参赛队员姓名: 付鑫雨

中学: 重庆南开中学

省份: 重庆

国家/地区: 中国

指导教师姓名: 黄婧

论文题目: Multi-Scale Adaptive Fusion Network for Skin Cancer Recognition

本参赛团队声明所提交的论文是在指导老师指导下进行的研究工作和取得的研究成果。尽本团队所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果。若有不实之处，本人愿意承担一切相关责任。

参赛队员：付鑫雨　指导老师：黄婧

2020年9月14日

Multi-Scale Adaptive Fusion Network for Skin Cancer Recognition

付鑫雨

shirlyfxywl@outlook.com

重庆市南开中学

Multi-Scale Adaptive Fusion Network for Skin Cancer Recognition

Abstract

Skin cancer is a very common cancer in today's society. The incidence of skin cancer covers all age groups. The diagnosis process of skin cancer is very complicated, requiring doctors to observe and judge first, and then perform a biopsy under a microscope. Therefore, more accurate skin disease classification algorithms will provide great help for the timely diagnosis of skin cancer. Existing skin cancer recognition methods directly borrow from natural image recognition models, and do not consider the scale variability and sample imbalance of skin cancer images, so the recognition accuracy is limited. To solve this problem, a multi-scale adaptive fusion network (MSAF-Net) is proposed for skin cancer recognition. The main novelty of the MSAF-Net is the consideration of the scale variability and sample imbalance of medical images. Specifically, the revised VGG16 model is exploited to learn multi-scale features, and the designed multi-scale fusion step is used to adaptively fuse them, so as to obtain discriminant line features containing multi-scale information, avoiding the impact of scale variability on recognition accuracy influences. In addition, a weighted loss function is designed for solving the problem of sample category imbalance by reducing the weight of simple samples. Based on the disclosed skin lesion image HAM10000 challenge dataset, this paper adds a class of normal skin images and constructs a HAM10000+ dataset to verify the proposed MSAF-Net method. Experimental results on the HAM10000+ dataset show that the MSAF-Net can effectively learn multi-scale information and solve the problem of unbalanced samples of medical images, thereby further improving the accuracy of skin cancer recognition.

Key Words

Skin Cancer Recognition; Multi-Scale; Adaptive Fusion; Weighted Loss; Transfer Learning

1. Introduction

As the largest organ in human body, the skin plays important roles in protecting people from injury, infection, and sunlight. Skin cancer is the most common cancer nowadays, globally accounting for 40% of cancer cases [3], [10]. Exposure to ultraviolet (UV) radiation may lead to the skin cancer of melanoma [7]. External stimulation, in addition to the above-mentioned ultraviolet radiation, also includes other skin injuries and repeated mechanical stimulation, so in the palm, soles of the feet, mucous membranes, nails and other areas often subject to friction accounted for 60% to 75% of the disease site [21]. Basal-cell *etc.* are the popular skin cancers [8]. Exposure to sunlight will increase risk for all three types of cancer [9]. Compared to non-melanoma skin cancer, which is more easily treated, melanoma tends to be more malignant and its incidence is increasing as well [5]. Most common diagnosis includes biopsy and histopathological examination [11]. Early detection could ensure higher survival rate [12]. However, relying on the naked eye to detect and diagnose tumors is subjective, and the reproducibility of diagnosis results is not ideal. Although skin mirror serves as a possible solution, the complexity of the skin mirror image itself, such as blurred skin boundary, in-group variation and intergroup similarity, present great technical challenges to melanoma detection [16]. For physicians, who also have the ability to diagnose, the diagnosis results of the same case are subjectively different and the reproducible effect is not ideal. The complexity of the skin mirror picture itself, the characteristics of small gap between classes and large gap between classes, make diagnosis prone to omission and misdiagnosis, resulting in the overall accuracy rate is not high enough [23].

Existing methods can be divided into two categories: 1) skin cancer recognition methods on the basis of hand-crafted features, and 2) skin cancer recognition methods on the basis of deep learning features. The former methods use traditional classifier, such as SVM and KNN to categorize via color, texture, and other features. For example, Codella et al. proposed a method that integrates sparse coding and CNN for skin cancer skin cancer recognition. Deep learning is mainly by designing a CNN to study features of skin cancer and categorize. All these findings lay solid foundation for later improvements in recognition. Automatic diagnosis and diagnosis technology not only helps patient complete health checks more quickly, but also helps medical image technicians and doctors reduce manual analysis time, improve work efficiency, and reduce the likelihood of misdiagnosis. However, although the accuracy of many of these methods is comparable to dermatologists and some even outperform them, some considerations still remain.

These methods have prompted the development of the skin cancer recognition. However, they directly borrow from natural image recognition models, and do not consider the scale variability and sample imbalance of skin cancer images, so the recognition accuracy is limited. In addition, these hand-crafted-based methods involve complex and cumbersome processes, have low generalization capabilities, and are not very applicable in clinical practice. Unlike methods that rely on artificial functions, in the past few years, CNN have had significant advantages in image recognition tasks [13]. Its superiority is the ability for automatically learning high-level features for various tasks [15]. In reality, medical images contain plenty of information about anatomical structure and pathology, so a CNN specially designed for medical images is needed. Moreover, there is no specific solution for disproportionate samples of different categories.

To solve the above problems, a multi-scale adaptive fusion network (MSAF-Net) is proposed for skin cancer recognition. Specifically, the proposed MSAF-Net can be divided into three steps: 1) multi-scale feature extraction, 2) multi-scale feature fusion, and 3) skin cancer recognition. First, we revised the original VGG16 model to learn multi-scale features from pooling1, pooling2, pooling3, pooling4, and conv5-3 layers. The learned multi-scale features contain different information with multiple receptive fields. Second, we combine the characteristics of multi-scale features by a scale-by-scale fusion. This step can adaptively integrate different information based on the designed reduction-attention module and densely connected module. Third, as for the fused feature, the dense connected layer is adopted to further learn the discriminative representation and infer the corresponding semantic label. In addition, to solve the problem of sample category imbalance, a weighted loss function is designed. Under the constraint of the designed weighted loss function, the whole network is well trained.

Overall, our contributions in this study involve four aspects:

1) We proposed a new skin cancer recognition framework, named multi-scale adaptive fusion network (MSAF-Net) for skin cancer diagnosis and improved the classification accuracy to nearly 98%.

2) Multi-scale features are learned to solve the scale variability problem of skin cancer images, so as to integrate different information with multiple receptive fields.

3) To solve the problem of imbalanced and insufficient samples in medical images, we proposed a new weighted loss function.

4) An amplified HAM10000+ dataset is developed for training the proposed MSAF-Net to distinguish between normal skin and skin cancer images.

This paper is arranged as follows. Section 2 describes related works of previous researchers in this area.

Section 3 discusses the method we proposed. Section 4 shows the experiments and result. In section 5, the conclusion is presented.

2. Related Works

In the past several years, many skin cancer recognition methods have been proposed. These methods can be classified into two main categories based on their characteristics. 1) skin cancer recognition methods on the basis of handcrafted features and 2) skin cancer recognition methods on the basis of deep learning features.

2.1. Skin Cancer Recognition Based on Handcrafted Features

Early attempts were made to use manual features to categorize skin cancer. As early as 2001, Ganster applied a supervised method for the recognition of skin cancer images. The main steps are to recognize the skin cancer image based on the combination of color and gradient features, where the specificity and sensitivity of the final recognition results are 84% and 77%, respectively. Later, Radu et al. used fractal and texture as features to analyze and discriminate between genign and maglignant tumors [36]. Besides, Machine learning is also applied to improve skin cancer recognition. Common machine learning algorithms have also made various attempts on this problem, such as SVM [17], KNN [18] or the integration of multiple algorithms. Besides, Celebi and others used a threshold-based segmentation method to split the skin loss area and then used a support vector machine to classify images that extracted color, texture, and shape features [23]. Sumithra's tissue segmentation of the skin mirror image was identified using KNN and SVM, resulting in an F-measure value of 0.61[24]. Codella et al. proposed a to integrates SVM and sparse coding for the skin cancer recognition [25]. Prachya et al. designed a segmentation scheme based on combination of snake model and SVM, and applied SVM in finding an appropriate initial curve and parameters for snake algorithm [32]. Ritesh et al. used normalized symmetrical GLCM to acquire texture features based on vector machine as a model for skin cancer classification [35].

