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/ACCESS.2020.2997710, IEEE Access

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Automatic Skin Cancer Detection in Dermoscopy Images based on Ensemble Lightweight Deep Learning Network

LISHENG WEI[1], KUN DING[2], HUOSHENG HU[3], Member, IEEE

[1]Anhui Key Laboratory of Electric Drive and Control, Anhui Polytechnic University, Wuhu 241000, CN.
[2] Anhui Key Laboratory of Electric Drive and Control, Department of Electrical Engineering, Anhui Polytechnic University, Wuhu 241000, CN.
[3]School of Computer Science and Electric Engineering, University of Essex, Colchester CO4 3SQ, U.K.

Corresponding author: KUN DING (e-mail: 58475885@163.com).

This work was supported by the Natural Science Foundation of Anhui Province (1908085MF215), the Middle-aged and young talents project of Anhui University of Engineering (2016BJRC008).

ABSTRACT The complex detection background and lesion features make the automatic detection of dermoscopy image lesions face many challenges. The previous solutions mainly focus on using larger and more complex models to improve the accuracy of detection, there is a lack of research on significant intra-class differences and inter-class similarity of lesion features. At the same time, the larger model size also brings challenges to further algorithm application; In this paper, we proposed a lightweight skin cancer recognition model with feature discrimination based on fine-grained classification principle. The propose model includes two common feature extraction modules of lesion classification network and a feature discrimination network. Firstly, two sets of training samples (positive and negative sample pairs) are input into the feature extraction module (Lightweight CNN) of the recognition model. Then, two sets of feature vectors output from the feature extraction module are used to train the two classification networks and feature discrimination networks of the recognition model at the same time, and the model fusion strategy is applied to further improve the performance of the model, the proposed recognition method can extract more discriminative lesion features and improve the recognition performance of the model in a small amount of model parameters; In addition, based on the feature extraction module of the proposed recognition model, U-Net architecture, and migration training strategy, we build a lightweight semantic segmentation model of lesion area of dermoscopy image, which can achieve high precision lesion area segmentation end-to-end without complicated image preprocessing operation; The performance of our approach was appraised through widespread experiments comparative and feature visualization analysis, the outcome indicates that the proposed method has better performance than the start-of-the-art deep learning-based approach on the ISBI 2016 skin lesion analysis towards melanoma detection challenge dataset.

INDEX TERMS Dermoscopy Images, Skin cancer detection, Lightweight deep learning network, Fine-grained feature.

I. INTRODUCTION

Skin cancer is one of with high mortality forms of cancer, according to cancer statistics released by the American Cancer Society, the mortality rate of patients with skin cancer is as high as 75% [1], [2], and the melanoma with the highest mortality rate is still increasing with an incidence of 14%. Fortunately, if the disease can be found and treated in time in the early stage, the probability of survival is very high [3], [4].

As a non-trauma skin imaging technique, dermoscopy is widespread used in the identification of melanoma. Although the accuracy of using dermoscopy to detect melanoma is higher than that without auxiliary observation [5], however, the diagnostic accuracy depends on the experience and professional skills of dermatologists. Even if dermatologists make the diagnosis, the accuracy of melanoma diagnosis can only reach 75-84% [6], and the diagnosis results of different doctors are different and have poor repeatability. Therefore, using the advantages of artificial intelligence to assist doctors in non-contact automatic diagnosis is of great practical significance.

Using dermoscopy images to automatically identify melanoma is a very challenging task, there are many interference factors in the dermoscopy image, such as hair on

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/ACCESS.2020.2997710, IEEE Access

the skin surface, related solutions used to enhance the clarity of skin lesions, and different-colored discs used for auxiliary identification, as follows Fig. 1.

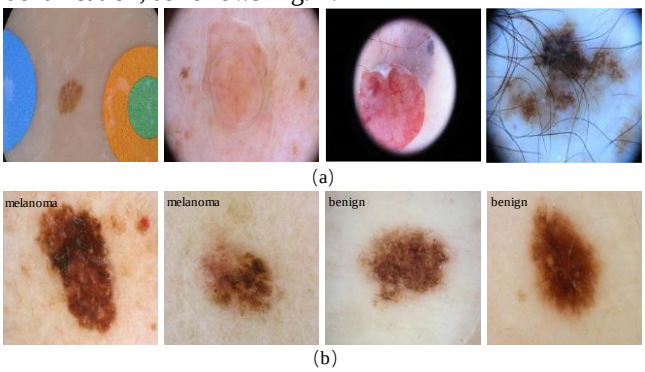

FIGURE 1. Automatic detection of melanoma is interfered by many factors. (a) shows some interferences unrelated to pathological factors. (b) shows melanoma and benign are not easy to distinguish in appearance.

II. RELATED WORK

To improve the accuracy of automatic detection, many related experts and scholars have conducted extensive research. Early automatic detection of lesions in dermoscopy images is usually based on hand-designed low-level features, combined with well-designed classifiers for training, and ultimately achieves the purpose of recognition. These features include color [7-9], shape [10], and texture [11]. Some researchers use feature fusion strategies to combine two or more features to elevated the robustness of the model [12-15]. However, the low-level feature expression ability based on the manual design is insufficient, which can't effectively deal with the problems of the large with-class feature difference and small between-class feature difference, moreover, the patterns concerned by these features are relatively fixed, and the generalization ability of the model is weak. Other researchers have proposed a method to segmentation the lesion area first and then base on the segmentation result to recognize melanoma [16-17]. These methods use image segmentation to extract image areas containing only lesions, so the extracted features are more representative, but due to the limitations of the low-level features, the final disease recognition rate is slightly improved, in addition, whether the type of lesion is only related to the lesion area unable be verified. Other researchers use the bag of feature (BOF) as the lesion feature for identification [18], but the pattern of melanoma and non-melanoma lesion features is complex and changeable, and an effective visual dictionary cannot be effectively constructed. Therefore, the robustness of the disease classification model based on BOF is weak.

Convolution neural networks because of its robust feature representation capability, has been extensive utilization in medical image analysis in the last several years and has achieved remarkable results. These applications include medical image segmentation [19-21], classification [22-24] and detection [25], [26]. Kawahara et al. designed a full convolution neural network based on AlexNet [27] model to recognize the melanoma [28]. Ge et al. Extracted high-level features from ResNet [29] and VGG [30] networks for bilinear merging, and trained using SVM classifiers achieved the best current recognition results on multiple test sets [31]. Esteva1 et al. Combined data-driven technology to train nearly 13W dermatological pictures on the InceptionV3 [32], and achieved excellent results comparable to professional dermatologists on the test set [33]. However, these methods do not consider the problem of poor discrimination between various skin cancer features. Yu et al. proposed a phased melanoma recognition method based on the deep residual network [34], through Melanoma recognition based on skin lesion segmentation they won the first place in the ISBI 2016 *Skin Lesion Analysis Towards Melanoma Detection* [35] (hereinafter referred to as ISBI 2016) classifier task, because of the final recognition needs to be carried out step by step, it is not an end-to-end solution. Combined with the adaptive sample learning strategy, Guo et al. Designed a multi convolution neural network to copy with the intra-class discrepancy of melanoma and related noise interference [36]. To extract more distinguishable pathological features. Yu et al. combining pre-trained model weights to encode the output features of the deep residual network into Fisher vectors, and trained SVM to reach the purpose of recognition, through further the integration strategy achieved the optimum performance in the ISBI 2016 classification test set [37]. Similarly, this method is not an end-to-end solution and the model construction is complex. Zhang et al. Designed a multi-CNN collaborative training dermoscopy image lesion recognition model, improves the robustness of lesion identification and verified the effectiveness of the proposed method on related data sets [38]. In order for the model to learn more powerful and more distinguishing feature representation capabilities, Zheng et al. Proposed a framework for automatic skin lesion recognition using cross-net based aggregation of multiple convolutional networks, and verified the proposed method through extensive experiments Superiority [39].

To make the model better segmentation the lesion areas of different scales in the original image, Li et al. designed a semantic segmentation model based on multi-scale full convolution to extract the lesion areas of skin disease [40]. To improve the segmentation precision of skin lesion boundary. Deng et al. based on VGG-16 and hole convolution, design a fully convolutional neural network that can simultaneously extract global features and local features [41]. Li et al. By constructing the residual network with different scale input and the calculation unit of the lesion index, the rough segmentation of lesion degree in the area of skin injury are realized [42]. Tang et al. designed a multi-stage semantic segmentation model combined with context information to achieve the end-to-end accurate segmentation of skin lesion [43]. In order to improve the robustness and accuracy of lesion boundary segmentation, Xie. et al. Designed a multi-branch fusion network with mixed feature

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/ACCESS.2020.2997710, IEEE Access

inputs, and verified the effectiveness of the proposed method through extensive comparison experiments [44]. Pour M P et al. Proposed a method with CIELAB color space and transform Domain dermoscopy image lesion segmentation network achieves high segmentation performance without using additional data and preprocessing techniques [45].

III. METHODOLOGY

In this paper, we proposed an efficient and lightweight melanoma classification network based on MobileNet [46], DenseNet [47]. Different from the previous solutions, we introduced the fine-grained classification principle in the lightweight melanoma classification network to improving the feature discrimination ability, recognition accuracy of lightweight networks and keep a small number of model parameters, meanwhile, we use focal loss [48] method for comparison experiments. Besides, we design a lightweight U-Net [49] model based on the feature extraction module of the classification network to accurately segmentation skin lesion area, our method can achieve high segmentation accuracy without complicated image preprocessing technology while ensuring the small number of model parameters. In the end, compared with the method of getting the start-of-the-art result on ISBI 2016 test set, the proposed method obtains better performance and verifies its effectiveness.

