

Received February 22, 2020, accepted April 5, 2020, date of publication April 9, 2020, date of current version April 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2986916

Dermoscopy Image Classification Based on StyleGANs and Decision Fusion

AN GONG¹, XINJIE YAO¹, AND WEI LIN^{2,3}

¹College of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266580, China

²School of Engineering Science, University of Chinese Academy of Sciences, Beijing 100049, China

³Department of Earth and Planetary Science, University of California at Berkeley, Berkeley, CA 94720, USA

Corresponding author: Xinjie Yao (Z18070013@s.upc.edu.cn)

This work was supported in part by the National Science and Technology Major Project of China under Grant 2017ZX05013-001.

ABSTRACT Skin cancer is one of the most common cancers in humans in recent years, affecting people of all ages. If skin cancer is treated in time, the curative effect and prognosis are favorable. At present, dermoscopy is an effective way for the early diagnosis of skin cancer. However, manual detection is highly dependent on the clinical experience of doctors, and the complexity of the dermoscopy image itself poses a great challenge to the classification. Therefore, we propose a decision fusion method. Through transfer learning, based on multiple pre-trained convolutional neural networks (CNNs), we use the block to combine multiple CNNs and finally make decisions through multiple blocks. The method of decision fusion can solve the generalization capability of a individual convolutional neural network (CNN) model, and is more robust and stable than the traditional fusion strategy. Based on ISIC 2019 dataset, we use StyleGANs to generate high-quality images to alleviate the problem of less and uneven distribution of the dermoscopy image dataset and improve the classification effect of CNNs. Our proposed method can improve the accuracy of dermoscopy image classification and provide help for dermatologists.

INDEX TERMS Dermoscopy images, decision fusion, convolutional neural networks, StyleGANs.

I. INTRODUCTION

The skin is the largest organ of the human body, and it is also the first line of defense of the human body. It has the functions of protecting, secreting, excreting and regulating body temperature. With the changes in lifestyle and environment, various skin diseases affect the normal life of human beings. Many people across the globe are suffering with skin diseases and the number of skin cancer cases are more compared to any other classes of cancer [1]. In the United States alone, there were 5.4 million new cases of skin cancer each year [2]. From all skin cancers, melanoma cases were present in just 5% of cases, but 75% of times, it might lead to death [3], [4]. Melanoma mortality rates in Spain increased during the last decades of the 20th century [5]. The prognosis for advanced melanoma remains poor. However, if detected during the earliest stages, the curative effect and prognosis are better. Early detection of melanoma by means of accurate screening is an important step towards reduction in mortality and prolongation of patients' survival [6]. In recent years, dermatologists around the world have devoted considerable

The associate editor coordinating the review of this manuscript and approving it for publication was Wu-Shiung Feng.

attention to the study of dermoscopy. Esteva *et al.* showed that the convolutional neural network (CNN) achieves performance on par with all tested experts across both classification tasks, demonstrating an artificial intelligence capable of classifying skin cancer with a level of competence comparable to dermatologists [7]. Sinz *et al.* also showed that the specificity of dermoscopic diagnosis of malignant melanoma even higher than clinical diagnosis [8]. Dermoscopy images are acquired by dedicated instruments, and histological images are acquired by invasive biopsy and microscopy. Both models produced highly standardized images. At present, dermoscopy is an effective way for the early diagnosis of skin cancer. Dermoscopy images can provide high image clarity, as shown in Fig. 1.

In order to assist doctors in the diagnosis, dermoscopy is introduced clinically [8]. But manual detection is highly dependent on the clinical experience of doctors, and the complexity of the dermoscopy image itself, such as intra-group variation, inter-group similarity and skin lesion boundary blurring, has made a huge impact on detection. Therefore, the diagnosis and treatment of patients with computer vision have gradually become an important research direction in the development of the medical field. Digital dermoscopy can

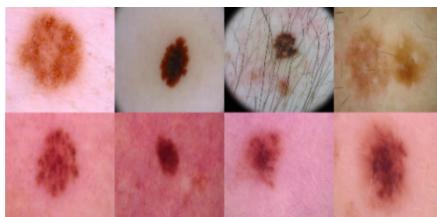


FIGURE 1. Examples of dermoscopy images.

help dermatologists diagnose and treat patients. Digital dermoscopy also facilitates the comparison of different lesions, medical consultations, presentation of interesting cases, and continual education [7]. Computer vision systems classified dermoscopy images with accuracy that exceeded some but not all dermatologists [9].

There has been considerable interest in developing computer vision systems for diagnosis, but few groups have directly compared different ensemble strategies. Prof. Finlay has publicly stated that he is fascinated in computer-aided diagnosis [10]. Research on automatic classification of skin lesions images had already appeared in the literature [11]. Early research focused on the manual extraction of shapes [12], textures [13], [14], and color [15], [16] to fuse these low-level features to improve classification accuracy [4], [17]. Schaefer *et al.* used an automatic border detection approach [18] to segment the lesion area and then assembled the extracted features, i.e., shape, texture and color, for melanoma recognition [19]. Because of the large intraclass differences between melanoma and non-melanoma and the small differences between classes, most of the time, the effect of manual extraction of features is not ideal. In addition, most of the methods based on manual extraction of features are complicated, resulting in low applicability and insufficient generalization ability, etc. in clinical practice. Previous work in dermatological computer-aided classification [20]–[22] lacked the generalization ability of medical practitioners due to insufficient data and attention to standardized tasks such as dermoscopy examination [23], [24] and histological image classification [25]–[27].

A sufficiently rich dataset is very meaningful for research, especially in the field of deep learning. However, medical image datasets are often unevenly distributed and less. Therefore, image classification research based on deep learning is largely limited. In 2014, Goodfellow proposed a new confrontation process to evaluate the effects of the generated model [28]. Generative adversarial networks (GANs) are mainly divided into two parts: the generation model and the discriminant model. The function of the generated model is to simulate the distribution of real data. The role of the discriminant model is to judge whether a sample is a real sample or a generated sample. The goal of GANs is to train a generated model to perfectly fit the real data distribution so that the discriminant model cannot be distinguished. By using GANs, the problem of less and uneven distribution

