# Step-wise Integration of Deep Class-specific Learning for Dermoscopic Image Segmentation

Lei Bi<sup>a</sup>, Jinman Kim<sup>a,\*</sup>, Euijoon Ahn<sup>a</sup>, Ashnil Kumar<sup>a</sup>, Dagan Feng<sup>a,c</sup> and Michael Fulham<sup>a,b,d</sup>

<sup>a</sup> School of Information Technologies, University of Sydney, NSW, Australia

<sup>b</sup> Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, NSW, Australia

<sup>c</sup> Med-X Research Institute, Shanghai Jiao Tong University, Shanghai, China

<sup>d</sup> Sydney Medical School, University of Sydney, NSW, Australia

\*Corresponding author: jinman.kim@sydney.edu.au

*Abstract*— The segmentation of abnormal regions on dermoscopic images is an important step for automated computer aided diagnosis (CAD) of skin lesions. Recent methods based on fully convolutional networks (FCN) have been very successful for dermoscopic image segmentation. However, they tend to overfit to the visual characteristics that are present in the dominant non-melanoma studies and therefore, perform poorly on the complex visual characteristics exhibited by melanoma studies, which usually consists of fuzzy boundaries and heterogeneous textures. In this paper, we propose a new method for automated skin lesion segmentation that overcomes these limitations via a novel deep class-specific learning approach which learns the important visual characteristics of the skin lesions of each individual class (melanoma vs non-melanoma) on an individual basis. We also introduce a new probability-based, step-wise integration to combine complementary segmentation results derived from individual class-specific learning models. We achieved an average Dice coefficient of 85.66% on the ISBI 2017 Skin Lesion Challenge (SLC), 91.77% on the ISBI 2016 SLC and 92.10% on the PH2 datasets with corresponding Jaccard indices of 77.73%, 85.92% and 85.90%, respectively, for the same datasets. Our

experiments on three well-established public benchmark datasets demonstrate that our method is more effective than other state-of-the-art methods for skin lesion segmentation.

Index Terms— Dermoscopic, Melanoma, Segmentation, Fully Convolutional Networks (FCN)

# I. INTRODUCTION

Melanoma (also known as malignant melanoma) has one of the most rapidly increasing incidences in the world with considerable mortality [1, 2]. Early diagnosis is particularly important because melanoma can be cured with prompt excision [3, 4]. Dermoscopy is a non-invasive imaging technique used for the in vivo evaluation of pigmented skin lesions and has an important role in the early diagnosis of melanoma [3]. Dermoscopy uses optical magnification, liquid immersion, and low angle-of-incidence lighting or cross-polarized lighting to make the contact area translucent, thus increasing the visibility of subsurface structures when compared to conventional clinical images. The identification of melanoma using human vision alone, is subjective, not reproducible and can be inaccurate, even among experienced dermatologists [5, 6]. These limitations can be attributed to the skin lesions having various sizes and shapes, fuzzy boundaries, low contrast when compared to the skin, heterogeneity and skin hair (Figure 1) [7]. These problems have motivated the development of computer aided diagnosis (CAD) systems that can assist dermatologists' clinical diagnoses [8, 9]. Skin lesion segmentation is a fundamental requirement for these CAD systems. Manual or semi-automatic segmentation methods, such as through seed selection, are usually subjective, time-consuming, not reproducible and as a consequence, unreliable for a CAD system. Not surprisingly, research efforts have been directed at the development of fully automated segmentation methods.



Figure 1. Examples of skin lesions with fuzzy boundaries (a, b, c, d), hair (c), inhomogeneity (a, c, d) and low-contrast (b).

#### A. Related Work

A number of fully automated skin lesion segmentation methods have been recently proposed and they can be separated into 3 main groups: (1) unsupervised methods that do not use training data; (2) traditional supervised methods that use trained classifiers; and (3) deep learning based methods that segment using a trained deep learning model. For a more detailed discussion of the field, the reader is referred to the comprehensive skin lesion segmentation survey papers by Celebi et al. [10, 11].

Unsupervised segmentation methods mainly focus on thresholding [12-14], energy functions [15-17] and iterative/statistical region-merging [18, 19]. Thresholding methods attempt to separate skin lesions based on a threshold value, which is generally calculated by analyzing pre-defined image features, e.g., intensity histogram. Energy functions methods identify the lesion boundaries by minimizing a well-defined cost (energy) function defined by image characteristics such as edges, smoothness, and statistical distributions. Iterative/statistical region-merging methods recursively merge pixels or regions in a hierarchical manner. More recently, saliency [20, 21], Delaunay triangulation [22], sparse coding with dynamic rule based refinement (SCDRR) [23] and multi-scale superpixels with cellular automata (MSCA) [24] have also been applied to skin lesion segmentation. Unsupervised methods, however, have a limited capacity to accurately segment challenging skin lesions where the lesions touch the image boundary and where there are nearby

artifacts (Figures 1b, 1d). Thresholding based methods are further limited by the inhomogeneous intensity distribution of the lesion and may fail if the distribution contains multiple peaks (Figure 1a).

Traditional supervised segmentation methods mainly focus on extracting pixel or region features such as pixel-level Gaussian [25], RGB color [26, 27] and texture features [28] and then use various classifiers, such as the Bayes classifier [25], wavelet network [26] or support vector machines [28], to separate the skin lesions from surrounding healthy skin. All of these methods, however, rely on low-level features, such as color and texture, which do not capture high-level image-wide semantic information. Consequently, these methods cannot provide accurate segmentation results for the difficult lesions. In addition, their performance depends heavily on correctly tuning a large number of parameters and effective pre-processing techniques such as hair removal and illumination correction, which limit their generalizability.

