ASNet: An adaptive scale network for skin lesion segmentation in dermoscopy images

Zhu, Liangjiu, Feng, Shuanglang, Zhu, Weifang, Chen, Xinjian

Liangjiu Zhu, Shuanglang Feng, Weifang Zhu, Xinjian Chen, "ASNet: An adaptive scale network for skin lesion segmentation in dermoscopy images," Proc. SPIE 11317, Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging, 113170W (28 February 2020); doi: 10.1117/12.2549178

SPIE.

Event: SPIE Medical Imaging, 2020, Houston, Texas, United States

ASNet: An adaptive scale network for skin lesion segmentation in dermoscopy images

Liangjiu Zhu [1,#] , Shuanglang Feng [1,#], Weifang Zhu [1,2] , Xinjian Chen [1,3,*]

[1]School of Electronics and Information Engineering, Soochow University, Suzhou 215006, China
[2]Collaborative Innovation Center of IoT Industrialization and Intelligent Production, Minjiang University, Fuzhou 350108, China
[3]State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou 215123, China

ABSTRACT

Dermoscopy is a non-invasive dermatology imaging and widely used in dermatology clinic. In order to screen and detect melanoma automatically, skin lesion segmentation in dermoscopy images is of great significance. In this paper, we propose an adaptive scale network (ASNet) for skin lesion segmentation in dermoscopy images. A ResNet34 with pre-trained weights is applied as the encoder to extract more representative features. A novel adaptive scale module is designed and inserted into the top of the encoder path to dynamically fuse multi-scale information, which can self-learn based on spatial attention mechanism. Our proposed method is 5-fold cross-validated on a public dataset from Challenge Lesion Boundary Segmentation in ISIC-2018, which includes 2594 images from different types of skin lesion with different resolutions. The Jaccard coefficient, Dice coefficient and Accuracy are 82.15±0.328%, 88.88±0.390% and 96.00±0.228%, respectively. Experimental results show the effectiveness of the proposed ASNet.

KEYWORDS: Skin lesion segmentation, dermatology, adaptive scale module, parallel convolutional layers, spatial attention mechanism

1. INTRODUCTION

Nowadays, the percentage of both melanoma and non-melanoma skin cancers is increasing worldwide. Dermoscopy is a non-invasive dermatology imaging method which can greatly help the specialists to inspect the pigmented skin lesions and diagnose malignant melanoma at the early stage. So the automated segmentation of skin lesion in dermoscopy images is a very important task.

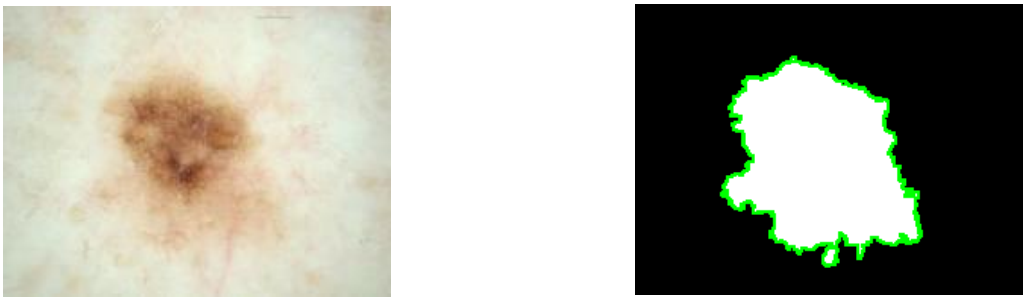
Recently, many methods based on convolutional neural networks (CNN) have been proposed for skin lesion segmentation. Sarker et al. [1] proposed a joint loss to perform the skin lesion segmentation. MultiResUNet[2] introduced multiple residual connection into the skip-connection of U-Net[3]. However, there are still many challenges due to the inhomogeneity of dermoscopy images, the influence of dense hair, and the blurred boundaries of lesions. Many multi-scale CNNs have been proposed to capture the multi-scale information. But they all ignore an important fact that CNNs should be similar to human visual system, which means that the receptive fields of the network should be dynamically adjusted with the change of target size and should not be integrated in a fixed way. In this paper, we design an adaptive scale network (refereed as ASNet) to solve above problems.

