Melanoma Segmentation and Classification in Clinical Images Using Deep Learning

Yunhao Ge
Robotics Institute
Shanghai Jiao Tong University
Shanghai, China
gyhandy@sjtu.edu.cn

Bin Li
Robotics Institute
Shanghai Jiao Tong University
Shanghai, China
lbin_sjtu@sjtu.edu.cn

Weixin Yan
State Key Lab of Mechanical System
and Vibration
Shanghai Jiao Tong University
Shanghai, China

ABSTRACT

In this paper, a deep learning computer aided diagnosis system (CADs) is proposed for automatic segmentation and classification of melanoma lesions, containing a fully convolutional neural network (FCN) and a specific convolutional neural network (CNN). FCN, which consists of a 28-layer neural structure, is designed for segmentation and with a mask for region of interest (ROI) as its output. Later, the CNN only uses the segmented ROI of raw image to extract features, while the DLCM features, statistical and contrast location features extracted from same ROI are merged into CNN features. Finally, the combined features are utilized by the fully connected layers in CNN to obtain the final classification of melanoma, malignant or benign. The training of FCN and CNN are separated with different loss functions. Publicly available database ISBI 2016 is used for evaluating the effectiveness, efficiency, and generalization capability with evaluating indicator, such as accuracy, precision, and recall. Preprocessing methods, such as data argumentation and balancing are utilized to make further improvements to performance. Experiments on a batch size of 100 images yielded an accuracy of 92%, a specificity of 93% and a sensitivity of 94%, revealing that the proposed system is superior in terms of diagnostic accuracy in comparison with the state-of-the-art methods.

Keywords

Deep learning; Melanoma; Segmentation; Classification; Benign or malignant.

1. INTRODUCTION

Skin cancer is the most common human malignancy, and there are 5.4 million new cases confirmed in the United States every year [1]. The death rate of malignant melanoma is very high, but melanoma detected in the early stage is curable in most cases. The low contrast as well as the irregular and fuzzy lesion borders makes the automatic melanoma segmentation and classification in clinical images a challenging task [2]. Although dermoscopy has been shown to lead to increased diagnostic accuracy compared to the conventional ABCD criteria [3], proper interpretation of dermoscopic images is normally time-consuming and complex.

Recently, deep learning has become one of the most powerful tools in machine learning and computer vision, increasing in popularity for the segmentation and classification of skin cancer. CNN is a type of deep learning method applied on the raw input images and used to automatically extract a set of complex high-level features. CNNs have also showed promising performance in various medical image computing problems, such as mitosis detection on histology images [4], as well as body parts recognition on CT images [5]. Adopting the traditional k-means classifier method to obtain the segmentation mask and CNN network, a CAD system has also been described to detect the melanoma, simplifying the classification process and achieving an accuracy of 0.8 [6]. How to detect the lesion of skin and accurately cut off the useless area in raw image, reserving the region of interest (ROI) are necessary to improve the final classification performance. However, different from the multi-class object detection in vast image, our aim is to detect a single melanoma and classify it, which obtains high ratio of area in the clinical image. Further, the first segmentation has a lower requirement for the quality of mask boundary.

Our contributions in this paper consists three parts. Firstly, we propose a fully automated melanoma detection CAD system that includes segmentation and classification using FCN and CNN. We extensively evaluate the effectiveness, efficiency, and the generalization capability of the proposed model using ISBI databases. Second, the specific FCN [7], [8] is proposed to target the medical image segmentation problem. The FCN segmentation classifies each pixel into either the foreground or background in combination with full resolution output. We use a loss function based Jaccard distance that is tailed to the medical image segmentation problem [9]. Merge data argumentation and data balancing are used to ensure effective and efficient learning with limited training data. We proposed the four layers of CNN to extract the convolutional feature from the ROI mask generated by FCN, while the GLCM and location features were simultaneously calculated. Finally, all the above-mentioned features were used as an input to fully connected layers in CNN to make the classification.

