Convolutional Neural Network Based on Regional Features and Dimension Matching for Skin Cancer Classification*

Zhichao SHA[†], *Nonmember*, **Ziji MA**[††], *Member*, **Kunlai XIONG**[†a)], **Liangcheng QIN**[†], and **Xueying WANG**[†], *Nonmembers*

SUMMARY   Diagnosis at an early stage is clinically important for the cure of skin cancer. However, since some skin cancers have similar intuitive characteristics, and dermatologists rely on subjective experience to distinguish skin cancer types, the accuracy is often suboptimal. Recently, the introduction of computer methods in the medical field has better assisted physicians to improve the recognition rate but some challenges still exist. In the face of massive dermoscopic image data, residual network (ResNet) is more suitable for learning feature relationships inside big data because of its deeper network depth. Aiming at the deficiency of ResNet, this paper proposes a multi-region feature extraction and raising dimension matching method, which further improves the utilization rate of medical image features. This method firstly extracted rich and diverse features from multiple regions of the feature map, avoiding the deficiency of traditional residual modules repeatedly extracting features in a few fixed regions. Then, the fused features are strengthened by up-dimensioning the branch path information and stacking it with the main path, which solves the problem that the information of two paths is not ideal after fusion due to different dimensionality. The proposed method is experimented on the International Skin Imaging Collaboration (ISIC) Archive dataset, which contains more than 40,000 images. The results of this work on this dataset and other datasets are evaluated to be improved over networks containing traditional residual modules and some popular networks.

key words: *skin cancer, medical image classification, convolutional neural network, ResNet, ISIC Archive, dimension and region feature matching*

1.   Introduction

The causes of skin cancer are diverse, and the main factors include exposure to ultraviolet radiation, exposure to ionizing radiation, genetic factors [1], and environmental triggers, etc. Early detection of skin cancer has better treatment effect for the later stage of skin cancer [2]. Recently, the diagnosis of skin tumors through dermatoscope imaging is a common and basic method in medicine. Since the diverse and ununiform appearance of skin tumors, it is often difficult to accurately determine the exact result based on the knowledge and experience of the medical professional alone [3]. In recent years, models obtained by training CNNs have surpassed dermatologists in distinguishing melanomas and nevi. Their advantages, such as short decision time, high ac-

curacy, low cost, and ease of system updates and upgrades, are propelling the development of medical detection systems to a new stage [4], [5].

At the beginning of the 21st century, Hinton proposed the deep neural network algorithm, which greatly improved the capability of artificial neural networks and became the core cornerstone of artificial intelligence. In 2015, the residual network (ResNet) proposed by He et al. [6] from Microsoft Research has achieved extraordinary results in many fields of ImageNet Large Scale Visual Recognition Challenge (ILSVRC), and further promoted the neural network to a deeper network.

With the rapid development of deep learning and other technologies, the ability of Artificial Intelligence (AI) is increasing, and the applicable fields are becoming more and more extensive [7]–[9]. The application of AI algorithms in the diagnosis of medical images has gradually become widespread.

2.   Related Works

Nowadays, humans have been battling with cancer for many years and invoking AI techniques can help not only to improve detection but also help explore the task of finding relevant drugs and predictions [10], [11]. Wei et al. proposed a CNN for automatic segmentation of retinal blood vessels [12]. The CNN used genetic algorithm to search for fewer parameters corresponding to better architecture, and obtained a very compact model, which was conducive to clinical application. However, the widespread and comprehensive use of AI in healthcare still needs to be promoted through further development because there are some problems behind it that need to be addressed. These include the existence of irregularities in data creation and diagnostic procedures, the inadequate use of medical data, and suboptimal results in AI diagnostics, among others [13], [14].

The main work of this paper is to classify the images of skin lesions. We focus on two ways to address the problems of poor fusion feature reinforcement and loss of data information extraction in ResNet. 1). For the difference in the dimensionality of the information content of the two propagation paths of the building block, adding convolutional layers on the branch to make it equal to the dimensionality of the main path. 2). For the problem of ignoring relevant features for convolution with stride 2, the convolution kernels are divided into multiple groups. Different groups of

convolution kernels are applied to different regions of the feature map to extract features.

This paper presents dimension feature matching and multi-dimension feature matching structures. The ResNet using these algorithmic structures are named dimensional matching ResNet (DM ResNet), regional feature matching ResNet (RFM ResNet), dimensional feature matching ResNet (DFM ResNet) and multiple dimensional feature matching ResNet (MDFM ResNet). The optimized residual models achieved better effects for classification of both skin cancer dataset and other datasets. The overall structure of this paper is presented as follows: in the next second section we present the specific structure of the algorithm proposed in this paper. In the fourth section we give experimental results. In the fifth section we compare the analysis of the results between different databases and networks and draw conclusions.

3. Proposed Algorithms

By introducing branch structure in the forward propagation process of the model, ResNet makes the weight parameters return directly through the direct link in the process of back propagation, thus reducing the occurrence of gradient disappearance and explosion, and the network model can enter a deeper layers. The ResNet network model, regardless of its depth, is composed of stacked and connected residual blocks. These blocks typically consist of two or three layers (more in deeper networks) connected through convolutional layers and shortcut connections. The problem that the size of the output feature map is incompatible with the output feature map of the short-circuit link due to the convolutional layer with a step size of 2 in the forward propagation path, it can be solved by adjusting the output of the short-circuit link using a 1*1 convolutional kernel.

