# Fundus Image Based Retinal Vessel Segmentation Utilizing a Fast and Accurate Fully Convolutional Network

Junyan Lyu, Pujin Cheng, and Xiaoying Tang[✉]

Southern University of Science and Technology, Shenzhen, China
tangxy@sustech.edu.cn

**Abstract.** Monitoring the condition of retinal vascular network based on a fundus image plays a vital role in the diagnosis of certain ophthalmologic and cardiovascular diseases, for which a prerequisite is to segment out the retinal vessels. The relatively low contrast of retinal vessels and the presence of various types of lesions such as hemorrhages and exudate nevertheless make this task challenging. In this paper, we proposed and validated a novel retinal vessel segmentation method utilizing *Separable Spatial and Channel Flow* and *Densely Adjacent Vessel Prediction* to capture maximum spatial correlations between vessels. Image pre-processing was conducted to enhance the retinal vessel contrast. Geometric transformations and overlapped patches were used at both training and prediction stages to effectively utilize the information learned at the training stage and refine the segmentation. Publicly available datasets including DRIVE and CHASE_DB1 were used to evaluate the proposed approach both quantitatively and qualitatively. The proposed method was found to exhibit superior performance, with the average areas under the ROC curve being 0.9826 and 0.9865 and the average accuracies being 0.9579 and 0.9664 for the aforementioned two datasets, which outperforms existing state-of-the-art results.

**Keywords:** Retinal vessel segmentation · Fully convolutional network · Dense Adjacently Vessel Prediction · Separable Spatial and Channel Flow · Fundus image

## 1 Introduction

Retinal vascular network is the only vasculature which can be visualized and photographed in vivo. Retinal vascular imaging is able to provide clinically prognostic information for patients with specific cardiovascular and ophthalomologic diseases [1]. Segmenting out the retinal vessels is a prerequisite for monitoring the condition of retinal vascular network. Currently, retinal vessel segmentation highly relies on the manual work of experienced ophthalmologists, which is tedious, time-consuming, and of low reproducibility. As such, a fully-automated and accurate retinal vessel segmentation method is urgently needed to reduce the

workload on ophthalmologists and provide objective and precise measurements of retinal vascular abnormalities.

Several factors make this task challenging. The lengths and calibers of the vessels vary substantially from subject to subject. The presence of various types of lesions including hemorrhages, exudate, microaneurysm and fibrotic band can be confused with the vessels, so do the retinal boundaries, optic disk, as well as fovea. Furthermore, the relatively low contrast of the vessels and the low quality of some fundus images further increase the segmentation difficulties.

Numerous methods have been proposed for retinal vessel segmentation, both unsupervised and supervised. Unsupervised methods typically rely on mathematical morphology and matched filtering [2]. In supervised methods, ground truth data is used to train a classifier based on pre-identified features to classify each pixel into either vessel or background [3]. In the past few years, deep learning methods have seen an impressive number of applications in medical image segmentation, being able to learn sophisticated hierarchy of features in an end-to-end fashion. For example, Ronneberger et al. proposed U-Net to perform cell segmentation, which has become a baseline network for biomedical image segmentation, including retinal vessel segmentation [4]. Liskowski et al. used a deep neural network containing Structured Prediction trained on augmented retinal vessel datasets for retinal vessel segmentation [5]. Oliveria et al. combined the multi-scale analysis provided by Stationary Wavelet Transform with a fully convolutional network (FCN) to deal with variations of the vessel structure [6]. These approaches applying deep learning methods have significantly outperformed previous ones, achieving higher segmentation accuracies and computational efficiencies.

