This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2973614, IEEE Journal of Biomedical and Health Informatics

An End-to-end Multi-task Deep Learning Framework for Skin Lesion Analysis

Lei Song, Jianzhe Lin, *Student Member, IEEE,* Z. Jane Wang, *Fellow, IEEE,* Haoqian Wang

***Abstract*—Automatic skin lesion analysis of dermoscopy images remains a challenging topic. In this paper, we propose an end-to-end multi-task deep learning framework for automatic skin lesion analysis. The proposed framework can perform skin lesion detection, classification, and segmentation tasks simultaneously. To address the class imbalance issue in the dataset (as often observed in medical image datasets) and meanwhile to improve the segmentation performance, a loss function based on the focal loss and the jaccard distance is proposed. During the framework training, we employ a three-phase joint training strategy to ensure the efficiency of feature learning. The proposed framework outperforms state-of-the-art methods on the benchmarks *ISBI 2016 challenge dataset towards melanoma classification* and *ISIC 2017 challenge dataset towards melanoma segmentation*, especially for the segmentation task. The proposed framework should be a promising computer-aided tool for melanoma diagnosis.**

***Index Terms*—Skin lesion analysis, end-to-end multi-task framework, deep learning, melanoma segmentation, convolution neural networks.**

I. INTRODUCTION

MELANOMA is the most lethal skin cancer and faces an increasing incidence over the past years. For instance, about 100,000 new diagnosed cases and 7000 deaths are expected in the United States in 2019. Fortunately, melanoma treatment benefits significantly from early detection, i.e., it has a 92% 5-year relative survival rate, while the survival rate in the later stage is only 26% [1]. Therefore early detection and diagnosis of melanoma is of great importance.

Dermoscopy, as a common medical imaging tool for melanoma diagnosis, can enhance the distinction between skin lesions and other normal skin areas by removing surface glare. Dermoscopy is promising to detect skin lesion early. However, advanced image processing and learning methods are still needed for automatic melanoma/skin lesion analysis using dermoscopy images. Currently, the lack of standard diagnosis rules brings difficulties for skin lesion analysis. In addition, there are several challenges in demorscopy images analysis:

This work is partially supported by the NSFC grant 61831014, in part by the Shenzhen Science and Technology Project under Grant numbers (JCYJ20180228175315535, JCYJ20180508152042002, GGFW2017040714161462).

Lei Song is with the Department of Automation, Tsinghua Shenzhen International Graduate School, Shenzhen 518055, China (e-mail: songl17@mails.tsinghua.edu.cn).

Jianzhe Lin and Z. Jane Wang are with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: zjanew@ece.ubc.ca).

Haoqian Wang is with the Tsinghua Shenzhen International Graduate School, Shenzhen 518055, China and Shenzhen Institute of Future Media Technology, Shenzhen 518071, China (e-mail: wanghaoqian@tsinghua.edu.cn).

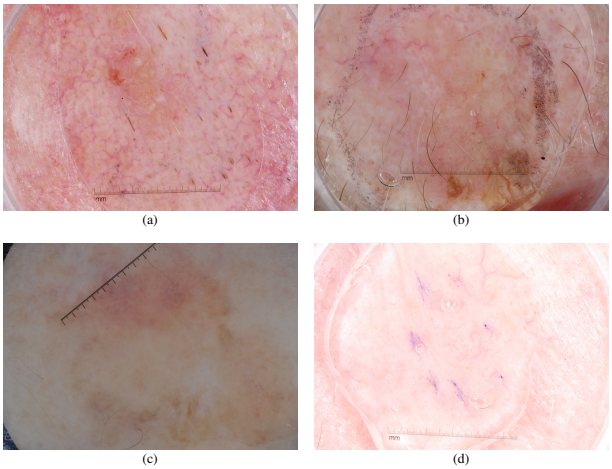

Fig. 1. Illustrative image examples to show difficulties in dermoscopy image analysis. (a) hard to identify with veins; (b) the presence of artifacts; (c) obscure and irregular boundaries; and (d) low contrast in visual impression.

Melanoma shares a high visual similarity with non-melanoma lesions even for demorscopy images; Besides obscure boundaries and the presence of artifacts, obstructions like hairs and veins make the skin lesion analysis more challenging. Some example images are shown in Fig. 1 to illustrate the challenges.

Many efforts have been dedicated to address these challenges by developing efficient image processing methods. Such methods can be classified broadly into two categories: One is conventional image processing and machine learning algorithms based on hand-crafted features; and the other is deep learning related methods, with focus on convolution neural networks(CNNs). For the first type, certain low level hand-crafted features including shape [2], color [3] [4] or texture [5] are investigated for melanoma diagnosis. On this basis, different feature selection algorithms are proposed to choose the most representative skin lesion features and such features can be further combined to improve the performance [6] [7]. These methods generally suffer from poor robustness. When the input image is changed (the relevant feature varies among different images), the algorithm performance may be degraded intensely.

Recently, the deep learning approach has been attracting increasing attention because of its automatic feature extraction ability. Such methods have shown effectiveness in image classification [8] [9], object detection [10] [11] [12], segmentation [13] [14] [15] and so on. Esteva *et al.* [16] were the first to achieve the dermatologist-level classification performance of skin cancer using deep neural networks. Yu *et al.* [17] trained

2168-2194 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northeastern University. Downloaded on June 03,2020 at 02:43:04 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2973614, IEEE Journal of Biomedical and Health Informatics

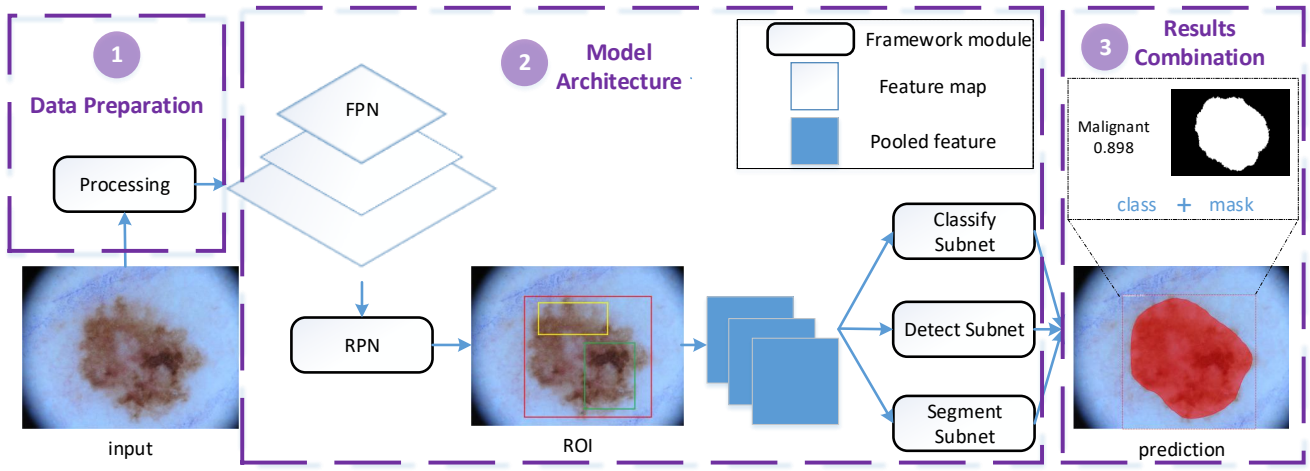

Fig. 2. Flowchart of the proposed skin lesion analysis framework. The dermoscopy images are processed by data preparation and model architecture in sequence. In results combination, predictions of three convolution subnets are fused to produce final results.