2.2. Skin Cancer Recognition Based on Deep Learning Features

Recently, some researchers also hope to apply the level of learning ability of CNN to improve the recognition ability of skin cancer. Krizhevsky et al. used convolutional neural networks to classify big data (1.2 million images) in 1,000 images and develop a new technology (AlexNet) with the best results [31]. Li Huang et al. proposed to classify melanoma and non-melanoma by mainly extracting local convolution features from the deep residual network to form more complex expressions, using key factors that may affect the performance, including image preprocessing, data enhancement, and network architecture [20]. The Deepmole model proposed by Pomponiu et al. uses the idea of knowledge migration: train the DNN network with natural images, and then classify the network as a feature extractor, which solves the problem of a smaller sample image [27]. Kawahara et al. used AlexNet-based full-volume neural networks (FCNs) to classify ten categories of dermatological images using multi-scale features of non-skin mirror melanoma images [28]. Hashebe Younis et al. adjusted MobileNet for training the skin cancer recognition model [29]. Ahmet et al. used deep learning architectures for the skin cancer recognition task, and obtained 84.09% and 87.42% accuracy rate respectively [34]. Pooyan et al. exploited the GAN model to generate unseen skin cancer images for training the skin cancer recognition model, which enhanced accuracy for 18% [35]. Brij et al. used the Deep Residual learning model with three parametric layers to detect skin cancer and achieved an accuracy of 77% on the ISIC [37].

3. Proposed Method

The purpose of this study is to solve the problems of scale variation and sample imbalance in medical images and to realize accurate recognition of various skin cancer images. Multi-scale adaptive fusion network is proposed for skin cancer recognition. As depicted in Figure 1, the MSAF-Net has four aspects. Data preprocessing, multi-scale feature extraction, multi-scale feature fusion, and skin cancer recognition. The rest of this chapter details these four aspects and introduces the optimization process for the entire model.

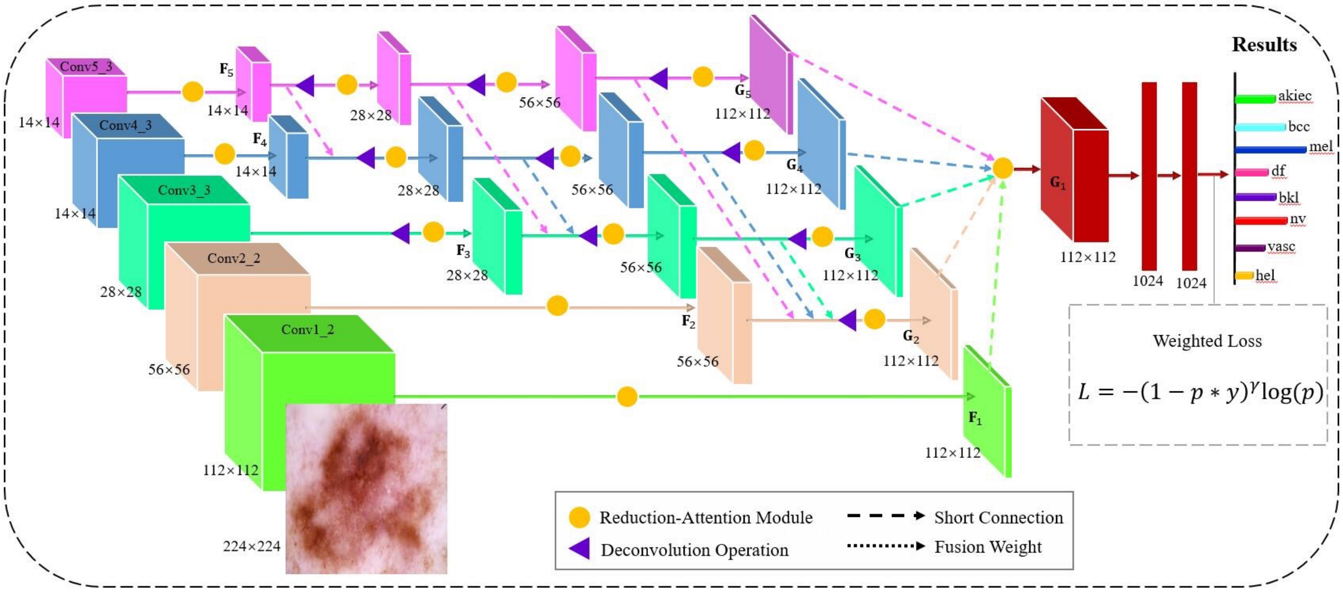

Figure 1. The framework of the MSAF-Net. First, multi-scale features are learned from the revised VGG16. Second, a densely connected module is devised to adaptively fuse multi-scale features scale-by-scale. Finally, the fused features are exploited to predict the semantic category of the input skin cancer image. The proposed method is trained with a designed weighted loss in an end-to-end manner.

3.1. Data Preprocessing

In order to obtain efficient training data, we have done two preprocessing operations: 1) image scale normalization, 2) data enhancement.

3.1.1. Image Scale Normalization

In order to train the proposed MSAF-Net for skin cancer recognition, the input size needs to be uniformed. As a result, the image size is normalized. Traditional image normalization methods, such as bilinear interpolation, are with the following disadvantage. When the aspect ratio of the target image and original image are different, these methods will deform the image, and the useful information of the image will be greatly compressed. The deformation and compression of the detailed features of the target will affect the classification accuracy.

To solve it, this article adopts an image scale normalization method. This method makes N different copies for each image with a different aspect ratio, where each copy is a part of the image, which owns the same aspect ratio with the original sin cancer image. The specific steps are:

1) Obtain the width W_{in} (pixels) and height H_{in} (pixels) of the input image, and calculate the aspect ratio of the input image:

$$A_{in} = \frac{H_{in}}{W_{in}} . \tag{1}$$

2) Get the width W_{go} and height H_{go} of the target image Calculate the aspect ratio of the target image:

$$A_{go} = \frac{H_{go}}{W_{go}} . \tag{2}$$

3) If $A_{in} = A_{go}$, the size of the image is interpolated directly using the nearest neighbor interpolation method; otherwise, proceed to steps 4)-6).

4) Calculate the short and long sides of the input image:

$$edg_{sh} = \min\left\{H_{in}, W_{in}\right\}, \tag{3}$$

$$edg_{lo} = \max\left\{H_{in}, W_{in}\right\}. \tag{4}$$

5) Take edg_{sh} as a measure and intercept N segments of length len in the long side edg_{lo} with step as the unit:

$$len = \begin{cases} edg_{sh} \times A_{go} & edg_{sh} = W_{in} \\ \dfrac{edg_{sh}}{A_{go}} & edg_{sh} = H_{in} \end{cases} . \tag{5}$$

The image whose length and width are both len is intercepted, so that the input image is converted into N images with an aspect ratio equal to the target image aspect ratio A_{go}.

6) The image obtained above is transformed into equal proportions and normalized to a target image of the same size.

The target image size in this paper is set as 224×224 pixels, step length is set as $step = (edg_{lo} - edg_{lo})/3,$ and $N = 4$. The normalization effect of flower images whose aspect ratio is not equal to the target aspect ratio. The images generated after scale normalization in this way retain a common aspect ratio, which can avoid the phenomenon of image distortion.

3.1.2. Data Enhancement

The limited amount of data restricts the training of neural networks. In order to enrich the amount of data and avoid over-fitting, this article mainly uses histogram equalization, rotation, and scaling to enhance the dataset. Since skin cancer images do not have rotation invariance, we need to pay attention when using image rotation to enhance the dataset, and only rotate in a small range.

