IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 30, 2019, accepted February 23, 2019, date of publication March 7, 2019, date of current version March 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2903587

Multi-Resolution CNN and Knowledge Transfer for Candidate Classification in Lung Nodule Detection

WANGXIA ZUO[1,2], FUQIANG ZHOU[1], ZUOXIN LI[1], AND LIN WANG[1]
[1]School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100083, China
[2]College of Electrical Engineering, University of South China, Hengyang 421001, China

Corresponding author: Fuqiang Zhou (zfq@buaa.edu.cn)

This work is supported by the National Natural Science Foundation of China (NSFC) under Project 61471123.

ABSTRACT The automatic lung nodule detection system can facilitate the early screening of lung cancer and timely medical interventions. However, there still exist multiple nodule candidates produced by initial rough detection in this system, and how to determine authenticity is a key problem. As this work is often challenged by the radiological heterogeneity of the computed tomography scans and the variable sizes of lung nodules, we put forward a multi-resolution convolutional neural network (CNN) to extract features of various levels and resolutions from different depth layers in the network for classification of lung nodule candidates. Through the use of knowledge transfer, the method can be divided into three steps. First, we transfer knowledge from the source CNN model which has been applied to edge detection and improve the model to a new multi-resolution model which is suitable for the image classification task. Then, the knowledge is transformed from source training progress so that all of the side-output branches in the model will be considered in the calculation. Moreover, the loss function and objective equation are improved to be image-wise calculation rather than pixel-wise. Finally, samples production and data enhancement are performed to train and test a classifier tailored for classification of lung nodule candidates. The experimental results on the LUNA16 data set show that our method gets an accuracy of 0.9733, a precision of 0.9673, and an AUC of 0.9954 while being used for lung nodule candidate classification, which is higher than the scores obtained by most of the state-of-the-art approach. In addition, when the test samples with three different sizes of 26*26, 36*36, and 48*48 are used to test the multi-resolution CNN, the accuracy rate of all three experiments exceed 92.81%, which demonstrates that the proposed model is insensitive to input scales.

INDEX TERMS Convolutional neural network, lung nodule candidate classification, multi-resolution model, knowledge transfer.

I. INTRODUCTION

Lung cancer is one kind of cancer among those which have caused the most deaths. According to the official report, the death rate caused by lung cancer has grown from the top 9 to the top 6 compared with those generated by other diseases over the past 20 years [42]. And the death toll is still increasing year by year.

Early pulmonary nodules screening is an effective means for timely medical intervention and lung cancer reduction. As a result, a large number of medical slices will be produced. Considering lung spiral CT scanning for example, almost 100 to 300 slice images will be produced for each patient after being scanned. Looking over and judging such numerous pictures will inevitably increase doctors' workload and even cause error accidents.

As is known to all, computer aided diagnosis systems (CADs) can make use of the knowledge of more than one experienced doctor and existing resources to assist radiologists for making medical diagnostics [38], [39]. The diagnosis process of lung nodules using CADs often includes rough nodular detection and nodule candidate judgment [29]. Besides, the rough nodule detection often results in multiple initial nodule candidates. Some of these nodule candidates are shown in Fig.1 (b) and (c), which are coarsely detected and segmented from CT slices similar to Fig.1(a). These initial

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

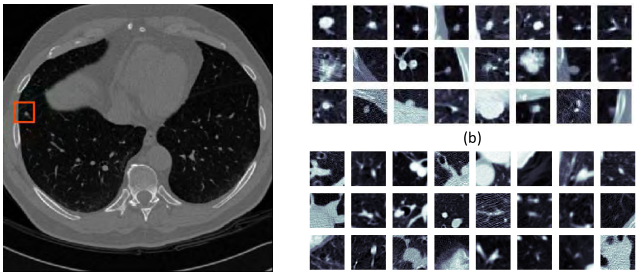


FIGURE 1. One slice image from lung spiral CT scan and some nodule candidates produced from this kind of image. (a) The CT slice with a nodule. (b) Some true nodules among the candidates. (c) Some false positives among the candidates.

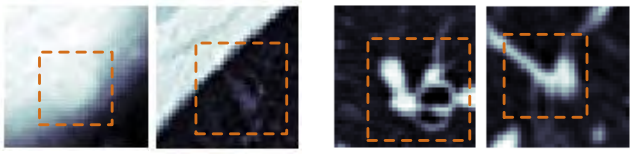


FIGURE 2. The inconsistent highlight caused by radiological heterogeneity. (a) The weakened real nodule regions. (b) The enhanced lymphatic nodes.

TABLE 1. Diameter distribution of the Lung Nodules in the LUNA16 data set.

Diameter(mm)	3=<d<6	6=<d<9	9=<d<12	12=<d<15	d >=15
Proportion(%)	42.33	29.34	10.54	7.08	10.71

nodule candidates include true nodules (Fig.1(b)) and false nodules (Fig.1(c)) which look like the true ones. Thus a decision is required to select the true nodules. And the work for selecting the true nodules is the candidate classification of pulmonary nodules, which is exactly what this paper needs to study.

To solve the classification problem of lung nodule candidates, two types of challenging issues must be paid attention to:

1) The radiological heterogeneity [12] might result in the invisibility of some nodules whereas other non-nodules are highlighted. As shown in Fig.2, where (a) is the weakened real nodule regions whereas (b) is the enhanced lymphatic nodes. This inconsistent highlight is easy to give rise to the difficulty of identifying nodules and non-nodules, directly leading to an increase in false positive candidates and false negative candidates.

2) The lung nodules are often in different sizes and variable shapes. As Table 1 shows, the diameter distribution of the lung nodules in the LUNA16 [29] data set changes in a wide range, which allows the large nodules for better identification whereas the small nodules show lower degree of identification when both are identified by a network model. It is just like that the lung nodules show different resolutions while facing the network. This will undoubtedly bring challenges to the discrimination of the network. Furthermore, the variable shapes of lung nodules also pose a challenge to the discrimination ability of the network, as mentioned in Fig.1, some lung tissues are very similar to the real nodules in shape which may lead to an increase in false positives.

To address those problems, we transfer and improve a multi-resolution CNN for lung nodule candidate classification by the way of knowledge transfer. With this method, both small nodules seemingly in low resolution and large nodules seemingly in high resolution can be recognized.

The knowledge transferred is from [24] which is viewed as the source domain. The network model in source domain can extract the features with different sizes and levels from the hidden layers of various depths, so that it can retain the main objectives described in low frequency characteristics as well as strengthen the weakened details with high frequency characteristics. By transferring this knowledge to the target domain, it is equivalent to mapping nodule candidates into features of different resolutions. Thus the algorithm in the target domain can overcome the feature representation difficulties caused by the inconsistent sizes and shapes of lung nodules and the radiological heterogeneity.

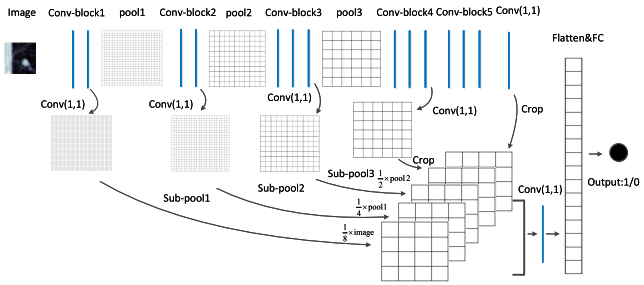


FIGURE 3. The multi-resolution depth feature mapping from multi-resolution CNN model. The grid rectangular block reflects the size of feature map. The sparser the grid is, the smaller the size of the feature map.

