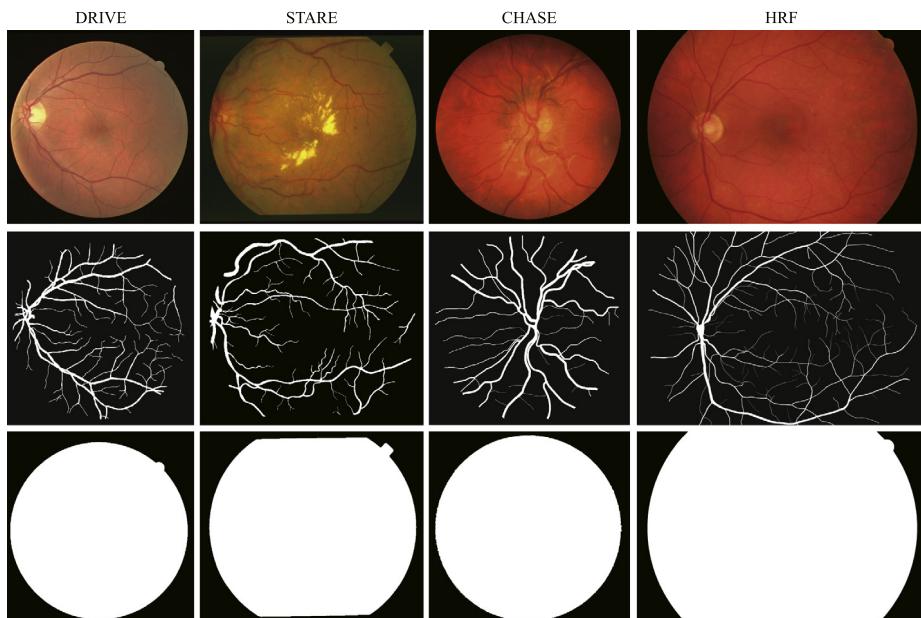
### 3.1. Datasets and material

Performances are evaluated on public datasets: DRIVE, STARE, CHASE\_DB1 (CHASE) and High-Resolution Fundus (HRF) dataset. The DRIVE (Digital Retinal Images for Vessel Extraction) contains 40 colored fundus photographs which were obtained from a diabetic retinopathy (DR) screening program in the Netherlands [46]. The plane resolution of DRIVE is 565 × 584. STARE (Structured Analysis of the Retina) dataset, which contains 20 images, is proposed to assist the ophthalmologist to diagnose eye diseases [47]. The plane resolution of STARE is 700 × 605. The

<sup>1</sup> <https://github.com/RanSuLab/DUNet-retinal-vessel-detection>.



**Fig. 1.** The pipeline of the three networks. (a) Original image; (b) Training samples; (c) Snapshots of proposed DUNet and compared models. Note that blue blocks refer to deformable convolution and the white ones represent regular convolution; (d) Inference results; (e) Re-composition of segmentation results. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

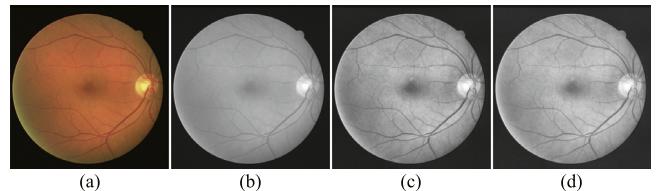


**Fig. 2.** Original retinal images (upper row), corresponding ground truth (middle row), field of view (bottom row) examples from DRIVE, STARE, CHASE and HRF sequentially.

CHASE dataset contains 28 images corresponding to two per patient for 14 children in the program Child Hear And Health Study in England [48]. The plane resolution of CHASE is  $999 \times 960$ . The HRF database contains three sets of fundus images: 15 images of healthy patients, 15 images of patients with diabetic retinopathy and 15 images of glaucomatous patients [49]. The ground truth is available for each image. The plane resolution of HRF is  $3504 \times 2336$ . Experts' manual annotations of the vascular are available as the ground truth (Fig. 2). Because the binary field of view (FOV) mask is not present for the STARE and CHASE datasets, we have manually created the masks for these two datasets as described in [50], which are presented in Fig. 2.

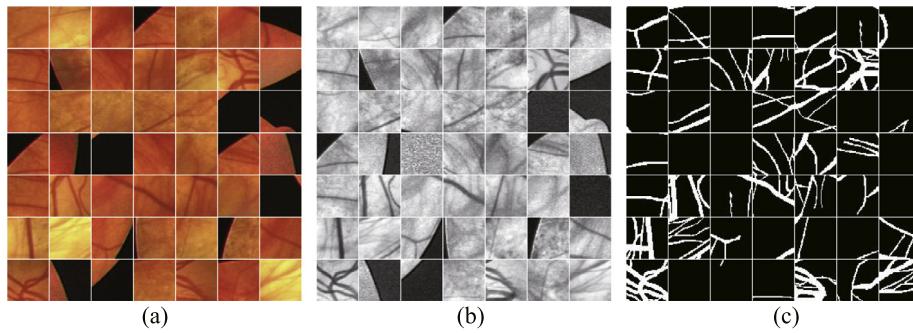
### 3.2. Image preprocessing and dataset preparation

The deep neural network has the ability to learn from un-preprocessed image data effectively. While it tends to be much more efficient if appropriate preprocessing has been applied to



**Fig. 3.** Typical images after each preprocessing step. (a) Original image; (b) Normalized image; (c) Image after CLAHE operation; (d) Image after Gamma correction.

the image data. In this study, three image preprocessing strategies were employed. Single channel images show the better vessel-background contrast than RGB images [51]. Thus, raw RGB images were converted into single channel ones. Normalization and Contrast Limited Adaptive Histogram Equalization [52] (CLAHE)



**Fig. 4.** Typical  $48 \times 48$  patches selected for model training. (a) shows the patches from the original images; (b) shows the patches from the preprocessed image; (c) shows the patches from the corresponding ground truth.

were used over the whole data set to enhance the foreground–background contrast. Finally, gamma correction was introduced to improve the image quality much further. Intermediate images after each preprocessing step are shown in Fig. 3.

To reduce overfitting problem, our models were trained on small patches which were randomly extracted from the images. In order to reduce the calculation complexity and ensure the surrounding local features, we set the size of the patch to  $48 \times 48$  for DUNet and U-Net empirically. It is worth mentioning that due to the high-resolution of HRF, we set the size of the patch to  $96 \times 96$ , which is different from the other datasets. The corresponding label for that patch was decided based on the ground truth images (Fig. 4).

All the datasets were divided into a training set, a validation set, and a test set. The training set was used for adjusting the weights. The validation set was used for selecting the best weight while the test set was used for performance evaluation. For DRIVE dataset, 20 images were used for training and validating purpose and the rest 20 images were used for testing. Since no splitting of training or test was provided for STARE/CHASE/HRF, we adopted “leave-one-out” method for STARE, from which training was conducted on the 19 samples, and the rest one image was tested iteratively [51]. For CHASE, there was no identical data partition strategy. After referring to [2,24], we used the first 14 images for training and the rest 14 images for testing. For HRF, we adopted the split strategy as proposed by Orlando et al. [53]. We used the first five images in each set (healthy, diabetic retinopathy and glaucoma) for training and validation, and the remaining 30 images for testing. From each training/validating image on DRIVE/STARE/CHASE/HRF, 10,000 patches were randomly sampled including 9,000 for training and 1,000 for validating. Therefore, DRIVE/STARE/CHASE/HRF had 180,000/171,000/126,000/135,000 patches as the training set and 20,000/19,000/14,000/15,000 patches as the validation set. Note that, the input patches are randomly sampled over the entire image (including those outside the FOV).

### 3.3. Deformable U-Net (DUNet)

Inspired by U-Net [9] and deformable convolutional network (Deformable-ConvNet) [11], we proposed the Deformable U-Net (DUNet) for retinal vessel segmentation task. The proposed network has a U-shaped architecture with encoders and decoders on two sides, and some of the original convolutional layers are replaced by the deformable convolutional block. The new model is trained to integrate the low-level features with the high-level features, and the receptive field and sampling locations are trained adaptively to vessels’ scale and shape, both of which enable precise segmentation.

Fig. 5 illustrates the architecture of DUNet. Detailed design of a deformable convolutional block is shown in the dashed

window. The architecture consists of a convolutional encoder (left side) and a decoder (right side) in a U-Net framework. In each encoding and decoding phase, deformable convolutional blocks are used to model retinal vessels of various shapes and scales through learning local, dense and adaptive receptive fields. Each deformable convolutional block consists of a convolution offset layer, which is the kernel concept of deformable convolution, a convolution layer, a batch normalization layer [54] and an activation layer. At the bottom of the DUNet, we use normal convolutional layers instead of the deformable blocks because a large number of parameters will be introduced without substantial performance improvement. With this architecture, DUNet can learn discriminative features and generate the precise retinal vessel segmentation results.