Directly using the network model to classify skin cancer images can get better results. However, to further improve the generalization ability in the case of a small network and a small amount of target data, this paper adopts the idea of transfer learning through skin cancer. The global feature of the direction field of the image is pre-trained for classification, and then the network model is fine-tuned with the skin cancer image. The direction field is both a global and a basic feature, which can accurately reflect the characteristics of skin cancer. Most skin cancer classification and recognition algorithms use the direction field as the research basis. There are many methods for obtaining the direction field of skin cancer, and the classical method based on gradient estimation is the most used and effective. Although more new methods have appeared one after another, the tests on CNN have confirmed that the model performance based on gradient estimation method is better. The method can be divided into three steps: obtaining Gaussian smoothing, gradient calculation, and direction field estimation. First, given a Gaussian filter, perform Gaussian smoothing on the original image to get the smoothing result G, namely:

$$G(x,y)=\frac{1}{2\pi\sigma^2}e^{\frac{-x^2+y^2}{2\sigma^2}} , \tag{6}$$

where σ is the standard deviation, x,y is the location of pixels. Then, we adopt the Sobel operator to compute the gradient of the image in different directions, namely:

$$S_x=\begin{bmatrix}-1 & 0 & 1\\ -2 & 0 & 2\\ -1 & 0 & 1\end{bmatrix}, \tag{7}$$

$$S_y=\begin{bmatrix}-1 & -2 & -1\\ 0 & 0 & 0\\ 1 & 2 & 1\end{bmatrix}, \tag{8}$$

where S_x and S_y are the gradient in different directions. Finally, estimate the image orientation field as follows:

$$\tan^{-1}\left[\frac{S_y}{S_x+\varepsilon}\right], \tag{9}$$

where ε is a very small constant. This article hopes to reduce the manual operation process of the image as much as possible, and the obtained direction field picture is a vector direction field picture. After obtaining the direction field map, first perform micro-rotation enhancement processing to obtain a large dataset; then input the network for pre-training to obtain the final pre-trained network model.

3.2. Multi-Scale Feature Extraction

In this paper, we revise the VGG16 for learning multi-scale features [40]. The VGG16 model includes the different functional layers [38], [39]. Among them, the convolutional layer plays a very important role, through the realization of local perception and parameter sharing. The convolution kernel is the core of the convolution layer. Based on it, the shape of the same object located at different positions in the image can be extracted, which can reduce the dimensionality. In the pooling layer, a pooling filter is used. The last is the dense connected layer, which functions as a classifier. Since the data to be classified this time is different from the original VGG16 classification data, the parameters of the dense connected layer are erased, and the last three layers are determined by retraining. Layer parameter information to achieve the classification goal.

Transfer learning shows that using the similarity between data, tasks or models, the model learned in the previous field is appropriately transferred to the target task, making the model more personalized and the task more concise. The simplest migration method in deep network migration learning is fine-tuning, that is, adjusting the trained network model for the tasks of this article. The advantages of fine-tuning are obvious. First, it reduces time consumption to a large extent. Second, the model is pre-trained on a huge dataset and has stronger robustness and generalization capabilities. Using existing big data and optimizing model training target data to get better results is a process of transfer learning.

This paper uses initialize our network by the weights of the VGG16 model, and fine-tune the model to learn the extraction of specific features of medical images through a small amount of skin cancer images. Specifically, we improve VGG16 to extract features of multiple scales, because VGG16 can extract various visual features elegantly, simply and effectively. The original VGG16 convolutional neural network model consists of 19 layers. The size of the input image is $224\times224\times3$, the size of the initial convolution kernel is $3\times3\times3$, the stride size is 1, the effective padding size is 1, and the pooling layer pooling uses 2×2 maximum pooling function max pooling method. The process of convolution in the model is: first use two convolution processing of 64 convolution kernels, then perform a pooling layer pooling, after completion, pass through the convolution of 128 convolution kernels twice, and use one pooling Layer pooling, after three convolutions of 256 convolution kernels, a pooling layer pooling is used, and finally after three 512 convolution kernels convolutions are repeated twice, a pooling layer pooling is performed. After the convolutional layer is processed, it is the three-time fully connected layer. At this time, the parameters that need to be processed in the network have been reduced a lot [5]. However, the classification speed of traditional VGG16 is still slow, and the accuracy of recognition needs to be improved. In this article, we have improved the original VGG16 network. We neglect the FC layer in the original model, and use the output of the pooling layer of the first four convolution blocks and the output of the third convolution layer of the last convolution block as the multi-scale convolution features.

As shown in Figure 1, suppose I denotes the input skin cancer image. The input and output of m-th convolutional layer are I_m and O_m, respectively. Let W_m and b_m denote the weight and deviation term of the convolutional kernel. Then, the operation in each layer is written as:

$$O_m = ReLU\left(I_m * W_m + b_m\right), \tag{10}$$

where * stands for convolution operation. $ReLU$ stands for rectified linear unit activation function, which can be represented as:

$$ReLU(x) = \max(0, x) \quad . \tag{11}$$

In addition, there is a maximum pooling layer behind each convolution block. We obtained five different scale features. These features will be sent to the multi-scale feature fusion step for further processing, so as to obtain discriminative medical image-specific features for skin cancer recognition.

3.3. Multi-Scale Feature Fusion

The starting point of the method in this paper is that multi-scale features can capture receptive fields of different sizes, so as to perform good information extraction on lesions of different sizes and achieve better recognition performance. Therefore, for this purpose, we designed a multi-scale feature fusion step, including reduction-attention module and adaptive fusion module. The former is used to alleviate the problem of dimensional disasters, and the latter is used to fuse multi-scale features adaptively. We will introduce these two parts in detail below.

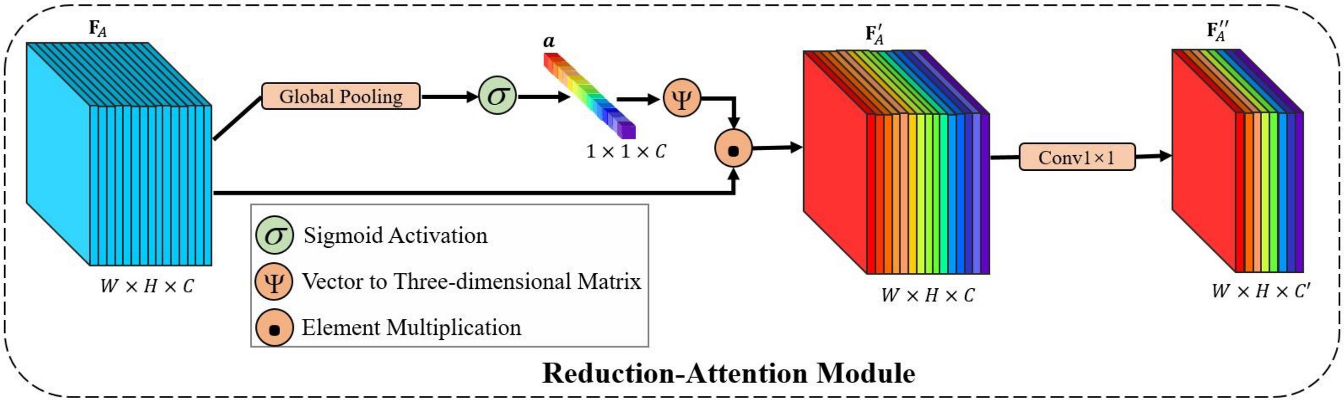

Figure 2. The proposed reduction-attention module.

3.3.1 Reduction-Attention Module

The earliest CNN essentially belongs to a network with spatial attention. Hence, it applies a uniform weight to each channel, resulting in only learning the local connection mode. Recent work has proved that the information in each channel is various [41]. Therefore, it is necessary for us to assign weights that match the amount of information of each channel. Based on the SE network [42], we designed a new reduction-attention module, which assign each channel different weight by learning the channel descriptors. Besides, in order to alleviate the problem of dimensionality disaster in learning, our reduction-attention module also includes an adaptive dimensionality reduction architecture.

