IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 16, 2020, accepted January 23, 2020, date of publication February 3, 2020, date of current version February 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2971225

Medical Image Classification Using a Light-Weighted Hybrid Neural Network Based on PCANet and DenseNet

ZHIWEN HUANG, XINGXING ZHU, MINGYUE DING, (Senior Member, IEEE), AND XUMING ZHANG

Key Laboratory of Molecular Biophysics, Department of Biomedical Engineering, School of Life Science and Technology, Ministry of Education, Huazhong University of Science and Technology, Wuhan 430074, China

Corresponding author: Xuming Zhang (zxmboshi@hust.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB1303100, and in part by the National Natural Science Foundation of China under Grant 61871440.

ABSTRACT Medical image classification plays an important role in disease diagnosis since it can provide important reference information for doctors. The supervised convolutional neural networks (CNNs) such as DenseNet provide the versatile and effective method for medical image classification tasks, but they require large amounts of data with labels and involve complex and time-consuming training process. The unsupervised CNNs such as principal component analysis network (PCANet) need no labels for training but cannot provide desirable classification accuracy. To realize the accurate medical image classification in the case of a small training dataset, we have proposed a light-weighted hybrid neural network which consists of a modified PCANet cascaded with a simplified DenseNet. The modified PCANet has two stages, in which the network produces the effective feature maps at each stage by convoluting inputs with various learned kernels. The following simplified DenseNet with a small number of weights will take all feature maps produced by the PCANet as inputs and employ the dense shortcut connections to realize accurate medical image classification. To appreciate the performance of the proposed method, some experiments have been done on mammography and osteosarcoma histology images. Experimental results show that the proposed hybrid neural network is easy to train and it outperforms such popular CNN models as PCANet, ResNet and DenseNet in terms of classification accuracy, sensitivity and specificity.

INDEX TERMS Medical image classification, hybrid neural network, PCANet, DenseNet.

I. INTRODUCTION

During the process of disease diagnosis, doctors need to exam such numerous medical images as X-ray images, magnetic resonance (MR) images and ultrasound images. Medical image classification is a highly non-trivial task. Computer aided image classification can avoid subjectivity and save labor, and thus it plays an important role in clinical diagnosis [1], [2]. Image feature extraction is a crucial step for image classification. In the traditional image classification technologies, such hand-designed feature descriptors as the local binary pattern (LBP) [3], the scale-invariant feature transform (SIFT) [4], the histogram of oriented gradients (HOG) [5] and Zernike moment magnitudes [6] were generally used to extract the image features. Some variants of the LBP were applied to retinal disease screening [7] and mammographic image classification [8]. The LBP combined with minimum redundancy maximum relevance feature selection was employed to recognize Parkinson's disease [9] and classify tumors from mammograms [10]. The SIFT features were used to realize the classification of neuroblastoma histological images [11]. The HOG features were utilized for risk estimation of breast cancer development [12]. Furthermore, the HOG features were combined with the SIFT features to realize the classification of brain disease in MR images [13]. Zernike moments were also applied to detecting Alzheimer's disease [14] and computer-aided diagnosis (CAD) of mammograms [15]. However, the above hand-designed feature descriptors can only describe the low-level image features, which are difficult to represent the

The associate editor coordinating the review of this manuscript and approving it for publication was Inês Domingues.

complicated features of medical images. Meanwhile, these descriptors lack in generality in that they are designed for certain kinds of images and they may not produce the good feature extraction results for other images.

Compared with the hand-designed feature descriptors, the deep learning (DL) models can automatically learn image features from the massive data. The DL models can fit the high dimensional non-linear functions by means of the massive connected neurons, and serve the classification purpose by adjusting the weights. A variety of supervised DL models have been proposed to realize medical image classification with better performance than the traditional methods [1], [16]. Artificial neural networks (ANNs) were utilized for brain tumor diagnosis [17], confocal corneal image classification [18], breast cancer diagnosis [19], and auxiliary diagnosis for lung cancer [20]. The ANNs with many layers have better representation ability than those with fewer layers, but they are easily affected by the over-fitting problems. Besides, the training process of the ANNs is time-consuming. To address these problems, Hinton *et al.* [21] proposed the deep belief nets (DBNs) that can be pre-trained layer by layer in an unsupervised way and then fine-tuned with the error back propagation (BP) algorithm [22]. The DBN was employed to recognize breast cancer [23] and assess the pain levels of the patients during surgery [24]. The feature vectors obtained by the DBN were also used as the input of such classifiers as K-nearest neighbors (KNN) and support vector machine (SVM) to recognize retina-based diseases [25]. Although the unsupervised training strategy of the DBN can alleviate the difficulty in training the deep neural networks and help to reduce the reliance on the labeled data to some extent, the time-consuming problem still remains and the fine-tuning process of the DBN still needs too much image data with the corresponding labels.

The convolutional neural network (CNN) model proposed by Lecun *et al.* [26] was specially designed for image processing by introducing such distinctive characteristics as the local receptive fields and the shared weights, which can help to significantly reduce the number of parameters and computational complexity. The well-known CNN models include AlexNet [27], VGG [28], ResNet [29], DenseNet [30] and so on. These CNN networks and their variations were widely applied to polyp detection in colonoscopy videos [31], interstitial lung nodule detection [32], HEp-2 cell image classification [33], breast masses classification [34], disease biomarkers detection in cerebral small vessels [35] and skin cancer classification [36]. Although the CNN generally performs better than the traditional feature extraction methods, it requires numerous image data for its long supervised training process, and it involves the difficulty in determining the number of convolution kernels and the vulnerability to training parameters and manners. However, in the field of medical diagnosis, the number of available image samples is generally insufficient. There are two common solutions to address this issue, i.e., data augmentation and transfer learning. The former can enlarge the number of data by rotation or translation and so forth, but many of generated images are still quite alike, thereby affecting the effectiveness of data augmentation method. Transfer learning works well for the networks with massive adjustable weights such as AlexNet [31] and ResNet-50 [33], but the parameter setting and fine-tuning still require many trials.

To reduce the difficulty in training the CNNs for image classification tasks, a novel principal component analysis network (PCANet) was proposed by Tsung-Han Chan et al. [37]. The PCANet maintains the network framework of the CNN, but it involves an unsupervised learning process in which the convolution kernels are simply learned from the image patches by the cascaded principal component analysis (PCA) algorithm instead of the iterative process of adjusting the weights. The PCANet was combined with the random binary hashing and the low-rank bilinear classifier to realize histopathological image classification [38]. Lee *et al.* [39] applied the PCANet combined with SVM to classify electrocardiogram signal for personal identification. Although the PCANet can easily learn the convolution kernels with the low computational complexity, its performance is greatly affected by the kernels. For the PCANet, the number of kernels should be no greater than the number of weights in one kernel. Therefore, the use of small sized kernels will lead to the inadequate number of kernels while the use of large sized kernels tends to produce the blurry image features. Moreover, the PCANet generally needs to be combined with the traditional machine learning classifiers such as SVM and KNN to realize the image classification task. However, these traditional classifiers are inferior to CNNs in terms of classification performance.

The hybrid neural network which combines an unsupervised CNN with a supervised one has become a research hotspot in the field of image classification. Oyallon [40] designed a hybrid network by combining the scattering network with the wide residual network. When compared with the regular CNNs, this network can produce the competitive image classification results using by far fewer parameters. In this paper, we have designed a light-weighted hybrid neural network for medical image classification by combining a modified PCANet and a simplified DenseNet model. The hybrid neural network utilizes a modified two-stage PCANet to extract the low-level features and applies a simplified DenseNet model to extracting the high-level features for accurate medical image classification. Here, the modified PCANet is utilized because its good feature extraction ability paves the way for the following network, and its unsupervised and interpretable learning strategy can alleviate the requirement for labeled data and render the hybrid neural network more reliable than the regular CNNs in medical image classification in the case of a small number of training samples. At each stage of the modified PCANet, the input images are split into image blocks which will be clustered using the K-means method, and the convolution kernels are learned from each clustering of the input data. In this way, the modified PCANet can produce the adequate number of

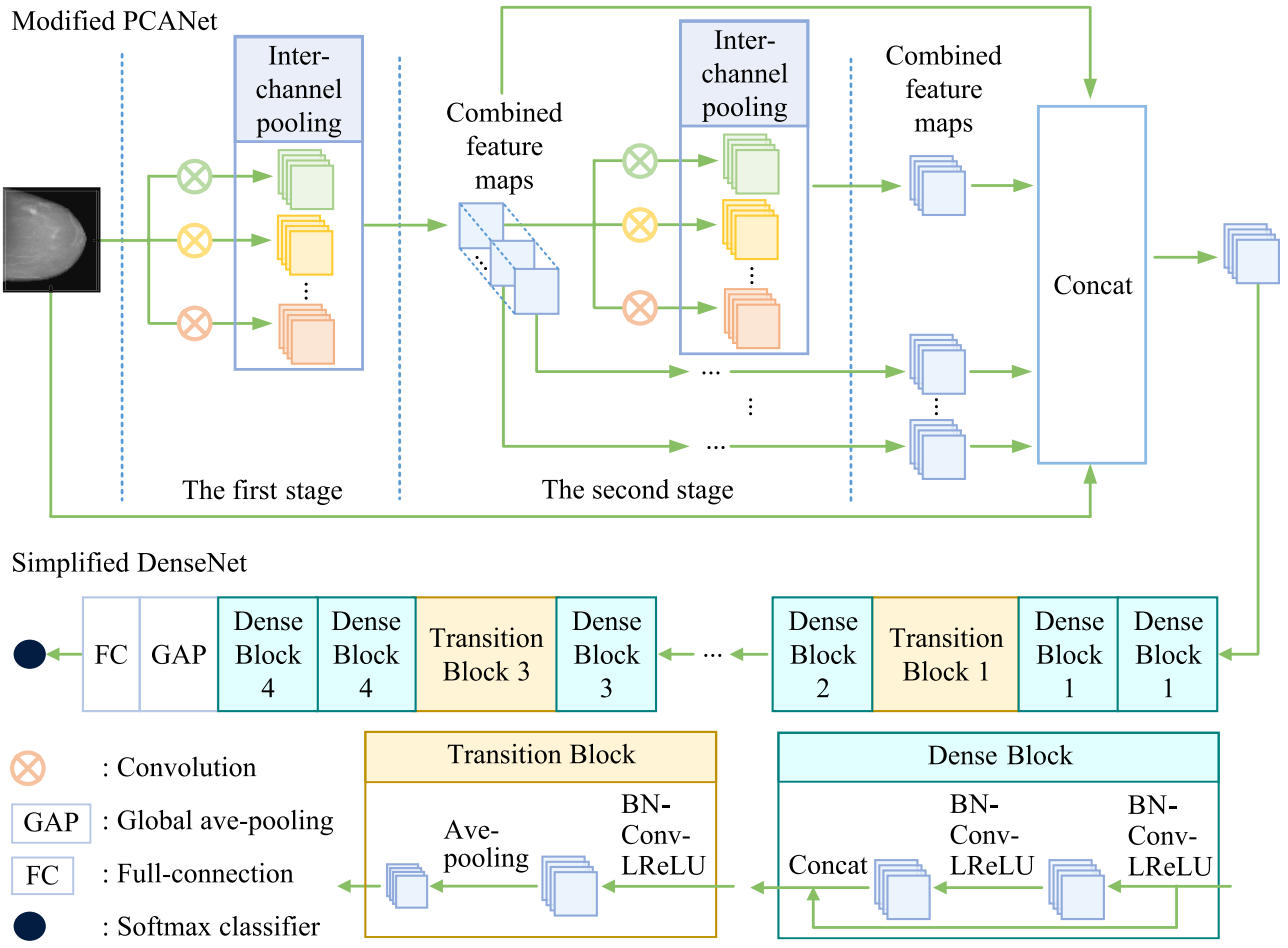

FIGURE 1. Diagram of the hybrid neural network for testing. the symbols "⊗" with different colors mean the convolution with different groups of kernels.