#### 3.3.1. Basic architecture: U-Net

Our customized implementation of U-Net has an overall architecture similar to the standard U-Net, consisting of an encoder and a decoder symmetrically on the two sides of the architecture. The encoding phase is used to encode input images in a lower dimensionality with richer filters, while the decoding phase is designed to do the inverse process of encoding by upsampling and merging low dimensional feature maps, which enables the precise localization. In order to solve the internal covariate shift problem and speed up the training processing, a batch normalization layer was inserted after each unit.

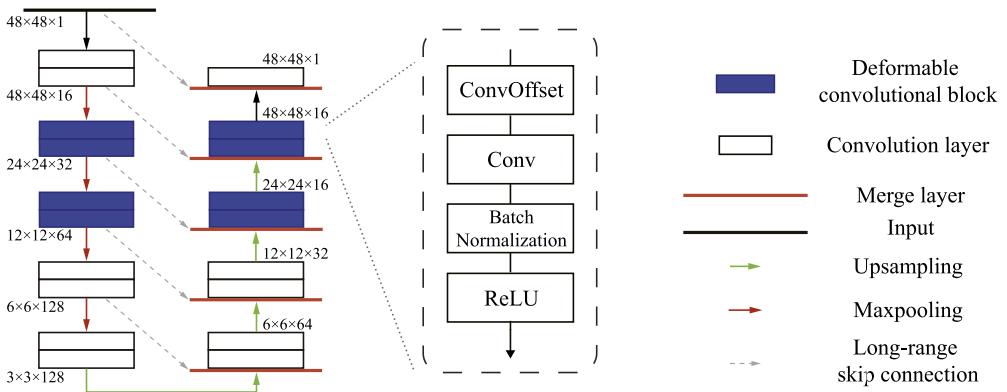
#### 3.3.2. Deformable convolution

A big challenge in vessel segmentation is to model the vessels with various shapes and scales [11]. Traditional methods such as the steerable filter [55], Frangi filter [56] exploit the vessel features through a linear combination of responses at multiple scales or direction, which may bring bias. The deformable convolutional network (Deformable-ConvNet) solved this problem by introducing deformable convolutional layers into the traditional neural networks. Inspired by the idea from Deformable-ConvNet that the various shapes and scales can be captured via deformable receptive fields, which are adaptive to the input features, we integrated the deformation convolution into the proposed network.

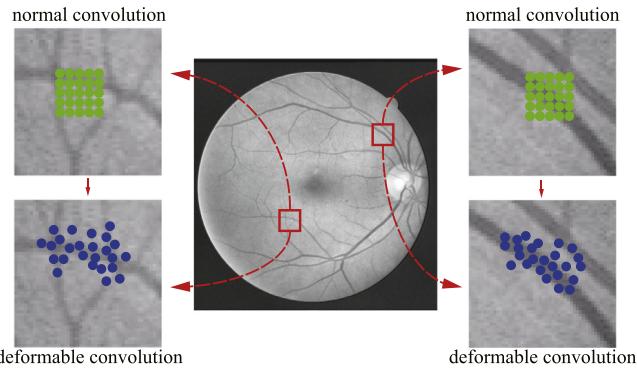
In the deformable convolution, offsets are added to the grid sampling locations which are normally used in the standard convolution. The offsets are learned from the preceding feature maps produced by the additional convolutional layers. Therefore, the deformation is able to adapt to different scales, shapes, orientation, etc. We take the  $5 \times 5$  deformable convolution as an example in Fig. 6.

As Fig. 6 shows, for a  $5 \times 5$  sized kernel with grid size 1, the normal convolution grid  $G$  can be formalized as:

$$G = \{(-2, 2), (-2, -1), \dots, (2, 1), (2, 2)\}, \quad (1)$$



**Fig. 5.** DUNet architecture with convolutional encoder and decoder based on U-Net architecture. Output size of feature map is listed beside each two layers.



**Fig. 6.** Illustration of the sampling locations in  $5 \times 5$  normal and deformable convolutions. The upper row stands for normal convolution and the corresponding deformable convolution is in the bottom row. Each sampling location has an offset to generate a new sampling location.

thus each location  $m_0$  from output feature map  $\mathbf{y}$  can be formalized as

$$\mathbf{y}(m_0) = \sum_{m_i \in G} \mathbf{w}(m_i) \cdot \mathbf{x}(m_0 + m_i), \quad (2)$$

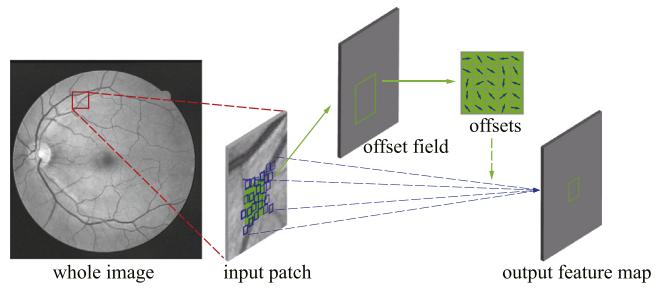
where  $\mathbf{x}$  denotes the input feature map,  $\mathbf{w}$  represents the weights of sampled value and  $m_i$  means the locations in  $G$ . While in deformable convolution, normal grid  $G$  is enhanced by the offset  $\Delta m_i$ , we have

$$\mathbf{y}(m_0) = \sum_{m_i \in G} \mathbf{w}(m_i) \cdot \mathbf{x}(m_0 + m_i + \Delta m_i). \quad (3)$$

Because offset  $\Delta m_i$  is usually not an integer, bilinear interpolation is applied to determine the value of the sampled points after migration. As mentioned above, the offset  $\Delta m_i$  is learned by an additional convolution layer. This procedure is illustrated in Fig. 7. Compared to the regular U-Net, DUNet may incur some computation cost in order to perform in a more local and adaptive manner.

#### 3.4. Comparison with U-Net and Deformable-ConvNet

We compared our proposed model with two state-of-the-art networks. One is the customized implementation of U-Net, which we have introduced above; The other is the deformable convolutional network (Deformable-ConvNet). Deformable-ConvNet was originally introduced to distinguish whether a pixel belongs to vessel or not. In this model, vessel segmentation was considered as a classification task. A pixel's feature can be extracted based



**Fig. 7.** Illustration of a  $5 \times 5$  deformable convolution. Offset field comes from the input patches and features while the channel dimension is  $2N$  corresponding to  $N$  2D offsets. The deformable convolutional kernel has the same resolution as the current convolution layer. The convolution kernels and the offsets are learned at the same time.

on its neighborhood defined as the patch centered on this pixel. In order to reduce the calculation complexity and to capture the local features, the size of the patch was set to  $29 \times 29$ . The Deformable-ConvNet mainly consists of a normal convolutional layer, three deformable convolutional blocks, etc. The deformable convolutional blocks are the same ones as those in DUNet. The architecture of the Deformable-ConvNet is shown in Fig. 8. The output size  $H \times W \times C$  of feature map is listed below the layer, where  $H, W, C$  are the height, width, channel of feature map respectively.

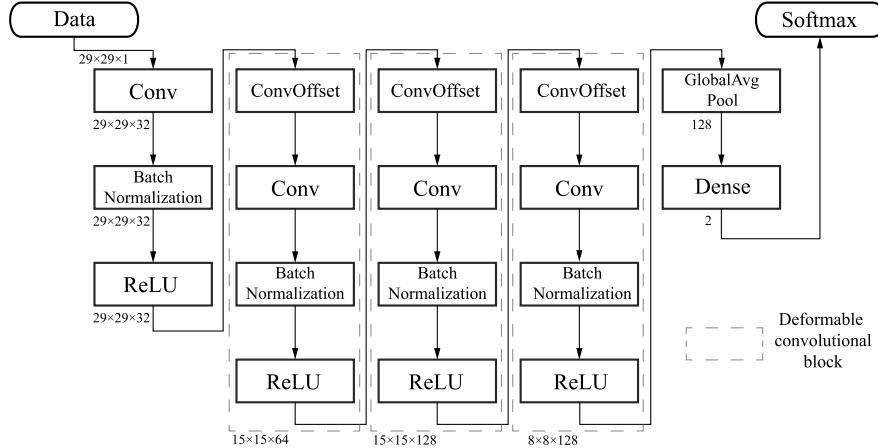
#### 3.5. Performance evaluation metrics

