Coarse to Fine Retinal Vessel Segmentation via Convolutional Neural Networks with Attention Model

Yijun Yuan

***Abstract*—Retinal vessel segmentation is a fundamental task for various ocular imaging applications. In this paper, we propose a coarse-to-fine strategy for retinal vessel segmentation via Convolutional Neural Networks with attention model. Our framework consists of a coarse stage and a fine stage, where the coarse stage proposes a rough estimation of the retinal vessel, and the fine stage refines the output of the coarse stage by reducing the false positive. We further deploy an attention model into the Convolutional Neural Networks by taking the importance of features from different layers into consideration. Extensive experiments on three publicly available datasets demonstrate the effectiveness of our method for automatic retinal blood vessel segmentation.**

***Index Terms*—Retinal vessel segmentation, Attention model, Coarse-to-fine**

I. Introduction

It is key for ophthalmologists to understand retinal vessels, when evaluating and monitoring various eye diseases, such as diabetes, hypertension, glaucoma, macular degeneration, and hypertension, among others. However, manual segmentation of retinal vessels is both tedious and time-consuming. Therefore, many approaches have been proposed to do segmentation of retinal blood vessels from fundus images automatically.

Early blood vessel segmentation works [1][2][3][4][5] mainly use handcrafted features or heuristic assumptions including image filters, vector geometry, statistical distribution studies, and photon distribution models for vessel detection. In recent, [6] proposed a multi-scale line detection scheme to compute vessel segmentation. [7] proposed a structured-output support vector machine with a fully-connected Conditional Random Field (CRF).

It has been increasingly popular to use deep networks for Retinal Image Understanding. Recently, in the realm of retinal vessel segmentation, [8] proposed cross-modality data transformation from retinal image to vessel map, and outputted the label map of all pixels for a given image patch. [9] proposed a multi-scale and multi-level CNN with Conditional Random Field (CRF) to solve retinal vessel segmentation problem. [10] used an ensemble of deep convolutional neural networks to segment vessel and non-vessel areas of a color fundus image. [11] presents Deep Retinal Image Understanding (DRIU), a unified framework of retinal image analysis that provides both retinal vessel and optic disc segmentation.

Motivated by the successes of DNNs based retinal blood segmentation, in this paper, we employ a coarse to fine strategy for retinal vessel. The coarse stage, take the fundus images to produce a vessel probability map. Based on the prediction from the coarse stage, fine stage combines the vessel map with original fundus image, and reuse the network in coarse stage to learn the mapping between these combinations and final retinal vessel map.

In recent years, attention models have been applied widely for computer vision tasks [12][13][14]. With the help of attention, the models could pay attention to the most needed features, rather than a whole image. In this work, we employ the attention model for vessel segmentation in spatial dimension.

The contributions of our paper are two-fold: i) To our knowledge, it is the first work to use coarse to fine strategy for retinal vessel segmentation; ii) In each stage, we employ attention model to train retinal vessel segmentation. Experimental results demonstrate the effectiveness of our framework.

II. Our Method

We here employed a Coarse To Fine strategy for retinal vessel segmentation. The Coarse stage is used to estimate approximate retinal vessel maps, and then Fine stage is designed to refine the predictions with the help from previous estimations.

A. Coarse Stage

The overall structure is shown in Fig. 1 and it consists of a series of 4 convolution blocks, attention model and CRF layer.

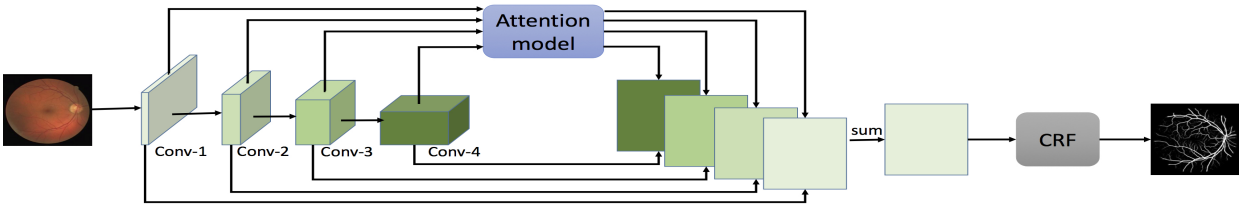


Fig. 1. The structure of Coarse stage. In the Fine stage, we replace the input of raw retinal image with the concatenation of retinal image and coarse stage output.