two deep residual networks where one was used to obtain the skin lesion mask, while the other output the classification results according to the learned mask. Li *et al.* [18] proposed a rotation equivariant network for skin lesion segmentation. Specially, they employed G-upsampling and G-projection to increase the network efficiency. Venkatesh *et al.* [19] presented a deep residual network based on UNET [13] for melanoma segmentation. However, the above deep learning based approaches share certain common problems: First, some methods are not end-to-end pipelines, and usually require complex post-processing steps (e.g., conditional random field, adjacent region smooth to refine the obtained results). Second, most of these methods are proposed for a single task either classification or segmentation, while performing multiple tasks at the same time is desired in skin lesion analysis. Finally, medical image datasets often contain many normal samples with much less abnormal ones, leading to the so-called class imbalance problem. The class imbalance problem affects the performance of these approaches severely due to the use of cross entropy loss function. The cross entropy loss function usually assumes that all samples (including positive and negative samples) contribute equally to the loss computation. Taking the skin lesion problem as an example, the network trained by the cross entropy loss is hard to identify the malignant class, which has much less training samples than the predominant class (the benign class). An overview comparison of these deep learning based approaches is shown in Tab. I.

To address such concerns, we propose a novel end-to-end deep learning multi-task framework for automatic skin lesion analysis. The framework takes the original dermoscopy image as the input and outputs the classification, detection and segmentation results of melanoma jointly. Specifically, the proposed framework can classify the melanoma as benign or malignant, meanwhile locate the position of the melanoma and generate an accurate mask of each instance. An effective loss function based on the focal loss [30] and the jaccard distance is incorporated into our framework, which alleviates the class imbalance problem in dermoscopy images and improves the segmentation accuracy. We evaluate the performance of the proposed framework extensively on the ISBI 2016 [31] and ISIC 2017 challenge datasets [24]. The contributions of this paper are three-fold:

- We propose a novel end-to-end deep learning framework for automatic skin lesion analysis. The framework doesn't require any additional post-processing steps.
- The proposed framework can deal with multiple tasks at the same time, including classification, detection and segmentation for melanoma. To the best of our knowledge, this is the first attempt to develop a real multi-task framework for skin lesion analysis.
- We design an effective loss function based on the focal loss and the jaccard distance to alleviate the class imbalance problem for skin image datasets.

The rest of this paper is organized as follows. We describe the proposed framework in Section II. The specific experiments and corresponding results are reported in Section III, and further discussions are given in Section IV. Finally, we conclude our study in Section V.

TABLE I
AN OVERVIEW OF SKIN LESION ANALYSIS IN DERMOSCOPY IMAGES AMONG DIFFERENT DEEP LEARNING METHODS

Methods	Task	End-to-end	Loss Function	Model
Esteva 2017 [16]	classfication(cls)	YES	cross entropy(CE)	Inception v3
Yu 2017 [17]	cls	NO	CE	FCRN+DRN
Lopez 2017 [20]	cls	NO	CE	VGG16
Vasconcelos 2017 [21]	cls	NO	CE	GoogleNet
Yu Z 2017 [22]	cls	NO	CE	AlexNet+SVM
Li 2018 [18]	segmentation(seg)	YES	CE	ResNet34
Venkatesh 2018 [19]	seg	YES	CE	UNET
Yuan Y 2017 [23]	seg	NO	Jaccard Loss(JL)	CDNN29
Matt Berseth 2017 [24]	seg	NO	CE	UNET
Bi L 2017 [25]	seg	YES	CE	ResNet+FCN
Vesal 2018 [26]	seg	NO	CE	Faster-RCNN
Menegola A 2017 [27]	seg	NO	Dice Coefficient	VGG16
Yang 2017 [28]	cls+seg	NO	CE	GoogleNet
Chen 2018 [29]	cls+seg	NO	CE	ResNet101
Ours	cls+detection+seg	YES	Focal Loss+JL	FPN+RPN+FCN

2168-2194 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northeastern University. Downloaded on June 03,2020 at 02:43:04 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2973614, IEEE Journal of Biomedical and Health Informatics

II. METHOD

Automatic skin lesion analysis is challenging due to high visual similarity and intra-class variation between melanoma and non-melanoma lesions. The lack of standard diagnosis rules makes it necessary to develop an accurate and efficient framework for melanoma analysis. The proposed framework is shown in Fig. 2. In this section, we first present the data preparation procedure and then describe each component of the model architecture. At last, we elaborate on the proposed loss function used in the framework.

A. Data Preparation

Data preparation is the first step in the proposed framework, where dermoscopy images with arbitrary sizes are fed as the inputs. The entire preparation flow is composed of image size organization, zero-center normalization as well as data augmentation in sequence.

1) Image Size Organization: The original dermoscopy images show a wide resolution range varying from 540×576 to 6688×6780. Usually, existing papers cut the original image into several patches equally [13] [32] or put a sliding window on the original image and randomly crop patches [33] [34]. Such approaches could damage the neighborhood relationships in objects, and patches only covering the background might disturb the model training. In our approach, we take the resized entire image as the input instead of image patches.

Based on a simple statistic study of dermoscopy images in terms of image size and color, as shown in Fig. 3 and Tab. II, we can see that the major image size is around 2016×3024. Considering the trade-off between limited GPU memory and enough information for training, we choose the size 1024×1024. To avoid the distortion caused by aspect ratio change, we resize the original image proportionally to make the long side to be 1024, and we then pad the resized image periphery with zero to obtain a 1024×1024 image. Image size organization arranges the irregular size of an image into the universal 1024×1024, which is convenient for efficient training in batches.

2) Zero-center Normalization: Following [35], we perform zero-center normalization on the resized images by subtracting pixel means in R, G, B channels separately. This step can help reduce the effects of the illuminance, and also help alleviate the exploding gradients problem in network training caused by the inputs with large variability [36] [37].

