

# Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks with Jaccard Distance

Yading Yuan\*, Ming Chao, and Yeh-Chi Lo

**Abstract**—Automatic skin lesion segmentation in dermoscopic images is a challenging task due to the low contrast between lesion and the surrounding skin, the irregular and fuzzy lesion borders, the existence of various artifacts, and various imaging acquisition conditions. In this article, we present a fully automatic method for skin lesion segmentation by leveraging a 19-layer deep convolutional neural networks (CNNs) that is trained end-to-end and does not rely on prior knowledge of the data. We propose a set of strategies to ensure effective and efficient learning with limited training data. Furthermore, we design a novel loss function based on Jaccard distance to eliminate the need of sample re-weighting, a typical procedure when using cross entropy as the loss function for image segmentation due to the strong imbalance between the number of foreground and background pixels. We evaluated the effectiveness, efficiency, as well as the generalization capability of the proposed framework on two publicly available databases. One is from ISBI 2016 *Skin Lesion Analysis Towards Melanoma Detection Challenge*, and the other is the PH2 database. Experimental results showed that the proposed method outperformed other state-of-the-art algorithms on these two databases. Our method is general enough and only needs minimum pre- and post-processing, which allows its adoption in a variety of medical image segmentation tasks.

**Index Terms**—deep learning, fully convolutional neural networks, image segmentation, jaccard distance, melanoma, dermoscopy

## I. INTRODUCTION

Malignant melanoma is one of the most rapidly increasing cancers all over the world, with the estimated new cases of 76,380 and estimated death of 10,130 in the United States in 2016 [1]. Nowadays, despite the fact that some advanced treatment techniques, such as radiation therapy and immunotherapy, are increasingly combined with surgery in clinical practice [2], the five year survival rate for advanced stage melanoma is still as low as 15% while it is over 95% for early-stage melanoma [3]. This discrepancy highlights the critical importance of timely diagnosis and treatment of melanoma for patient survival.

Dermoscopy is a popular *in vivo* non-invasive imaging tool that uses polarized light to aid dermatologists in examining pigmented skin lesions based on a set of morphological features. Although dermoscopy has been shown to lead to increased diagnostic accuracy comparing to the conventional

\* Y. Yuan, M. Chao and Y-C Lo are with the Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, NY, USA, e-mail: yading.yuan@mssm.edu, ming.chao@mountsinai.org, yeh-chi.lo@mountsinai.org.

Copyright (c) 2017 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

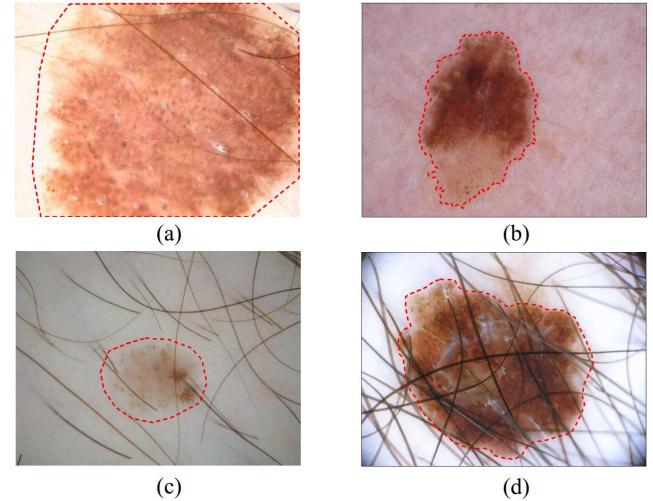


Fig. 1. Challenges of automated lesion segmentation on dermoscopic images. (a) large variety of sizes; (b) irregular and fuzzy borders; (c) low contrast between lesion and the surrounding skin; and (d) artifacts. Red dash line represents the skin lesions contoured by dermatologists.

ABCD criteria [4], proper interpretation of dermoscopic images is normally time-consuming, complex, and prone to suffer from inter- and intra-observer variabilities. Therefore, computerized analysis methods have been developed to assist dermatologists in improving their efficiency and objectivity of visual interpretation of dermoscopic images. A review of the state-of-the-art methods up to 2012 is presented in [5].

Automatically segmenting melanoma from the surrounding skin is an essential step in computerized analysis of dermoscopic images [6] [7]. However, this task is not trivial because melanoma usually has a large variety of appearance in size, shape, and color along with different types of skin and texture. Meanwhile, some lesions have irregular and fuzzy borders, and in some cases the contrast between lesion and the surrounding skin is pretty low. In addition, artifacts and intrinsic cutaneous features, such as hairs, frames, blood vessels and air bubbles can make the automatic segmentation more challenging, as illustrated in Figure 1.

## II. PREVIOUS WORK

In the past decades, researchers have developed various computer algorithms to conquer these challenges. These algorithms can be broadly classified as clustering, thresholding, region merging and splitting, active contour models and supervised learning. The advantages and drawbacks of each method

have been discussed and compared in many articles such as [5], [8], [9].

In some recent efforts to tackle this problem, Celebi et al. [10] developed an automated method for skin lesion border detection by ensembling four different thresholding methods. Peruch et al.[11] proposed a Mimicking Expert Dermatologists' Segmentation (MEDS) method, in which Principal Component Analysis (PCA) is firstly applied to project the RGB images onto the first principal component of the color histogram, then a thresholding technique that mimicks the cognitive process of human dermatologists is used to cluster the colors (and thus pixels) into lesional and non-lesional skin. Thresholding-based methods rely on the histogram distribution of image color (or its variants), which may be undesirably altered in cases when significant amount of hair and bubbles are present. In order to improve the effectiveness of the classic GVF snake in the presence of distractions or noise in the vicinity of a real boundary, Zhou et al. [12] integrated a mass density function into the optimization objective functional, which can be solved with the support of mean shift estimation. The optimization procedure, however, involves a large amount of computation to achieve convergence. In [13], Sadri et al. introduced a fixed-grid wavelet networks in which orthogonal least squares was used to calculate the network weights and to optimize the network structure in a supervised way. Another supervised method was proposed by Xie et al. in [14], where a self-generating neural network is combined with the genetic algorithms for skin lesion segmentation. These supervised methods, along with the aforementioned unsupervised ones, employ hand-crafted features that require specialized domain knowledge. Meanwhile, pre-processing steps, such as color space transform, contrast enhancement, lesion localization and artifact removal are usually introduced to facilitate the lesion segmentation procedure [15].

Recently, convolutional neural networks (CNNs) [16] have become one of the most powerful tools in machine learning and computer vision. These models have the capability of learning hierarchical features from raw image data, and thus the hand-crafted features are not needed. Besides the success in natural image recognition tasks such as image classification [17], CNNs have also showed promising performance in various medical image computing problems such as mitosis detection on histology images [18], body parts recognition on CT images [19], cerebral microbleeds detection on MR images [20], 2D/3D image registration [21] and skin cancer classification [22]. Specifically for medical image segmentation, CNNs have been applied to segment brain tumor on MR images [23], urinary bladder on CT urography [24] left ventricle on cardiac MRI [25], and skin lesion from non-dermoscopic images [26] among others. These segmentation algorithms perform patch-based image classification, in which the entire input image is divided into small patches and the CNN model is applied onto each patch to predict if it is within the target (foreground) or outside of the target (background). Since each patch only represents a local region of the image, this approach only incorporates limited contextual information contained in the patch. Enlarging patch size can increase the contextual information, but excessively large patch size may

lose fine details in the final segmentation. Sliding-window [18] can also be implemented to incorporate both local and global context into segmentation, however, this procedure is computationally demanding and inefficient due to the highly overlapping patches.

