IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 31, 2019, accepted January 11, 2020, date of publication January 17, 2020, date of current version January 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2967238

Pulmonary Nodules Detection Based on Deformable Convolution

JUNHUA GU[1,2], ZEPEI TIAN[3], AND YONGJUN QI[1,4]

[1]State Key Laboratory of Reliability and Intelligence of Electrical Equipment, Hebei University of Technology, Tianjin 300401, China
[2]Hebei Province Key Laboratory of Big Data Calculation, Hebei University of Technology, Tianjin 300401, China
[3]School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China
[4]Information Technology Center, North China Institute of Aerospace Engineering, Langfang 065000, China

Corresponding author: Yongjun Qi (qiyongjun.lf@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 6170215, in part by the NSF of Hebei Province through the Key Program under Grant F2016202144, in part by the National Natural Science Foundation of China under Grant 41804118, and in part by the Scientific and Technological Research Projects of Hebei Province under Grant QN2018085.

ABSTRACT Early detection of malignant pulmonary nodules is of great help to the treatment of lung cancer. Yet it is difficult to establish a general diagnostic standard because of the two main characteristics of pulmonary nodules: different sizes and irregular shapes. To address this problem effectively, an improved pulmonary nodule detection model based on deformable convolution is proposed. Specifically, by adding a branch network to obtain the offsets, the process of feature extraction is more suitable with the shape of nodule itself. Besides, a simple but effective strategy is proposed for the size variability of pulmonary nodules, which is combined with the multilevel information as well as the fusion of different sizes feature maps. Compared with the two-dimensional convolution neural network and other advanced technologies, our method has a significant improvement, and its mean average precision can achieve 82.7%.

INDEX TERMS Computer-aided detection, Lung cancer, Pulmonary nodules.

I. INTRODUCTION

Lung cancer is the most common malignant tumor, which poses a great threat to human health. If lung cancer can be found and treated early, the survival rate of patients can be significantly improved. The early lesions of lung cancer are in the form of pulmonary nodules. Therefore, the key to finding lung cancer is to accurately extract the features of pulmonary nodules and locate the location of the lesions, which is of great significance to the diagnosis of lung cancer. The NLST (National Lung Screening Trial) experiment showed that the mortality rate of lung cancer decreased by 20% after 7 years by using low-dose chest CT (Computed Tomography) to examine the high-risk population 3 times a year. This experiment proved that CT images have a high tissue resolution, are the most effective imaging method for the diagnosis of lung diseases, and have an important value for the detection and diagnosis of lung cancer [1], [2]. However, research data show that the annual growth rate of CT images data is about 30%, while the annual growth rate of radiologists is only about 4.1% [3]. Comparatively speaking, the number of radiologists is far less than that number of images data. Therefore, it is necessary to use computer technology to assist doctors to detect and identify pulmonary nodules.

Since the 1990s, the computer-aided system has been applied to the research of medical imaging. In the early lung cancer screening aspect, the system can catch quantitatively the features from the CT images and locate the suspicious nodule region, which can not only provide reliable reference information for radiologists, but also reduce the workload of doctors reading CT images [4]. At present, experts and scholars at home and abroad have proposed many effective lung nodule detection algorithms. The traditional method is to extract the gray and texture features of pulmonary nodules, and then use SVM or other classification algorithms to train the extracted features for detecting the suspected pulmonary nodules [5]. The shape, size, and texture of pulmonary nodules are highly variable, so there are some challenges in the process of screening lung cancer by using accurate CAD technology:

1) There are enough individual differences between normal tissue and pathological changes in CT images.
2) It was found that the shapes of pulmonary nodules are not a regular circle, some were long and thin, some