of medical image data can be alleviated to some extent. With the appearance of GANs, more and more people pay attention to GANs. Radford *et al.* proposed DCGANs, which have clear structural constraints and indicate that they have strong credibility for unsupervised learning and that they are generalized most of the time [29]. Liu proposed CoGANs [30], which can learn a joint distribution unsupervised by adding weights to the network to share constraints and solving distributed solutions. NVIDIA described a new GANs training method whose core idea is to gradually increase the capabilities of the generator and discriminator: starting with low resolution, adding a new layer of continuous modeling fine detail as a training process [31]. NVIDIA proposed another perspective on the GANs framework, creating a generator architecture that learns to generate advanced attributes and random variables in images [32]. With the development of research, deep learning techniques represented by convolutional neural networks (CNNs) have achieved good results in the fields of natural images and medical images analysis. Unlike methods that rely on manual extraction of features, deep learning techniques such as CNNs have significant advantages [33], [34]. The main advantage of the CNN is that it automatically learns the feature representations [35] required for the corresponding detection or classification tasks through the dataset, and has a good performance in many applications [36]. Codella *et al.* established a system combining recent developments in deep learning and machine learning approaches for skin lesion segmentation and classification [37]. Kawahara *et al.* presented a fully CNN based on AlexNet [34] to extract representative features of melanoma [38]. Yu *et al.* proposed a novel method based on very deep CNNs to meet the challenges of automated melanoma recognition in dermoscopy images, which consists of two steps: segmentation and classification [39]. Yang *et al.* used the segmented lesion region to guide the classification by region average pooling [40]. On utilizing age and sex information for classification, Matsunaga *et al.* observed slight effects [41]. iGessert *et al.* combined the crops both by simple averaging and a meta learning strategy to improve the accuracy of model classification [42]. Wei *et al.* proposed a novel Attention Based DenseUNet network with adversarial training for skin lesion segmentation [43]. Gupta *et al.* investigated the application of the Gaussian mixture model based on the analysis and classification of skin diseases from their visual images using a Mahalanobis distance measure [44]. Burlina *et al.* used GANs to generate fundus images of retinal diseases and used deep learning for discriminative tasks in ophthalmology [45]. Although some research has been done on the diagnosis of skin diseases by computer-aided doctors, there is still much room for improvement in classification evaluation metrics.

Based on the above analysis, we use StyleGANs to enhance the dataset and propose a method of decision fusion to automatically classify the dermoscopy image.

The main contributions of our work can be summarized as follows:

- 1) We propose a decision fusion method. Based on multiple pre-trained CNNs, we use block idea to combine multiple CNNs and finally make decisions through multiple blocks. The method of decision fusion can solve the generalization capability of a CNN model, and is more robust and stable than the traditional fusion strategy.
- 2) We train StyleGANs based on the ISIC 2019 dataset and generate high-quality images with trained StyleGANs. The images that are favorable for CNNs training are selected from the generated images and added to the dataset, which improves the classification effect of CNNs.

II. MATERIALS AND METHODS

A. DATA AUGMENTATION BY STYLEGANs

The International Skin Imaging Collaboration (ISIC) 2019 dataset has 25331 dermoscopy images [46]–[48], totaling 8 classes (i.e., actinic keratosis (AKIEC), basal cell carcinoma (BCC), benign keratosis (BKL), dermatofibroma (DF), melanoma (MEL), melanocytic nevus (NV), vascular lesion (VASC) and squamous cell carcinoma (SCC)). In this experiment, sixty percent of the dataset is defined as the training set, and twenty percent of the dataset is defined as the validation set. The remaining dataset is used as the test set. The training set is used for training to obtain a CNN model. Then we use the validation set to verify the validity of the model and pick the model that gets the best results until we get a satisfactory model. Finally, we use the test set to evaluate the final effect of the model. The test set is used only during model evaluation. It is worth mentioning that the division of the dataset should be as consistent as possible in the distribution of the data to avoid affecting the final result due to the additional bias introduced by the data division process.

Although the CNN has strong feature expression ability, it requires a large number of datasets for training. Most medical image datasets are less and unevenly distributed, which will reduce the training effect of CNNs and easily lead to over-fitting. In order to alleviate the above problems, data augmentation based on GANs are adopted. Since the advent of GANs, a wide variety of new GANs have emerged. We generated the dermoscopy images by selecting StyleGANs, as shown in the Fig. 2.

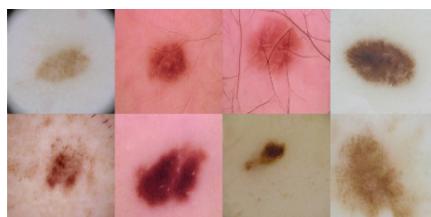


FIGURE 2. Examples of dermoscopy images generated by StyleGANs.

Based on the eight dermoscopy image classes, we divide the training set into eight classes. We train StyleGANs with each class of training set as input, so that we can get eight different StyleGANs. Different StyleGANs generate corresponding dermoscopy images. 200 generated images of each class are included in the training set. As can be seen from Fig. 2, StyleGANs have a good performance in generating dermoscopy images. StyleGANs result in the automatic, unsupervised separation of high-level attributes and random variations in the generated image, and the ability to control the composition intuitively and at a specific scale. However, not all dermoscopy images generated by StyleGANs are useful for training CNNs. StyleGANs may generate some blurry or distorted dermoscopy images, as shown in the Fig. 3. These poorly generated images are not conducive to our training of CNNs. Therefore, we need to manually select clear and undistorted dermoscopy images into the training set, as shown in the Fig. 2.



FIGURE 3. Examples of less effective dermoscopy images generated by StyleGANs.

We generated a large number of dermoscopy images with the help of StyleGANs, and selected the high-quality dermoscopy images and added them to the original dataset for CNNs training, as shown in the Fig. 4. The selection is made manually by visual assessment.

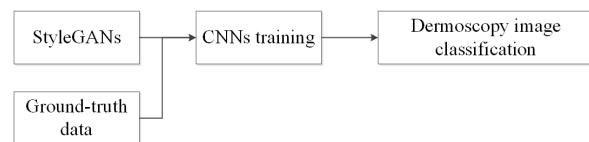


FIGURE 4. The flow chart of the dermoscopy image classification based on data augmentation by StyleGANs.

B. IMAGE SIZE

CNNs usually use fixed size and square images as input, and different size images can get different feature extraction effect. All images are cropped to the desired size before entering the CNN, but most cropping causes skin lesions to deform, and shape contour information is an important basis for discriminating skin cell damage classes. Relatively complete dermoscopy images are helpful for CNNs to extract features. However, if the image is large, the time and computational cost of training CNNs is high. We try to preprocess and crop the image without affecting the recognition

target, and adjust the image size with the center as the origin. By comparing three sizes of dermoscopy images, we choose a size of 600×600 as input.

C. LOSS FUNCTION

The CNN transforms the input image through multiple hidden layers to achieve high-level feature extraction. This end-to-end automatic learning feature process has a good performance in feature extraction. CNNs have better generalization ability and less computational complexity than fully connected neural networks. In this experiment, the transfer learning was used to fine-tune the CNN model pre-trained on the ImageNet to extract the characteristics of skin lesions. The final convolutional layer features of the CNN model are extracted due to the differences between the pre-trained dataset and the dermoscopy image dataset and the ability to express deep features. Considering the uneven distribution of medical image datasets, this paper sets the weight coefficient w in the softmax loss function to multiply the large sample by the smaller weight and the small sample by the larger weight, so as to alleviate the problem of uneven distribution in the dataset. Hence, the weighted softmax loss function is defined as:

$$J(x, y, \theta) = -\frac{1}{N} \left[\sum_{i=1}^N \sum_{j=1}^k w_j \mathbb{1}\{y_i = j\} \log \frac{\exp(\theta_j^T x_i)}{\sum_{i=1}^k \exp(\theta_i^T x_i)} \right] \quad (1)$$

where $w_j = N/N_j$ denotes weight of loss function, N denotes total number of samples, and N_j denotes the number of samples for class j , x_i denotes the test image, y_i denotes the mark of each image, θ denotes the parameters of the model, k denotes the total number of image classes, $\mathbb{1}\{y_i = j\}$ denotes the value is only 1 if the value in the braces is true, otherwise it is 0.

