Improving Skin Lesion Segmentation via Stacked Adversarial Learning

Lei Bi[1], Dagan Feng[1,4], Michael Fulham[1,2,3], Jinman Kim[1]
[1]School of Information Technologies, University of Sydney, Australia
[2]Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, Australia
[3]Sydney Medical School, University of Sydney, Australia
[4]Med-X Research Institute, Shanghai Jiao Tong University, China

ABSTRACT

Segmentation of skin lesions is an essential step in computer aided diagnosis (CAD) for the automated melanoma diagnosis. Recently, segmentation methods based on fully convolutional networks (FCNs) have achieved great success for general images. This success is primarily related to FCNs leveraging large labelled datasets to learn features that correspond to the shallow appearance and the deep semantics of the images. Such large labelled datasets, however, are usually not available for medical images. So researchers have used specific cost functions and post-processing algorithms to refine the coarse boundaries of the results to improve the FCN performance in skin lesion segmentation. These methods are heavily reliant on tuning many parameters and post-processing techniques. In this paper, we adopt the generative adversarial networks (GANs) given their inherent ability to produce consistent and realistic image features by using deep neural networks and adversarial learning concepts. We build upon the GAN with a novel stacked adversarial learning architecture such that skin lesion features can be learned, iteratively, in a class-specific manner. The outputs from our method are then added to the existing FCN training data, thus increasing the overall feature diversity. We evaluated our method on the ISIC 2017 skin lesion segmentation challenge dataset; we show that it is more accurate and robust when compared to the existing skin state-of-the-art methods.

Index Terms— Segmentation, Fully Convolutional Networks (FCN), Skin Lesion

1. INTRODUCTION

Malignant melanoma has one of the most rapidly increasing incidences in the world with a considerable mortality rate. Early diagnosis is particularly important since melanoma can be cured with prompt excision. Dermoscopy plays an important role in the non-invasive early detection of melanoma [1]. However, melanoma detection using human vision alone is subjective, can be inaccurate and poorly reproducible even among experienced dermatologists [2]. This is attributed to the challenges in interpreting images with diverse characteristics including lesions of varying sizes and shapes, lesions that may have fuzzy boundaries, different skin colors and the presence of hair [2]. Motivated by these difficulties, there has been a great interest in developing computer-aided diagnosis (CAD) systems that can assist the dermatologists' clinical evaluation [1, 2].

Segmentation of skin lesions is an important step for a melanoma CAD. However, traditional methods [3, 4] that use edges, regions and shape models, rely on hand-crafted features and a priori knowledge that limit widespread application. Recently, deep learning methods based on fully convolutional networks (FCNs) have been successful in natural image segmentation related challenges [5]. This success is primarily attributed to the ability of a FCN to leverage large datasets to hierarchically learn the features that best correspond to the appearance as well as the semantics of the images [5]. In addition, FCNs can be trained in an end-to-end manner for efficient inference, i.e., images are taken as inputs and the segmentation results are directly outputted. However, there is a scarcity of annotated medical imaging training data due to the large cost and human manpower required [6]. So in the situation where training data cannot account for skin lesions from different patients with large differences in textures/size/shape, FCNs do not provide accurate results. Data augmentation approaches, such as random crops, flips and color jittering, have been applied to increase the overall volume of the training data, but they simply duplicate existing training features rather than add a variety of new features for leaning.

Some researchers have used specific cost functions and post-processing algorithms to refine the coarse boundaries of the results to improve FCN skin lesion segmentation. For example, Yuan et al [7] replaced the cross-entropy loss used in traditional FCN with a Jaccard distance loss for training. Bi et al [8] used cellular automata algorithm as a post-processing algorithm to refine the FCN segmentation outcomes. Unfortunately, data specific cost functions have limited generalizability to different datasets. In addition, the reliance on post-processing algorithms could override the FCN outcomes because the post-processing is usually unsupervised and cannot fully describe the training data.