In the residual structure block, there are two paths after the input, indicating that there are two kinds of mapping relationships. The one after multi-layer convolution is the residual mapping, which is the residual to be learned, and the other is the identity mapping, which determines whether adjustment is needed according to the dimensionality of the output. X_i as input is mapped by two weight layers and then merged with identity mapping. The output after activation is y.

$$y = f(H(X_i)) * h(X_i)) \tag{1}$$

where y is the input of the next residual block and the overall summation mapping of the previous residual block, and $H(X_i)$ is the residual mapping of the input before the summation, and $L(X_i)$ represents the mapping before activation, which is what the network needs to learn, thus,

$$H(X_i) = L(X_i) - h(X_i) \tag{2}$$

When $H(X_i)$ is set to 0, identity mapping is generated.

$$L(X_i) = h(X_i), (H(X_i) = 0) \tag{3}$$

$$y = H(X_i) + X_i \tag{4}$$

For back propagation, let the loss function be l. Then the back propagation chain is that,

$$\frac{\partial l}{\partial X_0} = \frac{\partial l}{\partial X_k}(1 + \frac{\partial}{\partial X_0} \sum_{i=0}^{i=k-1} H(X_i, W_i)) \tag{5}$$

3.1 Dimensional and Region Feature Matching Algorithms

Considering that the feature maps with different convolution layers in the residual block of the original residual network carry different dimensions of information, the features after direct fusion are not prominent enough. Moreover, the number of 1*1 convolutional layers used in the direct connection layer is one while the step length is two, resulting in multiple rows and multiple columns of pixels being directly ignored, never participating in the convolution calculation, and the pixel points are not fully utilized [15]–[17]. Therefore, this paper proposes that the 1*1 convolution kernel in the short-circuit link is also divided into multiple groups to extract features in different regions to further utilize the ignored information. At the same time, it is also considered to add a convolution layer to improve the information dimension of the final output feature graph, so that the features after the fusion of the two paths become more prominent. Therefore, we propose two basic algorithms, region feature matching and dimensional matching.

3.1.1 Dimension Matching

To highlight the characteristics of the fused information, a corresponding convolution layer is added to the short-circuit links to improve the dimensionality of the output information. The residual module in Fig. 1, the branch adds a convolutional layer compared to the original residual fast, which raises the feature dimension level and matches the features of the main path more closely. Dimension matching considers lifting the number of convolutional layers of a branch to the same or similar number as the main path. Based on the concept of dimensional matching, we further give the concept of multiple dimensional matching.

For the network model with multi-branch structure, we want the dimensionality of the feature information output from each branch to be as close as possible to reduce the fusion of mismatched information tracts to appear new error information. To enhance the information features in deeper networks, we fuse the outputs of each convolutional layer of each path, so the information features of different dimensions are enhanced. In the residual module in Fig. 2, after a convolutional layer is added to the branch, the results of the first convolutional layer of both paths are done a fusion to strengthen the features. Figure 1 and Fig. 2 both take the two-layer convolution structure as an example. The rectangular block in Fig. 1 represents the feature map after convolution or fusion, $M * N$ is its size, D is its depth, and $*4$ is used to denote 4 groups of convolution kernels or feature maps, and

SHA et al.: CONVOLUTIONAL NEURAL NETWORK BASED ON REGIONAL FEATURES AND DIMENSION MATCHING FOR SKIN CANCER CLASSIFICATION
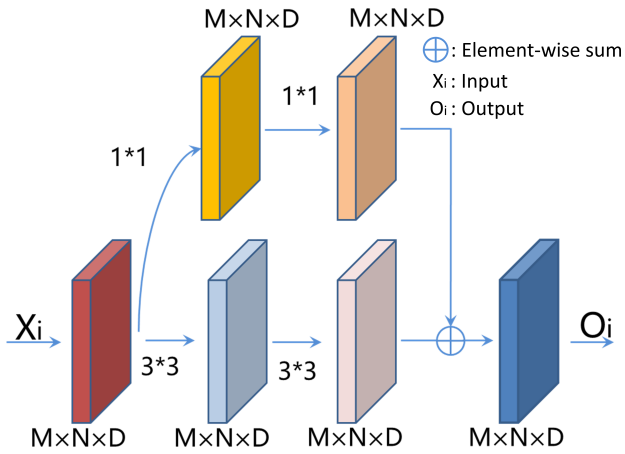
1321

Fig. 1 The input feature map (red) is convolved by two layers to form higher-level features, and the branches are similarly convolved by the same number of convolution layers to form the same or similar level features. The results of the two paths are then fused into the output.
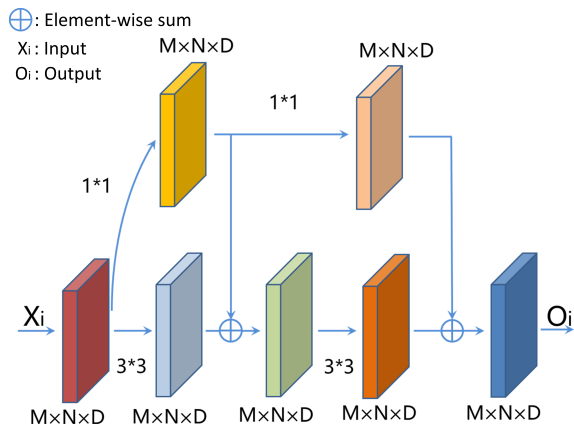

Fig. 2 Multi-dimensional matching structure diagram. The input feature map (red) is convolved in two layers and the branches are also convolved in two layers. The first layer convolution results of the two paths are fused once, and then the respective second layer convolution results are fused again and output. Different coloured blocks represent the results of different operations, followed by the same.

the same meaning in the subsequent figures.