small sized kernels for the extraction of diverse and elaborate features. These feature maps generated at the two stages are combined by a non-linear mapping and then input into the following layer and the following simplified DenseNet. The simplified DenseNet frequently uses shortcut connections among convolutional layers in order to make full use of middle feature maps to discover the meaningful features with few adjustable weights that will be learned by the BP algorithm. Some experiments have been done on mammography and osteosarcoma histology images to determine the impact of the key parameters in the proposed method and test its effectiveness. Experimental results show that the proposed method is easy to train and it can provide better medical image classification results in terms of accuracy, specificity and sensitivity than the compared CNN based methods.

The main contributions of this work are summarized as follows:

- A light-weighted hybrid neural network has been presented for medical image classification by combining an unsupervised PCANet with a supervised DenseNet.
- An effective learning strategy has been presented for the modified PCANet in which several groups of kernels are learned in parallel based on the different image contents instead of the whole images to ensure that the learned kernels are adequate and discriminative.
- A simplified structure of the DenseNet model has been designed, and the sequence of layers in the basic convolution units has been adjusted to ensure the effectiveness of this shallow network.

The remainder of this paper is organized as follows. In Section II, the structure of the hybrid neural network is introduced and its training is explained in detail. The classification experiments on three kinds of medical images are performed and the results of all evaluated methods are compared and analyzed in Section III. Finally, the conclusions are given in Section IV.

II. METHOD

A. STRUCTURE OF THE HYBRID NEURAL NETWORK

The diagram of the proposed network is shown in Fig. 1. The architecture of the entire network can be seen as two components. One component in Fig. 1 is a modified two-stage PCANet. At each stage of the modified PCANet, several groups of kernels will be learned, and each input image is

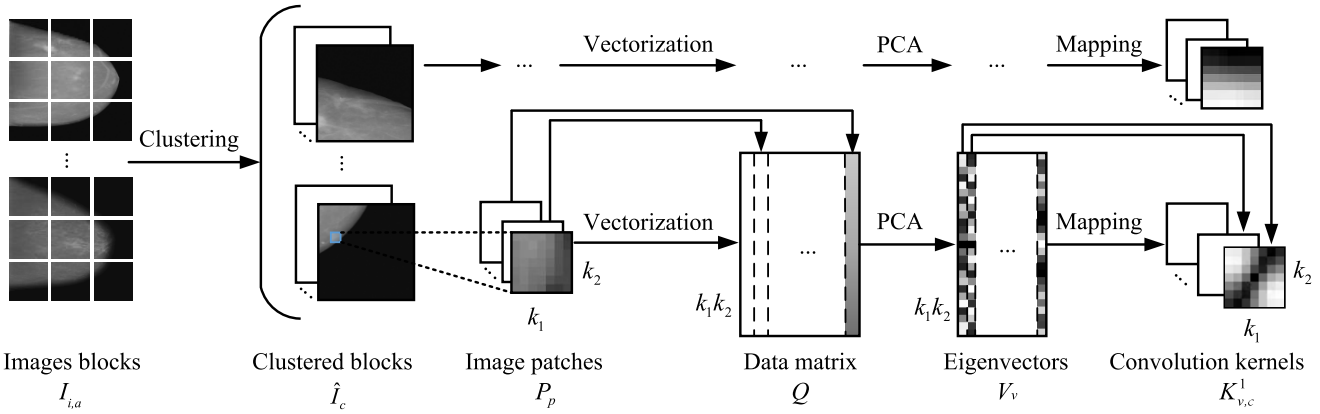

FIGURE 2. Diagram of the kernel learning process at the first stage of the modified PCAnet.

convoluted with the kernels, respectively. The ''inter-channel pooling'' in Fig. 1 is specifically designed for the modified PCANet to combine the obtained feature maps using a non-linear function in order to control the number of produced feature maps and integrate the obtained image information. Taking the output combined feature maps of the modified PCANet as the inputs, the following simplified DenseNet extracts the higher-level features for image classification. The structure of the simplified DenseNet, as another component in Fig. 1, consists of cascaded alternating dense blocks and transition blocks. Compared with the original DenseNet, the simplified DenseNet adopts a different implementation of dense blocks and uses much fewer layers and weights. In the hybrid neural network, the PCANet and the DenseNet will be trained in order using two different strategies.

B. THE MODIFIED PCANET

1) THE FIRST STAGE

At the first stage, the kernels are key to the feature extraction by the modified PCANet. The process of kernel learning is shown in Fig. 2. Let the image size be $m_1 \times m_2$, the N training images including all classes be $I = \{I_i,\ i = 1, 2, \ldots, N\}$, and the size of each convolution kernel be arbitrarily determined as $k_1 \times k_2$. Each input image I_i is split into α^2 neighboring blocks $\{I_{i,a},\ i = 1, 2, \ldots, N,\ a = 1, 2, \ldots, \alpha^2\}$ of size $s_1 \times s_2$ without overlapping, where $I_{i,a} \in \mathbb{R}^{s_1 \times s_2}$, $s_1 = \lfloor m_1/\alpha \rfloor$, and $s_2 = \lfloor m_2/\alpha \rfloor$ with $\lfloor \cdot \rfloor$ denoting the floor function. Using K-means clustering, all the blocks $I_{i,a}$ are partitioned into several clusters denoted as $\hat{I} = \{\hat{I}_c,\ c = 1, 2, \ldots, C\}$ where C means the number of clusters which need to be set manually.

For each cluster $\hat{I}_c$, the steps of learning kernels are the same. Here, we will take the c-th cluster $\hat{I}_c$ as an example. Assuming that there are n_b image blocks belonging to the c-th cluster, image patches $P = P_p \in \mathbb{R}^{k_1 \times k_2}$ $(p = 1, 2, \ldots n_p)$ of size $k_1 \times k_2$ can be collected from all these blocks pixel by pixel, where n_p is the number of patches and computed as $n_p = (s_1 - k_1 + 1) \times (s_2 - k_2 + 1) \times n_b$. Each P_p is vectorized into $\overrightarrow{P}_p$. Then all patch vectors $\{\overrightarrow{P}_p\}$ minus their mean vector $\overline{P}$ are assembled column by column into a matrix of size $k_1k_2 \times n_p$ denoted as $Q = \left[\overrightarrow{P}_1 - \overline{P}, \overrightarrow{P}_2 - \overline{P}, \ldots, \overrightarrow{P}_{n_p} - \overline{P}\right]$. Then the PCA algorithm is implemented on Q, which is to solve the minimization problem:

$$\min_{V \in \mathbb{R}^{k_1k_2 \times L^1}} \| Q - VV^TQ \|_F^2, \quad \text{s.t. } V^TV = I_{L^1}, \tag{1}$$

where L^1 denotes the number of kernels for each cluster at the first stage, I_{L^1} means an identity matrix of size $L^1 \times L^1$, and $\| \cdot \|_F$ means the Frobenius norm. The solution $V = \left[V_1, V_2, \ldots, V_v, \ldots, V_{L^1}\right]$ of (1) consists of L^1 principal eigenvectors corresponding to the first L^1 largest eigenvalues of the matrix QQ^T.

All the eigenvectors in V are mapped into the matrixes of size $k_1 \times k_2$ to produce the convolution kernel set $\{K^1_{v,c} \in \mathbb{R}^{k_1 \times k_2} |\ v = 1, 2, \ldots, L^1\}$ of the c-th cluster. The mapping function is expressed as:

$$\begin{aligned} &K^1_{v,c}(x, y) = V_v((y-1) \times k_1 + x), \\ &\text{s.t. } 1 \le x \le k_1, \quad 1 \le y \le k_2,\ x, y \in \mathbb{Z}, \end{aligned} \tag{2}$$

where (x, y) are the coordinates of the value $K^1_{v,c}(x, y)$ in the kernel.

It should be noted that to construct the matrix Q, our modified PCANet removes the mean of $\{\overrightarrow{P}_p\}$ instead of the mean of each $\overrightarrow{P}_p$ in the original PCANet. The reason is explained in this way. According to the properties of eigenvalues and eigenvectors, there is:

$$QQ^TV_v = \lambda_v V_v, \tag{3}$$

$$\begin{bmatrix} \sum_p q_1^p q_1^p & \cdots & \sum_p q_1^p q_{k_1k_2}^p \\ \vdots & \ddots & \vdots \\ \sum_p q_{k_1k_2}^p q_1^p & \cdots & \sum_p q_{k_1k_2}^p q_{k_1k_2}^p \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_{k_1k_2} \end{bmatrix} = \lambda_v \begin{bmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_{k_1k_2} \end{bmatrix}, \tag{4}$$

where q_i^p is the i-th element in the p-th column of Q, and $\mathbf{v}_i \in V_v$ $(i = 1, 2, \ldots, k_1k_2$ and $v = 1, 2, \ldots, L^1)$. According

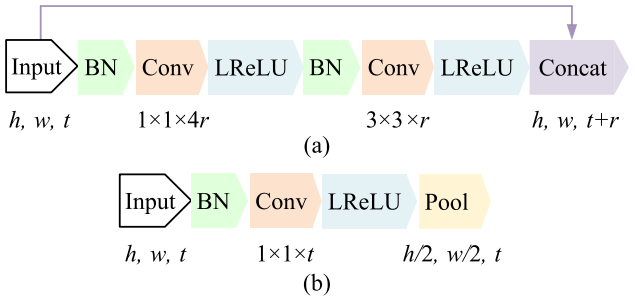

FIGURE 3. Details of the dense block (a) and the transition block (b). *h*, *w*, *t* are the height, width and number of the input feature maps, respectively.

to (4), we have

$$\sum_j \sum_i \left(\mathbf{v}_i \sum_p q_j^p q_i^p \right) = \lambda_v \parallel V_v \parallel_1 \tag{5}$$

$$\sum_i \left(\mathbf{v}_i \sum_p \left(q_i^p \sum_j q_j^p \right) \right) = \lambda_v \parallel V_v \parallel_1 \tag{6}$$

In the original PCANet, the mean value $\overline{P}_p$ of each $\overrightarrow{P}_p$ is removed from $\overrightarrow{P}_p$ itself to construct the matrix Q. Therefore, $\sum_i q_i^p = \parallel \overrightarrow{P}_p \parallel_1 - k_1 k_2 \overline{P}_p = 0$. It follows that $\parallel V_v \parallel_1 = 0$ or $\lambda_v = 0$. Because λ_v ($v \leq L^1 - 1$) values are generally not zeros, $\parallel V_v \parallel_1 = 0$. The convolution with these corresponding kernels $K_{v,c}^1$ represents the weighted subtraction among the pixels in image patches. Therefore, most of the kernels in the original PCANet can only be used to extract the gradient features. Our proposed network can avoid this problem by removing the mean of the whole set $\{\overrightarrow{P}_p\}$ to ensure that $\sum_i q_i^p \neq 0$ for most of image patches.