Current deep learning based methods mainly use fully convolutional networks (FCN) [29, 30]. The success of FCN is primarily attributed to their ability to leverage large datasets to derive a feature representation that combines low-level appearance information with high-level semantic information [30]. FCN can also be trained in an end-to-end manner for efficient inference, i.e., images are taken as inputs and the segmentation results are outputted directly. Yuan et al. [31] replaced the cross-entropy loss used in traditional FCN with a Jaccard distance loss for skin lesion segmentation. Yu et al. [32] used a 50-layer deep residual network, where the residual blocks proposed by He et al. [33] were used to increase the overall depth of the networks (number of layers) and enable segmentation based on more meaningful image features. In our prior work [34], we presented a multi-stage segmentation approach to address the segmentation challenges due to the variability in visual appearance within a skin lesion; multiple FCNs (16-layer VGGNet [35]) learned complementary visual features, with early-stage FCNs learned coarse appearance and localization information while late-stage FCNs learned the subtle characteristics of the skin lesion boundaries. Experimental results showed that our method achieved state-of-the-art segmentation results on two public datasets (2016 ISBI Skin Lesion Challenge dataset [36] and PH2 dataset [37]). These FCN-based methods, however, had inconsistent performance (segmentation results) across different types of skin lesions; they were usually overfitted to the dominant non-melanoma studies with a poorer performance on melanoma. In addition, the trained model was also sub-optimal for non-melanoma studies because the melanoma studies usually had fuzzy boundaries and inhomogeneous textures, which also influenced the learned features for

non-melanoma studies. Hence, these segmentation methods were sub-optimal for both non-melanoma and melanoma studies.

#### B. Our Contribution

We propose a new fully automated skin lesion segmentation method for dermoscopic images to overcome the limitations mentioned above. We refer to our method as deep class-specific learning with probability based step-wise integration (DCL-PSI). Our experiments show that our method improves on the state-of-theart FCN based segmentation methods and is specifically optimized for skin lesion segmentation. In contrast to the imbalanced segmentation results of FCN and its variants, our proposed method refines FCN segmentation across different classes (melanoma and non-melanoma) and then iteratively integrates the refined segmentation outputs based on the image classification results. Our method adds the following contributions to the state-of-the-art:

- We propose a deep class-specific learning (DCL) approach for accurate dermoscopic image segmentation. We trained our ResNet [33] based FCN model independently across different known classes. Compared with traditional FCN based methods, which only use a single-class (group) and a shallow architecture (e.g., VGGNet [35]) for training, our method greatly improves the segmentation performance with high balanced segmentation results across melanoma and non-melanoma studies.
- We propose a probability based step-wise integration (PSI) approach for segmentation refinement. Based on the image classification probability, we integrated the complementary segmentation results produced by different DCL models to iteratively maximize label agreement between neighboring pixels. Consequently, we ensured that the appearance of segmented skin lesion was spatially consistent and have better segmentation results of the skin lesion boundaries.
- Compared with all the existing methods, our method is particularly effective for segmenting challenging melanoma studies, which usually have fuzzy boundaries and heterogeneous textures.

The proposed method is also significantly different from our previous study [34]. In particular, (1) we introduce a deep classification architecture that guides the FCN-based segmentation process, which

influences the segmentation based upon whether the image is classified as a melanoma or non-melanoma study; (2) we use a deeper (101-layer vs 16-layer) neural network within our segmentation framework as a means of learning more meaningful features for the most complex skin lesion cases; and (3) we propose a novel PSI approach to iteratively integrate and refine segmentation results by maximizing the consistency between neighboring pixels. In addition, we also performed a more thorough evaluation, through experiments on a more challenging public image dataset (2017 ISBI Skin Lesion Challenge dataset). Our new method achieves better segmentation accuracy than our prior work, especially in the most complex cases for which there are relatively fewer training samples.

# II. METHODS



## A. Overview of the Proposed Method

Figure 2. Flow diagram of our proposed deep class-specific learning with probability based step-wise integration (DCL-PSI) method.

The flow diagram of our proposed segmentation method is shown in Figure 2. During the training stage, training dermoscopic images together with annotations were used to train 3 class-specific residual based fully convolutional networks (ResFCN) one-by-one (N-Model, M-Model and O-Model). An additional residual based convolutional neural networks (ResCNN) was also trained for separating melanoma and non-melanoma dermoscopic images. During the testing stage, we applied deep class-specific learning (DCL) to the input dermoscopic image to obtain class-specific segmentation maps ( $Y_N$ ,  $Y_M$ ,  $Y_O$ ) of the skin lesion. Then the input dermoscopic image was fed into the ResCNN to estimate a probability  $\omega$ , derived from the classification score corresponding to if the input dermoscopic image depicted melanoma or non-melanoma. Finally, this probability  $\omega$  was used within the probability based step-wise integration to iteratively combine the segmentation results generated by different class-specific learning models ( $Y_N$ ,  $Y_M$ ,  $Y_O$ ) to produce the final segmentation results.

# B. Materials

We used three well-established public benchmark datasets to train and test the effectiveness of our method.

- The 2017 and 2016 ISBI Skin Lesion Challenge (denoted as ISBI 2017 [38] and ISBI 2016 [36]) datasets are a subset of the large International Skin Imaging Collaboration (ISIC) archive, which contains dermoscopic images acquired on a variety of different devices at numerous leading international clinical centers. The ISBI 2017 challenge 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 ISBI 2016 challenge dataset provides 900 training images (727 non-melanoma and 173 melanoma) and a separate test dataset of 379 images (304 non-melanoma and 75 melanoma). Image size varies from 566×679 pixels to 2848×4288 pixels. Detailed distribution of these two datasets are listed in Figure 3.
- The PH2 public dataset [37] was collected by the Universidade do Porto, Técnico Lisboa, and the Dermatology service of Hospital Pedro Hispano in Matosinhos, Portugal. All 200 dermoscopic images (160 non-melanoma and 40 melanoma) were obtained under the same conditions using a Tuebinger Mole Analyzer system using a 20-fold magnification. Image size varies from 553×763 pixels to 577×769

pixels.



All datasets provided ground truth segmentations based on manual delineations by clinical experts.

Figure 3. Image size distribution of (a) ISBI 2017, (b) ISBI 2016 and (c) PH2 datasets, where the color bar indicates the number of images with a particular width and height combination.

#### C. Deep Class-specific Learning via ResNet based Fully Convolutional Networks

#### 1) Fully Convolutional Networks

The traditional FCN architecture was converted from convolutional neural networks (CNNs) for efficient dense inference [29] and contains downsampling and upsampling parts. The downsampling part has stacked convolutional layers to extract high-level semantic information and has been routinely used in CNNs for image classification tasks [39]. The upsampling part has stacked deconvolutional layers, which are transposed convolutional layers that upsample the feature maps derived from the downsampling part to output the score masks (segmentation maps) [40].

