IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 9, 2019, accepted August 21, 2019, date of publication September 3, 2019, date of current version September 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2939167

Multi-View Convolutional Neural Networks for Mammographic Image Classification

LILEI SUN[1], JUNQIAN WANG[2], ZHIJUN HU[1,3], YONG XU[3], AND ZHONGWEI CUI[4,5]
[1]College of Computer Science and Technology, Guizhou University, Guiyang 550025, China
[2]Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China
[3]State Key Laboratory of Public Big Data, Guizhou University, Guiyang 550025, China
[4]School of Mathematics and Big Data, Guizhou Education University, Guiyang 550018, China
[5]Big Data Science and Intelligent Engineering Research Institute, Guizhou Education University, Guiyang 550018, China

Corresponding author: Yong Xu (laterfall286@yahoo.com)

This work was supported in part by the Foundation of State Key Laboratory of Public Big Data under Grant 2018BDKFJJ001, in part by the Guizhou Science Contact Major Program under Grant [2018]3001, and in part by the Key Supported Disciplines of Guizhou Province—Computer Application Technology under Grant QianXueWeiHeZi ZDXK[2016]20.

ABSTRACT In recent years, deep learning has been widely applied for mammographic image classification. However, most of the existing methods are based on a single mammography view and cannot sufficiently extract discriminative features, thereby resulting in an unsatisfactory classification accuracy. To solve this problem and improve the mammographic image classification performance, we propose a novel multi-view convolutional neural network (CNN) based on multiple mammography views in this paper. Considering that the images acquired from different perspectives contain different and complementary breast mass information, we modify the CNN architecture to exploit the complementary information from the various views of mammography. The new architecture can extract discriminative features from the mediolateral oblique (MLO) and craniocaudal (CC) views of the mammographic images and can effectively incorporate these features for mammographic images. The dilated convolutional layers enable the feature maps extracted from the multiple breast mass views to capture information from a large ''field of vision''. Moreover, multi-scale features are obtained by using the convolutional and dilated convolutions. In addition, we incorporate a penalty term into the cross entropy loss function, which enables the model evolution to reduce the misclassification rate by enhancing the contributions of the samples misclassified in the training process. The proposed method was evaluated and compared with several state-of-the-art methods on the open Digital Database for Screening Mammography (DDSM) and Mammographic Image Analysis Society (MIAS) datasets. The experimental results show that the proposed method exhibits a better performance than those of the state-of-the-art methods.

INDEX TERMS Medical image processing, mammographic image, deep learning, convolutional neural network.

I. INTRODUCTION

Breast cancer is one of the worst cancers due to its highest rates of morbidity and fatality for women. Fortunately, more than 90% of patients can be cured in the early stage of breast cancer. Accordingly, the early detection and treatment of breast cancer are essential for saving the lives of women. In the field of breast cancer diagnosis, medical screenings, such as magnetic resonance imaging (MRI), ultrasonic imaging and molybdenum target X imaging, are the most popular medical imaging methods for breast cancer detection [1]. Among these medical screenings, molybdenum mammography has the advantages of low cost, convenience of operation, and low harm to patients.

Automatic computer-aided diagnosis of breast cancer with mammography can not only help radiologists accelerate the diagnostic process of breast examination but can also increase the accuracy of breast cancer detection and save precious medical resources. Considering the performance of potential features that have been extracted from mammographic images [2], the use of machine learning methods has been a controversial issue for benign or malignant breast mass

The associate editor coordinating the review of this article and approving it for publication was Jingchang Huang.

classification in the study of medical image processing. The machine learning methods can be divided into traditional and deep learning methods. The traditional machine learning methods mainly include sparse representation [3], support vector machine (SVM) and K-nearest neighbor (KNN) algorithms. For example, Alarabeyyat *et al.* [4] generated four coefficient matrices from each mammographic image by using wavelet decomposition. Reyad *et al.* [5] divided a full-breast image into $N \times N$ grids, extracted the texture feature of the breast image from each grid using local binary pattern (LBP) [6], and generated the statistical feature from the mammographic image based on discrete wavelet transform (DWT). Classification via the SVM algorithm on the features that are generated by fusing the texture feature and the statistical feature yields excellent classification performance. The traditional machine learning algorithms [7]–[9] are more efficient than nonmachine-learning image feature extraction algorithms, such as the histogram of oriented gradient (HOG) [10] and LBP. However, the feature extraction method must be designed by hand for the traditional machine learning algorithms; therefore, it is difficult to generate a highly effective feature extraction algorithm due to the diversity of the samples [11]. In recent years, the deep learning technique has attracted wide attention [12]–[14]. Due to its powerful automatic image feature extraction capability, the CNN has been widely used in the field of mammography processing. For instance, Jaffar [15] used the CNN to extract features from a full mammographic image and classified the extracted features by SVM. Gardezi *et al.* [16] extracted features from mammographic image by using VGG-16; and then classified the features as benign and malignant breast mass via traditional machine learning algorithms, such as the SVM, logistic classifier and KNN classifier; and they achieved promising mammographic image classification performance. Antropova *et al.* [17] used the method reported in [18] to extract features from mammographic images, then they also extracted features from mammographic images by using VGG19 [19] for mammographic image classification.

Although the abovementioned approaches have obtained outstanding performance in the mammographic image classification task, the following problems still exist: (1) the existing methods have not extracted adequately and essential features from the insufficient mammographic image dataset. (2) There is no network that is specially designed for mammographic image feature extraction according to the characteristics of the images. To overcome these problems, Lévy and Jain [20] used data augmentation and transfer learning methods to weaken the dependence of the deep learning model training. This approach can extract more discriminative features from the augmentative mammographic image dataset. Chougrad *et al.* [21] augmented a mammographic image dataset by randomly transforming and rotating the samples. The CNN exhibits strong performance in the automatic feature extraction from images [22]–[24]. However, training the model on a large-scale image dataset is necessary for making full use of the advantages of the CNN. Although the dependence of the CNN on the large-scale dataset can be weakened and the performance can be improved by introducing data augmentation and transfer learning methods [25], these approaches cannot sufficiently exploit the intrinsic features of the mammographic image dataset. The low-level layer in the CNN can extract the superficial characteristics of an object from the images, such as the edge or corner information, and the high-level layer can extract abstract features, such as the type of objects [26]. Jiao *et al.* [27] extracted features from different layers of a CNN to identify the diverse characteristics of a mammographic image. These researchers extracted features from the Conv-5 layer and Fc-7 layer and subsequently classified the extracted features by using the SVM. Because these methods cannot extract the specific features of mammographic images, the improvement of the accuracy is limited by the insufficient samples for the above methods.