2. METHODS

In this section, the proposed method will be described in five parts: overall structure of the proposed ASNet, the adaptive scale module, the scale-aware module, loss function and implementation details.

*Corresponding author: E-mail:xjchen@suda.edu.cn, # indicates these authors contributed equally to this work

Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging, edited by Andrzej Krol, Barjor S. Gimi, Proc. of SPIE Vol. 11317, 113170W · © 2020 SPIE · CCC code: 1605-7422/20/$21 · doi: 10.1117/12.2549178



(a) (b)

Fig.1. Example of dermoscopy image and skin lesion. (a) Original dermoscopy image. (b)The corresponding ground truth of skin lesion (areas enclosed by the green line).

2.1 Overall structure of the proposed ASNet

Fig.2 demonstrates our proposed ASNet, which is an encoder-decoder structure based CNN. To capture more representative features, the ResNet34[4] with pre-trained weights is employed as the encoder to capture high-level semantic information gradually by residual blocks. A simple down-top decoder is designed to recover the spatial information from the coarse information stage by stage. Meanwhile, multiple skip-connections between decoder and encoder are utilized to make up for the fine information loss caused by downsampling. Note that a novel adaptive scale module (ASM) is proposed and inserted at the top of encoder to dynamically capture multi-scale context.

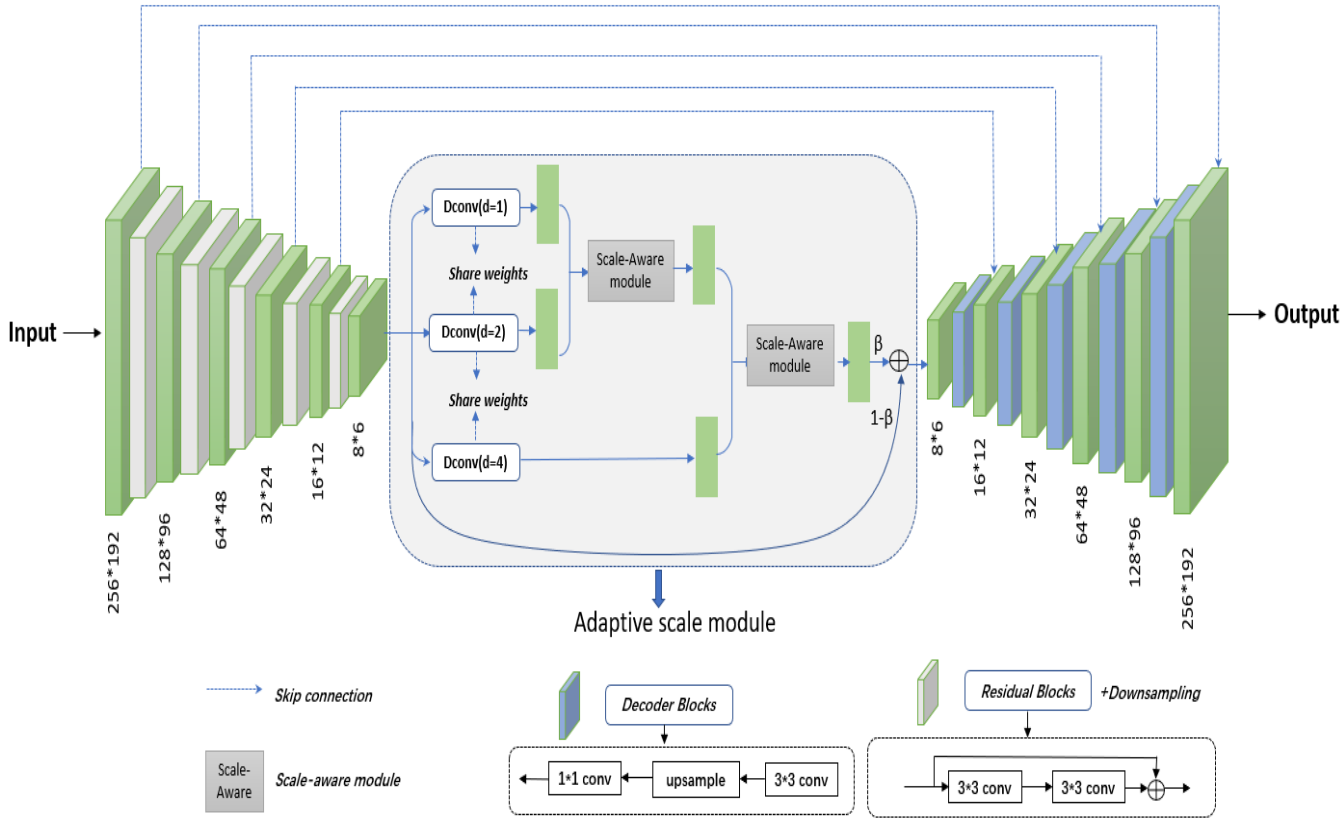


Fig.2. The illustrations of overall structure of the ASNet. The original image is fed into the encoder composed of pre-trained Resnet34 to obtain the high-level features, and then the multi-scale information is captured and dynamically merged by the proposed adaptive scale module (ASM). Next, the features are recovered by the decoder which consists of 3×3 convolution, bilinear interpolation up-sampling, and 1×1 convolution. Finally, the predicted score map is obtained.