As shown in Figure 2, first, we use a 1×1 convolution operation to reduce the dimensionality of the original input feature $O_A \in \mathbf{R}$. Meanwhile, after the operation of the 1×1 convolutional layer, the interaction of different channels is further enhanced. Then, a channel descriptor is obtained by compressing the information in each channel. Second, the obtained channel descriptor are transformed as channel weights $\mathbf{a} = [a_1, a_2, \boxed{?}, a_{C'}] \in \mathbf{R}^{C'}$. To this end, we need to capture the correlation between each channel and learn the non-mutual relationship between them. The $c'-\mathrm{th}$ term in $\mathbf{a}$ is represented as:

$$a_{c'} = sigmoid\left(\frac{1}{W\times H}\sum_{s=1}^{W}\sum_{k=1}^{H} o'_{c'}(s,k)\right), \tag{12}$$

where $o'_{c'}(s,k)$ represents the spatial response of $o'_{c'}$ at (s,k).

Finally, the channel-weighted features can be expressed as:

$$O''_A = \Upsilon\left(O_A\right) = \Gamma(\mathbf{a}) \boxed{?} O'_A = \overline{W} \quad \boxed{?}O'_A, \tag{13}$$

where Υ stands for reduction-attention operation, $\boxed{?}$ stands for Hadamard product, Γ stands for the operation of expanding the vector into a 3-D matrix, and $\overline{W}$ stands for the weight matrix.

To demonstrate the effectiveness of the reduction-attention module, a comparative experiment is conducted (see Table 2).

3.3.2 Adaptive Fusion Module

As depicted in Figure 2, inspired by the DenseNet [43], the adaptive fusion module is proposed. First, we obtain 5-scale features $\left[F_1,F_2,F_3,F_4,F_5\right]$ from the multi-scale feature extraction step. In this adaptive fusion module, we obtain the final fusion feature G_1 through the concatenation manner in different scale. Specifically, in order to learn G_1, we first reduce the number of channels of features $F_1,F_2,F_3,F_4,$ and F_5 through the designed reduction-attention module. Then, the multi-scale features other than F_1 are gradually up-sampled by deconvolution operation. Finally, F_1 and other up-sampling features are adaptively fused to obtain the final fusion feature G_1. Assuming there are *J* branches in total, the process of obtaining fusion features through densely connected modules is expressed by:

$$G_k=\begin{cases}\Lambda^{[k-2]}\left(F_j\right), & if\ k=J\\ \Upsilon\left(\sum_{i=j}^{J-1}w_i^k R_i^{i-k}+R_J^{J-1-k}\right), & else\ if\ k=1\\ \Lambda^{[k-1]}\left(\sum_{i=j}^{J-1}w_i^k R_i^{i-k}+R_J^{J-1-k}\right), & otherwise\end{cases}, \tag{14}$$

where $\Lambda^{[k-2]}$ and $\Lambda^{[k-1]}$ represent the cascade of deconvolution and reduction-attention operation, and the superscript represents the number of the operation. R_i^{i-k} indicates the activation characteristics of F_i after $\Lambda^{[i-k]}$. R_i^{J-1-k} represents the activation characteristics of F_i after $\Lambda^{[J-1-k]}$. w_i^k represents the connection weight. It should be noted that $\Lambda^{[0]}=1$, $R_k^0=F_k$.

To demonstrate the effectiveness of the adaptive fusion module, a comparative experiment is conducted (see Table 2).

3.4. Skin Cancer Recognition

After obtaining the multi-scale fusion feature through the above steps, we adjust its spatial dimension to 1 through global average pooling, which is converted into a feature vector. Then, it is input into 3 FC layers for the skin cancer recognition. Specifically, the neuron number of the FC layers are 1024, 1024, and 8, respectively.

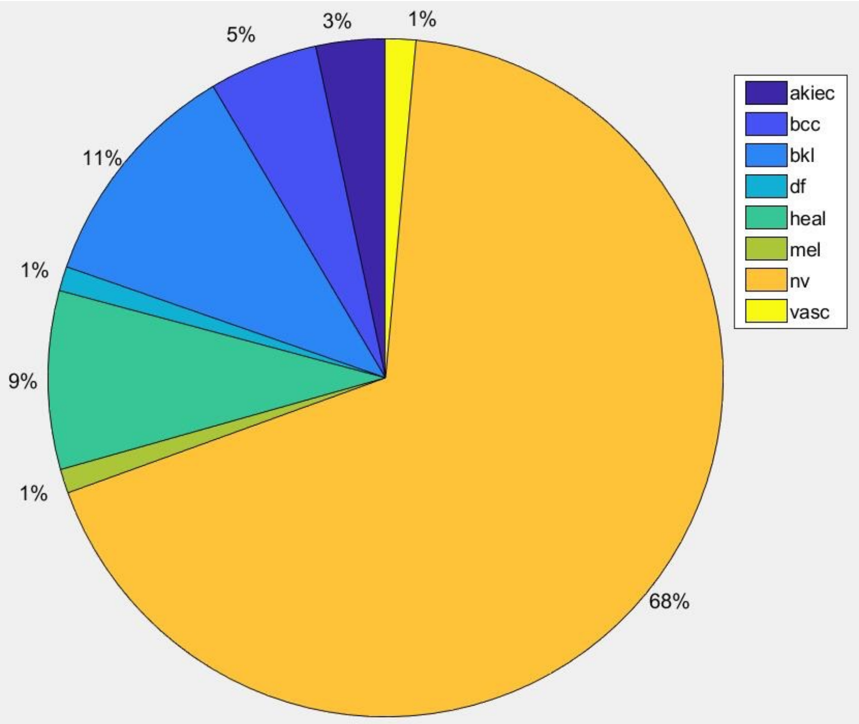

Figure 3. The data distribution in HAM10000+ dataset.

Considering that skin cancer data has serious imbalance in sample categories, as shown in Figure 3, we must solve this problem to improve recognition performance. Traditionally, the cross-entropy loss is utilized to train the CNN model. The formula of the cross-entropy loss can be represented as follows:

$$L_{CE} = -\sum_{k=1}^{K} y_k \log(p_k) , \tag{15}$$

where *K* represents the total number of categories, y_k represents the probability distribution of the true label, and p_k represents the predicted probability distribution. As shown in the following formula, when *k* belongs to the real label, $y_k = 1$; otherwise, $y_k = 0.$.

$$y_k = \begin{cases} 1 & (k = true \quad label) \\ 0 & (k \neq true \quad label) \end{cases} . \tag{16}$$

To solve the problem of unbalanced sample categories in the skin cancer recognition, we designed a weight loss function. Our weight loss function loss aims to solve the problem of sample category imbalance by reducing the weight of simple samples. The contribution to the loss is decided by the number of each class. The few classes of samples are more focused. Specifically, we introduce a weighting factor $(1 - p \times y)^{\gamma}$, where the weighting parameter $\gamma \geq 0$ is adjustable. The designed weight loss is written as follows:

$$L = (1 - p \times y)^{\gamma} \log(p) , \tag{17}$$

where the weight parameter γ is used to control the weight reduction rate of simple samples. In particular, when $\gamma = 0$, L is equal to L_{CE} . The importance of the weighting factor $(1 - p \times y)^{\gamma}$ increases with γ . The loss function has two important characteristics: 1) When a sample is classified falsely, the weighting factor is nearly to 1. Then, the loss can not be affected. 2) When it increases to 1, the weighting factor is nearly to 0, and the loss of properly classified samples is improved. In general, the weighted loss function is to use a more appropriate function to count the contribution of samples that are difficult to classify and easy to classify to the total loss, increase the contribution of the easily classified skin cancer image to the total loss, and reduce the number of comparisons. The contribution of other skin disease images that are large, but easy to align to the total loss. We use the weighted loss of $\gamma = 2$ in our experiment, which greatly improves the accuracy of skin cancer recognition.