To solve the abovementioned problems, we propose a novel and effective algorithm based on CNN and multiple mammography views for breast mass classification. Compared with the existing classification algorithms for mammographic images, the proposed method has the following contributions:

1) We propose a new method for simultaneously extracting features from multiple views of the same breast mass. Compared with the breast mass classification method based on a single view, the features extracted from multiple views of breast mass have a stronger discriminative ability. In addition, the multiple views also contain more complementary information of the same mass, which is more conducive to improve the accuracy of the breast mass classification.

2) We develop an improved architecture based on a CNN. The new architecture utilizes two subnetworks to extract features from the MLO and CC views of the breast mass images.

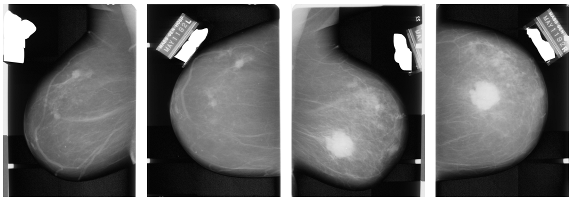
3) We embed the dilated convolution layers with different factors into the proposed novel architecture. The embedded layers expand the receptive fields and increase the diversity of the breast mass images.

4) We add a penalty term to the cross-entropy loss function to improve the performance of the deep learning model. The penalty term increases the difference between the benign and malignant breast mass images, which helps reduce the misclassification rate and improves the classification performance.

The remainder of this paper is organized as follows. Section 2 introduces the relevant literature. Section 3 details the method that presented in this paper. Section 4 conducts experiments and analyzes the experimental results. Section 5 concludes the method presented in this paper.

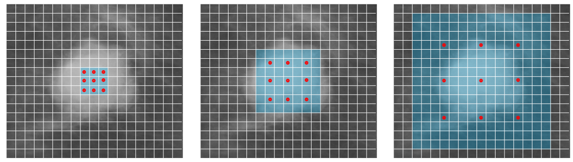
II. RELATED WORK

Over the past few years, several approaches for mammography classification have been proposed to increase the mammographic image classification performance. In this section, we review some advanced algorithms related to the proposed



(a) (b) (c) (d)

FIGURE 1. Mammographic images from a patient. (a) and (b) show the left breast images with the MLO and CC views, respectively. (c) and (d) show the right breast images with the MLO and CC views, respectively.



(a) (b) (c)

FIGURE 2. Receptive field expansion without loss of resolution or coverage on mammographic images. (a) The detailed local information of the mammographic image can be extracted by a dilated convolutional kernel with a small factor of 1; (b) the overall information of the breast mass can be extracted by a dilated convolutional kernel with a medium factor of 2; (c) the overall information of the breast mass with the context information can be extracted by a large dilated convolutional kernel with a factor of 4.

method, which are divided into two types: multi-view and dilated convolution mammographic image feature extraction.

A. MULTI-VIEW MAMMOGRAPHIC IMAGES

The images that provided with two views of the same object contain complementary information [28], in contrast to the images pertaining to a single view. The complementary information is used to describe the breast mass shape, edge morphology, density, degree of calcification and location distribution in increased detail from various perspectives. Consequently, by extracting features from the mammographic images with two views, it is possible to explore superior discriminative information. Fortunately, mammographic images are typically acquired from the same breast with two views: a lateral view called mediolateral oblique (MLO) view and a top head-to-toe view called craniocaudal (CC) view. Fig. 1 shows the mammographic images with the MLO and CC views obtained from a patient. From Fig. 1, we can find that images acquired from different views of the breast mass provide more visual features than those obtained from a single view.

Several scholars have tried to explore the features from the mammographic images with different views. Samulski and Karssemeijer [29] extracted breast masses from the mammographic images with the MLO and CC views by using the geometry-based region matching method, then classified the breast masses by using the multi-view classifier. Velikova *et al.* [30] classified the mammographic images with each view by using Bayesian networks, respectively. Then making decision on the outputs of their Bayesian networks by using logistic regression. Bekker *et al.* [31] used two ANNs to extract the features from the mammographic images with the MLO and CC views. Then they fused the features of the mammographic images from two ANNs for benign or malignant classification. Although the method proposed in the literature [31] can extract features from the mammographic images with various views, it cannot extract features with a strong discriminative performance due to the loss of spatial information caused by the feature transformation from image to a long dimensional vector. To capture complementary information from mammographic images, Carneiro *et al.* [32] proposed a CNN based algorithm to extract features from four full-breast images from the same patient simultaneously. Since a full-breast image contains many anatomical organs, such as pectoralis muscles, milk glands and fat, the features that are extracted from full-breast images cannot capture the crucial information due to the presence of many interfering objects.

B. DILATED CONVOLUTION

The dilated convolution is widely used in image feature extraction due to its good performance in extracting powerful features with various receptive fields from image [33]. The dilated convolution applies a filter with different convolutional regions by employing different dilation factors, which can expand the receptive fields without losing the resolution or coverage. Since the dilated convolution operation expands the receptive field without increasing the connections of the model, the parameters of the model are not increased. Fig. 2 shows the dilated convolutions with different factors on a mammographic image.

For mammographic image classification, Yu and Koltun [34] and Yang *et al.* [35] embedded the dilated convolutional layers into a CNN to capture multi-scale contextual information. Li *et al.* [36] proposed a detection method for identifying the pedestrians, which generates feature maps with various receptive field scales by using the dilated convolutional kernels with various factors. For the full-mammogram diagnosis, Hang *et al.* [37] applied dilated convolutions to maintain a higher resolution of mammographic image while keeping the receptive field size consistent.

In this paper, we design a network architecture to make full use of the characteristics of the mammographic images such that more discriminative information can be naturally captured. The proposed network can extract the complementary information from two perspectives of the same breast mass by using two feature extraction subnetworks, and it can extract various "field of vision" by using the dilated convolutional neural networks with various factors. More details of the proposed approach are described below.