Despite their significant progress, existing deep learning approaches are facing the dilemma of effectively extracting vessels with small calibers versus maintaining high accuracy. Aforementioned approaches suffer from low capabilities of detecting thin vessels. Zhang et al. introduced an edge-aware mechanism by adding additional labels, which yielded a considerable improvement on predicting thin vessels but a decreased overall accuracy [7]. This is due to the fact that FCNs do not make use of spatial information in the pixel-wise prediction stage, but deploy a fully connected layer to each pixel separately. In such context, we propose a novel method for fundus image based retinal vessel segmentation utilizing a FCN together with *Separable Spatial and Channel Flow (SSCF)* and *Dense Adjacently Vessel Prediction (DAVP)* to capture maximum spatial correlations between vessels. Geometric transformations and overlapped patches are used at both training and prediction stages to effectively utilize the information learned at the training stage and refine the segmentation. Our method is quantitatively and qualitatively evaluated on the Digital Retinal Images for Vessel Extraction (DRIVE) [8] and Child Heart and Health Study in England (CHASE_DB1) datasets [9].

## 2   Method

### 2.1   Image Pre-processing

Although convolutional neural networks (CNNs) can effectively learn from raw image data, clear information and low noise enable CNNs to learn better. For a fundus image, its green channel is often used since it shows the best contrast in physiological structures with blood. And the blue channel contains relatively little physiological information. As such, simple but effective pre-processing is applied via $p_{i,j} = 0.25r_{i,j} + 0.75g_{i,j}$, where $p_{i,j}$ denotes the resulting pixel value at position $(i,j)$ and $r_{i,j}$, $g_{i,j}$ respectively stand for the red channel value and the green channel value. Data augmentation is conducted to enlarge the training set by rotating each image patch by $90°$, $180°$ and $270°$.

### 2.2   Patch Extraction

Several studies have shown that CNNs can benefit from using overlapped patches extracted from large images for tasks that ignore contextual information [4, 6,7,11]. Given that the mean and variance of small image patches within a fundus image differ little, overlapped patches can also be applied to retinal vessel segmentation. Furthermore, a large amount of image patches can boost a CNN's performance by enlarging the sample size. In this work, a total of 1000 image patches of size $48 \times 48$ are randomly sampled from each training image. Center sampling is used and zero padding is performed if the center is located on image boundaries.

### 2.3   Fully Convolutional Network

FCNs can take input of arbitrary size and produce an output of the corresponding size with efficient inference and training by local connectivity and weight sharing [4]. FCNs typically have both down-sampling and up-sampling modules, which are used to respectively extract multi-scale features and reconstruct spatial information. In this work, we use U-Net as our baseline framework, which employs multiple skip connections to refine the detailed information lost in up-sampling modules [4]. As shown in Fig. 1, our overall architecture includes five stages. Extraction stage extracts low-level information from input images. Projection stage gradually projects multi-scale features into low-resolution feature maps that lie in high dimensional spaces. Mapping stage performs several non-linear mappings to explore more semantic information, providing guidance for pixels with low contrast and intensity. Refinement stage embeds spatial information into feature maps that have rich high-level information. By concatenating feature maps, this stage refines the semantic boundary. Reconstruction stage utilizes refined features to perform predictions, producing segmentation results.
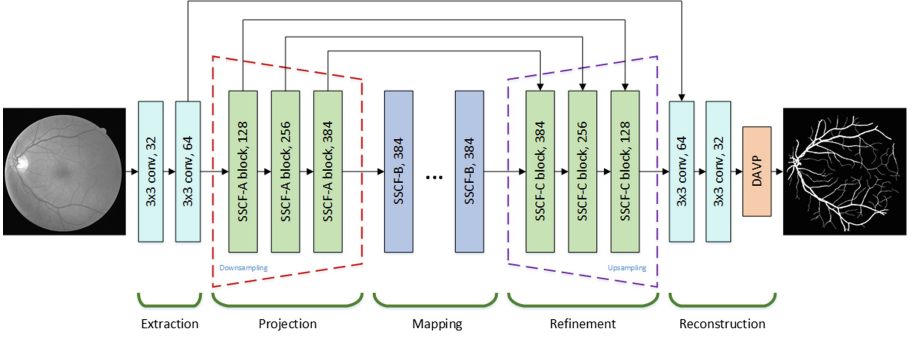
**Fig. 1.** Overall architecture of the proposed network. *SSCF-A*, *SSCF-B*, *SSCF-C* are detailed in Fig. 2. Please note, at Mapping stage, there can be any number of *SSCF-B* blocks, and we use two in our proposed network.