Below is a summary of the main contributions of our work:

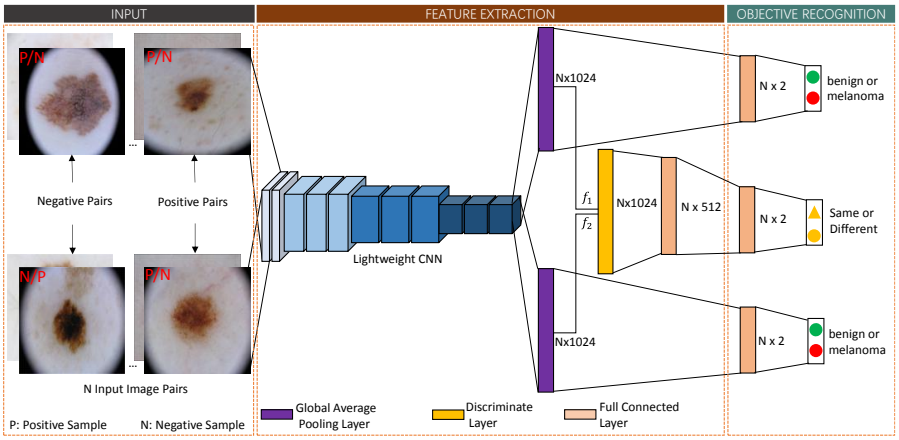

FIGURE 2. The proposed method framework for melanoma recognition. Given N pairs of 224 × 224 images as the feature extraction module input, the lightweight CNN will output two different 1024 dimensional feature vectors f_1 and f_2, then f_1 and f_2 will be used as lesion features in the training of melanoma recognition network, and through introducing a non-parametric discriminant layer we build a network which can verify whether the images corresponding to f_1 and f_2 belong to the same category.

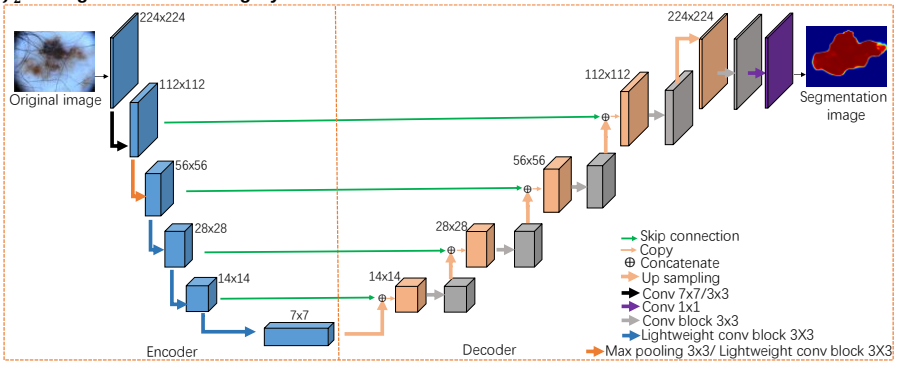

FIGURE 3. The proposed method framework for skin lesion segmentation. The proposed framework of full convolution skin lesion area semantic segmentation based on U-Net, which includes three parts: Encoder, Skip connection and Decoder. In the Encoder part, we used the feature extraction modules of the MobileNetV1 or DenseNet-121 networks to build. In the Decoder part, we use the same structure, which includes up sampling(stride=2), skip connection and 3 x 3 convolution, the skip connection uses the way of channel concatenate.

1) CLASSIFICATION TASK

The proposed dermoscopy image lesion recognition method includes three steps: image preprocessing, model construction and model training, and model fusion. Image preprocessing involves training set image augmentation and construction positive and negative sample pair training sets. It is mainly used to alleviate the overfitting of the model and make the input data format meet the model training input requirements; model construction involves lightweight recognition network and feature discrimination network construction, loading of pre-training weights, joint training of lightweight recognition networks and feature discrimination networks, It is mainly used to improve the model's feature discrimination ability, recognition performance, and reduce the number of model parameters; model fusion includes the extraction of different lightweight recognition networks that have been trained and fusion, It is mainly used to further improve the overall performance of the model. The related flowchart is shown in the figure below:

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/ACCESS.2020.2997710, IEEE Access

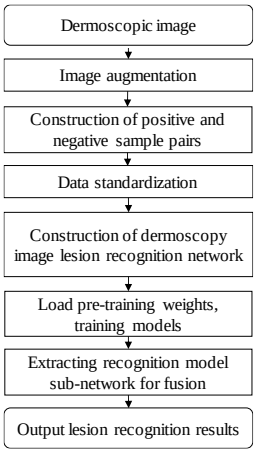

FIGURE 4. Flow chart of dermoscopy image lesion recognition method

The proposed recognition model uses two different features of the output of the lightweight CNN (see Table I and Table II for specific structure) as the input of the feature discrimination networks to determine whether the two input images belong to the same type, so as to enhance the model ability to distinguish similar features between the melanoma and non-melanoma. Compared with the original lightweight model, the model combined with the principle of fine-grained classification can extract more discriminant features and achieve higher accuracy. We call the proposed model DC-MobileNetV1 (Discriminant MobileNetV1), DC-DenseNet121(Discriminant DenseNet-121).

TABEL I
LIGHTWEIGHT CNN ARCHITECTURE BASE ON MOBILENETV1

Operation	Parameters	Input Size
Convolution	3×3 conv, s=2	224×224
Depthwise Separable Block1	$\begin{bmatrix} 3\times3\ dw\ conv \\ 1\times1\ conv \end{bmatrix}$, s=1 $\begin{bmatrix} 3\times3\ dw\ conv \\ 1\times1\ conv \end{bmatrix}$, s=2	112×112
Depthwise Separable Block2	$\begin{bmatrix} 3\times3\ dw\ conv \\ 1\times1\ conv \end{bmatrix}$, s=1 $\begin{bmatrix} 3\times3\ dw\ conv \\ 1\times1\ conv \end{bmatrix}$, s=2	56×56
Depthwise Separable Block3	$\begin{bmatrix} 3\times3\ dw\ conv \\ 1\times1\ conv \end{bmatrix}$, s=1 $\begin{bmatrix} 3\times3\ dw\ conv \\ 1\times1\ conv \end{bmatrix}$, s=2	28×28
Depthwise Separable Block4	$\begin{bmatrix} 3\times3\ dw\ conv \\ 1\times1\ conv \end{bmatrix} \times 5$, s=1 $\begin{bmatrix} 3\times3\ dw\ conv \\ 1\times1\ conv \end{bmatrix}$, s=2	14×14
Depthwise Separable Block5	$\begin{bmatrix} 3\times3\ dw\ conv \\ 1\times1\ conv \end{bmatrix}$, s=1	7×7

TABEL II
LIGHTWEIGHT CNN ARCHITECTURE BASE ON DENSENET-121

Operation	Parameters	Input Size
Convolution	7x7 conv, s=2	224×224
pooling	3x3 max pool, s=2	112×112
Dense Block1	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 6$, s=1	56×56
Transition Layer1	$1 \times 1\ conv$	56×56
	2x2 average pool, s=2	56×56
Dense Block2	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 12$, s=1	28×28
Transition Layer2	$1 \times 1\ conv$	28×28
	2x2 average pool, s=2	28×28
Dense Block3	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 24$, s=1	14×14
Transition Layer3	$1 \times 1\ conv$	14×14
	2x2 average pool, s=2	14×14
Dense Block4	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 16$, s=1	7×7

2) SEGMENTATION TASK

The proposed dermoscopy image lesion area segmentation method includes three steps: image preprocessing, model construction and model training, and model fusion. Image preprocessing is mainly to augmentation the original training set to alleviate the overfitting of the model; model construction and model training mainly includes build a lightweight encoder, standard decoder, pre-training weight loading under the U-Net architecture, and model training based on mixed loss function, it is mainly used to reduce the amount of parameters of the model while maintaining a high model segmentation accuracy, the related flowchart is shown in the Fig. 5. We call the proposed model U-MobileNetV1(base on MobileNetV1), U-DenseNet121(base on DenseNet-121).

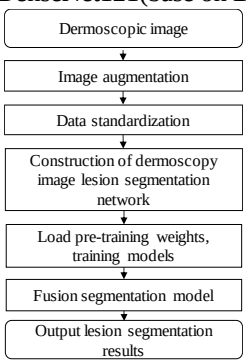

FIGURE 5. Flow chart of dermoscopy image lesion area segmentation method.

The proposed segmentation model is different from the original U-Net. We used the lightweight CNN (see Table I, II) to replace the encoder part of the original U-Net network. For U-MobileNetV1, the lightweight conv block 3×3 of encoder corresponds to the Depthwise Separable Block. For U-DenseNet121, the lightweight conv block 3×3 of encoder includes a Dense Block and a Transition Layer, see Fig. 6 for specific structure.

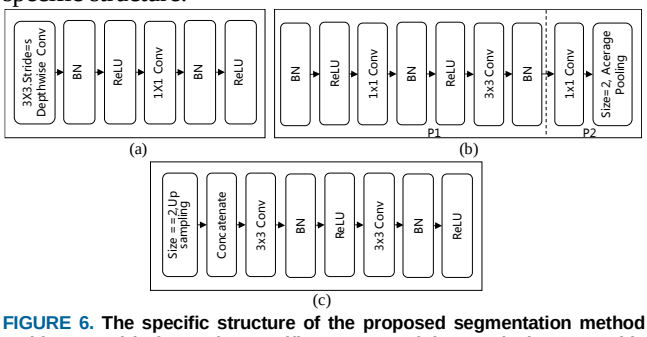

FIGURE 6. The specific structure of the proposed segmentation method architecture. (a) shows the specific structure of the Depthwise Separable block. (b)-P1, P2 respectively show the specific structure of the Dense Block and Transition Layer. (c) shows the main structure of the proposed segmentation architecture decoder.