We evaluated our model using several metrics: Accuracy (ACC), Positive Predictive Value (PPV), True Positive Rate (TPR), True Negative Rate (TNR) and the Area Under Curve (AUC) of Receiver Operating Characteristic (ROC). ACC is a metric for measuring the ratio between the correctly classified pixels and the total pixels in the dataset. PPV, which is also called precision, indicates the proportion of the true positive samples among all the predicted positive samples. TPR, also known as sensitivity, measures the proportion of positives that are correctly identified. TNR, or specificity, measures the proportion of negatives that are correctly identified. These metrics have forms as following:

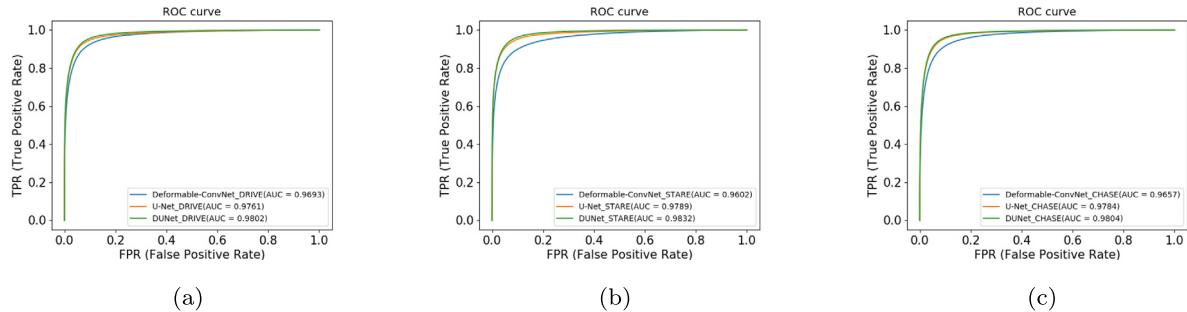
$$ACC = \frac{TP + TN}{TP + FP + TN + FN}, \quad (4)$$

$$PPV = \frac{TP}{TP + FP}, \quad (5)$$

$$TPR = \frac{TP}{TP + FN}, \quad (6)$$



**Fig. 8.** The architecture of the Deformable-ConvNet. It is mainly composed of convolution layers (Conv), deformable convolutional offset layers (ConvOffset), batch normalization layers and activation layers (ReLU). Layers in the dashed window consist in a deformable convolutional block.



**Fig. 9.** ROC curves of different models. (a) ROC curves on DRIVE; (b) ROC curves on STARE; (c) ROC curves on CHASE.

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (7)$$

where TP represents the number of the true positive samples; TN stands for the number of the true negative samples; FP means the number of the false positive samples; FN means the number of the false negative samples.

Additionally, performance was evaluated with F-measure ( $F_1$ ) [57] and Jaccard similarity (JS) [58] to compare the similarity and diversity of testing datasets. Here GT refers to the ground truth and SR refers to the segmentation result.

$$F_1 = 2 \cdot \frac{\text{PPV} \cdot \text{TPR}}{\text{PPV} + \text{TPR}} \quad (8)$$

$$\text{JS} = \frac{|\text{GT} \cap \text{SR}|}{|\text{GT} \cup \text{SR}|} \quad (9)$$

#### 4. Experimental result

We systematically compared the DUNet with Deformable-ConvNet and U-Net on the DRIVE, STARE and CHASE datasets. Then, the retinal vessel segmentation results on the test set were presented. Secondly, we compared DUNet with some other recently published approaches, most of which were under deep neural network framework and the others were standard segmentation approaches. What is more, we evaluated the performance of the DUNet on the High-Resolution Fundus Image dataset. Additionally, we tested on the other two extra datasets, i.e., SYNTHIA, WIDE, to evaluate the extendibility of DUNet. Furthermore, we used the cross-training strategy [24] and compared with some approaches to show the competitive generalization

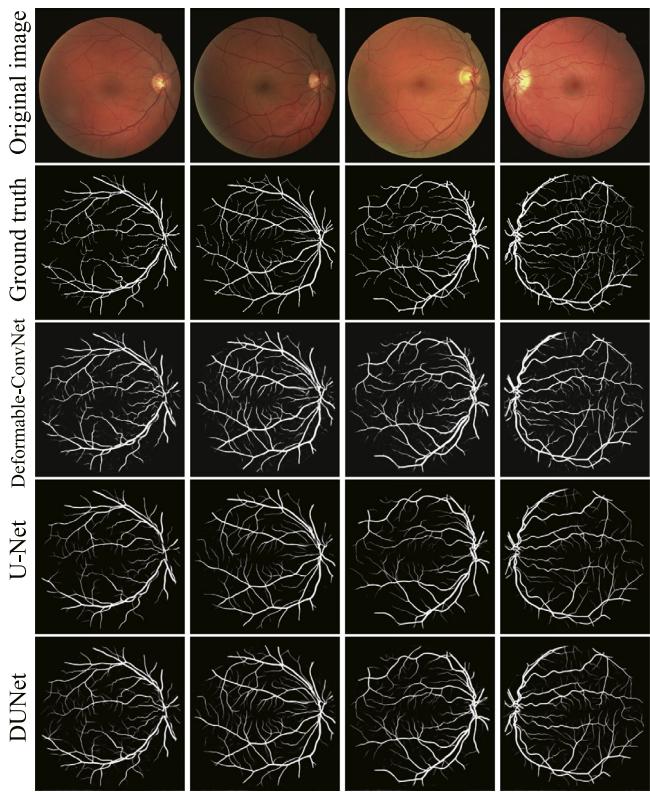
of DUNet. Experiments were conducted under the PyTorch [59] frameworks using an NVIDIA GeForce GTX 1080Ti GPU.

##### 4.1. Comparisons with Deformable-ConvNet and U-Net

We compared the three models, Deformable-ConvNet, U-Net and DUNet based on the DRIVE, STARE and CHASE datasets. As described in Section 3, we split the data into a training set, a validation set, and a test set. We used Adam as the optimizer and binary cross-entropy as the loss function and set the batch size to 60, total training epochs to 100. To ensure a quick convergence and avoid overfitting, we used a dynamic learning rate reduction strategy to adjust the learning rate values. The initial learning rate was set to 0.001. If the loss values remained stable after  $m_e$  epochs, the learning rate was reduced 10 times. Here  $m_e$  was set to 15 empirically. The validation accuracy and loss values were recorded during the training phase.

We evaluated the model using the test data. PPV, TPR, TNR, ACC,  $F_1$ -scores, JS and AUC were compared and shown in Tables 1–3. It shows from the tables that the DUNet achieves the highest values for most of the metrics. The global accuracy for Deformable-ConvNet, U-Net, and DUNet is 0.9486/0.9554/0.9566 on DRIVE, 0.9521/0.9637/0.9641 on STARE and 0.9536/0.9578/0.9610 on CHASE, respectively.

What is more, we evaluated the models using ROC curves, which is shown in Fig. 9. The closer the ROC curve to the top-left border is in the ROC coordinates, the more accurate a model is. These results show that the curves of DUNet are the most top-left one among the three models while the Deformable-ConvNet curve is the lowest one of the three. Besides, the figures also show that the DUNet obtains the largest area under the ROC curve (AUC).