After obtaining the convolution kernels for all the clusters, there are C groups of L^1 feature maps $F_{i,v,c}^1$ for each input image I_i. The feature maps can be obtained by convoluting I_i with the kernels $K_{v,c}^1$ of all the clusters:

$$\begin{aligned} &F_{i,v,c}^1 = I_i \otimes K_{v,c}^1, \\ &\text{s.t. } c = 1, 2, \ldots, C, \quad v = 1, 2, \ldots, L^1, \end{aligned} \tag{7}$$

where $\otimes$ denotes the 2D convolution. The combined feature maps $F_{i,v}^1$ of the first convolutional stage in the modified PCANet can be obtained by performing the ''inter-channel pooling'' on the absolute values of $F_{i,v,c}^1$:

$$\begin{aligned} F_{i,v}^1 = &\max_{c \in [1,2,\ldots,C]} \left(abs\left(F_{i,v,c}^1 \right) \right), \\ &\text{s.t. } i = 1, 2, \ldots, N, \quad v = 1, 2, \ldots, L^1 \end{aligned} \tag{8}$$

The function of calculating the absolute values and the max-pooling operation among groups is to reduce computational complexity and enhance the ability of non-linear mapping for the modified PCANet.

2) THE SECOND STAGE

For the second stage, all the combined feature maps $F_{i,v}^1$ are used as the inputs. The steps of kernel learning and forward propagation are similar to those at the first stage except that such parameters as the size $k_3 \times k_4$ of image patches and the number L^2 of kernels at the second stage can be different from $k_1 \times k_2$ and L^1. Besides, since the clustering is based on the content of input images, there is no need to perform the clustering algorithm on the generated block sets $\{F_{i,a,v}^1\}$ by splitting $\{F_{i,v}^1\}$. Accordingly, we can use a simple method as the clustering scheme for the second stage, which can be expressed as:

$$\hat{I}_c' = \left\{ F_{i,a,v}^1 | v = 1, 2, \ldots, L^1, I_{i,a} \in \hat{I}_c \right\}, \tag{9}$$

where $\hat{I}_c'$ means the *c-th* cluster of blocks at the second stage. For the *c-th* cluster, a group of kernels $K_{v',c}^2$ can be obtained by collecting the image patches, solving (1), and mapping the obtained eigenvectors according to (2), where $v' = 1, 2, \ldots, L^2$ represents the index of convolution kernels at the second stage. The feature maps $F_{i,v,v',c}^2$ at the second stage result from the convolution of $F_{i,v}^1$ with $K_{v',c}^2$. After applying the ''inter-channel pooling'' to $F_{i,v,v',c}^2$, there are $L^1 \times L^2$ output combined feature maps $F_{i,v,v'}^2$ for the image I_i at the second stage. Since the feature maps at different stages provide different information, the input image I_i and all the combined feature maps produced at the two stages will be concatenated as the final output O_i of the modified PCANet. The output O_i can be expressed as:

$$\begin{aligned} &O_i = Concat(I_i, F_{i,v}^1, F_{i,v,v'}^2), \\ &\text{s.t. } v = 1, 2, \ldots, L^1, \quad v' = 1, 2, \ldots, L^2 \end{aligned} \tag{10}$$

C. THE SIMPLIFIED DENSENET

The simplified DenseNet is constructed with the alternating dense blocks and the transition blocks followed by a fully connected layer and a softmax classifier. The dense block is made up of two cascaded convolutional units which consist of a batch normalization (BN) layer, a convolutional layer and a leaky rectified linear unit (LReLU) layer in order (denoted as ''BN-Conv-LReLU''). As shown in Fig. 3, the first ''BN-Conv-LReLU'' in the dense block uses the kernels of size 1×1 to produce $4r$ output feature maps while the second ''BN-Conv-LReLU'' generate r feature maps using 3×3 kernels, where r is a pre-defined constant and it is fixed to be 32 in this paper. The dense block concatenates r output feature maps with the input maps, thereby leading to the increasing number of maps. As for the transition block, it consists of a convolutional unit followed by an average pooling layer with a pool size of 2×2. The function of the transition block is to find an effective composition among feature maps produced from different convolutional layers and reduce the computational complexity. In this paper, the number of output feature maps in the transition block is set to be the same as that of its inputs.

TABLE 1. Architecture of the simplified DenseNet.

Module	Detail		Output
Convolution	$1 \times 1 \times 64$ conv		$128 \times 128 \times 64$
Max-pooling	2×2 pool		$64 \times 64 \times 64$
Dense block 1	$1 \times 1 \times 128$ conv $3 \times 3 \times 32$ conv	$\times 2$	$64 \times 64 \times 128$
Transition block 1	$1 \times 1 \times 128$ conv 2×2 pool		$32 \times 32 \times 128$
Dense block 2	$1 \times 1 \times 128$ conv $3 \times 3 \times 32$ conv	$\times 3$	$32 \times 32 \times 224$
Transition block 2	$1 \times 1 \times 224$ conv 2×2 pool		$16 \times 16 \times 224$
Dense block 3	$1 \times 1 \times 128$ conv $3 \times 3 \times 32$ conv	$\times 4$	$16 \times 16 \times 352$
Transition block 3	$1 \times 1 \times 352$ conv 2×2 pool		$8 \times 8 \times 352$
Dense block 4	$1 \times 1 \times 128$ conv $3 \times 3 \times 32$ conv	$\times 2$	$8 \times 8 \times 416$
Global average pooling	–		$1 \times 1 \times 416$
Full-connection	$416 \times \theta$ full-connection		θ
Softmax	softmax classifier		θ

Supposing the size of input images is 128×128, the details of structure and output for the simplified DenseNet are listed in Table 1. In Table 1, "conv" means a convolutional unit; "$1 \times 1 \times 64$ conv" means that the convolutional layer in the unit uses the kernels of size 1×1 to produce 64 output feature maps; "2×2 pool" denotes an average pooling layer with a pool size of 2×2; θ denotes the number of the final output classes; "$[\cdot] \times 2$" means the structure "$[\cdot]$" is repeatedly cascaded for 2 times. As shown in Table 1, there are 23 convolutional layer and one fully connected layer in the simplified DenseNet, which is much shallower than the original DenseNet with 121 convolutional layers.

The above design scheme can reduce the number of adjustable weights and produce the effective feature maps due to the following strategies.

- The small sized kernels such as 1×1 or 3×3 are adopted in the convolutional layers to adjust the number of feature maps and learn to produce the meaningful feature maps.
- The shortcut connections are repeatedly used since it can make use of middle feature maps and facilitate the training of the DenseNet by means of effective error feedback.
- The global average pooling (GAP) technique is utilized between the last transition block and the fully-connected layer to avoid flattening the output of the last convolutional layer so that a large amount of weights will not be introduced. Moreover, the GAP technique can be used to analyze the final feature maps of the network [41], which makes the proposed hybrid network more interpretable.
- The fully-connected layer is only applied once as the last layer before the softmax classifier to produce the specified number of outputs.

Note that the adopted convolutional unit in the simplified DenseNet is different from that in the original DenseNet in that the latter uses the structure of "BN-ReLU-Conv". Fig. 4 shows the structure of two kinds of convolutional units and some sampled feature maps. According to the structure of "BN-ReLU-Conv", the ReLU activation function is applied right after the BN layer which produces the normalized feature maps. Therefore, the ReLU maps the values of about half of elements in the normalized feature maps to be zeros, thereby leading to the loss of much information in the feature maps as shown in Fig. 4(a).

To design an effective shallow network, the shallow layers should transfer more useful information to the deeper layers. Accordingly, we have changed the order of layers in the convolutional unit to address the above problem resulting from the structure "BN-ReLU-Conv". Moreover, the LReLU is used as activation function to replace the ReLU to retain some details even if some values in feature maps are less than 0. Moreover, it can help to speed up the network convergence. The comparison of feature maps in Fig. 4 shows that there are more details in the feature maps of the activation layer and the convolutional layer in "BN-Conv-LReLU" than those in "BN-ReLU-Conv", which indicates that the structure "BN-Conv-LReLU" is more suitable for the shallow network.

D. THE TRAINING PHASE OF THE HYBRID NEURAL NETWORK

In the hybrid neural network, the modified PCANet needs to learn the kernels firstly and the training process of the modified PCANet is irrelevant to the simplified DenseNet. As described in Section II.B, the modified PCANet learns the kernels stage by stage. At the first stage, the input images are split into blocks, and then clustered into several groups. For each group, the image patches are collected from the blocks, and then the PCA algorithm (1) is used to produce the eigenvectors which represent some components of the image data. Finally, all these obtained eigenvectors are mapped into the convolution kernels according to the mapping function (2). To learn the kernels at the second stage, the input maps at this stage need to be produced firstly by convoluting the input images with all the learned kernels at the first stage according to (7), and combining the generated features maps based on (8). Then, the input maps are split and clustered according to (9). The subsequent procedure of the kernel learning is similar to that at the first stage. Please note that such operations as splitting of inputs and K-means based clustering will not be implemented at the testing phase of the modified PCAnet.