3) Data Augmentation: We perform data augmentation to enhance the model generalization ability. These five operations include flipping images horizontally and vertically, applying the affine transformation, multiplying each image with a random value between 0.8 and 1.5, and Gaussian blurring. Due to limited memory at our machine, we adopt a different data augmentation strategy to improve the diversity of training data: the number of samples after data augmentation stays the same as that of the original training dataset (12666 benign samples, 1084 malignant samples). To be more specific, for each input image, it is selected randomly from six images (one original image version and five transformed versions) with equal probability.

B. Deep Learning Model Architecture

After data preparation, the processed images are sent to the deep learning model architecture. The proposed model architecture consists of feature pyramid network (FPN), region proposal network (RPN) and three convolution subnets, where these subnets are used for classification, detection and segmentation separately, as shown in Fig. 2. FPN utilizes the multi-scale features to improve the model representation ability and RPN extracts region of interest (ROI) in dermoscopy images.

1) Feature Pyramid Network(FPN): The model needs to classify and localize objects over a wide range of scales. If only detecting features on a certain scale, the model performance will decrease dramatically. Here we utilize FPN proposed by Lin *et al.* [10] to handle such problem. Our framework employs ResNet [9] as the model backbone, which takes the prepared 1024×1024 images as the inputs and reduces the feature map size in half when passed by every level. That is, the size of feature map in level 1 to 5 are 512×512, 256×256, 128×128, 64×64, 32×32, respectively. We then do a 1×1 convolution on feature map in level 2 to 5 and denote as C2-C5. We use C5 as the top layer output of the FPN (P5), and after that, the second layer output (P4) is constructed by merging C4 and the feature map that up-samples P5 double. Similarly, the P3 and P2 can be gained in sequence. Finally, we take P2-P5 after a 3×3 convolution as the FPN outputs and feed into the next RPN.

Similar to standard feature pyramid in image processing, the feature map in low resolution can provide more strong features including semantic and contour information while the high resolution feature map describes more on details. FPN employed in our framework integrates features among different scale and enhances the model scale invariance and equivariance.

2) Region Proposal Network(RPN): Inspired by Faster R CNN [11], we use RPN for distinguishing foreground objects and background in dermoscopy images. RPN is similar to a binary classification network based on the softmax. Firstly, it puts many anchors on the inputs to scan every region in image area, where anchors are rectangle boxes generated by different size and aspect ratio. Secondly, for each of these anchors, we do the classification to pick up the highest score bounding box that contains foreground objects. Then, bounding box regression is applied on picked anchors to make the anchor cover the ground truth precisely, and choose positive and negative anchors according to the intersection-over-union (IOU) higher than 0.7, lower than 0.3, respectively. In the end, proposals are acquired after applying non-maximum suppression algorithm on positive and negative anchors.

RPN eliminates much irrelevant background information and increases the model learning efficiency of salient regions. Moreover, we set a fixed ratio between positive and negative anchors in RPN training (1:3 in our experiment), that means we only keep a proper number of negative anchors whatever the number of positive anchors changes. It eases the class imbalance problem in medical images for network training, and laying a foundation of subsequent three convolution subnets working in accuracy.

2168-2194 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northeastern University. Downloaded on June 03,2020 at 02:43:04 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2973614, IEEE Journal of Biomedical and Health Informatics

TABLE II
TRAINING DATASET: STATISTICAL INFORMATION ABOUT DERMOSCOPY IMAGES IN IMAGE SIZE AND COLOR

	mean	median	min	max
#Height	2163.18	2016.00	540.00	6688.00
#Width	3220.65	3024.00	576.00	6780.00
#Color(R,G,B)	(175.65,147.35,129.03)	(174.54, 147.22, 128.17)	(49.74, 27.59, 23.12)	(254.15, 236.11, 230.88)

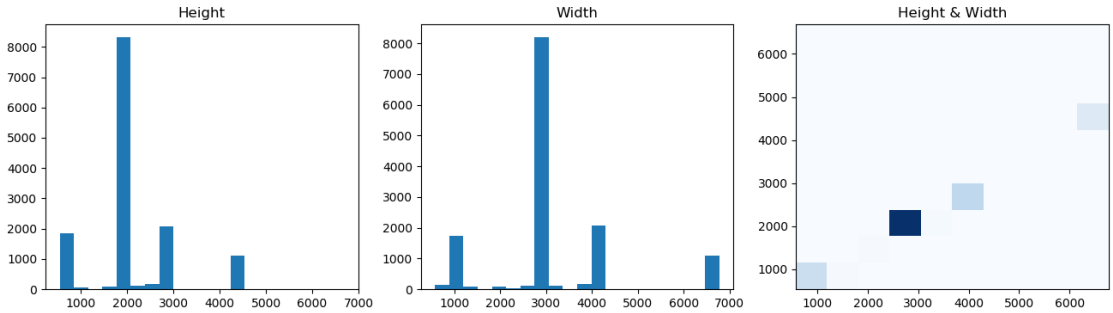

Fig. 3. Statistical information of dermoscopy images regarding the image size. For the left and medium charts, the horizontal axis denotes the number of pixels and the vertical axis stands for occurrence frequency. For the right chart, horizontal and vertical axes represent the width and the height (pixel sizes) of the images respectively, and a darker color means a higher frequency.

3) Convolution Subnets: After RPN outputs the ROI in images, we do feature pooling firstly to get the fixed size of proposal feature map (we use 7×7 in our experiment), because of full-connected layers used in the Classify and Detect subnets. Then the pooled feature maps are fed into three convolution subnets separately. One is the Classify subnet, another two are the Detect subnet and the Segment subnet. The Classify subnet is composed of two fully-connected layers and computes the probability towards a specific melanoma type based on the softmax (benign vs. malignant). The Detect subnet does bounding box regression to refine proposals aiming for more accurate localization. The Segment subnet uses the mask head proposed by He *et al.* [14] to generate segmentation masks for each skin lesion. It consists of four convolution layers and one $2\times$ up-sample layer and one sigmoid activation layer, where the sigmoid layer maps neuron values to binary mask form. Once classification, detection and segmentation results are obtained from these three convolution subnets, the framework shows them on the same dermoscopy image all together, namely results combination.

C. Loss Function

The proposed model architecture can be divided into two stages: Generating the ROI of the input dermoscopy image at the first stage, and then performing classification, detection, segmentation tasks based on the generated ROI at the second stage. The framework total loss is the sum of the RPN loss and the convolution subnets losses, as described below.

1) RPN Loss Function based on the Focal Loss: RPN classifies the foreground from all anchors and does the bounding box regression to obtain a precise location. Hence, its loss function should include classification and regression losses, and we employ the following loss function:

$$L_{rpn}\left(p_i, t_i\right) = \frac{1}{N_{cls}} \sum_i L_{cls}\left(p_i, p_i^*\right) + \rho \frac{1}{N_{reg}} \sum_i p_i^* \cdot L_{reg}\left(t_i, t_i^*\right). \tag{1}$$

where p_i and t_i stand for the prediction probability of anchor i being foreground and its estimated coordinates respectively, while p_i^* and t_i^* denote the corresponding ground truth information, and N_{cls} and N_{reg} are the number of anchors used for computing the classification loss $L_{cls}(p_i, p_i^*)$ and the regression loss $L_{reg}(t_i, t_i^*)$ respectively. ρ is a balancing parameter to make L_{cls} and L_{reg} terms roughly equally weighted, and we set ρ=10 in our experiments.