D. IMAGE NORMALIZATION

In order to reduce the interference caused by uneven light in medical images, this study normalized the dermoscopy image. The common image normalization method is to subtract the image pixel value from the average pixel value calculated on the whole dataset. However, the illumination, skin color, and viewing angle of the dermoscopy image in the dataset vary widely, and subtracting the uniform average does not well standardize the illumination of individual images. In response to the above problem, this experiment normalizes each dermoscopy image by subtracting the average intensity value of the channel calculated on a individual image from the image. By giving a dermoscopy image X , the normalized image X_{norm} is calculated and defined as:

$$\begin{cases} X_{\text{norm},R} = X_R - u(X_R) \\ X_{\text{norm},G} = X_G - u(X_G) \\ X_{\text{norm},B} = X_B - u(X_B) \end{cases} \quad (2)$$

where $u(X_R)$, $u(X_G)$, and $u(X_B)$ denote the average pixel values of the 3 color channels respectively.

E. CLASSIFICATION FRAMEWORK BASED ON DECISION FUSION

Ensemble learning is to build a set of base classifiers from the training dataset, and then get the prediction results through each base classifier, and finally classify by voting. Ensemble learning is not a classifier, but a combination of classifiers. Based on the traditional ensemble strategies, we proposed a CNNs image classification framework, as shown in the Fig. 5.

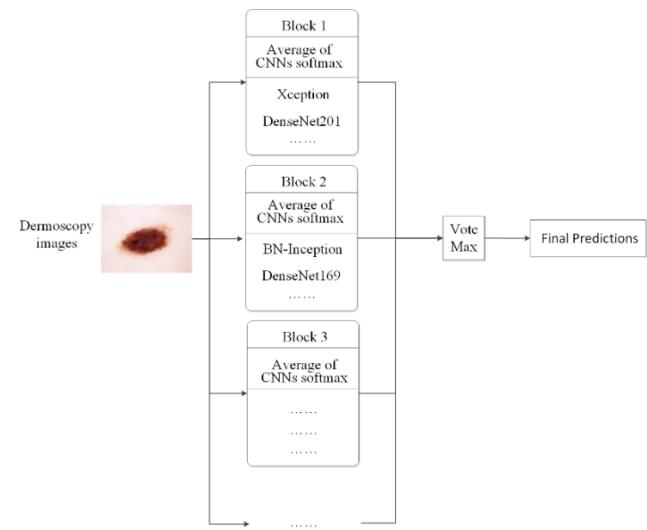


FIGURE 5. The flow chart of the classification framework based on decision fusion.

There are many evaluation metrics for medical image classification, but it is difficult for a individual CNN to perform well on multiple evaluation metrics. Therefore, we integrated multiple CNNs. Each fusion is called a block, and each block contains n CNNs, and then the n CNNs respectively classify the input dermoscopy image. At this point, each CNN will have a softmax output, which will add up the softmax output of each CNN, that is, the sum of the softmax of n CNNs is m . There are k classes of dermoscopy images in the experiment, each softmax contains k values, which represent the probability of being possible for each class of dermoscopy image for the input dermoscopy image. Next, divide m by n to get the mean of the softmax in the block. There are 8 classes of dermoscopy images in this experiment, k is 8. Considering the cost of calculation, each block contains 5 CNNs, and n is 5. The flow chart is presented in Fig. 6.

In CNNs, the output of the last fully connected layer is usually used as the input of softmax. Through the softmax, we map real numbers between 0 and 1. The softmax function is defined as follows:

$$y_j = \text{softmax}(z_j) = \frac{\exp(z_j)}{\sum_{i=1}^k \exp(z_i)} \quad (3)$$

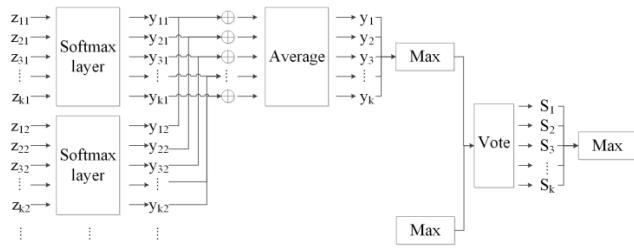


FIGURE 6. The flow chart of the block.

where y_j , z_j and k denote the probability of class j through softmax mapping, output of class j through the last fully connected layer, total number of dermoscopy image classes.

After obtaining the mean value of softmax in the block, the maximum value can be obtained from the k values. By finding the maximum value corresponding dermoscopy image class as a block of the predicted result. By analogy, the results of other blocks are grouped together. Finally, voting is performed, and the dermoscopy image class corresponding to the item with the most votes is used as the result predicted by the entire framework. If the number of votes is the same, then we randomly select one of them as the prediction result. For k classes, we can get k voting results, which is S_k . Considering the cost of calculation, 6 blocks are selected here. The difference between the proposed decision fusion and the traditional fusion method is that the fusion methods are different. The traditional fusion method is to simply take the average of the prediction probabilities of multiple CNNs or simply accumulate the prediction results of multiple CNNs. The proposed decision fusion method mainly uses blocks.

Based on the analysis above, we designed the framework in four steps:

- 1) Step1. Select the dermoscopy images: We choose clear and non-deformed dermoscopy images as input for CNNs.
- 2) Step2. Calculate softmax value: We use the softmax function to map the output of each CNN between 0 and 1 to obtain the probability that each dermoscopy image belongs to each class.
- 3) Step3. Calculate the result of each block: We calculate the average value of the softmax output of the CNNs in each block, and select the dermoscopy image class corresponding to the maximum value as the output of this block.
- 4) Step4. Perform image classification: By superimposing the output results of multiple blocks, the maximum value is taken from the final accumulation result as the final prediction result of the entire framework.

III. EXPERIMENT RESULTS AND ANALYSIS

Experimental images are from ISIC 2019 Skin Lesion Analysis Towards Melanoma Detection Challenge dataset. This dataset consists of 25331 dermoscopy images and the ground

truth is also available from the challenge, which is based on the manual delineation by clinical experts. 0.5 is used as the classification threshold according to ISIC 2019. For the optimizer, Adam, SGD, and RMSprop were tried. It was found that different optimizers had little effect on the classification of dermoscopy images, and finally a more stable Adam was used as the optimizer. Experiments were implemented with PyTorch library on a computer equipped with four NVIDIA Tesla P100 GPUs with 16GB of memory. We obtain weights through some existing pre-trained models. The CNNs performed 200 epoch of training. In the experiment, the batch size was 32, the learning rate was 0.0001.