For the following simplified DenseNet, it takes the outputs of the trained modified PCANet produced by (10) and the corresponding labels as the inputs, and it is trained using the

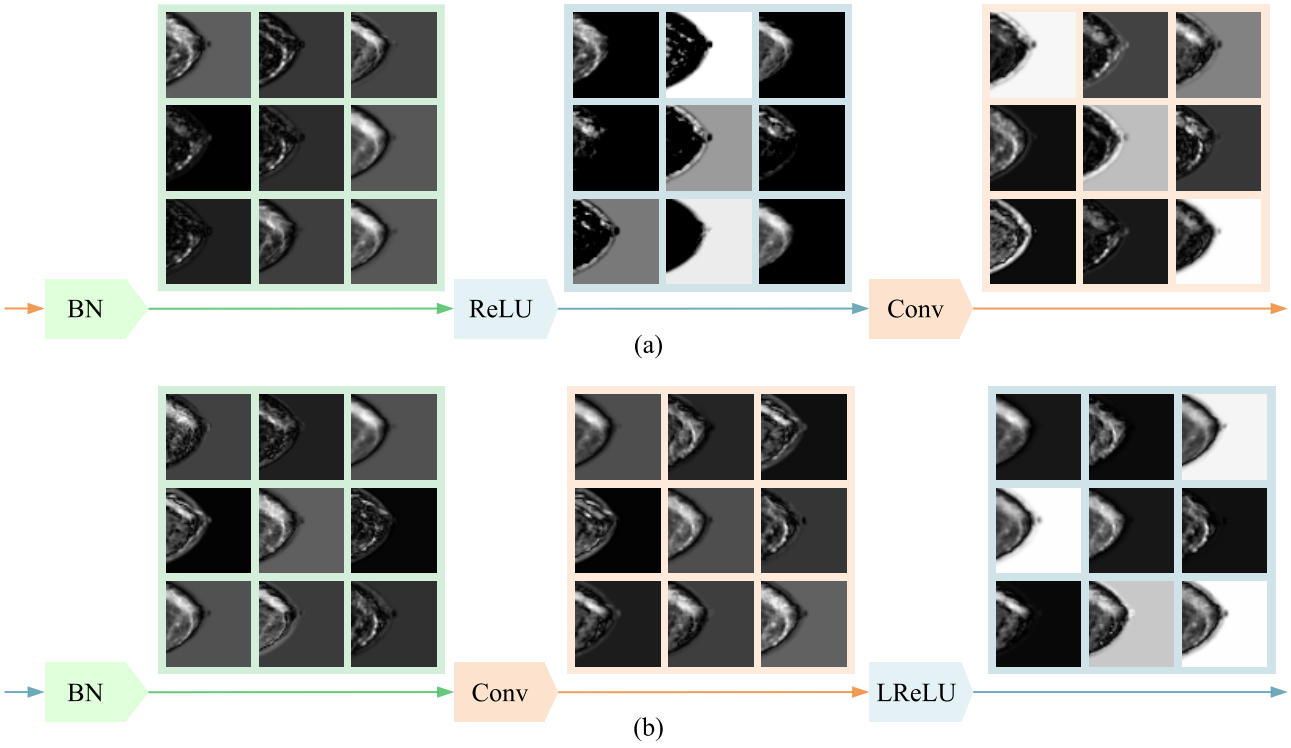

FIGURE 4. Comparison of two kinds of convolutional units. (a) "BN-ReLU-Conv" unit. (b) "BN-Conv-LReLU" unit.

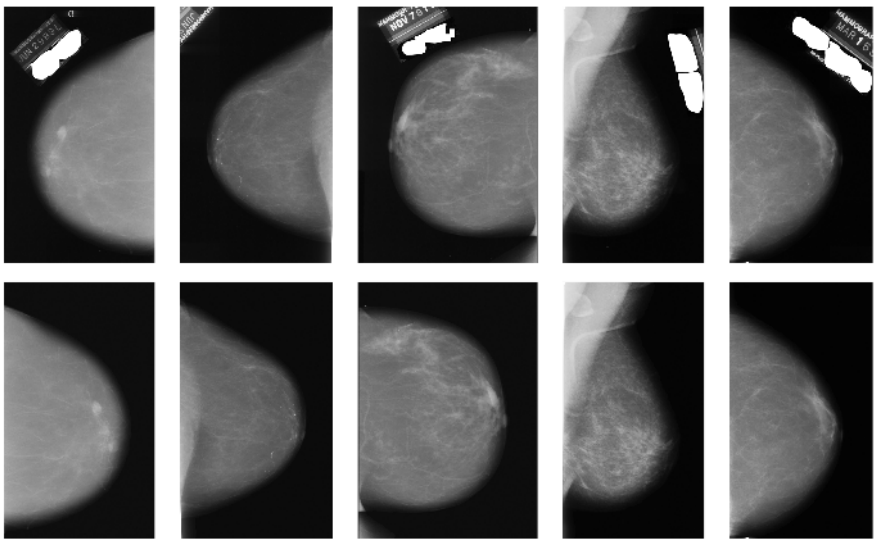

FIGURE 5. Examples of some original images (the first row) and pre-processed mammography images before sub-sampling (the second row) in the DDSM.

regular BP algorithm. Besides, the softmax layer is used as the classifier and the cross-entropy function is used as the cost function:

$$Loss = -\frac{1}{N}\sum_{i}^{N}\sum_{j}^{\theta} y_{i,j} log(x_{i,j}), \tag{11}$$

where $x_{i,j}$ is produced by the softmax function and it is also the predicted probability that the *i-th* sample belongs to the *j-th* class; $y_{i,j}$ is the corresponding label indicating whether the *i-th* sample belongs to the *j-th* class. The Adam optimizer [42] is adopted to minimize the cost function.

III. RESULT AND DISCUSSION

In this section, we will firstly discuss the determination of the key parameters in our method and then make the comparisons between the classification performance of our method and that of other evaluated methods.

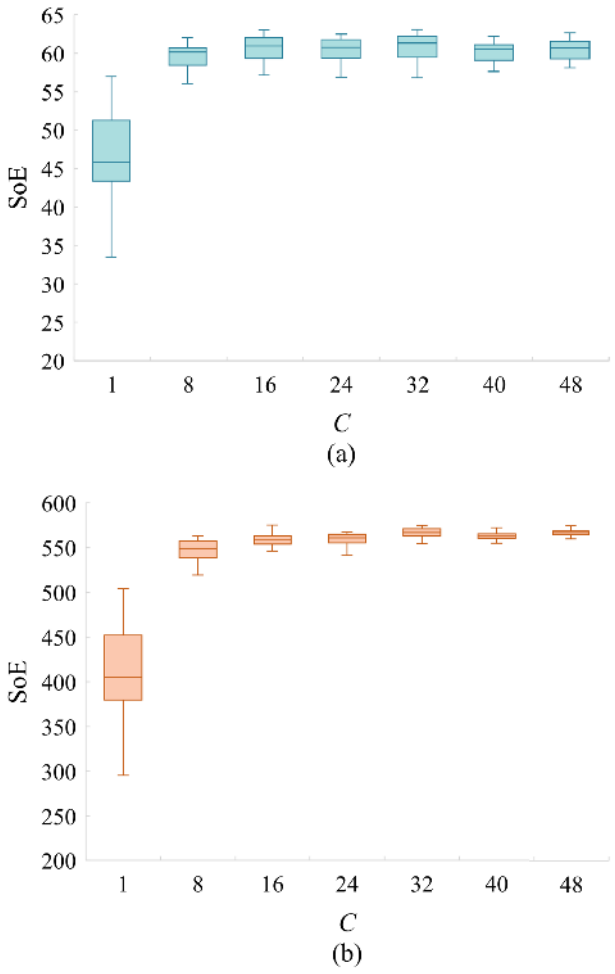

FIGURE 6. Sum of entropy of combined feature maps for different C at the first stage (a) and the second stage (b).

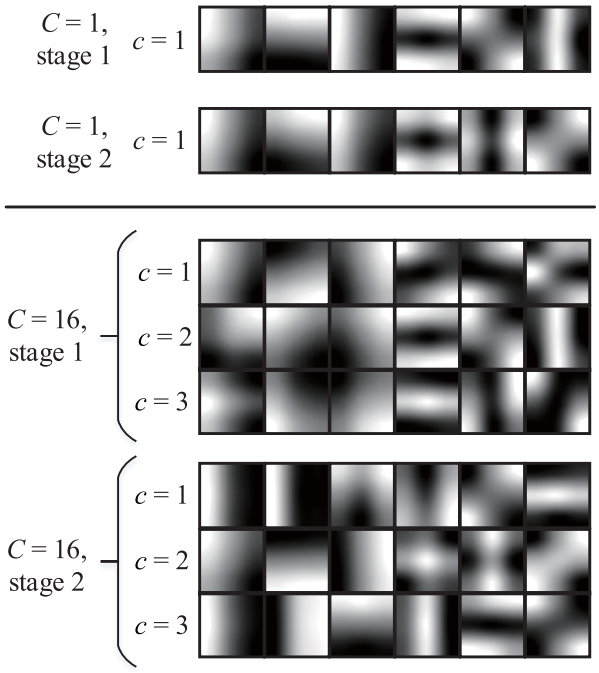

FIGURE 7. Some kernels in the modified PCANet with $C = 1$ and $C = 16$.

A. DETERMINATION OF THE KEY PARAMETERS

The modified PCANet involves some key parameters including the block number α^2 for each image, the kernel sizes $k_1 \times k_2$ and $k_3 \times k_4$, the numbers L^1, L^2 of kernels for each cluster at the two stages and the number C of clusters. α^2 will be set according to the image size and the image content, and it will fixed to be 9 empirically for the following experiments. Small sized kernels are preferred in CNN models for extracting the elaborate features using the fewer weights than the large sized ones. Therefore, we set $k_1 = k_2 = 3$, $k_3 = k_4 = 3$. Inspired by the Gabor filters [43] which have eight orientations, we have fixed $L^1 = L^2 = 8$.

As regards the number C of clusters, it directly decides the number and diversity of convolution kernels. If C is too large, the differences among the clusters will be very small, thereby leading to the redundant kernels. On the other hand, a too small C cannot meet the need for the number of kernels. To explore the influence of the number of clusters, the images from the Digital Database for Screening Mammography (DDSM) [44], [45] which can be downloaded from http://marathon.csee.usf.edu/Mammography/Database.html are used to construct the datasets. The training dataset consists of 4641 images including 1980 normal images and 2661 abnormal images. Meanwhile, 1160 images are chosen from the DDSM as the test dataset which contains 488 normal images and 672 abnormal images. We have performed some pre-processing on the DDSM, which includes filtering the background noise, removing the tags of patients' information in the images by filling the tag area with the mean of pixel intensities sampled from the background, cropping the white borders, and reflecting some images to ensure that all the breasts are on the same side. Each processed image is sub-sampled into 128×128 in order to reduce the calculation complexity and maintain the uniform size for all the images. Some pre-processed images are shown in Fig. 5.