For the classification loss, existing methods usually compute the cross entropy between positive and negative samples, which assumes that every sample contributes equally to the total loss. In the skin lesion case, much less malignant samples can cause the network fail to learn melanoma features. Inspired by the focal loss [30], we propose an effective way to deal with this problem by the following loss function:

$$L_{cls}\left(p_i, p_i^*\right) = -\alpha \cdot \Big\{ \left(1-p_i\right)^{\beta} \cdot p_i^* \cdot \log p_i + p_i^{\beta} \cdot \left(1-p_i^*\right) \log\left(1-p_i\right) \Big\}. \tag{2}$$

where α is a linear scale factor, and β is an exponential factor maintaining the balance between positive and negative samples. In our experiment, we set $\alpha = 0.25$, $\beta = 2$. The benign lesions are dominant (p_i is higher), so the corresponding weight $(1-p_i)^{\beta}$ gets lower. Similarly, the probability of malignant lesions $(1-p_i)$ is lower and its weight (p_i^{β}) gets higher. Such a loss function magnifies the effect of fewer positive samples (the malignant class) while lessens the impact of excessive negative samples (the benign class).

2168-2194 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northeastern University. Downloaded on June 03,2020 at 02:43:04 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2973614, IEEE Journal of Biomedical and Health Informatics

The smooth L1 loss is adopted here to compute the regression loss due to its insensitivity to outliers [11] [12]. It is defined as:

$$L_{reg}\left(t_i, t_i^*\right) = \begin{cases} 0.5 \cdot \left(t_i - t_i^*\right)^2 & if\ \left|\left(t_i - t_i^*\right)\right| < 1 \\ \left|\left(t_i - t_i^*\right)\right| - 0.5 & otherwise \end{cases} \quad (3)$$

2) Convolution Subnets Loss Function based on Jaccard Distance: Three convolution subnets are used for classification, detection, and segmentation, respectively. We design a multi-task loss function to integrate them and it can be described as:

$$L_{subnets} = L_{cls} + L_{reg} + L_{mask}. \quad (4)$$

where L_{cls} is the loss of Classify subnet, L_{reg} is the loss of Detect subnet and L_{mask} is the loss of Segment subnet. The computation of L_{cls} and L_{reg} are the same as above-mentioned in RPN loss function. The Segment subnet classifies every pixel into object or non-object, and then binarizes the pixel values to generate the mask for each $m \times m$ positive ROI. The jaccard distance measures the dissimilarity between two sets and its complimentary jaccard index is an evaluation metric for medical image segmentation. Here we introduce the jaccard distance in our loss function. It serves as a regularization term to improve the segmentation performance. Moreover, intuitively the proposed loss function with the additional jaccard distance term can make the network devote to output the results that are more similar to the ground truth instead of focusing on the less-informative background too much, and therefore alleviate the impact of the lesion-background imbalance. The specific loss is defined as:

$$L_{mask} = -\frac{1}{m^2} \sum_{1 \le i,j \le m} \left[y_{ij}^* \cdot \log y_{ij} + \left(1 - y_{ij}^*\right) \log\left(1 - y_{ij}\right) \right] + \delta \cdot \left[1 - \frac{y_{ij}^* \cdot y_{ij} + \delta}{\left(y_{ij}^* + y_{ij} - y_{ij}^* \cdot y_{ij}\right) + \delta} \right]. \quad (5)$$

where y_{ij}^* is the label of the mask pixel (i, j), y_{ij} is the corresponding prediction value, and δ is the smooth factor. Here we heuristically set $\delta = 100$ in the experiment.

Based on the above defined losses, the overall loss in the framework is defined as the sum of the RPN loss and the convolution subnets losses.

$$L_{total} = L_{rpn} + L_{subnets} \quad (6)$$

The proposed multi-task learning problem is formulated to minimize the above L_{total}, as described in Section III B.

III. EXPERIMENTS AND RESULTS

A. Datasets

We evaluate the proposed framework on a portion of ISIC images. Since the proposed framework is for multi-task, each training sample must contain multiple label information at the same time, i.e., melanoma type for classification, bounding box for detection and mask for segmentation. Considering there is no existing bounding box annotation, we manage the label information as follows: We gather all images with a melanoma mask (13780 in total), and then draw a minimum rectangle box to cover the whole mask for each melanoma and use this rectangle box as its detection label. A simple statistic of these selected images shows that these samples include 12666 benign, 1084 malignant and 30 undetermined lesions. We drop off 30 undetermined samples and keep the rest 13750 images (benign and malignant) as our network training dataset. For testing, we evaluate the method on the *ISBI 2016 challenge dataset towards melanoma classification* [31] and *ISIC 2017 challenge dataset towards melanoma segmentation* [24].

B. Implementation

The proposed framework is implemented in Python based on the Tensorflow library. In terms of the model architecture, we take ResNet50 or ResNet101 as the main backbone to extract melanoma features, and initialize the network with different pre-trained weights including IMAGENET [8] and COCO [38]. Stochastic gradient descent (SGD) is employed to train the model, where we set the initial learning rate as 0.001, the momentum as 0.9 and the weight decay as 0.0001. Moreover, we employ the three-phase training strategy to make sure that the network learns melanoma features efficiently. Specifically, we train the three convolution subnets for 40 epochs with the initial learning rate 0.001 firstly, and then train the layers from the RPN and up for 80 epochs with the learning rate 0.001, finally finetune all layers in the model architecture for 40 epochs with the learning rate 0.0001. Training and validation steps per epoch are set as 1000 and 50, respectively. The batch size equals to the number of images trained on each GPU multiply by the GPU counts. In our experiment, we set the number of images per GPU as 2, which can be adjusted according to GPU memory and image sizes.

C. Evaluation Metrics

We investigate different performance metrics to evaluate the proposed framework, mainly for classification and segmentation. Notably, we don't report the detection performance due to two reasons: The lack of the detection ground truth in the dataset makes the evaluation impossible; and if we simply use the hand-drawn bounding box by ourselves as the evaluation reference, it will be unconvincing and inaccurate. Also, the segmentation performance reflects on the detection accuracy to some degree since the generated mask contains position information towards detection.