The rest of this article is arranged as follows. In this Section, we introduce the implementation details of the proposed method, and in Section Ⅳ, we carry out extensive experimental verification and comparison. Finally, we discussed and concluded the proposed method in Section V and VI, respectively.

A. Basis of Lightweight Convolutional Neural Network

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/ACCESS.2020.2997710, IEEE Access

At present, some methods based on CNN have made positive progress in the task of automatic detection of melanoma [33-41]. Meanwhile, some researchers have focused on the automatic detection task of melanoma on lightweight deep learning models [50-52], which significantly reduced the number of model parameters and achieved better detection results. Inspired by these works, we redesigned the lightweight deep learning model so that it can adapt to the current task and improve the ability of the network to discriminate similar features by introducing feature discrimination layers. In addition, to solve the problem that the semantic segmentation model has a large number of parameters and is difficult to deploy to the mobile end, we proposed a full convolution semantic segmentation network based on a lightweight model and U-Net structure to achieve end-to-end lesion segmentation. In this part, we will briefly introduce the basis of the lightweight model related to this article.

The lightweight model is usually composed of a series of basic modules and each module is composed of a small number of specific network layers stacked. These network layers contain operations on spatial dimension and channel dimension. These operations include depthwise separable convolution, group convolution, channel separation, channel scrambling, residual connection, and channel concatenate. The principle of standard convolution and depthwise separable convolution is shown in the following figure:

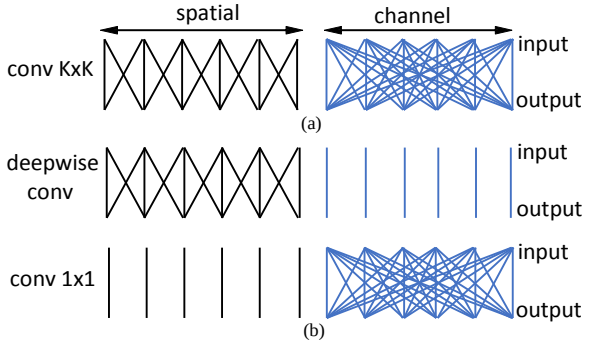

FIGURE 7. Schematic diagram of standard convolution (a) and depthwise separable convolution (b).

Taking the standard convolution shown in Fig. 7(a) as an example, assuming that the input feature map size is $H \times W \times N$, the output feature map size is $H \times W \times M$, and the convolution kernel size is $K \times K$, then the parameters of a standard convolution layer are $K^2 \times N \times M$, and the parameters of a depthwise separable convolution layer are $K^2 \times N + 1^2 \times N \times M$. It can be found that the depthwise separable convolution is only $1/M + 1/K^2$ of the standard convolution parameters. In the MobileNetV1, the use of deep separable convolutions instead of standard convolutions has significantly reduced the size of the model. In DenseNet-121, the parameters of the model are mainly reduced by using 1×1 convolution in the dense connected blocks and transition blocks of the model. The dense connected blocks are implemented as follows:

$$x_l = H_l([x_0 + x_1, \cdots, x_{l-1}]) \quad (1)$$

where x_l represents the feature map of the layer l in the network, H_l represents a combination function, which mainly includes operations such as BN, $ReLU$, $conv\ 1 \times 1$, and conv 3×3 (see Fig. 6 (b)-P1) for the specific structure).

B. Classification of melanoma

1) DATA AUGMENTATION

Considering that the ratio of melanoma to non-melanoma in the ISBI 2016 dermoscopy image dataset is 1:4, there is a significant category imbalance problem, and the amount of data is small, so it is necessary to perform data augment processing. We applied rotation (90,180,270), mirroring, center cropping, brightness change, random occlusion operations on the original image to reduce the possible over fitting of the model and enhance the robustness of the model. Finally, through the data enhance, we increased the original training set images from 900 to 4430 and separated the enhanced dataset into a training set and a validation set according to a ratio of 0.8: 0.2 for use in training the model.

2) DATA PREPROCESSING

Since the melanoma classification network needs to be provided with positive and negative sample pairs as input when training it, therefore the input data needs additional processing before network training. Note that to make the model better distinguish data with different categories but similar feature, we used two different data processing methods alternately during training, as follows:

Method1 constructs input data that contains more positive sample pairs: In each round of network training, the input comes from a group of randomly shuffled batch data. The batch data contains two parts, the input image x_1 and the correspond label y_1, where the shape of x_1 is $N \times W \times H \times C$(Set to $32 \times 224 \times 224 \times 3$), and the shape of y_1 is $N \times 2$ (one-hot coding). Let $y_{label} = max\_id(y_1)$, where max_id indicates the maximum value position index of the element in y_1, the y_{label} shape is $N \times 1$, which is composed of 0/1, there, 0 is regarded as non-melanoma and 1 is regarded melanoma. Then construct a variable id_{x_2} with the same dimension as y_{label}. By traversing the y_{label}, find in turn the data pairs with the same elements and the data pairs with different elements, simultaneously, exchange the corresponding position index and assign it to id_{x_2}. Finally, construct variables x_2, y_2 with the same dimensions as x_1 and y_1, and use the value of the element in id_{x_2} as an index to traverse x_1, y_1, find the elements corresponding to the index positions of x_1 and y_1, and assign these elements to x_2 and y_2 in turn, thus, two sets of input data x_1, y_1, x_2 and y_2 containing more positive sample pairs are constructed.

Method2 constructs input data that contains more negative sample pairs: Since the original input data is randomly shuffled before each training, therefore, we only need to read the batch twice before training, and by the corresponding values of the two batches are assigned to x_1, y_1, x_2 and y_2 in

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/ACCESS.2020.2997710, IEEE Access

turn. Thus, two sets of inputs data x_1, y_1, x_2 and $y_2$containing more negative sample pairs are constructed.

Note that during each iteration of training, Method1 and Method2 are used at a frequency of 1: 4 in this article, this is to make the model pay more attention to the feature difference between melanoma and benign.

3) CLASSIFICATION NETWORK ARCHITECTURE

For the proposed recognition model, we use MobileNetV1 or DenseNet-121 feature extraction module as the component of lightweight CNN, then the training data containing positive and negative sample pairs are input into the lightweight CNN respectively to get two different feature outputs, after the two outputs, two global average pooling layer and two fully connected layer of size 2 (using the Softmax activation function) were added to form two different classification network branches, respectively. By measuring the similarity of the outputs feature of the two global average pooling layers in the discriminate layers, and then apply the ReLU activation function fully connected layer of size 512 and an apply the Softmax activation function fully connected layer of size 2 were added. Finally, a similarity discriminate network is constructed to determine whether two groups of input images belong to the same type, therefore, the proposed classification network architecture includes three branch networks. The feature similarity measurement function used by the discriminate layer is as follows:

$$D(Y_1, Y_2) = (Y_1 - Y_2)^2 \tag{2}$$

where Y_1、Y_2 represent the output features of the two global average pooling layer, and the subtraction and square operation in the formula is carried out element by element, so the output size of the discriminate layer is consistent with the input size.

4) TRAINING PROCEDURE

To improve the efficiency of model training, we load pre-training weights (trained on Imagenet [53]) in the feature extraction part of the classification branch network(lightweight CNN), and take 0.0001 as the initial learning rate for Adam optimizer to training total network layers, the total number of iterations of network training is set to 50, and during the training, if the loss from the verification set does not reduce the specified value in three consecutive iterations, then the learning rate decays to 1/2 of the current value. Note that the two classification networks use cross-entropy loss function, and we also use the focal loss as a comparison experiment, the cross-entropy loss function is used to the feature discriminate network, and the relevant equation is as follows:

$$L = \begin{cases} -y\ln p & , y = 1 \\ -(1-y)\ln(1-p) & , y = 0 \end{cases} \tag{3}$$

$$FL = \begin{cases} -\alpha(1-p)^{\gamma} y\ln p & , y = 1 \\ -(1-\alpha)p^{\gamma}(1-y)\,ln(1-p) & , y = 0 \end{cases} \tag{4}$$

$$L_{total} = \begin{cases} AL + BL + CL, & use\ L \\ AFL + BFL + CL, & use\ FL \end{cases} \tag{5}$$

where L represent the cross-entropy loss function, y means the real category label, and p represent the prediction category label probability value. FL means the focal loss function, α is the factor that to balance the contribution of negative and positive samples to the loss function value, γ is the factor used to balance the contribution of hard and easy samples to the loss function value, the values of α, γ are set to 0.25 and 0.75. L_{total} is the total loss of the training network, A, B represents the loss function weight value of the two classification network in the total loss function, C represent the loss function weight value of the feature discriminate network in the total loss function, the values of A, B, and C are set to 1, 1, and 0.5 respectively.

During the test, we extracted the model with the better model performance from the two trained classification networks and then perform model performance evaluation on the test set after further model fusion.

C. Segmentation of skin lesion area

1) DATA PREPROCESSING

The data enhancement method used in the segmentation network is the same as that used in the classification network, the difference is that the original image is enhanced and the corresponding segmentation mask is enhanced in the same way, in addition, because the whole segmentation network only needs a group of inputs, no additional model data input preprocessing operation is required. Finally, through the data augment technology, we increased the original training set images from 900 to 6900, and separated the enhanced dataset into a training set and a validation set according to a ratio of 0.8: 0.2 for use in training the model.