Our basic idea about the transferred multi-resolution model is shown in Fig.3. The feature mapping of different resolutions can be obtained from the side-outputs after each convolution block (each convolution block contains multiple layers of convolution) and corresponding pooling layers at different depths. And then all of these features are pooled or cropped into the same size and are merged together to train the classifier. The technical detail will be described in section III.

Our contribution can be concluded as followed:

1) We transfer and improve a multi-resolution convolution neural network to complete the extraction of different frequency characteristics of the nodule candidates. It is the first time for this kind of structure being used for lung nodule candidate classification. The experimental results show that it is effective to overcome the obstacle caused by the large variation in sizes and shapes of lung nodules and the radiological heterogeneity.

2) The network prototype applied in the edge detection of images is ''image-input/image-output'' pattern, which is for pixel-wise classification. We successfully resolve the critical technical problems of transforming ''image-input/image-output'' pattern to ''image-input/label-output'' pattern,

meanwhile, giving a successful attempt to convert the pixel-wise classification into an image-wise classification either. In addition, it bridges the technical gap between different fields of image application.

3) We improve the training and testing scheme of the source network to make the network performance better in target domain. As a result, we get a much more obvious distinction between classes and achieve higher scores on most of the classification metrics on lung nodule candidate classification. During the training phase, multiple classifiers are added to the branches in different depths respectively for loss calculation, and multi-level gradient iteration is carried out at the same time to jointly act on the update of weight value, thus speeding up the convergence of the network. While at the testing phase, the classification performance is improved by averaging the prediction results of the classifiers connected at the fusion layer and the deepest layer, whereas the prediction results of shallow classifiers with low reliability are ignored.

The rest of the paper is organized like this: Section II presents the related work on CADs of lung nodule classification. In section III, the details of the knowledge transfer method for network structure design and sample making are introduced. And then in section IV, the comparison of various experiments and corresponding results are presented. Based on these experiences, we make a further analysis about some key issues of the proposed approach in section V. And in the end, we give the conclusion in section VI.

II. RELATED WORKS

Among all the current applications of CAD, the task of classifying lung nodules is most similar to that of this paper. Therefore, we conducted some investigations on the current state of the art of lung nodule classification methods.

To the best of our knowledge, there are two main types of methods for the automatic classification of lung nodule currently, namely, traditional methods [2]–[6] and methods of classification using convolutional neural networks and deep learning [7]–[17].

The hand-crafted features are often used in tradition methods for nodule auto-classification. These features may include those primary level features like the texture [3], shape [2]–[4] and size [3] of lung nodules. And it may also include the high level characteristics abstracted from primary level features such as small waves [4], Local Binary Patterns (LBP) [5] and Histogram of Oriented Gradients (HOG) [6] etc. The features used are self- designated, lack of self-learning ability like computers, and the traditional methods are not intelligent enough to work in an ''end-to-end'' way.

However, the method using convolutional neural networks (CNN) can handle the above situation. As a more automated approach, the CNN method uses raw image data as input and can be directly classified as output, which is an ''end to end'' way of working. Currently, there are two main ways of classifying lung nodules through convolution neural network: 1) The 3D volume data is directly used by building a 3D-depth learning model to extract features. 2) The 3D data is first transformed into a series of 2D images, and then features are extracted from these 2D images by building a 2D depth learning model.

In the case of the 3D CNN [7]–[10], [32], it's utilized more often to segment and locate lung nodules. For example, both Chen et al. and Dou et al. have used 3D CNN for segmentation work [7], [8]. But the recent work in the paper of Anirudh *et al.* [9] shows that a method of using weakly labeled data has completed the classification of lung nodules successfully by training a 3D network. In the method, some fixed-size rectangular regions around those points marked as nodules are determined and then used as inputs to train a four-layer 3D CNN to complete the classification task. It's more efficiently for this method to make use of the spatial correlation information between CT slices. But the computational cost and the waste of memory capacity make it not suitable to be designed too complicated in the network structures of 3D CNN [30], [31], which affects the accuracy of the extracted features.

To avoid the above mentioned situation, 2D CNN methods are often used in nodules classification tasks [11]–[17], [34], [35], [41]. Among these 2D CNN methods, Li *et al.* [14] proposed a 7-layer convolutional neural network based on AlexNet structure [18] to classify lung nodules, and Kumar *et al.* [15] proposed a 5-layer auto-encoder to extract deep features for lung nodules classification. What's more, Hua *et al.* [16] proposed a deep belief network and a convolutional neural network to complete the same lung nodule classification task.

While, in these 2D CNN approaches, some cases are similar to ours in form, which all are committed to solving the challenging problem of lung nodules classification mentioned above. For example, Shen *et al.* have put forward two different models of multi-scale CNN [11] and multi-crop CNN [12] to describe the characteristics of the nodules at different levels. The multi-scale CNN method takes images with three different sizes as inputs to train three CNNs respectively. And the output features from these three CNNs are then concatenated together for classifier training. This treatment improves the classification performance in a limited degree.

The multi-crop CNN [12] is an improvement on the multi-scale CNN [11]. Instead of employing the multi-scale information of input, the multi-crop CNN takes advantage of the information from the hidden layer. That is, the central regions of feature map from one certain layer are cropped into pieces in different sizes, which are then combined together to form a new feature map to add to the next layer so as to improve the network structure. The results show that this method works the best when the feature map comes from the shallowest layer. This shows that, just like the multi-scale CNN methods, this approach is mainly based on low-level features in shallow layer which is near the input layer. However, the improvement on classifier performance is very limited only by carrying out changes on the shallow layer of the network, which is proved by the comparison experiment in section IV.

The multi-view CNN proposed by Liu and Kang [13] is another improvement on the multi-scale CNN. In the method, the input images in different sizes are resized to the same size to make multi-view areas. By taking these multi-view areas as the multiple input channels, the CNN is trained and realizes the function of extracting multi-scale features. But as we can see, this approach does not change the way at all by making a change only on the input compared with the previous one.

Considering the state of the art, we proposed our multi-resolution 2D CNN model by means of knowledge transfer. Compared with the above 2D CNN methods, our model has a great improvement in structure, which makes the classification effect competitive. The specific structural differences are discussed in section V. For specific performance improvement, please refer to the experimental part in section IV.

III. METHODOLOGY

In this section, we describe the details of knowledge transfer method applied in the classification of lung nodule candidates.

The idea of knowledge transfer is similar to that of transfer learning [1], [36], [37]. According to the definition of transfer learning, it aims to help improve the learning of the target predictive function $f_T(\bullet)$ in target domain with the knowledge in resource domain and resource task [1]. Our knowledge transfer method is also intended for this purpose. But, unlike the transfer learning in which the knowledge is often directly inherited from the source domain without change, our knowledge transfer approach aims to transfer knowledge from the source domain but make corresponding improvements to obtain similar information to facilitate tasks in the target domain. To do that, we first carry out knowledge transfer from the model in source domain to get the similar model in target domain. Then, we transfer knowledge from the training progress in source domain into the target domain to complete the classification algorithm for target domain. After that, the samples production and data enhancement in target domain are completed to retrain the transferred model because the feature space in target domain is different from that in the source domain. As described below, we will introduce our methods from the above three aspects respectively.