In this paper, our aim is to improve the segmentation performance of FCNs via stacked adversarial learning (SAL). We leverage generative adversarial networks [9] (GANs) and add a stacked adversarial learning architecture to iteratively

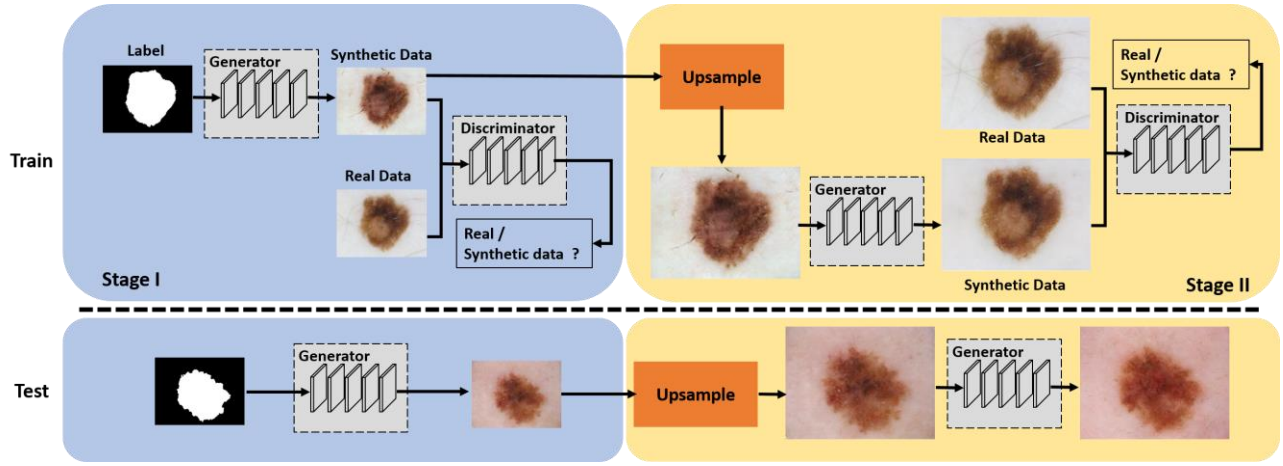

Figure 1. Overview of our proposed stacked adversarial learning (SAL).

learn skin lesion features in a class-specific manner e.g., melanoma and non-melanoma classes, and then added the learned skin lesion features into the existing FCN training data. Our hypothesis is that this approach will increase the overall feature diversity that then allows the FCN to learn and then improve the accuracy of the segmentation.

2. METHODS AND MATERIALS

2.1. Fully Convolutional Networks (FCNs)

The FCN architecture was converted from convolutional neural networks (CNNs) for efficient dense inference [5]. It contains downsampling and upsampling components. The downsampling part has stacked convolutional layers to extract high-level semantic information and has been routinely used in CNNs for image classification related tasks [11, 12]. The upsampling part has stacked deconvolutional layers, which are transposed convolutional layers that upsample the feature maps derived from the downsampling component to output the segmentation results. For skin lesion segmentation, the FCN architecture can be trained end-to-end by minimizing the overall loss function (e.g., cross-entropy loss) between the predicted results and the ground truth annotation of the training data. The FCN parameters (weights) can then be updated iteratively using e.g., a stochastic gradient descent (SGD) algorithm.

2.2. Stacked Adversarial Learning for Skin Lesion Features

Adversarial Learning (also known as generative adversarial networks (GANs) [9]) has 2 main components: a generative model G (the generator) that captures the data distribution and a discriminative model D (the discriminator) that estimates the probability of a sample that came from the training data rather than G. The generator is trained to produce outputs that are difficult to be distinguished from the real data by the adversarially trained discriminator, while the discriminator is trained to detect synthetic data created by the generator.