The parameter C varies in the set $\{1, 8, 16, 32, 40, 48\}$. To determine the suitable C value, we will compare the sum of entropy (SoE) of all the combined feature maps at each stage respectively. Ideally, the proper C should provide a big SoE so that the modified PCANet can transfer more information of input images to the simplified DenseNet. The values in the feature maps are normalized to be integers between 0 and 255 to produce the corresponding images, then the SoE is calculated as:

$$\mathrm{SoE} = -\sum_{i=1}^{\dot{N}} \sum_{j=0}^{255} x_j^i log(x_j^i), \tag{12}$$

where x_j^i means the probability of pixels whose values are equal to j in the i-th produced image, and $\dot{N}$ is the number of produced images. The results are shown in Fig. 6. The observation from Fig. 6 shows that the SoE firstly

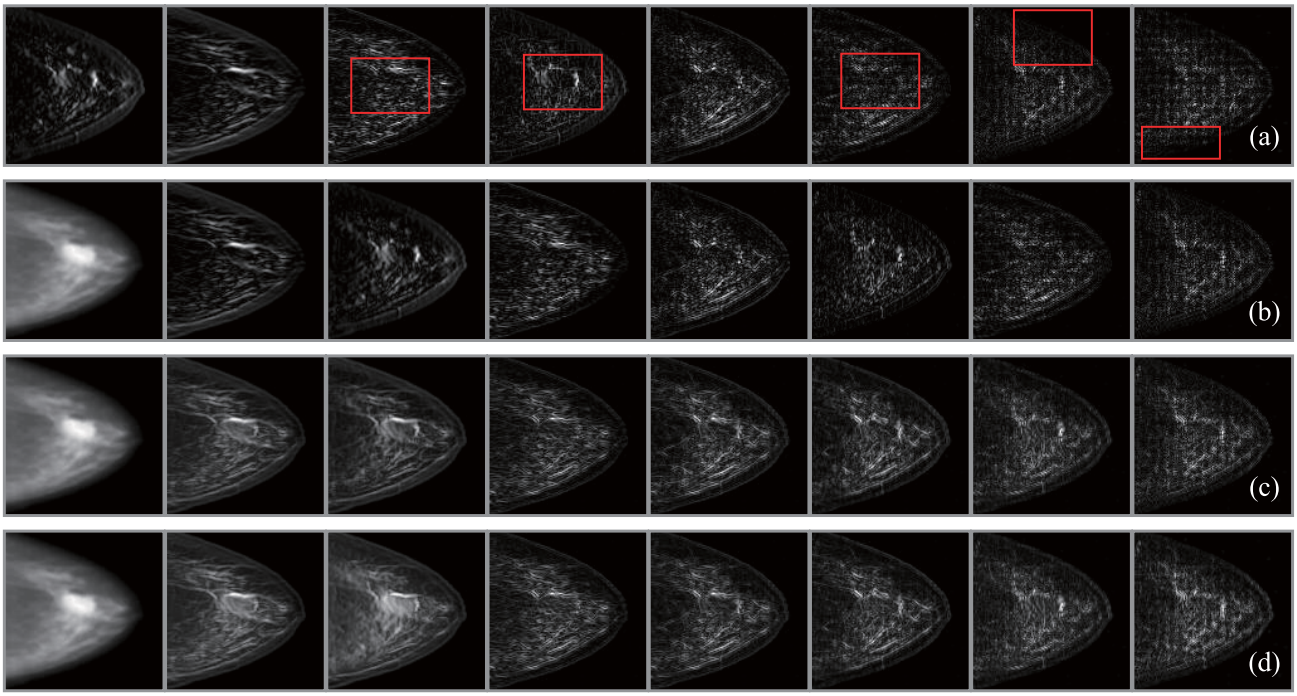

FIGURE 8. Examples of feature maps at the first stages in the two kinds of PCANet. (a) the feature maps produced by the original PCANet. (b) to (d) the combined feature maps produced by the modified PCANet with $C = 1$, 16, and 32, respectively.

increases with the increasing C, and is nearly stable when $C \geq 16$. Some examples of convolution kernels in the modified PCANet are shown in Fig. 7. The kernels above the line are learned with $C = 1$. Note that the first four filters for the two stages are quite similar, which indicates that the main features of the inputs at the first stage are similar to those at the second stage. As regards $C = 16$, much more kinds of kernels can be obtained to facilitate producing richer image features. The above comparison indeed indicates the advantage of using $C = 16$ over using $C = 1$ in the proposed method.

The combined feature maps at the first stage in the proposed modified PCANet are shown in Fig. 8, where the absolute value of each feature map is computed and normalized into [0, 255] for visualization. Clearly, there is a significant difference between the feature maps for the original PCANet and those for our proposed network. Some edges and textures in the feature maps for the original PCANet are unclear and incomplete as marked by the red boxes. By comparison, for the proposed modified PCANet using $C = 16$ and $C = 32$, the feature maps can be seen as the combination of clear and relatively complete edge and texture information and gray-level information of images. The ability of preserving gray-level information results from the advantage of the proposed method in maintaining the mean values of image patches as mentioned in Section II.B. Besides, it can be seen that the differences of feature maps between $C = 16$ and $C = 32$ are not as remarkable as those between $C = 16$ and $C = 1$. Based on the above analysis, we will fix $C = 16$ in the following experiments.

B. COMPARISONS OF CLASSIFICATION PERFORMANCE AMONG THE POPULAR NETWORKS

To demonstrate the performance of our proposed network, comparisons are made among the popular convolution neural networks such as VGG, ResNet-50, DenseNet-121, the original PCANet, our simplified DenseNet (sDenseNet-24) without the modified PCANet as the input and the proposed hybrid neural network (HybridNet) operating on the DDSM dataset, the osteosarcoma histology images [46], [47] and the mammographic image analysis society (MIAS) dataset [48]. All the networks are realized with Python based on TensorFlow 1.9.0 and Keras 2.2.4 on a Ubuntu 16.04, and they are run on a computer with a Core I7-6950X CPU and 96G RAM. The NVIDIA GTX 1080Ti GPU with CUDA 10.1 is used for acceleration. To evaluate the performance of the proposed method, the total classification accuracy ACC, sensitivity SEN and specificity SPE are utilized as metrics which are defined as:

$$ACC = \frac{TP + TN}{N'} \in [0, 1], \tag{13}$$

$$SEN = \frac{TP}{TP + FN} \in [0, 1], \tag{14}$$

$$SPE = \frac{TN}{TN + FP} \in [0, 1], \tag{15}$$

where N' is the number of test images, TP, TN, FP and FN mean the number of true positive, true negative, false positive and false negative cases, respectively. The sensitivity represents the ability of correctly recognizing the lesion images while specificity denotes the capacity of correctly classifying

TABLE 2. The number of parameters and FLOPs for the six evaluated networks.

Metrics	AlexNet	VGG-13	ResNet-50	DenseNet-121	sDenseNet-24	HybridNet ($C = 16$)
Trainable parameters	25.46M	59.75M	23.51M	6.95M	0.91M	0.96M
FLOPs	127.29M	298.74M	117.85M	35.09M	4.65M	4.86M

the normal images. For these metrics, the higher values mean the better classification effect.

The leaky rate of the leaky ReLU functions is set as 0.3 for both the sDenseNet and the HybridNet. For the original PCANet, we set $k_1 = k_2 = 3, k_3 = k_4 = 3, L^1 = L^2 = 8$ and it has 144 trainable parameters. Meanwhile, the linear SVM is used as the extra classifier for this network since the original PCANet can only extract features and it has not classification ability. The input images are sampled into 128×128 for all the compared networks. The computational cost and the number of parameters of the supervised networks are listed in Table 2. From Table 2, it can be seen that our HybridNet has 0.05M more parameters and 0.23M more FLOPs compared with the sDenseNet-24 due to the introduction of the combined feature maps produced by the modified PCANet. However, the computational cost of the HybridNet is obviously less than that of other popular networks. For example, the HybridNet involves about 1/24 of parameters and FLOPs in the ResNet-50.

In the following experiments, the training images are augmented with the slight translation and/or rotation, the batch size is setting as 24, and the maximum number of epochs is chosen to be 400 for all supervised networks. The learning rate is initialized as 0.001 and decreased to be 30% of the previous value at the chosen epochs, which are adjusted for each supervised network to ensure that the loss function will converge properly. During the training phase, some images chosen from the test dataset are used as the validation dataset to monitor the training process. For each supervised network, the training process will be early-stopped if the change of the loss for the validation dataset is less than 10^{-4} for successive 30 epochs.

1) TEST ON THE DDSM DATASET

The training process of the HybridNet is shown Fig. 9. Obviously, the training process is stable for the HybridNet and it converges quickly. The accuracy curves for all the networks operating on the validation dataset are shown in Fig. 10. It should be noted that we have trained a VGG-13 instead of the popular VGG-16 because the latter does not converge when it is directly trained with the DDSM dataset or fine-tuned based on the weights trained from ImageNet dataset [49]. From Fig. 10, it can be seen that our proposed network converges to a better solution faster and its training process is more stable than other networks. Compared with the ResNet-50 and the DenseNet-121, the shallower sDenseNet-24 has similar classification performance due to the similar structure but with much smaller solution space, which indicates that some specifically designed small-scale networks may

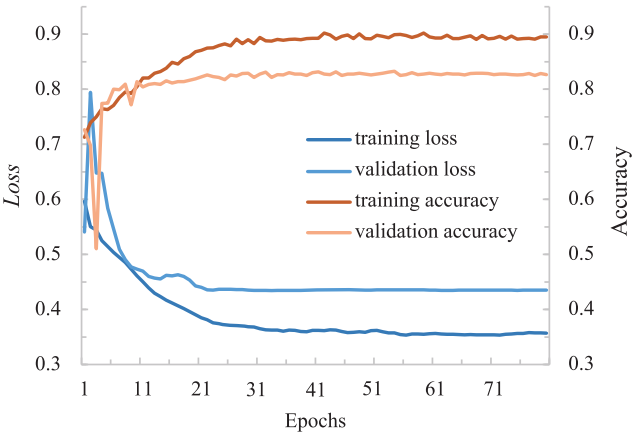

FIGURE 9. Loss and accuracy curves for the HybridNet on the training and validation dataset.

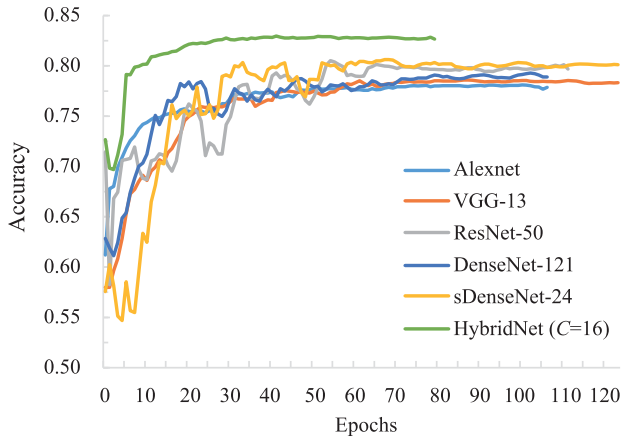

FIGURE 10. Accuracy curves for all evaluated networks on the validation dataset.

be more suitable for medical image classification than the heavy-weighted networks in the case of a small dataset. To verify the effectiveness of the HybridNet further, Fig. 11 shows some class activation maps of the HybridNet produced using the technology in [41]. These maps in Fig. 11 highlight the calcification and mass in the mammography images. Obviously, the hybrid network has indeed learned some meaningful information for the identification of suspicious lesions based on the differences between the normal images and the abnormal ones.