Inspired by the latest advances in deep learning research, we propose a novel framework based on deep CNNs to automatically segment skin lesions in dermoscopic images. Instead of developing sophisticated pre- and post-processing algorithms and hand-crafted features, we focus on designing appropriate network architecture and effective training strategies such that our deep learning model can handle images under various acquisition conditions, which makes our method highly scalable. In this study, we exploit a model known as fully convolutional network (FCN) [27], [28] that extends the convolution process across the entire image and predicts the segmentation mask as a whole. As the result, the redundant computation encountered in the patch-based methods can be avoided. In the current applications of FCN on segmentation, binary cross entropy between predicted values and targets is usually used as the loss function. However, since the skin lesion usually occupies a small portion of the dermoscopic image, the minimization of cross entropy tends to be biased toward background. In order to address this issue, we propose a novel loss function based Jaccard distance that is tailed to medical image segmentation problem. We evaluate our model on two publicly available databases. One is ISBI 2016 challenge dataset for skin lesion analysis towards melanoma detection [29], and the other is the PH2 database [30]. Our model shows superior segmentation performance than those state-of-the-art methods on both databases.

Our contributions in this paper are three fold. Firstly, we introduce a fully automated method for skin lesion segmentation by leveraging the discriminative power of a 19-layer deep FCN. To the best of our knowledge, our work is among the first few attempts to use deep neural networks to tackle this challenging problem [31], [32]. This model does not rely on the prior knowledge of data and is trained in an end-to-end fashion. We investigate a set of training strategies to ensure effective and efficient learning with limited training data. Secondly, we design an appropriate loss function that naturally handles the lesion-background imbalance of pixel-wise classification for medical image segmentation. Our results show that this loss function can further improve the segmentation performance. At last, We extensively evaluate the effectiveness, efficiency and the generalization capability of the proposed model using two large databases. Our model can be easily generalized to other challenging medical image segmentation problems.

This paper is organized as follows. we introduce the details of the proposed FCN-based segmentation method in Section III, then report the experimental design and results in Section IV. Finally, we discuss the results in Section V and conclude our study in Section VI.

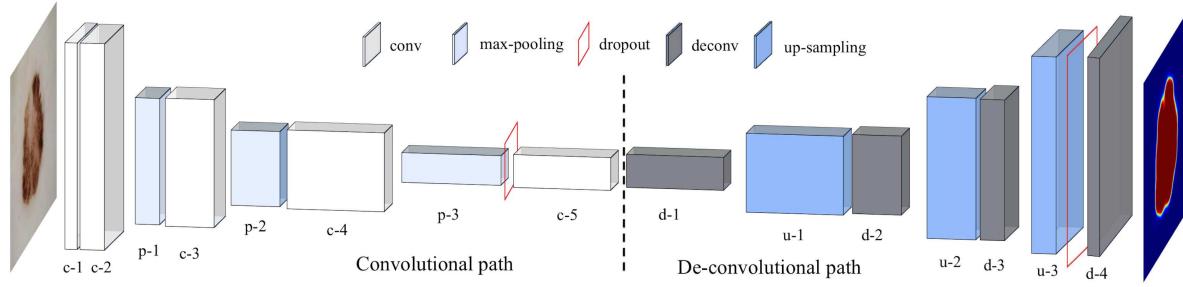


Fig. 2. Architecture of the proposed fully convolutional network (FCN). This model consists two pathways, in which contextual information is aggregated via convolution (c) and pooling (p) in the convolutional path and full image resolution is recovered via deconvolution (d) and up-sampling (u) in the deconvolutional path. The architectural details are described in Table I

TABLE I  
ARCHITECTURAL DETAILS OF THE PROPOSED FCN MODEL.

Conv	Filter	Output	Deconv	Filter	Output
c-1	$5 \times 5$	$188 \times 252 \times 8$	d-1	$5 \times 5$	$21 \times 29 \times 64$
c-2	$3 \times 3$	$186 \times 250 \times 16$	u-1	$2 \times 2$	$42 \times 58 \times 64$
p-1	$2 \times 2$	$93 \times 125 \times 16$	d-2	$4 \times 4$	$45 \times 61 \times 32$
c-3	$4 \times 4$	$90 \times 122 \times 32$	u-2	$2 \times 2$	$90 \times 122 \times 32$
p-2	$2 \times 2$	$45 \times 61 \times 32$	d-3	$4 \times 4$	$93 \times 125 \times 16$
c-4	$4 \times 4$	$42 \times 58 \times 64$	u-3	$2 \times 2$	$186 \times 250 \times 16$
p-3	$2 \times 2$	$21 \times 29 \times 64$	d-4	$3 \times 3$	$188 \times 252 \times 8$
c-5	$5 \times 5$	$17 \times 25 \times 64$	output	$5 \times 5$	$192 \times 256 \times 1$

### III. FULLY CONVOLUTIONAL NETWORKS

#### A. Network architecture

We train a FCN to map from an RGB input to a posterior probability map using the architecture shown in Figure 2. The network contains 19 layers with 290,129 trainable parameters, and Table I describes the architectural details. We fix the stride as 1 and use Rectified Linear Units (ReLUs) as the activation function [17] for each convolutional/deconvolutional layer. For output layer, we use *sigmoid* function as the activation function. Pixel-wise classification is performed and FCN is essentially 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.

With the successive convolution and pooling layers, CNNs can integrate contextual information from regional to global scales, resulting in reduced resolution in the output layer. In contrast, image segmentation calls for classifying each pixel into either foreground or background in combination with full-resolution output. In order to address this conflict between multi-scale information aggregation and full-resolution pixel-wise classification, we implement a strategy of using up-sampling and deconvolutional layers to recover lost resolution while carrying over the global perspective from pooling layers [33].

The up-sampling layer performs the reverse operation of pooling and reconstructs the original size of activation, and the deconvolutional layer densifies the coarse activation map obtained from up-sampling through swapping the forward and backward passes of a convolution, thus a single input activation is projected into multiple outputs after deconvolution, yielding an enlarged and dense feature map. Echo to the convolution path where image information is aggregated from fine details

to global concept, a hierarchical structure of deconvolutional layers is used to recover image details at different levels, with the lower layers encoding overall image information and higher layers capturing fine details regarding skin tumors. In this way, the network can take both global information and fine details into account for tumor segmentation.

#### B. Loss function

For any pixel in the input image  $x_{ij}$ , the corresponding FCN output  $p(\mathbf{w}|x_{ij})$  represents an estimated *posterior* probability that this pixel belongs to the skin tumor. Since FCN essentially performs a pixel-wise classification, cross entropy is usually used as the loss function [27][28]:

$$L(\mathbf{w}) = -\frac{1}{N_i \times N_j} \sum_{i,j} [t_{ij} \ln p(\mathbf{w}|x_{ij}) + (1 - t_{ij}) \ln(1 - p(\mathbf{w}|x_{ij}))] \quad (1)$$

where  $t_{ij} \in \{0, 1\}$  is the actual class of  $x_{ij}$  with  $t_{ij} = 1$  for tumor and  $t_{ij} = 0$  for background.

In dermoscopy, a skin tumor usually occupies a small region in the whole dermoscopic image. As the result, the pixel-wise classification tends to be biased towards the background and therefore there is a higher chance for tumor to be partially segmented or even missed. A typical solution is to assign a weight to each pixel during training that compensates the different frequencies of pixels from each class, so the contributions of tumor and background are re-balanced [28]. However, this pixel-wise re-weighting procedure brings additional computation cost, especially when image augmentations are employed in the training procedure.