The training process of FCN and CNN is desperate. The ground truth mask and origin image are used to train the FCN network; Further, the segmented ROI and the category of melanoma were used to train the parameter of CNN. After the FCN and CNN were all trained, FCN was connected together with CNN to form a CAD automatic classification system. Our model has high accuracy, with both high sensitivity and specificity.

The rest of the paper is organized as follows. In Section III, we introduce the details of the proposed FCN-based segmentation method. In Section IV report the experimental design and results. Finally, the results are discussed in Section V, and Section VI concludes the study.

2. MATERIALS and METHODS

2.1 Dataset

In this study, we test the performance of FCN and CNN to classify the exact melanoma abnormality that are obtained by camera. Our data source is ISBI 2016 challenge dataset for skin lesion analysis towards melanoma detection [10], which consists of 900 training images (727 benign and 173 malignant) and corresponding ground truth and labels. In ISBI 2016 challenge dataset, melanoma images are divided into two classes: benign and malignant.

2.2 Proposed method

As shown in Fig.1, the framework of the deep learning CAD system includes three key components: (1) obtaining ROI based on mask after the FCN segmentation model; (2) automated feature extraction by CNN and extracting complementary features; (3) feature combination and classification based on CNN.

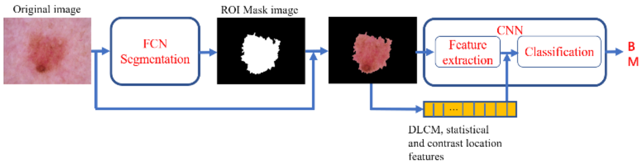

Fig. 1. The skeleton of proposed segmentation and classification CAD system.

2.2.1 *Segmentation based on FCN*

Figure 2 shows the architecture of FCN, in which an RGB image is used as the input and a posterior probability map as the output. The network contains 28 layers with 290129 trainable parameters, where Table.1 describes the architectural details. The stride of kernel is fixed as 1 and Rectified Linear Units (ReLU) are used as the activation function for each convolutional/deconvolutional layer [11]. For the output layer, we used sigmoid function as the activation function. Pixel-wise classification was performed and FCN served as a filter that projects the entire input image to a map where each element represents the probability that the corresponding input pixel belongs to the tumor. Up-sampling and deconvolutional layers were used to recover lost resolution while carrying over the global perspective from pooling layers [12]. The up-sampling layer performs the reverse operation of pooling and reconstructs the original size of activation, while the deconvolutional layer densifies the coarse activation map obtained from up-sampling by swapping the forward and backward passes of a convolution.

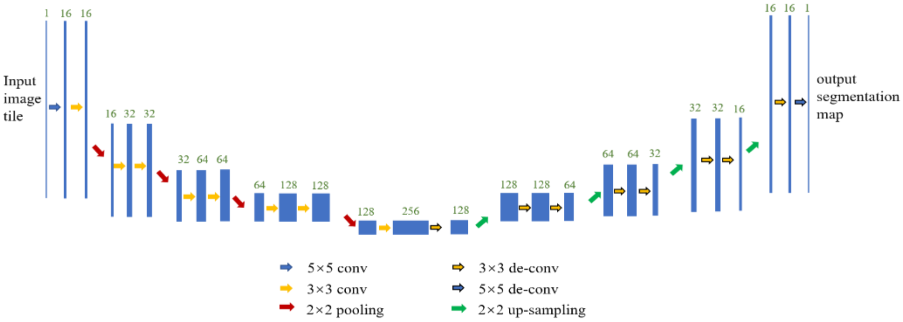

Fig.2 Architecture of the proposed fully convolutional network (FCN).

Proposed FCN model consists two pathways, in which contextual information is aggregated via convolution(c) and pooling (p)in the convolutional path and full image resolution is recovered via deconvolution (d) and up-sampling (u) in the deconvolutional path. The architectural details are described in Table 1.