After each of the four convolution blocks, there is a maxpool layer. We use ReLU as activation function.

In the HED[15] architecture, the intermediate CNN feature should be combined before further processing. Here we use attention model to merge the multi-level features output from CONV-1 to CONV-4. Suppose the CNN features has several levels $s = 1, ..., S$. And each of these features is upsampled to the same size of original image. f_i^s indicates a pixel i ($i = 1, ..., N$) in the feature of level s, where N is the number of pixels for a resized feature. U_i denotes the weighted sum of features at pixel i for all levels:

$$U_i = \sum_{s=1}^{S} \omega_i^s \cdot f_i^s \tag{1}$$

The weight w_i^s is

$$\omega_i^s = \frac{\exp(h_i^s)}{\sum_{t=1}^{S} \exp(h_i^t)} \tag{2}$$

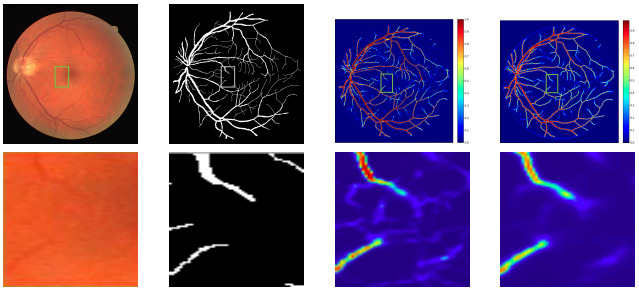
where h_i^s is the last layer output before softmax produced by attention model at pixel i for scale s.

Here our attention model takes the features that have been upsampled and have $16\times$ 4 channels in total as input. And it consists of two layers. The first layer has 16 filters with kernel size 3×3 and the second has 4 filters with kernel size 1×1.

The CRF stage takes the context into acount, to learn the non-local pixel correlation. The CNN has some shortages on structured prediction. CNN has convolutional kernels with large receptive field and will provide a coarse prediction. Moreover, since the CNN does not consider the inter-pixel correlation, the probability map may contain spurious regions in the segmentation result[9]. Therefore, in accordance with the fully-connected CRF model [16], the correlation between each pair of nodes will be concerned. Our proposed net use the implementation of [17] where a RNN layer can be utilized as CRF in end to end neural network.

Considering the retinal image has imbalanced foreground and background rate, following HED [15], we use a class-balanced cross-entropy loss function:

$$L(W) = -\frac{|Y^-|}{|Y|} \sum_{i \in Y^+} \log(\sigma(\alpha_i)) - \frac{|Y^+|}{|Y|} \sum_{i \in Y^-} \log(1 - \sigma(\alpha_i)) \tag{3}$$



(a) Fundus images (b) ground truth (c) Coarse stage result (d) Fine stage result

Fig. 2. The qualitive comparison between the results of coarse stage and fine stage. Row1: full size image, Row2: ROI(Green Box region of above image); From left to right are (a)fundus image, (b)groundtruth, (c)Vessel map of our Coarse stage and (d)Vessel map of our Fine stage. To have a better visualization, we use Density plot to draw the output map.

where W denote the standard set of parameters of the Network. Y^+ and Y^- denote the vessel and background pixels in the ground truth Y , and $\sigma(\alpha_i)$ is the sigmoid function on pixel i, α_i of the predicted map $X, , i = 1, ..., |Y|$.

B. Fine Stage

The Coarse stage takes a whole retinal image as input, and obtains an approximate estimation of retinal vessel maps. As shown in Fig.2, the approximate estimations tend to contain the noise from background. Sequently, we employ a FineNet to obtain the final pixel-accurate segmentation prediction to deal with this challenge.

In FineNet, we consider the 4 channel combination of original image(RGB) and "strong suggestion"(coarse stage result) as input. Then reuse the network illustrate in Fig.1 to provide a refined segmentation. The improvement is shown in Fig.2, we can observe that FineNet base on the previous result to provide a refined segmentation that restrained the appearance of background noise.