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

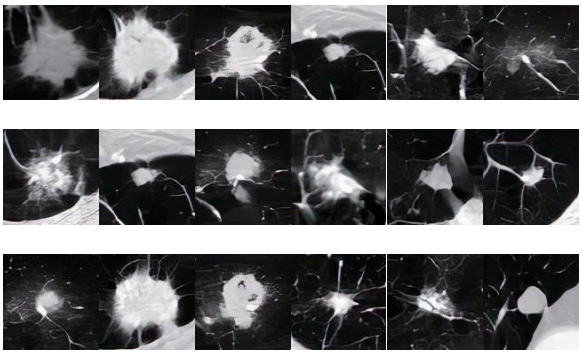

FIGURE 1. From the picture, the pulmonary nodule varies in shape. They are round, oval, slender and so on. They are distributed in different positions of the lung parenchyma, which makes the task of pulmonary nodules detection challenging.

were oval, and some were other shapes. The edge of the nodule is very similar to the surrounding lung parenchyma [6]. Fig. 1. depicts the diversity of nodule shapes

3) The size of pulmonary nodules is between 3-40 mm, which belongs to small target detection.

With the increase in the number of samples, the traditional machine learning methods gradually become difficult to adapt to the complex samples. Due to the limited representation ability of complex functions and weak learning ability, only primary features can be extracted. At the same time, the traditional machine learning method can not effectively catch the rich information contained in the CT images for the complex steps of selecting features manually sometimes. Deep learning algorithm is a new field of machine learning method, which is a kind of deep neural network. By simulating the human brain to build hierarchical model, it has a strong ability of automatic feature extraction and efficient feature expression, which eliminates the need for hand-made features. The CT images are directly entered into the network model, and the recognition results are directly obtained in the output layer, so as to improve the detection rate of pulmonary nodules, reduce the complexity of the algorithm and save the detection time. Based on this, we start from the difficulty of task detection, aiming at the relatively different size of pulmonary nodules and the diversity of pulmonary nodule shape, and propose an improved pulmonary nodule detection model based on deformable convolution. We draw the following three innovations:

1) In view of the diversity of pulmonary nodule shapes, we propose a deformable convolution structure based on convolution neural network. Its core is centering on the addition of a branch network to get the offsets, and learning the offsets from the object through training. The sampling point is no longer the original square, but changes according to the shape of the target. For example, if the shape of pulmonary nodule is slender, by adding offset to the sampling point over the input feature map, the original square sampling position becomes similar to that slender, which makes the feature extraction process more consistent with the characteristics of the target itself, so as to ensure a better overall and effective feature extraction.
2) A simple but effective strategy is proposed for the size variability of pulmonary nodules. High-level feature mapping has higher resolution and stronger semantics, although low-level feature mapping can learn sensitive local features. Therefore, this paper deconvolute high-level feature map and combine it with low-level feature map, which has rich feature information function. Then these feature maps form a feature pyramid as the prediction layer.
3) We verify and evaluate the proposed model on widely recognized public datasets, and also compare the variants of our model and discuss the impact when taking different local structures of the network, which be constructive and help for future pulmonary nodules detection research.

The rest of this paper is organized as follows. In Section II, the related works for pulmonary nodules detection research are reviewed. In Section III, the architecture and the details of the proposed model are described. Experimental results are provided in Section IV. Section V makes a brief conclusion and outlines future work.

II. RELATIONWORK

With the improvement of computing ability, the convolution neural network(CNN) can directly realize the feature learning from low-level to high-level for the input image, and demonstrate superior performance in several computer vision applications [7]. That also inspired some researchers to employ CNNs in automated pulmonary nodule detection. Some of them endeavored to design the network structure according to the representative characteristics of pulmonary nodules. For example, Setio *et al.* [8] designed three different convolution neural network models based on solid, subsolid and large solid nodules of pulmonary nodules, using the full connection to fuse the results, so as to improve the accuracy of detection results. Based on the different sizes of pulmonary nodules, Dou *et al.* [9] designed three kinds of 3D convolution neural network models, and finally combined the results according to the proportion of datasets. Based on the small data sets of pulmonary nodules, Zhu *et al.* [10] proposed the combination of 3D convolution network with Expectation Maximization idea, and applied weak supervision method to the task of pulmonary nodule detection. In a single CT image, the pulmonary tissues such as blood vessels and trachea may resemble pulmonary nodules, while in a 3D image, the shape of pulmonary nodules is similar to that of spheres, and the blood vessels and trachea are extended. In view of this point, Ypsilantis *et al.* [11] proposed a ReCTnet network model, which adds residual blocks and Long Short-Term Memory (LSTM) to extract slice image features. LSTM module is used to learn the dependence between slices and fully considers all the information of the image. Liao *et al.* [12] proposed