Convolutional layers are defined on a translation invariance basis and have shared weights across different spatial locations. The input and the output of convolutional layers are feature maps and are calculated by convolving convolutional kernels:

$$\boldsymbol{f} = \boldsymbol{W} *_{\boldsymbol{S}} \boldsymbol{X} + \boldsymbol{b} \tag{1}$$

where W denotes the kernel,  $*_s$  represents convolution operation with stride s, X and b are the input feature map and the bias, respectively. As a result, the resolution of the output feature map f is downsampled by a factor of s. Convolutional layers are usually interleaved with max-pooling layers and ReLU layers. Max-

pooling layers are form of non-linear downsampling. It partitions the input feature map into a set of nonoverlapping sliding windows and then calculates and outputs the maximum value of each sliding window [29]. Consequently, it allows to further improve translation invariance and representation capability [41]. ReLU layers were introduced to improve the non-linear properties of the network and is calculated based on the non-saturating activation function [42]. The ReLU layer allows efficient computation and minimizes the risk of having gradient exploding or vanishing problems [42, 43].

#### 2) Downsampling Part via ResNet

Traditional FCN architecture is based on the VGGNet architecture [35] and its downsampling part usually has limited capacity to add additional layers. Experimental data has shown that beyond certain depths, adding extra layers results in higher training errors and therefore, it is challenging to optimize very deep networks with many layers [44]. Deep residual networks (ResNet) have overcome these limitations and achieved state-of-the-art results in image classification and object detection [33] tasks. The ResNet architecture consists of a number of residual blocks with a skip connection (illustrated in Figure 4) that bypasses (shortcuts) a few convolutional (Conv), batch normalization (BN) and ReLU layers at a time. The ResNet architecture, therefore, is able to overcome the limit of the number of layers by adding shortcut connections that are aggregated with the output of the convolutional layers. ResNets can also be considered as an ensemble of many shallower networks [44], with different networks connected via these shortcuts, and thus allow optimal results by averaging the output of the different networks. To this end, a 101-layer ResNet consist of 33 residual blocks was adopted in the downsampling part for robust visual feature learning and representations.



Figure 4. A sample residual block used in our method.

## 3) Upsampling Part via RefineNet Module

In the upsampling part, we adopted the RefineNet module [45] to upsample the feature maps and to produce a segmentation. Our design utilizes multiple RefineNet modules to refine high-level features by incorporating low-level features in an iterative manner. The RefineNet module (Figure 5) comprises multiple residual convolution units (RCU), a multi-resolution fusion (MRF) and a chained residual pooling (CRP) unit. The RCU is an adaptive convolution set that fine-tunes the pre-trained ResNet weights for skin lesion segmentation. The MRF unit fuses the multi-path input into a higher-resolution feature map. The CRP unit consists of multiple max-pooling (Pool) layers at different scales, which were fused iteratively. Therefore, the CRP unit allows the extraction and encoding of contextual information from different sized regions.



Figure 5. Details of the RefineNet module and the building units.

4) Deep Class-specific Learning (DCL)



Figure 6. Overview of the ResNet based FCN (ResFCN).

After modifying the upsampling and downsampling parts, the overview of the ResNet based FCN (ResFCN) is shown in Figure 6 and can be defined as:

$$Y = U_{S}(F_{S}(I;\theta);\varphi)$$
(2)

where Y is the output prediction, I is the input image,  $F_S$  represents the stacked residual blocks with a list of strides, and  $U_S$  denotes the RefineNet modules that upsample the feature map by a list of strides S and maps the output Y to have the same scale with the input I (that is, the same height and width). The parameters  $\theta$  and  $\varphi$  are the learned weights for the downsampling and upsampling processes.

For skin lesion segmentation, the ResFCN network can be trained end-to-end by minimizing the overall loss between the predicted skin lesions area and the ground truth annotation of the training data:

$$\underset{\boldsymbol{\theta},\boldsymbol{\varphi}}{\operatorname{argmin}} \sum \mathcal{L}(\boldsymbol{Y},\boldsymbol{Z}|\boldsymbol{\theta},\boldsymbol{\varphi}) \tag{3}$$

where  $\mathcal{L}$  measures the cross-entropy loss of the expert-provided annotation  $\mathcal{Z}$  and the predicted results Y. The ResFCN network parameters ( $\boldsymbol{\theta}$  and  $\boldsymbol{\varphi}$ ) can then be updated iteratively using the stochastic gradient descent (SGD) [46] algorithm. To reduce the overfitting to non-melanoma (denoted as N) images while improving the learned features for melanoma images (denoted as M), we first trained one ResFCN model using non-melanoma and melanoma images for 100 epochs with a batch size of 1 at a learning rate of 0.0005 (denoted as the O-Model), ensuring that the O-Model converged (the loss no longer decreases) [47]. Then, we trained two other ResFCN models separately: one for melanoma (denoted as M-Model) and one for non-melanoma (denoted as N-Model) images; both used the same training batch size and learning rate as the O-Model. Our architecture was implemented using the MatConvNet [48] library.

There is a scarcity of medical images with annotations for use as training data, due to the cost and complexity of the acquisition procedures [47]. In contrast, there are many more data available in the field of general images [40, 49]. Existing research shows that the problem of insufficient training data can be alleviated by fine-tuning, where the lower layers of the fine-tuned network are more general filters (trained on general images) while those in the higher layers are more specific to the target problem [47, 50]. Therefore, we used the pre-trained 101-layer-ResNet (trained on ImageNet [49]) network for initialization. Since there is a small amount of images within each class (e.g., melanoma images) compared to the full dataset, we trained the M-Model and N-Model using the O-Model as the initialization. In addition, there is a marked variation in the size of dermoscopic image (Figure 3) so we pre-processed and resized (nearest-neighbor interpolation) all the training dermoscopic images to be at a maximum of 1,000 pixels in the longest axis while preserving the aspect ratios. We used data augmentation techniques, including random cropping and flipping , to improve robustness of training [50, 51]. On average, each model took around 48 hours to train on a 12GB Nvidia Maxwell Titan X GPU.

## D. Probability based Step-wise Integration



Figure 7. A sample of neighborhood pixels used in our probability based step-wise integration.