III. Experimental Evaluation

A. Experimental Setup

1) Datasets and Measurements: The datasets used in this experiment are ARIA, DRIVE, STARE. For ARIA, 134 labeled samples are provided, and it is utilized to fine tune our model. For DRIVE, 40 labeled samples are published and it provides 20 for training and the rest 20 for test. For STARE, note that there is no training set, it will only be used for testing.

TABLE I
COMPARISONS OF DIFFERENT COMPONENTS IN OUR FRAMEWORK ON STARE(For Acc and Spe, higher is better. For FPR, lower is better.)

	Acc	Spe	FPR
Our Network w.o attention	0.9648	0.9811	0.0189
Our Network w attention	0.9652	0.9832	0.0168
Coarse stage	0.9652	0.9832	0.0168
Fine stage	0.9670	0.9871	0.0129

TABLE II
COMPARISON WITH STATE-OF-ART METHODS ON DRIVE

	Acc	Spe	FPR
2nd human observer	0.9472	0.9725	0.0275
Staal [2]	0.9441	0.9773	0.0227
Mendonca [18]	0.9452	0.9764	0.0236
Marin [19]	0.9452	0.9801	0.0199
Fraz [20]	0.9480	0.9807	0.0193
Nguyen [21]	0.9407	-	-
Zhao [22]	0.9477	0.9789	0.0211
Melinscak [23]	0.9466	0.9785	0.0215
Azzopardi [24]	0.9442	0.9704	0.0296
Roychowdhury [25]	0.9494	0.9782	0.0218
Fu [9]	0.9523	-	-
Our method	0.9566	0.9890	0.0110

We follow [26] use Accuracy ($Acc = \frac{TP+TN}{TP+FN+TN+FP}$), Specificity ($Spe = \frac{TN}{TN+FP}$) and False Positive Rate($FPR = \frac{FP}{FP+TN}$) to evaluate the performance of our network for retinal vessel segmentation, where TP , TN , FP and FN represent the number of true positives, true negatives, false positives and false negatives, respectively. Higher Acc and Spe correspond to better performance. Lower FPR corresponds to better performance.

2) Parameters: Our method is implemented based on the CAFFE framework developed by Jia *et al.* [27]. During training, the first four convolution blocks (CONV-1 to CONV-4) of our network are initialized using 5stage-vgg.caffemodel provided by [15]. The rest parameters in our network are initialized with xavier. To avoid over-fitting and to improve the robustness of our model, we rotate each training retinal image in 8 angle.

During the training of Coarse stage, we follow [9] to employed a two-step approach. First we enforce the images of ARIA to have the same size of images as DRIVE and the network is finetuned on ARIA. Then we use DRIVE training set to fine tune sequentially to get the model. It take about 1 day(3000 iterations with batch size 4) to train the CoarseNet. With the help of the estimation from Coarse stage, we train Fine stage network both on ARIA and DRIVE datasets. It also takes about 1 day(2600 iteration with batch size 4).

B. Evaluation of Different Components in Our Method

The following experiment is designed to measure the functions of different components in our framework. We follow the experimental setup as described in III-A. All the following experiments follow the same setting unless a different setting is specified. To generate a binary map, we threshold the output probability map by 0.5, for which the positive decision(Vessel) is made if the probability is greater than the 0.5 threshold; otherwise negative decision is made(Background).

1) With Attention vs. Without attention: In order to understand the effect of attention model, we train relevant parts of the network independently for the network with attention model and one without attention model. We compare the results from these independently trained models on STARE dataset(0.5 thresholding). As shown in Tab. I, it illustrates that network with attention model performs better on those metrics than the other one for retinal vessel segmentation. This is because the network with attention leverages the advantages of the feature information from different levels.

2) CoarseNet vs. FineNet: The quantitative comparison of CoarseNet and FineNet is shown in Tab. I(0.5 thresholding). This clearly shows that Fine Stage improves upon the performance of Coarse Stage. Fig. 2 shows some examples of qualitative comparison between coarse stage and fine stage. We can observer the fine stage could help remove the unnecessary predicted branch to provide a more accurate prediction. The FineNet takes a "Strong Suggestion" from Coarse stage, and the performance of FineNet strongly depends on the estimations of Coarse stage.