Table 1. Architecture Details of The Proposed FCN Model

Conv	Kernel	Output	De-conv	Kernel	Output
c-1	5×5	718×538×16	d-1	3×3	41×30×128
c-2	3×3	716×536×16	u-1	2×2	82×60×128
p-1	2×2	358×268×16	d-2	3×3	84×62×128
c-3	3×3	356×266×32	d-3	3×3	86×64×64
c-4	3×3	354×264×32	u-2	2×2	173×128×64
p-2	2×2	177×132×32	d-4	3×3	175×130×64
c-5	3×3	175×130×64	d-5	3×3	177×132×32
c-6	3×3	173×128×64	u-3	2×2	354×264×32
p-3	2×2	86×64×64	d-6	3×3	356×266×32
c-7	3×3	84×62×128	d-7	3×3	358×268×16
c-8	3×3	82×60×128	u-4	2×2	716×536×16
p-4	2×2	41×30×128	d-8	3×3	718×538×16
c-9	3×3	39×28×256	output	5×5	722×542×1

We used a loss function based on Jaccard distance [9].

$$L_{dJ} = 1 - \frac{\sum_{i,j}(t_{ij}\, p_{ij})}{\sum_{i,j} t_{ij}^2 + \sum_{i,j} p_{ij}^2 - \sum_{i,j}(t_{ij}\, p_{ij})} \qquad (1)$$

The output of the FCN model is a posterior probability map where each pixel value represents the probability that the pixel belongs to the lesion. We altered the output to the binary image which similar to the mask ground truth in label. After making multiplications between original image and ROI mask, the ROI image are obtained and resized to 224×224 using bi-linear interpolation. RGB channels were kept as the input to the FCN model and each channel was rescaled to [0, 1].

2.2.2 *Classification based on CNN and multi-feature combination*

After the ROI image generated by the FCN network, we adopted two methods to simultaneously extract the features in the ROI image. First, the CNN network was used to extract a set of complex high level convolutional features. Second, the twelve artificial features consisted of GLCM features, location features, and artificial features. Finally, we combined these features together to make classification.

2.2.2.1 *Convolution features:*

In this experiment, a CNN model was used to extract the convolution features and the model structure is detailed as follows.

The first layer consists of:

• Convolutional layer: input is a 2D image (3 channels of dimension 64×64 of RGB input). It outputs 64 feature maps with a size of 64 × 64 (filter size is 5 × 5);

- ReLU non-linearity layer.

The second layer consists of:

- Convolutional layer: input is 64 feature maps. It outputs 32 feature maps with a size of 64 × 64 (filter size is 5 × 5);
- ReLU non-linearity layer.
- Max-pooling layer which down-samples the 64 feature maps to a dimension of 32 × 32;

The third layer consists of:

- Convolutional layer: input is 32 feature maps. It outputs 16 feature maps with a size of 32 × 32 (filter size is 3 × 3);
- ReLU non-linearity layer.

The final layer consists of:

- Convolutional layer: input is 16 feature maps. It outputs 16 feature maps with a size of 32 × 32 (filter size is 3 × 3);
- ReLU non-linearity layer.
- Max-pooling layer which down-samples the 16 feature maps to a dimension of 16 × 16;

Table 2. Architecture of The CNN Network Used For Training in Our Experiment

CNN Layer	1	2	3	4
# of Channels	64	32	16	16
Filter Size	5x5	5x5	3x3	3x3
Input size	64x64	64x64	32x32	32x32

2.2.2.2 Artificial features:

Before the extraction of artificial features, the ROI image need to be converted to the gray size. The artificial features consist of statistical parameters, GLCM features and location features. Below is a list of twelve features extracted from the mask area of the original image; they were selected to be combined with the output of the final convolutional layer to train and test our method.