In this work, we propose a novel loss function based on Jaccard distance. Jaccard index [34], also known as the Jaccard similarity coefficient, is one of the most frequently used evaluation measures in medical image segmentation. The Jaccard distance, which measures dissimilarity between two sets, is complementary to the Jaccard index. Let  $M$  represent the ground truth of segmentation, which is normally a manually-identified tumor region, and  $C$  represent a computer-generated mask, the Jaccard distance is defined as:

$$d_J(M, C) = 1 - J(M, C) = 1 - \frac{|M \cap C|}{|M| + |C| - |M \cap C|}. \quad (2)$$

$d_J(MC)$  itself is not differentiable, which makes it difficult to be directly applied into backpropagation. Even some minimization methods for non-differentiable functions can be introduced, it would be computationally expensive to generate a binary mask from continuous FCN output for each iteration during optimization. Thus, we designed the following loss function:

$$L_{d_J} = 1 - \frac{\sum_{i,j} (t_{ij} p_{ij})}{\sum_{i,j} t_{ij}^2 + \sum_{i,j} p_{ij}^2 - \sum_{i,j} (t_{ij} p_{ij})}, \quad (3)$$

With this loss function, a weight map is not needed to re-balance pixels from tumor region and background. Meanwhile, the proposed loss function is differentiable:

$$\begin{aligned} \frac{\partial L_{d_J}}{\partial p_{ij}} &= -\frac{t_{ij} [\sum_{i,j} t_{ij}^2 + \sum_{i,j} p_{ij}^2 - \sum_{i,j} (t_{ij} p_{ij})]}{[\sum_{i,j} t_{ij}^2 + \sum_{i,j} p_{ij}^2 - \sum_{i,j} (t_{ij} p_{ij})]^2} \\ &\quad + \frac{(2p_{ij} - t_{ij}) [\sum_{i,j} (t_{ij} p_{ij})]}{[\sum_{i,j} t_{ij}^2 + \sum_{i,j} p_{ij}^2 - \sum_{i,j} (t_{ij} p_{ij})]^2}, \end{aligned} \quad (4)$$

which can be efficiently integrated into backpropagation during network training.

### C. FCN Training

Training a deep network model with a limited number of samples is a challenging task. As described in the previous section, our model has 19 layers and 290,129 parameters to be learned. Even though the number of training samples available in this study is considered as a large-scale study in the field of digital dermoscopic image analysis, it is still relatively small as compared to the size of the network. In order to address this issue, we firstly apply FCN model to convert image segmentation into a pixel-wise classification problem. As the result, each pixel can be considered as an independent sample during network training, which greatly boosts the number of equivalent training samples. Then, We initialize the network weights using Xavier's technique [35] and employ the following strategies to improve the efficiency of network training while reducing overfitting.

1) *Adam stochastic optimization*: Stochastic gradient descent (SGD) with mini-batch is usually employed as the optimization algorithm for neural network training [36]. It is well known that learning rate is one of the critical hyperparameters that have a significant impact on classification performance. However, selecting proper learning rate and strategy can be fairly challenging. One commonly used strategy is to anneal the learning rate at each iteration  $t$  as  $\frac{\alpha}{t_0+t}$  where  $\alpha$  and  $t_0$  dictate the initial learning rate and the time when the annealing starts respectively, but it tends to have slow convergence when the loss function is highly sensitive to some directions in parameter space while insensitive to others. The momentum algorithm [37] can mitigate this issue, but with the expense of introducing another hyperparameter.

We adopt adam optimization algorithm [38], or adaptive moments, to adjust the learning rate based on the first and

the second-order moments of the gradient at each iteration. Here the momentum is incorporated as an estimate of the first moment and the effective stepsize at each iteration  $t$  depends on the ratio between the bias-corrected first and second-order moments as  $\Delta_t = \alpha \cdot \hat{m}_t / \sqrt{\hat{v}_t}$ , with a smaller value indicating that there is greater uncertainty about whether the direction of  $\hat{m}_t$  corresponds to the direction of the true gradient. We found adam is fairly robust to the choice of hyperparameters, and set the learning rate  $\alpha$  as 0.003 to speed up the training procedure in this study.

2) *Batch normalization*: We employ the batch normalization [39] to reduce the internal covariate shift by normalizing the input distribution of every layer to the standard Gaussian distribution for each training mini-batch. For this purpose, we add a batch normalization layer to the output of every convolutional and de-convolutional layer. The batch normalization is performed over all locations in the same feature map such that different elements in the same map are normalized in the same way. We set batch size as 18 in this study.

3) *Dropout*: Dropout [40] provides a powerful but computationally inexpensive way to reduce overfitting when training a very deep FCN model with limited data. This technique sets the output of each neuron in a given layer to zero with probability  $p$ , thus removes the contribution of those "dropped out" neurons from both forward pass and back-propagation. The subset of disabled neurons is drawn independently for each mini-batch and forms a different network architecture, then dropout trains the ensemble of all sub-networks that have different architectures but share weights in one epoch. In this way, a neuron cannot rely on the presence of particular other neurons and it is, therefore, forced to learn more robust features that are useful among different random subsets. This makes the trained FCN model more robust and improves the generalization ability.

We use dropout with  $p = 0.5$  before *conv5* and *deconv4* layers in Figure 2. Although it roughly doubles the number of iterations for convergence, dropout does reduce the overfitting substantially in our experiments.

4) *Image augmentation*: In order to improve the robustness of the proposed FCN model under a wide variety of image acquisition conditions, we employ image augmentation to further reduce model overfitting by artificially enlarging the training dataset with various image transformations. For the tumor segmentation on dermoscopic images, we primarily look for invariance to mild geometric transformation as well as robustness to pixel value variations, so we implement the following two types of image augmentations. Note that these augmentations only require little extra computation, so the transformed images are generated from the original images for every mini-batch within each iteration.

One type of image augmentation consists of a series of geometric transformations. Each original image in a mini-batch is firstly flipped horizontally with probability of 0.5, then flipped vertically with another probability of 0.5. The rotation operation is applied to rotate the image by angle  $\theta$ , which is randomly sampled from a Gaussian distribution in the range of  $[-40^\circ, 40^\circ]$ . This image is then rescaled by a Gaussian random factor in the range of  $[0.7, 1.3]$ , and shifted by a

uniformly-distributed random  $2d$  displacement vector within the range of 10 pixels. Although the resulting training samples are highly dependent, this scheme significantly increases the size of training set and allows us to use a deeper FCN without substantial overfitting.

The other type of image augmentation focuses on randomly normalizing the contrast of each channel in the training images. Specifically, given  $S_c$  as the  $c$ -th channel of the input image, we introduce  $l_L$  and  $l_H$ -th percentile of  $S_c$ , where  $l_L$  and  $l_H$  are randomly selected from [0, 20] and [90, 100] respectively. Then the new value at each pixel in this channel can be determined as:

$$S'_c = \begin{cases} 0 & S_c < S_c(l_L) \\ \frac{S_c - S_c(l_L)}{S_c(l_H) - S_c(l_L)} & S_c(l_L) \leq S_c \leq S_c(l_H) \\ 1 & S_c > S_c(l_H) \end{cases}. \quad (5)$$

This scheme simulates the scenario that the input images are acquired with various intensities, contrasts and color illuminations. At test time,  $l_L$  and  $l_H$  are fixed as 10 and 95, respectively.

#### D. Implementation

Our FCN-based lesion segmentation algorithm was implemented with Python based on Theano [41] and Lasagne packages. The experiments were conducted on a Dell XPS 8900 desktop with Intel(R) i7-6700 3.4 GHz CPU and a GPU of Nvidia GeForce GTX 1060 with 6GB GDDR5 memory.

Our aim is to have the FCN model automatically learn those features that are useful for lesion segmentation, so a simple pre-processing is employed to facilitate the following learning procedure while preserving the original image information. By observing most of images in the training set have a height to width ratio of 3 : 4, we resize the images to  $192 \times 256$  using bi-linear interpolation. RGB channels are kept as the input to the FCN model and each channel is rescaled to [0, 1].