A. KNOWLEDGE TRANSFERRING FROM MODEL AND IMPROVEMENTS ON STRUCTURE

The source model in [24] can simultaneously extract the high frequency and low frequency information of the edge while being applied to the edge detection task. This particular function inspires us. As we mentioned earlier, nodules of different sizes which look like in multiple resolutions need to be analyzed with multi-level features. Thus, we can transfer knowledge from model in source to the target, that is, we retain the main structure of the source model in the target domain but make the following modifications:

1) The network model in [24] remains four pooling layers in its backbone which are from pool1 layer to pool4 layer. We removed the pool4 layer and only retained the first three pooling layers. This is because our input images are relatively small, too many pooling layers may lose their information, resulting in missing important details.

2) As the modified model will be adapted to the classification task, the labels feed to the output sides are no longer images, but a series of single-valued vectors (“1” or “0”), so we do not need to interpolate the compressed features back. Instead, we will remove the up-sampling section of the side-output branches but add the pooling parts to them.

3) We add a certain size of padding to the first convolutional layer and the third pooling layer so that we can obtain the uniform size of feature maps with different levels. This conduct brings two huge benefits: a) It is convenient for the subsequent layer to concatenate all of these feature maps of the same size. b) It contributes to the selection of the kernel size of the pooling portion of each side-output branch, which can be selected twice as large as the size of its stride, so that the mapping scope of the pooling operation can be covered with each other without missing important information.

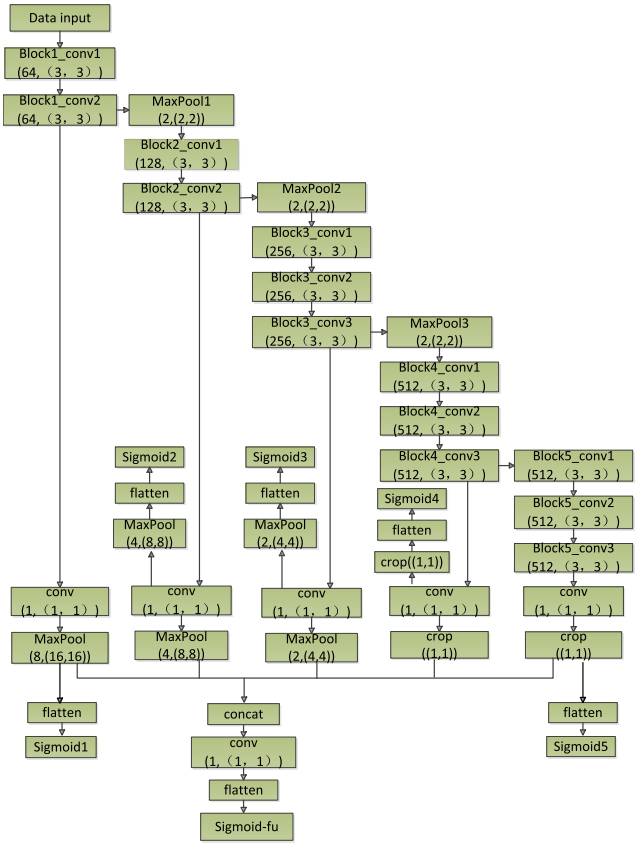


FIGURE 4. The structure of the multi-resolution CNN model. The numbers in each rectangular block represent the number and size of the filter kernel.

After the above improvements, the final model structure is shown in Fig.4. Similar to the model referred in [24], we reserve five side-output parts and their corresponding sigmoid classifiers which are named as sigmoid with an

appended number from 1 to 5 (such as: sigmoid5), each side-output branch is connected to different layers from the backbone of the network. The receptive field of each side-output branch can be obtained by calculating the receptive field of the corresponding layer in the backbone of the network.

According to this calculation, the corresponding relationship between each branch and the receiving field is shown in Table 2. As can be seen from Table 2, each side-output branch has different receptive field size. The deeper the layer, the larger the receptive field is. We can finally complete the extraction of different resolution features by integrating all of these feature maps in different receptive field sizes.

TABLE 2. The sizes of receptive fields (rf) and output feature maps (fm) in the backbone of the network.

Layer	C1_2	P1	C2_2	P2
RF size	5	6	14	16
FM size	46	23	23	11
Side-outout	Sigmoid1	-	Sigmoid2	-
Layer	C3_3	P3	C4_3	C5_3
RF size	40	44	92	140
FM size	11	6	6	6
Side-outout	Sigmoid3	-	Sigmoid4	Sigmoid5

B. KNOWLEDGE TRANSFERRING FROM TRAINING PROGRESS AND TRICKS ON ALGORITHM

It can be observe the source task in [24], which is ostensibly for edge detection, but is actually a kind of pixel-wise classification task. That is to say, the main idea of the algorithm in source domain is similar to the general classification task. Therefore, knowledge transfer can be carried out from the training process of the source task into the target algorithm to help improve the learning of the target predictive function. However, since the classification method of the target field is based on the whole image rather than on pixels, the loss function must be improved, such as, being calculated over the whole image. Besides that, in the test phase, in order to meet the requirements of image-wise classification accuracy for the target domain, the test method is improved too. The specific algorithm in the training and testing phase is as follows:

1)Training phase

During the training period, we reserve the loss values from 5 side-output branches and 1 fusion layer and adopt the method of ''joint training with separate calculation''. That is, these loss values are calculated independently in the forward recursion, but they work together for upgrading the weight when the gradient values are passed backward by using the stochastic gradient descent algorithm. The specific process is as follows:

Suppose that the training set contains m sample pairs $\{(X^{(i)}, Y^{(i)}), i = 1, 2, \cdots, m\}$. Where $X^{(i)}$ represents the ith raw image which works as the input sample, $Y^{(i)}$represents the corresponding label of $X^{(i)}$. Then the loss function of the nth side-output branch can be defined as:

$$l^n_{side}(W, w^n) = -\sum_{i=1}^{m} Y^{(i)} logP(Y^{(i)} = 1|X^{(i)}, W, w^{(n)}) - \sum_{i=1}^{m}(1 - Y^{(i)}) logP(Y^{(i)} = 0|X^{(i)}, W, w^{(n)}) \quad (1)$$

where W represents the parameter set of network standard layers, $w^{(n)}(n = 1, 2, \cdots, 5)$ represents the parameter set of those layers which only present in the nth branch. $P(Y^{(i)} = 1|X^{(i)}, W, w^{(n)})$ and $P(Y^{(i)} = 0|X^{(i)}, W, w^{(n)})$ represent the probability of ''1'' and ''0'' of the sample $X^{(i)}$ which are predicted by the classifier in the ith branch respectively. We put the parameters W and $w^{(n)}$ together and denote them as W', then the probability values can be calculated according to the formula (2) and formula (3):

$$P(Y^{(i)} = 1|X^{(i)}, W') = 1/\left(1 + e^{-W'^T X^{(i)}}\right) \quad (2)$$

$$P(Y^{(i)} = 0|X^{(i)}, W, w^{(n)}) = 1 - P(Y^{(i)} = 1|X^{(i)}, W, w^{(n)}) \quad (3)$$

Similarly, the loss function of fusion layer can be obtained, as shown in formula (4). Because that is the same kind of classifier used in the fusion layer and the side-output branches, the formula (1) and formula (4) are similar in form.

$$l_{fuse}\left(W, W^N, w^{fuse}\right) = -\sum_{i=1}^{m} Y^{(i)} logP(Y^{(i)} = 1|X^{(i)}, W, W^N, w^{fuse}) - \sum_{i=1}^{m}(1 - Y^{(i)}) logP(Y^{(i)} = 0|X^{(i)}, W, W^N, w^{fuse}) \quad (4)$$

In the formula, $W^N = (w^{(1)}, w^{(2)}, w^{(3)}, w^{(4)}.w^{(5)})$ is the integrated representation of the parameters of those layers in each branch, and w^{fuse} is the fusion weight obtained through adaptive training of the network. The model structure shows that the loss value of fusion layer has relationship both with the network parameters in each branch and with what in fusion layer.