For classification, we calculate the classification accuracy, the average precision (AP), the area under the ROC curve (AUC), the sensitivity (SE) and specificity (SP) to measure classification performances. For segmentation, there are five metrics suggested in ISIC 2017 challenge, including Jaccard index (JA), Dice coefficient (DI), pixel accuracy (AC), sensitivity (SE) and specificity (SP). We compute these metrics for

2168-2194 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northeastern University. Downloaded on June 03,2020 at 02:43:04 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2973614, IEEE Journal of Biomedical and Health Informatics

each tested dermoscopy image and then report the averaged values as the final results. These metrics are defined as:

$$\begin{aligned} JA &= \frac{TP}{TP+FN+FP}, \\ DI &= \frac{2 \cdot TP}{2 \cdot TP+FN+FP}, \\ AC &= \frac{TP+TN}{TP+FP+TN+FN}, \\ SE &= \frac{TP}{TP+FN}, \\ SP &= \frac{TN}{TN+FP}. \end{aligned} \tag{7}$$

where TP, TN, FP, FN refer to the number of true positive, true negative, false positive and false negative, respectively. In our classification evaluation, we denote the case of correctly classified benign samples as TP, wrongly classified malignant samples as FN, correctly classified malignant samples as TN and wrongly classified benign samples as FP. In segmentation, a lesion pixel is marked as true positive if its prediction label is a lesion, otherwise it is marked as false negative. A non-lesion pixel is marked as true negative if its prediction label is a non-lesion, otherwise it is marked as false positive.

D. Classification Performance of Proposed Framework

1) Classification Results with Different Loss Functions: We propose a loss function based on focal loss and jaccard distance in the proposed framework. In order to verify its rationality, we use the cross entropy loss in our framework as the benchmark, and then we employ the focal loss and jaccard loss in the proposed framework, respectively. Further the proposed compound loss function based on both focal loss and jaccard loss is investigated. Evaluation results are reported and compared in Tab. III. All experiments are conducted with the same settings except the loss function (i.e., initialized by IMAGENET, using ResNet101 and with the batch size 1×2). From experimental results, we have the following observations: If we replace the original cross entropy by the focal loss, AC increases from 0.788 to 0.801, SP increases from 0.573 to 0.730 and AP increases from 0.796 to 0.804, which indicates that the focal loss can help alleviate the class imbalance problem to some degree. When we apply the jaccard loss alone in the proposed framework, AC increases from 0.788 to 0.793, AUC increases from 0.754 to 0.783 and AP increases from 0.796 to 0.815. We also add the dice loss experiments and it yields a similar performance improvement when compared with the original loss, but worse than the jaccard loss overall. Moreover, we report the performance combing dice loss, focal loss and jaccard loss for comparison. The corresponding results are worse than that of the focal loss and jaccard loss combination. We think it might be related to the interaction effect of dice loss and jaccard loss. The loss function combining focal loss and jaccard loss yields the highest classification accuracy 0.813, the highest AUC 0.794, SP 0.731 and AP 0.823 at the cost of a slight SE degradation, which supports the effectiveness of our proposed loss function in dealing with the skin lesion problem. It is worth noting that a lower SE and a higher SP mean that the method cares more about relatively fewer malignant cases rather than the dominant benign samples in our classification experiments.

TABLE III
ISIC 2017 TEST DATASET: CLASSIFICATION RESULTS FROM DIFFERENT LOSS FUNCTIONS (WITH HYPERPARAMETERS IMAGENET, RESNET101, 1×2)

Type	AC	AUC	SE	SP	AP
Original	0.788	0.754	0.991	0.573	0.796
Dice Loss(DL)	0.790	0.765	0.990	0.664	0.787
Focal Loss(FL)	0.801	0.749	0.986	0.730	0.804
Jaccard Loss(JL)	0.793	0.783	**0.991**	0.647	0.815
FL+JL	**0.813**	**0.794**	0.977	**0.731**	**0.823**
FL+JL+DL	0.806	0.790	0.979	0.727	0.818

2) Classification Results with Different Network Hyperparameters: We further investigate the effects of certain network hyperparameters on the classification performance, including different weight initialization, model backbone and batch size in network training. Tab. IV reports the classification results, where all results are obtained by employing the same loss function (the combined focal loss with jaccard loss). These different metrics are complementary and they together reflect the framework performance comprehensively. Ideally, a method is definitely better than other methods if all metrics are better. However, it is often hard to achieve this in practice due to the trade-off between different performance metrics. From Tab. IV, we can see that a larger batch size yields a higher AC (but lower AUC and AP) when the initialization and backbone are the same. For example, if we choose the batch size 2×2 under COCO initialization and ResNet50, AC increases from 0.851 to 0.905, AUC decreases from 0.776 to 0.708 and AP decreases from 0.820 to 0.806. Moreover, the performance seems affected by different weight initializations. For the ISBI 2016 Challenge, AP is used as the main ranking metric. When we initialize the network with COCO, use ResNet101 as the experimental backbone and training in the batch size 1×2, we obtain the highest AP 0.825.

We need to compare the classification performance of the proposed framework with state-of-the-art methods. Since for ISIC 2017 test dataset, we can't make the comparisons directly because of the different classification tasks in our paper and in ISIC 2017. To make a fair comparison, we apply the proposed framework on ISBI 2016 test dataset (the same binary classification task as in our paper) and report the corresponding results. Here we use the COCO initialization, ResNet101, batch size 1×2 as our experimental hyperparameters setting. The performance comparisons with other methods are given in Tab. V, where we show the top five results of skin lesion classification challenge on ISBI 2016. It is noted that the proposed method achieves the highest AC 0.891, SE 0.996, SP 0.723, AP 0.815 among all methods. In terms of AC, SP and AP, there is a huge performance gain that demonstrates the effectiveness of the proposed framework in handling such a class imbalance dataset (malignant:benign=75:304). Generally, our framework provides a comparable performance in classification.

2168-2194 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northeastern University. Downloaded on June 03,2020 at 02:43:04 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2973614, IEEE Journal of Biomedical and Health Informatics

TABLE IV
ISIC 2017 TEST DATASET: CLASSIFICATION RESULTS WITH DIFFERENT NETWORK HYPERPARAMETERS (WITH THE FOCAL LOSS + JAC LOSS)

Initialization	Backbone	Batchsize	AC	AUC	SE	SP	AP
#COCO	ResNet50	1×2	0.851	0.776	0.986	0.802	0.820
#		2×2	0.905	0.708	**0.991**	0.880	0.806
#COCO	ResNet101	1×2	0.838	0.731	0.981	**0.968**	**0.825**
#		2×2	**0.918**	0.696	0.978	0.893	0.804
#IMAGENET	ResNet50	1×2	0.859	0.769	0.988	0.821	0.803
#		2×2	0.905	0.610	0.987	0.915	0.809
#IMAGENET	ResNet101	1×2	0.813	**0.794**	0.977	0.731	0.823
#		2×2	0.868	0.760	0.984	0.880	0.816