The output of FCN model is a posterior probability map where each pixel value represents the probability that the pixel belongs to the lesion. We employ a dual-threshold method to obtain a binary tumor mask from this probability map. Specifically, a relatively high threshold ( $th_H = 0.8$  in our experiments) is firstly applied to the FCN output to generate a series of tumor candidates. The mass of each candidate region, which is the area weighted by the pixel values within this region, is calculated. The tumor center is chosen as the centroid of the region that has the largest mass among these candidates. Then a lower threshold ( $th_L = 0.5$ ) is applied to the probability map. After filling small holes with morphological dilation, the final tumor mask is determined as the region that embraces the tumor center. Since this post-processing only involves thresholding and morphological operations, the whole lesion segmentation can be done within two seconds.

Finally, a bagging-type ensemble strategy is implemented to combine outputs of different FCNs to further improve the image segmentation performance on the testing images. For each fold in the 5-fold cross validation, the total number of epochs is set as 500, and the optimal epoch that yields the best

performance on the validation dataset is saved. Also saved is the corresponding trained FCN model. After cross validation is completed, another FCN model is trained using the entire training data and the total number of epochs is set as the maximum value among the optimal epochs obtained from cross validation. The outputs of these six FCN models are then averaged to predict the final segmentation.

## IV. EXPERIMENTAL DESIGN AND RESULTS

### A. Databases

We used two dermoscopy databases in our experiments. One is provided by the International Skin Imaging Collaboration (ISIC) for the 2016 International Symposium on Biomedical Imaging (ISBI 2016) challenge titled “Skin lesion analysis toward melanoma detection” [29]. This database includes a training dataset with 900 annotated dermoscopic images (173 melanomas), and a testing dataset with 379 images (75 melanomas). They are 8-bit RGB images and the image size ranges from  $542 \times 718$  to  $2848 \times 4288$ . Our model was trained and validated on the training set, and evaluated on the testing set. In order to further evaluate the generalization ability of our model, we also tested with an independent database called PH2 [30] collected from Dermatology Service of Hospital Pedro Hispano in Portugal, which contains 200 dermoscopic images including 80 common nevi, 80 atypical nevi, and 40 melanomas. The image size in PH2 is fixed as  $560 \times 768$ .

### B. Performance evaluation

The output of the FCN model is binarized to a lesion mask. The performance of the proposed segmentation algorithm was assessed by comparing the computer-generated lesion mask with the ground truths created by expert clinicians. We used the evaluation metrics suggested in the ISBI 2016 challenge for benchmarking [29], including pixel-wise accuracy (AC), sensitivity (SE), specificity (SP), dice coefficient (DI), and Jaccard index (JA). Let  $TP$ ,  $TN$ ,  $FP$ ,  $FN$  refer to the number of true positives, true negatives, false positives, and false negatives respectively, the evaluation metrics are defined as:

- $AC = (TP + TN) / (TP + FP + TN + FN)$
- $SE = TP / (TP + FN)$
- $SP = TN / (TN + FP)$
- $DI = 2 \cdot TP / (2 \cdot TP + FN + FP)$
- $JA = TP / (TP + FN + FP)$ .

### C. Key component validation

As the first step, we analyzed the effect of some key components of the proposed model to the lesion segmentation performance. These components included input image size, optimization method, image augmentation, as well as cost function. In each experiment, we replaced one component with an alternative setting while keeping other components unchanged, then trained and evaluated the modified model using 5-fold cross validation on the ISBI training set, where 720 images were used for training and the rest 180 images were used for validation in each fold.

TABLE II  
RESULTS WITH DIFFERENT INPUT SIZES

Image size	AC	DI	JA	SE	SP
96 × 128	0.963	0.918	0.855	<b>0.931</b>	0.964
192 × 256	<b>0.963</b>	<b>0.922</b>	<b>0.861</b>	0.926	0.971
384 × 512	0.953	0.900	0.833	0.894	<b>0.976</b>

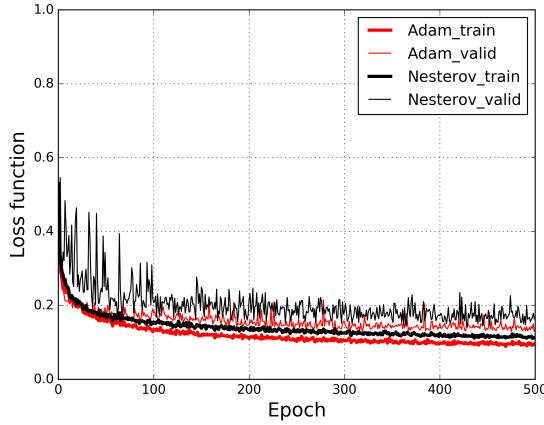


Fig. 3. Comparison between Adam and SGD with Nesterov momentum for training and validating of the proposed model.

1) *Image size*: We compared the segmentation performance using three different image sizes: 96 × 128, 192 × 256, and 384 × 512. The results, as summarized in Table II, show that the Jaccard index was improved from 0.855 to 0.861 when increasing image size from 96 × 128 to 192 × 256, but reduced to 0.833 when further increasing the image size to 384 × 512. This performance reduction was caused by the fact that the quadrupled number of pixels made the model training more challenging. While increasing the model capacity, such as adding more layers or features, may improve the performance for 384 × 512 image size, it would be very time-consuming. Currently, one epoch training with 384 × 512 took 110 seconds, while it was 27 seconds for 192 × 256 and 8 seconds for 96 × 128. These data indicate that image size of 192 × 256 yielded a good balance between segmentation performance and computation cost, so we resized all the images to this dimension before putting them into the FCN model.

2) *Adam optimization*: We compared the adam optimization method to SGD with Nesterov momentum, in which the learning rate was set as 0.003 and momentum as 0.9. Figure 3 shows how the loss function is minimized on both training and validation datasets as network training processes - one with Adam optimization (red lines) and the other with Nesterov momentum (black lines). It is clear to see that Adam makes rapid progress lowering the loss function and consistently converges faster than Nesterov momentum on both training and validation datasets. Meanwhile, the validation curve of Nesterov momentum is quite noisy, which makes it less reliable than Adam when picking a model for testing phase. As shown in Table III, the Jaccard index was reduced to 0.838 when using SGD with Nesterov momentum.

3) *Image augmentation*: We implemented two types of image augmentations in this work, with one consisting of

TABLE III  
RESULTS WITH DIFFERENT OPTIMIZATION METHODS

Optimization	AC	DI	JA	SE	SP
SGD+Nesterov	0.954	0.907	0.838	0.910	<b>0.974</b>
Adam	<b>0.963</b>	<b>0.922</b>	<b>0.861</b>	<b>0.926</b>	0.971

TABLE IV  
RESULTS WITH DIFFERENT IMAGE AUGMENTATION STRATEGIES

Augmentation	AC	DI	JA	SE	SP
w/o	0.946	0.894	0.820	0.899	0.965
CRN only	0.951	0.904	0.834	0.915	0.963
GT only	0.957	0.911	0.844	0.914	<b>0.973</b>
full	<b>0.963</b>	<b>0.922</b>	<b>0.861</b>	<b>0.926</b>	0.971

geometric transformations (GT) and the other focusing on image contrast random normalization (CRN). In order to investigate the effect of each image augmentation strategy to the final segmentation performance, we conducted the following three experiments. In the first experiment, the FCN model was trained with the original images and no image augmentation was applied, which yielded an average Jaccard index of 0.820 with training time of 11 seconds per epoch. We then artificially increased training set by altering the contrast of each image channel in the second experiment and applying geometric transformations in the third experiment, respectively. As can be observed from Table IV, both image augmentation strategies improved the segmentation performance, with 1.7% gain by altering contrast and 2.8% gain by geometric transformations, respectively. The training time per epoch was 12 seconds when including contrast alteration and 26 seconds when including geometric transformations. Since geometric transformations include image flipping (vertically and/or horizontally), rotation and scaling, this augmentation strategy can potentially generate more diverse training data as compared to simple contrast alteration, thus yield a better segmentation performance. Finally, the combination of these two strategies improved the Jaccard index by about 5%.