To demonstrate the effectiveness of the designed weighted loss, we conducted a comparative experiment, and the experimental results are seen in Table 2.

3.5. Optimization Strategy

Based on the weighted loss, the proposed MSAF-Net can be optimized as follows. The parameters of the revised VGG16 is initialized by the VGG16 model. The other parameters are randomly initialized from the truncated_normal distribution. During each training epoch, the optimization process is composed of five main steps. First, the training skin cancer images are input into the multi-scale feature extraction step to obtain multi-scale feature, which contain different information with multiple receptive fields. Second, the extracted multi-scale features are used as the input of the multi-scale feature fusion step to adaptively fuse them. The fused feature integrates different information and is more discriminative. Third, the semantic label of the input skin cancer image is predicted. Fourth, calculate the weighted loss based on the predicted semantic label and true semantic label according to Eq. 3. Finally, update all the parameters by minimizing the weighted loss. When a fixed iterative epoch is reached, the training phase process is terminated. Then, all the parameters are used to predict the semantic label of each testing skin cancer images. The detailed information about the optimization process of the MSAF-Net can be seen in Algorithm 1.

```
Algorithm 1 The proposed MSAF-Net
Input:
    Training connection $D^{tr}=\{I_t^{tr}, y_t^{tr}\}_{t=1}^{T}$ of the skin cancer images and the corresponding true semantic label;
    Testing connection $D^{te}=\{I_{t'}^{te}\}_{t'=1}^{T'}$ of the skin cancer images;
    Learning rate $lr$ and iterative epoch $E$.
Output:
    All the to-be-learned parameters of the proposed method;
    The predicted semantic label $\{p_{t'}^{te}\}_{t'=1}^{T'}$.
Initialization:
    The weights of the revised VGG16 are initialized from the original VGG16 model, and other weights are initialized from the truncated_normal distribution.
Repeat:
    1: Calculate multiscale features $F_j$ by forward propagation of the revised VGG16 model according Eq. 1;
    2: Fuse multi-scale features using the densely connected module according to Eq. 2;
    3: Predict the semantic label of the input skin cancer image;
    4: Calculate the weighted loss according to Eq. 3;
    5: Update all the parameters using the SGD-Prop.
Until: A fixed iterative epoch $E$.
Return: All the to-be-learned parameters and the predicted semantic label $\{p_{t'}^{te}\}_{t'=1}^{T'}$.
```

4. Experiments and Results

This section introduced the dataset, evaluation metrics, training details, and experimental results in details.

4.1. Dataset

Based on the released skin lesion image HAM10000 challenge dataset, this article adds a class of normal skin images and creates a HAM10000+ dataset to verify the method. The HAM10000 dataset is the world's most widely released dermal replication image dataset. The dataset contains 10015 skin cancer images. In this article, we downloaded 838 images of normal skin tissues including moles through a search engine, and trained the model to distinguish normal skin images from skin cancer. Therefore, the HAM10000+ dataset contains 7 types of skin cancer and normal skin tissue types, a total of 8 categories: photokeratosis and intraepithelial carcinoma (akiec), benign. Corneal lysis (bkl), basal cell carcinoma (bcc), dermatofibroma (df), vascular skin disease (vasc), melanoma (Nv), melanoma (mel), and health (heal). In this article, we randomly select 80% as the training set, and 20% as the testing set. Figure 4 shows some of the HAM10000+ dataset.

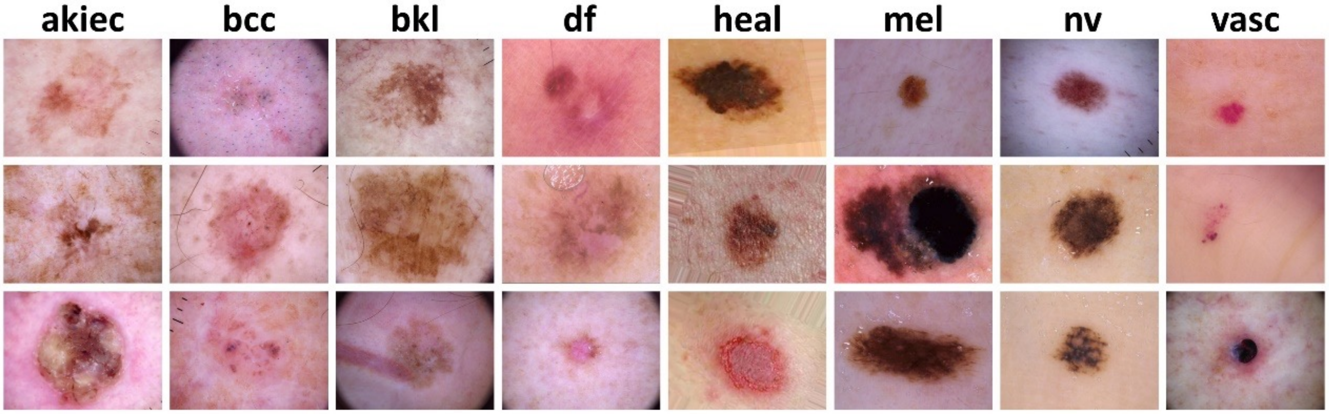

Figure 4. Some cases of the HAM10000+ dataset.

4.2. Evaluation Metrics

The evaluation metrics we use are accuracy, sensitivity, specificity, and confusion matrix, which are commonly adopted in image classification. To evaluate the classification quality of a classifier, four basic metrics, including TP, TN, FP, and FN, are calculated. Based on the four basic metrics, accuracy can be computed as (TP + TN) / (TP + TN + FP + FN), which means total correctly classified samples divided to total samples. Sensitivity can be computed by TP / (TP + FN). This means that the number of passing generals is fixed, and the number of regular samples is the same as the number of regular samples. Speciality can be computed by TN / (TN + FP) calculation. This means that the number of available data that can be used for passing is fixed, and that the number of data is the same as the number of data that can be used. The confusion matrix demonstrates confusions and error in different classes by counting the total number of correct and incorrect classification images of each type.

4.3. Implementation Details

The detailed settings in the proposed MSAF-Net are introduced as follows. The learning rate is 0.005. For every 1000 iterations, the learning rate is divided by 10. The momentum and weight decay parameters in the training phase are 0.0002 and 0.8, respectively. The batch size is 64. The iteration period E is set to 4000. The parameters of the revised VGG16 are initialized by the pre-trained VGG16 model. Other parameters are randomly initialized from the truncated_normal distribution. All experiments are performed on a PC with TITAN X GPU equipped with tensorflow.

4.4. Results and Discussion

In this section, we will introduce the experimental results and conduct some analysis. Specifically, the experimental results are presented in three aspects: 1) model convergence, 2) comparison with other methods, and 3) analysis of important factors. The details are given as follows.

4.4.1 Model Convergence

To verify the convergence of the MSAF-Net, we recorded the loss changes and accuracy changes of the model during the training process. Figure 5 shows the changing curve of training loss, which reveals the proposed MSAF-Net can converge quickly, approximately at 3500 iterations. Meanwhile, Figure 5 shows the changing curve of training accuracy. From the figure, we can also find the proposed MSAF-

Net can converge quickly. In the end, the training accuracy is converge at 98.5%. In summary, we can conclude that our model can converge quickly.