## 2.4   Separable Spatial and Channel Flow

Convolutional layers are designed to learn filters in a 3-dimension space (two spatial dimensions and one channel (grayscaled image intensity) dimension). Thus, a single convolution kernel should perform spatial and channel transformations jointly. However, filters in a single kernel usually conduct these two tasks implicitly, which may be vague and inefficient for high dimensional spaces. As such, we decouple the mappings of spatial correlations and cross-channel correlations sufficiently by factoring a kernel into a series of operations to perform those two mappings separately [10]. Specifically, we propose a block called *Separable Spatial and Channel Flow (SSCF)* and apply it to Projection, Mapping and Refinement stages, as shown in Fig. 2. Three depth-wise separable convolutional layers and one residual connection are contained in a *SSCF* block. Each depth-wise separable convolutional layer performs a depth-wise convolution followed by a point-wise convolution. A depth-wise convolution works as:

$$\mathbf{p}_{i,j,k} = \sum_{0 \le l,m < s} \mathbf{x}_{i+l-\lceil \frac{s}{2} \rceil, j+m-\lceil \frac{s}{2} \rceil, k} W_{i+l-\lceil \frac{s}{2} \rceil, j+m-\lceil \frac{s}{2} \rceil, k} \tag{1}$$

where $\mathbf{x}_{i,j,k}$, $\mathbf{p}_{i,j,k}$ respectively denote the input and result at position $(i, j)$ and channel $k$, $W_{i,j,k}$ denotes the corresponding weight for $\mathbf{x}_{i,j,k}$ and $s$ denotes the filter size. A point-wise convolution is written as:

$$\mathbf{y}_{i,j,c} = \sum_{k=1}^{C} \mathbf{p}_{i,j,k} W_{i,j,k,c} \tag{2}$$

where $C$ denotes the number of input channels, $\mathbf{p}_{i,j,k}$ stands for the input at position $(i, j)$ and channel $k$, $\mathbf{y}_{i,j,c}$ denotes the output at position $(i, j)$ and channel $c$ which can be any integer no larger than the total number of output channels. $W_{i,j,k,c}$ stands for the corresponding weight at position $(i, j)$ for input channel $k$ and output channel $c$. By stacking depth-wise separable convolutional

layers, *SSCF* not only enables spatial and channel information within the feature maps to flow separately, but also reduces redundant parameters and computational complexity. Furthermore, a residual shortcut between feature maps with smaller semantic and resolution gap can provide a better feature fusion.

## 2.5    Dense Adjacently Vessel Prediction

Despite the fact that the channel-wise spatial relations have an impact on prediction, pixels are classified individually in FCN. A pixel representing background is less likely to be surrounded by pixels of vessels since retinal vasculature is structurally continuous [5]. Inspired by this prior knowledge, we propose a dense prediction cell named *Dense Adjacently Vessel Prediction (DAVP)*, as shown in Fig. 2. In addition to a $1 \times 1$ convolutional layer, an extra $5 \times 5$ convolution with non-linearity is introduced to filter redundant information. Then another $5 \times 5$ convolution utilizes spatial relations between pixels to perform prediction, refining the result from a $1 \times 1$ convolution branch via an element-wise addition.
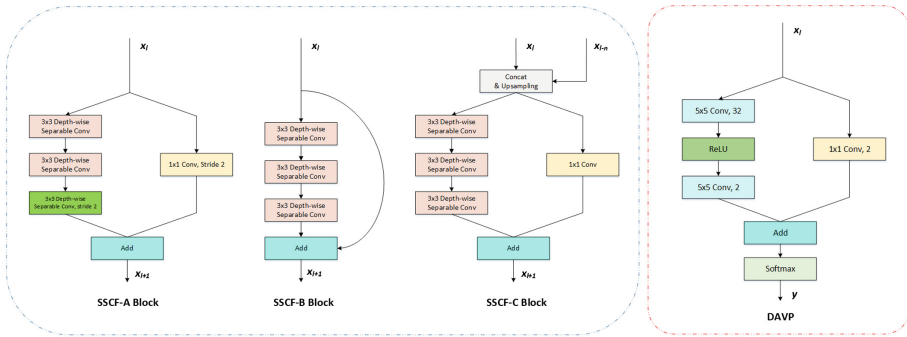