4) *Loss function*: We compared the proposed cost function based on Jaccard distance with the conventionally used cross entropy. The specificity using cross entropy as cost function was 0.972, which was slightly higher than the one using Jaccard distance (0.971). This is because the cross entropy models the contribution from both foreground and background pixels while Jaccard distance focuses on the foreground pixels, which is more related to the task of image segmentation. As the results shown in Table V, the cost function based on Jaccard distance yielded a higher sensitivity (0.926 vs. 0.918) and a higher Jaccard index (0.861 vs. 0.854).

#### D. Evaluation on ISBI 2016 testing database

After validating those key components in our model, we applied the trained network to the 379 dermoscopic images in ISBI 2016 testing dataset. Figure 4 shows several challenging examples of automatic segmentation on both melanoma and non-melanoma lesions, including cases with irregular shapes (Fig. 4(a) - (d)), cases with dark corners (Fig. 4 (b) and (c)), cases with inhomogeneous appearance (Fig. 4 (a), (b) and

TABLE V  
RESULTS WITH DIFFERENT LOSS FUNCTIONS

Loss function	AC	DI	JA	SE	SP
Cross entropy	0.961	0.916	0.854	0.918	<b>0.972</b>
Jaccard distance	<b>0.963</b>	<b>0.922</b>	<b>0.861</b>	<b>0.926</b>	0.971

TABLE VI  
RESULTS ON THE ISBI 2016 CHALLENGE TESTING DATASET WITH OUR MODEL COMPARED TO THE TOP THREE PERFORMERS

Method	AC	DI	JA	SE	SP
FCN 1	0.951	0.900	0.831	0.912	0.957
FCN 2	0.950	0.900	0.832	0.908	0.960
FCN 3	0.951	0.896	0.829	0.913	0.962
FCN 4	0.951	0.898	0.831	0.897	0.966
FCN 5	0.948	0.895	0.826	0.904	0.958
FCN 6	0.953	0.903	0.836	0.904	0.967
FCN ensemble	<b>0.955</b>	<b>0.912</b>	<b>0.847</b>	<b>0.918</b>	0.966
EXB	0.953	0.910	0.843	0.910	0.965
CUMED [32]	0.949	0.897	0.829	0.911	0.957
Mahmudur	0.952	0.895	0.822	0.880	<b>0.969</b>

(d)), cases with heavy hair artifacts (Fig. 4 (e) and (f)), cases with color charts (Fig. 4 (f) and (g)), as well as cases with extremely low contrast (Fig. 4 (g) and (h)). Overall, our model delineated the skin lesion accurately and is robust to various image acquisition conditions.

Figure 5 plots the distribution of the segmentation results on 304 non-melanoma and 75 melanoma lesions. The Jaccard index for non-melanoma lesions ranged from 0.181 to 0.971 with the mean of 0.843, while it ranged from 0.526 to 0.967 with the mean of 0.861 for melanoma lesions.

Table VI compares the results of the proposed FCN model with the top three performers on the ISBI 2016 Challenge testing dataset. Here FCN 1 - 5 are models trained during 5-fold cross validation (720 images for training and 180 images for validation), and FCN-6 was trained using the entire 900 images. Five out of these six models achieved performance that place them as the second rank in the 2016 ISBI challenge on melanoma lesion segmentation, with FCN-6 performing best to yield a Jaccard index of 0.836. Ensembling these FCN models further boosted the segmentation performance in terms of Jaccard index to 0.847, which outperformed the top performer of ISBI 2016 challenge.

#### E. Evaluation on PH2 database

In order to further evaluate the performance of the proposed FCN model, we conducted the following two experiments on PH2, an independent dermoscopic image database that is widely used for algorithm validation and benchmark [30]. In the first experiment, we directly applied the pre-trained model to perform image segmentation on PH2. Note that only FCN-2, the best cross-validated model, was considered in this section to simplify the evaluation procedure. Only the lower threshold ( $th_L$ ) was experimentally changed to 0.6 by visually evaluating the segmentation performance on randomly selected 10 images. In the second experiment, we fine-tuned the weights of FCN-2 by further training the network with PH2 data. Specifically, the PH2 database was randomly divided into

TABLE VII  
PERFORMANCE EVALUATION ON THE PH2 DATABASE

	DM	IWSL	SSLS	RSSLS	FCN	FCN-retrained
DI	—	0.925	0.913	0.911	0.915	<b>0.938</b>
DV(%)	13.92	14.99	17.48	16.45	18.48	<b>12.81</b>

two groups (I and II), each of which included 100 images. We then trained FCN-2 with 50 images and validated with the rest 50 images in group I. The total number of epochs was set to 100. The model performing best on the validation images was selected and applied to perform image segmentation in group II. The same procedure was repeated by using images in group II as training and validation, and the tuned model was tested on group I.

Table VII compares our model with several recently published methods, including a deformable model (DM) by Ma et al. [42], an image-wise supervised learning (IWSL) by Bi et al. [43], and a saliency-based method (SSLS) [44] and its robust variation (RSSLS) by Ahn et al. [45]. Besides DI, we also used divergence value (DV) to evaluate the segmentation performance, which measures the percentage of segmentation errors and can be obtained by:

$$DV = \frac{M \oplus C}{M} \times 100\% = \frac{FP + FN}{TP + FN} \times 100\%, \quad (6)$$

where  $M \oplus C$  is the XOR operation between the ground truth and the computer-generated lesion mask.

The direct application of FCN-2 on PH2 achieved a reasonably good segmentation result with  $DI$  of 0.915 and  $DV$  of 18.48%, even though they are slightly inferior to the results of those state-of-the-art methods. By further training the FCN model with PH2 data, our retrained FCN model achieved the best performance with  $DI$  of 0.938 and  $DV$  of 12.81%. Because we only used 100 images for training and validation, a model retraining can be completed within 5 minutes.

It should be noted that not all the images in PH2 were used in the first three of the fore-mentioned studies due to various reasons. Ma employed 160 images with 8 images diagnosed as melanoma and 152 as either typical or atypical nevi, and Bi excluded 40 images if they either had multiple lesions, or were heavily affected by illumination or hair artifacts. Ahn excluded 40 images where the lesions connected to the image edges in their initial study of SSLS. On the contrary, our study employed all the 200 images to evaluate the proposed FCN model under different conditions. Figure 6 shows a few segmentation results on some challenging cases, including cases affected by hair artifacts (Fig. 6 (a) and (c)) and by illumination variations (Fig. 6 (b)), cases whose boundaries reached the image edges (Fig. 6 (b) and (d)), and two melanoma cases (Fig. 6 (c) and (d)). Our method achieved promising results on these challenging cases, demonstrating that the proposed FCN model is fairly robust to various patient populations and can be quickly adapted to new patient population via transfer learning.

## V. DISCUSSIONS

Delineating lesion from surrounding skin regions often serves as one of the prerequisite steps for computerized