III. PROPOSED METHOD

As mentioned previously, exploiting a mammographic image with multi-view and multi-scale are useful to extract the

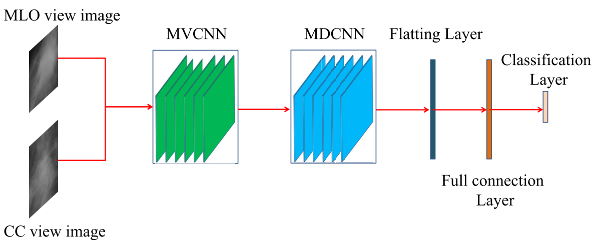


FIGURE 3. Architecture of the MVMDCNN.

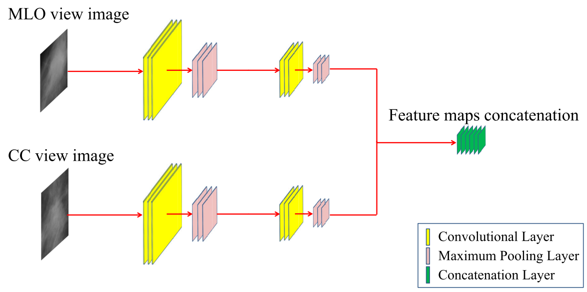


FIGURE 4. Architecture of multi-view mammographic image convolutional neural subnetwork (MVCNN).

features with strong discriminative performance. Inspired by these motivations, we propose a novel approach to increase the classification accuracy of mammographic images. We improve the classification performance of the mammographic images from two phases: improving the architecture of the network and the loss function of the deep learning model. The former is described in Sections 3.1, 3.2 and 3.3, and the latter is described in Section 3.4.

A. ARCHITECTURE OF NETWORK

In this section, we propose a novel CNN based architecture for breast mass classification, which integrates the multi-view mammographic image convolutional neural subnetwork and the multi-dilated convolutional neural subnetwork (MVMDCNN). The architecture of the MVMDCNN is illustrated in Fig. 3. The MVMDCNN contains two inputs, a multi-view mammographic image convolutional neural subnetwork (MVCNN) and a multi-dilated convolutional neural subnetwork (MDCNN). For the proposed network, two breast mass images with the MLO and CC views are simultaneously fed into the MVCNN. Then MVCNN generates the feature maps from multiple views and feeds them into the MDCNN. The MDCNN convolutes the inputted feature maps by using the dilated convolutional layers with factors of 1, 2 and 4 to generate the feature maps with multiple receptive fields. Then, the feature maps that are extracted from the MDCNN are flattened and inputted into a fully connected layer to extract the features of the breast mass image. Finally, inputting the features into a softmax layer to classify the breast mass. MVMDCNN receives two mammographic images with two views and outputs the classification result. As a result, it realizes an end-to-end mammographic image classification algorithm.

The MVMDCNN uses the MVCNN to extract the features from breast mass images with the MLO and CC views. Because the mammographic images are acquired from different perspectives, the complementary breast mass information can be exploited from various views of the mammographic images and are conductive to obtain features that can describe the characteristics of the breast mass more completely. The MDCNN generates the feature maps with multi-scales information from the mammographic images by using the dilated convolutional kernels with different factors. The extracted feature maps can represent the characteristics of the breast mass with a multi-granularity level, which is helpful to extract powerful discrimination information of the breast mass images.

The inputted images are resized to a size of 180×180 pixels. In order to put the feature maps generated by the MDCNN into a full connection layer, we use a flatting layer to flatten the 60 feature maps extracted from previous layers into a one-dimensional vector. A full connection layer contains 1,000 nodes is used to extract the features from the one-dimensional vector generated by the flattening layer. At the end of the architecture of our network, we use a classification layer that contains two nodes to classify the image into benign or malignant mass.

B. MULTI-VIEW CONVOLUTIONAL NEURAL SUBNETWORK

To extract the features with the powerful discriminative information from the mammographic images, a novel feature extraction network, called the MVCNN, is designed, as described in this section. As illustrated in Fig. 4, the MVCNN contains two inputs and one output. It receives two breast mass images with the MLO and CC views simultaneously, and it uses two convolutional neural subnetworks to extract the features from the inputs. The MVCNN outputs the superposed feature maps that are generated by superposing the outputs from the two convolutional neural subnetworks. Because the breast mass images with the MLO and CC views contain complementary information of the breast mass, using two convolutional neural subnetworks to extract the features from these images can help obtain powerful discriminative information compared to that obtained when utilizing a single view.

Each convolutional neural subnetwork in MVCNN uses two convolutional layers to extract the convolutional features from a mammographic image. Each convolutional layer contains 30 3×3 convolutional kernels with a stride of 1 and followed by a maximum pooling layer with a 2×2 filter with a stride of 2. At the end of the network, a concatenation layer is adopted to concatenate the feature maps generated by the two convolutional neural subnetworks. The size of the inputted mammographic image is 180×180 pixels and the outputs of the MVCNN are 60 feature maps with 45×45 pixels.

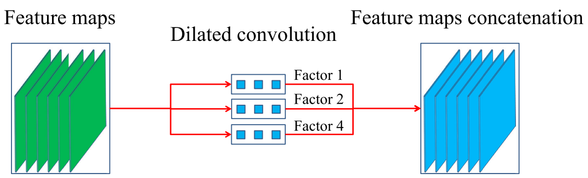


FIGURE 5. Architecture of the multi-dilated convolution subnetwork (MDCNN).

C. MULTI-DILATED CONVOLUTIONAL NEURAL SUBNETWORK

Breast masses vary substantially in terms of the shape, size, and texture. Using the convolutional kernels with different receptive fields to convolute the breast mass images can increase the diversity and complementary information of the features that are extracted from the breast mass images. Furthermore, we can extract more diverse information from the mammographic images. For instance, we can obtain the texture information of the breast mass from the dilated convolutional feature maps with a small factor of 1, and we can obtain the shape of the breast mass from the dilated convolutional feature maps with a large factor of 4.