A. EVALUATION METRICS

In order to comprehensively evaluate the classification performance of the model, accuracy (ACC), sensitivity (SE), specificity (SP), average precision (AP) and area under ROC curve (AUC) are used as evaluation metrics. In the experiment, the CNN model with better classification of dermoscopy images was selected based on the ACC and AUC as the main evaluation metrics. ACC, which is the ratio of correctly classified samples to all samples, can be used to evaluate the ability of the classifier to judge the overall sample. SE and SP are important metrics in medical diagnosis. SE is also called true positive rate or recall. The higher the value, the lower the probability of missed diagnosis. The SP is also called the true negative rate, and the higher the value, the higher the probability of diagnosis. The evaluation metric of AP is introduced, which obtains the final result by calculating the precision and recall. As a metric of the relationship between precision and recall, the value is between 0 and 1, the higher the value, the better the classifier. In order to more accurately evaluate the classifier for unbalanced data, AUC was introduced. AUC is defined as the area under the ROC (receiver operating characteristic) curve, which in turn is a graphical interpretation of true positive rate (TPR) and false positive rate (FPR). The metrics are defined as:

$$ACC = \frac{N_{tp} + N_{tn}}{N_{tp} + N_{fp} + N_{fn} + N_{tn}} \quad (4)$$

$$SE = \frac{N_{tp}}{N_{tp} + N_{fn}} \quad (5)$$

$$SP = \frac{N_{tn}}{N_{tn} + N_{fp}} \quad (6)$$

$$AP = \int_0^1 P(R) dR \quad (7)$$

$$AUC = \int_0^1 t_{pr}(f_{pr}) df_{pr} \quad (8)$$

where N_{tp} , N_{tn} , N_{fp} , N_{fn} , P , R , t_{pr} and f_{pr} denote the number of true positive, true negative, false positive, false negative, precision, recall, true positive rate and false positive rate, respectively, and they are all defined on the image level. A lesion image is considered as a true positive if its prediction is lesion; otherwise it is regarded as a false negative. A non-lesion image is considered as a true negative if its

prediction is non-lesion; otherwise it is regarded as a false positive.

B. COMPARISON OF CLASSIFICATION EFFECTS UNDER DIFFERENT DATASETS

In general, medical image datasets are less and unevenly distributed, which poses a huge challenge for CNNs training. Therefore, in this experiment, the dermoscopy image is generated by using StyleGANs, thereby enhancing the dataset and alleviating the problem that the dermoscopy image dataset is less and unevenly distributed. The comparison of CNNs classification effects before and after the enhanced dataset is shown in Table 1.

TABLE 1. Comparison of multiple datasets classification effects.

Model	Mean Value				
	ACC	AUC	AP	SE	SP
InceptionV3-StyleGANs	0.975	0.943	0.887	0.821	0.983
InceptionV3	0.966	0.908	0.818	0.742	0.976
ResNet50- StyleGANs	0.967	0.934	0.819	0.763	0.977
ResNet50	0.958	0.910	0.770	0.700	0.972
VGG16BN- StyleGANs	0.963	0.912	0.783	0.725	0.975
VGG16BN	0.956	0.896	0.755	0.692	0.969

Through experiments, we selected three CNNs, InceptionV3, ResNet50, and VGG16BN, to compare the dermoscopy image classification effects of the original dataset and the dataset enhanced by StyleGANs. Through experiments, it can be seen that the three CNNs trained in the dataset enhanced by StyleGANs have a certain degree of improvement in various evaluation metrics. Especially in the two important evaluation metrics of AUC and SE, InceptionV3 increased by 3.5% and 7.9%, ResNet50 increased by 2.4% and 6.3%, and VGG16BN increased by 1.6% and 3.3%, respectively.

C. COMPARISON OF CLASSIFICATION EFFECTS UNDER DIFFERENT IMAGE SIZES

Most cropping causes skin lesions to deform, and shape contour information is an important basis for discriminating skin cell damage classes. Therefore, we tried a variety of image sizes to the optimal image size, and the experimental results are shown in Table 2.

By comparing the experimental results of multiple image sizes as input, it is found that different CNN models have better evaluation metrics than image sizes of 400×400 and 224×224 with image size 600×600 as input. Experiments show that the relatively complete dermoscopy image information is beneficial to feature extraction, and the CNN model

TABLE 2. Comparison of multiple image sizes classification effects.

Model	Mean Value				
	ACC	AUC	AP	SE	SP
InceptionV3-600×600	0.975	0.943	0.887	0.821	0.983
InceptionV3-400×400	0.945	0.894	0.780	0.710	0.967
InceptionV3-224×224	0.904	0.721	0.498	0.450	0.937
ResNet50-600×600	0.967	0.934	0.819	0.763	0.977
ResNet50-400×400	0.951	0.893	0.740	0.712	0.969
ResNet50-224×224	0.916	0.785	0.491	0.482	0.950
VGG16BN-600×600	0.963	0.912	0.783	0.725	0.975
VGG16BN-400×400	0.942	0.846	0.697	0.640	0.963
VGG16BN-224×224	0.914	0.739	0.521	0.480	0.943

can be performed better to some extent. The comparison of the size of 600×600 and the size of 224×224 is the most obvious. In the three evaluation metrics of AUC, AP and SE, InceptionV3 is increased by 22.2%, 38.9%, 37.1%, and ResNet50 is increased by 14.9%, 32.8%, 28.1%, and VGG16BN is increased by 17.3%, 26.2%, 24.5%, respectively.

D. COMPARISON EFFECTS OF DIFFERENT CLASSIFICATION METHODS

We compared decision fusion with the individual CNN and traditional fusion methods. It contains 43 CNNs, and the traditional fusion methods include averaging and voting. The experimental results are shown in Table 3.

According to Table 3, CNNs have different performances on different evaluation metrics. It is difficult for a individual CNN to perform particularly well on all evaluation metrics. For example, in ACC, Xception can reach 97.7%. In AUC, DenseNet201 has a good performance, which can reach 94.9%. In AP, InceptionV3 can reach 88.7%. In SE, the outstanding performance is SE_ResNeXt50_32 × 4d, which can reach 83.5%. In SP, InceptionV3 performs well and can reach 98.3%. When we use a individual model to deal with problems, we often encounter problems in model generalization. The generalization ability of a model is limited due to some objective factors. In addition, after establishing a model, the model may perform better in solving certain problems, but the results may not be satisfactory when solving other datasets. Therefore, we use some methods to fuse excellent models to break the generalization ability of the individual model to unknown datasets, and combine the advantages of multiple models to obtain a better solution to the same datasets. For the generalization capability of a

TABLE 3. Comparison of classification effects.