2) SEGMENTATION NETWORK IMPLEMENTATION

For the proposed segmentation model, the decoder structure part of based on the MobileNetV1 feature extraction module is shown in TABEL I. When the 3×3 depthwise separable convolution stride is 1, this means that the size of the input feature map is the same as the output. When the stride is 2, it means that the output feature map is 1/2 of the input feature map size. In the encoder module, the skip connection with the decoder module is the output of the Convolution 3×3, Depthwise Separable Block2, 3, and Depthwise Separable Block4(s=1); The decoder structure part of based on the DenseNet-121 feature extraction module is shown in TABEL II. In the encoder, the skip connection with the decoder module is the output of the, Convolution 7×7, Dense Block1, Dense Block2, Dense Block3. It should be noted that the last lightweight conv block 3 × 3 in the U-MobileNetV1 encoder contains Depthwise Separable Block4 and Depthwise Separable Block5, while the last lightweight conv block 3 × 3 in the U-DenseNet121 encoder contains Transition Layer3, and Dense Block4. The overall model structure can be seen in Fig. 3.

3) TRAINING PROCEDURE

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/ACCESS.2020.2997710, IEEE Access

Before model training, we load pre-training weights (trained on Imagenet [53]) into the encoder part to make the model more efficient for training, and take 0.0001 as the initial learning rate for Adam optimizer to training total network layers, if the loss from the verification set does not reduce the specified value in three consecutive iterations, then the learning rate decays to 1/2 of the current value. The model training final executes 50 epochs, and the model loss function uses BCE Dice Loss. The specific formula is as follows:

$$S1 = \sum_{i=1}^{w}\sum_{j=1}^{h} P_{ij}Y_{ij} \quad (6)$$

$$S2 = \sum_{i=1}^{w}\sum_{j=1}^{h} Y_{ij} \quad (7)$$

$$S3 = \sum_{i=1}^{w}\sum_{j=1}^{h} P_{ij} \quad (8)$$

$$\mathrm{L_{Cross}} = -\frac{1}{\mathrm{w\times h}}\sum_{i=1}^{\mathrm{w}}\sum_{j=1}^{\mathrm{h}}\left[Y_{ij}\ln P_{ij} + \left(1-Y_{ij}\right)\ln\left(1-P_{ij}\right)\right] \quad (9)$$

$$L_{Dice} = 1 - (2 * S1 + \epsilon)/(S2 + S3 + \epsilon) \quad (10)$$

$$L_{BCE\ Dice} = L_{Dice} + L_{Cross} \quad (11)$$

where P_{ij} and Y_{ij} represent the output probability map of the segmentation network and the real segmentation map, respectively, both of which are represented by the matrix. $P_{ij}Y_{ij}$ represents the pixel-wise multiplication operation of the matrix elements, w, h is the width and height of training images, ϵ set to 1(avoid division zero).

D. Data set and System implementation

We validate our put forward approach on the ISBI 2016 challenge dataset, which is from the International Skin Imaging Association (ISIC) archives [1]. This is the most comprehensive collection of quality controlled dermatoscopy image databases on skin lesions. The ISBI 2016 challenge dataset contains 900 training set and 379 test set (including the original pictures and the corresponding lesion segmentation masks marked by professional doctors). Data set are divided into melanoma and benign. where nearly 80% of the data set are benign (test set contains 304, training set contains 727). In this section, we will introduce the challenge results provided by the organizers and the state-of-the-art results. In addition, the put forward approach is based on the Keras framework [2] and implemented under the NVIDIA Tesla K80 GPU (12G).

E. Evaluation indicators

We used the evaluation indicators specified by the official website of the challenge to evaluate our model performance. For the segmentation task, the evaluation indicators include accuracy (AC), jaccard index (JA), dice coefficient (DC), specificity (SP), and sensitivity (SE), these evaluation indicators are evaluated at the pixel level, and the final ranking is based on the JA score(the higher the better) on the test set. The relevant definitions are as follows:

$$AC = (S_{tp} + S_{tn})/(S_{tp} + S_{tn} + S_{fp} + S_{fn}),$$
$$SE = S_{tp}/(S_{tp} + S_{fn}),$$
$$SP = S_{tn}/(S_{tn} + S_{fp}),$$
$$JA = S_{tp}/(S_{tp} + S_{fp} + S_{fn}),$$
$$DI = 2S_{tp}/(2S_{tp} + S_{fp} + S_{fn}) \quad (12)$$

where S_{tp} indicates the number of true positive pixels in the segmentation result, S_{tp} indicates he number of true positive pixels, S_{fn} indicate the number of false negative pixels and S_{fp} indicate the number of false positive pixels.

For the classification task, the final ranking is based on the AP score and the AUC, AC score (the higher the better) is used as the reference. The detailed definition can be found in [35].

IV. EXPERIMENTS AND RESULTS

A. The Performance of our method in classification tasks

1) LIGHTWEIGHT CNN SELECTION IN THE PROPOSED METHOD

As a representative early lightweight CNN, SqueezNet [54] uses fire module to achieve compression of nearly 50X of AlexNet model parameters while maintaining approximate accuracy. After this, lightweight networks such as MobileNet and ShuffleNet [55], DenseNet were successively proposed. Under the same experimental conditions, we tested the performance of these lightweight networks and some other networks with small parameters and excellent performance on the data set used in this article, related test results are shown in Table III. Compared with SqueezeNet, ShuffleNet has more parameters and more complicated models, but it does not perform as well as SqueezNet in the overall evaluation index. Similarly, ResNet-18 has more parameters than MobileNetV1 and DenseNet-121, but only AC index is higher than MobileNetV1 0.2%, other indicators are lower than MobileNetV1 and DenseNet-121. Taken together, MobileNetV1 and DenseNet-121 are more prominent and better in various evaluation indicators, and are more suitable as a lightweight feature extractor for dermoscopy image lesion detection model.

TABLE III
PERFORMANCE COMPARISON BETWEEN DIFFERENT LIGHTWEIGHT CNN

Method	AC	AUC	AP	Parameters
SqueezeNet[54]	0.789	0.757	0.494	1,237,498
ShuffleNet[55]	0.760	0.717	0.436	3,616,946
ResNet18[29]	0.831	0.802	0.596	11,187,915
MobileNetV1[46]	0.829	**0.830**	**0.645**	3,230,914
DenseNet-121[47]	**0.837**	0.800	0.636	7,039,554

Table IV shows the effect of data augmentation processing on lightweight CNN performance. It can be found that after data augmentation, MobileNetV1 increases 1.1% and 2.7% on AUC and AP indicators, and DenseNet-121 increases 0.1% on AC and AUV indicators, respectively. 2.6%. Overall, data augmentation can improve the performance of the model to

[1] https://isic-archive.com
[2] https://keras.io/

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/ACCESS.2020.2997710, IEEE Access

a certain extent, but the improvement effect is relatively limited.

TABLE IV
COMPARISON WITH AND WITHOUT AUGMENTATION

Method	AC	AUC	AP
MobileNetV1[46]	0.836	0.819	0.618
MobileNetV1[46] (with data-aug)	0.829	**0.830**	**0.645**
DenseNet-121[47]	0.836	0.806	0.610
DenseNet-121[47] (with data-aug)	**0.837**	0.800	0.636

2) PERFORMANCE COMPARISON WITH AND WITHOUT DISCRIMINANT NETWORK

By introducing a discriminate network in classification model to improve the feature discriminate ability and accuracy of the lightweight model, we contrasted the model without discriminate network, Table V shows the experimental results. We can see that after the introduction of the discriminate network as a constraint, the main evaluation indicator of our method has been significantly improved, there is a margin of~2.8%, ~5.5%, and ~1.5% in m AC, AP, and AUC, respectively, and the amount of trainable parameters of the model did not increase significantly. In addition, the DenseNet-121 with deeper network layers has achieved better results in the model with the introduction of the discrimination layer, which also shows that the performance of the CNN model can be improved by increasing the network depth. The training loss curve of the proposed model is shown in Fig.8(the batch size and the number of epochs are 32, 50, respectively).

TABLE V
COMPARISON WITH AND WITHOUT DISCRIMINANT NETWORK

Method	AC	AP	AUC	Parameters
MobileNetV1[46]	0.829	0.645	0.830	3,230,914
DenseNet-121[47]	0.837	0.636	0.800	7,039,554
DC-MobileNetV1	**0.865**	0.665	0.832	3,736,902
DC-DenseNet121	0.855	**0.700**	**0.845**	7,483,782

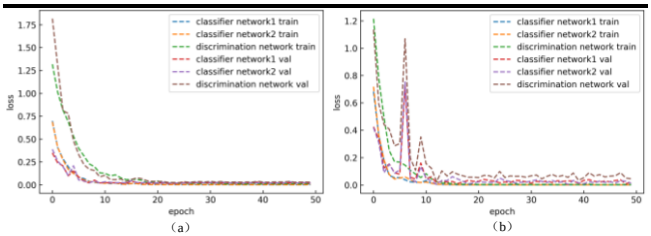

FIGURE 8. The training loss curve of proposed model. (a) DC-MobileNetV1. (b) DC-DenseNet121.

Table VI shows the result that use different loss functions to carry out the comparative experiment when use the discriminant network. Although the model using the Focal Loss function has acquired better results (unused model fusion), however, the score of each index is generally lower than the model using the cross-entropy loss function (unused model fusion), there is a margin of~1.1%, ~6.1%, and ~1.7% in AC, AP, and AUC, respectively. One possible reason is that the model based on focal loss has instability during training, which makes it difficult for the model to converge to the best effect. At the same time, because we use a data augmentation strategy to alleviate the imbalance of categories, the use of focal loss does not improve the model performance as expected.