Statistical parameters: Four statistical parameters of the ROI region were extracted: mean, variance, skewness, and kurtosis. The four parameters are calculated as follows:

$$mean: \mu = \sum_{k=1}^{N} f_k p_f(f_k) \tag{2}$$

$$\mathrm{var}iance: \sigma^2 = \sum_{k=1}^{N} (f_k - \mu)^2 p_f(f_k) \tag{3}$$

$$skewness: ske = \sum_{k=1}^{N} [(f_k - \mu)^3 p_f(f_k)] / \sigma^3 \tag{4}$$

$$kurtosis: kur = \sum_{k=1}^{N} [(f_k - \mu)^4 p_f(f_k)] / \sigma^4 \tag{5}$$

GLCM features: GLCM features consists of sum entropy (SE), sum average (SA), difference variance (DV), and difference entropy (DE). SE is a logarithmic of ROI in consideration. SA is calculated from the ROI and the size of gray scale. DV is a variance measure between the ROI intensities calculated as a function of the SE calculated previously. DE is an entropy measure which provides a measure of no uniformity while taking into consideration a difference measure obtained from the original image. And these four parameters are calculated as follows:

$$SE = -\sum_{i=2}^{2N_g} p_{x+y}(i) \log\{p_{x+y}(i)\} \tag{6}$$

$$SA = \sum_{i=2}^{2N_g} i p_{x+y}(i) \tag{7}$$

$$DV = \sum_{i=2}^{2N_g} (i - SE)^2 p_{x-y}(i) \tag{8}$$

$$DE = -\sum_{i=2}^{2N_g} p_{x-y}(i) \log\{p_{x-y}(i)\} \tag{9}$$

Location features: Four parameters about ROI location and shape were extracted: convexity (C(S)), compactness (C), aspect ratio (AR), area ratio (R_Area). The four parameters are calculated as follows:

$$C(S) = \frac{A}{Area(CH(S))} \tag{10}$$

$$C = \frac{P^2}{4\pi A} \tag{11}$$

$$AR = \frac{D_y}{D_x} \tag{12}$$

$$R\_Area = \frac{Area\_ROI(in\_pixels)}{Area\_window(in\_pixels)} \tag{13}$$

Where: S is a ROI, CH(S) is its convex hull and A is the ROI's area, P is the ROI's perimeter, and $Area\_window = D_x * D_y$, D_x is the width's ROI and D_y is the height's ROI.

2.2.2.3 Combination of all features:

The output of the fourth layer—including 16 channels of 1x1 array as well as the 12 features extract from origin image—is fully connected to an MLP classifier with a dropout rate of 0.8. The MLP classifier contains two hidden layers with a size of [1024, 64], and its output layer activation function adopted the softmax function and outputs the category of input clinical image.

3. RESULTS and DISCUSSION

3.1 Training and Classification Details

Table 3 illustrates the training method and parameter in FCN training. The FCN model has 28 layers and 1,055,008 parameters to be learned. The dataset is relatively small compared with the size of the network. We initialized the network weights using Xavier's technique [13].

Table 3. Training Methods of The FCN Network in Our Experiment

Training Methods	**Optimization Algorithm**	**Learning Rate (α)**	**Dropout (p)**
Option/value	Adam	0.003	0.5
Function	Adjust the learning rate	Speed up the training procedure	Reduce the overfitting substantially

Because the label for segmentation of each image is a mask, the mask should be merged into the raw image as an additional channel for RGB image to make data argumentation suitable for the FCN model. After the common argumentation for irregular medical image—flipping, rotating, zooming, horizontal and vertical translation—the mask and raw data must be separated from each other.

Cross entropy function was chosen to train the CNN network; Stochastic Gradient Descent (SGD) with RMSProp was employed [14], which is an adaption of R-Prop for SGD with Nesterov momentum [15]. The uniform weight filler was used, which has a learning rate of 0.0001, and the epochs are 3000 iterations.

The training of FCN and CNN are all implemented with Tensorflow framework, which can support GPU accelerate technology. Accordingly, it makes the training of architecture with millions of parameters feasible. All experiments were implemented on a workstation with two NVIDIA Titan X GPU that has 24 GB memory. Based on our implementation and hardware configuration, the entire model took 1.8 ms to conduct segmentation and classification for every image during testing.

