Guided M-Net for High-Resolution Biomedical Image Segmentation with Weak Boundaries

Shihao Zhang[1], Yuguang Yan[1], Pengshuai Yin[1], Zhen Qiu[1], Wei Zhao[2], Guiping Cao[2], Wan Chen[3], Jin Yuan[3], Risa Higashita[4], Qingyao Wu[1], Mingkui Tan[1(✉)], and Jiang Liu[5,6]

[1] South China University of Technology, Guangzhou, China
[2] CVTE Research, Guangzhou, China
[3] Zhongshan Ophthalmic Center Sun Yat-sen University, Guangzhou, China
[4] Tomey Corporation, Nagoya, Japan
[5] Southern University of Science and Technology, Shenzhen, China
[6] Cixi Institute of BioMedical Engineering, Ningbo Institute of Industrial Technology, Chinese Academy of Sciences, Beijing, China

Abstract. Biomedical image segmentation plays an important role in automatic disease diagnosis. However, some particular biomedical images have blurred object boundaries, and may contain noises due to the limited performance of imaging device. This issue will highly affects segmentation performance, and will become even severer when images have to be resized to lower resolution on a machine with limited memory. To address this, we propose a guide-based model, called G-MNet, which seeks to exploit edge information from guided map to guide the corresponding lower resolution outputs. The guided map is generated from multi-scale input to provide a better guidance. In these ways, the segmentation model will be more robust to noises and blurred object boundaries. Extensive experiments on two biomedical image datasets demonstrate the effectiveness of the proposed method.

1 Introduction

Biomedical image segmentation plays important role in automatic disease diagnosis. In particular, in glaucoma screening, correct optic disc (OD) and optic cup (OC) segmentation will help obtain an accurate vertical cup-to-disc ratio (CDR), which is commonly used for glaucoma diagnosis. Moreover, in cataract grading, lens structure segmentation helps to calculate the density of different lens parts, and the density quantification is a kind of cataract grading metric [11].

In recent years, Convolutional neural networks (CNNs) have shown strong power in biomedical image segmentation with remarkable accuracy. For example, [9] proposes a U-shape convolutional network (U-Net) to segment images

This work was done when S. Zhang and Y. Yan are interns at CVTE Research.

H. Fu et al. (Eds.): OMIA 2019, LNCS 11855, pp. 43–51, 2019.
https://doi.org/10.1007/978-3-030-32956-3_6

with precise boundaries by constructing skip connections to restore the information loss caused by pooling layers. [5] proposes an M-shape convolutional network, which combines multi-scale inputs and constructs local outputs to link the loss and early layers. In practice, however, some high-resolution biomedical images have noises and blurred boundaries, like the anterior segment optical coherence tomography (AS-OCT) images, which may hamper the segmentation performance, as shown in Fig. 1. Furthermore, suffering from the limitation of memory, existing methods usually receive down-sampled images as input and then up-sample the results back to the original resolution, which, however, may lead even worse segmented boundaries.

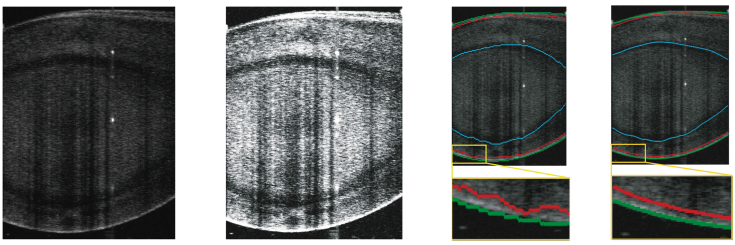

Fig. 1. (a): An AS-OCT image sample with weak nucleus and cortex boundaries. (b): corresponding histogram equalized image with a lot of noise. (c): segmentation results of M-Net with low-resolution input. (d): segmentation results of G-MNet.

To address the above issues and hence improve the segmentation performance, we seek to exploit guided filter to extract edge information from high-resolution images. In this way, high-quality segmentation results can be generated from low-resolution poorly segmented results. Moreover, precise segmented boundaries can be maintained after up-sampling. Guided filter [6] is an edge-preserving image filter and has been incorporated into deep learning on several tasks. For example, [12] formulates it into an end-to-end trainable module, [7] combines it with superpixels to decrease computational cost. Different from existing works which use guided filter as post-processing, we incorporate the guided filter into CNNs to learn better features for segmentation.