TABLE VI
THE PERFORMANCE OF WITH DIFFERENT CLASSIFICATION LOSS FUNCTION

Method		AC	AP	AUC
Focal Loss	DC-MobileNetV1	0.850	0.633	0.838
	DC-DenseNet121	0.854	0.639	0.803
Cross Entropy Loss	DC-MobileNetV1	**0.865**	0.665	0.832
	DC-DenseNet121	0.855	**0.700**	**0.845**

3) PERFORMANCE COMPARISON WITH OTHER METHODS

We have made a wide comparison with the advanced melanoma recognition methods, these methods include: based on feature fusion [31], segmentation first and then recognition [34], combining adaptive sample learning strategy with multi-CNN [36], combining deep residual network with Fisher coding [37], multi-CNN collaborative training model [38], combining Fisher Vector and multi-CNN fusion [39]. As seen in Table VII, our method was superior the rank one method [34] in the ISBI 2016 classification task, and the methods [39] 6.3% and 1.3% respectively in AP index score, After the fusion (weighted average) of the prediction results of two different classification networks, the put forward approach exceeds the state-of-the-art method [39] 3.7% in AP index. It is should be noted that some methods contain more intermediate steps or higher calculation amounts. For instance, using multiple CNN and network fine-tuning [36]. The method [39] include three steps: image processing, CNN training, fisher vector coding and SVM training. which cannot achieve end-to-end model training. Our framework has very few parameters, the final fusion model size is only with 42M (The model size of method [37], [39] are 97.6M, 179M, respectively). Furthermore, because our model also can be trained end-to-end and effectively, which can be easily used in the analysis of other medical tasks.

TABLE VII
COMPARISON WITH THE OTHER METHODS

Method	AC	AP	AUC
BL-CNN [31]	0.850	0.625	-
CUMED [34]	0.855	0.637	0.804
M-CNN [36]	0.852	-	0.810
DCNN-FV [37]	0.868	0.685	0.852
MC-CNN [38]	0.863	0.681	0.822
M-CNN-FV [39]	0.868	0.687	**0.858**
DC-MobileNetV1	0.865	0.665	0.832
DC-DenseNet121	0.855	0.700	0.845
Fusion（our）	**0.876**	**0.724**	0.854

4) FEATURE VISUALIZATION

To further illustrate that our trained classification model mainly focuses on the lesion area and extracts more discriminative lesion features, we have visualized the feature activation map (Fig. 9). The areas with color gradients in Fig. 9 (b), (c) indicate the areas of attention when the

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/ACCESS.2020.2997710, IEEE Access

classification model makes classification decisions. It can be found that the areas of attention of the DC-MobileNetV1 and DC-DenseNet121 models are mainly concentrated in the lesion location, which indicates that the model has learned an effective feature representation mode. In order to further illustrate that our classification model extracts more discriminative features, Fig. 10 shows the result that use TSNE algorithm to cluster the final features of the classification model output, it can be found that the distribution of the two types of data, melanoma and benign, is scattered in the original data sets, and there is no obvious clustering phenomenon. This also indirectly illustrates the problems of large differences within the same lesion type, small differences between different lesion types. In Fig. 10 (b), (c), both types of data appear clustering phenomenon, and the clustering effect based on DC-Densenet121 is more significant, on the one hand, it indicates that our approach can extract more distinguishable data features, on the other hand, this shows to a certain extent that the features that can be extracted with deeper network structure are more distinguished.

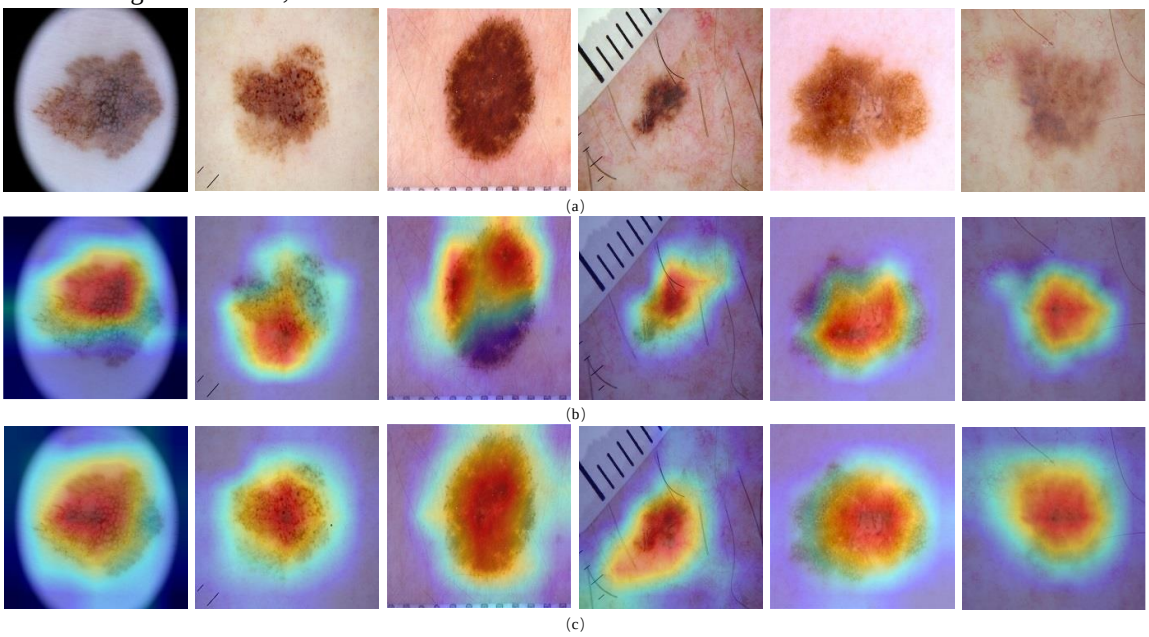

FIGURE 9. Visualization of the feature activation map. (a) the original dermoscopy image. (b) shows the feature activation map generated based on the weight value of the last separable convolutional layer of the DC-MobilenetV1 model. (c) shows the feature activation map generated based on the weight value of the last convolution layer of the DC-DenseNet121 model.

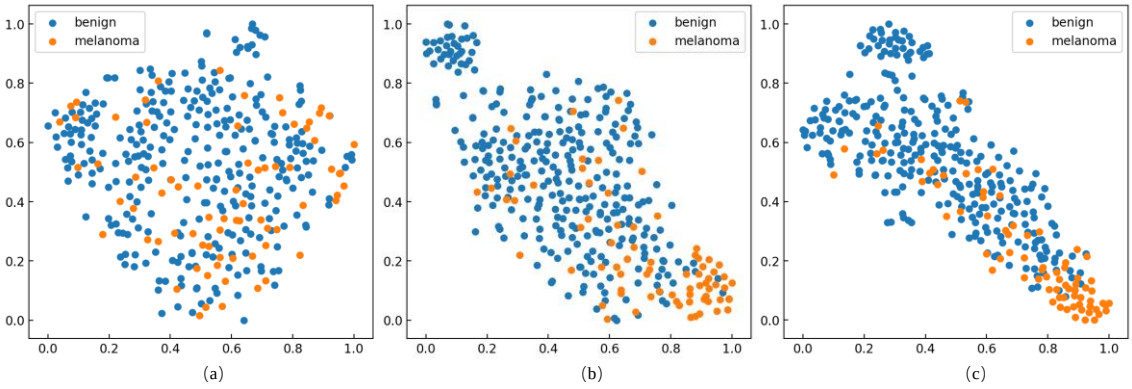

FIGURE 10. Visualization of the data feature clustering. (a) show the clustering effect of the original data features of the test set. (b), (c) show the clustering effect of the test set after feature extraction of DC-MobileNetV1 and DC-DenseNet121 models, respectively.

B. The Performance of our method in Segmentation tasks

1) EXPERIMENTS ON DIFFERENT FEATURE EXTRACTION MODEL UNDER THE U-NET ARCHITECTURE

The way to improve the effect of semantic segmentation is usually to use the model with a deeper and more complex network structure, but it also makes the parameters of the model become larger and training more difficult, at the same time, it also needs to consume more hardware resources. Therefore, we built a lightweight semantic segmentation model based on DenseNet-121 and MobileNetV1 under the U-NET architecture, we call it U-MobileNetV1, U-DenseNet121,

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/ACCESS.2020.2997710, IEEE Access

in addition, under the same conditions, we built different semantic segmentation model with large parameters based on VGG-16 and ResNeXt-50[56], we call it U-VGG16 and U-ResNeXt50, and then conducted a comparison experiment on the test set. Related experimental comparison data are listed in Table VIII. The results show that except for the SP indicator, the U-ResNeXt50 has attained the optimum score in other indicators, but at the same time, the model size also reached 125M. Compared with other models, the U-MobileNetV1 and U-DenseNet121 models are closer to U-ResNeXt50 in scores of various indicators, and the models are only respectively 32M and 48M. After a simple fusion of U-MobileNetV1 and U-DenseNet121, the model's main evaluation index scores exceeded ResNeXt50, there is a margin of~0.1%, ~0.1%, and ~0.3% in m AC, DC, and JA, respectively. The training loss curve of the proposed model is shown in Fig. 11(the batch size and the number of epochs are 32, 50, respectively).

TABLE VIII
THE PERFORMANCE OF WITH DIFFERENT FEATURE EXTRACTION MODEL UNDER THE U-NET ARCHITECTURE

Method	AC	DC	JA	SE	SP	Model Size
U-NET[49]	0.951	0.910	0.845	0.935	0.960	120M
U-ResNeXt50	0.961	0.922	0.864	**0.935**	0.972	125M
U-VGG16	0.955	0.909	0.843	0.927	0.968	73M
U-DenseNet121	0.958	0.918	0.859	**0.935**	0.971	48M
U-MobileNetV1	0.960	0.919	0.859	0.932	0.972	32M
Fusion(our)	**0.962**	**0.923**	**0.867**	0.934	**0.974**	80M

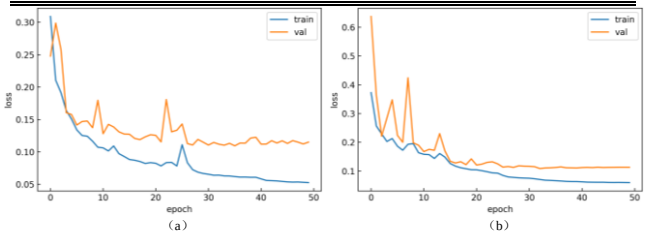

FIGURE 11. The training loss curve of proposed model. (a) U-MobileNetV1. (b) U-DenseNet121.