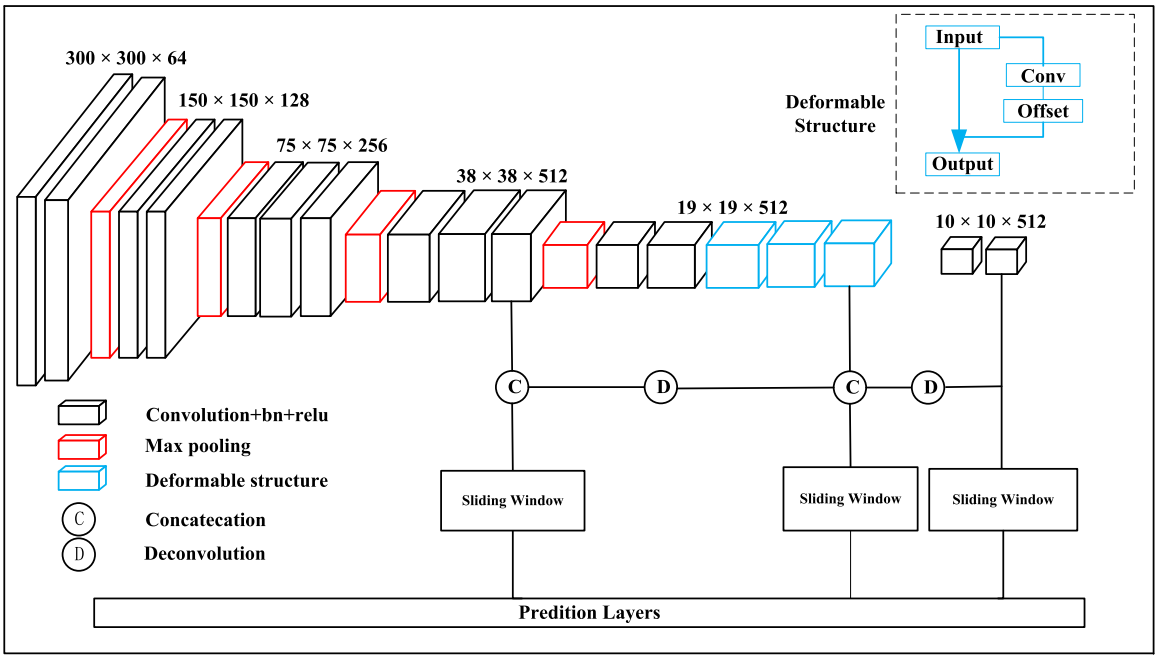

FIGURE 2. Different colors used represent different types of convolution layers. Our network is divided into VGG up to conv5_2, three deformable convolution layers, two convolution layers. And three feature maps in different locations are selected as the prediction layers. It is known that feature maps of different levels in the network have different receptive ïĄeld sizes. To determine the size of nodules, we design three anchors of different sizes for each sliding window: 3×3, 12×12, 24×24.

a adopting a three-dimensional deep neural network. In addition to convolution layers and pooling layers, this model also includes deconvolution, feature fusion and residual block structures. For feature extraction, Chen *et al.* [13] combined Dense Block with Residual Block to form a two-way network, which is conducive to accurately extract object features.