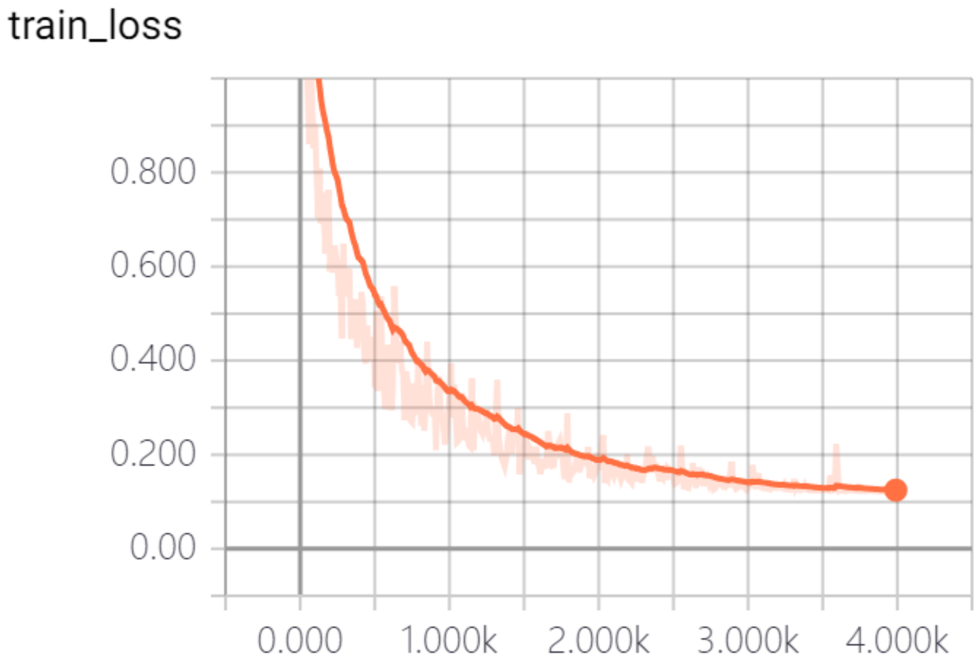

Figure 4. The changing curve of training loss.

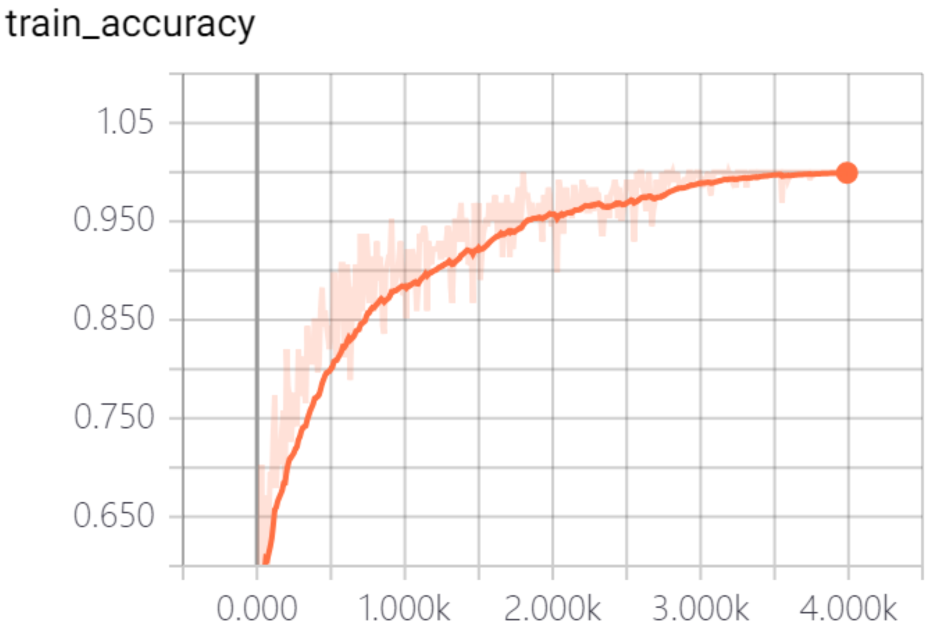

Figure 5. The changing curve of training accuracy.

4.4.2 Comparison with Other Methods

To prove the superiority of the proposed MSAF-Net, we report the comparison results. The comparison methods can be divided into two categories: 1) skin cancer recognition methods on the basis of handcrafted features, and 2) skin cancer recognition methods on the basis of deep learning features. The former methods include CTS [23], snake+SVM [32], and GLCM [35] methods. The latter methods include local-CNN [20], transfer-CNN [26], full-volume-CNN [27], MobileNet [29], ResNet-101 [33], Inception-v3 [33], and GAN [34] methods. The CTS method uses a threshold-based segmentation method to segment the skin loss area, and then uses SVM to classify the image. The snake+SVM method was designed as a segmentation scheme based on combination of snake model and SVM, and applied SVM in finding an appropriate initial curve and parameters for snake algorithm. The GLCM method adopted normalized symmetrical GLCM to learn texture features based on vector machine as a model for skin cancer classification. The local-CNN method extracted local convolution features from the deep residual network for the skin cancer recognition. The transfer-CNN method used the idea of knowledge migration to train the DNN network for the skin cancer recognition. The full-volume-CNN method used AlexNet-based full-volume neural networks to classify non-skin mirror melanoma images. The MobileNet method was trained over the HAM1000 skin lesion dataset. The ResNet-101 and Inception-v3 methods are trained for the skin cancer recognition task. The GAN method produced unseen skin cancer images for training the skin cancer recognition model. The quantitative result is reported in Table 1. The qualitative result is shown in Figure 6.

Table 1. The quantitative result compared with other methods.

Method	Sen(%)	Spec(%)	ACC(%)
CTS	31.10	92.75	79.53
snake+SVM	31.23	92.41	80.06
GLCM	32.34	93.39	81.11
local-CNN	54.37	89.56	86.52
transfer-CNN	56.88	90.87	85.86
full-volume-CNN	57.18	91.85	88.37
MobileNet	61.87	94.82	92.63

ResNet-101	62.41	94.16	94.29
Inception-v3	63.83	94.15	93.13
GAN	65.16	95.52	95.26
MSAF-Net	**69.32**	**98.56**	**97.46**

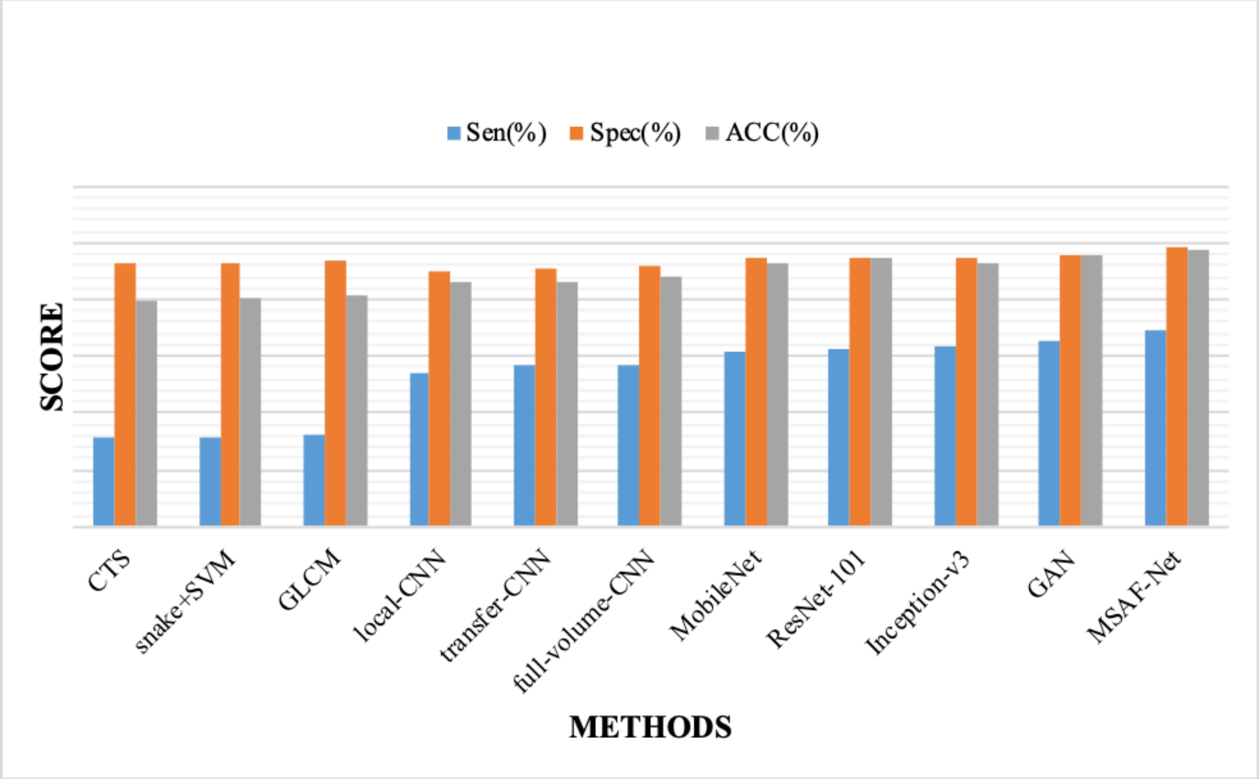

Figure 6. The qualitative result compared with other methods.