Unfortunately, the performance of the guided filter will be affected by noises and blurred boundaries in images. Therefore, better guidance rather than the original image is required. In this sense, we design a guided block to produce an informative guided map, which helps to alleviate the influence of noises and blurred boundaries. Besides, multi-scale features and multi-scale inputs are also combined to make model more robust to noise. Thorough experiments on two benchmark datasets, namely CASIA-2000 and ORIGA datasets, demonstrate the effectiveness of our method. Our method also achieves the best performance on CASIA-2000 dataset and outperforms the state-of-the-art OC and/or OD segmentation methods on ORIGA dataset.

2 Methodology

In this section, we provide an overview of our guide-based model, named G-MNet, in Fig. 2. Then introduce its three components: an M-shape convolutional network (M-Net) to learn hierarchical representations, a guided block for better guidance, and a multi-guided filtering layer to filter multi-scale low-resolution outputs. Our G-MNet firstly generates multi-scale side-outputs by M-Net, then these side-outputs are filtered to high-resolution through the multi-guided filtering layer. The guided block is exploited to provide better guidance for the multi-guided filtering layer. After that, an average layer is employed to combine all the high-resolution outputs. At last, the multi-guided filter receives the combined outputs and produces the final segmentation result.

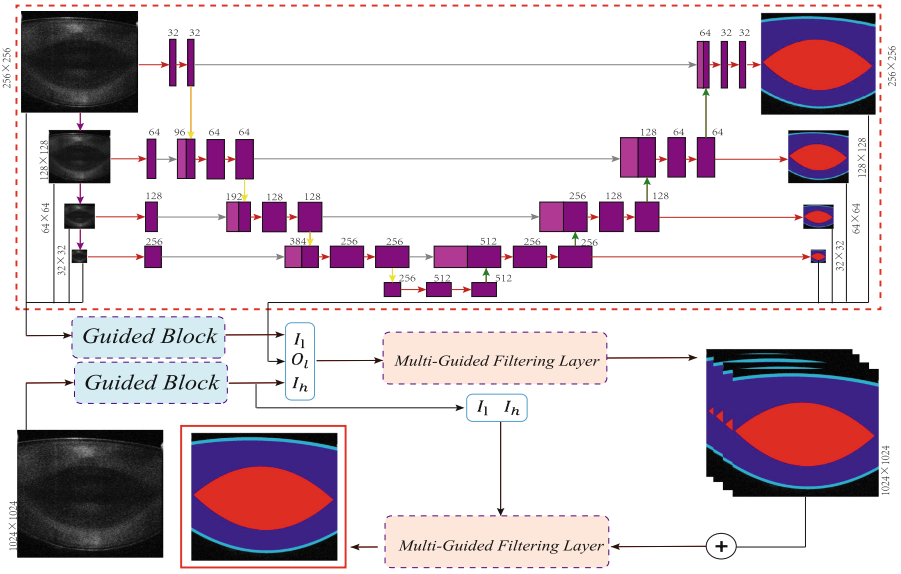

Fig. 2. Overview of the proposed deep architecture. Firstly, multi-scale side-outputs are generated by M-Net. Then the multi-guided filtering layer filters these side-output to high-resolution with the guidance from the guided map. At last, an average layer is employed to combine all the outputs, and the result is then guided to produce the final segmented output.

2.1 M-Shape Convolutional Network

We choose the M-Net [5] as the main body of our method, as shown by the red dashed box in Fig. 2. The M-Net includes a U-Net used to learn a rich hierarchical representation. Besides, multi-scale input and side-output are combined to better leverage multi-scale information.

2.2 Guided Block

In order to provide better guidance and reduce the impact of noise, we design a guided block to produce guided maps. The guided maps contain the main structure information extracted from the original images and also remove the noisy components. Figure 3 shows the architecture of the guided block. The guided block contains two convolution layers, between which are an adaptive normalization layer and a leaky ReLU layer. After the second convolution layer, an adaptive normalization layer [3] is added. The guided block is jointly trained with the entire network, thus the produced guided maps cooperate better with the rest of the model compared with the original image.