Transfer learning is also one of the common methods. Its idea is to adjust the tasks of one's own by using the classical object detection model. These object detection models have a good performance in the natural image data sets. The results are remarkable and have been carefully studied by scholars and tested by practice. Based on these models, we apply them to specific fields by modifying or extending them. That approach may yield better performance than recreating the network model. Region proposals with convolution neural network (R-CNN) introduces CNN method into the field of object detection, which greatly improves the effect of object detection, and is a milestone in the application of CNN to the problem of object detection. Four basic steps of nodule feature extraction, candidate region generation, location regression and classification are unified into a complete deep network framework, which can ensure a high recall rate [14]. For instance, Ding *et al.* [15] used two-dimensional Faster Region-based Convolution Neural Network (Faster R-CNN) combined with three-dimensional CNN to reduce false positive rate to detect pulmonary nodules. The other is the method of object detection based on regression, which uses an end-to-end network to directly predict object classes and locations [16]. Khosravan and Bagci [17] used a single feed-forward pass of a single network for pulmonary nodule detection, which adopts Densely Connected Convolutional Networks (DenseNet) to extract image features. Through the above papers, the application of the deep convolution neural network in the detection and recognition of pulmonary nodules has a great auxiliary role in the early stage of lung cancer screening, and is also the development trend of medical image processing.

III. OUR METHOD

The proposed model is similar to SSD model [18], using the basic network model VGG16 [19] to extract features. VGG16 network consists of thirteen 3 × 3 convolution kernels, three full connection layers and five 2 × 2 pooling layers. Compared with other networks, VGG16 uses more continuous and smaller convolution kernels, which increases more nonlinear transformation and the depth of network, and improves performance in a certain degree under the condition of ensuring the same receptive field. This paper keep layers of VGG16 from conv1_1 to conv5_3, converting fc6 and fc7 to convolution layers, then add extra convolution layers behind them. In addition to the basic network structure, the innovation of our structure is to propose a deformable convolution structure because of the irregular shape of pulmonary nodules. At the same time, the size of pulmonary nodules is small, and the feature maps are fused to form a multi-level feature prediction structure. The whole network structure is shown in the Fig. 2. The detailed description will be given below.

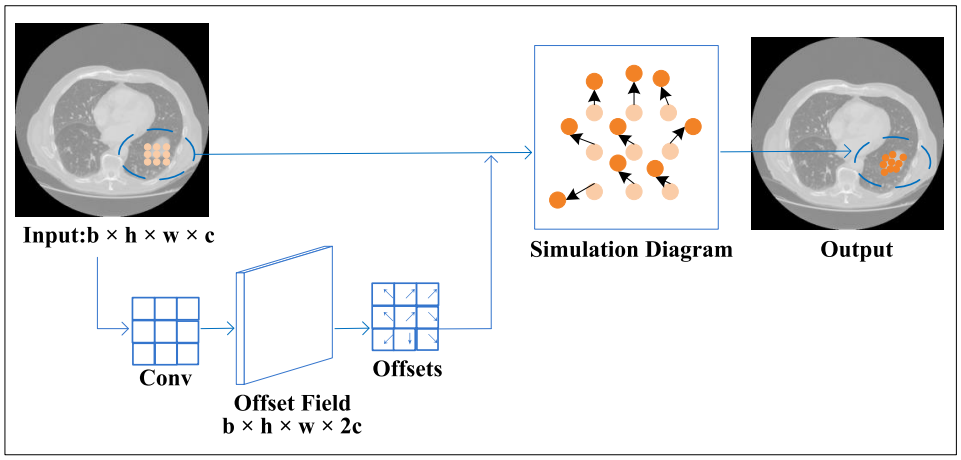

FIGURE 3. As shown in the figure above, this figure represents the deformable convolution process. The offsets are obtained through the side branch convolution, that is, the offset field, which represents the offset of each point in the X and Y directions. The sampling points on the input feature map are added to the offset so that the position of the sampling points changes. Finally, irregular sampling points are obtained in the output image.