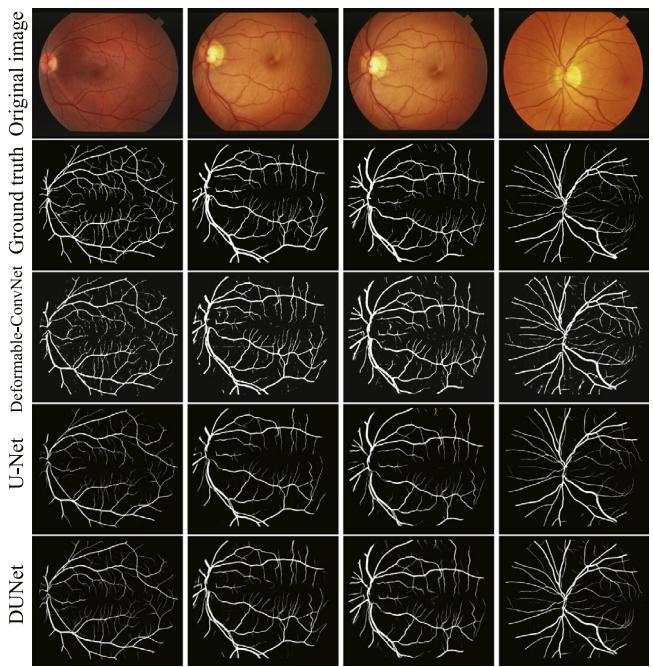


**Fig. 10.** Segmentation results using the different models on DRIVE.

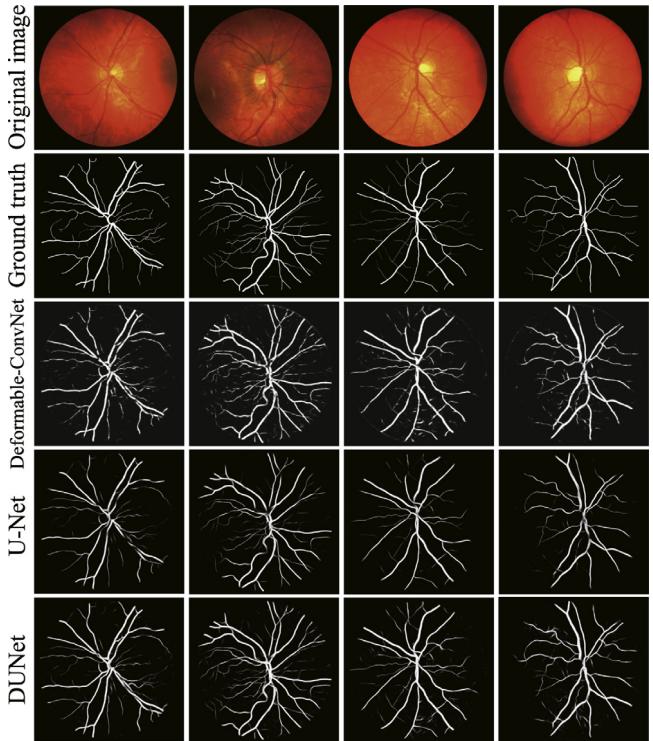
As Table 4 shows, the computation time of segmenting a retinal fundus image varies from each other. These results provide further support for the hypothesis that introducing an extra convolutional layer to learn the deformable offset will take the overhead of computation. In fact, a forward propagation with deformed convolution takes a long time, while regular convolution costs less. This aspect could be considered as a potential limitation of the proposed method. Another important point is that the DUNet requires fewer parameters, which is 0.88 million (M) compared to our customized U-Net, while reaching a better performance by introducing the deformable convolutional blocks.

In order to further observe the segmentation results which came from the three models, the probability maps for retinal vessel segmentation are displayed in Figs. 10–12. From the figures, it can be seen subjectively and objectively that DUNet produces more distinct vessel segmentation results. The proposed DUNet can detect weak vessel or vessels that are tied up which may be lost in U-Net and Deformable-ConvNet, thus it is more powerful to preserve more details.

The details of the segmentation results of the three models are presented in Fig. 13, it shows the local magnification view of the vascular junctions, where several vessels are tied up and close to each other, and the tiny vessels of DRIVE, STARE and CHASE respectively. Due to the complicated vascular trees, segmentation algorithms are difficult to proceed precisely with such complicated structures. In the junction region of vessels, Deformable-ConvNet and U-Net extracted coarse information due to the limitation of networks. It is worth mentioning that Deformable-ConvNet extracted vessels with more details than U-Net in some junction regions, which showed its ability to capture retinal vessels of various shapes. With the help of deformable convolutional blocks, the DUNet successfully segmented the tied vessels. In the tiny vessel regions, U-Net showed its limitation in handling details. However, Deformable-ConvNet picked them up somewhere.



**Fig. 11.** Segmentation results using the different models on STARE.



**Fig. 12.** Segmentation results using the different models on CHASE.

As a result, the DUNet obtained desirable segmentation results in those tiny and weak vessels.

With this structure, the DUNet is able to distinguish different vessels and present a better performance than the other models. Experimental results arrival at a conclusion that DUNet architecture has a more desirable performance in dealing with complicated and weak vessel structures among the three models mentioned above.

**Table 1**  
Performance of the three models tested on DRIVE.

Models	DRIVE						
	PPV	TPR	TNR	ACC	F <sub>1</sub>	JS	AUC
Deformable-ConvNet	0.8446	0.7311	0.9803	0.9486	0.7937	0.9486	0.9693
U-Net	<b>0.8531</b>	0.7849	<b>0.9802</b>	0.9554	0.8175	0.9555	0.9761
<b>DUNet</b>	0.8529	<b>0.7963</b>	0.9800	<b>0.9566</b>	<b>0.8237</b>	<b>0.9566</b>	<b>0.9802</b>

**Table 2**  
Performance of the three models tested on STARE.

Models	STARE						
	PPV	TPR	TNR	ACC	F <sub>1</sub>	JS	AUC
Deformable-ConvNet	0.7873	0.7371	0.9770	0.9521	0.7614	0.9522	0.9602
U-Net	0.8692	<b>0.7640</b>	0.9867	0.9637	0.8133	0.9637	0.9789
<b>DUNet</b>	<b>0.8777</b>	0.7595	<b>0.9878</b>	<b>0.9641</b>	<b>0.8143</b>	<b>0.9642</b>	<b>0.9832</b>

**Table 3**  
Performance of the three models tested on CHASE.

Models	CHASE						
	PPV	TPR	TNR	ACC	F <sub>1</sub>	JS	AUC
Deformable-ConvNet	0.7410	0.7364	0.9748	0.9536	0.7387	0.9537	0.9657
U-Net	0.7300	<b>0.8355</b>	0.9698	0.9578	0.7792	0.9578	0.9784
<b>DUNet</b>	<b>0.7630</b>	0.8155	<b>0.9752</b>	<b>0.9610</b>	<b>0.7883</b>	<b>0.9610</b>	<b>0.9804</b>

**Table 4**

Computation time comparison of the three models. For Deformable-ConvNet, computation time was calculated over all the  $29 \times 29$  sized patches of a retinal image. For the rest models, it was calculated over all the  $48 \times 48$  sized patches with the stride width/height of 5 from a retinal image. Params: Number of parameters.

Models	Params	DRIVE		STARE		CHASE	
		(565 × 584)	(700 × 605)	(999 × 960)	(999 × 960)	(999 × 960)	(999 × 960)
Deformable-ConvNet	0.27 M	58.7 s	77 s	179.9 s			
U-Net	3.35 M	<b>3.1 s</b>	<b>3.8 s</b>	<b>9.7 s</b>			
<b>DUNet</b>	<b>0.88 M</b>	15.3 s	20 s	47.7 s			

**Table 5**

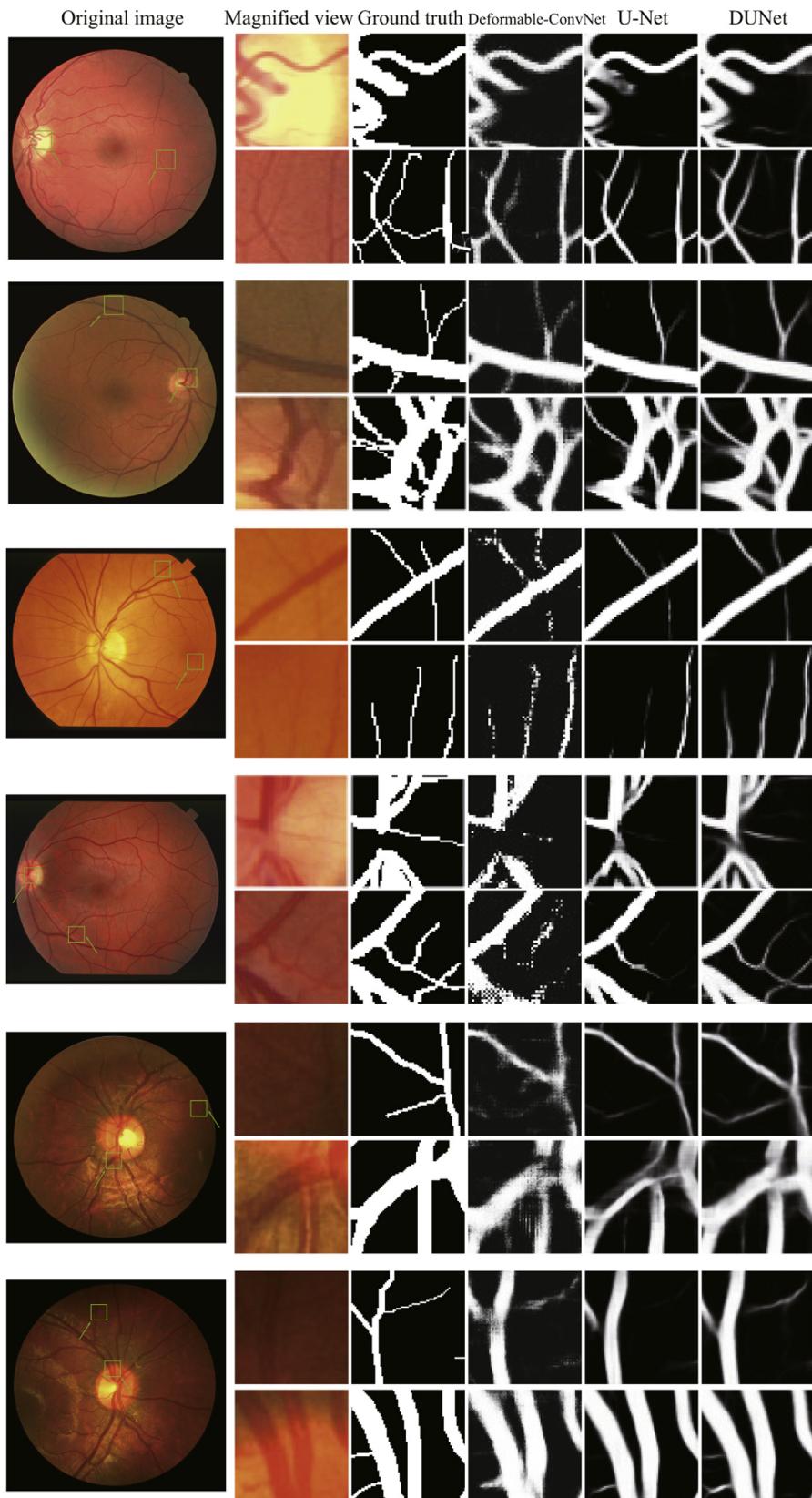
Comparisons against existing approaches on DRIVE, STARE and CHASE datasets.