3.2 Performance Evaluation

The output of the entire model is the classification of benign or malignant in relation to the image. There are four possibilities of results based on the actual class for the true states and predicted class for predicted states. If the training example is positive and the prediction is positive, it is referred to as a true positive (TP); if the prediction is negative, it is denoted as a false negative (FN). On the other hand, if the training example is negative and it is classified as negative, it is called a true negative (TN); otherwise, it is a false positive (FP).

In order to evaluate the performance and discriminative power of the whole model, measurements for overall classification accuracy, sensitivity, and specificity were calculated as follows:

$$Sensitivity = \frac{TP}{TP+FN} \quad (13)$$

$$Specificity = \frac{TN}{TN+FP} \quad (14)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (15)$$

3.3 ROI image results using FCN

Recall that the FCN works as ROI image extraction, where the results are shown in Fig. 3.

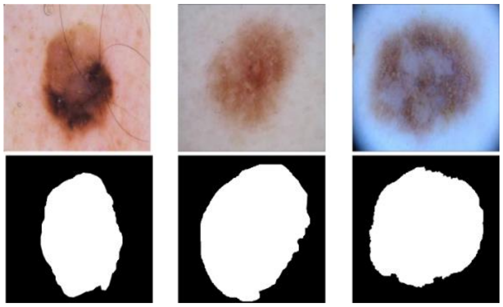

Fig. 3. Segmentation results for examples in the ISBI 2016 testing dataset

The automatic segmentation result shows the performance of the FCN model, in which the convolution layers aggregated the image information from fine details to global concept; a hierarchical structure of deconvolutional layers performed well in recovering image details regarding skin tumors. The FCN network can obtain both global information and local details to obtain the segmentation mask of the raw image. Overall, the FCN model accurately delineated the skin lesion.

3.4 CNN classification results

In this section, the proposed method is evaluated on the ISBI 2016 challenge dataset. This database consists of 900 images (173 malignant and 727 benign). The training set of CNNs must be sufficient enough, and data augmentation is conducted to increase the number of images from 900 origin images to 9000 images. The training data and testing data is split into two randomly selected groups with a rate of 0.8 and 0.2, respectively. Thus, there is no overlap between testing and training samples.

The confusion matrix of a batch size of 100 test images (50 malignant, 50 benign) is obtained using the entire model trained by 5820 training images (1390 malignant, 5820 benign), then the accuracy, sensitivity, and specificity of cases can be calculated as follows.

$$Sensitivity = \frac{46}{50} = 0.92$$

$$Specificity = \frac{47}{50} = 0.94$$

$$Accuracy = \frac{47+46}{100} = 0.93$$

Table 4. Confusion Matrix For Augmented ISBI 2016 challenge dataset Test Set Predictions Using Proposed CADs

		Predict		
		Benign	**Malign**	**Total**
Actual	**Benign**	47	3	50
	Malign	4	46	50
	Total	51	49	100

For comparing the proposed method with other existing methods, three works that have reported their results on the same dataset are studied: [16], [17] and [18]. It should be noted that the classification in [17] is based on the area, where the physician who had already conducted examinations. For the fairness of comparison, only the automatic CAD systems are reported here. Further, the evaluation results are shown in Table 5. From Table 5, it can be observed that the proposed CAD system has the highest accuracy and specificity compared with other state of the art methods. The result of this study indicates that deep learning has promising potential in the field of intelligent medical image diagnosis practice.

Table 5. Quantitative comparison of diagnostic results, best results are bold

Methods	**Test Dataset**		
	Accuracy	**Sensitivity**	**Specificity**
MED-NODE texture descriptor [16]	0.62	0.85	0.76

MED-NODE color descriptor [16]	0.74	0.72	0.73
C. Muntean et.al [18]	0.46	0.87	0.70
Nasr-Esfahani [6]	0.81	0.80	0.81
Proposed	**0.92**	**0.94**	**0.93**