**Fig. 2.** Details of *SSCF* and *DAVP* employed in the proposed method.

# 3    Experiments

## 3.1    Datasets

We evaluated our method on the DRIVE and CHASE_DB1 public datasets. DRIVE consists of 40 fundus images of size $584 \times 565$ taken from both healthy adults and adults with mild diabetic retinopathy. There are 20 images for training and 20 images for testing [8]. CHASE_DB1 consists of 28 fundus images of size $999 \times 960$ taken from 14 10-year old children. For both datasets, gold standard segmentations are available. Since there is no official division into training and testing sets for CHASE_DB1 [9], we performed a 4-fold cross-validation in this case. The field of view masks for both datasets are publicly available [6], on which our quantitative evaluations are conducted.

## 3.2   Training

During the training process, Adam Optimizer was used to minimize cross entropy loss:

$$J(y, p) = -\sum_{k=1}^{C} y_k log(p_k) \tag{3}$$

where C refers to the number of classes, $p$ and $y$ respectively denote the probabilistic prediction and ground truth. The learning rate decayed by half for every 10 epochs, with an initial value of 0.001. The network was trained for 50 epochs, taking less than 1 h.

## 3.3   Implementation Details

The proposed method was implemented utilizing Keras with Tensorflow backend. All training and testing experiments were conducted on a workstation equipped with NVIDIA GTX Titan Xp.

## 3.4   Quantitative Results

To compare with other state-of-the-art results, we used four metrics for evaluation: accuracy (Acc), sensitivity (Sn), specificity (Sp) and area under the ROC curve (AUC-ROC). AUC-ROC is the key metric in retinal vessel segmentation considering the imbalance of classes. To obtain the binary vessel segmentation, a threshold of 0.5 is applied to the probability map.

Table 1 demonstrates the performance gains obtained from *SSCF* and *DAVP*, as evaluated on the DRIVE dataset. By decoupling the mappings of cross-channel correlations and spatial correlations, *SSCF* boosts the performance of U-Net, achieving an improvement on AUC-ROC by 0.09%. An incorporation of *DAVP* further improves the predictions by taking neighboring pixels into consideration during classification. These results imply *SSCF* and *DAVP* are helpful for embedding more spatial information between vessels. Figure 3 visualizes how *SSCF* and *DAVP* work.



**Fig. 3.** A test image patch from the DRIVE dataset. From left to right: one of the original image patches, the segmentation results from the baseline model, the baseline + *SSCF* model, the baseline + *SSCF* + *DAVP* model (the proposed) and the ground truth.

**Table 1.** Performance comparisons of different models on DRIVE.

| Method | AUC-ROC |
|---|---|
| Baseline(U-Net) | 0.9796 |
| Baseline+preprocessing | 0.9809 |
| Baseline+preprocessing+$SSCF$ | 0.9818 |
| Baseline+preprocessing+$SSCF$+$DAVP$ | **0.9826** |

We also compare our method with several other state-of-the-art methods in Tables 2 and 3. Our method outperforms all the other methods in terms of both accuracy and AUC-ROC. Also, reducing redundant parameters via the depthwise separable convolutions strikingly shortens the training and inference time, making the proposed method being 90% faster than existing methods. Figure 4 shows representative segmentation results obtained from the proposed method on CHASE_DB1.