The final objective equation is:

$$(W, W^N, w^{fuse})^* = argmin(\sum_{n=1}^{5} \alpha_n l^n_{side}(W, w^{(n)}) + \beta l_{fuse}(W, W^N, w^{fuse})) \quad (5)$$

where α_n and β are the weight coefficient allocated to each classifier, and their numerical setting is shown in section IV.

In the training phase, the parameters W, W^N and w^{fuse} can be updated with the stochastic gradient descent method at the end of each batch of data for training, until they are trained to meet accuracy requirements. The experimental results are shown in section IV.

2)Testing phase

After the training phase is completed, the multi-resolution CNN can be used to predict the lung nodule candidates.

Denote the predicted result as $\hat{Y}$, $\hat{Y} \in (0, 1)$. Since six classifiers from five side-output branches and one fusion layer are designed in this network. Then, the predicted results $\left\{\hat{Y}_{side}^{(n)}, n = 1, 2, \ldots, 5\right\}$ and $\hat{Y}_{fuse}$ can be obtained from the side-output branches and fusion layer respectively while being given the input image X. Where n represents the number of side-output branches, which ranks from shallow to deep according to the position connected to the backbone, and $n = 5$ indicates the deepest layer.

The final unified output classification results can be obtained by combining those predictions. Different from the treatment in [24], we just integrate the predictions of the deepest branch and the fusion layer as the final predicted result. The classifiers of other branches are used only for auxiliary training instead of contributing to the test results. Specific reasons are displayed in section IV. And the final output classification results are:

$$\hat{Y}_{out} = Average(\hat{Y}_{fuse}, \hat{Y}_{side}^{(5)}) \tag{6}$$

C. SAMPLES PRODUCTION AND DATA ENHANCEMENT

The source and the target are different in feature spaces, so we need to make our own samples to retrain the transferred network.

The data set we are using comes from the Grand Challenge on Lung Nodule Analysis 2016 (LUNA16) [29]. Through the use of the publicly available LIDC/IDRI database [25], [26] after being screened, the LUNA16 data set retains 888 CT scans from 888 different patients. These CT scans are all consistent in annotations made by at least three experts. The LUNA16 data set also provides document containing annotated information about the location of the nodules in those CT scans. In the annotation file, a list of 1186 nodules' coordinates and their diameter values is given. By use of these locations in the annotation file, we make our own samples.

The samples production process is divided into two stages. In the first stage, the central coordinates of nodules and non-nodules are extracted. In the second stage, we take the coordinates extracted in the first stage as the centers, expand the space around the centers to form some square patches, and then segment the square patches and label them with the corresponding class labels to get samples.

In the first stage, we extracted the center point coordinates of the positive samples on the basis of the world coordinates of the 1186 nodules provided in the annotation file. We convert the world coordinates of these nodules into voxel coordinates first, then take out the coordinate points every few pixels along the x-axis, y-axis and z-axis respectively. And these coordinates are counted as center coordinates of the positive samples. This method, in fact, is equivalent to taking a series of target points near each nodule location, which is an expansion of the number of positive samples.

As for the extraction of the central coordinates of negative samples (non-nodules), we take the points that are randomly sampled in the lung area. For the chest CT images contain the lung areas and bone areas, the negative sample coordinates obtained are more practical by limiting the sample range in the lung areas.

After the first phase, we get all the points expressed in voxel coordinates. In the second stage, we will find those slices from the CT scans where those points are located according to their z coordinate. In each of these slices, we take the point (x, y) which is defined by x and y coordinates as center point and crop a certain size of square patch around the center point. These square patches are the samples we need. The sample extraction process is shown in Fig.5.

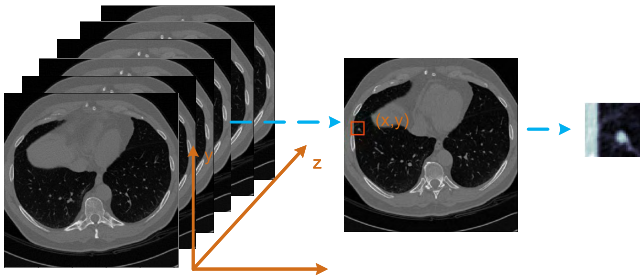


FIGURE 5. The process of the sample extraction. First, the slice position can be found according to the z coordinate. Then the sample can be cut out according to the given x and y coordinates.

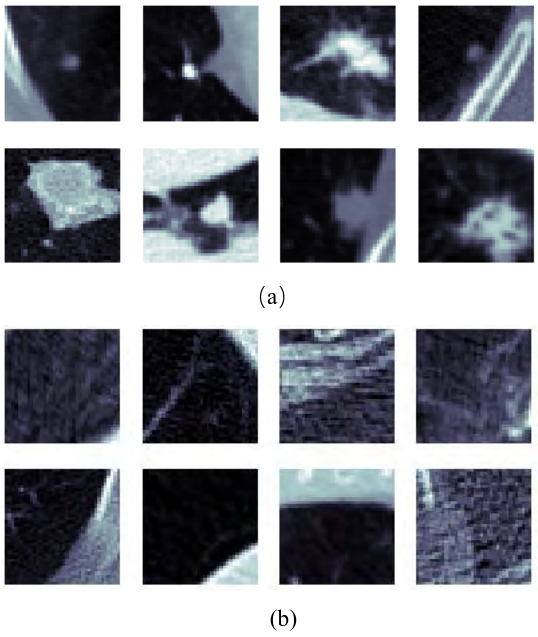


FIGURE 6. Positive and negative sample patches. (a) The positive sample patches with spherical contours in irregular shapes and different sizes. (b) The negative samples without typical spherical contours.

Based on the above method, we obtained 71,160 negative sample patches in size of 36*36 and 55752 positive sample patches in the same size. We put these samples together to form a sample set including 126,912 samples. Fig.6. shows these sample patches. We can see that there are some highlighted portions with spheroidal contours contained in the positive samples which are known as nodules or part of the nodules, but we cannot obviously find those kinds of contours in the negative samples.

IV. EXPERIMENTS AND RESULTS

In this section, a series of comparative experiments and corresponding evaluation results will be introduced. For example, in order to prove that the network structure and training scheme of this article works better than that of others, we conducted a comparison between the relevant network structures and the training schemes. In order to prove that our method is insensitive to image scales and robust to multi-resolution problems, we compared the classification results of the images with different input sizes and resolutions. In addition, we compared our method with the state-of-art approach to verify that ours is more competitive in dealing with the classification of lung nodule candidates with different resolutions.

The details of these experiments are described below.

A. THE EXPERIMENT SETTINGS AND EVALUATE METRICS

We train our network model on one GTS1080 Ti GPU and use the keras as the deep learning framework which bases on the backend of Tensorflow. In order to process medical image formats, we also use professional plug-ins like pydicom and SimpleITK.

Our experimental process is divided into training and testing phases. During the training phase, the training parameters are mainly selected through experience. We set the learning rate to 0.0001, the momentum to 0.9, the batch size to 256, and complete the training through an iteration with 50 epochs. In both of the two phases, we choose the hyper-parameters of α_n and β mainly through empirical testing. That is, the loss-weight α_n for each side-output branch is set to 1, and the fusion layer weight β set to 1 too.