The receiver operating characteristic (ROC) curves of the evaluated networks implemented on the DDSM test dataset are shown in Fig. 12. Clearly, the proposed HybridNet has the best ability to distinguish the normal images and

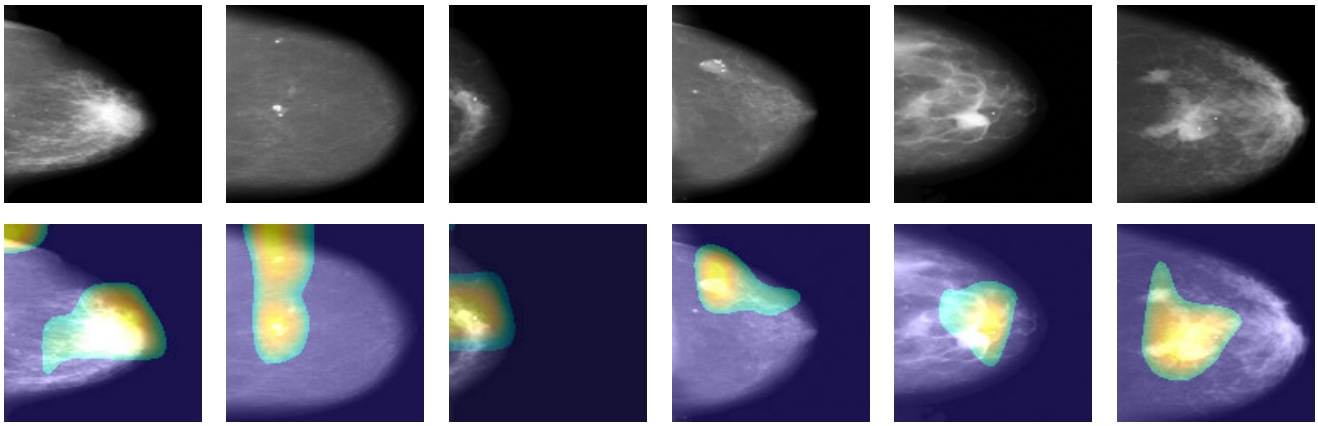

FIGURE 11. Some processed images from the DDSM dataset (the first row) and the corresponding class activation maps of the HybridNet (the second row).

TABLE 3. Metrics for all the networks on the DDSM dataset.

Metrics	AlexNet	VGG-13	ResNet-50	DenseNet-121	PCANet	sDenseNet-24	mPCANet+ ResNet-50	mPCANet+ DenseNet-121	HybridNet ($C = 16$)
SEN	0.832	0.813	0.887	0.848	0.847	0.839	0.844	**0.893**	0.862
SPE	0.723	0.758	0.705	0.760	0.725	0.762	**0.789**	0.723	0.787
ACC	0.786	0.790	0.810	0.811	0.794	0.807	0.821	0.822	**0.830**
AUC	0.844	0.865	0.877	0.875	0.860	0.888	0.864	0.891	**0.897**

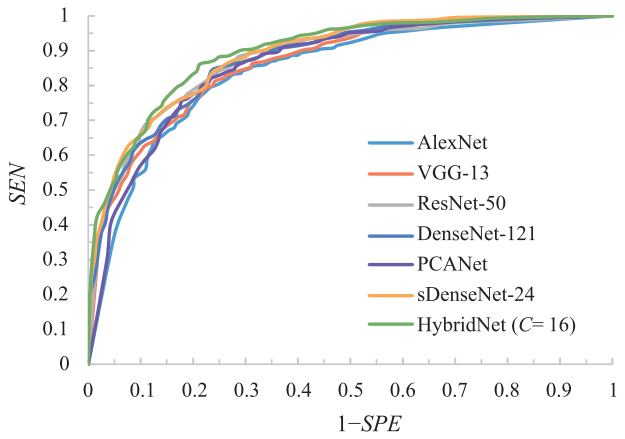

FIGURE 12. ROC curves of all the compared networks on the DDSM dataset.

the abnormal ones compared with other networks. To further demonstrate the advantage of combining the modified PCANet with the CNN model, we have included the other two hybrid networks which combine the modified PCANet with the ResNet-50 (mPCANet+ResNet-50) and DenseNet-121 (mPCANet+DenseNet-121), respectively. The metrics of seven evaluated networks and two extra hybrid networks are list in Table 3. Among these compared networks, the ResNet-like networks can provide relatively better results compared with AlexNet-12 and VGG-13 because the shortcut connections of the former indeed help to extract the useful deep features. Moreover, these hybrid networks obtain better classification accuracy compared with the corresponding CNN models without embedding the modified PCANet. The reason is that the modified PCANet can extract the meaningful features for image classification by using the various kernels to decompose the input images into different components and using ''inter-channel pooling'' to highlight the main components of the images and remove redundant information. Although the mPCANet+DenseNet-121 provides the best sensitivity (0.893) and the mPCANet+ResNet-50 provides the highest specificity (0.789), the former has a low specificity (0.723) and the latter has a relatively low sensitivity (0.844). By comparison, our HybridNet not only provides the second best specificity (0.787) which is very close to the best one (0.789), but also obtains the best accuracy (0.830) and the best result for the area under the curve (AUC) (0.897). The reason why our shallow HybridNet outperforms the other hybrid networks in most cases is that it is much easier to train due to its fewer parameters and it has better ability to avoid the over-fitting problem.

TABLE 4. Composition of the osteosarcoma histology image dataset.

Image dataset	Non-tumor	Necrotic tumor	Viable tumor
Training dataset	486	216	295
Test dataset	200	200	200

2) TEST ON THE OSTEOSARCOMA HISTOLOGY IMAGE DATASET

To further verify the superiority of our method, some experiments have been conducted on the osteosarcoma

FIGURE 13. Sample images in the osteosarcoma histology image dataset.

histology images downloaded from the cancer imaging archive (TCIA) [50]. The osteosarcoma histology images provide valuable information for the assessment of viable and necrotic tumor. The osteosarcoma data are classified as three classes including non-tumor (NO) images, necrotic tumor (NE) images and viable tumor (VT) images. The number of osteosarcoma histology images is much less than that of DDSM and all these images are RGB images. The composition of the constructed dataset is listed in Table 4. Some sample images are shown in Fig. 13. The training dataset is augmented with translation and/or rotation just in a similar way to the DDSM training dataset. The test dataset is constructed to include 200 images for each class by rotation for 90°, −90° and 180° of the original 50 test images in each class.

To evaluate all the methods for multi-class classification problem, we have calculated the average sensitivity $\overline{SEN}$, the average specificity $\overline{SPE}$ and the average $\overline{AUC}$. Fig. 14 shows the ROC curves for all evaluated methods and an extra 9-layer CNN model (CNN-9) presented in [46]. The metrics for all the methods are shown in Table 5. Please note that the average sensitivity $\overline{SEN}$ is equal to accuracy ACC in this experiment according to:

$$\begin{aligned}\overline{SEN} &= \frac{1}{3}\times\left(\frac{TP_1}{TP_1+FN_1}+\frac{TP_2}{TP_2+FN_2}+\frac{TP_3}{TP_3+FN_3}\right)\\ &= \frac{1}{3}\times\frac{1}{200}\times(TP_1+TP_2+TP_3)\\ &= ACC,\end{aligned} \tag{16}$$

where TP_i and FN_i ($i \in \{1, 2, 3\}$) mean the number of true positive and false negative cases respectively when the *i-th* class is considered as the positive class and the rests as the negative classes.

From Fig. 14 and Table 5, it can be seen that the performance of all the compared networks on the osteosarcoma histology image dataset is better than that on the DDSM dataset because the histology images provide richer information in RGB images and the key features to distinguish the different classes are distributed across the whole images, which makes it easier for classification. The original PCANet provides an unsatisfactory result in that the SVM is not good at multi-class classification as the CNNs. As for the sDenseNet-24, it gains the second best $\overline{AUC}$ (0.949) and ACC (0.840), which proves the advantage of its structure. The HybridNet outperforms other networks for all the metrics. Especially, its ACC surpasses that of the DenseNet-121 by 4.2%, which

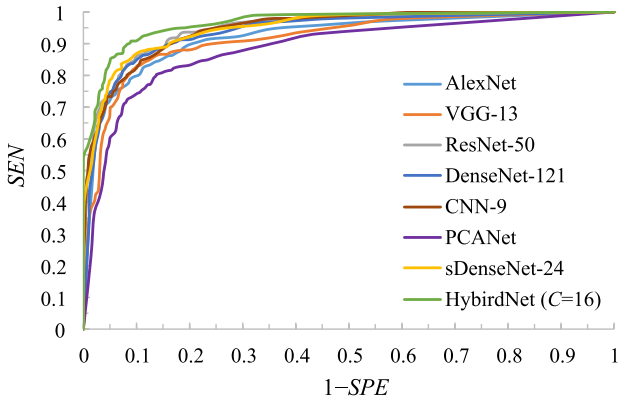

FIGURE 14. ROC curves of all the networks on the osteosarcoma histology image dataset.

verifies the superiority of combining the modified PCANet and a well-designed light-weighted CNN.

Fig. 15 shows the confusion matrixes of all the compared networks on the osteosarcoma histology image dataset. From Fig. 15, we can see that the necrotic tumor images are most difficult to distinguish among the non-tumor, viable tumor and necrotic tumor images. The reason is that the number of the necrotic tumor images in training dataset is insufficient. Both the DenseNet and the PCANet can provide the second best recognition rate 74.5% for the necrotic tumor images. However, the PCANet misclassifies 22.0% of viable tumor images as non-tumor images and the DenseNet-121 provides relatively lower recognition rate for non-tumor compared with the other methods. The sDenseNet achieves the best recognition rate 90.0% for the non-tumor images and provides the third best recognition rate 95.0% for the viable tumor images. Nevertheless, it can only correctly recognize 67.0% of necrotic tumor images. By comparison, our HybridNet can provide the third best recognition rate of 88.0% for the non-tumor images and the highest recognition rates 75.5% and 98.0% for the necrotic tumor images and the viable tumor images, respectively.

3) TEST ON THE MIAS IMAGE DATASET

To validate the universality of our network, an experiment has been conducted on MIAS database which consists of 322 mammography images of 161 patients. This database has been used for classifying breast tissues into the fatty (FA), fatty-glandular (FG) and dense-glandular (DG) classes. Breast tissue density classification can help doctor to design

TABLE 5. Metrics of all the networks on the osteosarcoma histology image dataset.