2.2 The adaptive scale module

As shown in Fig.2, we propose the ASM module, which consists of three parallel dilated convolutions[5] with different dilation rates of 1, 2 and 4, to capture different scale information. Note that these different dilated convolutions have shared weights, which can reduce the number of model parameters and the risk of overfitting, and keep the transformation be consistent and the differences between these three branches only lie in the receptive fields. We employ two cascaded scale-aware modules (which will be introduced below) to get the final fusion feature of three branches. Then a residual connection with learnable parameter β is employed to obtain the output of the whole module.

2.3 The scale-aware module

We design a scale-aware module (SAM) to fuse different scale feature. As shown in Fig.3, a spatial attention mechanism is introduced to dynamically select the appropriate scale features and fuse them by self-learning. Specifically, two different scale features M_A and M_B pass through a series of convolutions and obtain two feature maps $\mathbf{A}, \mathbf{B} \in \mathrm{R}^{H \times W}$. Then pixel-wise attention maps a_A, a_B are generated by softmax operator on the spatial-wise values:

$$a_{A_i} = \frac{e^{A_i}}{e^{A_i} + e^{B_i}}, a_{B_i} = \frac{e^{B_i}}{e^{A_i} + e^{B_i}}, i = [1, 2, 3 \ldots H \times W] \tag{1}$$

Finally, two weighted features are added together as the fusion feature map:

$$M_{fusion} = a_A \odot M_A + a_B \odot M_B \tag{2}$$

Where the element-wise product operations ($\odot$) are performed between the attention maps and two scale features. We employ two cascaded scale-aware modules to get the final fusion feature of three branches.

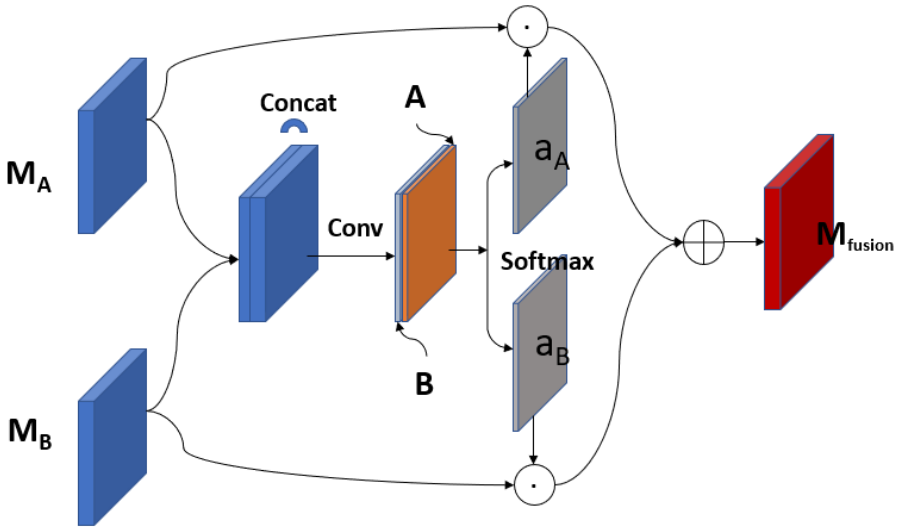


Fig.3. The illustrations of scale-aware module. Different weight maps are weighted to the features of different receptive fields through the spatial attention mechanism.