Methods	Type	Year	DRIVE			STARE				CHASE				
			TPR	TNR	ACC	TPR	TNR	ACC	AUC	TPR	TNR	ACC	AUC	
Azzopardi et al. [5]	STA	2015	0.7655	0.9704	0.9442	0.9614	0.7716	0.9701	0.9497	0.9563	0.7585	0.9587	0.9387	0.9487
Li et al. [60]	DNN	2015	0.7569	0.9816	0.9527	0.9738	0.7726	0.9844	0.9628	0.9879	0.7507	0.9793	0.9581	0.9716
Liskowski et al. [40]	DNN	2016	0.7763	0.9768	0.9495	0.9720	<b>0.7868</b>	0.9754	0.9566	0.9785	–	–	–	–
Fu et al. [41]	DNN	2016	0.7603	–	0.9523	–	0.7412	–	0.9585	–	0.7130	–	0.9489	–
Dasgupta et al. [45]	DNN	2017	0.7691	0.9801	0.9533	0.9744	–	–	–	–	–	–	–	–
Roychowdhury et al. [61]	STA	2017	0.7250	<b>0.9830</b>	0.9520	0.9620	0.7720	0.9730	0.9510	0.9690	0.7201	<b>0.9824</b>	0.9530	0.9532
Chen et al. [62]	DNN	2017	0.7426	0.9735	0.9453	0.9516	0.7295	0.9696	0.9449	0.9557	–	–	–	–
Yan et al. [63]	DNN	2018	0.7653	0.9818	0.9542	0.9752	0.7581	0.9846	0.9612	0.9801	0.7633	0.9809	0.9610	0.9781
Yan et al. [24]	DNN	2018	0.7631	0.9820	0.9538	0.9750	0.7735	0.9857	0.9638	<b>0.9833</b>	0.7640	0.9806	0.9607	0.9776
<b>DUNet</b>	DNN	<b>2018</b>	<b>0.7963</b>	0.9800	<b>0.9566</b>	<b>0.9802</b>	0.7595	<b>0.9878</b>	<b>0.9641</b>	0.9832	<b>0.8155</b>	0.9752	<b>0.9610</b>	<b>0.9804</b>

#### 4.2. Comparison against existing methods

We also compared our method with several published state-of-the-art approaches which were proposed recently. Among them, some were standard segmentation algorithms (denoted with STA) while the others were all based on deep neural networks (denoted with DNN). Table 5 summarizes the type of algorithm, year of publication, and performance on DRIVE, STARE and CHASE dataset. Generally speaking, the STA based methods largely relied on handcrafted features and the performance was worse than DNN based methods. As can be seen from the table, the DUNet performed the best among those methods on DRIVE and CHASE. It achieved the highest global accuracy of 0.9566/0.9610 and the highest AUC of 0.9802/0.9804, which indicated that the DUNet exhibited state-of-the-art performance comparing both standard segmentation methods and deep neural network based methods. Although DUNet performed not better than Liskowski et al.'s method [40] and Yan et al.'s method [24]

on STARE, DUNet used less training patch samples than Liskowski et al.'s method [40], and the AUC was very close to the best result, i.e., Yan et al.'s method [24]. Concretely, Li et al. [60] extracted an image patch with a size of  $16 \times 16$ , which might fall into the flat area, thus it was not a good choice for a high-resolution dataset. Liskowski et al. [40] treated the segmentation task as a patch-based binary classification problem. Although this method reached new state-of-the-art performance in some metrics, it took a long time to predict a new retinal image compared with FCN based method, which was not practical for real application. Fu et al. [41] used the CRF for post-processing, which was not purely CNN-based method. Dasgupta et al. [45] and Chen et al. [62] tested only on one or two datasets, the generalization of this method remained unknown. Yan et al. [63] first trained a segment-level loss branch, then trained a joint loss model. Although the performance was desirable, the training procedure was complicated. Yan et al. [24] had the same shortcoming in



**Fig. 13.** Magnified view of green-boxed patches predicted by different models on DRIVE (two rows above), STARE (two rows middle) and CHASE (two rows below).

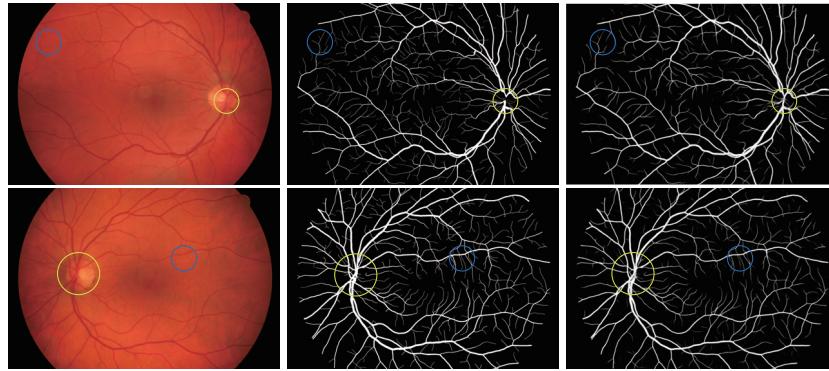
training. In summary, the DUNet architecture is simple to train and contains only 0.88 M parameters while it reaches the best overall performance comparing to those listed methods.

#### 4.3. Performance on high-resolution dataset

We also evaluated the performance of the DUNet on the HRF database. In the test phase, we cropped the HRF dataset into

**Table 6**  
Comparisons against existing approaches on HRF dataset.

Method	Type	Year	PPV	TPR	TNR	ACC	AUC
Odstcilik et al. [64]	STA	2013	0.6950	0.7794	0.9650	–	–
Zhang et al. [65]	STA	2016	–	<b>0.7978</b>	0.9717	0.9556	–
Orlando et al. [53]	DNN	2017	0.6630	0.7874	0.9584	–	–
Laibacher et al. [2]	DNN	2018	–	–	–	0.9635	–
Yan et al. [63]	DNN	2018	0.6647	0.7881	0.9592	0.9437	–
<b>DUNet</b>	<b>DNN</b>	<b>2018</b>	<b>0.8593</b>	0.7464	<b>0.9874</b>	<b>0.9651</b>	<b>0.9831</b>



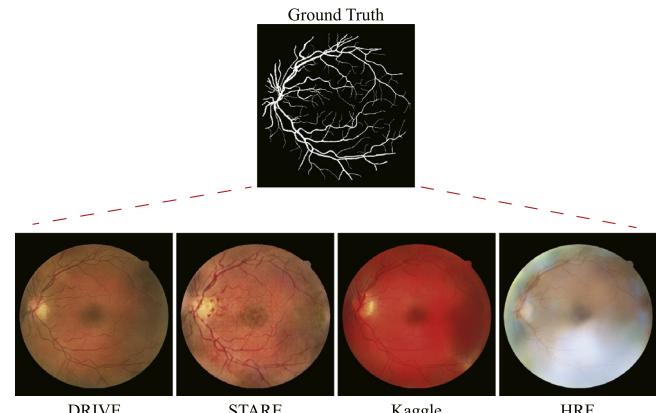
**Fig. 14.** Two typical results on the HRF dataset. The first column: the fundus images. The second column: the ground truth. The third column: the segmentation results generated by DUNet. Typical thick vessels are in yellow circles. Typical thin vessels in low contrast are in blue circles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

192 × 192 sized patches. Compared to the state-of-the-art methods, DUNet reached the accuracy/AUC of 0.9651/0.9831, which was the best among them (Table 6). As for TPR, lower compared to other methods, a possible explanation for this might be that the training patch size is not large enough compared to a high-resolution image. Two typical segmentation results are presented in Fig. 14. It shows that the DUNet may fail in segmenting some thick vessels (yellow circles), these results are likely to be related to the low-resolution patches we used for training, larger resolution patches might improve the performance, i.e., 388 × 388, 512 × 512. On the contrary, the DUNet successfully segmented the microvessels (blue circles) in Fig. 14. The DUNet makes it possible for real-world application on high-resolution datasets, which is not highlighted in some of the state-of-the-art methods.

