Automatic Skin Lesion Segmentation by Feature Aggregation Convolutional Neural Network

Yuanfeng Ji[1], Xiaomeng Li[2], Gong Zhang[1], Di Lin[3], and Hao Chen[1,2]

[1] Imsight Medical Technology, Co.Ltd
[2] Department of Computer science and engineering, The Chinese University of Hong Kong.
[3]Shen Zhen University

Abstract. Melanoma is the most common form of cancer and the automatic melanoma segmentation is critical in the skin cancer diagnosis systems. In this paper, we present the feature aggregation convolutional neural network for skin lesion segmentation, where the low-level features and high-level features are carefully aggregated to enrich the feature extraction. To further improve the segmentation performance, the auxiliary loss is integrated in the encoder part. We participated the skin lesion segmentation challenge of the ISIC 2018 - Skin Lesion Analysis Towards Melanoma Detection and our method achieved 80.0% (Jaccard Index) on the validation dataset.

1 Introduction

Skin lesion segmentation is one of the essential steps in the computer-aided diagnosis for the various skin diseases. The automatic lesion segmentation in dermoscopic images is very challenging due to large variations in lesion size, location, shape, and color over different patients and the presence of artifacts such as hairs and veins. Moreover, the low contrast between the lesion and the surrounding textures also impedes the automatic lesion segmentation. Recently, the deep convolutional neural networks (DCNNs) based on the FCN [2] have significantly improved the skin lesion segmentation result. In this work, we present feature aggregation convolution neural network for skin lesion segmentation, where the low-level and high-level features are carefully integrated together to extract discriminative features. Also, we employed the auxiliary supervision to the convolution blocks between the encoder module and decoder module, which mitigates the difficulty in the network optimization. The experimental results are 80.0% Jaccard Index by a signal-model basis on the validation dataset.

2 TASKS AND METHODS

2.1 Dataset

The purpose of task 1 in the ISIC 2018 *Lesion Boundary Segmentation* is to recognize the melanoma from the dermoscopic images with accurate shape and

location. The challenge organizer provides 2594 dermoscopic images with binary masks as the training dataset [6] while the validation dataset contains 100 dermoscopic images. In the test stage, 1000 dermoscopic images are provided for algorithm evaluation and the final rank is determined according to Jaccard Index.

2.2 Method

Network Architecture Figure 1 shows our feature aggregation network architecture. We employ the ResNet34 [1] as the backbone network to build the encoder module while the decoder part consists of several deconvolution operators to recover the spatial resolution of the feature maps. To enrich features between the encoder and decoder part, we build dense connection modules to help the flow between the high-level feature and the low-level feature. However, different from the direct connection like U-net architecture [5], the features from the encoder part are first multiplied together, then add with the features from the decoder part. Such dense connection can bridge the gap between low-level features and the effective aggregation contributes to the features extraction, after which the decoder part can recover the spatial information with rich context information. To further enrich the context information and ease the difficulty in the training, we employ the auxiliary loss at the encoder part of our network. Finally, we use the Sigmoid function to produce the final score map. Both the main loss and auxiliary loss is the weighted combination of the binary cross entropy loss and dice loss, as described in equation 1. In our experiments, λ is set to 0.6.

$$loss = \lambda * \frac{2\sum t(\hat{y}) * y}{\sum t(\hat{y}) + \sum y} + (1-\lambda) * \frac{1}{n}\sum -y * log(\hat{y}) \tag{1}$$

where y denotes the ground truth label, $\hat{y}$ is the prediction score and $t(\cdot)$ denotes thresholding by 0.5. We omit the summation over pixels for simplification.

Training We use the SGD with momentum to optimize our network and the initial learning rate is 0.005. Dropout is added at the final convolution layer and the dropout ratio is 0.2. Data augmentation includes randomly flipping, shifting, rotation as well as scaling. In addition, we also randomly enlarge the image contrast, brightness, saturation within scale (-0.5,0.5). We train the network for 50 epochs and the batch size is set to 8. To evaluate our method, we conduct 5-fold cross-validation on the training dataset. Specifically, we select the training and validation data by the KFold module from the Sklearn [4] python package. Table 1 shows the Jaccard index results on each validation fold.