**Table 2.** Performance comparison with state-of-the-art methods on DRIVE, where $t_T$ and $t_I$ respectively denote time consumptions in training and inference.

| Method | AUC-ROC | Acc | Sn | Sp | $t_T$ | $t_I$ |
|---|---|---|---|---|---|---|
| 2nd observer [8] | N.A. | 0.9473 | 0.7760 | 0.9725 | N.A. | N.A. |
| Gu et al. [12] | 0.9779 | 0.9545 | **0.8309** | N.A. | N.A. | N.A. |
| Liskowski et al. [5] | 0.9790 | 0.9535 | 0.7811 | 0.9807 | 8 h | 92 s |
| Wu et al. [11] | 0.9807 | 0.9567 | 0.7844 | 0.9817 | 16 h | 10 s |
| Oliveria et al. [6] | 0.9821 | 0.9576 | 0.8039 | 0.9804 | N.A. | N.A. |
| Proposed method | **0.9826** | **0.9579** | 0.7940 | **0.9820** | **1 h** | 1.3 s |

**Table 3.** Performance comparison with state-of-the-art methods on CHASE_DB1.

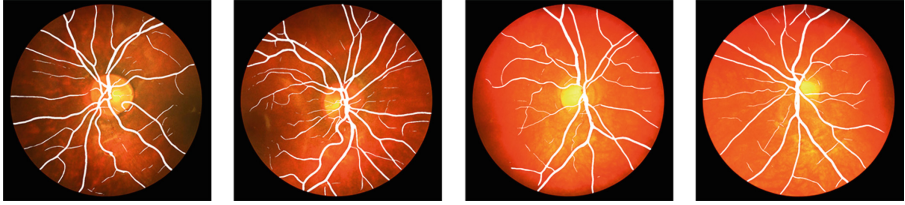| Method | AUC-ROC | Acc | Sn | Sp |
|---|---|---|---|---|
| 2nd observer [9] | N.A. | 0.9560 | 0.7686 | 0.9779 |
| Wu et al. [11] | 0.9825 | 0.9637 | 0.7538 | 0.9847 |
| Oliveria et al. [6] | 0.9855 | 0.9653 | 0.7779 | 0.9864 |
| Proposed method | **0.9865** | **0.9664** | **0.7878** | **0.9865** |

**Fig. 4.** Representative segmentation results superimposed on the original fundus images, obtained from the proposed method on CHASE_DB1.

## 4   Conclusion

In this paper, we proposed a novel FCN by incorporating *SSCF* and *DAVP* into U-Net for segmenting retinal vessels. The proposed *SSCF* and *DAVP* blocks can capture maximum spatial correlations between vessels, being able to solve the dilemma of maintaining high segmentation accuracy versus effectively extracting thin vessels. We demonstrated that the proposed method has state-of-the-art segmentation performance and high computational efficiency, which are essential in practical clinical applications. Future work will involve applying the proposed method to large-scale clinical studies.

## References

1. Patton, N., et al.: Retinal vascular image analysis as a potential screening tool for cerebrovascular disease: a rationale based on homology between cerebral and retinal microvasculatures. J. Anat. **206**(4), 319–348 (2005)
2. Hoover, A., Kouznetsova, V., Goldbaum, M.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans. Med. Imaging **19**(3), 203–210 (2000)
3. Zhang, J., et al.: Retinal vessel delineation using a brain-inspired wavelet transform and random forest. Pattern Recogn. **69**, 107–123 (2017)
4. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
5. Liskowski, P., Krawiec, K.: Segmenting retinal blood vessels with deep neural networks. IEEE Trans. Med. Imaging **35**(11), 2369–2380 (2016)
6. Oliveira, A., Pereira, S., Silva, C.A.: Retinal vessel segmentation based on fully convolutional neural networks. Expert Syst. Appl. **112**, 229–242 (2018)
7. Zhang, Y., Chung, A.C.S.: Deep supervision with additional labels for retinal vessel segmentation task. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 83–91. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_10

8.  Staal, J., et al.: Ridge-based vessel segmentation in color images of the retina. IEEE Trans. Med. Imaging **23**(4), 501–509 (2004)

9.  Owen, C.G., Rudnicka, A.R., Mullen, R., Barman, S.A., et al.: 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), 2004–2010 (2009)

10. Chollet, F.: Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on CVPR, pp. 1251–1258 (2017)

11. Wu, Y., Xia, Y., Song, Y., Zhang, Y., Cai, W.: Multiscale network followed network model for retinal vessel segmentation. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 119–126. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_14

12. Gu, Z., et al.: CE-Net: context encoder network for 2D medical image segmentation. IEEE Trans. Med. Imaging (2019, in press)