We first used a step-wise refinement approach on the results produced by the DCL to further enhance the segmentation results, e.g., the contour of the lesions. We optimized individual DCL model produced segmentation maps by exploiting local similarity given by neighborhood pixels and global similarity given by different model produced results. More specifically, for each step, individual pixels of the segmentation map propagate according to their local and global agreement (the neighbors' current states), and constrain the boundary of the segmentation map. Consequently, the spatially corresponding pixels of the segmentation maps have equal influence in determining the pixels' segmentation value in the following step. The spatially corresponding pixels are defined as the neighborhood pixels of the one DCL result across all DCL models (see Figure 7). Therefore, the optimal results are achieved when the segmentation map converges into a stable state. We used the hierarchical evolving model proposed by Qin et al. [52] to achieve this, defined as:

$$\boldsymbol{Y}_{g}^{(i+1)} = \boldsymbol{Y}_{g}^{(i)} + \boldsymbol{\gamma} \cdot \boldsymbol{H}_{g}^{(i)} \tag{4}$$

where *i* represents number of steps;  $Y_g$  represents output from the DCL ( $g \in T = \{O, M, N\}$ );  $\gamma$  is a constant value that balances the importance of the foreground (skin lesion) and background, and was empirically set to 0.05 to encourage the pixel to follow the neighborhood pixels.  $H_g^{(i)}$  is defined as:

$$\boldsymbol{H}_{g}^{(i)} = \sum_{z}^{T} \begin{pmatrix} \boldsymbol{\emptyset}_{z}^{(i)}(r,c) & \cdots & \boldsymbol{\emptyset}_{z}^{(i)}(\cdot) \\ \vdots & \ddots & \vdots \\ \boldsymbol{\emptyset}_{z}^{(i)}(\cdot) & \cdots & \boldsymbol{\emptyset}_{z}^{(i)}(\cdot) \end{pmatrix}$$
(5)

$$\phi_{z}^{(i)}(r,c) = \sum_{\delta} sign(p_{z}^{(i)}(r,c) - \mu_{z})$$
(6)

where  $\begin{pmatrix} & \dots & \\ \vdots & \ddots & \vdots \\ & \dots & \end{pmatrix}$  represents the segmentation map,  $p_z(r, c)$  is the segmentation value (the probability  $\dots$ 

that the pixel at location (r, c) in the z-th segmentation map belongs to the skin lesion),  $\delta$  is the corresponding neighborhood system and  $sign(\cdot)$  is a signum function. The parameter  $\mu_z$  represents the adaptive threshold level (Otsu threshold level [53]) for the z-th segmentation map; it was computed from the initial segmentation map (i = 0) and does not change over time. The total number of steps *i* was set to 3.

After all the segmentation maps evolved into a stable state, we used a probability based integration approach to produce the final segmentation result. In general, the DCL results were complementary to each other; the deep trained M-model was intended to be responsible for the melanoma studies while the N-model was intended to be responsible for the rest of the studies. Therefore, we trained another ResNet based convolutional neural networks (ResCNN) architecture [33, 54] for melanoma classification purpose. We modified the last layer of 101-layer ResCNN to have 2 neurons to match with the non-melanoma and melanoma classes. All the training images were downscaled to 224 pixels (shorter axis and nearest-neighbor interpolation). During training, we fine-tuned the ResCNN model, pre-trained on ImageNet dataset for 150 epochs using a linear schedule learning rate (the learning rate was updated for every 50 batches. For each update, the new learning rate will be calculated as  $base_{lr} \times \left(1 - \frac{current_u}{total_u}\right)$ , where  $base_{lr} = 0.01$  is the base learning rate,  $total_u = 6600$  is the total number of updates and  $current_u$  is the current updating number). The same data augmentation as mentioned above including random crops and flips were used to improve the robustness of the model. It took around 1 day to train on a 12GB Nvidia Maxwell Titan X GPU with a batch size of 45. During testing (segmentation of unseen data), the input dermoscopic image was fed into the ResCNN to estimate the probability that the skin lesion depicted was melanoma ( $\omega_M$ ) and the probability that the skin lesion depicted was non-melanoma ( $\omega_N$ ). These probabilities were derived from the outputs of the softmax layer of the ResCNN.

The final integrated segmentation map  $\rho$  after *i* number of steps was calculated as:

$$\boldsymbol{\rho} = \omega_N \cdot \boldsymbol{Y}_N^{(i)} + \omega_M \cdot \boldsymbol{Y}_M^{(i)} + \boldsymbol{Y}_O^{(i)}.$$
(7)

## E. Post-segmentation Refinement

The final integrated segmentation map  $\rho$  was normalized into the range [0,1] and converted into a binary segmentation result via thresholding at 0.5. We followed the process described by Celebi et al. [10, 11]: a morphological dilation process (disk size of 1) was used to smooth the boundary and fill holes while connected thresholding was used to remove isolated single pixels (regions that were 1 pixel in size). Finally, we resized the binary segmentation result to the size of the original image using nearest-neighbor interpolation.

# **III.** EXPERIMENTS AND RESULTS

#### A. Experiment Setup

We performed the following experiments on the three datasets: (a) comparison of the overall performance of our method with several baseline and state-of-the-art methods; and (b) analysis of the performance of each component in our proposed method. For experiments using the ISBI 2017 and ISBI 2016 datasets, we used the specified training and test dataset. For experiments on the PH2 dataset, we followed the protocol used by Yuan et al. [31]: training on the ISBI 2016 training dataset and tested on the PH2 dataset.

The baseline skin lesion segmentation methods include: (1) SSLS [20, 21] – saliency based skin lesion segmentation; (2) MSCA [24] – multi-scale superpixel based cellular automata; (3) FCN [29] – fully convolutional networks (FCN-8s, at stride size 8). The state-of-the-art lesion segmentation methods were the top 5 results out of 21 teams for the ISBI 2017 challenge [38], the top 5 out of 28 teams for the ISBI 2016 challenge [36], and the following specific algorithms: (1) DT [22] – skin lesion segmentation using Delaunay triangulation; (2) SCDRR [23] – sparse coding with dynamic rule-based refinement; (3) JCLMM [55] – skin lesion segmentation by joining circular-linear distributions with mixture models; (4) M-FCN [34] – dermoscopic image segmentation via multi-stage fully convolutional networks; and (5) J-FCN [31] – a Jaccard distance based fully convolutional network for skin lesion segmentation.