A. DEFORMABLE CONVOLUTION THEORY

Inspired by the ideas of space transformation network [20] and deformable parts model [21], [22]. Deformable convolution can generally be divided into two steps. The first step is to get the offsets by learning from the bypass convolution. And then adopt bilinear interpolation to determine the value of the sampling points. Grid R represents the size and dilation of the receptive field. For example, $R = \{(-1,-1),(-1,0),\ldots,(1,0),(1,1)\}$, defining a 3×3 kernel. $x(p_0+p_n)$ represents the different positions of the input feature map. $w(p_n)$represents the weight of the convolution kernel. $y(p_0)$represents the different positions of the output feature map. So according to the principle of convolution, for each location p_0 of the output feature map y. We can represent it as

$$y(p_0) = \sum_{p_n \in R} w(p_n) * x(p_0 + p_n) \tag{1}$$

In deformable convolution, supposing that the input feature map $(b \times h \times w \times c)$, and the temporary values $(b \times h \times w \times 2c)$ are obtained by the bypass convolution, which is what we called the offset field, where 2c represents the offsets in both x and y directions. Then the sample points over the input feature map are added with the offsets to change the convolution position, so that the convolution process is more suitable for the object itself. Fig. 3. illustrates the process intuitively. As a result, the equation (1) becomes

$$y(p_0) = \sum_{p_n \in R} w(p_n) * x(p_0 + p_n + \Delta p_n) \tag{2}$$

Then, considering that the non-integer coordinates can not be used for discrete image data, as the offset Δp_n may be typically fractional. The equation (2) is implemented via bilinear interpolation as:

$$x(p) = \sum_{q} W_i(q_i, p) \times x(q_i) \tag{3}$$

$$W_i(q_i, p) = w_i\left(q_{i_x}, p_x\right) \times x\left(q_{i_y}, p_y\right) \tag{4}$$

where $x(p)$ represents the sampling point after processing. q_i enumerates all integral spatial locations in the feature map x, and w_i refer to the corresponding weight of each location. The equation (4) represents the product relationship between the all integral spatial locations and the optimized point, $w_i\left(q_{i_x}, p_x\right)$ is the expression of the integral spatial locations and the X direction of the optimized point. $\left(q_{i_y}, p_y\right)$ is the integral spatial locations and the Y direction of the optimized point. In particular, after using the bilinear interpolation method to determine the position of the sampling point, the whole process can realize the standard back-propagation for end-to-end training.

B. MULTI-LEVEL FEATURE PREDICTION STRUCTURE

For different sizes of pulmonary nodules, a multi-level feature mapping layer strategy is proposed. The largest pulmonary nodule is about 40 mm, which belongs to small object detection. Although low-level feature mapping can learn sensitive local features, high-level feature mapping has higher resolution and stronger semantics [23]. Therefore, by deconvoluting the high-level feature maps, we can get more semantic information in the deconvolution feature maps, and combine it with the low-level feature maps, which has richer feature information function, eliminating the impact of small object as much as possible. We use these feature maps to form the feature pyramid as the prediction layer, as shown in Fig. 4. Pulmonary nodules vary in size, ranging from 3 to 40 mm, these feature maps of different level layers are fused by concatenation. We use different proportions of anchor size on different level feature maps can ensure that different size

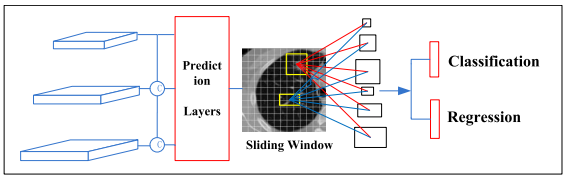

FIGURE 4. After fusion, multi-level feature maps form the prediction layer together. Three different basic sizes are set for the sliding window over the prediction layers. According to each size, they are transformed into extra corresponding rectangles, so all of which have six sizes, such as the red and blue lines in the graph. After that, each anchor is classified and regressed.

pulmonary nodules can be detected, then make classification judgment and location regression for each anchor.