3.1.2 Regional Feature Matching

This paper presents novel optimization strategies to tackle certain challenges encountered in the aforementioned residual networks. As for the advantages brought by the convolution layer with convolution step size of 2, we need to further solve the problems caused by convolution step size of two while maintaining the advantages. The convolution kernel with a step size greater than 1 skips some regions due to the step length jump during the convolution process. To ensure maximum participation of each pixel in the feature map within the convolutional feature extraction, without diminishing the stride size, this paper proposes partitioning a group
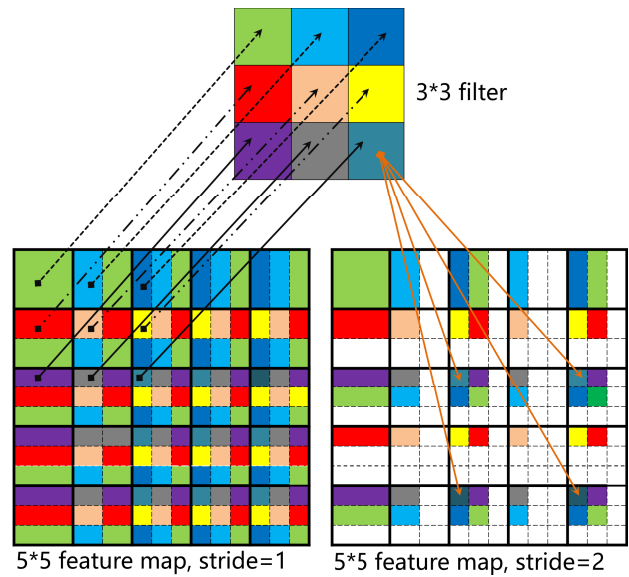

Fig. 3 The upper part shows the 3*3 convolution kernel, and the lower part shows the participation of each pixel of the feature map in the convolution operation when the step size is 1 or 2, with the step size of 1 on the left and 2 on the right. The black arrow indicates the starting convolution position, and the orange arrow indicates the position through which the convolution process passes.

of convolutional kernels, dedicated to convolving the same region, into multiple subsets responsible for extracting features from regions traversed by larger strides. Each group of convolution kernels extracts features in regions complementary to each other with a step length of 2. The total number of convolution does not change, so it no longer consumes additional computing resources. Region feature matching considers that the features of branch and main path extracted by convolution of grouping have regional differences, and the fusion of corresponding regions reduces the loss of valid information. In the following, we take the 3*3 convolution kernels in the residual block as an example to qualitatively show the number of times each pixel in the same feature map is involved in the convolution operation with the convolution kernel step size of 1 or 2, using a 5*5 feature map of size depth of 1. For the 3*3 convolution kernel in the upper part of Fig. 3, different colors represent different weight coefficients. The number of colours in the dotted line segmentation indicates the number of times the pixel point has been involved in the computation throughout the convolution process, which also corresponds to the 9 positional colours in the convolution kernel, i.e., different coloured regions in a pixel represent the weights of the corresponding colours of the convolution kernel that have been involved in the computation. The left panel below Fig. 3 shows the number of times each pixel participates in the calculation when the step size of the convolution kernel is 1. The lower right side of Fig. 3 shows the calculation of each pixel when the step size of the convolution kernel is 2, and the number of participations in the white area is counted as 0. In this paper, we apply multiple sets of convolution kernels to the white area shown in the

bottom of Fig. 3 to extract feature information from the additional area while maintaining the advantage of interval step size. For larger step size convolution, the number of pixels involved in the convolution process shows a rapid downward trend. Insufficient inclusion of pixels in the computation can result in inadequate feature extraction, thereby impeding the network's ability to fully unleash its performance potential. Based on the $M * N$ size feature map, the stride of convolution kernels is S ($2 < S < min(M, N)$), the number of convolution kernels in the current residual block convolution layer is k_s, and the number of convolution kernels used in each newly assigned region is $K_N = f/(S^2 - 1)$ as well as the number of groups is S^2. Let the shape of the input feature map be (M_i, N_i, C_i), C is the number of channels, and the calculation method of the size of the output feature graph (M_o, N_o, C_o) is as follows.

$$M_o = \lfloor \frac{(M_i + 2 * padding - D * (k_s - 1) - 1)}{stride} + 1 \rfloor \quad (6)$$

$$N_o = \lfloor \frac{(N_i + 2 * padding - D * (k_s - 1) - 1)}{stride} + 1 \rfloor \quad (7)$$

where D is the distance parameter of dilation convolution and $\lfloor * \rfloor$ denotes rounding down.

From Fig. 4 and Fig. 5, it can be seen that the convolution step size is 1, the convolution step size is 2, and the group convolution step size is 2. The left figure shows the convolution process of a regular set of 3*3 convolution kernels with step 1, from left to right and then from top to bottom. The right side of Fig. 4 shows a regular set of 3*3 convolution kernels spaced convolution process, which is more sparse. The right side of Fig. 5 shows the convolution process of dividing this set of convolution kernels into 4 sets of convolutions with interval step length, which further increases the density of the convolution process. Staggered grouped convolution improves the feature extraction density without increasing the computational effort.