Model	Mean Value				
	ACC	AUC	AP	SE	SP
AlexNet	0.959	0.910	0.758	0.638	0.971
BN-Inception	0.974	0.936	0.867	0.798	0.980
Caffe-ResNet101	0.970	0.931	0.837	0.780	0.977
DenseNet121	0.971	0.946	0.864	0.799	0.978
DenseNet161	0.970	0.937	0.855	0.801	0.977
DenseNet169	0.973	0.948	0.871	0.813	0.979
DenseNet201	0.973	0.949	0.876	0.833	0.979
DPN68b	0.969	0.920	0.838	0.825	0.976
DPN68	0.967	0.916	0.812	0.810	0.975
DPN92	0.957	0.892	0.756	0.764	0.969
DPN98	0.954	0.892	0.732	0.759	0.968
DPN107	0.946	0.876	0.672	0.701	0.964
DPN131	0.948	0.871	0.680	0.703	0.964
FbResNet152	0.965	0.934	0.816	0.734	0.975
InceptionResNetV2	0.974	0.943	0.870	0.803	0.979
InceptionV3	0.975	0.943	0.887	0.821	0.983
InceptionV4	0.974	0.944	0.877	0.822	0.980
NASNet-A_Large	0.955	0.888	0.744	0.701	0.967
NASNet-A_Mobile	0.968	0.907	0.832	0.731	0.975
PNASNet-5_Large	0.959	0.892	0.744	0.721	0.970
PolyNet	0.959	0.910	0.770	0.701	0.970
ResNet18	0.968	0.938	0.837	0.770	0.977
ResNet34	0.972	0.937	0.862	0.798	0.978
ResNet50	0.967	0.934	0.819	0.763	0.977
ResNet101	0.967	0.931	0.829	0.758	0.975
ResNet152	0.964	0.926	0.790	0.737	0.973
ResNeXt101_32x4d	0.969	0.926	0.825	0.772	0.977
ResNeXt101_64x4d	0.963	0.917	0.773	0.750	0.973
SE_ResNet152	0.967	0.930	0.832	0.794	0.975
SE_ResNeXt50_32x4d	0.977	0.933	0.883	0.835	0.980
SE_ResNeXt101_32x4d	0.973	0.928	0.853	0.801	0.977
SENet154	0.966	0.915	0.809	0.773	0.974
SqueezeNet1_0	0.941	0.834	0.575	0.434	0.964

TABLE 3. (Continued.) Comparison of classification effects.

SqueezeNet1_1	0.943	0.844	0.599	0.449	0.964
VGG11	0.961	0.901	0.774	0.687	0.970
VGG11BN	0.966	0.914	0.805	0.746	0.973
VGG13	0.960	0.902	0.765	0.684	0.970
VGG13BN	0.965	0.915	0.809	0.734	0.973
VGG16	0.958	0.906	0.736	0.637	0.970
VGG16BN	0.963	0.912	0.783	0.725	0.975
VGG19	0.955	0.888	0.696	0.596	0.969
VGG19BN	0.964	0.910	0.782	0.711	0.973
Xception	0.977	0.934	0.874	0.831	0.981
Averaging1	0.981	0.961	0.938	0.921	0.979
Vote1	0.957	0.902	0.847	0.440	0.990
Decision Fusion1	0.986	0.953	0.926	0.914	0.989
Averaging2	0.989	0.978	0.977	0.972	0.991
Vote2	0.977	0.972	0.966	0.753	0.994
Decision Fusion2	0.994	0.984	0.977	0.973	0.995
Averaging3	0.995	0.980	0.979	0.972	0.992
Vote3	0.978	0.972	0.970	0.765	0.996
Decision Fusion3	0.995	0.989	0.984	0.983	0.996

individual CNN model, we propose a method of decision fusion and compare it with the traditional fusion strategy. In general, the classification effect through fusion is better than a individual CNN. However, the CNNs selected during the fusion process are very important. If the selected CNNs are not well classified, the classification effect of the fusion model may be worse. According to Table 3, the CNNs selected in each block perform as well as possible on various evaluation metrics, especially on ACC and AUC. In addition, the CNNs in each block are as diverse as possible, that is, there are differences between different CNNs. In this experiment, three groups of CNNs were selected for averaging, voting and decision fusion. The first group includes AlexNet, VGG19BN, ResNet101, DPN107, DPN131, VGG11, VGG11BN, VGG13, VGG13BN, VGG16, VGG16BN, VGG19, ResNet152, SqueezeNet1_1, SqueezeNet1_0. The second group includes SENet154, ResNet18, ResNet34, DenseNet121, DenseNet161, DenseNet169, DenseNet201, InceptionV3, InceptionV4, SE_ResNeXt101_32 × 4d, ResNet50, ResNet101, ResNet152, VGG16BN, SE_ResNet152. The third group includes DenseNet121, DenseNet169, DenseNet201, InceptionV3, DPN68b,

Xception, InceptionV4, InceptionResNetV2, ResNet18, ResNet34, ResNet50, SE_ResNeXt50_32 \times 4d, SE_ResNeXt101_32 \times 4d, BN-Inception, SE_ResNet152. Because our proposed method requires multiple different CNNs and different CNNs will have different effects on our proposed method. We mainly compare the overall impact of each group of CNNs on our method, so not all CNNs need to be used. Therefore, the three groups of CNNs we selected have different classification effects as a whole. Our idea is to use 6 blocks with 5 CNNs in each block. First we divided the selected CNNs into three groups, each group has 15 CNNs. We number these 15 CNNs in sequence. The first three CNNs of each group are put into these 6 blocks. The missing two CNNs in each block are selected from the remaining 12 CNNs and cannot be repeated. Averaging1, Averaging2, Averaging3, Vote1, Vote2, Vote3, Decision Fusion1, Decision Fusion2, and Decision Fusion3 represent the averaging, voting, and decision fusion strategies based on the first, second, and third groups of CNNs, respectively. According to Table 3, based on the first group of CNNs, the classification of the model using the voting method is not good, and the SE is only 44%. Because most of the first group of CNNs are not very effective, they affect the effect of the fusion to some extent. The third group of CNNs selected CNNs with better classification effect than the first group and the second group. In ACC, Averaging3 and Decision Fusion3 performed very well, reaching 99.5%. In SP, Vote3 and Decision Fusion3 can reach 99.6%. In other evaluation metrics, Decision Fusion3 performed better than other CNNs and fusion models. For example, in AUC, Decision Fusion3 is 4% better than DenseNet201. In SE, Decision Fusion3 is increased by 54.3% compared to Vote1. Therefore, it is meaningful to adopt a fusion strategy, but it is necessary to pay attention to the classification effect when selecting the merged CNNs. From the perspective of the third group of CNNs, Decision Fusion3 and Averaging3 performed equally well in ACC, but Decision Fusion3 has some improvement in other evaluation metrics. For example, in terms of AUC and SE, Decision Fusion3 increased by 0.9% and 1.1%, respectively. Decision Fusion3 and Vote3 behave similarly on the SP, but in other respects Decision Fusion3 has a significant improvement, especially in the SE aspect increased by 21.8%. The decision fusion proposed by us has a certain improvement compared with the traditional fusion strategy.