2.4 Loss function

A main challenge in medical image segmentation is class distribution imbalance. In order to optimize our model further, we employ a joint loss consisting of Dice loss and cross-entropy loss to perform this segmentation task, which is described as:

$$L(P,\ P') = -\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{2}P_i \log P_i' + \frac{2P_iP_i'}{P_i + P_i'}\right) \tag{3}$$

Where Pi and Pi′ represent the flatten predicted probabilities and the flatten ground truths of i^{th} image respectively, and N indicates the batch size.

2.5 Implementation details

We use the 'poly' learning rate policy. The basic learning rate is set to 0.01 , and the power is set to 0.9. Besides, stochastic gradient descent (SGD)[6] is adopted to optimize our model, in which momentum and weight decay are set to 0.9 and 0.0001 respectively. The batch size is set to 12. To better verify the performance of our network, we performed 5-fold cross validation both in ablation experiments and contrast experiments. To improve the computational efficiency of the model, we resized the image to 256×192 while maintaining the average aspect ratio. Online randomly left-right flipping was applied for data augmentation.

3. RESULTS

3.1 Dataset

The demoscopy image dataset was acquired from a public challenge: Lesion Boundary Segmentation in ISIC-20182. The data for this challenge were extracted from the ISIC-2017[7] dataset and the HAM10000 dataset [8], which were collected from different leading clinical centers internationally and acquired from different types of devices. The dataset includes 2594 images containing different types of skin lesions with different resolutions.

3.2 Evaluation metrics

To objectively evaluate the performance of the proposed method, three official evaluation metrics, including Jaccard index (Jac)[9], Dice coefficient (Dice)[10] and Accuracy (Acc), are adopted. In Table 1, TP, FP, TN and FN represent true positive, false positive, true negative and false negative, respectively.

Table 1
Evaluation metrics adopted in skin lesion segmentation experiments

Jac	TP/(TP+FP+FN)
Dice	2TP/(2TP + FP + FN)
Acc	(TP + TN)/(TP + FP + TN + FN)

3.2 Results

In order to verify the effectiveness of ASM module, we conduct a series of ablation experiments, which is shown in Table 2. The basic U-shape model with the pre-trained ResNet34 backbone is taken as the Baseline method. We first insert an ASM module without dilation convolution (ASM_w/o_Dl) into the Baseline and the Jaccard index decreases by 0.59% than the complete ASM, which implies that the capture of multi-scale context information is necessary. Second, we insert

an ASM without Scale-Aware module (ASM_w/o_SA) into Baseline, which also causes a performance reduction of 0.36% compared to the complete ASM and indicates that the dynamical selection of multi-scale contextual information is more conducive to image segmentation. Besides, to verify that the performance improvement of the proposed model is not caused by increasing the model complexity, we design a network with similar complexity to our model by adding multiple residual blocks in decoder, which is called Baseline-Wide[11] in Table 2. The results show that the proposed network, which adopts a novel adaptive scale module, can dynamically fuse multi-scale information through self-learning based on spatial attention mechanism.

To prove the superiority of the proposed network, we compare the segmentation results with that of U-Net and MultiResU-Net (an existing method evaluated on the same dataset), which are also shown in Table 2. As can be seen from Table 2, the proposed ASNet is 3.45% and 1.85% better than U-Net and MultiResU-Net in Jaccard index, respectively, which indicates that our ASNet can dynamically capture multi-scale context and achieve good performance. Fig.4 shows the visualization results of different models. As can be seen from Fig. 4, the proposed ASNet can predict more true positives and less false positives, which thanks to the network's ability of dynamical selection of receptive fields according to different target sizes.