Implementation details We resize all the training images to 512×512 and we do not employ other pre-processing and additional dataset. The whole framework was implemented by Pytorch [3] package and the experiments were conducted with 4 Titan XP GPUs. The inference time for one dermoscopic image is 0.14 s.

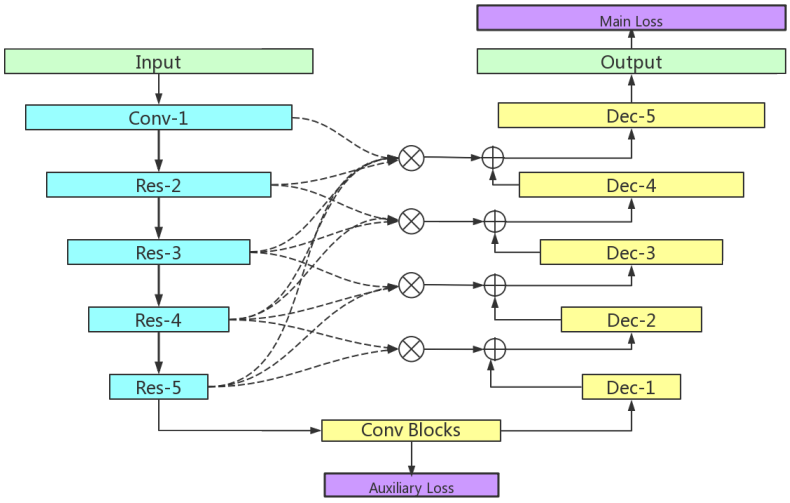

Fig. 1. Our network architecture. The feature maps in the encoder part are connected to the decoder part through U-net connections. Before each U-net connection, feature maps in the encoder part are multiplied together to preserve low-level features; see the dash line. Such dense feature aggregation can contribute to the feature extraction. The whole network is optimized by the main and auxiliary loss function, both of which are the weighted combination of cross-entropy loss and dice loss.

3 RESULTS

Our single model can achieve an average Jaccard index of 80.0% on the validation dataset. Fig. 2 shows some representative results on the validation set and we can see that the lesions with different shape and location can be well segmented by our method.

References

1. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)

Table 1. The cross validation results on the training dataset.

Folds	Jarrard index with threshold (%)
0	81.87
1	79.98
2	78.33
3	77.84
4	79.57
Average	79.51

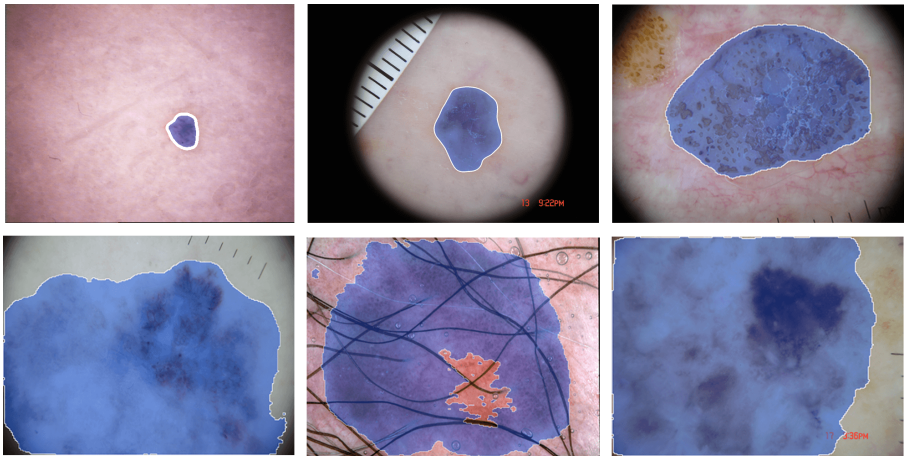

Fig. 2. Examples of skin lesion segmentation results on the ISIC 2018 validation datasets. The blue mask denotes the lesion segmentation result.

2. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015)
3. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch (2017)
4. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
5. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
6. Tschandl, P., Rosendahl, C., Kittler, H.: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data **5**, 180161 (2018). https://doi.org/10.1038/sdata.2018.161