4. CONCLUSION

In this study, a novel CAD system based on deep learning algorithm is proposed for automatically making segmentation and classify-cation in melanoma clinical images. Preprocessing methods, data argumentation, and balancing preprocessing are used to avoid overfitting and enhance the performance based on the ISBI 2016 dataset. The FCN segmentation model achieves pixel-wise classification using up-sampling and deconvolutional layers to reconstruct the original size of activation. Based on the output of FCN, the ROI of raw image was obtained to eliminate the useless skin region. CNN automatically extracts complex high-level features, other complementary features, while DLCM features, as well as statistical and contrast location features are also extracted. The combined features are utilized by fully connected CNN layers to make the final classification. The evaluation experiment on ISBI database shows an accuracy of 92%, specificity of 93%, and a sensitivity of 94%. The suitable robustness performance implies our system can be easily generalized to other challenging medical image segmentation and classification problems.

5. ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China under Grant No. 51475305 and 61473192.

6. REFERENCES

[1] Rogers, H. W. et al. Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the US population, 2012. *JAMA Dermatology* 151.10, 1081–1086 (2015).

[2] A. F. Jerant, J. T. Johnson, C. Sheridan and T. J. Caffrey, “Early detection and treatment of skin cancer,” American family physician, vol. 62, no. 2, pp. 357-386, 2000.

[3] F. Nachbar, W. Stolz, T. Merkle, A. B Cognetta, T. Vogt, M. Landthaler, P. Bilek, O. B.-Falco, and G. Plewig, “The abcd rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions,” Journal of the American Academy of Dermatology, vol. 30, no. 4, pp. 551–559, 1994.

[4] D. C. Cire¸san, A. Giusti, L. M. Gambardella, and J. Schmidhuber, “Mitosis detection in breast cancer histology images with deep neural networks,” in Proc. MICCAI. 2013, pp. 411–418.

[5] Z. Yan et al., “Multi-instance deep learning: Discover discriminative local anatomies for bodypart recognition,” IEEE Trans.Med.Imag., vol. 35, no. 5, pp. 1332–1343, May 2016.

[6] Nasr-Esfahani, E., et al. (2016). Melanoma Detection by Analysis of Clinical Images Using Convolutional Neural Network. 2016 38th Annual International Conference of the Ieee Engineering in Medicine and Biology Society. J. Patton, R. Barbieri, J. Ji et al.: 1373-1376.

[7] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in Proc. CVPR, Jun. 2015, pp. 3431–3440.

[8] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. MICCAI, 2015, pp. 234–241.

[9] Yuan, Y., et al. (2017). "Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks With Jaccard Distance." IEEE Transactions on Medical Imaging 36(9): 1876-1886.

[10] https://challenge.kitware.com/#challenge/560d7856cad3a57cfde481ba

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Proc. Adv. Neural Inf. Process. Sys., 2012, pp. 1097–1105.

[12] H. Noh, S. Hong, and B. Han, “Learning deconvolution network for semantic segmentation,” in Proc. ICCV, Sep. 2015, pp. 1520–1528.

[13] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in Proc. Aistats, vol. 9. 2010, pp. 249–256.

[14] Dauphin, Y. N., de Vries, H., Chung, J., Bengio, Y., 2015. Rmsprop and equilibrated adaptive learning rates for non-convex optimization. arXiv:150204390.

[15] Sutskever, I., Martens, J., Dahl, G., Hinton, G., 2013. On the importance of initialization and momentum in deep learning. International Conference on Machine Learning, pp. 1139-1147.

[16] I. Giotis, N. Molders, S. Land, M. Biehl, M. F. Jonkman and N. Petkov, “MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images,” Expert Systems with Applications, Elsevier, vol. 42, no. 19, pp. 6578-6585, 2015.

[17] C. Munteanu and S. Cooclea, “Spotmole – melanoma control system,” 2009. Available: http://www.spotmole.com/

[18] E. Zagrouba and W. Barhoumi, “A preliminary approach for the automated recognition of malignant melanoma, ” Image Analysis & Stereology, vol. 23, no. 2, pp. 121-135, 2004.