Table IX shows the effect of data augmentation processing on the proposed model performance. It can be found that after data augmentation, U-MobileNetV1 increases 1.2%, 0.4%, 0.4%, 0.2%, and 0.2% on JA, SP, AC, DC and SE indicators, respectively, and U-DenseNet-121 increases 1.5% 1.1%, 0.3%, 0.2% and 0.1% on SE, JA, AC DC and SP indicators, respectively. Overall, data augmentation can improve the performance of the model to a certain extent, but the improvement effect is relatively limited.

TABLE IX
COMPARISON WITH AND WITHOUT AUGMENTATION

Method	AC	DC	JA	SE	SP
U-MobileNetV1	0.956	0.917	0.847	0.930	0.968
U-MobileNetV1 (with data-aug)	**0.960**	**0.919**	**0.859**	0.932	**0.972**
U-DenseNet121	0.955	0.917	0.848	0.920	0.970
U-DenseNet121 (with data-aug)	0.958	0.918	0.859	**0.935**	0.971

Table X shows the performance difference of the model when using different loss functions (BCE Dice and Focal loss, $\partial = 0.75, \gamma = 1.25$). It can be found that the model based on focal loss achieve a higher score on the sensitivity index, because the area ratio of the non-lesion area to lesion area is about 1:3.6(statistical results of training set images), therefore, the category imbalance adjustment factor of focal loss plays a certain role in improving the SE index, but the scores in the other four indicators are generally lower than using BCE Dice loss, one possible reason for the model is that the single focal loss has numerical instability during training, which makes it difficult to converge to a better effect. At the same time, the excessive adjustment effect of factor (γ) may also bring negative effects and cause the related index score to be lower than the BCE Dice loss.

TABLE X
THE PERFORMANCE OF WITH DIFFERENT SEGMENTATION LOSS FUNCTION

Method		AC	DC	JA	SE	SP
Focal Loss	U-MobileNetV1	0.959	0.917	0.848	0.937	0.969
	U-DenseNet121	0.957	0.914	0.843	**0.939**	0.964
BCE Dice Loss	U-MobileNetV1	**0.960**	**0.919**	**0.859**	0.932	**0.972**
	U-DenseNet121	0.958	0.918	0.859	0.935	0.971

2) COMPARISON WITH DIFFERENT METHODS

We conducted extensive comparisons of the proposed approach with ranked number one approaches [35] (in ISBI 2016 challenge) and other advanced methods on ISBI 2016 segmentation test set. these methods include the FCN method using Jaccard distance loss [20], the FCN method combining multi-scale input [21], the method of post-processing after segmentation [35], the method of hybrid FCN [38], the VGG-16 method combining hole convolution [39], the method using hybrid U-Net [41], multi-stage U-Net architecture [43], multi-attention segmentation mechanism [44] and combine transform domain and CIELAB color space [45]. Table XI shows the final comparison results, it is seen in that our approach is better than the method [35], and the state-of-the-art methods [44] in main indicators JA by margins of ~2.4%, and ~0.9%, respectively. In addition to 4% lower than method [45] in the evaluation index SE, our method has the best score in other indexes. It is noteworthy that because our model is small and performs well in the segmentation index, it can be efficiently deployed to mobile end or applications other medical image analysis tasks.

TABLE XI
COMPARISON WITH OTHER METHODS

Method	AC	DC	JA	SE	SP
JD-FCN [20]	0.955	0.912	0.847	0.918	0.966
MS-FCN [21]	0.951	0.902	0.834	0.918	0.952
EXB [35]	0.953	0.910	0.843	0.910	0.965
MFCN-PI [40]	0.955	0.912	0.845	0.922	0.955
GL-FCN [41]	0.953	0.907	0.841	0.938	0.952
MS-UNet [43]	0.959	0.915	0.853	0.927	0.964
HR-CNN[44]	0.938	0.918	0.858	0.870	0.964
TD-CNN[45]	0.961	0.921	0.852	**0.974**	0.949

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/ACCESS.2020.2997710, IEEE Access

U-DenseNet121	0.958	0.918	0.859	0.935	0.971
U-MobileNetV1	0.960	0.919	0.859	0.932	0.972
Fusion(our)	**0.962**	**0.923**	**0.867**	0.934	**0.974**

3) VISUALIZATION ANALYSIS OF SEGMENTATION RESULT

We conduct a visual analysis from the perspective of the final probability map of the output of the segmentation network to further specify the effectiveness of the proposed segmentation model. As shown in Fig. 12, the red letters "A-E" indicate the image number, and the ISBI 2016 segmentation task stipulates that the probability value of the segmentation image pixel is less than 0.5 is the non-lesion area, otherwise, as lesion area. It can be found that the regions segmented by different models mainly focus on the location of lesions, which shows that the model has learned an effective feature representation mode of lesions. From the specific effect of the segmentation probability map, after the model fusion of U-MobileNetV1 and U-DenseNet121, the previous orange non-lesion area in the output probability map(Fig. 12 (b)-E) becomes light blue(Fig. 12 (d)-E), Combined with Fig. 12 (e), the probability value of this light blue area is less than 0.5, so the model will correctly determine this area as a non-lesion area, which shows that the fusion model can makes a more accurate classification of the pixels in the segmentation area and has better robustness. The contour difference map of the final segmentation results of the proposed method on partial challenging images are shown in Fig. 13, such as, containing with dense hair (Fig. 13 (a), (b), (c) and (d)), auxiliary marker (Fig. 13 (d), (e), (f) and (i)), low contrast (Fig. 13 (g), (h) and (i)), and irregular shapes Fig. 13 (j). Our method has achieved satisfactory segmentation results in these challenging cases, which proves that the semantic segmentation network based on the lightweight deep learning model is an efficacious way to deal with the challenge of skin lesion segmentation.

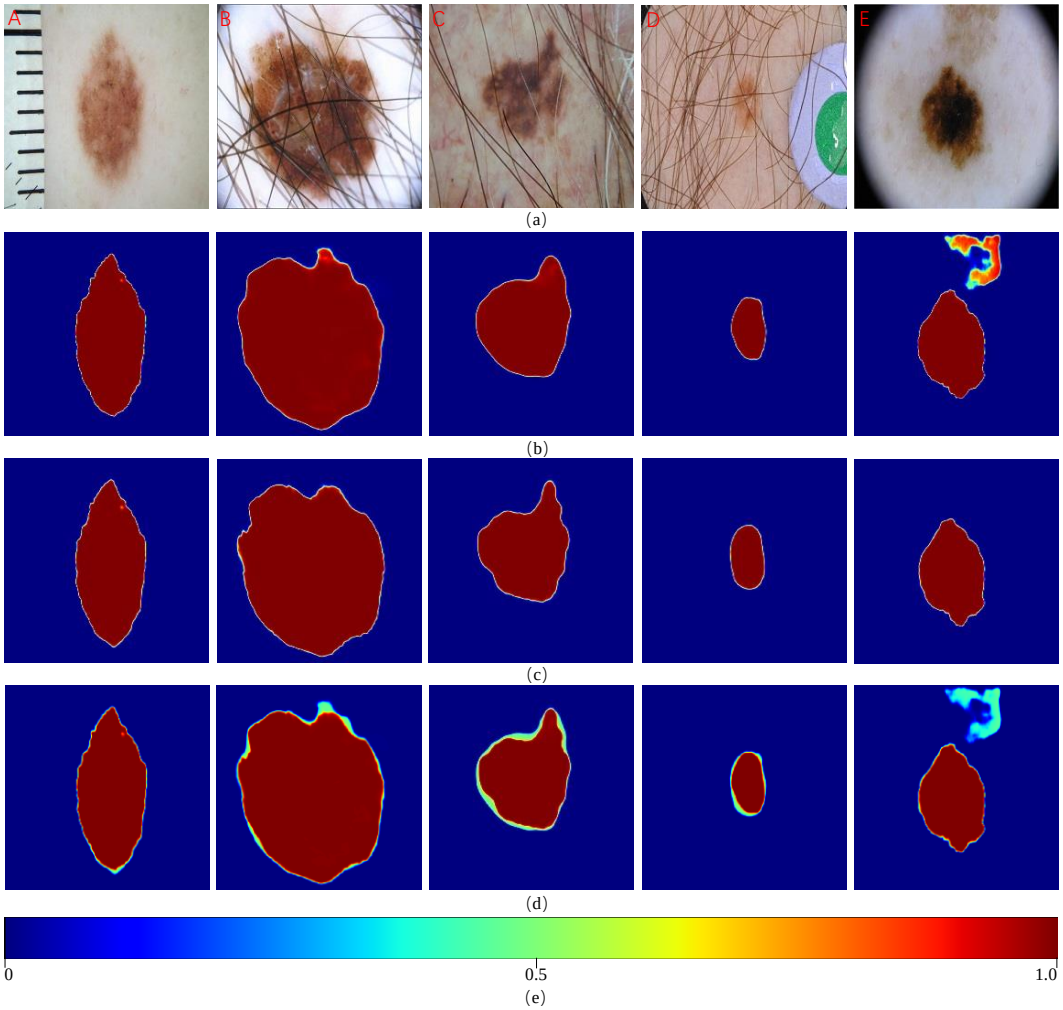

FIGURE 12 **Probability map of segmentation results of different models under U-Net structure. (a) Original image. (b), (c) and (d) are respectively the U-MobileNetV1, U-DenseNet121, U-MobileNetV1 and U-DenseNet121 fusion model output probability maps. (e) represent the pixel probability value distribution in the segmentation map.**