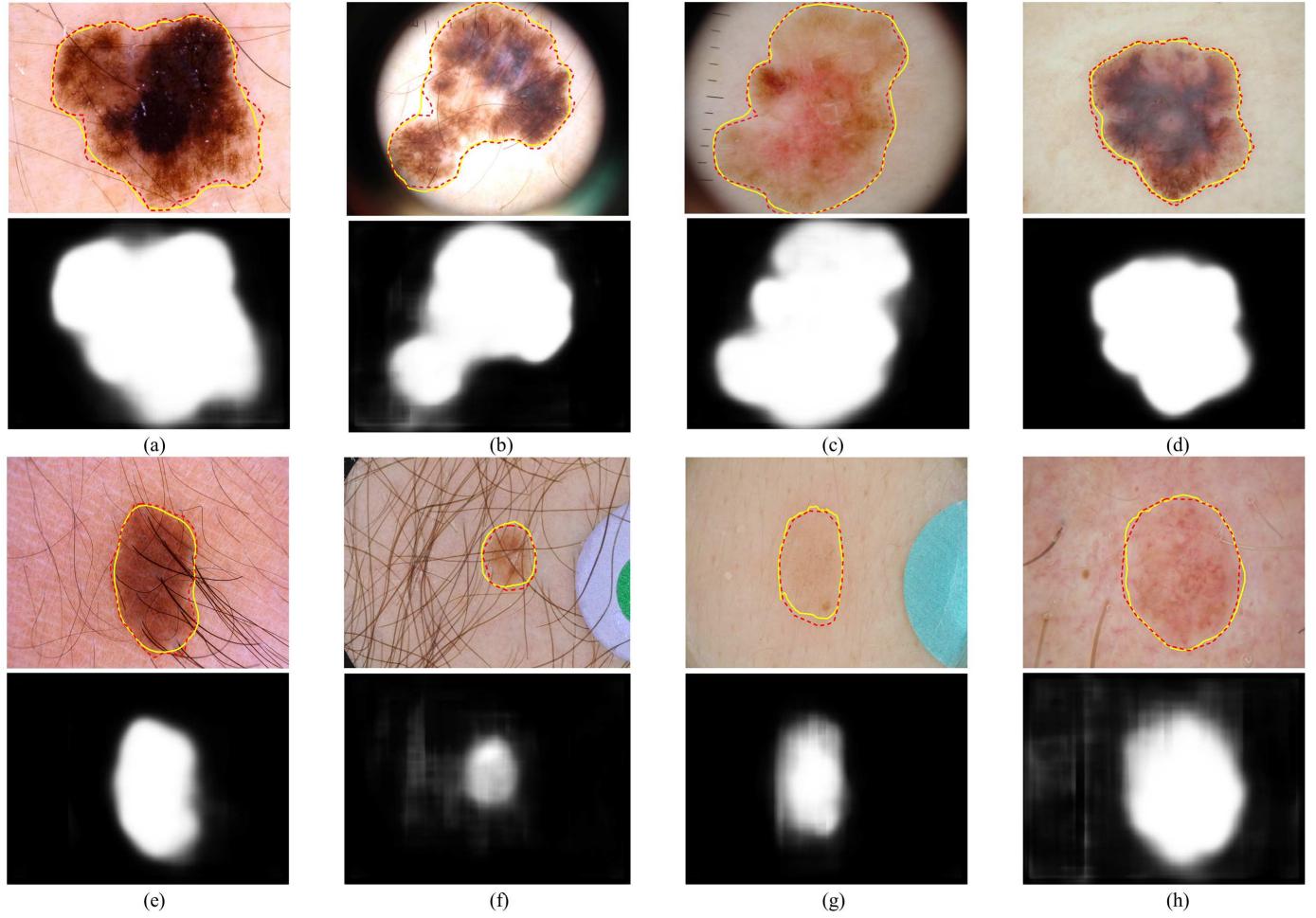


Fig. 4. Segmentation results for eight examples in the ISBI 2016 testing dataset, where (a) - (d) are melanoma and (e) - (h) are non-melanoma lesions. In each figure, the red dash line indicates the outline contoured by dermatologist and the yellow solid line is the result of automatic segmentation. The second and fourth rows are the corresponding probability maps generated by the proposed FCN model.

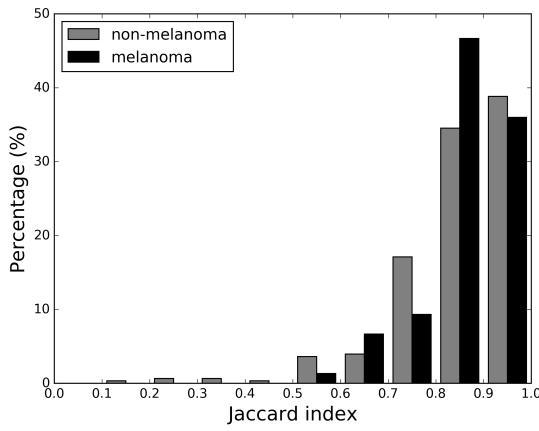


Fig. 5. Distribution of segmentation performance in terms of Jaccard index on ISBI 2016 testing dataset.

melanoma recognition on dermoscopic images. Despite the large body of work in the literature, automatic lesion segmentation remains a challenging task due to the large variations of appearance on dermoscopic images, the lack of

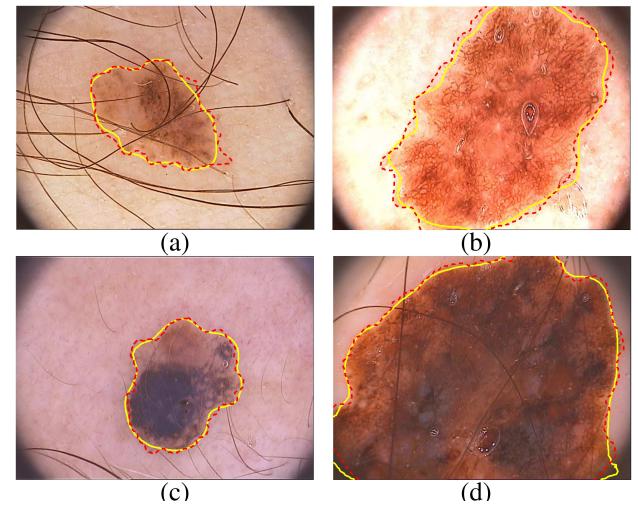


Fig. 6. Segmentation results of some challenging cases in the PH2 database, including cases affected by hair artifacts (a) and by illumination variation (b), a case with low contrast (c) and one whose boundary reaches the image edges (d). The top row includes two non-melanoma and the bottom row are two melanoma lesions. The red dash and yellow solid contours indicate the ground truth and the segmentation results of fine-tuned model, respectively.

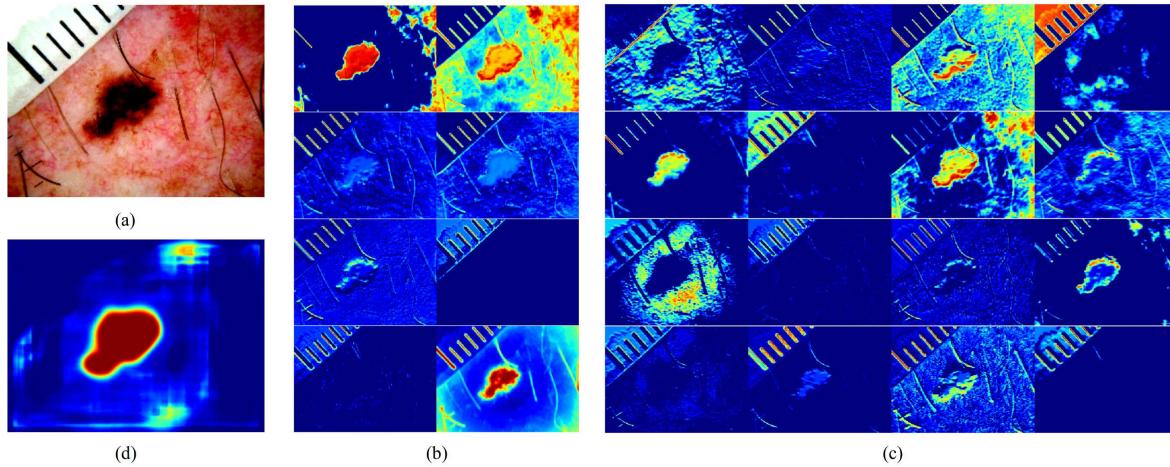


Fig. 7. Visualization of activations in our network. The image in (a) is the input to our model after contrast re-normalization, and (d) is the predicted probability map. (b) and (c) are the outputs of the first two convolutional layers where each activation map shows how the network responds to different image features.

distinct boundaries to adjacent skin, as well as the presence of various artifacts. In this work, we presented a fully deep convolutional network architecture for automatic skin lesion segmentation that was found to outperform the state-of-the-art on two publicly available databases. Efficiently training a deep network is a challenging task for skin lesion segmentation. When a network goes deeper, problems such as vanish gradient may become more prominent, which makes it more difficult to tune the parameters of the early layers [35]. Meanwhile, a deeper network usually requires more training samples to reduce overfitting, however, the dataset available in skin lesion segmentation is fairly limited as compared to natural images. This dilemma further exacerbates the difficulties in training the deep network. We exploited the following strategies to address these issues: 1) Instead of patch-wise supervised learning, we employed a fully convolutional model to predict the lesion mask in a pixel-wise fashion, which greatly boosted the number of equivalent training samples; 2) Batch normalization reduced the internal covariate shift and Adam optimization allowed us to use a fairly large learning rate to speed up the training procedure; 3) Dropout and image augmentation efficiently alleviated overfitting with limited training data, and 4) We used Jaccard distance as the loss function during network training, which is more related to the task of image segmentation as compared to cross entropy, yielding an improved segmentation performance. It should be noted that these training strategies are general enough to be extended to other medical image segmentation problems.