In order to evaluate the results of the classification experiment, we used the evaluation metrics such as accuracy, precision, recall [28] and AUC [27].The detailed description about these metrics refer to [27].

In addition, in order to evaluate the experimental results in the same data set, we also employed the evaluation metrics provided by the LUNA16 challenge, which are the sensitivities at 1/8, 1/4, 1/2, 1, 2, 4 and 8 FPs per scan and the Competition Performance Metric (CPM) respectively [29].

B. CLASSIFICATION PERFORMANCE UNDER DIFFERENT NETWORK CONFIGURATIONS AND PREDICTION SCHEMES

After observing the structure of our multi-resolution CNN model described in section III, we discover that the side-output branches of the network seems to make more contributions to the processing of multi-resolution information. So we try to change the network configuration by removing these branches or their corresponding classifiers, and conduct the lung nodular classification experiments with the modified configurations and corresponding training methods. These comparisons are described in the item 1) below.

It can be noticed that, in section III, we have used the average outputs of multi-resolution CNN's deepest branch and fusion layer as the final prediction in the prediction phase. We explain the reasons by experiment and try to compare this with other prediction schemes as well, which are introduced in the item 2) below.

1) Network configurations and training schemes comparison

In order to verify that the multi-resolution CNN network has a better performance, based on the network structure mentioned in section III, we have designed two training schemes with two structural changes. The first option is to remove all of the side-output branches of the multi-resolution CNN network, retaining the main part and the classifier in the deepest layer only. We named this scheme as MRC_trunk. The second option is just like the description in section III. All of the five side-output branches and the final fusion layer are connected to their respective classifiers. While training, there are six loss functions corresponding to their respective six classifiers being calculated at the same time to promote the weights updating. This scheme is named as MRC_6loss.

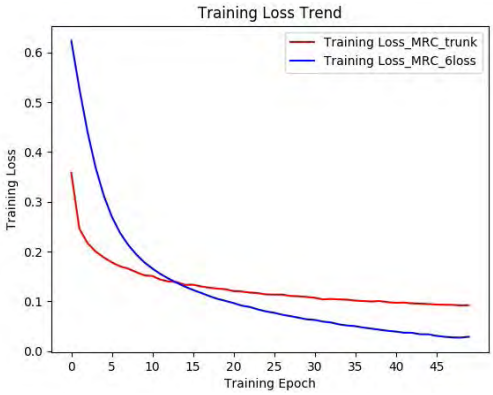


FIGURE 7. The training loss curve of the MRC_trunk scheme and our MRC_6loss scheme.

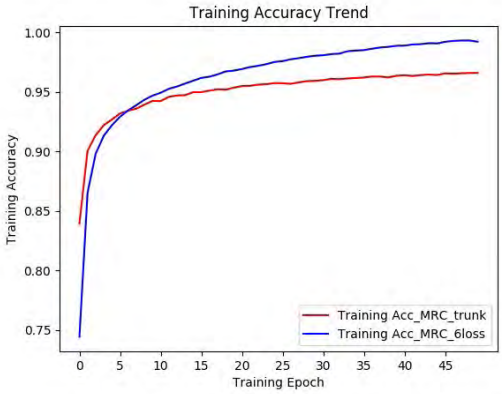


FIGURE 8. The training accuracy curve of the MRC_trunk scheme and our MRC_6loss scheme.

We set the training cycles of the two experiments to 50 epochs, and depict the loss values and accuracy values in the training process as curves shown in Fig.7 and Fig.8 respectively. As can be seen from the figures, although all of these experiments could eventually converge, the MRC_6loss

with a faster decline trend for its loss value has fallen to 0.1010 only in the 19th epoch, whereas the loss value of MRC_trunk falls to the same level almost after the 35th epoch. Similarly, in the training accuracy curve of the three schemes, the training accuracy of MRC_6loss is shown to be a faster growth too, and it is completely beyond that of MRC_trunk only after the 10th epochs. This is because each of the side-output branches of MRC_6loss could contribute to the updating of the parameters of its previous layers when the gradient iteration is carried out by using the stochastic gradient descent method. In this way, owing to those shallow layers, each update of their weight will be more accurate, which makes the network easier to converge.

Moreover, when two types of networks are used for a classification test of multi-resolution lung nodule candidates respectively, results obtained are shown in Table 3. It's worth noting that MRC_6loss has obtained a higher score on all the metrics. And especially on recall, it gets a score of 0.9726 which is far higher than that of MRC_trunk. It indicates that when all branches and corresponding classifiers are retained, multi-resolution CNN can make the training process easier to converge and the classification performance being better.

TABLE 3. Comparison of different network configurations and training schemes on their scores over the classification metrics.

Metrics	Accuracy	Precision	Recall	AUC
MRC_trunk	0.9651	0.9638	0.9572	0.9944
MRC_6loss	**0.9733**	**0.9673**	**0.9726**	**0.9954**

2) Prediction schemes comparison

As mentioned before, there are six classifiers in the MRC_6loss which give six predicted confidences correspondingly. So, we can adopt two prediction schemes for comparison, that is: (1) Just like what described in [24], we regard the average value of all these six predicted confidences from all the branches and fusion layer as the final prediction and name this prediction scheme as MRC_6loss_avg6. (2) Only take the average value of predicted confidences from the fusion layer and the deepest branch as the final predicted result, as described in section III. This prediction scheme is named as MRC_6loss_avg2.

Fig.9 shows the class confidences comparison between the MRC_6loss_avg6 scheme and the MRC_6loss_avg2 scheme under ten random test samples. It can be seen that in the MRC_6loss_avg2 scheme a real nodule in ground truth is predicted to be true nodule with a confidence above 95% whereas a non-nodule is predicted to be a true one with a confidence below 2%. The obvious confidence gap between various classes just shows that different categories are easily distinguishable in the MRC_6loss_avg2 scheme than that in MRC_6loss_avg6, as mentioned above, the MRC_6loss_avg6 scheme has a more ambiguous confidence gap for the confidence coefficient is about 62% and 30% respectively.

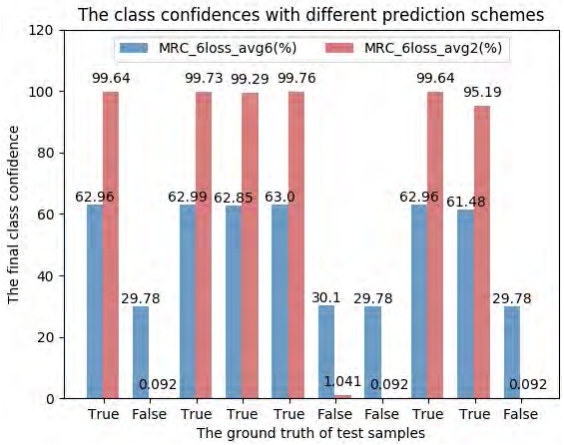


FIGURE 9. Prediction confidences comparison on different prediction schemes.

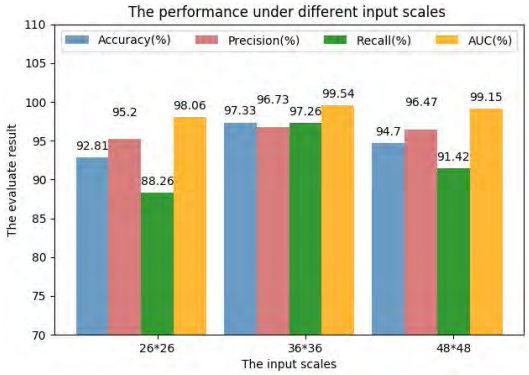


FIGURE 10. The classification performance of the multi-resolution CNN under different input scales and resolutions.