#### B. Evaluation Metrics

The standard skin lesion segmentation evaluation metrics were used for comparison including: dice similarity coefficient (DSC), Jaccard index (Jac.), sensitivity (Sen.), specificity (Spec.) and accuracy (Acc.). They are defined as:

$$DSC = \frac{2|GT \cap AP|}{|GT| + |AP|} \tag{8}$$

$$Jac. = \frac{|GT \cap AP|}{|GT \cup AP|} \tag{9}$$

$$Sen. = \frac{|TP|}{|TP|+|FN|} \tag{10}$$

$$Spec. = \frac{|TN|}{|TN| + |FP|} \tag{11}$$

$$Acc. = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|} \quad (12)$$

where AP is the algorithm predicted segmentation result, GT denotes the ground truth, TP is the true positive pixels (skin lesions), TN is the true negative pixels (background), FN is the false negative pixels and FP is the false positive pixels. In addition, we calculated the precision-recall (PR) curves for additional comparisons, which have been widely used for object detection and segmentation problems on general images [56].

# C. Results on ISBI 2017 Dataset

Figure 8 and Table 1 to Table 3 show that our DCL-PSI method achieved the best overall performance across all the measurement and improved upon the existing methods with a large margin for both non-melanoma and melanoma studies (>1% increase in Jaccard measure) when compared with the best performing techniques in the ISBI 2017 skin lesion challenge.

 Table 1. Comparison of segmentation results on the ISBI 2017 dataset for all studies, where \* represents the best results.

ISBI 2017 – Overall	DSC	Jac.	Sen.	Spec.	Acc.
Team - Mt.Sinai	84.90	76.50	82.50	97.50	93.40
Team - NLP LOGIX	84.70	76.20	82.00	97.80	93.20
Team - BMIT (Our Previous)	84.40	76.00	80.20	98.50	93.40
Team - BMIT (Our Previous)	84.20	75.80	80.10	98.40	93.40
Team - RECOD Titans	83.90	75.40	81.70	97.00	93.10
MSCA	58.71	47.93	54.59	92.65	83.12
SSLS	57.49	44.77	46.29	99.40*	83.92
FCN	81.64	73.12	83.57	95.96	92.86
Our (DCL-PSI)	85.66*	77.73*	86.20*	96.71	94.08*

Table 2. Comparison of segmentation results on the ISBI 2017 dataset for non-melanoma studies.

ISBI 2017 – Non-melanoma	DSC	Jac.	Sen.	Spec.	Acc.
Team - Mt.Sinai	85.81	77.78	83.85	97.69	94.18
Team - NLP LOGIX	86.07	78.02	83.99	97.93	94.23
Team - BMIT (Our Previous)	85.50	77.60	81.85	98.57	94.34
Team - BMIT (Our Previous)	85.38	77.45	81.81	98.54	94.31
Team - RECOD Titans	85.11	77.00	83.46	97.04	94.02
MSCA	60.46	49.74	57.09	92.26	84.19
SSLS	59.17	46.54	48.17	99.37*	85.42
FCN	82.01	73.74	84.30	96.30	93.55
Our (DCL-PSI)	86.63*	79.07*	87.54*	97.35	95.05*

Table 3. Comparison of segmentation results on the ISBI 2017 dataset for melanoma studies.

ISBI 2017 – Melanoma	DSC	Jac.	Sen.	Spec.	Acc.
Team - Mt.Sinai	81.04	71.20	76.99	96.88	89.96
Team - NLP LOGIX	79.07	68.82	73.75	97.52	89.03
Team - BMIT (Our Previous)	79.62	69.28	73.40	97.99	89.64
Team - BMIT (Our Previous)	79.48	69.06	73.26	97.97	89.56
<b>RECOD</b> Titans	79.08	68.78	74.35	96.66	89.37
MSCA	51.49	40.46	44.27	94.27	78.70
SSLS	50.55	37.43	38.52	99.52*	77.73
FCN	80.10	70.55	80.56	94.54	90.02
Our (DCL-PSI)	81.65*	72.18*	80.67*	94.07	90.08*



Figure 8. PR curves of different methods on ISBI 2017 Skin Lesion Challenge dataset, where (i, ii, iii) represent overall, non-melanoma and melanoma studies, respectively.

# D. Results on ISBI 2016 Dataset

Table 4 to Table 6 and Figure 9 show that our DCL-PSI method outperformed all of the current methods on the ISBI 2016 dataset. When compared with the recently published works MFCN and J-FCN, our method achieved a better performance with an increase of ~1.2% in Jaccard measure (Table 4),

Table 4.	Com	parison	of	segmentation	results	on th	e ISBI	[ 2016	dataset	for	all	studies

ISBI 2016 - Overall	DSC	Jac.	Sen.	Spec.	Acc.
Team - ExB	91.00	84.30	91.00	96.50	95.30
Team - CUMED	89.70	82.90	91.10	95.70	94.90
Team - Rahman	89.50	82.22	88.00	96.90	95.20
Team - SFU	88.50	81.11	91.50	95.50	94.40
Team - TMU	88.80	81.10	83.20	98.70*	94.60
MSCA	75.88	66.19	78.30	91.31	85.68
SSLS	69.97	57.20	70.04	97.31	84.67
FCN	88.64	81.37	91.70	94.90	94.13
MFCN	91.18	84.64	92.17	96.54	95.51
J-FCN	91.20	84.70	91.80	96.60	95.50
Our (DCL-PSI)	91.77*	85.92*	93.11*	96.05	95.78*

ISBI 2016 – Non- melanoma	DSC	Jac.	Sen.	Spec.	Acc.
Team - ExB	91.18	84.64	91.12	97.22	95.78
Team - CUMED	89.68	82.95	90.82	96.55	95.30
Team - Rahman	89.44	82.04	87.84	97.51	95.70
Team - SFU	88.32	80.88	91.55	95.82	94.93
Team - TMU	88.58	80.73	82.89	99.04*	94.87
MSCA	75.11	65.57	78.59	91.14	85.84
SSLS	70.81	58.34	72.87	97.15	86.15
FCN	88.66	81.38	91.17	95.87	94.55
MFCN	90.97	84.34	91.63	97.20	95.71
Our (DCL-PSI)	91.78*	85.60*	92.95*	96.79	96.15*

Table 5. Comparison of segmentation results on the ISBI 2016 dataset for non-melanoma studies.