Instead of developing sophisticated pre- and post-processing techniques to account for various image acquisition conditions and to remove artifacts, we designed a fully trainable system with minimal pre- and post-processing to perform lesion segmentation in an end-to-end fashion. Figure 7 shows an example of how the trained network deals with these artifacts. Here (a) is the original image, (b) and (c) show the outputs of the first two convolutional layers where each activation map visualizes how the network responds to different image features. We can see that the lesion, skin hairs and the color

chart are activated in different maps, which greatly facilitates further processing in the following layers to predict a clear probability map as shown in (d). The strategy of relying on the trained network while simplifying pre- and post-processing also dramatically increased the speed of inference while maintaining the generalization of the proposed method. For a typical  $768 \times 1024$  image, the entire segmentation procedure only took about 0.14 second.

The generalization capability of our model was further validated by testing on the independent PH2 database. Without changing any parameters in the trained network, our method already achieved consistently promising performance on this new database. This is because our method only uses raw image data as input and allows the neural network to directly learn features from them. A well designed network tends to learn a more comprehensive and robust features as compared to hand-craft features that reflect the prior knowledge of the feature designers and their understanding of data. On the other hand, since the PH2 images are acquired at the Dermatology Service of Hospital Pedro Hispano in Portugal, the segmentation performance is inevitably affected by variabilities such as patient population, image acquisition conditions, as well as the dermatologists who contoured the lesions. By fine tuning the FCN model on the PH2 data, we were able to address these variations and further boost the segmentation performance. Transfer learning has been shown to be beneficial in medical image analysis tasks such as computer-aided detection (CADe) [46], we demonstrated in this work it is also an effective and efficient way to build a deep network for lesion segmentation with small dataset.

In order to achieve a good segmentation performance, a few parameters were experimentally determined using grid search during the 5-fold cross-validation with ISBI training dataset. For the proposed FCN model, the mainly tuned hyper-parameters included the number of layers, kernel size and the number of feature channels in each convolutional/deconvolutional layer. A set of hyper-parameters was selected if its model performed best on the validation dataset.

There is a trade-off between model complexity and its generalization ability, as well as the computational demand for model training. We found that increasing the capability of the FCN model by adding additional layers and doubling feature channels did not necessarily yield performance improvement while significantly increasing the time for model training. Another set of parameters included two thresholds that convert a probability map to a binary tumor mask. The selection criterion for the higher threshold was to make sure all the tumors could be correctly identified. Considering that a lot of datasets may not have labeled images, the operator normally needs to have dermatologists draw contours on several images then adjust the lower threshold to match manually-drawn contours. We mimicked this scenario when applying the trained FCN model to the PH2 database without retraining the model. It should be noted that we did not spend much effort to fine tune these two thresholds since it is not the focus of our study, but the segmentation performance might be improved by re-optimizing them for a new dataset.

Although our model has achieved promising segmentation accuracy on most of cases in two independent databases, there are cases where our model needs further improvement. Figure 8 (a) shows the most challenging case (ISIC\_0011338) among all the ISBI testing images where both our model and all the top three teams in the ISBI challenge performed poorly, while (b) shows a failed case that has very low contrast. Fig. 8 (c) and (d) are examples for lesion under- and over-segmentation. While the level of under/over segmentation can be controlled by adjusting the lower threshold on the probability map, it is not straightforward to automatically tune this parameter based on a specific image. One way to further improve the segmentation performance is to combine the proposed method with other post-processing techniques, such as levelsets [24]. Another option is to integrate Bayesian learning, such as conditional random field (CRF) [47], into our model to improve its discriminative capability. These interesting topics are worthy being investigated in the future study.

## VI. CONCLUSION

In this work, we proposed a fully automatic framework based on deep convolutional neural network for skin lesion segmentation on dermoscopic images. Several effective training strategies were implemented to tackle the challenges that training a deep network may face when only limited training data is available. We designed a novel loss function based on the Jaccard distance to further boost the segmentation performance. Compared to the conventionally used cross entropy, the Jaccard distance based loss function directly maximizes the overlap between the foreground of the ground truth and that of the predicted segmentation mask, and thus eliminates the needs of data re-balancing when the numbers of foreground and background pixels are highly unbalanced, such as binary medical image segmentation. Our approach outperformed the state-of-the-art methods when evaluating on an open challenge database of *Skin Lesion Analysis Towards Melanoma Detection* in ISBI 2016, as well as on another publicly available database of PH2. Our results also clearly demonstrated that

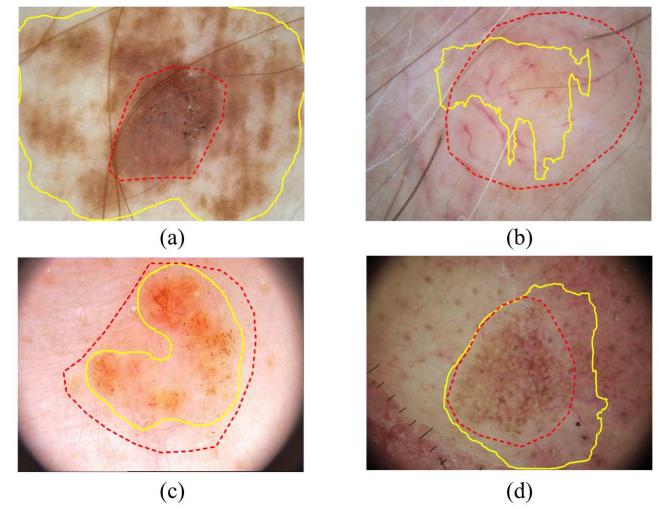


Fig. 8. Cases where our results were suboptimal. The red dash and yellow solid contours indicate the ground truth and the segmentation result, respectively. (a) shows the most challenging case in the ISBI 2016 testing dataset where both our method and all the top three teams in the ISBI 2016 challenge performed poorly. (b) is a failed case that has very low contrast. (c) and (d) are examples for lesion under- and over-segmentation, respectively.

the proposed method is robust to various image artifacts and imaging acquisition conditions while using minimum pre- and post-processing. We believe this method can generalize well to other medical image segmentation tasks.

## VII. ACKNOWLEDGMENT

The authors are grateful to International Skin Imaging Collaboration (ISIC), the organizers of 2016 International Symposium on Biomedical Imaging (ISBI 2016) challenge of “Skin lesion analysis towards melanoma detection”, and the Dermatology Service of Hospital Pedro Hispano in Portugal, who make the dermoscopic image data publicly available. They also thank the anonymous reviewers whose comments and suggestions helped improve this manuscript.