From Table 1 and Figure 6, we can apparently notice that the proposed MSAF-Net outperform other methods on all of the evaluation metrics, which indicates the multi-scale features and weighted loss is effective for improve the performance of skin cancer recognition. Meanwhile, it can be observed that compared with the skin cancer recognition methods on the basis of handcrafted features, the MSAF-Net improves about 16% in accuracy, which indicates that deep convolution features have more powerful expressive power than low-level manual feature descriptors. In addition, Figure 7 shows the confusion matrix of MSAF-Net for skin cancer recognition. It is obviously that the proposed method achieves considerable results, especially for the recognition of nv, heal, and vasc classes.

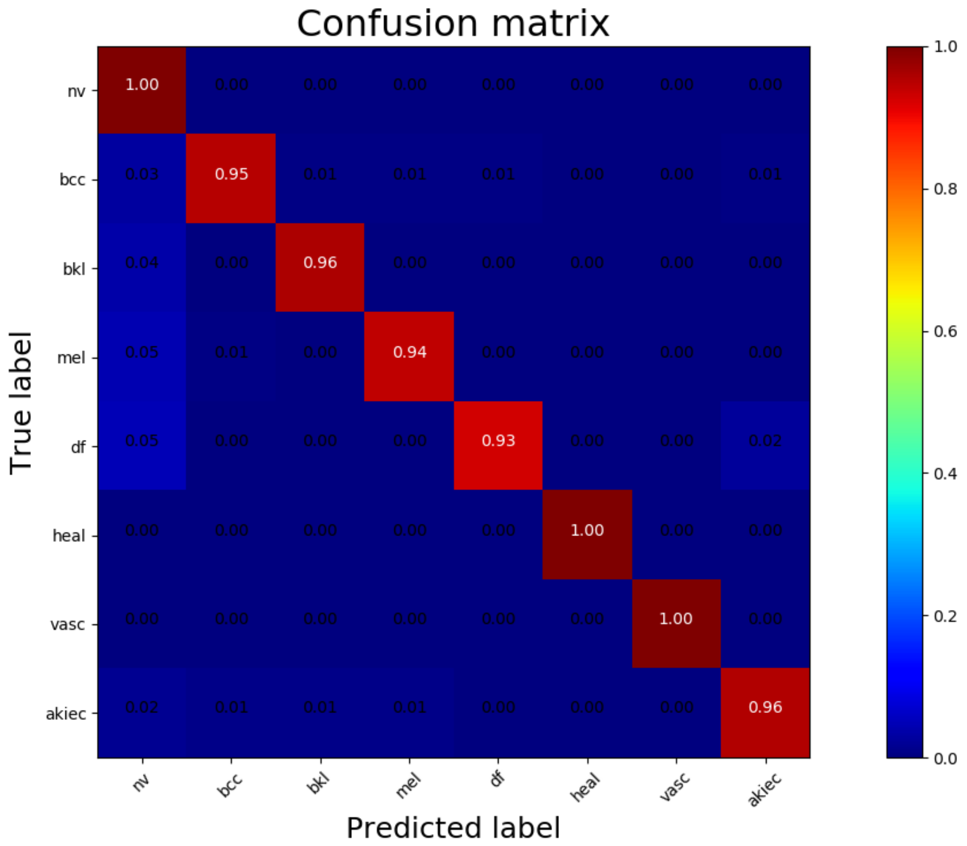

Figure 7. The confusion matrix of the proposed method for skin cancer recognition.

4.4.3 Analysis of Important Factors

In order to demonstrate the importance of the designed reduction-attention module, densely connected module, and weighted loss, we conduct three variations of the proposed method. Specifically, the reduction-attention module is omitted to prove its importance (O-RA). The densely connected module is replaced with concatenation operation to testify its effectiveness (O-DC). The weighted loss is replaced with cross entropy loss to testify its effectiveness (O-WL). The results are reported in Table 2.

Table 1. The quantitative result compared with other methods.

Method	Sen(%)	Spec(%)	ACC(%)
O-RA	60.37	84.51	80.25
O-DC	58.31	83.49	78.22
O-WL	64.43	91.57	90.40
MSAF-Net	**69.32**	**98.56**	**97.46**

From the observation to Table 2, we can find the proposed surpass other variations to a great extent. The accuracy improves from 80.25% to 97.46% when the reduction-attention module is adopted. The accuracy improves from 78.22% to 97.46% when the densely connected module is applied. The accuracy improves from 90.46% to 97.46% when the weighted loss is adopted. They demonstrated the effectiveness of the designed reduction-attention module, densely connected module, and weighted loss, respectively.

5. Conclusions

In this research, we propose a multi-scale adaptive fusion network for skin cancer recognition. The network solves the problem of scale variability in medical images by learning multi-scale features, and solves the problem of sample imbalance in medical images by designing a weighted loss function. Through these two improvements, the proposed model greatly improves the accuracy of skin cancer recognition. Through experiments on the amplified HAM10000+ dataset, the experimental results on the amplified HAM10000+ dataset show that the method can effectively learn multi-scale information, solve the problem of unbalanced medical image samples, and further improve the accuracy of skin cancer recognition.

References

[1] R. Kasmi and K. Mokrani, “Classification of malignant melanoma and benign skin lesions: implementation of automatic abcd rule,” IET Image Processing, vol. 10, no. 6, pp. 448-455. 2016.

[2] M. Celebi, H. 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.

[3] Craythorne, E., & Al-Niami, F. (2017). How to examine a patient with skin cancer. Medicine, 45(7),

429–430. doi:10.1016/j.mpmed.2017.04.002.

[4] Friedman RJ, Rigel DS, Kopf AW. Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. CA Cancer J Clin 1985; 35: 130e51.

[5] Taber, J. M., Dickerman, B. A., Okhovat, J.-P., Geller, A. C., Dwyer, L. A., Hartman, A. M., & Perna, F. M. (2018). Skin cancer interventions across the cancer control continuum: Review of technology, environment, and theory. Preventive Medicine, 111, 451–458. doi:10.1016/j.ypmed.2017.12.019.

[6] Guy G, Thomas C, Thompson T, Watson M, Massetti G, Richardson L. Vital signs: melanoma incidence and mortality trends and projections - United States, 1982-2030. MMWR Morbid Mortal Wkly Rep. 2015;64(21):591-596.

[7] Singer, S., Tkachenko, E., Yeung, H., & Mostaghimi, A. (2020). Skin Cancer and Skin Cancer Risk Behaviors Among Sexual and Gender Minority Populations: A Systematic Review. Journal of the American Academy of Dermatology. doi:10.1016/j.jaad.2020.02.013 "Skin Cancer Treatment (PDQ®)". NCI. 25 October 2013. Archived from the original on 5 July 2014. Retrieved 30 June 2014.

[8] Elgamal, M. (2013). Automatic skin cancer images classification. IJACSA) International Journal of Advanced Computer Science and Applications, 4(3), 287-294.

[9] Gallagher RP, Lee TK, Bajdik CD, Borugian M (2010). "Ultraviolet radiation". Chronic Diseases in Canada. 29 Suppl 1: 51–68. PMID 21199599.