Let's take an example of a residual block with two convolutional layers. After receiving the input feature map, it is still divided into two paths. The main path continues into the second convolution layer via the first convolution layer (which also includes BN layer, activation layer, and pooling layer), and the convolution step at this time is equal to 2, and the size is 3*3. The convolution kernels are divided into 4 groups on the feature map, and each group of convolution kernels starts from the first column of the first row, the second column of the first row, the first column of the second row, and the second column of the second row, respectively. The convolution operation is performed in steps of two, so that four feature maps can be obtained. The four feature maps are stacked in the lower right direction and fused with the output of the short-circuit connection. The short-circuit link also uses a 1*1 size convolution kernel for the input feature map, and extracts the features in groups with a step size of 2. The output feature maps of the two paths are the corresponding features extracted from the same region, so we named them regional feature matching, as shown in Fig. 6.
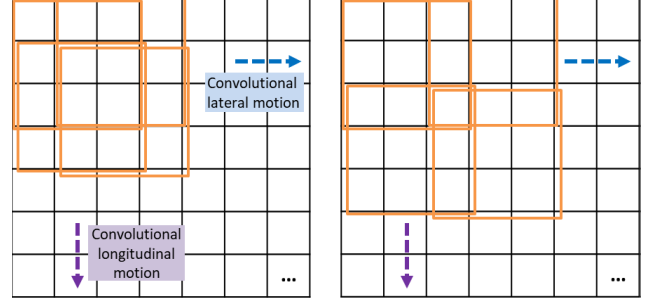

Fig. 4 Convolution step of 1 and 2. Conventional convolution process with step size 1 (3*3, left side) with dense convolution region. Conventional convolution process with step size 2 (3*3, right side) with sparser convolution region.
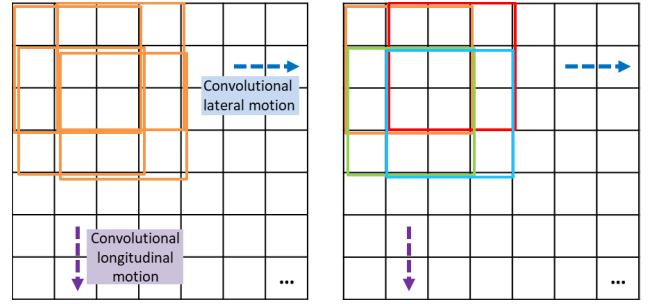

Fig. 5 Convolution kernel grouping and step of 2. Conventional convolution process with step size 1 (3×3, left side), grouped convolution process with step size 2 (3*3, right side), different colours are the starting areas of different groups of convolutions, and the domain convolution region remains dense.
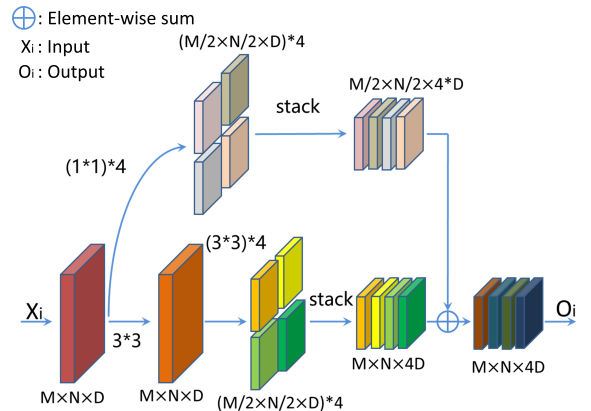

Fig. 6 Regional feature matching struture. For the input feature map (red), feature maps of different regions with the same branches are obtained by grouping convolutions with alternating non-repeating start positions in each group. Finally, the feature maps of two paths corresponding to the same region are fused and output.

3.2 Methodologies

3.2.1 Dimension Feature Matching Structure

We integrate the two fundamental improved algorithms dis-

cussed in Sect. 2.1 to create an enhanced structure applied to the network, encompassing dimension feature matching and its corresponding variant, multi-dimension feature matching. In terms of the information dimensions contained in the feature maps, we aim to keep the dimensions of the two paths as similar as possible. After the feature map is outputted from the preceding residual structure, we append a 1x1 convolutional layer to the subsequent residual block in the branched path. This convolutional layer adopts the same convolution pattern employed in the corresponding convolutional layer of the main path, with a stride of 1. Of course, the second convolution layer does not change the pattern in the next step, stride is 2, and is divided into multiple groups. The results of the two paths are combined and activated together as the output, because at this moment it is the information fusion of similar dimensions, which we name as dimension feature matching, as shown in Fig. 7. The forward propagation calculation of the dimensional feature matching structure block is described as follows,

$$O_i = \sum_1^n (W_2^n \sigma(W_1 X_i) + W_{s2}^n \sigma(X_i W_{s1})) \quad (8)$$

Where, W_{s1} is the first layer of convolution of the branch in the structure block, and W_{s2}^n is the nth convolution group of the second layer of convolution on the branch.

In the reverse update, we also calculate the loss function as C, then,

$$\frac{\partial C}{\partial X_i} = \sum_1^n \left(\frac{\partial C}{\partial h_2^n(X_i)} \frac{\partial h_2^n(X_i)}{\partial h_1(X_i)} \frac{\partial h_1(X_i)}{\partial X_i} + \frac{\partial C}{\partial H_2^n(X_i)} \frac{\partial H_2^n(X_i)}{\partial H_1(X_i)} \frac{\partial H_1(X_i)}{\partial X_i} \right) \quad (9)$$

Among them, $h_1(X_i)$ represents the first-level convolution result of the branch in the dimensional feature matching structure block, and $h_2^n(X_i)$ denotes the nth convolution group result of the second-level convolution on the branch.
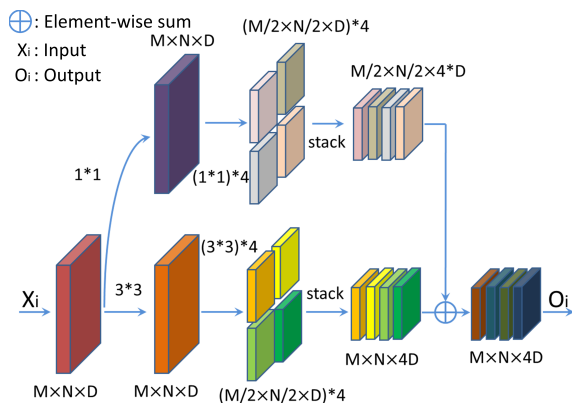
Combine the calculation of $\frac{\partial h_1(X_i)}{\partial X_i} = W_{s1}$ on the branch of forward propagation, therefore,

$$\frac{\partial C}{\partial X_i} = \sum_1^n \left(\frac{\partial C}{\partial h_2^n(X_i)} W_{s1} + \frac{\partial C}{\partial H_2^n(X_i)} \frac{\partial H_2^n(X_i)}{\partial H_1(X_i)} \frac{\partial H_1(X_i)}{\partial X_i} \right) \quad (10)$$