Table 6. Comparison of segmentation results on the ISBI 2016 dataset for melanoma studies.

ISBI 2016 - Melanoma	DSC	Jac.	Sen.	Spec.	Acc.
Team - ExB	90.11	82.94	90.57	93.84	93.23
Team - CUMED	89.98	82.90	92.47	92.34	93.21
Team - Rahman	89.93	82.65	88.72	94.44	93.22
Team - SFU	89.44	81.88	91.16	94.13	92.19
Team - TMU	89.68	82.31	84.62	97.48	93.43
MSCA	79.00	68.68	77.12	92.01	85.02
SSLS	66.55	52.59	58.58	97.94*	78.67
FCN	88.56	81.33	93.83	90.98	92.39
MFCN	92.03*	85.84*	94.34*	93.89	94.70*
Our (DCL-PSI)	91.72	85.62	93.77	93.05	94.29



Figure 9. PR curves of different methods on ISBI 2016 Skin Lesion Challenge dataset.

## E. Results on PH2 Dataset

Table 7 to Table 9 and Figure 10 show that our method achieved leading performance for both non-melanoma and melanoma studies on PH2 dataset. It also greatly improved over recently published methods with a large margin (Table 7): ~10% in Jaccard measure compared with SCDRR, an increase of ~5.64% in Accuracy

PH2 – Overall	DSC	Jac.	Sen.	Spec.	Acc.
SCDRR	86.00	76.00	-	-	-
DT	-	-	80.24	97.22	89.66
JCLMM	82.85	-	-	-	-
MSCA	81.57	72.33	79.87	95.57	88.75
SSLS	78.38	68.16	75.32	98.18*	84.85
FCN	89.38	82.15	93.14	93.00	93.48
MFCN	90.66	83.99	94.89	93.98	94.24
J-FCN	91.50	-	-	-	-
Our (DCL-PSI)	92.10*	85.90*	96.23*	94.52	95.30*

Table 7. Comparison of segmentation results on the PH2 dataset for all studies.

Table 8. Comparison of segmentation results on the PH2 dataset for non-melanoma studies.

PH2 – Non-melanoma	DSC	Jac.	Sen.	Spec.	Acc.
DT	-	-	86.79	97.47	93.74
MSCA	85.52	76.88	85.78	96.33	93.86
SSLS	84.72	75.52	83.96	98.05*	91.77
FCN	89.27	82.01	94.83	94.22	94.79
MFCN	90.77	84.15	95.64	95.12	95.61
Our (DCL-PSI)	92.26*	86.05*	97.11*	95.85	96.61*

Table 9. Comparison of segmentation results on the PH2 dataset for melanoma studies.

PH2 – Melanoma	DSC	Jac.	Sen.	Spec.	Acc.
DT	-	-	54.04	95.97	66.15
MSCA	65.77	54.13	56.25	92.49	68.31
SSLS	53.00	38.73	40.74	98.67*	57.16
FCN	89.81	82.72	91.39	88.16	88.25
MFCN	90.25	83.35	91.88	89.42	88.78
Our (DCL-PSI)	91.44*	85.33*	92.70*	89.19	90.05*



Figure 10. PR curves of different methods on PH2 dataset.

## F. Component Analysis

Figure 11 shows the PR curves of our method and its individual components. We also adopted our method with the traditional FCN for further comparison. It shows that the deep class-specific learning improved the segmentation performance for each specific class: the N-Model performed better for non-melanoma studies while M-Model performed better for melanoma studies and that the probability based step-wise integration approach improved upon the results produced by any of the individual models.



Figure 11. PR curves of our method at different stages on ISBI 2017 dataset. Top row shows the results when using traditional FCN architecture while bottom row shows the results when using the proposed ResFCN architecture.

#### **IV. DISCUSSIONS**

Our findings show that our DCL-PSI method achieved higher accuracy than all other methods on three well-established public datasets and we attribute this to DCL being able to learn important visual features of skin lesions and the PSI to fuse the segmentation results and further refine the segmentation.



Figure 12. Segmentation results from four examples: (a) input images; (b) ground truth (GT); (c) segmentation obtained from our DCL-PSI method; and (d) segmentations obtained by MSCA, SSLS, FCN and MFCN, respectively, from top to bottom.

The traditional methods we investigated – MSCA, SSLS, DT, SCDRR and JCLMM – had difficulty separating the skin lesions from color band artifacts and performed poorly on the challenging heterogeneous skin lesions (see example in Figure 12) because they were not able to identify the image-wide pixel variations and texture differences between artifacts and skin lesions. This resulted in ~20% (Table 1, Table 4, Table 7) lower values in the Jaccard measure when compared with our DCL-PSI method. These methods are also heavily reliant on tuning a large number of parameters and selecting appropriate pre-processing procedures and so have difficulty in adapting for different datasets.

The improvement of the FCN based methods - FCN, J-FCN and MFCN - over the traditional methods are

due to the FCN being able to combine deep semantic information (upper layers) and shallow appearance information (lower layers) in a hierarchical manner that enables them to encode image-wide location information and semantic characteristics. Table 4 compares our method with both J-FCN and MFCN on the ISBI 2016 dataset. Table 4 also illustrates that our method had better sensitivity when compared to J-FCN and MFCN and WFCN and we attribute this to their use of the VGGNet architecture, which limits their ability to add convolutional layers to learn deeper features for detection.

The next closest results to our method on ISBI 2017 dataset (Table 1) were the methods proposed by Mt.Sinai [57], NLP LOGIX [58] and our previous BMIT [59] method. Mt.Sinai used a 29-layer FCN architecture with color based data augmentation, NLP LOGIX used a U-Net [60] architecture, and our previous BMIT trained the ResFCN with an additional 8,000 (10,000 in total) dermoscopic images. We explain the inferior performance of these methods because they were trained and inferred in a single class (group, both melanoma and non-melanoma) and so had imbalanced segmentations for non-melanoma and melanoma studies. In contrast, our method enables to learn deeper class-specific visual features and infer according to the relevance of the dermoscopic image to the DCL models. Hence, our method retained good performances for both non-melanoma studies and challenging melanoma studies. For instance, our method was 1.29% higher than Mt.Sinai in Jaccard measure for non-melanoma studies (Table 2) and our method was 3.36% higher than NLP LOGIX and 2.90% higher than BMIT in Jaccard measure for melanoma studies (Table 3).