For learning skin lesion features, we embed the training label (annotation) for training and adoption as part of the formulation. During training, the generator takes the training label as the input to learn a mapping to synthesize the dermoscopic images that appear realistic. The discriminator then attempts to separate the real and synthetic dermoscopic images. Thus the loss function can be defined as conditional [13, 14] on the label l:

$$\mathcal{L}(G,D) = \mathbb{E}_{l,y}[logD(l,y)] + \mathbb{E}_{l,z}[\log(1 - D(l,G(l,z)))]$$

where y is the dermoscopy images and z is the input random noise. $D(\cdot)$ represents the probability that the input to $D(\cdot)$ came from the real data while $G(\cdot)$ represents the mapping to synthesize the real data. We used a stacked architecture to refine the output of the synthesized images, which can be defined as:

$$\mathcal{L}(G^*,D^*) = \mathbb{E}_{s,y}[logD^*(s,y)] + \mathbb{E}_{s,z}[\log(1 - D^*(s,G^*(s,z)))]$$

where s is output of the synthesized data and Fig. 1 shows the overall structure of the proposed stacked adversarial learning.

2.3. Adversarial Learning for Improving Segmentation

Fig. 2 presents our proposed approach to leverage the proposed SAL. Initially, we separated the training data (denoted as R) into melanoma (denote as M) and non-melanoma (denote as N) training sets. Afterwards, we used the label M_l and images M_y in M to train a stacked adversarial learning model (SAL) for deriving melanoma

features (denote as *M*-Model). The same approach was also used to train the non-melanoma training data (denote as *N*-Model). At the adoption stage, the trained *N*-Model was applied on the label M_l to produce the additional non-melanoma training data N_y^* while *M*-Model was applied on the label N_l to get the additional melanoma training data M_y^*. Finally, the original training data *R* together with the derived additional training data were used to train a new FCN. The reason we divided the training set into two disjoint sets is so that the individual sets could describe different class-specific attributes, e.g., melanoma and non-melanoma. Consequently, this allows the FCN to learn additional skin lesion features in a controlled manner.

We trained the first and the second stage of the SAL separately with the Torch library on a 12 GB Nvidia Maxwell Titan X GPU. For training the first stage we resized the image to 256×256 and to 512×512 for the second stage, while keeping the aspect ratio (the shorter axis of the image was padded with 0 values). The two stages were trained with a batch size of 1 at a learning rate of 0.0002 for 200 epochs.

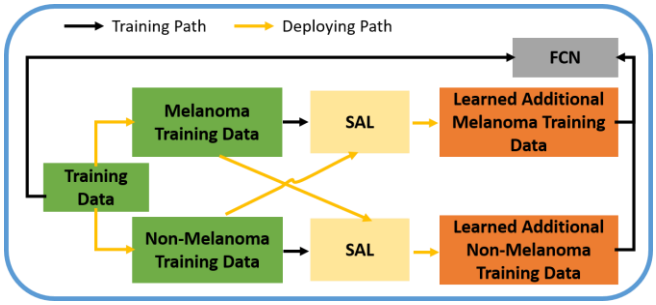

Figure 2. Outline of our approach.

3. RESULTS AND DISCUSSION

3.1. Materials and Experimental Setup

The 2017 ISIC Skin Lesion Challenge (denoted as ISIC 2017 [10]) dataset is a subset of the large International Skin Imaging Collaboration (ISIC) archive. It contains dermoscopic images acquired with a variety of different devices at numerous international clinical centers. The dataset provides 2,000 training images (1,626 non-melanoma and 374 melanoma) and a separate test dataset of 600 images (483 non-melanoma and 117 melanoma). Image size varies from 453×679 pixels to 4499×6748 pixels. The training dataset was used to train the deep models that were then applied on the test dataset. Manual delineations by clinical experts were used as the ground truth.

To evaluate the effect of including SAL in FCN segmentation, we applied it to two widely used FCN segmentation models based on classic VGGNet [17] architecture (denote as VGGNet), and the more recent FCN using residual network architecture (101-layer, denoted as ResNet [15, 16]). The top 3 results from 21 teams for the ISIC 2017 challenge [10] were used for further comparison. The common segmentation evaluation metrics including the dice similarity coefficient (Dice) and Jaccard (Jac.) were used.