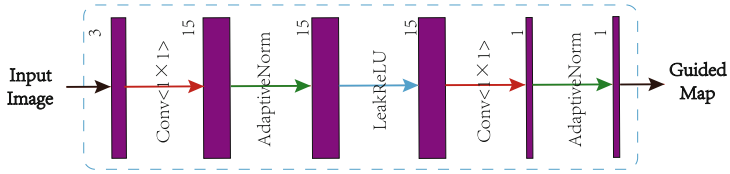

Fig. 3. Structure of the guided block. The guided block converts three-channel images to single-channel guided maps which reduce noise interference and provide better guidance.

2.3 Multi-guided Filtering Layer

The Multi-Guided Filtering Layer, take the advantages of guided filter, aims to transform the structure information contained in guided map and produce high-resolution filtered output (O_h). The inputs includes low-resolution output (O_l) the guided maps from the low (I_l) and high-resolution (I_h) input image.

Concretely, the guided filter is subjected to an assumption that the low-resolution filtered output $\hat{O}$ is a linear transform of guided map I in a square window w_k, which is centered at the position k with the radius being r. O_h is up-sampled from $\hat{O}$. The definition of $\hat{O}$ with respect to w_k is given as:

$$\hat{O}_{ki} = a_k I_{l_i} + b_k, \forall i \in w_k, \tag{1}$$

where (a_k, b_k) are some linear coefficients assumed to be constant in w_k and the radius of window is r.

a_k, b_k can be obtained by minizing the loss function:

$$E(a_k, b_k) = \sum_{i \in w_k} ((a_k I_{l_i} + b_k - O_{l_i})^2 + \epsilon a_k^2), \tag{2}$$

where ϵ is a regularization parameter penalizing large a_k.

Considering that each position i is involved in multiple windows $\{w_k\}$ with different coeffecients $\{a_k, b_k\}$, we average all the values of $\hat{O}_{ki}$ from different

windows to generate $\hat{O}_i$, which is equal to average the coefficients (a_k, b_k) of all the windows overlapping i, i.e.,

$$\hat{O}_i = \frac{1}{N_k} \sum_{k \in \Omega_i} a_k I_{l_i} + \frac{1}{N_k} \sum_{k \in \Omega_i} b_k = A_{l_i} * I_{l_i} + B_{l_i}, \tag{3}$$

where Ω_i is the set of all the windows including the position i, and $*$ is the element-wise multiplication. After upsampling A_l and B_l to obtain A_h and B_h, respectively, the final output is calcuted as (Fig. 4):

$$O_h = A_h * I_h + B_h. \tag{4}$$

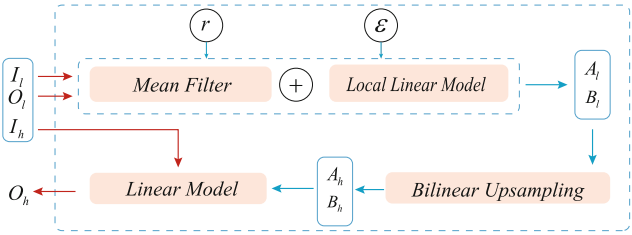

Fig. 4. Illustrations of multi-guided filtering layer. With low-resolution input I_l, O_l and hyperparameters r, ϵ, low-resolution A_l, B_l are calculated. By bilinear upsampling A_l, B_l, high-resolution A_h, B_h are generated which are then used to produce the final high-resolution output O_h with high-resolution guided map I_h.

3 Experiments

3.1 Datasets

(1) CASIA-2000: We collect high-resolution AS-OCT images with weak boundaries and noise from CASIA-2000 produced by Tomey Co. Ltd. The dataset contains 2298 images, including 1711 training images and 587 testing images. All the images are annotated by experienced ophthalmologists.

(2) ORIGA: It contains 650 fundus images with 168 glaucomatous eyes and 482 normal eyes. The 650 images are divided into 325 training images (including 73 glaucoma cases) and 325 testing images (including 95 glaucomas).