3) Time costs: We test the running time of our method on DRIVE datasets. Our algorithm is implemented on TITANX GPU. We run our program 20 times for each 565×584 image, obtain the average running time 0.459s for Coarse stage and 0.505s for Fine stage.

TABLE III
COMPARISON WITH STATE-OF-ART METHODS ON STARE

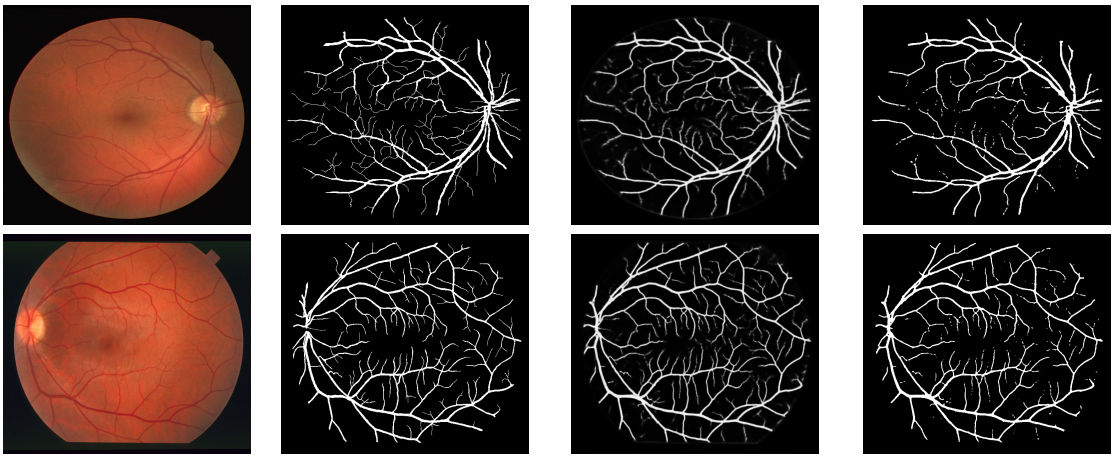
	Acc	Spe	FPR
2nd human observer	0.9349	0.9390	
Staal [2]	0.9516	0.981	0.0190
Mendonca [18]	0.9440	0.973	0.0270
Marin [19]	0.9526	0.9819	0.0181
Fraz [20]	0.9534	0.9763	0.0237
Nguyen [21]	0.9326	-	-
Zhao [22]	0.9509	0.9767	0.0233
Melinscak [23]	-	-	-
Azzopardi [24]	0.9497	0.9701	0.0299
Roychowdhury [25]	0.9560	0.9842	0.0158
Fu [9]	0.9585	-	-
Ours method	0.9670	0.9871	0.0129

C. Comparisons on Retinal Vessel Segmentation

We compare our method with several state-of-the-art vessel segmentation methods. We employ a 0.5 thresholding method on our final output to obtain the binary image to compare with those state-of-the-art. Tab. II and III list the performances on the two datasets. We also show some examples about the final predictions of our method in Fig. 3. Our method obtains better scores among the methods.

IV. CONCLUSION

In this paper, we have developed a retinal vessel segmentation method, based on a novel deep learning architecture. We employ coarse-to-fine strategy for retinal vessel segmentation. The attention model can leverage the advantages of the feature



(a) Fundus images (b) ground truth (c) Our final result map (d) Thresholded result

Fig. 3. Examples of final output. Row 1, row 2 are the fundus images from DRIVE and STARE, respectively. From left to right are (a)fundus image, (b)gound truth, (c)Our final result, (d)Thresholded result

information from different levels. We have demonstrated that our system produces state-of-the-art results on two publicly available datasets.

REFERENCES

[1] M. D. Abramoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Reviews in Biomedical Engineering*, vol. 3, p. 169, 2010.

[2] J. Staal, M. D. Abrmoff, M. Niemeijer, M. A. Viergever, and G. B. Van, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–9, 2004.

[3] M. Sofka and C. V. Stewart, "Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures," *IEEE Transactions on Medical Imaging*, vol. 25, no. 12, p. 1531, 2006.

[4] J. Jan, J. Odstrcilik, J. Gazarek, and R. Kolar, "Retinal image analysis aimed at blood vessel tree segmentation and early detection of neural-layer deterioration," *Computerized Medical Imaging & Graphics*, vol. 36, no. 6, p. 431, 2012.