#### 4.4. Generalization of the proposed DUNet

Additionally, we compared our method with Dasgupta et al.'s [45] method and Alom et al.'s [10] method on the other two datasets for qualitative and quantitative analysis. The first dataset, named WIDE, which was initially used for tree topology estimation, contains 15 high-resolution and wide-field RGB images. Each retinal image was taken from a different individual and captured as an uncompressed TIFF file at the widest setting [66]. The WIDE dataset does not contain ground truth for retinal vessel segmentation. We used the WIDE for qualitative analysis and compared the proposed method with the other two methods. The second dataset (denoted with SYNTHETE) is synthesized from generative adversarial nets [67]. The dataset contains 20 retinal images at 565 × 584 resolution, which includes DRIVE, STARE, Kaggle and HRF style. Each of these styles contains 5 retinal images generated by 5 corresponding ground truth images. Fig. 15 shows the SYNTHETE dataset, four distinct style retinal images are generated from the same vessel map.

To show the generalization of these three models, we used the weights well-trained on DRIVE and predicted on the WIDE dataset. We preprocessed and cropped these images into patches in the same way. From Fig. 16, it qualitatively indicates that DUNet produces competitive results.



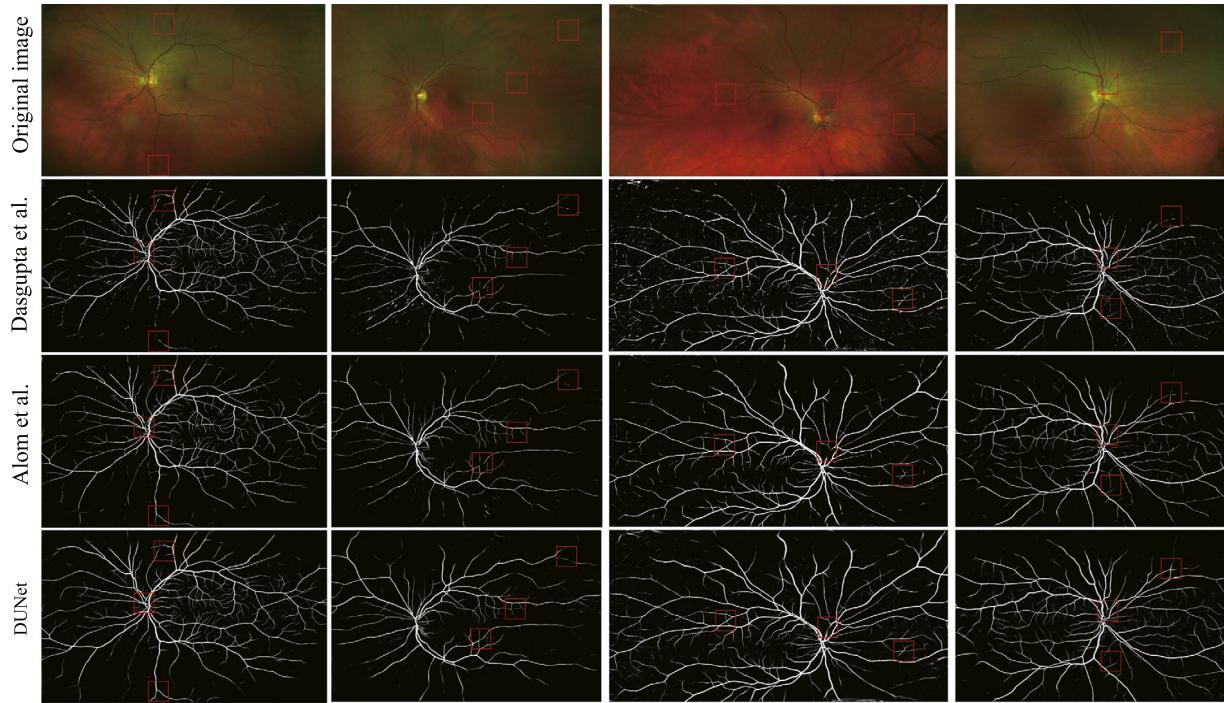
**Fig. 15.** Four distinct style of retinal images synthesized by generative adversarial nets.

To further validate quantitatively the generalization of these models, we also used the well-trained weights from DRIVE and tested on the SYNTHETE datasets. We mixed the four distinct style images together, preprocessed and generated segmentation results on SYNTHETE. The performances of the three models are summarized in Table 7. It shows that the DUNet achieved the highest TPR, JS and AUC while the PPV, TNR, ACC and F<sub>1</sub> are lower than those of Alom et al.'s [10]. However, the Alom et al.'s [10] method requires much more parameters than the DUNet, which is more difficult to train in real applications. Note that the Dasgupta et al.'s [45] and Alom et al.'s [10] methods were replicated in this paper. In summary, it proves quantitatively that the DUNet achieved a rather competitive performance in terms of generalization.

Last but not least, we conducted the cross-training evaluation to test the extendibility [24]. We tested the generalization of DUNet by applying the well-trained model on one dataset to another dataset. Different from [60] which retrained the model, we used the well-trained models described in Section 4.1. The

**Table 7**  
Performances of the three models tested on SYNTHIA using well-trained weights on DRIVE.

Models	SYNTHIA							
	Params	PPV	TPR	TNR	ACC	F <sub>1</sub>	JS	AUC
Dasgupta et al. [45]	<b>0.14 M</b>	0.8272	0.6320	0.9786	0.9304	0.7165	0.9304	0.9266
Alom et al. [10]	48.9 M	<b>0.8760</b>	0.6463	<b>0.9852</b>	<b>0.9381</b>	<b>0.7439</b>	0.9380	0.9345
DUNet	0.88 M	0.8510	<b>0.6488</b>	0.9816	0.9353	0.7363	<b>0.9654</b>	<b>0.9399</b>



**Fig. 16.** Detailed view of four images on WIDE. The DUNet outperforms the other two methods overall. Particularly, it can be seen in the red boxes that the DUNet achieves the best segmentation performance. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 8**  
Comparisons of the cross-training evaluation.

Dataset	Methods	Year	TPR	TNR	ACC	AUC
DRIVE (trained on STARE)	Fraz et al. [68]	2012	0.7242	0.9792	0.9456	0.9697
	Li et al. [60]	2015	<b>0.7273</b>	0.9810	<b>0.9486</b>	0.9677
	Yan et al. [24]	2018	0.7014	0.9802	0.9444	0.9568
	DUNet	2018	0.6505	<b>0.9914</b>	0.9481	<b>0.9718</b>
STARE (trained on DRIVE)	Fraz et al. [68]	2012	0.7010	0.9770	0.9495	0.9660
	Li et al. [60]	2015	0.7027	0.9828	0.9545	0.9671
	Yan et al. [24]	2018	0.7319	<b>0.9840</b>	<b>0.9580</b>	0.9678
	DUNet	2018	0.7000	0.9759	0.9474	0.9571
	DUNet <sup>a</sup>	2018	<b>0.8419</b>	0.9563	0.9445	<b>0.9690</b>

<sup>a</sup>The result is obtained by retrained model on DRIVE dataset using a preprocessing strategy without normalization and CLAHE.

results of the cross-training evaluation are summarized in Table 8. For DRIVE dataset, we obtained the highest AUC and TNR values while the TPR and accuracy were lower than some other methods. Contrarily, for STARE dataset, the TPR, TNR, ACC and AUC were lower than other methods. Possible explanations for this might be that the STARE dataset mainly contains thick vessels and the DRIVE dataset contains thin vessels [24], and the STARE is complicated than the DRIVE. To further evaluate the generalization, we provided a new result which was retained on DRIVE using a different preprocessing strategy. The result indicated that the DUNet produced the best TPR/AUC among the state-of-the-art methods and obtained an overall best performance.