To identify the features that have a strong discriminative performance from the breast mass images, a network called the MDCNN is designed, which is able to capture more valuable features by using the multiple dilated convolutional kernels with different factors. The architecture of the MDCNN is illustrated in Fig. 5. The MDCNN has one input and one output. It receives the feature maps that are generated by the previous layer, uses multiple dilated convolutional layers with dilated factors of 1, 2 and 4 to convolute the inputs, and obtains the dilated feature maps with various convolutional receptive fields. In the end, the MDCNN outputs the feature maps that are generated by superposing the dilated feature maps from the multiple dilated convolutional layers. Because the dilated convolutional kernels with different factors can extract the feature maps with various receptive fields, the feature maps with multi-granularity levels contain more detailed and richer feature information, which is beneficial to improve the classification performance of the mammographic images.

Considering that the use of a dilated kernel with a too big factor is not conducive to the extraction of the effective convolutional features from the breast mass images, we use the dilated convolutional kernels with small factors of 1, 2 and 4 to extract the feature maps from the inputs. The inputs of the MDCNN are 60 feature maps extracted from the previous layer. The three dilated convolutional layers in the MDCNN extract the feature maps from the inputs via 20 3×3 dilated convolutional kernels with factors of 1, 2 and 4. The outputs of the MDCNN are 60 feature maps obtained by concatenating the outputs of these dilated convolutional layers.

D. LOSS FUNCTION

In addition to improving the network architecture, we also modify the loss function to improve the breast mass image classification performance. Various loss functions, such as the mean square (MSE) and cross entropy loss functions, are widely used in the training process of the deep learning model. These loss functions can guide the evolution direction of the deep learning model to reduce the error rate on all the samples. Therefore, all the samples, including the samples with correct and incorrect classification in the dataset, contribute equally to the model training. In fact, a sample that is classified correctly has a smaller contribution than a sample that is misclassified to the model parameters. If we guide the evolution direction of the model to focus on the misclassified samples and reduce the contributions of the samples that have been classified correctly, that is, we reduce the misclassification rate of the model by amplifying the contribution weight of the previously misclassified samples, the discriminative performance of the model can be enhanced. We introduce a penalty term that guides the process of the model training to focus on the samples that are misclassified. The contributions of the misclassified samples are increased by the additional penalty term in the model training process. The penalty term is defined as follows:

$$P = \frac{\sum_{i=1}^{n} abs\,(label_i - prediction_i)}{n}, \tag{1}$$

where n is the number of misclassified samples in an epoch on the training dataset. The penalty term calculates the sum of the absolute differences between the labels and the prediction probabilities of all the misclassified samples, and it is used to increase the weights of the misclassified samples in the training process. The loss function is defined as follows:

$$J = CrossEntropy + \lambda P, \tag{2}$$

where $CrossEntropy$ is the cross entropy formula, P is the penalty term, and λ is a hyper-parameter that modulates the contribution weights of all the misclassified samples in the training process.

For improving the classification performance of the mammographic images, we train the MVMDCNN model with our proposed improved loss function J and refer to this approach as the MVMDCNN-LOSS.

IV. RESULTS AND DISCUSSIONS

We evaluate our approach on two public datasets: DDSM and MIAS [38]. In our work, we first describe the evaluation datasets, experimental setting and experimental implementation of the proposed approach. Then we analyze the experimental results of the proposed method and the state-of-the-art methods for mammographic image classification in detail.

A. DATASETS

The Digital Database for Screening Mammography (DDSM) [39] is a mammographic image dataset, which contains approximately 2,600 patients' breast medical images, each patient provides 4 images from the MLO and CC views of the left and right breasts, respectively.

In this section, a subset with 1,445 samples is selected from the DDSM database for evaluating the different classification methods. Specifically, the subset contains 747 benign samples and 698 malignant samples. The k-fold cross-validation scheme is widely used to evaluate the performance of the classification methods in the medical image processing and deep learning research [40], [41]. In our experiments, we adopt the seven-fold cross-validation strategy and choose the average of the accuracies of seven folds to evaluate all the mammographic image classification methods. The training and test datasets in the first six folds contain 85% and 15% of the samples, respectively. In other words, there exist 1,228 training samples and 217 test samples in the first six folds. In the last fold, 90% of the samples (i.e., 1,300 samples) are selected as the training samples and the remaining samples (i.e., 145 samples) are regarded as the test samples. Simple random sampling is performed on the subset by dividing the samples into seven portions by percentages without overlapping. Specifically, the first six training folds contain 635 benign and 593 malignant breast masses, and the corresponding test folds contain 112 benign and 105 malignant breast masses. The last training and test folds contain 672 benign and 628 malignant breast masses, 75 benign and 70 malignant breast masses, respectively.

B. EXPERIMENTAL SETTING

The breast mass images that are used as the training and test datasets are resized to 180 × 180 pixels. The weights of the network are randomly initialized and the learning rate is set as 10^{-4}. The contribution weights of all the misclassified samples in the proposed loss function λ is set as 10^{-3}. The batch size is set as 128 in the training process. The stop criterion of our network is set as the condition at which the training accuracy equals 100%. The maximum number of training epochs for our network is set as 200. All methods mentioned in this paper are optimized by Adam algorithm [42]. In this work, TensorFlow [43] is chosen as the framework to evaluate all compared methods and our proposed method in this paper. The proposed method is evaluated by using the following software environment: Ubuntu 16.04, python 3.5 and TensorFlow 1.8.0; the hardware environment is as follows: Intel i7-8700K, RAM 16G and NVIDIA GeForce GTX 1080 Ti GPU.

C. ACCURACY COMPARISON AND ANALYSIS

The LeNet-BN is based on LeNet, which adds a BN layer after every convolutional layer to normalize the outputs of the convolutional layer to the same data distribution. From Table 1, LeNet and LeNet-BN obtain accuracies of 0.7142 and 0.7742, respectively. This finding demonstrates that the BN layer used in the LeNet-BN is beneficial to improve the classification performance. Accordingly, we add a BN layer after every convolutional layer in our method. Specifically, from the experimental results listed in Table 1, the accuracies of our methods, namely, MVCNN, MVMDCNN and MVMDCNN-LOSS are 0.8018, 0.8156 and 0.8202, respectively. The accuracy improvements of our methods are 8.76%-10.60% over the baseline LeNet model and 2.76%-4.60% over the LeNet-BN. This aspect illustrates that the proposed multi-view based methods can extract more powerful features from two different views of the breast mass images simultaneously compared to those obtained by the LeNet and LeNet-BN. The breast mass images with the MLO and CC views from the DDSM are presented in Fig. 6, which shows the morphological characteristics of the same breast mass with different views. From the breast mass images with the MLO and CC views, we can extract more useful information. For instance, we can obtain the shape of the breast mass from the CC view of sample (c), determine the prominent texture of the breast mass from the MLO view of sample (c), and capture more detailed edge information from the MLO and CC views of sample (b) simultaneously.