3.2 Training Details

We train our G-MNet from scratch for 80 epochs using Adam optimiser with the learning rate being 0.001. For the experiments on CASIA-2000 dataset, we set $\epsilon = 0.01$ and $r = 5$. The original image size is 2130×1864. We crop the

lens area, which is about 1024×1024 pixels, and resize it into 1024×1024 and 256×256 for high- and low-resolution inputs. For the experiments on ORIGA dataset, we set $\epsilon = 0.9$ and $r = 2$. The original image size is 3072×2048. We train a LinkNet [1] on training set to crop the OD area, and then resize it into 256×256 for low-resolution inputs.

3.3 Results on CASIA-2000 Dataset

Segmentation on CASIA-2000 aims to evaluate capsule, cortex and nucleus segmentation performance. Following the previous work in AS-OCT image segmentation [15], we employ the normalized mean squared error (NMSE) between a predicted shape $S_p = \{\hat{x}_i, \hat{y}_i\}$ and the ground truth shape $S_g = \{x_i, y_i\}$, where the shapes are represented by the coordinates of pixels. NMSE is defined as

$$NMSE = \frac{1}{n_g} \sum_{i=1}^{n_g} \sqrt{(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2}, \tag{5}$$

where n_g is the number of annotation points. A lower NMSE indicates the network is performing better.

We compare our G-MNet with several state-of-the-art networks. To verify the efficacy of the guided map, we replace it by the original image in G-MNet, and named this model **G-MNet-Image**. To test the performance of guiding in multi-scale, we construct a special G-MNet, named **G-MNet-Single**, which only filters the final averaged result without filtering multi-scale side-outputs. Table 1 shows the performance of different methods. We have the following observations: Firstly, G-MNet-Single performs better than M-Net, which indicates that guided filter is able to improve the accuracy of segmentation. Secondly, G-MNet outperforms G-MNet-Single by 0.16, 0.20 and 0.17 in capsule, cortex and nucleus boudary, respectively. This demonstrates the effectiveness of the learning strategy in multi-scale. Lastly, G-MNet performs much better than G-MNet-Image, which is disturbed by noises. This verifies that guided maps are able to provide better guidance for reducing noises.

Table 1. Segmentation results on CASIA-2000.

Method	Capsule	Cortex	Nucleus
FCN-VGG16 [8]	3.08 ± 4.84	3.34 ± 3.14	11.03 ± 4.08
DeepLabV2-Res101 [2]	3.97 ± 4.08	6.18 ± 4.31	10.88 ± 8.04
PSPNet-Res34 [16]	1.37 ± 0.96	1.73 ± 0.75	8.20 ± 3.97
M-Net [5]	1.37 ± 2.62	1.60 ± 0.93	7.93 ± 3.65
G-MNet-Image (ours)	3.23 ± 1.46	4.39 ± 1.34	9.44 ± 2.75
G-MNet-Single (ours)	0.73 ± 0.72	1.17 ± 0.91	7.62 ± 3.29
G-MNet (ours)	$\mathbf{0.57 \pm 0.29}$	$\mathbf{0.97 \pm 0.60}$	$\mathbf{7.45 \pm 3.24}$

3.4 Results on ORIGA Dataset

Following the previous work [5], we evaluate the OD and/or OC segmentation performance and employ the following overlapping error (OE) as the evaluation metric:

$$OE = 1 - \frac{A_{GT} \bigcap A_{SR}}{A_{GT} \bigcup A_{SR}}, \tag{6}$$

where A_{GT} and A_{SR} denote the areas of the ground truth and segmented mask, respectively.

We compare our G-MNet to the state-of-the-art methods in OD and/or OC segmentation, including ASM [14], SP [4], SW [13], U-Net [9], M-Net [5], M-Net with polar transformation (M-Net + PT) and Sun's [10].

Following the setting in [5], we firstly localize the disc center, and then crop 640×640 pixels to obtain the input images. Inspired by M-Net+PT, Inspired by M-Net+PT [5], we provide the results of G-MNet with polar transformation, called G-MNet+PT. Besides, to reduce the impacts of changes in the size of OD, we construct a method G-MNet+PT+50, which enlarges 50 pixels of bounding-boxes in up, down, right and left, where the bounding boxes are obtained from our pretrained LinkNet.