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/ACCESS.2020.2997710, IEEE Access

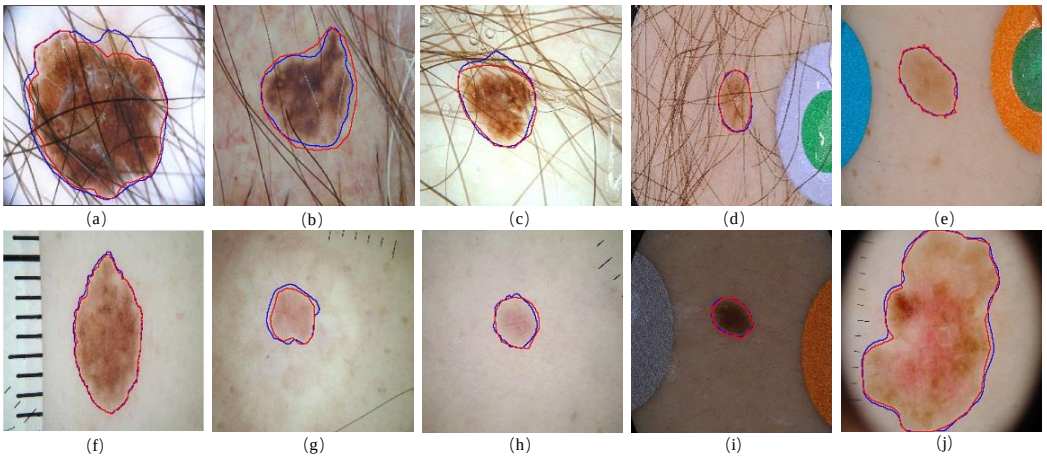

FIGURE 13. Some skin lesion segmentation contour renderings on test images. The blue and red outlines represent the segmentation results of the proposed method and the ground truth.

V. DISCUSSION

Based on the lightweight deep learning network, we propose a detection method to automatically recognizes skin cancer and segmentation skin lesion area present in dermoscopy images, at the same time, we have carried out a wide range of experiments to testing the effectiveness of the proposed method. In addition to the impressive results, there are also some influencing factors worthy of attention. On the other hand, although the data enhancement strategy can alleviate the data imbalance [56], [57], however, it is limited to improve the performance of the model through data enhancement strategy, the most direct way is to expand the size of the original dataset. Besides, by loading the weight of the pre-training model into the redesigned model, the training efficiency and performance of the model can be effectively improved, a recent study [58], [59] also explored this issue, but how pre-trained models effectively generalize the differences between medical and natural images is still lacking research. On the another hand, although the method of segmentation before recognition can make the model make better decisions based on the lesion area, but the good recognition effect depends largely on the accuracy of the segmentation network, and due to the classification network has certain requirements on the resolution of input image, result in the input image size of segmentation network is too large, which also puts forward higher computing resources and unsuitable for mobile end design requirements. Furthermore, with limited training data, it is difficult for us to fully excavate the discriminative ability of the lightweight deep learning network. So that even if our method can gain satisfactory results in most cases, but in some cases, the performance of the proposed method is still unsatisfactory, as shown in Fig. 14 (a) and Fig. 14 (b).

It should be noted that although focal loss has achieved good results in target detection, but some factors may also cause its effect to fail to meet expectations, such as: adjustment of hyperparameters, instability during training, etc. Therefore, further improvement or use in combination with other loss functions is one of the strategies to improve its performance. Finally, how to further improve the accuracy of the model and deploy the model to the mobile end or the web end for people to assist in the diagnosis of skin diseases and timely discover the potential lesion risk, this is undoubtedly another work that our will study in the future.

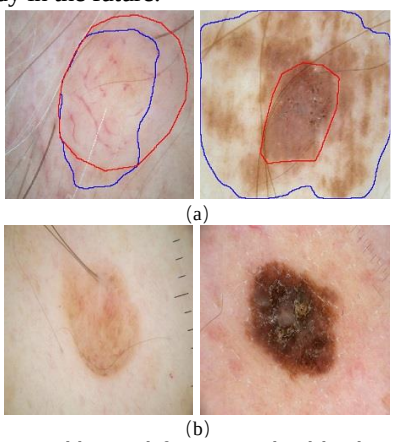

FIGURE 14. Cases with unsatisfactory results. (a) suboptimal cases of segmentation, the red curve and blue curve are representing the segmentation result of the ground truth and our approach in turn, (b) melanoma (left) and benign (right) were incorrectly identified as benign and melanoma by our method, respectively.

VI. CONCLUSION

In this paper, we have designed a discriminant dermoscopy image lesion recognition model. It uses a pre-trained lightweight network as a feature extractor to construct a dermoscopy image lesion classification branch network and lesion feature discriminant branch network, through the joint training of each branch network, the proposed model achieves the classification of lesion type and the similarity of lesion features at the same time, so it can extract more discriminative lesion features, Compared with the existing multi-CNN fusion method or the method based on local depth feature Fisher Vector coding, our framework can achieve an approximate or even higher model performance with a lower number of model parameters end-to-end; Meanwhile, Based on the feature extractor of the lesion recognition model of the proposed

dermoscopy image, we constructed a lightweight semantic segmentation model, by replacing the feature extraction module with a lightweight feature extraction module and combining with a migration training strategy, the proposed method achieves higher segmentation accuracy while maintaining small amount of model parameters.

We conducted systematic and extensive experiments to study some key factors that may affect the performance of our method, including network architecture, data enhancement, and loss function selection. Through extensive experimental comparisons with the state-of-the-art methods on the open challenge dataset of ISBI 2016 and validate the effectiveness and superiority of the proposed method. Further research includes the design of more effective feature discrimination networks, evaluating our method on more datasets and further development to facilitate cross-platform application deployment.

ACKNOWLEDGMENT

We would like to thank web of science for the help of literature search in the process of writing the paper.

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA Cancer J. Clin.*, vol. 66, no. 1, pp. 7–30, 2016.

[2] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA Cancer J. Clin.*, vol. 67, no. 1, pp. 7–30, 2017.

[3] A.F. Jerant *et al.*, "Early detection and treatment of skin cancer," *Amer Family Phys*., vol. 62, no. 2, pp. 357–386, 2000.

[4] C. M. Balch *et al.*, "Final version of the American joint committee on cancer staging system for cutaneous melanoma," *J. Clin. Oncol*, vol. 19, no. 16, pp. 3635–3648, 2001.

[5] M. L. Bafounta *et al.*, "Is dermoscopy (epiluminescence microscopy) useful for the diagnosis of melanoma? Results of a meta-analysis using techniques adapted to the evaluation of diagnostic tests," *Jama Dermat*, vol. 137, no. 10, pp.1343-1350, oct, 2001.

[6] G. Argenziano, H. P. Soyer, S. Chimenti, R. Talamini, R. Corona *et al.*, "Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet," *J. Am. Acd. Dermatol*, vol. 48, no. 5, pp. 679-693, May, 2003.

[7] C. Barata, M. E. Celebi and J. S. Marques, "Improving Dermoscopy Image Classification Using Color Constancy," *IEEE Trans. Inf. Technol. B*, vol. 19, no. 3, pp. 1146-1152, May 2015.

[8] F. Ercal, A. Chawla, W. V. Stoecker, H. C. Lee and R. H. Moss, "Neural network diagnosis of malignant melanoma from color images," *IEEE Trans. Biomed. Eng.*, vol. 41, no. 9, pp. 837-845, Sept. 1994.

[9] Y. Cheng, Ragavendar. S, S. E. Umbaugh, *et al.*, "Skin lesion classification using relative color features," *Skin. Res. Technol*, vol. 14, no. 1, p 53-64, Feb 2008.

[10] M. Ruela, C. Barata, J. S. Marques and J. Rozeira, "A system for the detection of melanomas in dermoscopy images using shape and symmetry features," *Comput. Method. Biomec. Imag. Vis*, vol. 5, no. 2, pp. 127-137. 2017.

[11] T. Tommasi, E. L. Torre, and B. Caputo, "Melanoma recognition using representative and discriminative kernel classifiers," *presented at the Proc. Int. Workshop Comput. Vis. Approaches Med. Image Anal*., Berlin. GER. 2006.

[12] M. E. Celebi *et al.*, "A methodological approach to the classification of dermoscopy images," *Comput. Med. Imag. Graph.*, vol. 31, no. 6, pp. 362–373, 2007.

[13] E. Almansour and M. A. Jaffar, "Classification of Dermoscopic Skin Cancer Images Using Color and Hybrid Texture Features," *Int. J. Comput. Science Network Security (IJCSNS)*, vol.16, no. 4, pp. 135-139, April 2016.

[14] B. Catarina, R. Margarida, F. Mariana, M. Jorge, M. Teresa, "Two Systems for the Detection of Melanomas in Dermoscopy Images Using Texture and Color Features," *IEEE. Syst. J*, vol. 8, no. 3, pp. 965-979. July. 2013.

[15] W. V. Stoecker *et al.*, "Detection of granularity in dermoscopy images of malignant melanoma using color and texture features," *Comput Med Imaging Graph*, vol. 35, no. 2, pp. 144-147, Mar, 2011.

[16] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler, "Automated melanoma recognition," *IEEE. Trans. Med. Imag.*, vol. 20, no. 3, pp. 233–239, Mar. 2001.

[17] S. Gerald, K. Bartosz, C. M. Emre, I. Hitoshi, "An ensemble classification approach for melanoma diagnosis," *Memet. Comput*, vol. 6, no. 4, pp. 223-240, Oct. 2014

[18] C. BarataEmail, M. Ruela, T Mendonca, J. S. Marques, "A Bag-of-Features Approach for the Classification of Melanomas in Dermoscopy Images: The Role of Color and Texture Descriptors," *Comput. Vis. Techni. Diagn. Skin Cancer*. Berlin, GER, Springer, 2014, ch. 3, pp. 49-69.