Our analysis of the individual components on both using traditional FCN and ResFCN architecture showed that the N-Model performed better for non-melanoma studies while the M-Model performed better for melanoma studies and that the proposed DCL-PSI method performed better than their individual models (Figure 11). The improvement over O-Model indicates the benefits in using probability based step-wise integration approach to iteratively constrain the boundary definitions of the skin lesions. The relatively small improvement of our M-Model, we suggest, was due to the small training dataset of melanoma studies, which limited our method's ability to learn subtler variations.

# V. CONCLUSIONS

In this paper, we proposed a method to automatically segment skin lesions in dermoscopic images. Our

method introduced a deep class-specific learning to quantify the subtle visual characteristics of different types of skin lesions and a probability-based step-wise integration of the derived class-specific segmentation maps to ensure visual consistency of the segmented regions. By learning and inferring important visual characteristics of the skin lesions with a deep ResNet based classifier, we were able to achieve accurate segmentation results across both non-melanoma and melanoma studies. In particular, we observed that our method's ability to extract and integrate complementary class-specific information, in an iterative manner, ensured visual consistency in the segmented region boundaries for skin lesions that are challenging to segment, such as those with fuzzy boundaries and/or low contrast to the background. Our benchmark experiments against 18 state-of-art methods on three well-established public datasets (ISBI 2017, ISBI 2016 Skin Lesion Challenge datasets and PH2 dataset) showed that our method is consistently more accurate than existing methods across different datasets.

Based on the superior performance of the proposed method, we plan to explore the adaptions of our method to different applications, including automatic skin lesion monitoring and tracking over multiple image sequences. We will also extend our method to support different medical imaging datasets which have similar imbalanced distribution problem such as in pathology affected lung segmentation in CT (computed tomography) images.

## VI. ACKNOWLEDGEMENT

This work was supported in part by Australia Research Council (ARC) grants.

#### VII. REFERENCES

- [1] D. S. Rigel, R. J. Friedman, and A. W. Kopf, "The incidence of malignant melanoma in the United States: issues as we approach the 21st century," *Journal of the American Academy of Dermatology*, vol. 34, pp. 839-847, 1996.
- C. Barata, M. E. Celebi, and J. S. Marques, "Development of a clinically oriented system for melanoma diagnosis," *Pattern Recognition*, vol. 69, pp. 270-285, 2017.
- [3] 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," *Computerized Medical Imaging and Graphics*, vol. 31, pp. 362-373, 2007.

- [4] G. Capdehourat, A. Corez, A. Bazzano, R. Alonso, and P. Musé, "Toward a combined tool to assist dermatologists in melanoma detection from dermoscopic images of pigmented skin lesions," *Pattern Recognition Letters*, vol. 32, pp. 2187-2196, 2011.
- [5] M. E. Celebi, H. Iyatomi, W. V. Stoecker, R. H. Moss, H. S. Rabinovitz, G. Argenziano, and H. P. Soyer, "Automatic detection of blue-white veil and related structures in dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 32, pp. 670-677, 2008.
- [6] Q. Abbas, M. E. Celebi, C. Serrano, I. F. GarcíA, and G. Ma, "Pattern classification of dermoscopy images: A perceptually uniform model," *Pattern Recognition*, vol. 46, pp. 86-97, 2013.
- [7] C. Barata, M. Emre Celebi, and J. S. Marques, "Improving dermoscopy image classification using color constancy," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, pp. 1146-1152, 2015.
- [8] 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, pp. 115-118, 2017.
- C. Serrano and B. Acha, "Pattern analysis of dermoscopic images based on Markov random fields," *Pattern Recognition*, vol. 42, pp. 1052-1057, 2009.
- [10] M. E. Celebi, H. Iyatomi, G. Schaefer, and W. V. Stoecker, "Lesion border detection in dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 33, pp. 148-153, 2009.
- [11] 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," *Dermoscopy Image Analysis*, pp. 97-129, 2015.
- [12] 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 Journal of Selected Topics in Signal Processing*, vol. 3, pp. 35-45, 2009.
- [13] R. Garnavi, M. Aldeen, M. E. Celebi, G. Varigos, and S. Finch, "Border detection in dermoscopy images using hybrid thresholding on optimized color channels," *Computerized Medical Imaging and Graphics*, vol. 35, pp. 105-115, 2011.
- [14] M. Emre Celebi, Q. Wen, S. Hwang, H. Iyatomi, and G. Schaefer, "Lesion border detection in dermoscopy images using ensembles of thresholding methods," *Skin Research and Technology*, vol. 19, pp. e252-e258, 2013.
- [15] H. Zhou, G. Schaefer, M. E. Celebi, F. Lin, and T. Liu, "Gradient vector flow with mean shift for skin lesion segmentation," *Computerized Medical Imaging and Graphics*, vol. 35, pp. 121-127, 2011.
- [16] J. Tang, "A multi-direction GVF snake for the segmentation of skin cancer images," *Pattern Recognition*, vol. 42, pp. 1172-1179, 2009.
- [17] X. Yuan, N. Situ, and G. Zouridakis, "A narrow band graph partitioning method for skin lesion segmentation," *Pattern Recognition*, vol. 42, pp. 1017-1028, 2009.
- M. Emre Celebi, H. A. Kingravi, H. Iyatomi, Y. Alp Aslandogan, W. V. Stoecker, R. H. Moss, J. M. Malters, J. M. Grichnik,
   A. A. Marghoob, and H. S. Rabinovitz, "Border detection in dermoscopy images using statistical region merging," *Skin Research and Technology*, vol. 14, pp. 347-353, 2008.