TABLE V
ISBI 2016 TEST DATASET: CLASSIFICATION PERFORMANCE COMPARISONS WITH STATE-OF-THE-ART METHODS

Methods	AC	AUC	SE	SP	AP
Ours	**0.891**	0.805	**0.996**	**0.723**	**0.815**
Ours(ONLY ISBI2016)	0.880	0.796	0.994	0.657	0.782
CUMED	0.855	0.783	0.931	0.547	0.624
Jordan Yap	0.844	0.775	0.993	0.240	0.559
BF-TB	0.834	**0.826**	0.961	0.320	0.598
Haebeom Lee	0.821	0.793	0.974	0.200	0.555
GTDL	0.813	0.802	0.872	0.573	0.619

TABLE VI
ISIC 2017 TEST DATASET: SEGMENTATION RESULTS FROM DIFFERENT LOSS FUNCTIONS (WITH HYPERPARAMETERS IMAGENET, RESNET101, 1×2)

Type	JA	DI	AC	SE	SP
Original	0.789	0.868	0.952	0.829	0.983
Dice Loss(DL)	0.794	0.873	0.948	0.830	0.972
Focal Loss(FL)	0.805	0.880	0.950	0.846	0.983
Jaccard Loss(JL)	**0.813**	**0.889**	0.946	**0.861**	0.976
FL+JL	0.799	0.878	**0.959**	0.831	**0.986**
FL+JL+DL	0.796	0.875	0.958	0.833	0.980

E. Segmentation Performance of Proposed Framework

1) Segmentation Results with Different Loss Functions: Similar to classification performance evaluation, we investigate the effects of different loss functions on segmentation. The corresponding results are reported in Tab. VI. When compared with the original loss function (cross entropy), in terms of JA, DI and SE metrics, FL, JL and FL+JL all yield higher values. However, both FL and JL perform worse in terms of AC and SP, while for the FL+JL loss function, all performance metrics are improved. For instance, we note that JA increases from 0.789 to 0.799, DI increases from 0.868 to 0.878, AC increases from 0.952 to 0.959, SE increases from 0.829 to 0.831, and SP increases from 0.983 to 0.986. Experimental results demonstrate that the loss function based on focal loss in RPN and jaccard distance in convolution subnets is beneficial for the studied segmentation task. Therefore, in the following experiments, we employ this loss function as our basic framework configuration.

2) Segmentation Results with Different Network Hyperparameters: We investigate the effects of different network hyperparameters on the segmentation performance of the proposed method. Different weight initializations, backbone and batch size are employed in experiments, and the results are reported in Tab. VII. With the setting (the COCO initialization, ResNet101, batch size 2×2), the highest segmentation performance is obtained, i.e., JA 0.849, DI 0.911 and SE 0.888. Such results surpass many existing methods, as shown in Tab. VIII. From the table, we can see that our proposed framework yields the best results in terms of all studied metrics. In terms of JA and DI (the main ranking metrics for ISIC 2017 challenge), a significant performance gain is noted, exceeding nearly six points than state-of-the-art methods. And for AC and SE, we get approximately two or three points gain, while SP is equivalent to that of the compared best method. We also report the classification results of using only the ISBI 2016 training part in Tab. V and the segmentation results of using only the ISIC 2017 training part in Tab. VIII. And we denote them as Ours(ONLY ISBI2016) and Ours(ONLY ISIC2017) respectively. The results are still better than other state-of-the-art methods, which demonstrates the superior classification and segmentation performances of our proposed multi-task framework. On the other hand, the results trained only by ISBI 2016 or ISIC 2017 dataset are slightly worse than that of the initial training dataset, which shows that our performances might be further improved if we use external samples.

The proposed end-to-end framework takes each tested dermoscopy image as the input and performs the classification, detection and segmentation tasks jointly. Some challenging examples and their corresponding outputs by the proposed framework are shown in Fig. 4, where (c), (d), (g) show obscure boundary and poor illumination, and (e), (f), (h) show the irregular shape of melanoma. In addition, strong hair artifacts are shown in (h). As we can see, the melanomas in irregular shape ((e), (f), (h)) are difficult to segment the boundary precisely, especially under the impact of strong hair artifacts ((h)). The lesion in obscure boundary ((g)) is difficult to detect its position (the detected bounding box doesn't cover the mask accurately). In addition, the malignant skin lesions are hard to classify because of the inadequate samples when compared with the benign lesions. Therefore, the possible future directions may include improving malignant lesion recognition with data augmentation, improving the segmentation of melanomas with irregular shape (especially samples with strong hair artifacts), and improving the detection of lesions with obscure boundary.

2168-2194 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northeastern University. Downloaded on June 03,2020 at 02:43:04 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2973614, IEEE Journal of Biomedical and Health Informatics

TABLE VII
ISIC 2017 TEST DATASET: SEGMENTATION RESULTS WITH DIFFERENT NETWORK HYPERPARAMETERS (WITH THE FOCAL LOSS + JAC LOSS)

Initialization	Backbone	Batchsize	JA	DI	AC	SE	SP
#COCO	ResNet50	1×2	0.814	0.887	0.956	0.850	0.985
#		2×2	0.843	0.906	0.958	0.879	0.984
#COCO	ResNet101	1×2	0.823	0.895	0.955	0.863	0.981
#		2×2	**0.849**	**0.911**	0.956	**0.888**	0.985
#IMAGENET	ResNet50	1×2	0.818	0.894	0.965	0.848	**0.986**
#		2×2	0.836	0.905	**0.968**	0.862	0.985
#IMAGENET	ResNet101	1×2	0.799	0.878	0.959	0.831	0.986
#		2×2	0.828	0.899	0.959	0.862	0.983

TABLE VIII
ISIC 2017 TEST DATASET: SEGMENTATION PERFORMANCE COMPARISONS WITH STATE-OF-THE-ART METHODS.

Methods	JA	DI	AC	SE	SP
Ours	**0.849**	**0.911**	**0.956**	**0.888**	**0.985**
Ours(ONLY ISIC2017)	0.816	0.883	0.950	0.871	0.982
Venkatesh [19]	0.764	0.856	0.936	0.830	0.976
Xiaomeng Li [18]	0.772	0.856	0.936	0.854	0.972
Yuan Y [23]	0.765	0.849	0.934	0.825	0.975
Matt Berseth [24]	0.762	0.847	0.932	0.820	0.978
Bi L [25]	0.760	0.844	0.934	0.802	0.985
Menegola A [27]	0.754	0.839	0.931	0.817	0.970

TABLE IX
ISIC 2017 TEST DATASET: MULTI-TASK ABLATION INVESTIGATION FOR CLASSIFICATION RESULTS (WITH HYPERPARAMETERS IMAGENET, RESNET101, 1×2, THE FOCAL LOSS + JAC LOSS)