Table 1: Segmentation results for the ISIC 2017 Skin Lesion Challenge dataset. **Red** represents the best and **Blue** the second best results.

		Dice	Jac.
Overall	**Team – Mt. Sinai [18]**	**84.90**	**76.50**
	Team – NLP LOGIX [19]	84.70	76.20
	Team – BMIT [20]	84.40	76.00
	VGGNet	80.87	71.56
	VGGNet+SAL	81.35	72.34
	ResNet	84.69	76.21
	ResNet+SAL	**85.16**	**77.14**
Non-Melanoma	**Team – Mt. Sinai [18]**	85.81	77.78
	Team – NLP LOGIX [19]	**86.07**	**78.02**
	Team – BMIT [20]	85.50	77.60
	VGGNet	81.74	72.88
	VGGNet+SAL	82.04	73.47
	ResNet	85.65	77.56
	ResNet+SAL	**85.87**	**78.17**
Melanoma	**Team – Mt. Sinai [18]**	**81.04**	**71.20**
	Team – NLP LOGIX [19]	79.07	68.82
	Team – BMIT [20]	79.62	69.28
	VGGNet	77.29	66.11
	VGGNet+SAL	78.50	67.67
	ResNet	80.74	70.60
	ResNet+SAL	**82.23**	**72.91**

3.2. Results and Discussions

Table 1 and Fig. 3 show that our method improves upon traditional VGGNet and ResNet based FCN segmentation methods and reflects the advantage of adding more variants of the skin lesion features to the original training data that then enables better learning. As expected, the differences between VGGNet and ResNet indicate the benefit of the deep residual architecture for segmentation, where residual blocks allowed the increase in overall depth of the network (101-layer in ResNet compared with 16-layer in VGGNet) and thus resulted in more meaningful image features.

Table 1 shows that our ResNet+SAL outperforms the state-of-the-art methods. Our approach improved on the Jaccard measure by 1.71% compared to team Mt. Sinai, by 4.09% for team NLP LOGIX and 3.63% for the BMIT. Generally, melanoma studies are more difficult to segment, due to the marked inhomogeneity and non-uniformity of the boundary patterns. We attribute our enhancement to using SAL to derive additional class-specific e.g., melanoma and non-melanoma characteristics of the skin lesions. Hence the FCN can segment non-melanoma studies as well as the more challenging melanoma studies and ensures a balanced segmentation performance across melanoma and non-melanoma studies.

In Fig. 4, we show how the derived output images from SAL have various melanoma and non-melanoma characteristics and further that the output from the second stage is markedly improved and artifacts have been minimized.

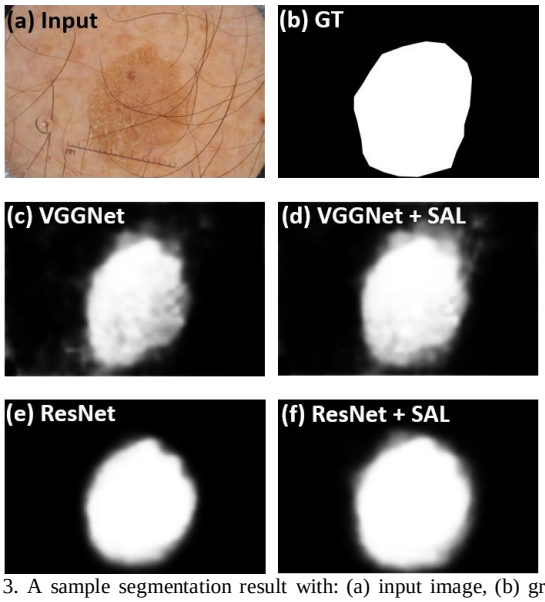

Figure 3. A sample segmentation result with: (a) input image, (b) ground truth annotation, (c-f) results from VGGNet, VGGNet+SAL, ResNet and ResNet+SAL.