Metrics	CNN-9	AlexNet	VGG-13	ResNet-50	DenseNet-121	PCANet	sDenseNet-24	HybridNet ($C = 16$)
$\overline{SEN}$	0.815	0.802	0.818	0.833	0.830	0.767	0.840	**0.872**
$\overline{SPE}$	0.908	0.900	0.909	0.917	0.909	0.883	0.920	**0.936**
ACC	0.815	0.802	0.818	0.833	0.830	0.767	0.840	**0.872**
$\overline{AUC}$	0.947	0.921	0.916	0.949	0.938	0.885	0.949	**0.965**

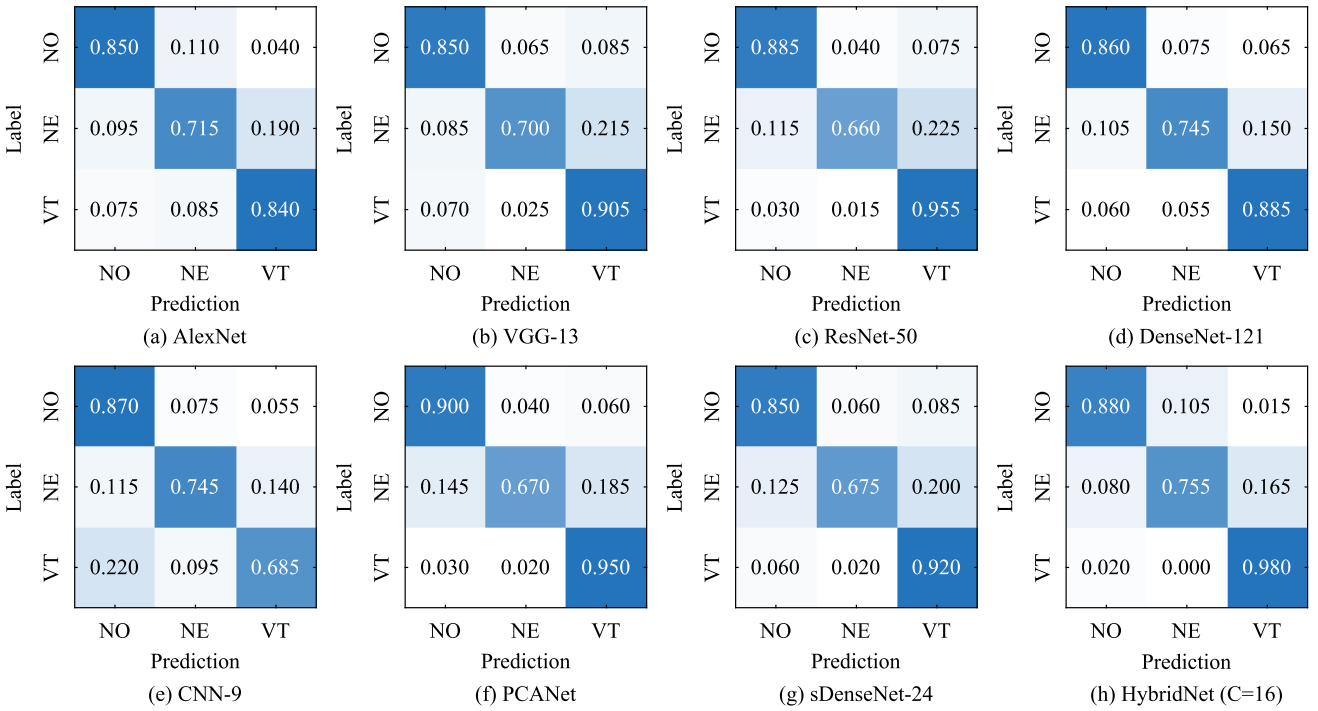

FIGURE 15. Confusion matrixes of all the compared networks on the osteosarcoma histology image dataset. "NO", "NE" and "VT" denote the non-tumor, necrotic tumor and viable tumor images, respectively.

TABLE 6. Composition of the MIAS dataset.

Image dataset	Fatty	Fatty-glandular	Dense-glandular
Training dataset	86	84	92
Test dataset	60	60	60

special diagnosis and treatment plans. In MIAS database, there are two breast images for each patient, which are highly consistent with the other. Therefore, when constructing the MIAS dataset, one should ensure that the images of one patient only belong to either the test dataset or the training dataset. The number of images in the constructed dataset is listed in Table 6.

All the images are still processed in the same way as DDSM images, and they have been augmented during the training process as mentioned in the above experiments. Since it is a multi-classification problem, the $\overline{SEN}$, $\overline{SPE}$, $\overline{AUC}$, and ACC are calculated as the metrics. The results are shown in the following Table 7, Fig.16 and Fig. 17. From Table 7 and Fig. 16, we can see that the HybridNet still obtains the best results in terms of all the metrics. Especially, the ACC of the HybridNet is 5.0% higher than the second best accuracy

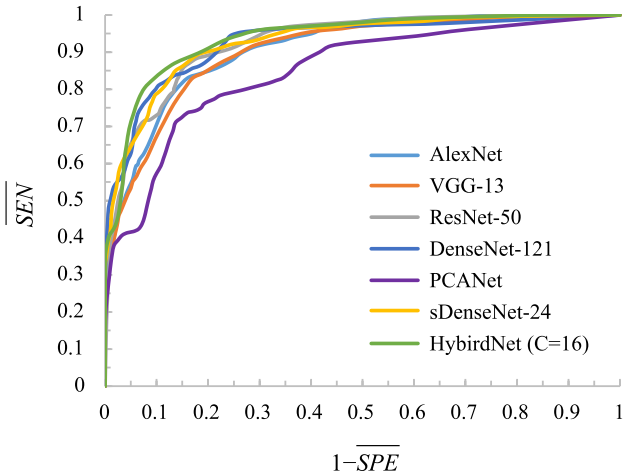

FIGURE 16. ROC curves of all the networks on the MIAS dataset.

gained by the sDenseNet-24 and the DenseNet-121, which validates that the feature maps of the modified PCANet help to improve the classification ability. Besides, the remarkable difference between the accuracy and ROC curves of

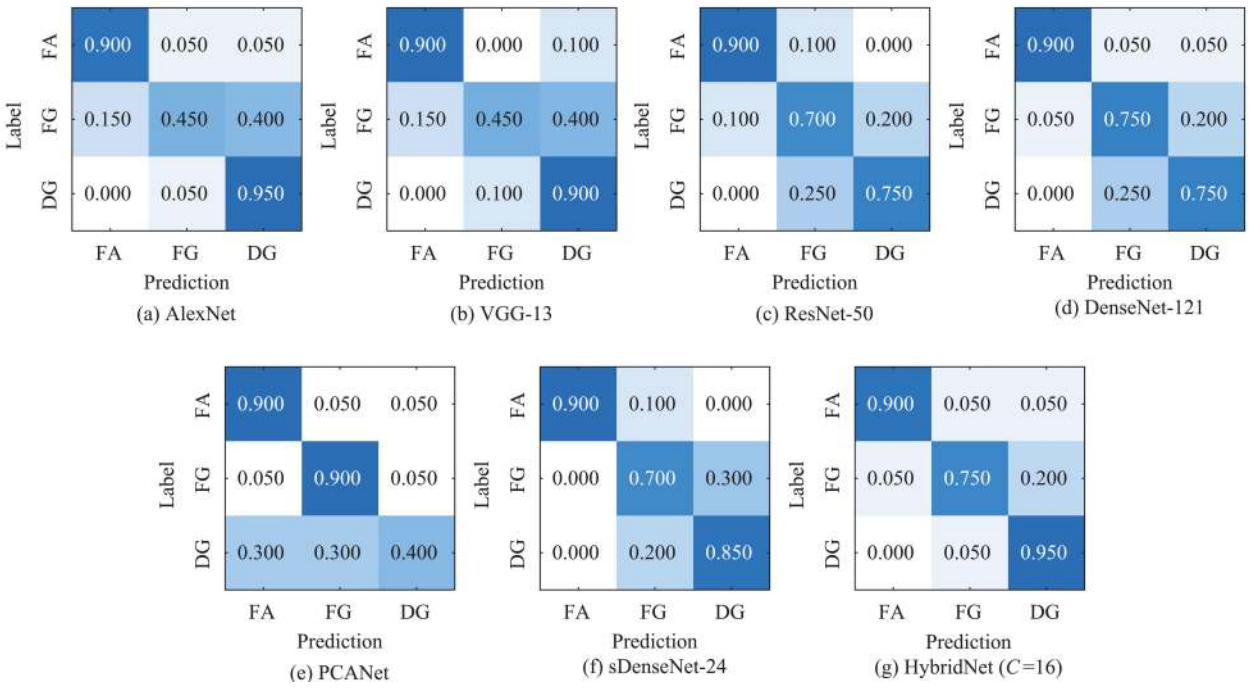

FIGURE 17. Confusion matrixes of all the compared networks on the MIAS dataset. "FA", "FG" and "DG" denote the fatty, fatty-glandular, and dense-glandular tissues, respectively.

TABLE 7. Metrics of all the networks on the MIAS dataset.

Metrics	AlexNet	VGG-13	ResNet-50	DenseNet-121	PCANet	sDenseNet-24	HybridNet ($C = 16$)
$\overline{SEN}$	0.767	0.750	0.783	0.800	0.733	0.817	**0.867**
$\overline{SPE}$	0.883	0.875	0.891	0.900	0.867	0.908	**0.933**
ACC	0.767	0.750	0.783	0.800	0.733	0.817	**0.867**
$\overline{AUC}$	0.910	0.904	0.925	0.927	0.850	0.928	**0.938**

the HybridNet and those of the PCANet demonstrates that a well-trained network is more advantageous than the SVM classifier. The results of the VGG-13 on the MIAS dataset are slightly worse than those of the AlexNet. The reason is that the VGG-13 can not achieve the global optimal parameter values after training process because there are not sufficient diverse training images.

The confusion matrixes in Fig. 17 show that all the networks are provided with 90.0% recognition rates on the images with the fatty tissues. The PCANet has the best recognition rate 90.0% for the images containing the fatty-glandular tissues while it can only distinguish 40.0% of dense-glandular tissues, which means that it will recognize some dense-glandular tissues mistakenly as the fatty-glandular tissues. In addition, both the DenseNet-121 and the HybridNet achieve the second best recognition rate 75.0% for the fatty-glandular tissues. The HybridNet obtains the highest recognition rate for the dense-glandular tissues. Overall, the HybridNet provides the best classification results for breast tissue density recognition.

IV. CONCLUSION

In this paper, we have proposed a novel light-weighted hybrid neural network for medical image classification in the case of a small amount of training image data. The proposed hybrid neural network consists of a modified PCANet and a simplified well-designed densely connected neural network. The modified PCANet overcomes the drawbacks of the original PCANet by providing more small-sized convolution kernels for the effective feature extraction. By using the feature maps of the modified PCANet as the inputs, the following simplified DenseNet can realize accurate medical image classification using a light-weighted structure with fewer adjustable weights than the traditional DenseNet. Experiments on the DDSM dataset, the osteosarcoma histology image dataset and the MIAS dataset show that the proposed hybrid neural network outperforms such popular networks as AlexNet, VGG, ResNet and DenseNet in terms of sensitivity, specificity and accuracy. The future work will be focused on designing the deeper PCANet or other unsupervised networks, and exploring the combination of feature maps in the training process.