3.2.2 Multiple Dimensional Feature Matching Structure

The final fused data is the output of the second convolution layer of the two paths after batch normalization (BN), while the output results of the first convolution layer of the two paths are not utilized. To further improve the efficiency of data utilization, we propose to merge the first convolution result of the branch with the first convolution result of the main path and continue forward propagation. Although this makes the bifurcated path more complicated, it can also provide more path choices for the network model during back propagation, so that the optimal path can be found to get better results. We named it multidimensional Feature matching, as shown in Fig. 8. In Fig. 8, the main path of input features is convolved by a layer of convolution and then li 4 groups of convolution are crossed to extract features in different regions respectively. The feature maps in the branch paths are also cross-extracted by the second layer of 4 groups of convolution to extract the features of the cross region, and the output is a two-way fusion.

We also give the following description of the output calculation for the multi-dimensional feature matching block.

$$O_i = \sum_1^n ((W_2^n \sigma(W_1 X_i) + X_i W_{s1}) + W_{s2}^n \sigma(X_i W_{s1})) \quad (11)$$

where, $W_1 X_i + X_i W_{s1}$ is the dimensional matching of the first layer of convolution in the two paths in the structure block.

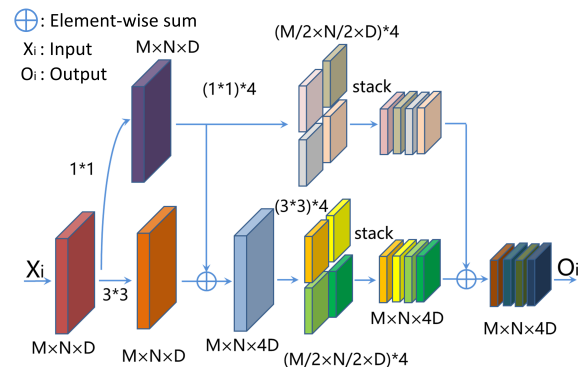The backward propagation is described as follows, counting the loss function as C,

Fig. 7 Dimensional feature matching structure diagram. For the input feature map (red), both paths contain two convolutional layers, the second convolutional layer of each path uses grouped alternating convolution to obtain feature maps of different regions, and the feature maps corresponding to the same region are fused as output.

Fig. 8 Multi-dimensional feature matching structure diagram. For the input feature map (red), both paths contain two convolutional layers, the result of the first convolution for each path is fused first, and then the second convolutional layer is fused as the output using grouped alternate convolutions, corresponding to the same region of the feature map.

Algorithm 1 Multiple dimensional feature matching

Input: $X_i = O_{i-1}$, B=block_size
   // i is the ith residual block and O is the output data
Output: O_i
  initialization;
  for i in B **do**
    if size of X_i != size of O_i **then**
      X_{o1}= Active(BN ($Conv2D(X_i$,strid=1))) X_{r1}= Active(BN ($Conv2D(X_i$,strid=1)));
      X_{i2}= Xo1+ Xr1;
      X_{o2-1}= BN($Conv2D_{3-1}(X_{i2}$,strid=2));
      X_{o2-2}= BN($Conv2D_{3-2}(X_{i2}$,strid=2));
      X_{o2-3}= BN($Conv2D_{3-3}(X_{i2}$,strid=2));
      X_{o2-4}= BN($Conv2D_{3-4}(X_{i2}$,strid=2));
      X_{o2}=stack(X_{o2-1}, X_{o2-2}, X_{o2-3}, X_{o2-4});
      X_{r2-1}= BN($Conv2D_{1-1}(X_{r1}$,strid=2));
      X_{r2-2}= BN($Conv2D_{1-2}(X_{r1}$,strid=2));
      X_{r2-3}= BN($Conv2D_{1-3}(X_{r1}$,strid=2));
      X_{r2-4}= BN($Conv2D_{1-4}(X_{r1}$,strid=2));
      X_{ro}=stack(X_{r2-1}, X_{r2-2}, X_{r2-3}, X_{r2-4});
      X_o=Active(X_{o2}+ X_{ro});
      return X_o;
    else
      X_{o1}= Active(BN ($Conv2D(X_i$,strid=1)));
      X_{o2}= Active(BN ($Conv2D(X_{i1}$,strid=1)));
      X_{ro}= X_i;
      X_o=Active(X_{i2}+ X_{ro});
      return X_o;
    end if
  end for
  O_i= X_o

$$\frac{\partial C}{\partial X_i} = \sum_1^n \left(\frac{\partial C}{\partial h_2^n(X_i)} \frac{\partial h_2^n(X_i)}{\partial h_1(X_i)} \frac{\partial h_1(X_i)}{\partial X_i} + \frac{\partial C}{\partial H_2^n(X_i)} * \right.$$
$$\left. \frac{\partial H_2^n(X_i)}{\partial h_1(X_i)} \frac{\partial h_1(X_i)}{\partial X_i} + \frac{\partial C}{\partial H_2^n(X_i)} \frac{\partial H_2^n(X_i)}{\partial H_1(X_i)} \frac{\partial H_1(X_i)}{\partial x_i} \right)$$
(12)

