A MULTI-TASK FRAMEWORK WITH FEATURE PASSING MODULE FOR SKIN LESION CLASSIFICATION AND SEGMENTATION

Sheng Chen[1], *Zhe Wang*[2], *Jianping Shi*[2], *Bin Liu*[1], *Nenghai Yu*[1]

[1]CAS Key Laboratory of Electromagnetic Space Information, University of Science and Technology of China
[2]SenseTime Group

ABSTRACT

Skin lesion classification and segmentation are highly correlated tasks. However, their relationship is not fully utilized in previous methods. In this paper, we propose a multi-task deep convolutional neural network architecture to solve the skin lesion classification and segmentation problem simultaneously. To take full advantage of features from different tasks and thus get richer knowledge about the sample, we design a feature passing module to pass messages between segmentation branch and classification branch. Since feature passing module is not always helpful and can be related with individual samples, gate functions are used for controlling messages transmission. Therefore, features from one task are learned and selectively passed to the other task, and vice versa, which effectively improves the performance of both tasks. We have evaluated the proposed method on ISBI-2017 challenge dataset, and the experimental results demonstrate the superiority and effectiveness of the proposed method, compared to our base model and other state-of-art methods.

Index Terms— skin lesion, convolutional neural networks, multi-task, feature passing

1. INTRODUCTION

With the development of medical technology, the diagnosis performance of skin diseases has been improved a lot. However, the shortage of dermatologists still makes early-stage detection and treatment of skin diseases difficult to achieve. Therefore, developing automatic analysis of skin lesion has drawn great attention in assisting dermatologists for enhancing their efficiency and objectivity of visual interpretation of dermoscopic images in clinics.

Recently, various algorithms have been proposed for automatic skin lesion assessment. For the skin disease recognition task, most previous methods adopt discriminative machine learning approaches, such as Support Vector Machine, Logistic Regression, etc. Recently, deep convolutional neural networks (CNN) [1] with hierarchical feature learning capability has been introduced into this field, and achieved better results compared with hand-crafted features. For example,

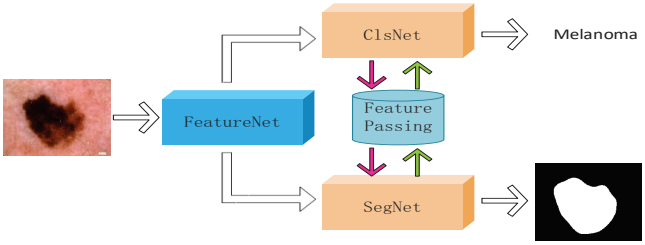

Fig. 1. Workflow of the proposed algorithm

Kawahara et al. [2] took deep features pretrained on ImageNet [1] as the input of a linear classifier. Ge et al. [3] proposed to use bilinear pooling technique [4] to extract the local features of VGG network [5] , and then combined it with global features extracted by the deep residual network (ResNet) [6]. As for skin lesion segmentation, [8] proposed a convolutional-deconvolutional neural network architecture and explored multiple color spaces to facilitate network training. Yuan et al. [9] designed a jaccard distance loss function to deal with strong imbalance between the number of foreground and background pixels of skin lesion images. Also, some researchers treated skin lesion classification and segmentation as a unified problem [10]. Yu et al. [11] proposed to segment the image to get the pathological part (foreground) firstly, and then crop out the foreground part as the input of classification network. This setting provides the classification network more representative and specific features. However, these methods did not make full use of the conjunct information existing in the features of different tasks.

Motived by [12], which used a gated bi-directional CNN to pass messages among multi scale features, we propose to integrate skin lesion classification and segmentation into a unified framework, using a multi-task architecture with feature passing module. In the proposed framework, the classification branch and the segmentation branch are jointly trained and share features at lower layers. It is observed that this multi-task technique can improve the efficiency in time and the prediction accuracy slightly. To dig deeper in latent dependencies of two tasks and make full use of additional messages that the classification (segmentation) branch provided

to the segmentation (classification) branch, we propose to use a feature passing module to transmit messages between two branches. Learnable gate functions are used to control message transmission, so that only useful information can get through. The experimental results show the superiority of our proposed method.

2. THE PROPOSED METHODS

The overview of the proposed algorithm is shown in Fig. 1. We first use the FeatureNet to extract features from the input dermoscopic image, and then feed them to ClsNet for classification and SegNet for segmentation. The proposed feature passing module is used to link ClsNet and SegNet so that information from the two tasks can flow to each other.