E. COMPARISON OF LOSS

In general, the larger the loss function value, the smaller the classification probability of the classifier on the real label, and the worse the performance. Through ten-fold cross validation, the loss is used to measure the quality of the model. Several typical CNN models were selected to compare the loss, which is presented in Fig. 7.

It can be seen from the loss curve that Densnet201 has the fastest convergence rate, and the loss is stable at 0.23 after about 25 epoch, and other CNN models converge after about 45 epoch. According to Table 3 and Fig. 7, DenseNet has

TABLE 4. Patient information and number statistics.

Patient information	Number of patients							
	MEL	NV	BCC	AK	BKL	DF	VASC	SCC
Female	1980	6379	1318	406	1134	117	125	203
Male	2461	6225	2000	458	1467	122	128	425
0 years old	0	45	0	0	5	0	4	0
5 years old	1	108	0	0	1	0	3	0
10 years old	1	135	2	2	0	0	2	0
15 years old	5	355	2	0	0	7	6	0
20 years old	18	349	10	0	7	0	4	0
25 years old	45	606	13	0	5	2	6	0
30 years old	102	955	61	19	43	11	8	0
35 years old	176	1324	55	0	57	19	17	3
40 years old	258	1656	125	15	132	32	23	5
45 years old	373	1832	170	6	133	29	26	16
50 years old	464	1503	200	52	197	28	18	27
55 years old	401	1090	292	63	233	32	28	31
60 years old	410	802	293	131	311	18	17	54
65 years old	478	713	382	108	292	24	9	69
70 years old	570	431	443	157	362	14	28	115
75 years old	383	363	431	119	365	9	34	92
80 years old	391	200	397	102	221	5	13	130
85 years old	361	91	441	90	234	9	7	86
Anterior torso	1331	3699	1114	92	445	43	78	113
Head/Neck	880	767	1131	598	1001	0	36	174
Lateral torso	14	34	0	0	6	0	0	0
Lower extremity	796	2876	462	66	412	138	58	182
Oral/Genital	19	24	0	0	14	0	2	0
Palms/Soles	201	168	0	6	7	0	3	13
Posterior torso	430	1888	186	4	230	2	22	25
Upper extremity	724	1325	358	79	234	52	23	115

a good performance and we found that the deep and more recent CNN model usually performs better in dermoscopy image classification, but not all.

F. DERMOSCOPY IMAGES CORRESPONDING TO PATIENT INFORMATION

Some of the dermoscopy images in the dataset have patient information such as age, gender, and lesion location. We counted various classes of information and the number

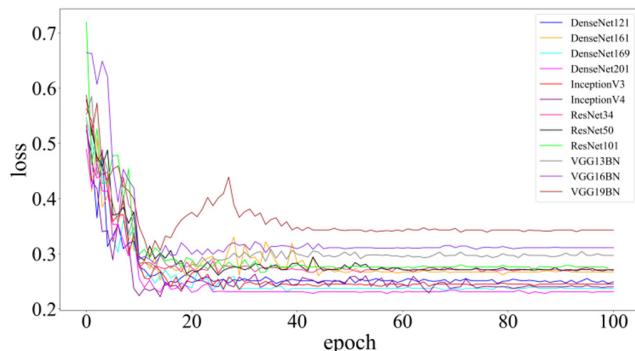


FIGURE 7. Loss curves of different CNNs.

of patients with various classes of skin cancer, as shown in Table 4.

According to Table 4, men with BCC and SCC were about 1.5 and 2 times as likely as women respectively. The location of lesions with NV is mainly concentrated in anterior torso and lower extremity. The location of BCC is mainly found in head, neck and anterior torso. DF is more likely to appear in lower extremity. Skin cancer patients are mainly middle-aged, and infants have a higher probability of having NV than other classes of skin cancer.

IV. DISCUSSION

Deep learning provides an end-to-end learning paradigm. The whole learning process does not divide the artificial sub-problems, but completely passes it to the deep learning model to directly learn the mapping from raw data to expected output. Traditional image recognition problems often use divide and conquer. It is divided into several steps such as preprocessing, feature extraction and selection, and classifier design. The motivation of the divide and conquer method is to divide the mother problem of image recognition into several small sub-problems that are simple, controllable and clear. Although the optimal solution can be obtained on the sub-problem, the optimal solution on the sub-problem does not mean that the final solution of the global problem can be obtained. In contrast, end-to-end learning eliminates the need for data annotations to be performed before each independent learning task is performed. The cost of labeling samples is expensive and error-prone. The quality of medical image datasets are important for deep learning, but medical image datasets are usually unevenly distributed and lacking, so that the model can not complete the training, which easily leads to over-fitting and thus reduces the effect. Although researchers now share data to alleviate this problem, medical image data involves issues such as personal privacy, which poses a huge challenge to obtaining a large, diverse, balanced set of medical image data. In addition, some medical image cases are few and less accessible. In the medical field, our proposed method can help doctors in diagnosis and treatment and enhance medical image datasets.

Considering the characteristics of deep learning and medical image datasets, GANs can solve this problem to some extent. The images generated by GANs do not involve sensitive issues such as personal privacy, and at the same time provide high-quality quality datasets for deep learning. As a more mature model in deep learning, CNNs have a outstanding performance in image processing. StyleGANs can learn advanced attributes and random variables in images to generate high-quality images. CNNs are able to extract key features from these generated high-quality images to improve the classification of images. In addition, we use a complete dermoscopy image to ensure that the image information is complete. CNNs can be trained better by setting the weight of loss function and preprocessing images. Although CNNs have strong feature extraction capabilities, a individual CNN has certain limitations, and it does not guarantee good performance on multiple issues. Therefore, we propose a method of decision fusion. Through transfer learning, based on multiple pre-trained CNNs, we use block idea to combine multiple CNNs and finally make decisions through multiple blocks. The method of decision fusion can solve the generalization capability of a individual CNN model, and is more robust and stable than the traditional fusion strategy. However, decision fusion has certain requirements for selecting CNNs, and selects CNNs with better effects as much as possible, so as to maximize the performance of decision fusion. Medical images contain a lot of information, shape contour information is an important basis for the classification of skin cell damage classes, and complete medical images are very meaningful for research. In addition, the patient information corresponding to the medical image also has a certain promotion and guidance for the research.

The computer plays an auxiliary role in the diagnosis. When encountering some important problems, it is necessary to combine the computer with the human experience. The significance of computer-aided diagnosis can not only alleviate the pathological reading problems of primary clinicians, but also enhance people's health awareness. The Internet and smart devices are very popular. Patients can be better detected and treated by transplanting computer-aided diagnostics to mobile devices. For skin cancer, if the treatment is timely and timely, the curative effect and prognosis are better.