## 5. Conclusion

Deep neural networks, which uses hierarchical layers of learned features to accomplish high-level tasks, has been applied to a wide range of medical processing tasks. In this study, we propose a fully convolutional neural network, named DUNet to handle the retinal vessel segmentation task in a pixel-wise manner. DUNet is an extension of the U-Net with some of convolutional layers replaced by the deformable convolution blocks. With the symmetric U-shape architecture, DUNet is designed to capture context by the encoder and enable precise localization by the decoder by combining the low-level feature maps with the high-level ones. With the deformable convolution blocks, DUNet is able to capture the retinal blood vessels at various

shapes and scales by adaptively adjusting the receptive fields according to the vessels' scales and shapes. By adding offsets to the regular sampling grids of standard convolution, the receptive fields are deformable and augmented, which improves the performance and reduces the training parameters as well. In order to test the performance of the proposed network, we have trained Deformable-ConvNet and customized U-Net from scratch for comparison on three public datasets: DRIVE, STARE and CHASE\_DB1. This is also the first time that DUNet being used to conduct the retinal vessel segmentation. Besides, comparisons with several standard segmentation algorithms and some other deep neural network based approaches are introduced. DUNet is tested on the High-Resolution Fundus Image dataset for further evaluation. To validate the generalization of our model, we test the DUNet on WIDE and SYNTHE datasets, and analyze qualitatively and quantitatively. Extensive cross-training evaluations are used to further assess the extendibility of DUNet. Results show that with the help of deformable convolutional blocks, more detailed vessels are extracted, and the DUNet exhibits state-of-the-art performance in segmenting the retinal vessels. However, The limitation of this architecture in this study is the cost of computation time in the train/test model stage, which is caused by introducing extra convolutions to learn deformable offsets.

The proposed method segments the healthy/pathological retinal vessels, which is an early stage of diagnosing some severe disease. None of the handcrafted features is needed in the architecture. Consequently, practical clinical applications of the automatic diagnosing system are wide, e.g., DR/DR monitoring during the lifetime. The DUNet provides a general, high-performance computing framework for retinal vessel segmentation. In the future, more retinal vessel data will be incorporated to validate the proposed end-to-end model. And an efficient implementation is needed to reduce the computation time. We also plan to extend our DUNet architecture to three dimensions, aiming to obtain more accurate results in medical image analysis tasks. This would be a fruitful area for further work.

## Acknowledgments

This work is supported by the National Natural Science Foundation of China [Grant No. 61702361], the Science and Technology Program of Tianjin, China [Grant No. 16ZXHGX00170], the National Key Technology R&D Program of China [Grant No. 2015BAH52F00] and the National Key Technology R&D Program of China [Grant No. 2018YFB1701700].

## References

- [1] T.J. Smart, C.J. Richards, R. Bhatnagar, C. Pavesio, R. Agrawal, P.H. Jones, A study of red blood cell deformability in diabetic retinopathy using optical tweezers, in: Optical Trapping and Optical Micromanipulation XII, vol. 9548, International Society for Optics and Photonics, 2015, 954825.
- [2] T. Laibacher, T. Weyde, S. Jalali, M2U-Net: Effective and efficient retinal vessel segmentation for resource-constrained environments, 2018, arXiv preprint [arXiv:1811.07738](https://arxiv.org/abs/1811.07738).
- [3] S. Irshad, M. Akram, Classification of retinal vessels into arteries and veins for detection of hypertensive retinopathy, in: Biomedical Engineering Conference, CIBEC, 2014 Cairo International, IEEE, 2014, pp. 133–136.
- [4] C.Y. Cheung, Y. Zheng, W. Hsu, M.L. Lee, Q.P. Lau, P. Mitchell, J.J. Wang, R. Klein, T.Y. Wong, Retinal vascular tortuosity, blood pressure, and cardiovascular risk factors, *Ophthalmology* 118 (5) (2011) 812–818.
- [5] G. Azzopardi, N. Strisciuglio, M. Vento, N. Petkov, Trainable COSFIRE filters for vessel delineation with application to retinal images, *Med. Image Anal.* 19 (1) (2015) 46–57.
- [6] J.H. Tan, U.R. Acharya, K.C. Chua, C. Cheng, A. Laude, Automated extraction of retinal vasculature, *Med. Phys.* 43 (5) (2016) 2311–2322.
- [7] L. Xu, S. Luo, A novel method for blood vessel detection from retinal images, *Biomed. Eng. Online* 9 (1) (2010) 14, [http://dx.doi.org/10.1186/1475-925X-9-14](https://doi.org/10.1186/1475-925X-9-14).
- [8] Z. Han, Y. Yin, X. Meng, G. Yang, X. Yan, Blood vessel segmentation in pathological retinal image, in: IEEE International Conference on Data Mining Workshop, 2014, pp. 960–967.
- [9] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
- [10] M.Z. Alom, M. Hasan, C. Yakopcic, T.M. Taha, V.K. Asari, Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation, 2018, arXiv preprint [arXiv:1802.06955](https://arxiv.org/abs/1802.06955).
- [11] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, Y. Wei, Deformable convolutional networks, in: International Conference on Computer Vision, 2017, pp. 764–773.
- [12] F. Zana, J. Klein, Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation, *IEEE Trans. Image Process.* 10 (7) (2001) 1010–1019.
- [13] T. Chanwimaluang, G. Fan, An efficient blood vessel detection algorithm for retinal images using local entropy thresholding, in: Proceedings of International Symposium on Circuits and Systems, 2003. ISCAS'03, vol. 5, 2003, pp. V-21, [http://dx.doi.org/10.1109/ISCAS.2003.1206162](https://doi.org/10.1109/ISCAS.2003.1206162).
- [14] M.M. Fraz, S.A. Barman, P. Remagnino, A. Hoppe, A. Basit, U. Uyyanonvara, A.R. Rudnicka, C.G. Owen, An approach to localize the retinal blood vessels using bit planes and centerline detection, *Comput. Methods Programs Biomed.* 108 (2) 600–616.
- [15] M. Martinez-Perez, S. Hughes, A. Bharath, K. Parker, Segmentation of blood vessels from red-free and fluorescein retinal images, *Med. Image Anal.* 11 (1) (2007) 47–61.
- [16] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, M.D. Abramoff, Comparative study of retinal vessel segmentation methods on a new publicly available database, in: Medical Imaging 2004: Image Processing, Vol. 5370, International Society for Optics and Photonics, 2004, pp. 648–657.
- [17] L. Zhang, M. Fisher, W. Wang, Retinal vessel segmentation using multi-scale textons derived from keypoints, *Comput. Med. Imaging Graph.* 45 (2015) 47–56.
- [18] D. Siva Sundhara Raja, V. Sivamurugesan, Automatic detection of blood vessels in retinal images for diabetic retinopathy diagnosis, *Comput. Math. Methods Med.* 2015 (2015) 419279, [http://dx.doi.org/10.1155/2015/419279](https://doi.org/10.1155/2015/419279).
- [19] G. Hassan, N. El-Bendary, A.E. Hassani, A. Fahmy, A.M. Shoeb, V. Snasel, Retinal blood vessel segmentation approach based on mathematical morphology, *Procedia Comput. Sci.* 65 (2015) 612–622.
- [20] W.S. Oliveira, J.V. Teixeira, T.I. Ren, G.D. Cavalcanti, J. Sijbers, Unsupervised retinal vessel segmentation using combined filters, *PLoS One* 11 (2) (2016) e0149943.
- [21] N. Jouandeau, Z. Yan, P. Greussay, B. Zou, Y. Xiang, Retinal vessel segmentation based on adaptive random sampling, *J. Med. Bioeng.* 3 (3) (2014) 199–202.
- [22] S. Garg, J. Sivaswamy, S. Chandra, Unsupervised curvature-based retinal vessel segmentation, in: IEEE International Symposium on Biomedical Imaging: From Nano To Macro, IEEE, 2007, pp. 344–347.
- [23] M. Zardadi, N. Mehrshad, S.M. Razavi, Unsupervised segmentation of retinal blood vessels using the human visual system line detection model, *J. Inf. Syst. Telecommun.* 4 (2016) 125–133, [http://dx.doi.org/10.7508/jist.2016.02.008](https://doi.org/10.7508/jist.2016.02.008).
- [24] Z. Yan, X. Yang, K.-T. Tim Cheng, A three-stage deep learning model for accurate retinal vessel segmentation, *IEEE J. Biomed. Health Inf. PP* (2018) 2168–2194, [http://dx.doi.org/10.1109/JBHI.2018.2872813](https://doi.org/10.1109/JBHI.2018.2872813).
- [25] Y. Hamamoto, S. Uchimura, M. Watanabe, T. Yasuda, Y. Mitani, S. Tomita, A gabor filter-based method for recognizing handwritten numerals, *Pattern Recognit.* 31 (4) (1998) 395–400.
- [26] V. Nguyen, M. Blumenstein, An Application of the 2D Gaussian filter for enhancing feature extraction in off-line signature verification, in: International Conference on Document Analysis and Recognition, 2011, pp. 339–343.
- [27] J. Staal, M.D. Abràmoff, M.A. Viergever, B. Van Ginneken, Ridge-based vessel segmentation in color images of the retina, *IEEE Trans. Med. Imaging* 23 (4) (2004) 501–509.
- [28] E. Ricci, R. Perfetti, Retinal blood vessel segmentation using line operators and support vector classification, *IEEE Trans. Med. Imaging* 26 (10) (2007) 1357–1365.
- [29] X. You, Q. Peng, Y. Yuan, Y.M. Cheung, J. Lei, Segmentation of retinal blood vessels using the radial projection and semi-supervised approach, *Pattern Recognit.* 44 (10) (2011) 2314–2324.
- [30] C. Sinthanayothin, J.F. Boyce, H.L. Cook, T.H. Williamson, Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images, *Br. J. Ophthalmol.* 83 (8) (1999) 902.
- [31] X. Li, L. Wang, E. Sung, AdaBoost with SVM-based component classifiers, *Eng. Appl. Artif. Intell.* 21 (5) (2008) 785–795.
- [32] S. Aslani, H. Sarnej, A new supervised retinal vessel segmentation method based on robust hybrid features, *Biomed. Signal Process. Control* 30 (2016) 1–12.