[19] P. Moeskops, M. A. Viergever, A. M. Mendrik *et al.*, "Automatic segmentation of MR brain images with a convolutional neural network," *IEEE Trans. Med. Imag*, vol. 35, no. 5, pp. 1252-1261, Mar. 2016.

[20] Y. D. Yuan, M. Chao, Y. C. Lo, "Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks with Jaccard Distance," *IEEE Trans. Med. Imag*, vol. 36, no. 9, pp. 1876-1886, Sep. 2017.

[21] L. Bi, J. Kim, E. Ahn, D. Feng and M. Fulham, "Semi-automatic skin lesion segmentation via fully convolutional networks," in *Proc. IEEE 14th Int Symp. Biomed. Imag (ISBI 2017)*, Melbourne, VIC, 2017, pp. 561-564.

[22] W. Shen, M. Zhou, F. Yang, C. Y. Yang, J. Tian, "Multi-scale Convolutional Neural Networks for Lung Nodule Classification," Information Processing in Medical Imaging, Cham, GER, Springer, 2015, ch. 63, pp. 588-599.

[23] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, andS. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Trans. Med. Imag*., vol. 35, no. 5, pp. 1207–1216, May 2016.

[24] C. Francesco *et al.*, "Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box," *Med. Imag. Anal*, vol. 26, no. 1, pp. 195-202, Dec. 2015.

[25] Q. Dou, H. Chen, L. Q. Yu, L. Zhao, Qin *et al.*, "Automatic Detection of Cerebral Microbleeds from MR Images via 3D Convolutional Neural Networks," *IEEE Trans. Med. Imag*, vol. 35, no. 5, pp. 1182-1195, May. 2016.

[26] A. A. Setio, C. Jacobs, J. Gelderblom, B. V. Ginneken, "Automatic detection of large pulmonary solid nodules in thoracic CT images," *Med. Phys*, vol. 42, no. 10, pp. 5642-5653, Oct. 2015.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst*., 2012, pp. 1097–1105

[28] J. Kawahara, A. BenTaieb and G. Hamarneh, "Deep features to classify skin lesions," in *Proc. Int Symps. Biomed. Imag (ISBI)*, 2016, pp. 1397-1400.

[29] K. He *et al.*, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*, 2015, pp. 770–778.

[30] S. Karen and Z. Andrew, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *Proc. Int. Conf. Learn. Representat* (ICLR). San Diego, USA, 2015.

[31] Z. Y. Ge *et al.*, "Exploiting local and generic features for accurate skin lesions classification using clinical and dermoscopy imaging," in *Proc. Int. Symp. Biomed. Imag (ISBI)*, Melbourne, VIC, 2017, pp. 986-990.

[32] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*, Boston, MA, 2015, pp. 1-9.

[33] A. Esteva, B. Kuprel, R. A. Novoa and J. Ko, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 546, no. 7660, pp. 686-686, Jun. 2017.

[34] L. Q. Yu, H. Chen, Q. Dou, J. Qin and P. Heng, "Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks," *IEEE Trans. Med. Imag*, vol. 36, no. 4, pp. 994-1004, April. 2017.

[35] D. Gutman *et al.*, (May 2016). "Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical

imaging (ISBI) 2016, hosted by the international skin imaging collabo-ration (ISIC)." [Online]. Available: https://arxiv.org/abs/1605.01397.
[36] Y. Guo, A. S. Ashour, L. Si and D. P. Mandalaywala, "Multiple Convolutional Neural Network for Skin Dermoscopic Image Classification," *IEEE Int. Symp. Signal Proc. Informat. Technol (ISSPIT)*, Louisville, KY, USA, pp. 365-369, 2018.
[37] Z. Yu *et al.*, "Melanoma Recognition in Dermoscopy Images via Aggregated Deep Convolutional Features," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 4, pp. 1006-1016, April 2019.
[38] J. P Zhang, Y. T Xie, Q Wu, et al. "Medical image classification using synergic deep learning" [J]. *Med Image Anal*, 2019, 54: 10-19.
[39] Y Zhen, Feng J, Feng Z, et al. "Convolutional descriptors aggregation via cross-net for skin lesion recognition"[J]. [Online]. Available: https://doi.org/10.1016/j.asoc.2020.106281
[40] L. Bi, *et al.*, "Dermoscopic image segmentation via multi-stage fully convolutional networks," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2065-2074, Sep. 2017.
[41] Z. Deng, *et al.*, "Segmentation of dermoscopy images based on fully convolutional neural network," in *Proc. IEEE Int. Conf. Imag. Proc.*, Beijing, China, pp.1732-1736, Sep. 2017.
[42] Y. X. Li, L. L. Shen, "Skin Lesion Analysis Towards Melanoma Detection Using Deep Learning Network," *Sensors*, vol. 18, no. 2, pp. 556-572, Feb. 2018.
[43] Y. Tang, F. Yang, S. Yuan and C. Zhan, "A Multi-Stage Framework With Context Information Fusion Structure For Skin Lesion Segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag (ISBI 2019)*, Venice, Italy, 2019, pp. 1407-1410.
[44] F. Y Xie. et al. "Skin lesion segmentation using high-resolution convolutional neural network." Comput Methods Programs Biomed. 2020, 186:105241.
[45] Pour M P, Seker H. "Transform domain representation-driven convolutional neural networks for skin lesion segmentation" [J]. Expert Systems with Applications, 2020, 144: 113129.
[46] G. Howard, M. L. Zhu, B. Chen, D. Kalenichenko, W. J. Wang, T. Weyand, M. Andreetto, and H. Adam., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv:1704.04861*, 2017.
[47] G. Huang *et al.*, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Patt. Recog*, Honolulu, Hawaii, USA, pp. 2261–2269, 2017.
[48] T. Y. Lin, P. Goyal, R. Girshick *et al.*, "Focal loss for dense object detection," in *Proc. IEEE Int. conf. Comput. Vis*, Venice, Italy 2017. pp. 2980-2988.
[49] O. Ronneberger *et al.*, "U-net: Convolutional networks for biomedical im-age segmentation," presented at the Conf. Med. Imag. Comput. Comput-Assist Intervention, Munich, Germany, 2015.
[50] P. Sahu, D. T Yu, H. Qin., "Apply lightweight deep learning on internet of things for low-cost and easy-to-access skin cancer detection," presented at the Imag. Inf. Healthcare , Houston, TX, USA, 2018.
[51] M. A. R Ratul, M. H. Mozaffari, E. Parimbelli, W. S. Lee., "Atrous Convolution with Transfer Learning for Skin Lesions Classification," [Online]. Available: https://doi.org/10.1101/746388
[52] A. M. Taqi, F. A. Azzo, A. Awad, M. Milanova, "Skin Lesion Detection by Android Camera based on SSD-Mobilenet and TensorFlow Object Detection API," *Int. J. Adv. Res*, vol.3, no. 1, pp. 5-11, Jul. 2019.
[53] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei., "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput.Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 248–255.
[54] Iandola F N, Han S, Moskewicz M W, et al. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size" [J]. *arXiv:1602.07360, 2016.*
[55] Zhang X, Zhou X, Lin M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]. in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.* Salt Lake City, US, 2018.
[56] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," in *Proc. Br. Mach. Vis. Conf.* Jubilee Campus, UK, 2014.
[57] A. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *CVPR Workshops*, Columbus, OH, USA, 2014.
[58] S. N. Xie, R. Girshick, P. Dollár, Z. W. Tu, & K. M. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn., Honolulu, Hawaii, USA, 2017.*
[59] A. Menegola, M. Fornaciali, R Pires, *et al.*, "Knowledge transfer for melanoma screening with deep learning," presented at the IEEE 14th Int. Symp. Biomed. Imag., Melbourne, VIC, Australia, 2017.

LISHENG WEI received his B.S. degree in automation from Anhui Polytechnic University in 2001, the MSc degree in flight vehicle design from China Aerospace Science and Industry Corporation 061 Base in 2004, and the PhD degree in control science and engineering from Shanghai University in 2009. Since 2017, he has been a professor with the School of Electrical Engineering, Anhui Polytechnic University. He is the author of more than 70 articles and holds 12 patents. His research interests include image recognition and application, embedded instrumentation and system, intelligent network control theory, system and simulation, data fusion. Prof. WEI was a recipient of Rookie of the Teaching Forum Award in 2013, and the Anhui Natural Science Outstanding Paper Second Award in 2010.

KUN DING (Corresponding author) received bachelor's degree from Chuzhou University in 2017. Now he is a postgraduate with the School of Electrical Engineering of Anhui Polytechnic University, China. His main research interest includes image processing, deep learning.

HUOSHENG HU received the MSc degree in industrial automation from Central South University, China in 1982, and the PhD degree in robotics from the University of Oxford, U.K. in 1993. He is currently a Professor with the School of Computer Science and Electronic Engineering, University of Essex, U.K., and the head of the Robotics Research Group. His research interests include behavior-based robotics, human–robot interaction, embedded systems, multi-sensor data fusion, machine learning algorithms, mechatronics, and pervasive computing. He has authored over 500 papers in journals, books, and conferences and received several best paper awards. He is a fellow of the Institute of Engineering and Technology and the Institute of Measurement and Control, U.K., and a Chartered Engineer. He has been the Program Chair or a member of the Advisory Committee of many IEEE international conferences, such as the IEEE ICRA, IROS, ICMA, and ROBIO. He currently serves as one of Editor-in-Chiefs for the International Journal of Automation and Computing, Editor-in-Chief of MDPI Robotics Journal, and Executive Editor of the International Journal of Mechatronics and Automation.