In this section, we first describe the multi-task architecture, including FeatureNet, ClsNet and SegNet. Then we get into the details of the feature passing module.

2.1. Multi-task architecture

The FeatureNet. The FeatureNet is taken from the part of the deep residual network (ResNet-101, until the $conv4\_10$ block) [6], aiming at extracting generic features of dermoscopic images for both tasks. A deep residual network contains a set of residual blocks, each of which consists of a few stacked layers, including convolutional layers, batch normalization (BN) layers and ReLU layers. Extra skip connections are introduced to each residual block to improve the information flow and alleviate the gradient vanishing problem largely. Let B_l to be the l-th residual block, H_{l-1} and H_l represent its input and output respectively. Then, the input-output relationship of block B_l can be written as:

$$H_l = f_l(H_{l-1}) + H_{l-1} \tag{1}$$

where $f_l(x)$ is the residual mapping function that the residual block B_l learns.

The ClsNet. The ClsNet is basically the latter part of ReNet-101 model, except that we replaces the output number of the last fully connected layer with the number of skin diseases 's categories.

The SegNet. The SegNet mainly follows the deeplab-ResNet101 [7,13], which proposed dilated convolution to obtain high-resolution predications. Dilated convolutions support exponential expansion of the receptive field without loss of resolution or coverage, hence it can improve predication accuracy cooperated with pooling function. If we refer to $*_l$ as dilation convolution, the dilation convolution function can be shown as below:

$$(F *_l k)(p) = \sum_{s+l\times t=p} F(s)k(t) \tag{2}$$

where F is the image matrix, k is the convolution filter, p, s, t are positions.

Another concern is that the deep imbalance between the foreground and the background classes of skin lesion images. To solve this problem, we use a weighted cross-entropy loss function [14] for skin lesion segmentation. It can be defined as:

$$L = -\beta \sum_{j\in Y_+} \log P(y_j = 1|X) - (1-\beta) \sum_{j\in Y_-} \log P(y_j = 0|X) \tag{3}$$

where X is The input image, $y_i \in \{0, 1\}, j = 1, .., |X|$ is the pixel-wise binary label of X, and Y_+, Y_- are the positive and negative labeled pixels with $\beta = |Y_-|/|Y_+|$. $P(\cdot)$ is obtained by employing a sigmoid function to the output layer.

2.2. Feature passing module

Since the category and the shape of the lesion are highly related, it is reasonable to expect the feature from one task to be useful for the other task. Therefore, we propose a feature passing module with gates to model the latent relationship between two tasks. With the feature passing module, the features from the segmentation branch are screened by the gate function, then useful information is selected to be passed and added to the classification branch, and vice versa. The gate functions are convolution layers followed by sigmoid functions to make the message passing rate in the range of (0,1). These gate functions are learnable so that the message passing rates can be controlled by the responses to particular visual patterns which are captured by gate filters and adapted to individual samples. We set this module at the $conv4\_22$ block.

Without the feature passing module, the relationship between inputs and outputs of this block can be represented in Eq. (1). Here we take x_{cls}, x_{seg} as inputs of classification and segmentation branches at this block. f_{cls}^{old}, f_{seg}^{old} are corresponding outputs, respectively.

With feature passing module, this block can be described as (here we ignore BN function for simplicity):

$$f_{cls}^{new} = op(f_{cls}^{old}, G_{seg2cls} \cdot \tilde{f}_{seg}) \tag{4}$$

$$f_{seg}^{new} = op(f_{seg}^{old}, G_{cls2seg} \cdot \tilde{f}_{cls}) \tag{5}$$

$$\tilde{f}_{seg} = \sigma(x_{seg} \otimes \tilde{w}_{seg} + \tilde{b}_{seg}) \tag{6}$$

$$\tilde{f}_{cls} = \sigma(x_{cls} \otimes \tilde{w}_{cls} + \tilde{b}_{cls}) \tag{7}$$

$$G_{seg2cls} = sig(x_{seg} \otimes w_{seg2cls} + b_{seg2cls}) \tag{8}$$

$$G_{cls2seg} = sig(x_{cls} \otimes w_{cls2seg} + b_{cls2seg}) \tag{9}$$

Where $G_{seg2cls}$ ($G_{cls2seg}$) is the gate function that controls the feature $\tilde{f}_{seg}$ ($\tilde{f}_{cls}$) from the segmentation (classification) branch to be passed to the classification (segmentation) branch. f_{cls}^{new} and f_{seg}^{new} are new outputs of the classification and segmentation branches at this block respectively. sig and σ are sigmoid and ReLU function. $\otimes$ is the convolution operation, and w and b are its learnable parameters. op can be element-wise summation, production or concatenation, and here we take it as element-wise summation operation.