TABLE 1. Comparison of different mammography classification methods based on LeNet on the DDSM.

Method	Accuracy
LeNet [44]	0.7142
LeNet-BN	0.7742
MVCNN	0.8018
MVMDCNN	0.8156
MVMDCNN-LOSS	0.8202

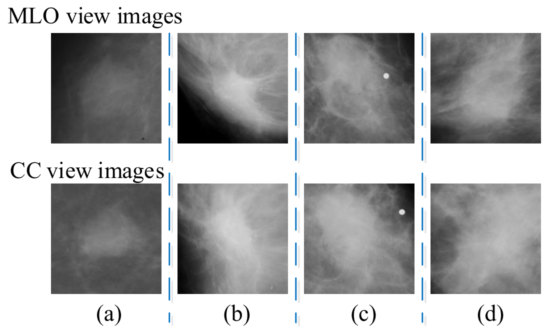


FIGURE 6. Comparison of the breast mass images with different perspectives. The images in the first and second rows are cropped images with the MLO and CC views of the mammography, respectively. The images in the different columns are cropped from different breast masses: columns (a) and (b) correspond to benign breast masses and (c) and (d) are malignant breast masses.

Benefiting from the feature maps extracted from the breast mass images by employing various receptive fields can represent diverse information of the breast mass, the MVMDCNN achieves an accuracy of 0.8156, which is 1.38% higher than that of the MVCNN. The better performance of the MVMDCNN in comparison with that of the MVCNN demonstrates that introducing a dilated convolutional layer is beneficial to improving the classification performance. Fig. 7 shows the morphological characteristics of the same breast mass with different factors dilated convolutional kernels. The figure also illustrates that the MVMDCNN can extract more diversified features from the breast mass image.

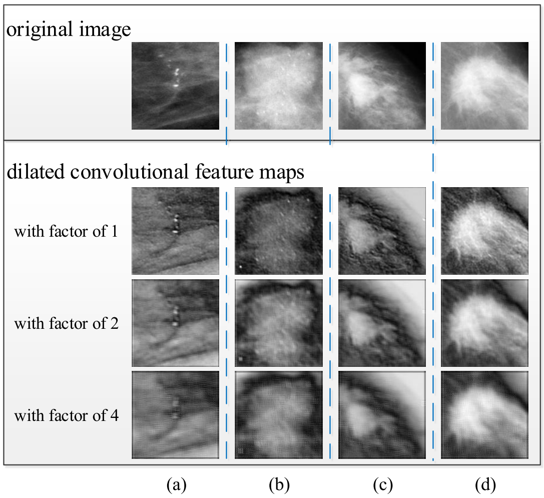


FIGURE 7. Comparison of the dilated feature maps captured with different factors from breast mass images. The images in the first row are the original breast mass images. The images in the second, third and fourth rows are the dilated feature maps captured with factors of 1, 2 and 4, respectively. The images in different columns are cropped from different breast masses: Columns (a) and (b) correspond to benign breast masses, columns (c) and (d) are the malignant breast masses.

TABLE 2. Comparison of different classification methods on the DDSM.

Method	Accuracy	Year
ANNS-MV [31]	0.5990	2016
CNN+SVM [15]	0.7372	2017
CNN(Conv5+Fc7)+SVM [27]	0.7511	2016
CNN(Conv5+Fc7)+BN+SVM	0.8018	-
AlexNet [45]	0.7244	2012
ZFNet [26]	0.7546	2014
VGG16 [19]	0.7617	2014
VGG19 [19]	0.7623	2014
ResNet101 [46]	0.7808	2016
DenseNet121 [47]	0.7789	2017
InceptionV4 [48]	0.7872	2017
MobileNetV2 [49]	0.7913	2018
ShuffleNetV2 [50]	0.7980	2018
MVMDCNN-LOSS	0.8202	-

Compared with the MVMDCNN which uses the standard cross entropy loss function, the MVMDCNN-LOSS obtains an accuracy improvement of 0.46%. This finding proves the effectiveness of the modified loss function in guiding the evolution direction of the classification model to reduce the misclassification error. In addition, Table 1 and Table 2 show that the accuracies of our improved LeNet based models, i.e., MVMDCNN and MVMDCNN-LOSS are 10.14% and 10.60% higher than that of the baseline model, respectively. In summary, the two modified approaches exploiting architecture improvement and the approach with the modified loss function are all effective and beneficial for further enhancing the performance of the breast mass classification task.

The experimental results of the proposed method and the compared state-of-the-art methods on the DDSM dataset are listed in Table 2. The ANNS-MV uses two ANNs to extract the features from the breast mass images with the MLO and CC views. Then, an ANN is used to classify the fusion features extracted from the aforementioned ANNs as a benign or malignant breast mass. The ANNS-MV achieves a low accuracy due to the loss of the spatial information which is caused by flattening a two-dimensional image to a vector.

The CNN+SVM uses the SVM to classify the features extracted from the breast mass images via the CNN. Since the CNN has a powerful automatic feature extraction capability and the SVM has a superior generalization performance for classification, the accuracy of the CNN+SVM is 2.30% higher than that of LeNet. The CNN(Conv5+Fc7)+SVM uses the SVM to classify the fusion feature by concatenating the outputs of the fifth computing layer (convolutional layer) and the last computing layer (full-connected layer). Since the features extracted from the various layers contain more diverse information, which is conductive to obtaining features with a strong discriminative performance, the CNN(Conv5+Fc7)+SVM achieves an accuracy improvement of 1.39% in comparison with that of the CNN+SVM that extracts features from only the last layer. The BN layer can improve the generalization performance of the CNN effectively. Therefore, we implement the CNN(Conv5+Fc7)+SVM by appending a BN layer after every convolutional layer, where the resulting network is denoted as the CNN(Conv5+Fc7)+BN+SVM. According to the experimental results, the accuracy of the CNN(Conv5+Fc7)+BN+SVM is 5.07% higher than that of the CNN(Conv5+Fc7)+SVM.