Type	AC	AUC	SE	SP	AP
Cls	0.763	0.752	0.968	0.705	0.787
Cls+Seg	0.798	0.786	0.977	0.720	0.815
Cls+Dec+Seg	**0.813**	**0.794**	**0.977**	**0.731**	**0.823**

TABLE X
ISIC 2017 TEST DATASET: MULTI-TASK ABLATION INVESTIGATION FOR SEGMENTATION RESULTS (WITH HYPERPARAMETERS IMAGENET, RESNET101, 1×2, THE FOCAL LOSS + JAC LOSS)

Type	JA	DI	AC	SE	SP
Seg	0.760	0.827	0.930	0.829	0.964
Cls+Seg	0.782	0.850	0.948	0.823	0.975
Cls+Dec+Seg	**0.799**	**0.878**	**0.959**	**0.831**	**0.986**

IV. DISCUSSION

Automatic skin lesion analysis for dermoscopy images is important but challenging due to high visual similarity and various artifacts. Here we present a novel end-to-end multi-task framework for automatic skin lesion analysis, which includes melanoma type diagnosis (benign or malignant), lesion area localization and segmentation. Compared with existing methods, the proposed framework has the following advantages: Firstly, it doesn't need any additional post-processing operations. Secondly, the framework can provide more comprehensive information for melanoma diagnosis since the outputs contain lesion type, position and boundary results. Thirdly, the loss function integrated in the framework can alleviate the severe class imbalance concern in real dermoscopy image datasets.

The proposed framework contains three major components: data preparation, model architecture and results combination. Data preparation makes it possible to take arbitrary size image as the framework input, while keeping the original feature as much as possible. As for model architecture, it can be divided into two stages: The RPN at the first stage extracts the foreground in an image, and then three separate convolution subnets at the second stage are responsible for the corresponding classification, detection and segmentation tasks. In the final combination, the results from three subnets are combined and displayed. In addition, We investigate the effect of the multi-task setting, and report the performances under different single-task and multi-task learning settings. Detailed results are shown in Tab. IX and Tab. X. We compare the single task performance (only training the Classify subnet or the Segment subnet) with the multi-task performance (training the Classify, Detect and Segment subnets together). In order to explore the impact of the detection task, we also report the results without using the Detect subnet in the multi-task (training the Classify and Segment subnets). We denote them as Cls/Seg, Cls+Seg and Cls+Dec+Seg respectively. Both classification and segmentation results in the multi-task setting are better than that in the single task setting. Intuitively different tasks can provide regularization effects to the feature learning and thus improve the performance when compared with the single task. In particular, by comparing the performances of Cls+Seg and Cls+Dec+Seg, we observe that including the detection task in the framework generally helps improve the classification and segmentation performances.

In our framework training, a three-phase joint training strategy is used to help the network learn features more efficiently. We train three convolution subnets for 40 epochs with the initial learning rate 0.001, then train the layers from the RPN and up for 80 epochs with the learning rate 0.001, and finally finetune all layers in the model architecture for 40 epochs with the learning rate 0.0001. Moreover, we propose an effective loss function based on the focal loss and the jaccard distance to alleviate the severe class imbalance concern in real dermoscopy images. We evaluate the proposed framework on the ISBI 2016 and ISIC 2017 datasets. The framework yields a comparable classification accuracy while achieves the highest segmentation performance when compared with several state-of-the-art deep learning based methods.

2168-2194 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northeastern University. Downloaded on June 03,2020 at 02:43:04 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2973614, IEEE Journal of Biomedical and Health Informatics

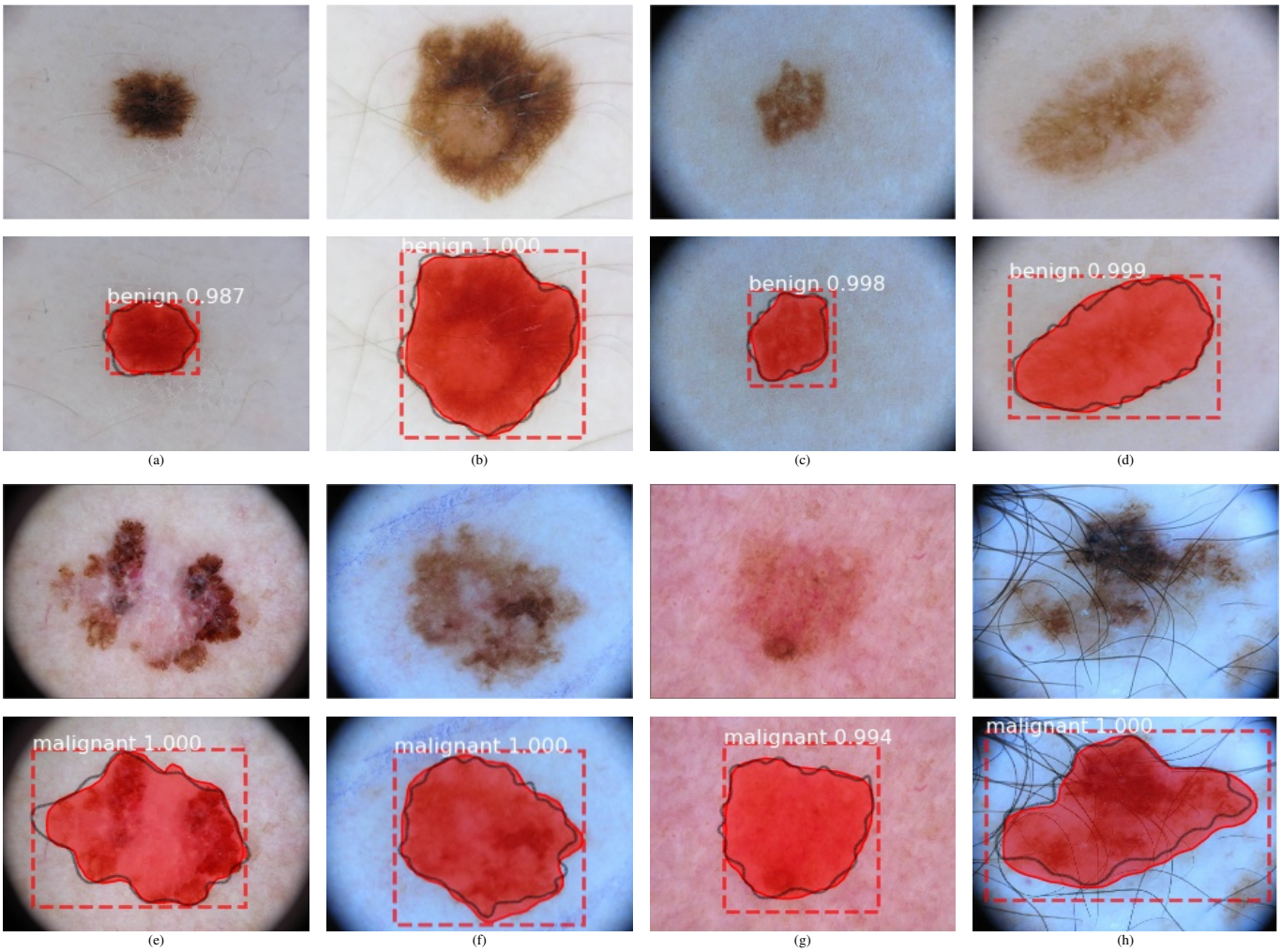

Fig. 4. Some image examples and corresponding outputs from the proposed framework. For each output, it consists of a probability of a certain class, a bounding box covering the skin lesion and the specific segmentation area. Here (a), (b), (c), (d) show prediction results of benign skin lesions, and (e), (f), (g), (h) show prediction results of malignant skin lesions. The red and black contours present the segmentation results and the ground truth, respectively.