- [19] R. Nock and F. Nielsen, "Statistical region merging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 1452-1458, 2004.
- [20] 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 Journal of Biomedical and Health Informatics*, vol. 21, pp. 1685 - 1693, 2017.
- [21] E. Ahn et al., "Automated Saliency-based Lesion Segmentation in Dermoscopic Images," in *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 3009 3012.
- [22] A. Pennisi, D. D. Bloisi, D. Nardi, A. R. Giampetruzzi, C. Mondino, and A. Facchiano, "Skin lesion image segmentation using Delaunay Triangulation for melanoma detection," *Computerized Medical Imaging and Graphics*, vol. 52, pp. 89-103, 2016.
- [23] B. Bozorgtabar, M. Abedini, and R. Garnavi, "Sparse Coding Based Skin Lesion Segmentation Using Dynamic Rule-Based Refinement," in *Proceedings of the International Workshop on Machine Learning in Medical Imaging*, 2016, pp. 254-261.
- [24] 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 *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)*, 2016, pp. 1059-1062.
- [25] P. Wighton, T. K. Lee, H. Lui, D. I. McLean, and M. S. Atkins, "Generalizing common tasks in automated skin lesion diagnosis," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, pp. 622-629, 2011.
- [26] A. R. Sadri, M. Zekri, S. Sadri, N. Gheissari, M. Mokhtari, and F. Kolahdouzan, "Segmentation of dermoscopy images using wavelet networks," *IEEE Transactions on Biomedical Engineering*, vol. 60, pp. 1134-1141, 2013.
- [27] F. Xie and A. C. Bovik, "Automatic segmentation of dermoscopy images using self-generating neural networks seeded by genetic algorithm," *Pattern Recognition*, vol. 46, pp. 1012-1019, 2013.
- [28] Y. He and F. Xie, "Automatic skin lesion segmentation based on texture analysis and supervised learning," in *Proceedings* of the Asian Conference on Computer Vision, 2012, pp. 330-341.
- [29] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431-3440.
- [30] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 640-651, 2017.
- [31] Y. Yuan, M. Chao, and Y.-C. Lo, "Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks with Jaccard Distance," *IEEE Transactions on Medical Imaging*, vol. 36, pp. 1876 - 1886, 2017.
- [32] Y. Lequan, H. Chen, Q. Dou, J. Qin, and P. A. Heng, "Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks," *IEEE Transactions on Medical Imaging*, vol. 36, pp. 994-1004, 2017.
- [33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.
- [34] L. Bi, J. Kim, E. Ahn, A. Kumar, M. Fulham, and D. Feng, "Dermoscopic Image Segmentation via Multi-Stage Fully Convolutional Networks," *IEEE Transactions on Biomedical Engineering*, vol. 64, pp. 2065-2074, 2017.

- [35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [36] 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.
- [37] 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 *Proceedings of the IEEE International Conference of Engineering in Medicine and Biology Society* (*EMBC*) 2013, pp. 5437-5440.
- [38] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, and H. Kittler, "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)," arXiv preprint arXiv:1710.05006, 2017.
- [39] X. Zhang, J. Zou, K. He, and J. Sun, "Accelerating very deep convolutional networks for classification and detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, pp. 1943-1955, 2016.
- [40] H. Chen, X. Qi, L. Yu, and P.-A. Heng, "DCAN: Deep Contour-Aware Networks for Accurate Gland Segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2487-2496.
- [41] L. Wang, L. Wang, H. Lu, P. Zhang, and X. Ruan, "Saliency detection with recurrent fully convolutional networks," in Proceedings of the European Conference on Computer Vision, 2016, pp. 825-841.
- [42] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, 2011, pp. 315-323.
- [43] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, 2015.
- [44] A. Veit, M. J. Wilber, and S. Belongie, "Residual networks behave like ensembles of relatively shallow networks," in Proceedings of the Advances in Neural Information Processing Systems, 2016, pp. 550-558.
- [45] G. Lin, A. Milan, C. Shen, and I. Reid, "Refinenet: Multi-path refinement networks with identity mappings for highresolution semantic segmentation," arXiv preprint arXiv:1611.06612, 2016.
- [46] J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, A. Senior, P. Tucker, K. Yang, and Q. V. Le, "Large scale distributed deep networks," in *Proceedings of the Advances in Neural Information Processing Systems*, 2012, pp. 1223-1231.
- [47] 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 Transactions on Medical Imaging*, vol. 35, pp. 1285-1298, 2016.
- [48] A. Vedaldi and K. Lenc, "Matconvnet: Convolutional neural networks for matlab," in *Proceedings of the ACM International Conference on Multimedia*, 2015, pp. 689-692.
- [49] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248-255.

- [50] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, "An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, pp. 31-40, 2016.
- [51] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Proceedings of the Advances in Neural Information Processing Systems, 2012, pp. 1097-1105.
- [52] Y. Qin, M. Feng, H. Lu, and G. W. Cottrell, "Hierarchical Cellular Automata for Visual Saliency," *arXiv preprint arXiv:1705.09425*, 2017.
- [53] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, pp. 62-66, 1979.
- [54] Z. Wu, C. Shen, and A. v. d. Hengel, "Wider or Deeper: Revisiting the ResNet Model for Visual Recognition," *arXiv preprint arXiv:1611.10080*, 2016.
- [55] A. Roy, A. Pal, and U. Garain, "JCLMM: A Finite Mixture Model for Clustering of Circular-Linear data and its application to Psoriatic Plaque Segmentation," *Pattern Recognition*, vol. 66, pp. 160-173, 2017.
- [56] X. Li, Y. Li, C. Shen, A. Dick, and A. Van Den Hengel, "Contextual hypergraph modeling for salient object detection," in Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 3328-3335.
- [57] Y. Yuan, M. Chao, and Y.-C. Lo, "Automatic skin lesion segmentation with fully convolutional-deconvolutional networks," arXiv preprint arXiv:1703.05165, 2017.
- [58] M. Berseth, "ISIC 2017-Skin Lesion Analysis Towards Melanoma Detection," arXiv preprint arXiv:1703.00523, 2017.
- [59] L. Bi, J. Kim, E. Ahn, and D. Feng, "Automatic Skin Lesion Analysis using Large-scale Dermoscopy Images and Deep Residual Networks," *arXiv preprint arXiv:1703.04197*, 2017.
- [60] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 234-241.