C. CLASSIFICATION PERFORMANCE UNDER INPUTS IN DIFFERENT SCALES AND RESOLUTIONS

We train the multi-resolution CNN with samples in size of 36*36. However, as described in [24], this kind of structure of the network which was transferred by the use of knowledge transfer is insensitive to input scales.

As we know, the image with big size contains more semantic information about the environment around the target, whereas the smaller one is more focused on the target itself and contains less semantic information. Therefore, for most general networks, the sizes of the input images will also affect the results of classification. To prove the multi-resolution CNN without this obvious disadvantage, we use the same way as described in section III to produce test samples with two different sizes of 26* 26 and 48*48 respectively, and employ the interpolation method to resize these samples into the same size of 36*36. Thus these test samples become test images with different resolutions but in the same size. Then we put these images with different resolutions together for experiments and compare these with test images in the same size from the original test set. The test results of these comparative experiments are shown in Fig.10 and Fig.11.

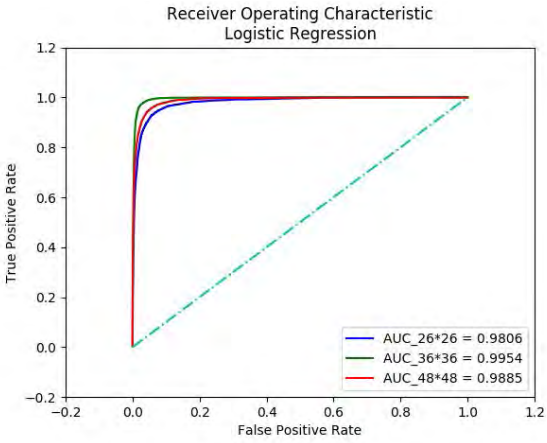


FIGURE 11. The ROC curve of multi-resolution CNN under different input scales and resolutions.

TABLE 4. Classification performance compared with the current similar approaches.

Metrics	Accuracy	AUC	Sensitivity	Specificity
Multi-scale[11]	0.8684	–	–	–
Multi-crop[12]	0.8714	0.9300	0.7700	0.9300
Multiview[13]	–	0.9810	0.8913	**0.9991**
Our work	**0.9733**	**0.9954**	**0.9726**	0.9738

As can be seen from these figures, although the multi-resolution CNN is trained with those images in size of 36* 36, which causes the classification accuracy of the images whose original size is 36* 36 slightly higher than that of the other images. But overall, all of these three experiments have an accuracy of over 92.81%. Their ROC curve also shows that the classification performance is generally good. This indicates that the multi-resolution CNN is not sensitive to the scales of inputs. What's more, since these test images have different resolutions, the experiments also prove that the multi-resolution CNN is robust to the classification of multi-resolution images.

D. CLASSIFICATION PERFORMANCE COMPARED WITH THE STATE-OF-THE-ART APPROACH

In this section, we compare our approach with the multi-scale CNN approach [11], the multi-crop CNN approach [12], and the multiview CNN approach [13]. The results are shown in Table 4. Although it is difficult to make a fair comparison because of the differences in sample size, it can be seen that our multi-resolution CNN still shows strong competitiveness on all the metrics. Among these current methods, our method is the only one to achieve its scores more than 90% in accuracy and sensitivity. And its AUC value is also the highest among four methods. Although the specificity metric of our method is slightly lower than that of the multi-view approach, it is significantly higher than that of the other methods. This shows that compared with the existing similar methods, our method has dramatically improved the performance of classification tasks related to pulmonary nodules.

Moreover, we also compared our method with the method using the same data set, such as the method mentioned in [40] and [10]. Since the LUNA16 competition also specifically provided 754,975 nodule candidates for "the false positive reduction track", by adopting the sampling method similar to that in [40], we produced 0.82 million samples from these candidates to retrain our model and then use the trained model to classify these candidates. The experimental results are shown in Fig.12 and Table 5.

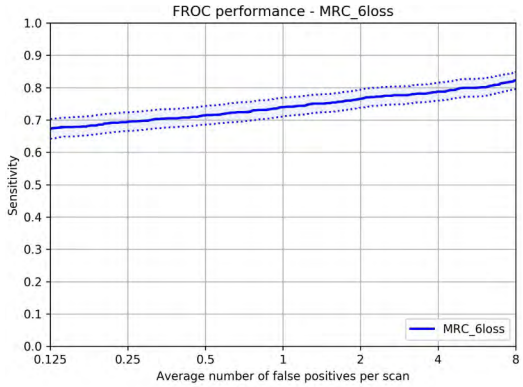


FIGURE 12. The FROC curve obtained on LUNA16 data set. The solid line is the mean sensitivity, and the dash lines are the upper bound and lower bound of sensitivity.

In Fig.12, the 95% confidence interval of the FROC curve is computed using bootstrapping with 1,000 bootstraps [29]. And in Table 5, the mean sensitivities (out of parentheses) and upper sensitivities (in parentheses) are given at each FPs/scan respectively. It can be seen that although 3D CNN gets better score in CPM, for it is better than 2D CNN in capturing the context information between slices, our model has more advantages when FPs/scan is lower. In fact, our method gained the best score at 1/8 FPs/scan among all the single CNN method. It proves that our method can achieve a good performance even with extremely low false positive rates.

However, since some methods combine multiple CNN for classification prediction, such as the Iitem03 method and the Faster R-CNNs method [40], compared with those, ours gets a higher score on CPM than the Iitem03 method and a higher sensitivity at 8 FPs/scan than the Faster R-CNNs method. But in general, the Faster R-CNNs approach performs slightly better than ours due to the idea of multi-CNNs combination strategy used in that method.

But on the other hand, as a single-CNN model, our model is more convenient to train and use than that of the multi-CNNs model. Moreover, our model can also be used as a base single model to provide reference for the multi-CNNs combination strategy.

V. DISCUSSION

A. THE ADVANTAGES OF THE PROPOSED METHOD

We used the knowledge transfer method to transfer and reconstruct a multi-resolution convolutional neural network to complete the classification task of lung nodule candidate. The network can map the image of lung nodule candidate into characteristics of different resolutions and scales while

TABLE 5. The experiment results compared with the method using the same luna16 data set. The figures in or out of parentheses represent the upper sensitivities and mean sensitivities respectively.

System name	CNN	Combination	0.125	0.25	0.5	1	2	4	8	CPM
MRC_6loss(ours)	2D	Single-CNN	**0.672**(0.703)	0.694	0.714	0.739	0.766	0.787	0.822(**0.848**)	0.742(0.762)
LungNess	2D	Single-CNN	0.453	0.535	0.591	0.635	0.696	0.741	0.797	0.635
Iitem03	2D	Two-CNNs	0.394	0.491	0.570	0.660	0.732	0.795	0.851	0.642
Faster R-CNNs[40]	2D	Three-CNNs	– (**0.734**)	–	–	–	–	–	– (0.834)	– (**0.790**)
LUNA16CAD	3D	Single-CNN	0.640	0.698	**0.750**	**0.804**	**0.847**	**0.874**	**0.897**	**0.787**

encountering the difficulty in characteristics description of pulmonary nodule for its radiological heterogeneity and variabilities in sizes and shapes, thus greatly reducing the false positive of classification task and improving the scores on classification metrics.