The state-of-the-art deep learning classification methods, such as AlexNet, ZFNet, VGG16, VGG19, ResNet, DenseNet, InceptionV4, MobileNetV2 and ShuffleNetV2 have a powerful feature extraction capability. From the experimental results listed in Table 1 and Table 2, we can find that the accuracies of the AlexNet and ZFNet are 1.02% and 4.04% higher than that of the baseline LeNet model. This finding indicates that using small filters to convolute the breast mass image can help retain more original image information. Because the deeper networks can extract more abstract image features than shallow networks and the abstract features facilitate the extraction of more discriminating features, the classification methods with deeper networks, i.e., VGG16, VGG19, ResNet and DenseNet achieve accuracy improvements of 4.75%-6.66% over that of the baseline LeNet model. The MobileNetV2 and ShuffleNetV2 are light-weight networks, which have fewer parameters than other typical CNN networks with complex architectures. The accuracies of the MobileNetV2 and ShuffleNetV2 are 7.71% and 8.38% higher than that of the baseline LeNet model, respectively.

The proposed method extracts features with multiple views and various receptive fields from breast mass images. This strategy can enhance the discriminate performance of the features effectively. Furthermore, the penalty term incorporated into the cross entropy loss function increases the weights of the samples that are misclassified in the training process.

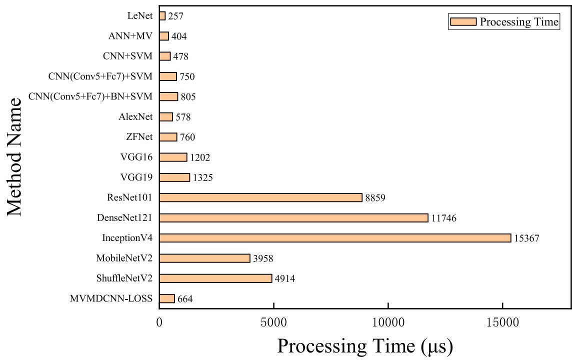


FIGURE 8. Processing time for breast mass image classification tasks: Comparison with the state-of-the-art classification methods on DDSM.

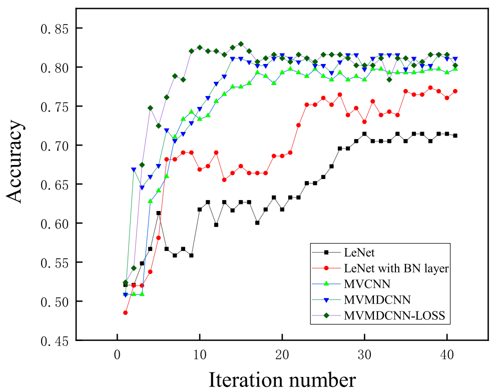


FIGURE 9. Convergence rates in the training processes of the baseline model and our methods on the DDSM dataset.

This strategy enables the model to reduce the misclassification rate. Consequently, the MVMDCNN-LOSS can identify the discriminative features effectively. The experimental results show that our method achieves the highest accuracy against the compared state-of-the-art methods for mammographic image classification on the DDSM.

The processing times of the different classification methods are shown in Fig. 8. Compared with the baseline LeNet model, the processing time of our MVMDCNN-LOSS method is higher than that of the baseline model. However, as listed in Table1 and Table 2, the accuracy of our method is 10.60% higher than that of the baseline model. In particular, from the experimental results listed in Table 1 and Table 2 and shown in Fig. 8, we can find that compared with some of the state-of-the-art classification methods, such as ResNet, DenseNet and InceptionV4, our method not only achieves the highest classification accuracy, but also is the most efficient due to our method contains less computing layers than the other methods. In summary, our method demonstrates a better performance than that of the state-of-the-art methods for the mammographic image classification mentioned in this paper.

D. CONVERGENCE RATE ANALYSIS

The baseline model and the proposed method have large differences in terms of their convergence rates in the training process. The convergence rates of the baseline model and the proposed method on the DDSM are plotted in Fig. 9.

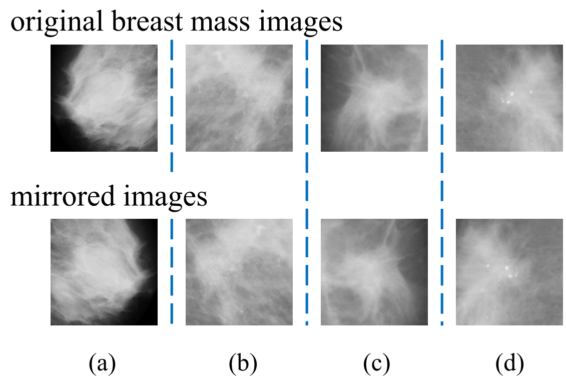


FIGURE 10. Comparison of the original breast mass images and their flipped images. The images in the first and second rows are the original images and the corresponding flipped images, respectively. The images in different columns are cropped from different breast masses: Columns (a) and (b) correspond to benign breast masses, columns (c) and (d) are malignant breast masses.

The LeNet converges at approximately iteration 30. The LeNet-BN converges faster than the LeNet. This finding illustrates that introducing the BN layer that follows by every convolutional layer is beneficial to reduce the difficulty of adapting multiple data distributions. The convergence rate of the LeNet-BN is slower than those of our methods. This phenomenon is mainly observed because the LeNet-BN diminishes the data consistency and increases the difficulty of classification by randomly inputting the breast mass images from different perspectives. The proposed methods converge at approximately iteration 20, which is faster than the convergence of the compared methods for the mammographic image classification.

E. ROBUSTNESS EVALUATION

To verify the robustness of the proposed method, we compare our method with the state-of-the-art methods for the mammographic image classification task on the MIAS. Since the MIAS contains only 322 mammographic images and the deep learning technique is a data-driven algorithm, it is difficult to realize its advantage on this small dataset. Consequently, we use the models that are trained on the DDSM to verify the samples that are extracted from the MIAS.