- [33] D. Marin, A. Aquino, M.E. Gegundezarias, J.M. Bravo, A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features, *IEEE Trans. Med. Imaging* 30 (1) (2011) 146–158.
- [34] H.A. Song, S.Y. Lee, Hierarchical representation using NMF, in: International Conference on Neural Information Processing, 2013, pp. 466–473.
- [35] S. Wang, Y. Yin, G. Cao, B. Wei, Y. Zheng, G. Yang, Hierarchical retinal blood vessel segmentation based on feature and ensemble learning, *Neurocomputing* 149 (2015) 708–717.
- [36] H. Fu, Y. Xu, D.W.K. Wong, J. Liu, Retinal vessel segmentation via deep learning network and fully-connected conditional random fields, in: IEEE International Symposium on Biomedical Imaging, 2016, pp. 698–701.
- [37] D. Maji, A. Santara, P. Mitra, D. Sheet, Ensemble of deep convolutional neural networks for learning to detect retinal vessels in fundus images, 2016, arXiv preprint [arXiv:1603.04833](https://arxiv.org/abs/1603.04833).
- [38] P. Jiang, Q. Dou, X. Hu, A supervised method for retinal image vessel segmentation by embedded learning and classification, *J. Intell. Fuzzy Systems* 29 (5) (2015) 2305–2315.
- [39] M.Z.C. Azemin, F.A. Hamid, M.I.B.M. Tamrin, A.H.M. Amin, Supervised retinal vessel segmentation based on neural network using broader aging dataset, in: IWBBIO, 2014, pp. 1235–1242.
- [40] P. Liskowski, K. Krawiec, Segmenting retinal blood vessels with deep neural networks, *IEEE Trans. Med. Imaging* 35 (11) (2016) 2369–2380.
- [41] H. Fu, Y. Xu, S. Lin, D.W.K. Wong, J. Liu, DeepVessel: Retinal vessel segmentation via deep learning and conditional random field, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2016, pp. 132–139.
- [42] J.H. Tan, H. Fujita, S. Sivaprasad, S.V. Bhandary, A.K. Rao, K.C. Chua, U.R. Acharya, Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network, *Inform. Sci.* 420 (2017) 66–76.
- [43] J.H. Tan, U.R. Acharya, S.V. Bhandary, K.C. Chua, S. Sivaprasad, Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network, *J. Comput. Sci.* 20 (2017) 70–79.
- [44] K. Xu, D. Feng, H. Mi, Deep convolutional neural network-based early automated detection of diabetic retinopathy using fundus image, *Molecules* 22 (12) (2017) 2054.
- [45] A. Dasgupta, S. Singh, A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation, in: International Symposium on Biomedical Imaging, IEEE, 2017, pp. 248–251.
- [46] J. Staal, M.D. Abràmoff, M. Niemeijer, M.A. Viergever, B. Van Ginneken, Ridge-based vessel segmentation in color images of the retina, *IEEE Trans. Med. Imaging* 23 (4) (2004) 501–509.
- [47] A. Hoover, V. Kouznetsova, M. Goldbaum, Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response, *IEEE Trans. Med. Imaging* 19 (3) (2002) 203–210.
- [48] C.G. Owen, A.R. Rudnicka, R. Mullen, S.A. Barman, D. Monekosso, P.H. Whincup, J. Ng, C. Paterson, Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (CAIAR) program, *Invest. Ophthalmol. Vis. Sci.* 50 (5) (2009) 2004–2010.
- [49] A. Budai, R. Bock, A. Maier, J. Hornegger, G. Michelson, Robust vessel segmentation in fundus images, *Int. J. Biomed. Imaging* 2013 (6) (2013) 1–11, <http://dx.doi.org/10.1155/2013/154860>.
- [50] J.I. Orlando, E. Prokofyeva, M. del Fresno, M.B. Blaschko, Convolutional neural network transfer for automated glaucoma identification, in: 12th International Symposium on Medical Information Processing and Analysis, vol. 10160, International Society for Optics and Photonics, 2017, p. 101600U, <http://dx.doi.org/10.1117/12.2255740>.
- [51] J.V. Soares, J.J. Leandro, R.M. Cesar, H.F. Jelinek, M.J. Cree, Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification, *IEEE Trans. Med. Imaging* 25 (9) (2006) 1214–1222.
- [52] S.M. Pizer, E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J.B. Zimmerman, K. Zuiderweld, Adaptive histogram equalization and its variations, *Comput. Vis. Graph. Image Process.* 39 (3) (1987) 355–368.
- [53] J.I. Orlando, E. Prokofyeva, M.B. Blaschko, A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images, *IEEE Trans. Biomed. Eng.* 64 (1) (2017) 16–27.
- [54] S. Ioffe, C. Szegedy, batch normalization: Accelerating deep network training by reducing internal covariate shift, in: International Conference on Machine Learning, 2015, pp. 448–456.
- [55] W.T. Freeman, E.H. Adelson, The design and use of steerable filters, *IEEE Trans. Pattern Anal. Mach. Intell.* 13 (9) (1991) 891–906.
- [56] A.F. Frangi, W.J. Niessen, K.L. Vincken, M.A. Viergever, Multiscale Vessel Enhancement Filtering, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, vol. 1496, 1998, pp. 130–137.
- [57] Y. Sasaki, The truth of the F-measure, *Teach. Tutor Mater.* (2007).
- [58] P. Jaccard, The distribution of the flora in the alpine zone, *New Phytol.* 11 (2) (2010) 37–50.
- [59] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, A. Lerer, Automatic differentiation in PyTorch, NIPS-W, 2017.
- [60] Q. Li, B. Feng, L.P. Xie, P. Liang, H. Zhang, T. Wang, A cross-modality learning approach for vessel segmentation in retinal images, *IEEE Trans. Med. Imaging* 35 (1) (2015) 109–118.
- [61] S. Roychowdhury, D.D. Koozekanani, K.K. Parhi, Blood vessel segmentation of fundus images by major vessel extraction and subimage classification, *IEEE J. Biomed. Health Inform.* 19 (3) (2017) 1118–1128.
- [62] Y. Chen, A labeling-free approach to supervising deep neural networks for retinal blood vessel segmentation, 2017, arXiv preprint [arXiv:1704.07502](https://arxiv.org/abs/1704.07502).
- [63] Z. Yan, X. Yang, K. Cheng, Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation, *IEEE Trans. Biomed. Eng.* 65 (9) (2018) 1912–1923.
- [64] J. Odstrcilik, R. Kolar, A. Budai, J. Hornegger, J. Jan, J. Gazarek, T. Kubena, P. Černosek, O. Svoboda, E. Angelopoulou, Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database, *IET Image Process.* 7 (4) (2013) 373–383.
- [65] J. Zhang, B. Dashtbozorg, E.J.E. Bekkers, J.P.W. Pluim, R.R. Duits, B.M.T.H. Romeny, Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores, *IEEE Trans. Med. Imaging* 35 (12) (2016) 2631–2644.
- [66] R. Estrada, C. Tomasi, S.C. Schmidler, S. Farsiu, Tree topology estimation, *IEEE Trans. Pattern Anal. Mach. Intell.* 37 (8) (2015) 1688–1701.
- [67] H. Zhao, H. Li, S. Maurerstroh, L. Cheng, Synthesizing retinal and neuronal images with generative adversarial nets, *Med. Image Anal.* 49 (2018) 14–26.
- [68] F. Muhammad Moazam, R. Paolo, H. Andreas, U. Bunyarit, A.R. Rudnicka, C.G. Owen, S.A. Barman, An ensemble classification-based approach applied to retinal blood vessel segmentation, *IEEE Trans. Biomed. Eng.* 59 (9) (2012) 2538–2548.