REFERENCES

[1] G. Cosma, D. Brown, M. Archer, M. Khan, and A. G. Pockley, "A survey on computational intelligence approaches for predictive modeling in prostate cancer," *Expert Syst. Appl.*, vol. 70, pp. 1–19, Mar. 2017.

[2] M. A. Nogueira, P. H. Abreu, P. Martins, P. Machado, H. Duarte, and J. Santos, "Image descriptors in radiology images: A systematic review," *Artif. Intell. Rev.*, vol. 47, no. 4, pp. 531–559, Apr. 2017.

[3] T. Ojala, M. Pietikainen, and T. Maenpaa, "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.

[4] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Diego, CA, USA, Jul. 2005, pp. 886–893.

[6] M. R. Teague, "Image analysis via the general theory of moments," *J. Opt. Soc. Amer.*, vol. 70, no. 8, pp. 920–930, Aug. 1980.

[7] S. Morales, K. Engan, V. Naranjo, and A. Colomer, "Retinal disease screening through local binary patterns," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 184–192, Jan. 2017.

[8] S. V. Da Rocha, G. Braz Junior, A. C. Silva, A. C. De Paiva, and M. Gattass, "Texture analysis of masses malignant in mammograms images using a combined approach of diversity index and local binary patterns distribution," *Expert Syst. Appl.*, vol. 66, pp. 7–19, Dec. 2016.

[9] B. Rana, A. Juneja, M. Saxena, S. Gudwani, S. S. Kumaran, M. Behari, and R. Agrawal, "Relevant 3D local binary pattern based features from fused feature descriptor for differential diagnosis of Parkinson's disease using structural MRI," *Biomed. Signal Process. Control*, vol. 34, pp. 134–143, Apr. 2017.

[10] T. Sadad, A. Munir, T. Saba, and A. Hussain, "Fuzzy C-means and region growing based classification of tumor from mammograms using hybrid texture feature," *J. Comput. Sci.*, vol. 29, pp. 34–45, Nov. 2018.

[11] S. Gheisari, D. Catchpoole, A. Charlton, Z. Melegh, E. Gradhand, and P. Kennedy, "Computer aided classification of neuroblastoma histological images using scale invariant feature transform with feature encoding," *Diagnostics*, vol. 8, no. 3, p. 56, Aug. 2018.

[12] F. J. Pérez-Benito, F. Signol, J.-C. Pérez-Cortés, M. Pollán, B. Pérez-Gómez, D. Salas-Trejo, M. Casals, I. Martínez, and R. Llobet, "Global parenchymal texture features based on histograms of oriented gradients improve cancer development risk estimation from healthy breasts," *Comput. Methods Programs Biomed.*, vol. 177, pp. 123–132, Aug. 2019.

[13] T. Li, W. Li, Y. Yang, and W. Zhang, "Classification of brain disease in magnetic resonance images using two-stage local feature fusion," *PLoS ONE*, vol. 12, no. 2, Feb. 2017, Art. no. e0171749.

[14] S.-H. Wang, S. Du, Y. Zhang, P. Phillips, L.-N. Wu, X.-Q. Chen, and Y.-D. Zhang, "Alzheimer's disease detection by pseudo Zernike moment and linear regression classification," *CNS Neurol. Disorders-Drug Targets*, vol. 16, no. 1, pp. 11–15, Jan. 2017.

[15] S. Sharma and P. Khanna, "Computer-aided diagnosis of malignant mammograms using Zernike moments and SVM," *J. Digit. Imag.*, vol. 28, no. 1, pp. 77–90, Feb. 2015.

[16] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[17] S. Huda, J. Yearwood, H. F. Jelinek, M. M. Hassan, G. Fortino, and M. Buckland, "A hybrid feature selection with ensemble classification for imbalanced healthcare data: A case study for brain tumor diagnosis," *IEEE Access*, vol. 4, pp. 9145–9154, 2016.

[18] M. Sharif, R. Qahwaji, S. Ipson, and A. Brahma, "Medical image classification based on artificial intelligence approaches: A practical study on normal and abnormal confocal corneal images," *Appl. Soft Comput.*, vol. 36, pp. 269–282, Nov. 2015.

[19] A. S. Becker, M. Marcon, S. Ghafoor, M. C. Wurnig, T. Frauenfelder, and A. Boss, "Deep learning in mammography: Diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer," *Investigative Radiol.*, vol. 52, no. 7, pp. 434–440, Jul. 2017.

[20] X. Duan, Y. Yang, S. Tan, S. Wang, X. Feng, L. Cui, F. Feng, S. Yu, W. Wang, and Y. Wu, "Application of artificial neural network model combined with four biomarkers in auxiliary diagnosis of lung cancer," *Med. Biol. Eng. Comput.*, vol. 55, no. 8, pp. 1239–1248, Aug. 2017.

[21] G. E. Hinton, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[22] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.

[23] A. M. Abdel-Zaher and A. M. Eldeib, "Breast cancer classification using deep belief networks," *Expert Syst. Appl.*, vol. 46, pp. 139–144, Mar. 2016.

[24] H. Lim, B. Kim, G.-J. Noh, and S. Yoo, "A deep neural network-based pain classifier using a photoplethysmography signal," *Sensors*, vol. 19, no. 2, p. 384, Jan. 2019.

[25] R. Arunkumar and P. Karthigaikumar, "Multi-retinal disease classification by reduced deep learning features," *Neural Comput. Appl.*, vol. 28, no. 2, pp. 329–334, Feb. 2017.

[26] 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.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Inter. Conf. Neural Inf. Process. Syst.*, Lake Tahoe, NE, USA, 2012, pp. 1097–1105.

[28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[30] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, Jul. 2017, pp. 2261–2269.

[31] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[32] W. Zuo, F. Zhou, Z. Li, and L. Wang, "Multi-resolution CNN and knowledge transfer for candidate classification in lung nodule detection," *IEEE Access*, vol. 7, pp. 32510–32521, 2019.

[33] H. Lei, T. Han, F. Zhou, Z. Yu, J. Qin, A. Elazab, and B. Lei, "A deeply supervised residual network for HEp-2 cell classification via cross-modal transfer learning," *Pattern Recognit.*, vol. 79, pp. 290–302, Jul. 2018.

[34] L. Sun, J. Wang, Z. Hu, Y. Xu, and Z. Cui, "Multi-view convolutional neural networks for mammographic image classification," *IEEE Access*, vol. 7, pp. 126273–126282, 2019.

[35] Y. Hsieh, Y. Luo, C. Pan, M. Su, and C. Chen, "Cerebral small vessel disease biomarkers detection," *Sensors*, vol. 19, no. 11, p. 2573, Jun. 2019.

[36] 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.

[37] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification?" *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5017–5032, Dec. 2015.

[38] J. Shi, J. Wu, Y. Li, Q. Zhang, and S. Ying, "Histopathological image classification with color pattern random binary hashing-based PCANet and matrix-form classifier," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 5, pp. 1327–1337, Sep. 2017.

[39] J.-N. Lee, Y.-H. Byeon, S.-B. Pan, and K.-C. Kwak, "An EigenECG network approach based on PCANet for personal identification from ECG signal," *Sensors*, vol. 18, no. 11, p. 4024, Nov. 2018.

[40] E. Oyallon, "Scattering networks for hybrid representation learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 9, pp. 2208–2221, Sep. 2019.

[41] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, Jun. 2016, pp. 2921–2929.

[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015.

[43] F. Bianconi and A. Fernández, "Evaluation of the effects of Gabor filter parameters on texture classification," *Pattern Recognit.*, vol. 40, no. 12, pp. 3325–3335, Dec. 2007.

[44] M. Heath, Jr., K. Bowyer, D. Kopans, R. Moore, and P. Kegelmeyer, Jr, "The digital database for screening mammography," in *Proc. 5th Int. Work. Digit. Mammogr.*, Toronto, ON, Canada, 2000, pp. 212–218.

[45] M. Heath, Jr., "Current status of the digital database for screening mammography," in *Digital Mammography.*, K. Nico, T. Martin, H. Jan, and E. van Leon, Eds., Norwell, MA, USA: Kluwer, 1998, pp. 457–460.

[46] R. Mishra, O. Daescu, P. Leavey, D. Rakheja, and A. Sengupta, "Convolutional neural network for histopathological analysis of osteosarcoma," *J. Comput. Biol.*, vol. 25, no. 3, pp. 313–325, Mar. 2018.

[47] P. Leavey, "Implementation of computer-based image pattern recognition algorithms to interpret tumor necrosis; A first step in development of a novel biomarker in osteosarcoma," *Pediatric Blood Cancer*, vol. 64, p. S52, Jun. 2017.

[48] J. Suckling, "The mammographic image analysis society digital mammogram database," *Expertise Medica, Int. Congr. Ser.*, vol. 1069, pp. 375–378, Jan. 1994.
[49] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and A. L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, Jun. 2009, pp. 248–255.
[50] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *J. Digit. Imag.*, vol. 26, no. 6, pp. 1045–1057, Dec. 2013.

ZHIWEN HUANG received the B.S. degree in biomedical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2015, where he is currently pursuing the Ph.D. in biomedical engineering. His researches focus on deep learning and medical image processing.

XINGXING ZHU received the B.S. degree in biomedical engineering from the Wuhan University of Technology, Wuhan, China, in 2016. He is currently pursuing the Ph.D. degree in biomedical engineering with the Huazhong University of Science and Technology, Wuhan. His research interests include deep learning, image registration, and surgical guidance.

MINGYUE DING (Senior Member, IEEE) received the B.S. degree from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 1982, the M.S. degree from the University of Electronics Science and Technology, Chengdu, China, in 1985, and the Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China, in 1988.

Since 2006, he has been a Professor with the Department of Biomedical Engineering, School of Life Science and Technology, Huazhong University of Science and Technology. He is also with the Key Laboratory of Molecular Biophysics, Ministry of Education, Huazhong University of Science and Technology. His research interests include medical image analysis, ultrasound imaging, and image-guided surgery and therapy.

XUMING ZHANG received the B.S. and M.S. degrees in material science and engineering from the Wuhan University of Technology, in 1998 and 2001, respectively, and the Ph.D. degree in material science and engineering from Shanghai Jiao Tong University, Shanghai, China, in 2005.

From 2006 to 2008, he was a Postdoctoral Researcher with the School of Mechanical Science and Engineering, Huazhong University of Science and Technology. From 2008 to 2009, he pursued the Postdoctoral research on high-precision optical measurement with the University of California at Davis, Davis, USA. Since 2009, he has been an Associate Professor with the Department of Biomedical Engineering, Huazhong University of Science and Technology. His research interests include medical image denoising, image registration, and image restoration.