Table 1. Performances compared with our base model and multi-task network without feature passing module.

Method	AC(cls)	mAP(cls)	JA(seg)	AC(seg)	DI(seg)
Base	0.772	0.699	0.779	0.940	0.862
Multi-task	0.779	0.712	0.780	0.940	0.863
Ours	0.801	0.747	0.787	0.944	0.868

3. EXPERIMENTS

3.1. Experimental setting

Dataset. We evaluated the proposed method on the ISBI-2017 challenge dataset[1]. This dataset contains 2000 training samples, 150 validation samples and 600 testing samples from three disease categories (melanoma, nevus and seborrheic keratosis).

Additional training data selection. There is a larger dataset named ISIC Data Archive[2] which contains over 10000 images. Since a large portion of the images in this dataset contain heavy noise, the performance even drops if all the images are added to our training dataset. Thus we proposed a simple method for data selection. We first extract convolutional features of the additional images and the original training images with a pretrained VGG network [10], and calculate the mean cosine distances between every additional sample and all the original training samples. An additional sample is selected only if it is close enough to all the original training samples in the feature space. Our intuition is that when the distance is close, this additional sample will be similar to the training sample space and suitable for our tasks. In this way, we choose 521 samples adding to the training set from this archive finally.

Data augmentation and Post processing. To prevent overfitting, we use on-the-fly data augmentation (crop, zoom, rotate, flip and add gaussian noise) during the training phase. During the testing phase, we also perform augmentation for each test image(crop, zoom, flip) and use average ensemble to get final results. For segmentation, we design to fill small holes of the output masks with morphological dilation since the pathological region of skin lesion images do not have holes.

Implementation details. We train our network using stochastic gradient descent (SGD) with a mini-batch of 16 images. First, we train the FeatureNet and ClsNet without SegNet for about 80 epochs at a learning rate of 0.00001. Then, we use the weight of ClsNet to initialize the SegNet (same as the initialization method in [7]) and train the SegNet with FeatureNet and ClsNet fixed for about 150 epochs at the learning rate of 0.0001. Finally we add the feature passing module between the ClsNet and SegNet, and train the whole network jointly for about 45 epochs at the learning rate of 0.0001. The input size of images is set to 233×233 , and it is observed that the input size has little impact on the performance.

[1]https://challenge.kitware.com/#challenge

[2]https://isic-archive.com/

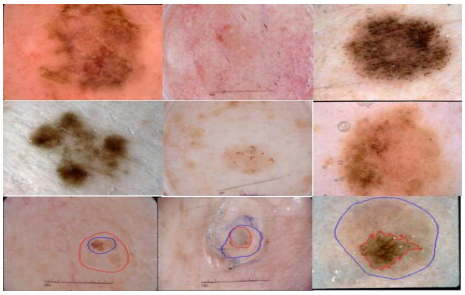

Fig. 2. Testing cases. Row 1: success for classification. Row 2: failures for classification. Row 3: failures for segmentation. The red dash and blue solid contours indicate the ground truth and the segmentation result, respectively.

3.2. Experimental results and evaluation

Evaluation metrics. Our segmentation task aims to segment out skin lesion region (foreground) from the background, which is the same as the challenge. So we adopted the challenge evaluation metrics, including Jaccard index (JA), accuracy (AC) and Dice coefficient (DI). For classification task, we treat it as multi-class classification (3 categories) while the challenge split the three-class classification into two binary classification. Thus we took the normal multi-class classification metrics as our evaluation metrics, including accuracy (AC) and mean average precision (mAP). The definition of mAP can be found in [15]. And other metrics are defined as:

$$\begin{aligned} AC &= \frac{N_{tp}+N_{tn}}{N_{tp}+N_{fp}+N_{fn}+N_{tn}} \\ \text{JA} = \frac{N_{tp}}{N_{tp}+N_{fn}+N_{fp}}&, \text{DI} = \frac{2 \cdot N_{tp}}{2 \cdot N_{tp}+N_{fn}+N_{fp}} \end{aligned} \tag{10}$$

where N_{tp}, N_{tp}, N_{tp} and N_{tp} denote the number of true positive, true negative, false positive and false negative, respectively.