Among these, one more path is added, combined with the calculation of $\frac{\partial h_1(X_i)}{\partial X_i} = W_{s1}$ on the forward propagation branch, we can get,

$$\frac{\partial C}{\partial X_i} = \sum_1^n \left(\frac{\partial C}{\partial h_2^n(X_i)} \frac{\partial h_2^n(X_i)}{\partial h_1(X_i)} W_{s1} + \frac{\partial C}{\partial H_2^n(X_i)} \frac{\partial H_2^n(X_i)}{\partial h_1(X_i)} * \right.$$
$$\left. W_{s1} + \frac{\partial C}{\partial H_2^n(X_i)} \frac{\partial H_2^n(X_i)}{\partial H_1(X_i)} \frac{\partial H_1(X_i)}{\partial x_i} \right)$$
(13)

4. Related Experiments and Analysis

4.1 Datasets

The experimental part uses image data from the ISCI Archive [18], a dataset consisting of more than 30,000 images, including benign and malignant skin diseases. The images with classification labels of 'uncertain' and 'unknown' were removed from the image data during the experiment. ISCI Archive is open source and can download all digital dermoscope images in bulk. There are many types of diseases that occur on the skin, and some tumor-like lesions
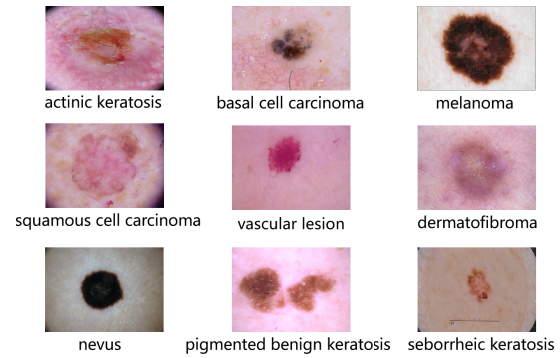

Fig. 9 Benign and malignant skin samples.

have similar visual characteristics to the human eye, which makes it difficult for medical personnel to identify them. Benign and malignant skin tumors can be further diagnosed as actinic keratosis, basal cell carcinoma, dermatofibroma, melanoma, nevus, pigmented benign keratosis, seborrheic keratosis, squamous cell carcinoma, Vascular lesions, some typical samples of which are shown in Fig. 9.

We also conducted experiments on CIRFAR 10 [19]. CIRFAR 10 contains 10 different substance classifications. The hardware used to train the model in this paper includes Intel Core i9-9900K (3.6 GHz, 8 cores), 16GB dual-channel DDR4 @ 2666 Mhz and 11 GB NVIDIA GeForce RTX 2080Ti graphics card. The software is based on the TensorFlow framework with NVIDIA's CUDA accelerated computing architecture.

4.2 Experimental Results

We present the results of the improved algorithms MDFM ResNet, DFM ResNet, FM ResNet and DM ResNet proposed in this article compared with the results of the unimproved classical convolutional neural networks such as RsNet, VGG 16 and inceptionNet for the analysis. In addition, we further compared the network ResNet D [20], which has been changed on the ResNet branch. To conduct experiments and compare results in the same environment, we trained all network models from scratch.

It can be seen from Fig. 10 that the improved MDFM algorithm in this paper has the best effect compared to the unimproved ResNet on ISIC Archive by 1.92%, and there is a similar conclusion on CIFAR10, which is an increase of 1.85%. The effects of the remaining two methods mentioned in this article are also better than Resnet 18 on the two data sets. When the number of network layers is low, there are more times of multi-dimensional matching and fusion, and feature enhancement is better. In order to further analyze the stability of the training process of each network, we take out the accuracy and loss of the last 1/10 epochs of training and draw their box plots as follows.

It can be seen from Fig. 11 that the accuracy of ResNet and the improved schemes FM ResNet and DFMResNet are almost symmetrically distributed, and the overall optimiza-
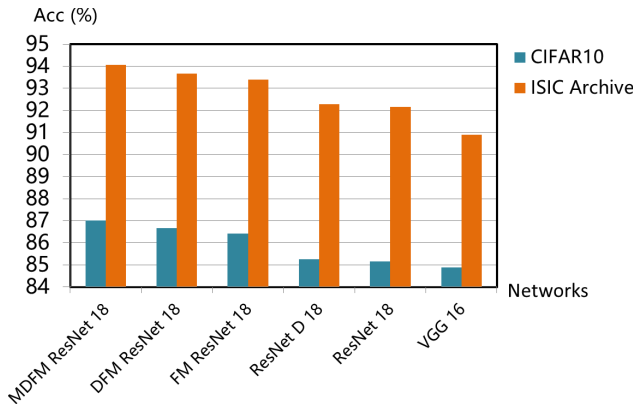
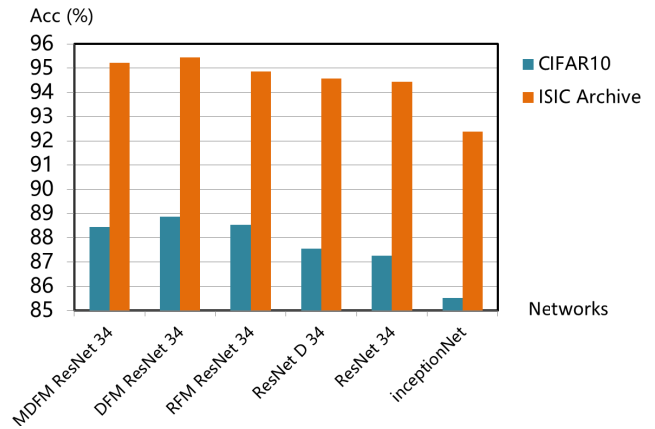**Fig. 10**   Accuracy of 18-layer and similar-ayer networks on datasets.