[5] D. Sheet, S. P. K. Karri, S. Conjeti, S. Ghosh, J. Chatterjee, and A. K. Ray, "Detection of retinal vessels in fundus images through transfer learning of tissue specific photon interaction statistical physics," in *IEEE International Symposium on Biomedical Imaging*, 2013, pp. 1452–1456.

[6] U. T. V. Nguyen, A. Bhuiyan, L. A. F. Park, and K. Ramamohanarao, "An effective retinal blood vessel segmentation method using multi-scale line detection," *Pattern Recognition*, vol. 46, no. 3, pp. 703–715, 2013.

[7] J. I. Orlando and M. Blaschko, "Learning fully-connected crfs for blood vessel segmentation in retinal images," in *Miccai*, 2014, pp. 634–641.

[8] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, "A cross-modality learning approach for vessel segmentation in retinal images," *IEEE transactions on medical imaging*, vol. 35, no. 1, pp. 109–118, 2016.

[9] H. Fu, Y. Xu, S. Lin, D. W. K. Wong, and J. Liu, "Deepvessel: Retinal vessel segmentation via deep learning and conditional random field," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 132–139.

[10] D. Maji, A. Santara, P. Mitra, and D. Sheet, "Ensemble of deep convolutional neural networks for learning to detect retinal vessels in fundus images," 2016.

[11] K. K. Maninis, J. Ponttuset, P. Arbelez, and L. V. Gool, "Deep retinal image understanding," 2016.

[12] V. Mnih, N. Heess, A. Graves *et al.*, "Recurrent models of visual attention," in *Advances in neural information processing systems*, 2014, pp. 2204–2212.

[13] K. Chen, J. Wang, L.-C. Chen, H. Gao, W. Xu, and R. Nevatia, "Abc-cnn: An attention based convolutional neural network for visual question answering," *arXiv preprint arXiv:1511.05960*, 2015.

[14] L.-C. Chen, Y. Yang, J. Wang, W. Xu, and A. L. Yuille, "Attention to scale: Scale-aware semantic image segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3640–3649.

[15] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1395–1403.

[16] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," in *Advances in neural information processing systems*, 2011, pp. 109–117.

[17] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr, "Conditional random fields as recurrent neural networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1529–1537.

[18] A. M. Mendonca and A. Campilho, "Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction," *IEEE transactions on medical imaging*, vol. 25, no. 9, pp. 1200–1213, 2006.

[19] D. Marín, A. Aquino, M. E. Gegúndez-Arias, and J. M. Bravo, "A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features," *IEEE Transactions on medical imaging*, vol. 30, no. 1, pp. 146–158, 2011.

[20] F. Muhammad Moazam, R. Paolo, H. Andreas, U. Bunyarit, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 9, pp. 2538–48, 2012.

[21] U. T. Nguyen, A. Bhuiyan, L. A. Park, and K. Ramamohanarao, "An effective retinal blood vessel segmentation method using multi-scale line detection," *Pattern recognition*, vol. 46, no. 3, pp. 703–715, 2013.

[22] Y. Q. Zhao, X. H. Wang, X. F. Wang, and F. Y. Shih, "Retinal vessels segmentation based on level set and region growing," *Pattern Recognition*, vol. 47, no. 7, pp. 2437–2446, 2014.

[23] M. Melinščak, P. Prentašić, and S. Lončarić, "Retinal vessel segmentation using deep neural networks," in *VISAPP 2015 (10th International Conference on Computer Vision Theory and Applications)*, 2015.

[24] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, "Trainable cosfire filters for vessel delineation with application to retinal images," *Medical image analysis*, vol. 19, no. 1, pp. 46–57, 2015.

[25] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, "Iterative vessel segmentation of fundus images," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 7, pp. 1738–1749, 2015.

[26] M. Kaur, "Retinal vessel extraction by using local entropy based thresholding and directional filters," 2015.

[27] Jia, Yangqing, Shelhamer, Evan, Donahue, Jeff, Karayev, Sergey, Long, and Jonathan, "Caffe: Convolutional architecture for fast feature embedding," *Eprint Arxiv*, pp. 675–678, 2014.