Table 2. Segmentation results on ORIGA.

Method	OE_{disc}	OE_{cup}
ASM [14]	0.148	0.313
SP [4]	0.102	0.264
SW [13]	–	0.284
Sun's [10]	0.069	0.213
U-Net [9]	0.115	0.287
M-Net [5]	0.083	0.256
M-Net+PT [5]	0.071	0.230
G-MNet (ours)	0.075	0.229
G-MNet+PT (ours)	0.069	0.213
G-MNet+PT+50 (ours)	**0.062**	**0.211**

Table 2 shows the segmentation results, the overlapping errors of other approaches come directly from the published results. Our method outperforms all the state-of-the-art OD and/or OC segmentation algorithms in terms of the aforementioned two evaluation criteria, which demonstrates the effectiveness of our model. Besides, Our G-Mnet outperforms M-Net by 0.008 and 0.027 in OE_{disc} and OE_{cup}, respectively. Simultaneously, Our G-Mnet+PT also performs better than M-Net+PT. These results indicate that our modification to M-Net has a great help to the performance.

4 Conclusions

In this paper, we propose a guide-based M-shape convolutional network, G-MNet, to segment biomedical images with weak boundaries, noise and high-resolution. Our G-MNet products high-quality segmentation results by incorporating guided filter into CNNs to learn better features for segmentation. It also benefit from the informative guided maps which provide better guidance and reduce the influence of noise by extracting the main feature from the original images. We further filter multi-scale side-outputs to construct the guided block more robust to noise and scaling. Thorough epxeriment on two benchmark datasets demonstrate the effectiveness of our method.

Acknowledgements. This work was supported by National Natural Science Foundation of China (NSFC) 61602185 and 61876208, Guangdong Introducing Innovative and Enterpreneurial Teams 2017ZT07X183, and Guangdong Provincial Scientific and Technological Fund 2018B010107001, 2017B090901008 and 2018B010108002, and Pearl River S&T Nova Program of Guangzhou 201806010081, and CCF-Tencent Open Research Fund RAGR20190103, and National Key R&D Program of China #2017YFC0112404.

References

1. Chaurasia, A., Culurciello, E.: LinkNet: exploiting encoder representations for efficient semantic segmentation. In: VCIP. IEEE (2017)
2. Chen, L.C., et al.: DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. TPAMI **40**, 834–848 (2018)
3. Chen, Q., et al.: Fast image processing with fully-convolutional networks. In: ICCV (2017)
4. Cheng, J., et al.: Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. TMI **32**, 1019–1032 (2013)
5. Fu, H., et al.: Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. TMI **37**, 1597–1605 (2018)
6. He, K., et al.: Guided image filtering. TPAMI **35**, 1397–1409 (2013)
7. Hu, P., et al.: Deep level sets for salient object detection. In: CVPR (2017)
8. Long, J., et al.: Fully convolutional networks for semantic segmentation. In: CVPR (2015)
9. 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
10. Sun, X., et al.: Localizing optic disc and cup for glaucoma screening via deep object detection networks. In: Stoyanov, D., et al. (eds.) OMIA/COMPAY -2018. LNCS, vol. 11039, pp. 236–244. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00949-6_28
11. Wong, A.L., et al.: Quantitative assessment of lens opacities with anterior segment optical coherence tomography. Br. J. Ophthalmol. **93**, 61–65 (2009)
12. Wu, H., et al.: Fast end-to-end trainable guided filter. In: CVPR (2018)

13. Xu, Y., et al.: Sliding window and regression based cup detection in digital fundus images for glaucoma diagnosis. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011. LNCS, vol. 6893, pp. 1–8. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23626-6_1
14. Yin, F., et al.: Model-based optic nerve head segmentation on retinal fundus images. In: EMBC. IEEE (2011)
15. Yin, P., et al.: Automatic segmentation of cortex and nucleus in anterior segment OCT images. In: Stoyanov, D., et al. (eds.) OMIA/COMPAY -2018. LNCS, vol. 11039, pp. 269–276. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00949-6_32
16. Zhao, H., et al.: Pyramid scene parsing network. In: CVPR, pp. 2881–2890 (2017)