The images in the MIAS were photographed with the MLO view only. To adapt the two inputs of the proposed methods, we generated the breast mass image with the CC view from each breast mass image with the MLO view in the MIAS. Since the two images of the MLO and CC views are acquired from the same breast mass with different perspectives, the images reflect multiple visual features of the same breast mass. Inspired by the mirror image generation method in [51], we generated the mirror images from the breast mass images of the MLO view by flipping them horizontally. The mirror image can be viewed as a representation of the breast mass that is opposite to the corresponding original breast mass image to some extent. The original breast mass images and the corresponding mirror images are shown in Fig. 10.

TABLE 3. Comparison with state-of-the-art mammography classification methods on the MIAS.

Method	Accuracy
LeNet	0.5760
CNN+SVM	0.6036
CNN(Conv5+Fc7)+SVM	0.6129
CNN(Conv5+Fc7)+BN+SVM	0.6221
AlexNet	0.6213
ZFNet	0.6295
VGG16	0.6203
VGG19	0.6272
ResNet101	0.6061
DenseNet121	0.5870
InceptionV4	0.5947
MobileNetV2	0.6222
ShuffleNetV2	0.6178
MVCNN	0.6396
MVMDCNN	0.6306
MVMDCNN-LOSS	0.6306

The experimental results of the different methods on the MIAS are listed in Table 3, we can find that the proposed method is more robust than the compared state-of-the-art methods. The good performance of the proposed methods is attributing to the effective feature extraction capability from the breast mass images of our architecture. The accuracies of the MVMDCNN and MVMDCNN-LOSS are 0.9% lower than that of the MVCNN. This finding indicates that as the complexity of the model increases, the generalization ability of the model decreases. In summary, our method obtains a higher classification accuracy than those of the compared state-of-the-art classification methods for mammographic images on the MIAS and exhibits the highest robustness on the DDSM and MIAS mammographic image datasets.

V. CONCLUSION

In this paper, we proposed a novel method for mammographic image classification. The proposed method integrates the MVCNN and MDCNN to extract features from multiple views of the breast mass images. The network can extract the features with complementary and diverse information, which can help increase the classification accuracy and robustness of the model. We modified the cross-entropy loss function by adding a penalty term that can enhance the contribution weights of the misclassified samples in the training process. The modified loss function guides the evolution direction of the model to minimize the misclassification error and increase the classification accuracy. We evaluated the proposed method and the compared state-of-the-art classification methods on the two well-known mammographic image datasets, i.e., DDSM and MIAS. The experimental results demonstrated that the proposed method can outperform the compared state-of-the-art methods for mammographic image classification.

REFERENCES

[1] S. Xu, H. Wu, and R. Bie, "CXNet-m1: Anomaly detection on chest X-rays with image-based deep learning," *IEEE Access*, vol. 7, pp. 4466–4477, 2018.
[2] Y. Akhtar and D. P. Mukherjee, "Detection of architectural distortion from the ridges in a digitized mammogram," *Signal, Image Video Process.*, vol. 12, no. 7, pp. 1285–1292, 2018.
[3] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: Algorithms and applications," *IEEE Access*, vol. 3, pp. 490–530, 2015.
[4] M. R. Al-Hadidi, A. Alarabeyyat, and M. Alhanahnah, "Breast cancer detection using k-nearest neighbor machine learning algorithm," in *Proc. 9th Int. Conf. Develop. eSyst. Eng. (DeSE)*, Aug./Sep. 2016, pp. 35–39.
[5] Y. A. Reyad, M. A. Berbar, and M. Hussain, "Comparison of statistical, LBP, and multi-resolution analysis features for breast mass classification," *J. Med. Syst.*, vol. 38, no. 9, p. 100, 2014.
[6] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
[7] Y.-Q. Song, Z. Liu, J.-M. Chen, F. Zhu, and C.-H. Xie, "Medical image segmentation based on non-parametric mixture models with spatial information," *Signal, Image Video Process.*, vol. 6, no. 4, pp. 569–578, 2012.
[8] J. Wen, Y. Xu, and H. Liu, "Incomplete multiview spectral clustering with adaptive graph learning," *IEEE Trans. Cybern.*, to be published.
[9] P. S. Mukherjee, "A multi-resolution and adaptive 3-D image denoising framework with applications in medical imaging," *Signal, Image Video Process.*, vol. 11, no. 7, pp. 1379–1387, 2017.
[10] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, vol. 1, no. 1, pp. 886–893.
[11] J. Wen, X. Fang, J. Cui, L. Fei, K. Yan, Y. Chen, and Y. Xu, "Robust sparse linear discriminant analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 2, pp. 390–403, Feb. 2019.
[12] C. Tian, Y. Xu, L. Fei, J. Wang, J. Wen, and N. Luo, "Enhanced CNN for image denoising," *CAAI Trans. Intell. Technol.*, vol. 4, no. 1, pp. 17–23, Mar. 2019.
[13] K. Guo, S. Wu, and Y. Xu, "Face recognition using both visible light image and near-infrared image and a deep network," *CAAI Trans. Intell. Technol.*, vol. 2, no. 1, pp. 39–47, 2017.
[14] T. Akilan, Q. M. J. Wu, W. Jiang, A. Safaei, and J. Huo, "New trend in video foreground detection using deep learning," in *Proc. IEEE 61st Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2018, pp. 889–892.
[15] M. A. Jaffar, "Deep learning based computer aided diagnosis system for breast mammograms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 7, pp. 286–290, 2017.
[16] S. J. S. Gardezi, M. Awais, I. Faye, and F. Meriaudeau, "Mammogram classification using deep learning features," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Sep. 2017, pp. 485–488.
[17] N. Antropova, B. Q. Huynh, and M. L. Giger, "A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets," *Med. Phys.*, vol. 44, no. 10, pp. 5162–5171, 2017.
[18] L. Zheng, Y. Zhao, S. Wang, J. Wang, and Q. Tian, "Good practice in CNN feature transfer," 2016, *arXiv:1604.00133*. [Online]. Available: https://arxiv.org/abs/1604.00133
[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556
[20] D. Lévy and A. Jain, "Breast mass classification from mammograms using deep convolutional neural networks," 2016, *arXiv:1612.00542*. [Online]. Available: https://arxiv.org/abs/1612.00542
[21] H. Chougrad, H. Zouaki, and O. Alheyane, "Deep convolutional neural networks for breast cancer screening," *Comput. Methods Programs Biomed.*, vol. 157, pp. 19–30, Apr. 2018.
[22] C. Yuan, Z. Xia, L. Jiang, Y. Cao, Q. M. J. Wu, and X. Sun, "Fingerprint liveness detection using an improved CNN with image scale equalization," *IEEE Access*, vol. 7, pp. 26953–26966, 2019.
[23] C. Tian, Y. Xu, L. Fei, and K. Yan, "Deep learning for image denoising: A survey," in *Proc. Int. Conf. Genetic Evol. Comput.*, 2018, pp. 563–572.
[24] Q. Yin, J. Wang, X. Luo, J. Zhai, S. K. Jha, and Y.-Q. Shi, "Quaternion convolutional neural network for color image classification and forensics," *IEEE Access*, vol. 7, pp. 20293–20301, 2019.
[25] R. K. Samala, H.-P. Chan, L. Hadjiiski, M. A. Helvie, J. Wei, and K. Cha, "Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography," *Med. Phys.*, vol. 43, no. 12, pp. 6654–6666, Dec. 2016.
[26] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 818–833.