B. TRAIN FROM SCRATCH OR PRE_TRAIN?

Our knowledge transfer method is somewhat similar to the idea of transfer learning. But for most transfer learning, the network structures tend to be unaltered. And just for this reason, it is often convenient for the transferred network to preload the weight of the target model before training, which amounts to pre-train the network. However, in contrast to that usual treatment, to make the model structure more suitable for the target task, we have made some improvements on the model structure. This treatment makes it difficult for preloading the existent weight to the transferred network, so the proposed method does not pre-train the network like the general transfer learning indeed. In spite of this, the experimental results shown in Fig.8 above have proved that without pre-training, multi-resolution CNN can still converge rapidly while being randomly weighted. The reason might be attributed to our sufficient samples and reasonable selection of training program.

C. THE ANTI-OVERFITTING MEASURES

It is often necessary for network training to solve the problem of overfitting. As far as I know, the network structure of the multi-resolution CNN with more parameters is more complicated than what others applied in similar applications, which makes measures to prevent overfitting. There are two kinds of anti-overfitting measures commonly used. One is through the network structure changing, like using a simpler network structure or adding a dropout layer into the network. The other is to expand the number of samples, such as, using sample enhancement strategies to diversify the sample. As for the first approach, it might hurt the network's classification performance by simplifying the network structure, plus the method of adding the dropout layer has been proved to be less effective for medical applications in [12]. That's why we make an attempt to expand the number of samples. As described in section III, we obtain the expanded samples by translating the nodule candidate blocks along the x, y and z axis respectively in the sample preparation phase.

D. DIFFERENCES AND IMPROVEMENTS COMPARED WITH SIMILAR METHODS

The method presented in this paper is similar to that proposed by Shen *et al.* [11], [12] and Liu and Kang [13] in form, but it has a great improvement in performance in comparison with their methods. There are two main reasons:

1) The 2D network structures used in the methods proposed by Shen *et al.* and Liu *et al.* are relatively simple and have fewer layers. But for classification tasks, the deeper features are more conducive to the classification of targets [33]. Therefore, the depth of the network model should reach a certain degree to achieve more accurate expression of the targets, which is also one of the reasons the naming of deep learning. In this sense, just like what are shown in Table 6, our model is made up of 16 layers and is deeper than the previous three kinds of 2D network models. This might make our network model more effective.

2) The methods proposed by Shen *et al.* and Liu *et al.* focus only on integrating the characteristics extracted from one layer of the network, but our method integrates the features of different resolutions extracted from different layers of the network. Fig.13 shows the differences in their optimization patterns, where the subfigures (a) and (c) show the two methods proposed by Shen *et al.*, whose fusion features are from the output layer and the shallowest hidden layer respectively.

TABLE 6. The number of layers for four different network models.

Model	Multi-scale[11]	Multi-crop[12]	Multiview[13]	Ours
Number of layers	5	7	5	16

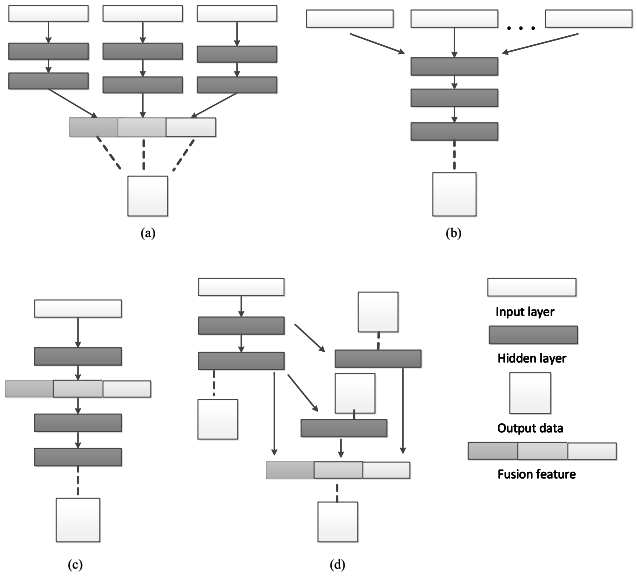


FIGURE 13. Current improvement patterns of CNN for the classification of multi-resolution lung nodules. (a) The multi-scale CNN [11]. (b) The multi-view CNN [13]. (c) The multi-crop CNN [12]. (d) Our method.

The subfigure (b) shows the method proposed by Liu *et al.*, whose fusion features are from the output layer. The subfigure (d) is our proposed method, which combines the characteristics of different levels from multiple layers. As our method focuses on the multiple analysis of multi-resolution features, it is more effective to solve the problem of multi-resolution lung nodule candidates classification.

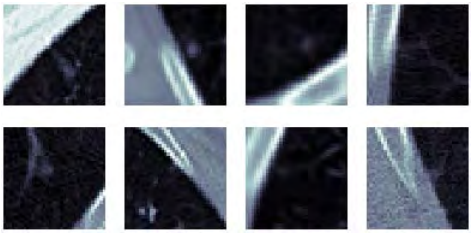


FIGURE 14. Some of the misidentified nodule candidates without being fully highlighted.

E. LIMITATIONS AND FUTURE WORKS

Although the characteristics of lung nodules can be analyzed in multiple resolutions by our proposed method, thus some of the less obvious nodules caused by radiological heterogeneity can be successfully identified as well. There are still some limitations that need to be addressed in this study. First, a complete lung nodule is often distributed on multiple slices. However, our 2D CNN method is limited in capturing the contextual information between slices, thus a 3D CNN model may be considered in our future work. Second, our model is trained on LUNA16 data set. But some types of nodules are not fully represented or not fully highlighted in the data set, which may lead to the false identification of nodules. Some of these misidentified nodule candidates are shown in Fig.14. We believe that a larger data set for sample preparation and effective sample preprocessing before training will help identify these candidates, which will also be part of our future research work.

VI. CONCLUSION

We transfer and improve a multi-resolution CNN to solve the challenging problems posed by radiological heterogeneity and variable sizes and shapes of nodules in the classification of lung nodule candidates using the method of knowledge transfer. We demonstrate the importance and effectiveness of bringing in multi-resolution feature information extracted from this network in the case of the classification task of lung nodule candidate. The experimental results on the LUNA16 data set show that the proposed method achieves a competitive score on the evaluation index of the classification performance. The network we have transferred is a network model that has been proved effective on edge extraction tasks, therefore in principle our approach is universal. As a classification method for lung nodule candidate, our method can effectively improve the accuracy of lung nodule screening combined with the effective initial rough detection method.

ACKNOWLEDGMENT

The authors would like to thank the LUNA16 challenge organizers for providing the dataset.

REFERENCES

[1] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010. doi: 10.1109/TTHZ.2016.2544142.

[2] X. Ye*, X. Lin, J. Dehmeshki, G. Slabaugh, and G. Beddoe, "Shape-based computer-aided detection of lung nodules in thoracic CT images," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 7, pp. 1810–1820, Jul. 2009.

[3] F. Liu, D.-H. Lee, R. Lagoa, and S. Kumar, "Computer-aided detection of lung nodules using outer surface features," *Bio-Med. Mater. Eng.*, vol. 26, no. s1, pp. 1213–1222, 2015.

[4] P. Aggarwal, R. Vig, and H. K. Sardana, "Patient-wise versus nodule-wise classification of annotated pulmonary nodules using pathologically confirmed cases," *J. Comput.*, vol. 8, no. 9, pp. 2245–2256, Sep. 2013.

[5] 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.

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Diego, CA, USA, Jun. 2005, pp. 886–893. doi: 10.1109/CVPR.2005.177.