V. Conclusion

In this paper, we propose a novel end-to-end multi-task framework that can classify, detect and segment dermoscopy images simultaneously for skin lesion analysis. The framework can take the image with arbitrary size as the input and outputs melanoma type, position and boundary without any additional post-processing operations. We also design an effective loss function based on the focal loss and the jaccard distance to deal with the class imbalance issue (the number disparity between benign and malignant samples) and meanwhile improving segmentation performance. Experimental results on the ISBI 2016 and ISIC 2017 challenge datasets demonstrate the effectiveness and accuracy of the proposed framework. Further investigations such as incorporating attention mechanism, dilated convolution and group normalization into the framework will be explored in the future.

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: a cancer journal for clinicians*, 2019.

[2] N. K. Mishra and M. E. Celebi, "An overview of melanoma detection in dermoscopy images using image processing and machine learning," *arXiv preprint arXiv:1601.07843*, 2016.

[3] R. J. Stanley, W. V. Stoecker, and R. H. Moss, "A relative color approach to color discrimination for malignant melanoma detection in dermoscopy images," *Skin Research and Technology*, vol. 13, no. 1, pp. 62–72, 2007.

[4] Y. Cheng, R. Swamisai, S. E. Umbaugh, R. H. Moss, W. V. Stoecker, S. Teegala, and S. K. Srinivasan, "Skin lesion classification using relative color features," *Skin Research and Technology*, vol. 14, no. 1, pp. 53–64, 2008.

[5] L. Ballerini, R. B. Fisher, B. Aldridge, and J. Rees, "A color and texture based hierarchical k-nn approach to the classification of non-melanoma skin lesions," in *Color Medical Image Analysis*. Springer, 2013, pp. 63–86.

[6] 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, no. 6, pp. 362–373, 2007.

[7] G. Schaefer, B. Krawczyk, M. E. Celebi, and H. Iyatomi, "An ensemble classification approach for melanoma diagnosis," *Memetic Computing*, vol. 6, no. 4, pp. 233–240, 2014.

[8] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[9] 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.

2168-2194 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northeastern University. Downloaded on June 03,2020 at 02:43:04 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2973614, IEEE Journal of Biomedical and Health Informatics

[10] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *CVPR*, vol. 1, no. 2, 2017, p. 4.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 6, pp. 1137–1149, 2017.

[12] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[14] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Computer Vision (ICCV), 2017 IEEE International Conference on*. IEEE, 2017, pp. 2980–2988.

[15] 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.

[16] 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, p. 115, 2017.

[17] L. Yu, 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, no. 4, pp. 994–1004, 2017.

[18] X. Li, L. Yu, C.-W. Fu, and P.-A. Heng, "Deeply supervised rotation equivariant network for lesion segmentation in dermoscopy images," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Springer, 2018, pp. 235–243.

[19] G. Venkatesh, Y. Naresh, S. Little, and N. E. OConnor, "A deep residual architecture for skin lesion segmentation," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Springer, 2018, pp. 277–284.

[20] A. R. Lopez, X. Giro-i Nieto, J. Burdick, and O. Marques, "Skin lesion classification from dermoscopic images using deep learning techniques," in *Biomedical Engineering (BioMed), 2017 13th IASTED International Conference on*. IEEE, 2017, pp. 49–54.

[21] C. N. Vasconcelos and B. N. Vasconcelos, "Increasing deep learning melanoma classification by classical and expert knowledge based image transforms," *CoRR, abs/1702.07025*, vol. 1, 2017.

[22] Z. Yu, D. Ni, S. Chen, J. Qin, S. Li, T. Wang, and B. Lei, "Hybrid dermoscopy image classification framework based on deep convolutional neural network and fisher vector," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. IEEE, 2017, pp. 301–304.

[23] Y. Yuan and Y.-C. Lo, "Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks," *IEEE Journal of Biomedical and Health Informatics*, 2017.

[24] M. Berseth, "Isic 2017-skin lesion analysis towards melanoma detection," *arXiv preprint arXiv:1703.00523*, 2017.

[25] 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.

[26] S. Vesal, S. M. Patil, N. Ravikumar, and A. K. Maier, "A multi-task framework for skin lesion detection and segmentation," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Springer, 2018, pp. 285–293.

[27] A. Menegola, J. Tavares, M. Fornaciali, L. T. Li, S. Avila, and E. Valle, "Recod titans at isic challenge 2017," *arXiv preprint arXiv:1703.04819*, 2017.

[28] X. Yang, Z. Zeng, S. Y. Yeo, C. Tan, H. L. Tey, and Y. Su, "A novel multi-task deep learning model for skin lesion segmentation and classification," *arXiv preprint arXiv:1703.01025*, 2017.

[29] S. Chen, Z. Wang, J. Shi, B. Liu, and N. Yu, "A multi-task framework with feature passing module for skin lesion classification and segmentation," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 1126–1129.

[30] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE transactions on pattern analysis and machine intelligence*, 2018.

[31] 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.

[32] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 424–432.

[33] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in neural information processing systems*, 2012, pp. 2843–2851.

[34] G. Papandreou, I. Kokkinos, and P.-A. Savalle, "Modeling local and global deformations in deep learning: Epitomic convolution, multiple instance learning, and sliding window detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 390–399.

[35] Y. Le, L. Bottou, G. Orr, and K. Muller, "Lecun, y. efficient backprop in neural networks: Tricks of the trade," 1998.

[36] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[37] N. Bjorck, C. P. Gomes, B. Selman, and K. Q. Weinberger, "Understanding batch normalization," in *Advances in Neural Information Processing Systems*, 2018, pp. 7694–7705.

[38] X. Chen, H. Fang, T.-Y. Lin, R. Vedantam, S. Gupta, P. Dollár, and C. L. Zitnick, "Microsoft coco captions: Data collection and evaluation server," *arXiv preprint arXiv:1504.00325*, 2015.

2168-2194 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northeastern University. Downloaded on June 03,2020 at 02:43:04 UTC from IEEE Xplore. Restrictions apply.