[27] Z. Jiao, X. Gao, Y. Wang, and J. Li, "A deep feature based framework for breast masses classification," *Neurocomputing*, vol. 197, pp. 221–231, Jul. 2016.
[28] T. Akilan, "A foreground inference network for video surveillance using multi-view receptive field," 2018, *arXiv:1801.06593*. [Online]. Available: https://arxiv.org/abs/1801.06593
[29] M. Samulski and N. Karssemeijer, "Optimizing case-based detection performance in a multiview CAD system for mammography," *IEEE Trans. Med. Imag.*, vol. 30, no. 4, pp. 1001–1009, Apr. 2011.
[30] M. Velikova, M. Samulski, P. J. F. Lucas, and N. Karssemeijer, "Improved mammographic CAD performance using multi-view information: A Bayesian network framework," *Phys. Med. Biol.*, vol. 54, no. 5, p. 1131, 2009.
[31] A. J. Bekker, H. Greenspan, and J. Goldberger, "A multi-view deep learning architecture for classification of breast microcalcifications," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 726–730.
[32] G. Carneiro, J. Nascimento, and A. P. Bradley, "Unregistered multiview mammogram analysis with pre-trained deep learning models," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 652–660.
[33] Y. Wang, Z. Yu, and L. Zhu, "Foreground detection with deeply learned multi-scale spatial-temporal features," *Sensors*, vol. 18, no. 12, p. 4269, Dec. 2018.
[34] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*. [Online]. Available: https://arxiv.org/abs/1511.07122
[35] Z. Yang, Z. Hu, R. Salakhutdinov, and T. Berg-Kirkpatrick, "Improved variational autoencoders for text modeling using dilated convolutions," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 3881–3890.
[36] J. Li, Y. Wu, J. Zhao, L. Guan, C. Ye, and T. Yang, "Pedestrian detection with dilated convolution, region proposal network and boosted decision trees," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 4052–4057.
[37] W. Hang, Z. Liu, and A. Hannun, "GlimpseNet: Attentional methods for full-image mammogram diagnosis," Stanford AI Lab Internal Rep., Stanford Univ., Stanford, CA, USA, 2017.
[38] S. J. P, "The mammographic image analysis society digital mammogram database," *Digit. Mammo*, pp. 375–386, Jul. 1994.
[39] M. Heath, K. Bowyer, D. Kopans, R. Moore, and P. Kegelmeyer, Jr., "The digital database for screening mammography," in *Proc. 5th Int. Workshop Digit. Mammography*, 2000, pp. 212–218.
[40] R. Ben-Ari, A. Akselrod-Ballin, L. Karlinsky, and S. Hashoul, "Domain specific convolutional neural nets for detection of architectural distortion in mammograms," in *Proc. IEEE 14th Int. Symp. Biomed. Imag.*, Apr. 2017, pp. 552–556.
[41] C. R. Naguri and R. C. Bunescu, "Recognition of dynamic hand gestures from 3D motion data using LSTM and CNN architectures," in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2017, pp. 1130–1133.
[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980
[43] M. Abadi *et al.*. "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," Mar. 2016, *arXiv:1603.04467*. [Online]. Available: https://arxiv.org/abs/1603.04467
[44] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[45] 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.
[46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[47] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 4700–4708.
[48] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4284.
[49] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
[50] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jun. 2018, pp. 116–131.
[51] Y. Xu, X. Li, J. Yang, and D. Zhang, "Integrate the original face image and its mirror image for face recognition," *Neurocomputing*, vol. 131, no. 5, pp. 191–199, May 2014.

LILEI SUN received the master's degree in computer technology from Guizhou University, in 2016, where he is currently pursuing the Ph.D. degree with the College of Computer Science and Technology. His research interests include pattern recognition and biomedical image processing.

JUNQIAN WANG received the master's degree from the Harbin Institute of Technology, in 2018. She is currently pursuing the Ph.D. degree with the Bio-Computing Research Center, College of Computer Science and Technology, Harbin Institute of Technology (Shenzhen). Her research interests include pattern recognition, computer vision, deep learning, and biomedical image processing.

ZHIJUN HU received the master's degree in computational mathematics from Chongqing University, in 2009. He is currently pursuing the Ph.D. degree with the College of Computer Science and Technology, Guizhou University. His research interests include machine learning and computer vision.

YONG XU received the Ph.D. degree in pattern recognition and intelligence systems from the Nanjing University of Science and Technology (NUST), China, in 2005. He is currently with Guizhou University. His current research interests include pattern recognition, biometrics, and machine learning.

ZHONGWEI CUI received the master's degree in computer application technology from Guizhou University, Guiyang, China, in 2008, where he is currently pursuing the Ph.D. degree. He has 11 years of teaching experience. Since December 2013, he has been an Associate Professor with the School of Mathematics and Big Data, Guizhou Education University, Guiyang. His research interests include machine vision and wireless networks.