## REFERENCES

- [1] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2016,” *CA Cancer J Clin.*, vol. 66, pp. 7–30, 2016.
- [2] C. A. Barker and M. A. Postow, “Combinations of radiation therapy and immunotherapy for melanoma: a review of clinical outcomes,” *Int J Radiation Oncol Biol Phys*, vol. 88, no. 5, pp. 986–997, 2014.
- [3] C. M. Balch, J. E. Gershenwald, and S.-J. S. et al., “Final version of 2009 AJCC melanoma staging and classification,” *J Clin Oncology*, vol. 27, no. 36, pp. 6199–6206, 2009.
- [4] M. Vestergaard, P. Macaskill, P. Holt, and S. Menzies, “Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting,” *Br J Dermatol.*, vol. 159, no. 3, pp. 669–676, 2008.
- [5] K. Korotkov and R. Garcia, “Computerized analysis of pigmented skin lesions: a review,” *Artif. intell. Med.*, vol. 56, no. 2, pp. 69–90, 2012.
- [6] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler, “Automated melanoma recognition,” *IEEE Trans. Med. Imag.*, vol. 20, no. 3, pp. 233–239, 2001.
- [7] M. E. Celebi, H. A. Kingravi, B. Uddin, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, and R. H. Moss, “A methodological approach to the classification of dermoscopy images,” *Comput. Med. Imag. Graph.*, vol. 31, no. 6, pp. 362–373, 2007.
- [8] M. Silveira, J. C. Nascimento, J. S. Marques, A. R. Marçal, T. Mendonça, S. Yamauchi, J. Maeda, and J. Rozeira, “Comparison of segmentation methods for melanoma diagnosis in dermoscopy images,” *IEEE J. Sel. Top. Signal Process.*, vol. 3, no. 1, pp. 35–45, 2009.

- [9] M. E. Celebi, H. Iyatomi, G. Schaefer, and W. V. Stoecker, "Lesion border detection in dermoscopy images," *Comput. Med. Imag. Graph.*, vol. 33, no. 2, pp. 148–153, 2009.
- [10] M. Emre Celebi, Q. Wen, S. Hwang, H. Iyatomi, and G. Schaefer, "Lesion border detection in dermoscopy images using ensembles of thresholding methods," *Skin Res. Technol.*, vol. 19, no. 1, pp. e252–e258, 2013.
- [11] F. Peruch, F. Bogo, M. Bonazza, V.-M. Cappelleri, and E. Peserico, "Simpler, faster, more accurate melanocytic lesion segmentation through meds," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 557–565, 2014.
- [12] H. Zhou, X. Li, G. Schaefer, M. E. Celebi, and P. Miller, "Mean shift based gradient vector flow for image segmentation," *Comput. Vis. Image Understand.*, vol. 117, no. 9, pp. 1004–1016, 2013.
- [13] A. R. Sadri, M. Zekri, S. Sadri, N. Gheissari, M. Mokhtari, and F. Kolahdouzan, "Segmentation of dermoscopy images using wavelet networks," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 4, pp. 1134–1141, 2013.
- [14] F. Xie and A. C. Bovik, "Automatic segmentation of dermoscopy images using self-generating neural networks seeded by genetic algorithm," *Pattern Recognition*, vol. 46, no. 3, pp. 1012–1019, 2013.
- [15] M. E. Celebi, Q. Wen, H. Iyatomi, K. Shimizu, H. Zhou, and G. Schaefer, "A state-of-the-art survey on lesion border detection in dermoscopy images," pp. 97–129, 2015.
- [16] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Adv. Neural Inf. Process. Sys.*, 2012, pp. 1097–1105.
- [18] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Proc. MICCAI 2013*. Springer, 2013, pp. 411–418.
- [19] Z. Yan, Y. Zhan, Z. Peng, S. Liao, Y. Shinagawa, S. Zhang, D. N. Metaxas, and X. S. Zhou, "Multi-instance deep learning: Discover discriminative local anatomies for bodypart recognition," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1332–1343, 2016.
- [20] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V. C. Mok, L. Shi, and P.-A. Heng, "Automatic Detection of Cerebral Microbleeds From MR Images via 3D Convolutional Neural Networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1182–1195, 2016.
- [21] S. Miao, Z. J. Wang, and R. Liao, "A CNN Regression Approach for Real-Time 2D/3D Registration," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1352–1363, 2016.
- [22] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [23] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1240–1251, 2016.
- [24] K. H. Cha, L. Hadjiiski, R. K. Samala, H.-P. Chan, E. M. Caoili, and R. H. Cohan, "Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets," *Med. Phys.*, vol. 43, no. 4, pp. 1882–1896, 2016.
- [25] M. Avendi, A. Kheradvar, and H. Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI," *Med. Image Anal.*, vol. 30, pp. 108–119, 2016.
- [26] M. H. Jafari, E. Nasr-Esfahani, N. Karimi, S. Soroushmehr, S. Samavi, and K. Najarian, "Extraction of skin lesions from non-dermoscopic images using deep learning," *arXiv preprint arXiv:1609.02374*, 2016.
- [27] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. CVPR*, 2015, pp. 3431–3440.
- [28] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI 2015*. Springer, 2015, pp. 234–241.
- [29] D. Gutman, N. C. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, and A. Halpern, "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)," *arXiv preprint arXiv:1605.01397*, 2016.
- [30] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. Marcal, and J. Rozeira, "PH 2-A dermoscopic image database for research and benchmarking," in *Proc. EMBC 2013*. IEEE, 2013, pp. 5437–5440.
- [31] N. Codella, Q.-B. Nguyen, S. Pankanti, D. Gutman, B. Helba, A. Halpern, and J. R. Smith, "Deep learning ensembles for melanoma recognition in dermoscopy images," *arXiv preprint arXiv:1610.04662*, 2016.
- [32] L. Yu, H. Chen, Q. Dou, J. Qin, and P. A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Trans. Med. Imag.*, vol. 36, no. 4, pp. 994–1004, 2017.
- [33] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. ICCV 2015*, 2015, pp. 1520–1528.
- [34] P. Jaccard, "The distribution of the flora in the alpine zone." *New phytologist*, vol. 11, no. 2, pp. 37–50, 1912.
- [35] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Aistats*, vol. 9, 2010, pp. 249–256.
- [36] L. Bottou, "Stochastic gradient descent tricks," in *Neural Networks: Tricks of the Trade*. Springer, 2012, pp. 421–436.
- [37] I. Sutskever, J. Martens, G. E. Dahl, and G. E. Hinton, "On the importance of initialization and momentum in deep learning," *Proc. ICML 2013*, vol. 28, pp. 1139–1147, 2013.
- [38] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [39] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.
- [40] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [41] T. T. D. Team, R. Al-Rfou, G. Alain, A. Almahairi, C. Angermueller, D. Bahdanau, N. Ballas, F. Bastien, J. Bayer, A. Belikov *et al.*, "Theano: A python framework for fast computation of mathematical expressions," *arXiv preprint arXiv:1605.02688*, 2016.
- [42] Z. Ma and J. M. R. Tavares, "A novel approach to segment skin lesions in dermoscopic images based on a deformable model," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 2, pp. 615–623, 2016.
- [43] L. Bi, J. Kim, E. Ahn, D. Feng, and M. Fulham, "Automated skin lesion segmentation via image-wise supervised learning and multi-scale superpixel based cellular automata," in *Proc. ISBI 2016*. IEEE, 2016, pp. 1059–1062.
- [44] E. Ahn, L. Bi, Y. H. Jung, J. Kim, C. Li, M. Fulham, and D. D. Feng, "Automated saliency-based lesion segmentation in dermoscopic images," in *Proc. EMBC 2015*. IEEE, 2015, pp. 3009–3012.
- [45] E. Ahn, J. Kim, L. Bi, A. Kumar, C. Li, M. Fulham, and D. Feng, "Saliency-based lesion segmentation via background detection in dermoscopic images," *IEEE J. Biomed. Health Inform.*, vol. PP, no. 99, 2017.
- [46] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [47] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," *arXiv preprint arXiv:1412.7062*, 2014.