V. CONCLUSION

The dermoscopy image presents a huge challenge to automatic classification due to its own complexity. In this paper, we use StyleGANs to train based on the ISIC 2019 dataset. The trained StyleGANs generate high-quality images in batches, and select more realistic images to add to the original dataset. Based on the enhanced dataset, we used 43 pre-trained CNNs for training. Three groups of CNNs were selected from the 43 trained CNNs for comparison of different fusion strategies. Through experiments, it is found that the selection of CNNs with better classification results is beneficial to the effect of fusion. The decision fusion has a significant improvement over multiple evaluation metrics for

a individual CNN and is more stable and improved in evaluation metrics than the traditional fusion strategy. The method we proposed not only can automatically classify dermoscopy images, but also enhance the dataset and alleviate the uneven distribution of dermoscopy images. In general, Skin cancer will have a better prognosis if it is found and treated earlier. Computer-aided diagnosis can not only help doctors, but also enhance people's health awareness.

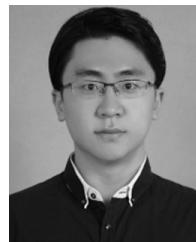
REFERENCES

- [1] *Skin Cancer Facts and Statistics*. [Online]. Available: <https://www.skincancer.org/skin-cancer-information/skin-cancer-facts>
- [2] H. W. Rogers, M. A. Weinstock, S. R. Feldman, and B. M. Coldiron, "Incidence estimate of nonmelanoma skin cancer (Keratinocyte Carcinomas) in the US population, 2012," *JAMA Dermatol.*, vol. 151, no. 10, p. 1081, Oct. 2015.
- [3] R. Kasmi and K. Mokrani, "Classification of malignant melanoma and benign skin lesions: Implementation of automatic ABCD rule," *IET Image Process.*, vol. 10, no. 6, pp. 448–455, Jun. 2016.
- [4] M. E. Celebi, H. A. Kingravi, B. Uddin, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, and R. H. Moss, "A methodological approach to the classification of dermoscopy images," *Computerized Med. Imag. Graph.*, vol. 31, no. 6, pp. 362–373, Sep. 2007.
- [5] E. Gutiérrez, G. González, G. L. Abente, N. Aragónés, M. Pollán, R. P. Barriuso, M. J. Sánchez, and B. Pérez, Gómez, "Trends in mortality from cutaneous malignant melanoma in spain (1982–2016): Sex–specific age–cohort–period effects," *J. Eur. Acad. Dermatol. Venereol.*, vol. 33, no. 8, pp. 1522–1528, Apr. 2019.
- [6] T. Fikrle and K. Pizinger, "Digital computer analysis of dermatoscopical images of 260 melanocytic skin lesions; perimeter/area ratio for the differentiation between malignant melanomas and melanocytic nevi," *J. Eur. Acad. Dermatol. Venereol.*, vol. 21, no. 1, pp. 48–55, Jan. 2007.
- [7] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.
- [8] C. Sinz, P. Tschandl, C. Rosendahl, B. N. Akay, G. Argenziano, A. Blum, R. P. Braun, H. Cabo, J. Y. Gourhant, J. Kreusch, and A. Lallas, "Accuracy of dermatoscopy for the diagnosis of nonpigmented cancers of the skin," *J. Amer. Acad. Dermatol.*, vol. 77, no. 6, pp. 1100–1109, 2017.
- [9] M. A. Marchetti, N. C. F. Codella, S. W. Dusza, D. A. Gutman, B. Helba, A. Kalloo, N. Mishra, C. Carrera, M. E. Celebi, J. L. DeFazio, N. Jaimes, A. A. Marghoob, E. Quigley, A. Scope, O. Yélamos, and A. C. Halpern, "Results of the 2016 international skin imaging collaboration international symposium on biomedical imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images," *J. Amer. Acad. Dermatol.*, vol. 78, no. 2, pp. 270–277.e1, Feb. 2018.
- [10] A. Y. Finlay, "Pioneers in dermatology and venereology: An interview with Prof. Andrew finlay," *J. Eur. Acad. Dermatol. Venereol.*, vol. 33, no. 4, pp. 633–636, Apr. 2019.
- [11] N. Cascinelli, M. Ferrario, T. Tonelli, and E. Leo, "A possible new tool for clinical diagnosis of melanoma: The computer," *J. Amer. Acad. Dermatol.*, vol. 16, no. 2, pp. 361–367, Feb. 1987.
- [12] N. K. Mishra and M. Emre Celebi, "An overview of melanoma detection in dermoscopy images using image processing and machine learning," 2016, *arXiv:1601.07843*. [Online]. Available: <http://arxiv.org/abs/1601.07843>
- [13] L. Ballerini, R. B. Fisher, B. Aldridge, and J. Rees, "A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions," in *Color Medical Image Analysis*. Dordrecht, The Netherlands: Springer, 2013, pp. 63–86.
- [14] T. Tommasi, T. E. La, and B. Caputo, "Melanoma recognition using representative and discriminative kernel classifiers," in *Proc. Int. Workshop Comput. Vis. Approaches Med. Image Anal.* Berlin, Germany: Springer, 2011, pp. 1–12.
- [15] R. J. Stanley, W. V. Stoecker, and R. H. Moss, "A relative color approach to color discrimination for malignant melanoma detection in dermoscopy images," *Skin Res. Technol.*, vol. 13, no. 1, pp. 62–72, Feb. 2007.
- [16] Y. Cheng, R. Swamisai, S. E. Umbaugh, R. H. Moss, W. V. Stoecker, S. Teegala, and S. K. Srinivasan, "Skin lesion classification using relative color features," *Skin Res. Technol.*, vol. 14, no. 1, pp. 53–64, 2008.
- [17] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler, "Automated melanoma recognition," *IEEE Trans. Med. Imag.*, vol. 20, no. 3, pp. 233–239, Mar. 2001.
- [18] M. E. Celebi, H. Iyatomi, G. Schaefer, and W. V. Stoecker, "Lesion border detection in dermoscopy images," *Comput. Med. Imag. Graph.*, vol. 33, no. 2, pp. 148–153, Mar. 2009.
- [19] G. Schaefer, B. Krawczyk, M. E. Celebi, and H. Iyatomi, "An ensemble classification approach for melanoma diagnosis," *Memetic Comput.*, vol. 6, no. 4, pp. 233–240, Dec. 2014.
- [20] A. Masood and A. Al-Jumaily, "Computer aided diagnostic support system for skin cancer: A review of techniques and algorithms," *Int. J. Biomed. Imag.*, vol. 2013, Dec. 2013, Art. no. 323268.
- [21] B. Rosado, "Accuracy of computer diagnosis of melanoma: A quantitative meta-analysis," *Arch. Dermatol.*, vol. 139, no. 3, pp. 361–367, Mar. 2003.
- [22] M. Burroni, R. Corona, G. Dell'Eva, F. Sera, R. Bono, P. Puddu, R. Perotti, F. Nobile, L. Andreassi, and P. Rubegni, "Melanoma computer-aided diagnosis: Reliability and feasibility study," *Clin. Cancer Res.*, vol. 10, no. 6, pp. 1881–1886, 2004.
- [23] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, "Diagnostic accuracy of dermoscopy," *Lancet Oncol.*, vol. 3, no. 3, pp. 159–165, 2002.
- [24] D. Gutman, N. C. F. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC)," 2016, *arXiv:1605.01397*. [Online]. Available: <http://arxiv.org/abs/1605.01397>
- [25] M. Binder, H. Kittler, A. Seeber, A. Steiner, H. Pehamberger, and K. Wolff, "Epiluminescence microscopy-based classification of pigmented skin lesions using computerized image analysis and an artificial neural network," *Melanoma Res.*, vol. 8, no. 3, pp. 261–266, Jun. 1998.
- [26] W. H. Clark, D. E. Elder, D. Guerry, L. E. Braitman, B. J. Trock, D. Schultz, M. Synnestvedt, and A. C. Halpern, "Model predicting survival in stage i melanoma based on tumor progression," *JNCI J. Nat. Cancer Inst.*, vol. 81, no. 24, pp. 1893–1904, Dec. 1989.
- [27] T. Schindewolf, W. Stolz, R. Albert, W. Abmair, and H. Harms, "Classification of melanocytic lesions with color and texture analysis using digital image processing," *Anal. Quant. Cytol. Histol.*, vol. 15, no. 1, pp. 1–11, 1993.
- [28] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
- [29] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: <http://arxiv.org/abs/1511.06434>
- [30] M. Y. Liu and O. Tuzel, "Coupled generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 469–477.
- [31] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," 2017, *arXiv:1710.10196*. [Online]. Available: <http://arxiv.org/abs/1710.10196>
- [32] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4401–4410.
- [33] Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, and T. Wang, "Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 10, pp. 2421–2433, Oct. 2015.
- [34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [35] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [36] O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 234–241.
- [37] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J. R. Smith, "Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images," in *Proc. Int. Workshop Mach. Learn. Med. Imag.* Cham, Switzerland: Springer, 2013, pp. 118–126.
- [38] J. Kawahara, A. BenTaieb, and G. Hamarneh, "Deep features to classify skin lesions," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 1397–1400.
- [39] L. Yu, H. Chen, Q. Dou, J. Qin, and P.-A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Trans. Med. Imag.*, vol. 36, no. 4, pp. 994–1004, Apr. 2017.