Fig. 11   Boxplot of network accuracy around layers of 18.

Fig. 12   Loss box plot of the network around the 18 layers.

Fig. 13   The accuracy of 34 layers of network modification and other networks on the data set.

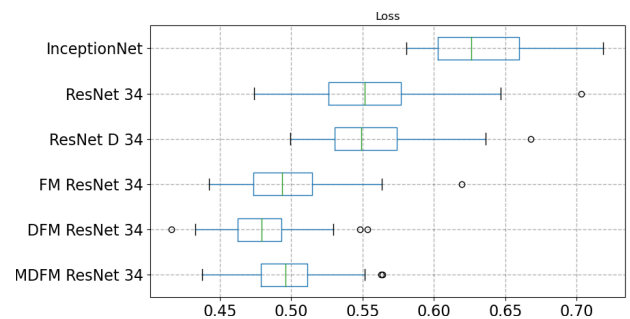**Fig. 14**   Accuracy box plot of the 34-layer improved network and other networks.

Fig. 15   Loss boxplot of layer 34 improved network and other networks.

tion schemes are higher than ResNet. The median line of MDF ResNet is to the right, which means that more epochs have better accuracy. The improvement schemes are all better than the ResNet for improvement. The improved scheme has overall lower losses, with the MDFM ResNet having the lowest median loss line to the left, where lower losses are concentrated, while the DFM ResNet has more scattered losses.

The experimental results on the network with more layers are shown in Fig. 14. Under this condition, dimension matching becomes more important, while excessive fusion will produce wrong information. DFM ResNet has the largest improvement on the dataset, achieving a 1.59%

improvement on CIFAR10, and is also better than other networks on ISIC Archive.

The conclusion of Fig. 14 can be reflected in Fig. 15. The median line of DFM ResNet is the highest, and its accuracy is 1.32%, 0.42% and 0.32% higher than the other three optimization schemes ResNet D, MDFM ResNet and FM ResNet, respectively, and 1.59% higher than that of unimproved ResNet. The overall range of fluctuations after convergence is higher for FM ReNet is larger, and only a few cycles fluctuate more after convergence of DFM ResNet. A similar conclusion can be seen in Fig. 15. The median loss

line of DFM ResNet is the lowest, and the overall distribution is symmetrical. From the comparison of the result graphs, the improved method based on ResNet proposed in this manuscript has improved accuracy compared to the original Resnet network, VGG network [21] and inceptionNet network [22] and other classic networks. Different improved algorithms have different effects on the overall improvement of the network. When the number of network layers is small, the effect of multi-dimensional feature matching algorithm is better than other networks. When the number of network layers increases, the dimensional feature matching algorithm will have more advantages in improving network performance, because the feature dimensionality of the deep network is deeper, and the features are strengthened after matching.

5. Conclusion and Prospect

Based on the residual network, this paper illustrates that the features of the fused feature graph are further enhanced by applying multiple sets of convolution kernels to extract features in different regions of the feature graph and using multiple convolution to generate features of the same dimension on branches. Multiple sets of convolutional kernels are decomposed from the original set of convolutional kernels, and the cross start position of each set of convolutional kernels enhances the extraction of information. Because of this, the amount of computation does not increase. For a multidimensional feature matching structure, the increase in propagation paths implies an increase in the number of coupling methods between information machines, which leads to an increase in effective features. The increase in the number of branching convolutional layers drives the fusion effect, and then the fusion of grouped cross-convolution and regional feature matching further improves the accuracy of the network in skin cancer type recognition to 95.45%.

Of course, there are some other aspects that are expected to have room for further research. First, for the pixels in the feature map, the beginning rows and columns of pixels are not involved in the convolution operation as many times as the later rows and columns of pixels (without considering the step size). Although this can be solved by filling 0 pixels around the feature map, it is not necessarily the optimal solution considering that 0 does not bring a practical effect and takes up memory space and computational resources while not providing a contribution. And whether the sub-pixel filled periphery computed using interpolation method will bring better results is to be further investigated. In addition, although the residual network is composed of multiple residual blocks in series, most of the residual blocks use a convolutional layer with a step size of 1, adding multiple 1*1 convolutions on the short-circuit link Whether the new link with multiple 1*1 convolutional layers will bring better results needs to continue to be studied.

References

[1] P. Fontanillas, B. Alipanahi, N.A. Furlotte, M. Johnson, C.H. Wilson, M. Agee, R.K. Bell, K. Bryc, S.L. Elson, D.A. Hinds, K.E. Huber, A. Kleinman, N.K. Litterman, J.C. McCreight, M.H. McIntyre, J.L. Mountain, E.S. Noblin, C.A.M. Northover, J.F. Sathirapongsasuti, O.V. Sazonova, J.F. Shelton, S. Shringarpure, C. Tian, J.Y. Tung, V. Vacic, S.J. Pitts, R. Gentleman, and A. Auton, "Disease risk scores for skin cancers," Nat. Commun., vol.12, no.1, p.160, Jan. 2021.

[2] N.A. Negbenebor, "The Power of a multidisciplinary tumor board: Managing unresectable and/or high-risk skin cancers," Cutis, vol.107, no.5, pp.E22–E23, May 2021.

[3] K.D. Shue-McGuffin and K. Powers, "Dermatologic simulations in nurse practitioner education: Improving skin cancer knowledge, confidence, and performance," J. Am. Assoc. NURSE Pract., vol.34, no.3, pp.489–498, March 2022.