Experimental results. To investigate whether the proposed multi-task framework with feature passing module has a positive impact on our tasks, we compared the proposed method with the base model (ResNet for classification and deeplab-ResNet for segmentation), and the multi-task network without feature passing. The results are showed in Table 1. As we can see from Table 1, compared with the base model, the proposed method achieved a better result of accuracy (0.801 vs. 0.772) for classification and Jaccard index (0.787 vs. 0.779) for segmentation. And the feature passing module improves the performance by 2.8% at accuracy for classification and 0.9% at Jaccard index for segmentation, which proves its effectiveness.

Also, we compared the proposed algorithm with other state-of-art methods on the testing dataset, and the results are shown in Table 2 and Table 3. For classification task, we compared with AlexNet [1], VGG16 [5] and ResNet101 [6], and our method performed best. For segmentation task, we compared with MtSinai (the first place in the challenge of segmentation) [8], U-net [16] and deeplab-resnet [13], it is clear that

Table 2. Classification performance compared with other state-of-art classification models.

Method	AC(cls)	mAP(cls)
AlexNet	0.775	0.706
VGG16	0.789	0.727
ResNet101	0.772	0.699
Ours	0.801	0.747

Table 3. Segmentation performance compared with other state-of-art segmentation models.

Method	JA(seg)	AC(seg)	DI(seg)
MtSinai	0.765	0.934	0.849
U-net	0.741	0.926	0.822
Deeplab-ResNet101	0.779	0.940	0.862
Ours	0.787	0.944	0.868

our method achieved better performance than those methods.

As can be seen in Fig. 2, we also provide some testing examples. From left to right are three categories: melanoma, seborrheic keratosis and nevus. The first two rows show the correct and failed samples for classification, respectively. It is clear that these failure cases differ with samples in its category for classification largely. The bottom row are failed samples for segmentation. Obviously, these samples have low contrast and artifacts around the lesions.

4. CONCLUSION

In this paper, we have proposed a multi-task framework with feature passing module, which solves skin lesion classification and segmentation simultaneously. The proposed feature passing module effectively transmit messages from the classification (segmentation) branch to the segmentation (classification) branch with gate functions, enhancing the performance of both tasks. The experimental results demonstrated the superiority of the proposed method.

5. REFERENCES

[1] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

[2] Kawahara J, BenTaieb A, Hamarneh G. Deep features to classify skin lesions[C]//Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on. IEEE, 2016: 1397-1400.

[3] Ge Z, Demyanov S, Bozorgtabar B, et al. Exploiting local and generic features for accurate skin lesions classification using clinical and dermoscopy imaging[C]//Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on. IEEE, 2017: 986-990.

[4] Lin T Y, RoyChowdhury A, Maji S. Bilinear cnn models for fine-grained visual recognition[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 1449-1457.

[5] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[6] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[7] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3431-3440.

[8] Yuan Y, Lo Y C. Improving Dermoscopic Image Segmentation with Enhanced Convolutional-Deconvolutional Networks[J]. arXiv preprint arXiv:1709.09780, 2017.

[9] Yuan Y, Chao M, Lo Y C. Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks with Jaccard Distance[J]. IEEE Transactions on Medical Imaging, 2017.

[10] Yang X, Zeng Z, Yeo S Y, et al. A Novel Multi-task Deep Learning Model for Skin Lesion Segmentation and Classification[J]. arXiv preprint arXiv:1703.01025, 2017.

[11] Yu L, Chen H, Dou Q, et al. Automated melanoma recognition in dermoscopy images via very deep residual networks[J]. IEEE transactions on medical imaging, 2017, 36(4): 994-1004.

[12] Zeng X, Ouyang W, Yan J, et al. Crafting gbd-net for object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.

[13] Chen L C, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. arXiv preprint arXiv:1606.00915, 2016.

[14] Xie S, Tu Z. Holistically-nested edge detection[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1395-1403.

[15] Gutman D, Codella N C F, Celebi E, et al. Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC)[J]. arXiv preprint arXiv:1605.01397, 2016.

[16] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015: 234-241.