- [40] J. Yang, F. Xie, H. Fan, Z. Jiang, and J. Liu, "Classification for dermoscopy images using convolutional neural networks based on region average pooling," *IEEE Access*, vol. 6, pp. 65130–65138, 2018.
- [41] K. Matsunaga, A. Hamada, A. Minagawa, and H. Koga, "Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble," 2017, *arXiv:1703.03108*. [Online]. Available: <http://arxiv.org/abs/1703.03108>
- [42] N. Gessert, T. Sentker, F. Madesta, R. Schmitz, H. Kniep, I. Baltruschat, R. Werner, and A. Schlaefer, "Skin lesion diagnosis using ensembles, unscaled multi-crop evaluation and loss weighting," 2018, *arXiv:1808.01694*. [Online]. Available: <http://arxiv.org/abs/1808.01694>
- [43] Z. Wei, H. Song, L. Chen, Q. Li, and G. Han, "Attention-based DenseUNet network with adversarial training for skin lesion segmentation," *IEEE Access*, vol. 7, pp. 136616–136629, 2019.
- [44] C. Gupta, N. K. Gondhi, and P. K. Lehana, "Analysis and identification of dermatological diseases using Gaussian mixture modeling," *IEEE Access*, vol. 7, pp. 99407–99427, 2019.
- [45] P. M. Burlina, N. Joshi, K. D. Pacheco, T. A. Liu and N. M. Bressler, "Assessment of deep generative models for high-resolution synthetic retinal image generation of age-related macular degeneration," *JAMA Ophthalmol.*, vol. 137, no. 3, pp. 258–264, 2019.
- [46] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, no. 1, Dec. 2018, Art. no. 180161.
- [47] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 168–172.
- [48] M. Combalia, N. C. F. Codella, V. Rotemberg, B. Helba, V. Vilaplana, O. Reiter, C. Carrera, A. Barreiro, A. C. Halpern, S. Puig, and J. Malvehy, "BCN2000: Dermoscopic lesions in the wild," 2019, *arXiv:1908.02288*. [Online]. Available: <http://arxiv.org/abs/1908.02288>



AN GONG received the B.S. degree in computer and applications, and the M.S. degree in oil and gas field development project from the China University of Petroleum (East China), Qingdao, China, in 1993 and 2001, respectively. He is currently an Associate Professor with the College of Computer Science and Technology, China University of Petroleum (East China). His research field includes big data intelligent processing.



XINJIE YAO received the B.S. degree in information security from North China Electric Power University, Baoding, China, in 2018. He is currently pursuing the M.S. degree in computer technology, with the China University of Petroleum (East China), Qingdao, China, under the supervision of Prof. A. Gong. He is the first inventor of 8 software. In 2016, he was designated as meritorious winner in Mathematical Contest in Modeling (MCM). His research field includes computer vision.



WEI LIN received the B.S. degree in petroleum engineering from the China University of Geosciences, Beijing, China, in 2015. He is currently pursuing the Ph.D. degree in fluid mechanics with the University of Chinese Academy of Sciences, Beijing, under the supervision of Prof. X. Li and Prof. Z. Yang. From 2018 to 2019, he started a 12-month visit studying with the Department of Earth and Planetary Science, University of California at Berkeley, Berkeley, under the supervision of Prof. M. Manga. His collaborative study is a combination of developing models and percolation theory for tight rocks and volcanic rocks, and performing and analyzing x-ray tomography images of rocks using advanced light source (ALS) with the Lawrence Berkeley National Laboratory. He has authored more than 33 articles, and is the first inventor of four patents and 1 software. His research interests include physical and chemical transport in porous materials, modeling and characterization of porous media, multiscale matter, tertiary oil recovery including oilfield chemistry, and microbial enhanced oil recovery. Dr. Lin's awards and honors include the CAS President Award (Chinese Academy of Sciences, CAS), the Chinese Government Scholarship (China Scholarship Council), and the National Scholarship for Doctoral Students (Ministry of Education, China). He is a Reviewer of *Marine and Petroleum Geology*, the *Journal of Petroleum Science & Engineering*, *IEEE ACCESS*, *Energies*, *RSC Advances*, the *Journal of Dispersion Science and Technology*, *Heliyon*, and the *International Journal of Mechanics Research*. He was the SPE President of Research Institute of Petroleum Langfang SPE Student Chapter. Since 2019, he has been serving as the Bentham Ambassador of Bentham Science Publishers Ltd: Sharjah.

• • •