[4] K. Yu, L. Tan, X. Shang, J. Huang, G. Srivastava, and P. Chatterjee, "Efficient and privacy-preserving medical research support platform against COVID-19: A blockchain-based approach," IEEE Consum. Electron. Mag., vol.10, no.2, pp.111–120, March 2021.

[5] Z. Jiang, Z. Ma, Y. Wang, X. Shao, K. Yu, and A. Jolfaei, "Aggregated decentralized down-sampling-based ResNet for smart healthcare systems," Neural Comput. Appl., vol.35, no.20, pp.14653–14665, July 2023.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.770–778, 2016.

[7] K. Yu, L. Tan, C. Yang, K.-K.R. Choo, A.K. Bashir, J.J.P.C. Rodrigues, and T. Sato, "A blockchain-based Shamir's threshold cryptography scheme for data protection in industrial internet of things settings," IEEE Internet Things J., vol.9, no.11, pp.8154–8167, June 2022.

[8] X. Shao, C. Wu, X. Chen, and W. Zhao, "Editorial: Intelligent mobility and edge computing for a smarter world," Mobile Netw. Appl., vol.27, no.6, pp.2215–2217, Dec. 2022.

[9] Z. Ma, K. Xu, Y. Teng, X. Shao, M. Dong, and Y. Wang, "A Model of extraction of Rail's vertical corrugation based on flexible virtual ruler," IEEE Trans. Intell. Transp. Syst., vol.23, no.2, pp.1097–1108, Feb. 2022.

[10] Y.S. Choi, "Artificial intelligence: Will it replace human medical doctors?," Korean Med. Educ. Rev., vol.18, no.2, pp.47–50, 2016.

[11] Y. Cui, H. Zhang, H. Ji, X. Li, and X. Shao, "Cloud-edge collaboration with green scheduling and deep learning for industrial internet of things," 2021 IEEE Global Communications Conference (GLOBECOM), 2021.

[12] J. Wei, G. Zhu, Z. Fan, J. Liu, Y. Rong, J. Mo, W. Li, and X. Chen, "Genetic U-Net: Automatically designed deep networks for retinal vessel segmentation using a genetic algorithm," IEEE Trans. Med. Imag., vol.41, no.2, pp.292–307, 2021.

[13] K. Yu, Z. Guo, Y. Shen, W. Wang, J. Lin, and T. Sato, "Secure artificial intelligence of things for implicit group recommendations," IEEE Internet Things J., vol.9, no.4, pp.2698–2707, Feb. 2022.

[14] Y.S. Choi, "Artificial intelligence: Will it replace human medical doctors?," Korean Med. Educ. Rev., vol.18, no.2, pp.47–50, 2016.

[15] C. Perera, C. Premachandra, and H. Kawanaka, "Enhancing feature detection and matching in low-pixel-resolution hyperspectral images using 3D convolution-based siamese networks," Sensors, vol.23, no.18, 8004, Sept. 2023.

[16] O. Clivio, F. Falck, B. Lehmann, G. Deligiannidis, and C. Holmes, "Neural score matching for high-dimensional causal inference," International Conference on Artificial Intelligence and Statistics, vol.151, 2022.

[17] K. Hui, X. Shen, S. Abhadiomhen, and Y. Zhan, "Robust low-rank representation via residual projection for image classification," Knowledge-Based Systems, vol.241, 108230, April 2022.

[18] V. Rotemberg, N. Kurtansky, B. Betz-Stablein, L. Caffery, E.

Chousakos, N. Codella, M. Combalia, S. Dusza, P. Guitera, D. Gutman, A. Halpern, B. Helba, H. Kittler, K. Kose, S. Langer, K. Lioprys, J. Malvehy, S. Musthaq, J. Nanda, O. Reiter, G. Shih, A. Stratigos, P. Tschandl, J. Weber, and H.P. Soyer, "A patient-centric dataset of images and metadata for identifying melanomas using clinical context," Sci. Data, vol.8, no.1, 81, March 2021.

[19] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Handbook of Systemic Autoimmune Diseases, vol.1, no.4, 2009.

[20] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li, "Bag of tricks for image classification with convolutional neural networks," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019), pp.558–567, 2019.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," ArXiv Preprint, arXiv:14091556, 2014.

[22] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.2818–2826, 2016.

Liangcheng Qin is a M.Sc. Student (College of Electronic Science and Technology) at National University of Defense Technology. He received his Bachelor's Degree (Science in Electronic Information Engineering) from Beijing Institute of Technology in June 2021. His main research interests include deep learning (specially recurrent neural networks), 2D computer vision.

Xueying Wang was born in 1987. He his B.S. degree in electronic information engineering from Beihang University in 2009. He is currently a associate professor in National University of Defense Technology, Changsha, China. His major research interest is spatial information acquisition and processing.

Zhichao Sha was born in 1985. He received his Ph.D. degree in communication and information system from National University of Defense Technology, Changsha, China, in 2013. He is currently a Associate researcher in National University of Defense Technology, Changsha, China. His major research interest is spatial information acquisition and processing.

Ziji Ma received the B.Sc. degree in electronic information engineering from Hunan University, Changsha, China, in 2001, and the Ph.D. degree in information science from the Nara Institute of Science and Technology, Nara, Japan, in 2012. He is currently an Associate Professor with the College of Electrical and Information Engineering, Hunan University. His research interests include machine vision, signal processing, and V2V communication. He is a member of IEICE.

Kunlai Xiong was born in 1986. He received his Ph.D. degree in communication and information system from National University of Defense Technology, Changsha, China, in 2015. He is currently a lecturer in National University of Defense Technology, Changsha, China. His research interests are array signal processing and blind signal separation.