[7] H. Chen, Q. Dou, X. Wang, J. Qin, J. C. Y. Cheng, and P.-A. Heng, "3D fully convolutional networks for intervertebral disc localization and segmentation," in *Proc. MIAR*, Bern, Switzerland, 2016, pp. 375–382.

[8] Q. Dou *et al.*, "Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1182–1195, May 2016.

[9] R. Anirudh, J. J. Thiagarajan, T. Bremer, and H. Kim, "Lung nodule detection using 3D convolutional neural networks trained on weakly labeled data," *Proc. SPIE*, vol. 9785, Mar. 2016, Art. no. 978532.

[10] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, "Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1558–1567, Jul. 2017.

[11] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," *Inf. Process. Med. Imag.*, vol. 24, pp. 588–599, 2015.

[12] W. Shen *et al.*, "Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification," *Pattern Recognit.*, vol. 61, pp. 663–673, Jan. 2017.

[13] K. Liu and G. Kang, "Multiview convolutional neural networks for lung nodule classification," *Int. J. Imag. Syst. Technol.*, vol. 27, no. 1, pp. 12–22, Mar. 2017.

[14] W. Li, P. Cao, D. Zhao, and J. Wang, "Pulmonary nodule classification with deep convolutional neural networks on computed tomography images," *Comput. Math. Method. Med.*, vol. 2016, Nov. 2016, Art. no. 6215085.

[15] D. Kumar, A. Wong, and D. A. Clausi, "Lung nodule classification using deep features in CT images," in *Proc. CRV*, Jun. 2015, pp. 133–138.

[16] K.-L. Hua, C.-H. Hsu, S. C. Hidayati, W.-H. Cheng, and Y.-J. Chen, "Computer-aided classification of lung nodules on computed tomography images via deep learning technique," *Onco Targets Ther.*, vol. 8, pp. 2015–2022, Aug. 2015.

[17] H. Lee, H. Lee, M. Park, and J. Kim, "Contextual convolutional neural networks for lung nodule classification using Gaussian-weighted average image patches," *Proc. SPIE*, vol. 10134, Mar. 2017, Art. no. 1013423.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NIPS*, Lakex Tahoe, NV, USA, 2012, pp. 1097–1105.

[19] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556

[20] C. Szegedy *et al.* (2014). "Going Deeper with Convolutions." [Online]. Available: https://arxiv.org/abs/1409.4842

[21] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. ECCV*, Amsterdam, The Netherlands, 2016, pp. 21–37.

[22] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, Munich, Germany, 2015, pp. 234–241.

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

[24] S. Xie and Z. Tu, "Holistically-nested edge detection," *Int. J. Comput. Vis.*, vol. 125, nos. 1–3, pp. 3–18, Dec. 2017.

[25] A. P. Reeves *et al.*, "The lung image database consortium (LIDC): A comparison of different size metrics for pulmonary nodule measurements," *Acad. Radiol.*, vol. 14, no. 12, pp. 1475–1485, Dec. 2007.

[26] S. G. Armato, III, *et al.*, "The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans," *Med. Phys.*, vol. 38, no. 2, pp. 915–931, 2011.

[27] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, 1997.

[28] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and F-score, with implication for evaluation," *Int. J. Radiat. Biol.*, vol. 51, no. 5, pp. 199–239, 2005.

[29] A. A. A. Setio *et al.*, "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge," *Med. Image. Anal.*, vol. 42, pp. 1–13, Dec. 2017.

[30] G. Urban, M. Bendszus, F. Hamprecht, and J. Kleesiek, "Multi-modal brain tumor segmentation using deep convolutional neural networks," in *Proc. MICCAI BraTS*, 2014, pp. 31–35.

[31] S. C. Turaga *et al.*, "Convolutional networks can learn to generate affinity graphs for image segmentation," *Neural Comput.*, vol. 22, no. 2, pp. 511–538, 2010.

[32] I. R. S. Valente, P. C. Cortez, E. C. Neto, J. M. Soares, V. H. C. de Albuquerque, and J. M. R. Tavares, "Automatic 3D pulmonary nodule detection in CT images: A survey," *Comput. Methods Programs Biomed.*, vol. 124, pp. 91–107, Feb. 2016.

[33] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 936–944.

[34] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[35] C. Li, G. Zhu, X. Wu, and Y. Wang, "False-positive reduction on lung nodules detection in chest radiographs by ensemble of convolutional neural networks," *IEEE Access*, vol. 6, pp. 16060–16067, Apr. 2018.

[36] J. Yang, S. Li, and W. Xu, "Active learning for visual image classification method based on transfer learning," *IEEE Access*, vol. 6, pp. 187–198, Feb. 2018.

[37] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang, "Transfer learning using computational intelligence: A survey," *Knowl.-Based Syst.*, vol. 80, pp. 14–23, May 2015.

[38] T. Messay, R. C. Hardie, and S. K. Rogers, "A new computationally efficient CAD system for pulmonary nodule detection in CT imagery," *Med. Image Anal.*, vol. 14, no. 3, pp. 390–406, Jun. 2010.

[39] M. Javaid, M. Javid, M. Z. U. Rehman, and S. I. AliShah, "A novel approach to CAD system for the detection of lung nodules in CT images," *Comput. Meth. Programs Biomed.*, vol. 135, pp. 125–139, Oct. 2016.

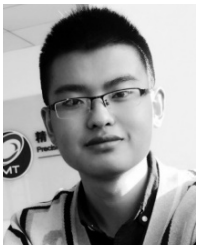
[40] X. Hongtao, D. Yang, N. Sun, Z. Chen, and Y. Zhang, "Automated pulmonary nodule detection in CT images using deep convolutional neural networks," *Pattern Recognit.*, vol. 85, pp. 109–119, Jan. 2019.

[41] M. Patrice, S. Qi, M. Xu, F. Han, X. Zhao, and W. Qian, "CNN models discriminating between pulmonary micro-nodules and non-nodules from CT images," *Biomed. Eng. Online*, vol. 17, no. 7, p. 96, Jul. 2018.

[42] World Health Organization. (May 2018). *The Top 10 Causes of Death*. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

WANGXIA ZUO received the B.S. degree in electric automatization from the Wuhan University of Hydraulic and Electric Engineering, Wuhan, China, in 2001, and the M.S. degree in control theory and control engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2006. She is currently pursuing the Ph.D. degree in measurement technology and instruments from Beihang University, Beijing, China. Her current research interests include image processing and deep learning.

FUQIANG ZHOU received the B.S., M.S., and Ph.D. degrees in instrument, measurement and test technology from Tianjin University, Tianjin, China, in 1994, 1997, and 2000, respectively. He joined the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China, as a Postdoctoral Research Fellow, in 2000, where he is currently a Professor with the School of Instrumentation and Optoelectronic Engineering. His current research interests include computer vision and image processing.



ZUOXIN LI received the B.S. degree in measurement technology and instruments from Beihang University, Beijing, China, in 2016, where he is currently pursuing the M.S. degree in measurement technology and instruments with the Key Laboratory of Precision Opto-Mechatronics Technology of Ministry of Education. His current research interests include machine learning, and deep learning applications and theories in computer vision.

LIN WANG received the B.S. degree in mathematics and applied mathematics from Weinan Normal University, Shannxi, China, in 2010, and the M.S. degree in mathematics from the North China University of Technology, Beijing, China, in 2015. She is currently pursuing the Ph.D. degree in measurement technology and instruments with the Key Laboratory of Precision Opto-Mechatronics Technology of Ministry of Education, Beihang University, Beijing, China. Her current research interests include image processing and deep learning.