[10] Dubas LE, Ingraffea A (February 2013). "Nonmelanoma skin cancer". Facial Plastic Surgery Clinics of North America. 21 (1): 43–53. doi:10.1016/j.fsc.2012.10.003. PMID 23369588.

[11] "General Information About Melanoma". NCI. 17 April 2014. Archived from the original on 5 July 2014. Retrieved 30 June 2014.

[12] Housman, T. S., Feldman, S. R., Williford, P. M., Fleischer Jr, A. B., Goldman, N. D.,

Acostamadiedo, J. M., & Chen, G. J. (2003). Skin cancer is among the most costly of all cancers to treat for the Medicare population. Journal of the American Academy of Dermatology, 48(3), 425-429.

[13] Song Youyi， Zhang Ling， Chen Siping， et al. Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning [J]. IEEE Transactions on Biomedical Engineering， 2015， 62(10): 2421- 2433.

[14] Krizhevsky A， Sutskever I， Hinton G. Imagenet classification with deep convolutional neural networks [C] / / The 19th International Conference on Neural Information Processing Systems (ICONIP 2012). Doha: Curran Associate Inc， 2012: 1097-1105.

[15] Lecun Y， Bengio Y， Hinton G. Deep learning [J]. Nature， 2015， 521(7553): 436-444.

[16] Melanoma Detection Techniques in Dermoscopy Images Based on Deep Learning, Bai Pengcheng, 201520131506.

[17] Celebi M E, Kingravi H A, Uddin B, et al. A methodological approach to the classification of dermoscopy images[J]. Computerized Medical Imaging & Graphics, 2007, 31(6):362- 373.

[18] Ballerini L, Fisher R B, Aldridge B, et al. A Color and Texture Based Hierarchical K-NN Approach to the Classification of Non-melanoma Skin Lesions[J]. 2013.

[19] Schaefer G, Krawczyk B, Celebi M E, et al. An ensemble classification approach for melanoma diagnosis[J]. Memetic Computing, 2014, 6(4):233-240.

[20] Melanoma Recognition in Dermoscopy Images via Deep Residual Network, Li Hang Yu Zhen, Vol. 37 No. 3, 2018.

[21] Gloster H M, Neal K W. Skin cancer in skin of color[J]. Journal of The American Academy of Dermatology, 2006, 55(5): 741-760.

[22] Argenziano G, Fabbrocini G, Carli P, et al. Epiluminescence Microscopy for the Diagnosis of Doubtful Melanocytic Skin Lesions: Comparison of the ABCD Rule of Dermatoscopy and a New 7-Point Checklist Based on Pattern Analysis[J]. Archives of Dermatology, 1998, 134(12):1563-1570.

[23] Celebi M E, Kingravi H A, Uddin B, et al. A methodological approach to the classification of dermoscopy images[J]. Computerized Medical Imaging & Graphics, 2007, 31(6):362- 373.

[24] Sumithra R, Suhil M, Guru D S. Segmentation and Classification of Skin Lesions for Disease Diagnosis [J]. Procedia Computer Science, 2015, 45:76-85.

[25] Codella N, Cai J, Abedini M, et al. Deep Learning, Sparse Coding, and SVM for Melanoma Recognition in Dermoscopy Images[J]. 2015:118-126.

[26] Pomponiu V, Nejati H, Cheung N M. Deepmole: Deep neural networks for skin mole lesion classification[C]// IEEE International Conference on Image Processing. IEEE, 2016:2623- 2627.

[27] Kawahara J, Bentaieb A, Hamarneh G. Deep features to classify skin lesions[C]// IEEE, International Symposium on Biomedical Imaging. IEEE, 2016:1397-1400.

[28] Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database[C]// Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 65 2009:248-255.

[29] Haseeb Younis et al. , Classification of Skin Cancer Dermoscopy Images using Transfer Learning, 978-1-7281-5404-6 2019 IEEE

[30] Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097-1105.

[31] H. Haenssle et al., "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," vol. 29, no.

8, pp. 1836-1842, 2018.

[32] Prachya et al. , Detection Skin Cancer Using SVM and Snake Model 978-1-5386-2615-3, 2018 IEEE

[33] Ahmet Demer et al., Early Detection of Skin Cancer Using Deep Learning Architectures: Resnet-101 and Inception-v3, 978-1-7281-2420-9, 2019 IEEE.

[34] Pooyan Sedigh et al., Generating Synthetic Medical Images by Using GAN to Improve CNN Performance in Skin Cancer Classification, 978-1-7281-6604-9 2019 IEEE.

[35] Ritesch Maurya et al., GLCM and Multi Class Support Vector Machine based Automated Skin Cancer Classification, 978-93-80544-12-0, 2014 IEEE.

[36] Radu Dobrescu et al., Medical images classification for skin cancer diagnosis based on combined texture and fractal analysis, ISSN: 1109-9518, 2010.

[37] Rokad, B., & Nagarajan, D. (2019). Skin Cancer Recognition using Deep Residual Network. arXiv preprint arXiv:1905.08610.

[38] Sharma, S. , & Mehra, R. . (2020). Conventional machine learning and deep learning approach for multi-classification of breast cancer histopathology images—a comparative insight. Journal of Digital Imaging, 33(3), 632-654.

[39] Wajire, P. , Angadi, S. , & Nagar, L. . (2020). Image Classification for Retail. 2020 International Conference on Industry 4.0 Technology (I4Tech). IEEE.

[40] Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. international conference on learning representations.

[41] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2019). ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. arXiv: Computer Vision and Pattern Recognition.

[42] Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2019). Squeeze-and-Excitation Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1-1.

[43] Huang, G., Liu, Z., Der Maaten, L. V., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. computer vision and pattern recognition.

致谢信

从高一阅读Brain facts开始对脑神经科学有了浓厚的兴趣，后来在黄婧老师的鼓励和帮助下开始了进一步的探索，在阅读探索脑和一些最新的书籍概念的时候，发现了自己对神经网络的兴趣，于是决定进一步探索并有所运用。

感谢黄老师对我的鼓励和引导，让我对将计算机和神经科学结合的课题有信心去学习。感谢百度百科上各种介绍神经网络比如CNN知识的文章。感谢张老师对我深度学习知识以及表达上的帮助和答疑，以及对学习困惑中的书籍推荐。通过学习计算机，尤其是深度学习的概念及运用，更好地帮助我理解脑神经中生物机制的一些概念。

最后，感谢丘成桐中学科学奖，让我能够在自己所热爱的领域不断地探索并且做出贡献。之前对于脑神经的了解仅限于生物机制上，而这次通过学习深度学习的知识和统计学习等方法，来提升一个图片分类中存在的现实问题，以运用的方式来加深我对知识的理解，让我收获颇丰。在对深度学习的学习和探索中，也了解到了计算机对于生物探索的重要帮助。

谢谢在这次课题研究中给我提供指导和支持的老师，朋友和家人们！

参赛队员介绍

付鑫雨，女，重庆南开中学国际部高三学生，喜欢探索，爱好广泛，对生物，物理，计算机，文学，哲学感兴趣，热爱脑神经科学。

- 2020 年 全国脑神经科学大赛China Brain Bee 一等奖
- 2019-2020 China HOSA 全国性学生组织社团管理 副主席
- 2020 年 英美生物奥林匹克竞赛银奖获得者
- 2020 John Locke 写作比赛法律组shortlist
- 2020 年 NSDA TOC 全美冠军赛 中国决赛全国第13名
- 2019-2020 年 NSDA 深圳赛区决赛原创演讲第三名
- 2019 年 澳大利亚奥林匹克竞赛 生物全球金奖，物理全球金奖，化学全球银奖，环境科学全球银奖
- 2019 年 Bpho 英国物理奥林匹克竞赛 铜奖一
- 2019 年 HOSA 国家地理学术测评 ATC 大学物理 全球 Shortlist（Top20）
- 2019 年 南开大学文学院优秀中学生论坛发言人代表