Table 2
The results of comparison experiments and ablation studies on skin lesion segmentation task (mean ± standard deviation)

Methods	Jaccard(%)	Dice(%)	Accuracy(%)
U-Net [3]	78.70±0.317	86.58±0.376	95.06±0.251
MultiResU-Net [2]	80.30±0.372	\	\
Baseline	81.12±0.551	87.90±0.561	95.69±0.617
Baseline _ASM_w/o_Dl	81.56±0.353	88.25±0.433	95.80±0.305
Baseline _ASM_w/o_SA	81.79±0.425	88.46±0.532	95.92±0.359
Baseline_Wide	81.73±0.376	88.41±0.438	95.90±0.342
ASNet	**82.15±0.328**	**88.88±0.390**	**96.00±0.228**

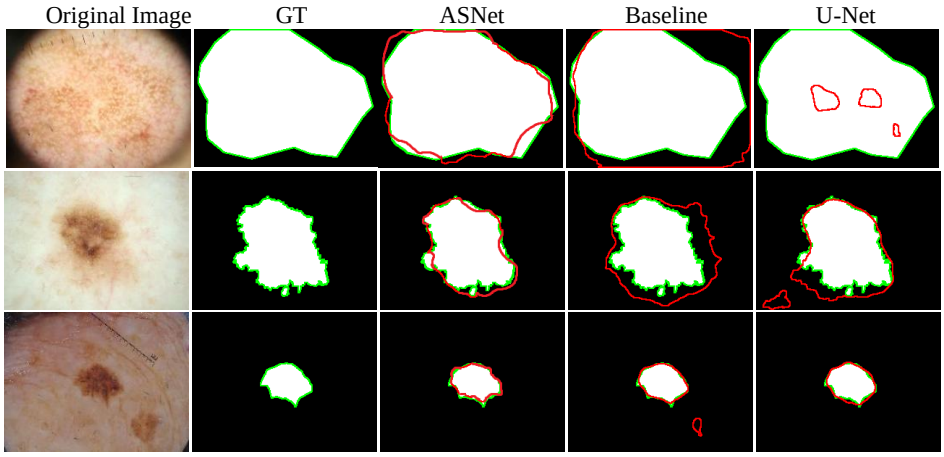


Fig.4. The visual examples of skin lesion segmentation. The areas outlined by green and red lines represents the ground truth and the prediction results, respectively. From the left to right: original image, ground truth (GT), the proposed ASNet, Baseline, U-Net.

4. CONCLUSIONS

In this paper, we propose an adaptive scale network (ASNet) for skin lesion segmentation in dermoscopy images. A novel adaptive scale module is designed and inserted into the top of the encoder path to dynamically fuse multi-scale information, which can self-learn based on spatial attention mechanism. Binary cross-entropy loss and Dice coefficient loss are effectively combined to constraint the model. We do comparison and ablation experiments to prove the superiority of the proposed network. As a result, our method achieves the best performances in Jaccard, Dice and Accuracy. The experimental results demonstrate the effectiveness and practicability of the method, which may help the specialists to inspect the pigmented skin lesions.

5. ACKNOWLEDGEMENTS

This work was supported in part by the National Key R&D Program of China under Grant 2018YFA0701700, and in part by the National Basic Research Program of China (973 Program) under Grant 2014CB748600, the National Natural Science Foundation of China (NSFC) under Grant 61622114, 81401472, and in part by Collaborative Innovation Center of IoT Industrialization and Intelligent Production, Minjiang University under Grant IIC1702.

6. REFERENCE

[1] Sarker, M. M. K., Rashwan, H. A., Akram, F., Banu, S. F., Saleh, A., Singh, V. K., ... & Puig, D. (2018, September). SLSDeep: Skin lesion segmentation based on dilated residual and pyramid pooling networks. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 21-29). Springer, Cham.
[2] Ibtehaz, N., & Rahman, M. S. (2020). MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Networks, 121, 74-87.
[3] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
[4] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
[5] Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122.
[6] Bottou, L. (2012). Stochastic gradient descent tricks. In Neural networks: Tricks of the trade (pp. 421-436). Springer, Berlin, Heidelberg.
[7] Codella, N. C., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., ... & Halpern, A. (2018, April). Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) (pp. 168-172). IEEE.
[8] Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data, 5, 180161.
[9] Niwattanakul, S., Singthongchai, J., Naenudorn, E., & Wanapu, S. (2013, March). Using of Jaccard coefficient for keywords similarity. In Proceedings of the international multiconference of engineers and computer scientists (Vol. 1, No. 6, pp. 380-384).
[10] Ye, S., & Ye, J. (2014). Dice similarity measure between single valued neutrosophic multisets and its application in medical diagnosis. Neutrosophic Sets and Systems, 6, 48-52.
[11] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (pp. 3-11). Springer, Cham.