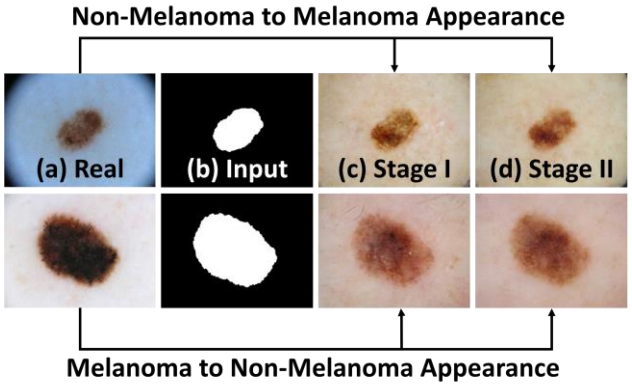

Figure 4. SAL feature learning results with: (a) real dermocopic images; (b) input label images; (c) and (d) derived additional training features at the first and the second stage of SAL.

4. CONCLUSION AND FUTURE WORK

We proposed a new method to improve FCN based segmentation for dermoscopic images via a stacked adversarial learning (SAL) approach. Our SAL learns skin lesion features iteratively in a class-specific manner e.g., melanoma and non-melanoma, and so improves the diversity of the features in the training data. Our experiments with the ISIC 2017 skin lesion challenge dataset show that our method improved the VGGNet and the recent ResNet based FCN segmentation methods. Further, when we coupled our SAL to ResNet it was the best performed method.

REFERENCES

[1] M. E. Celebi, et al., "A methodological approach to the classification of dermoscopy images," Comput. Med. Imag. Grap., 2007.

[2] M. E. Celebi, et al., "Automatic detection of blue-white veil and related structures in dermoscopy images," Comput. Med. Imag. Grap., 2008.

[3] E. Ahn, et al., "Saliency -based Lesion Segmentation via Background Detection in Dermoscopic Images," IEEE J. Biomed. Health Inform., 2017.

[4] M. Silveira, et al., "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images," IEEE Journal of Selected Topics in Signal Processing, 2009.

[5] J. Long, et al., "Fully convolutional networks for semantic segmentation," in CVPR, 2015.

[6] H.-C. Shin, et al., "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," IEEE Trans. on Med. Imag., 2016.

[7] Y. Yuan, et al., "Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks with Jaccard Distance," IEEE Trans. on Med. Imag., 2017.

[8] L. Bi, et al., "Dermoscopic Image Segmentation via Multi-Stage Fully Convolutional Networks, " IEEE Trans. Biomed. Eng., 2017.

[9] I. Goodfellow, et al., "Generative adversarial nets," In NIPS, 2014.

[10] Noel C. F. Codella, et al., "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)," 2017; arXiv:1710.05006.

[11] L. Bi, et al., "Step-wise Integration of Deep Class-specific Learning for Dermoscopic Image Segmentation," Pattern Recognition, 2018.

[12] A. Krizhevsky, et al., "Imagenet classification with deep convolutional neural networks," in NIPS, 2012.

[13] M. Mirza, et al., "Conditional generative adversarial nets," arXiv preprint arXiv:1411.1784, 2014.

[14] P. Isola, et al., "Image-To-Image Translation with Conditional Adversarial Networks, " In CVPR, 2017.

[15] K. He, et al., "Deep residual learning for image recognition, " In CVPR, 2016.

[16] G. Lin, et al., "Refinenet: Multi-path refinement networks with identity mappings for high-resolution semantic segmentation, " In CVPR, 2017.

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

[18] Y. Yuan, et al., "Automatic skin lesion segmentation with fully convolutional-deconvolutional networks," arXiv preprint arXiv:1703.05165, 2017.

[19] M. Berseth, "ISIC 2017-skin lesion analysis towards melanoma detection," arXiv:1703.00523, 2017.

[20] L. Bi, et al., "Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks," arXiv:1703.04197, 2017.