C. LOSS FUNCTION

Loss function includes Log loss for classification and smooth L1 for regression and is the weighted sum of the two. α is the weight of both formulas and set to 1 in this paper.

$$Loss = L_{class} + \alpha L_{reg} \tag{5}$$

l and g represent the default box and the ground truth box respectively. The localization loss is a Smooth L1 loss.

$$L_{reg}(l, g) = \sum_{i} smooth_{L1}(l_i - g_i) \tag{6}$$

in which

$$smooth_{L1} = \begin{cases} 0.5x^2, & if\ |x| < 1 \\ |x| - 0.5, & otherwise \end{cases} \tag{7}$$

is a robust L1 loss that is less sensitive to outliers than the L2 loss. Regression loss denotes that a $smooth_{L1}$ loss between the i-th default box and the ground truth box parameters. Classification loss is defined by:

$$L_{class} = plog\left(p'\right) + (1 - p)\, log\left(1 - p'\right) \tag{8}$$

p' and p represent the predicted probability and label respectively. Intersection over Union (IOU) is used to determine the threshold of each default box. Default boxes whose IOU with the target nodule larger than 0.5 are treated as positive samples.

IV. EXPERIMENT

A. DATA SETS

LIDC-IDRI (lung image database consortium and image database resource initiative) has been extensively used for pulmonary nodule detection, and a total of 1,018 research cases are included in the data sets. In each case, four authoritative radiologists performed two stages of diagnostic marking. In the first stage, each expert independently diagnosed and marked the nodule's location, with three categories: 1) nodules size $\geq$ 3mm, 2) nodules size $<$ 3mm, and 3) nodules size $\geq$ 3mm but non-nodules. In the second phase, each expert reviewed the labels of the other three experts independently and gave his final diagnosis [24]. This two-stage annotation can annotate all the results as completely as possible avoiding forcing consensus. A total of 36,378 nodules were labeled in LIDC-IDRI. The clinical treatment has shown that nodules under 8 mm are less likely to develop into lung cancer. At the same time, referring to the current research situation in this field, this paper chooses nodes larger than 3mm in the data sets. Nodes with a diameter $<$ 3 mm and non-nodules were not included. There were 11,509 nodules with diameter $<$ 3 mm and 19,004 non-nodule, so 5,765 were left. For these 5,765 nodules, if two nodules are too close to each other (the overlapping part of two nodules is larger than half of the nodule), the two nodules are combined. The combined center and radius are the mean value of the two nodules. After such processing, there are 2,290 nodules left, forming a new datasets. In the 2,290 nodules, 2,290 were labeled by at least one expert, 1,602 by at least two experts, 1,186 by at least three experts, and 777 by at least four experts. The distribution of data sets is shown in the Fig. 5.

Similar to the PASCAL VOC datasets format used in object detection tasks in computer vision. Each node consists of multiple point position coordinates. In order to ensure the fairness of the node positions and the accuracy of the training samples, the maximum and minimum values of the coordinates marked by each doctor are first taken, and then their average values are taken to obtain the coordinates of the two extreme points. Subtract two extreme points to form a matrix label. By processing the average value of the data, the labels are more fair and the accuracy of the data is guaranteed. At the same time, similar to SSD model [18], the data sets is enlarged by operations such as flipping and random slicing of pictures.

B. PREDICTION LAYERS

In this paper, multi-level feature maps as prediction layers are a feature pyramid formed by the fusion of feature maps of different sizes. In the model, the size of these convolution layers decreases gradually by convolution and pooling operation, forming multiple scales feature maps. In order to measure the advantage gained, we test each layer separately and compare the results. For a fair comparison, we adopt the same anchor every time, which conforms to the number and size of other layers. Table. 1. shows the accuracy for each layer. We select the feature maps with better effect, then add the feature fusion operation as the prediction layers. In prediction layers, for each scale (except last one), half of the feature maps are learned from the previous scale through a series of convolution layers, while the remaining half feature maps are directly up-sampled from the contiguous high-semantics feature maps with using deconvolution method, which actually brings a multi-level feature map for each scale from all of its contiguous scales. After that, we fuse the feature maps of two parts in three ways: sum, product, channel concatenation. The results of each approach are shown in Table. 2.

C. DEFORMABLE CONVOLUTION POSITION

This part mainly discusses how to add deformable convolution. A lot of experiments have been done using the different

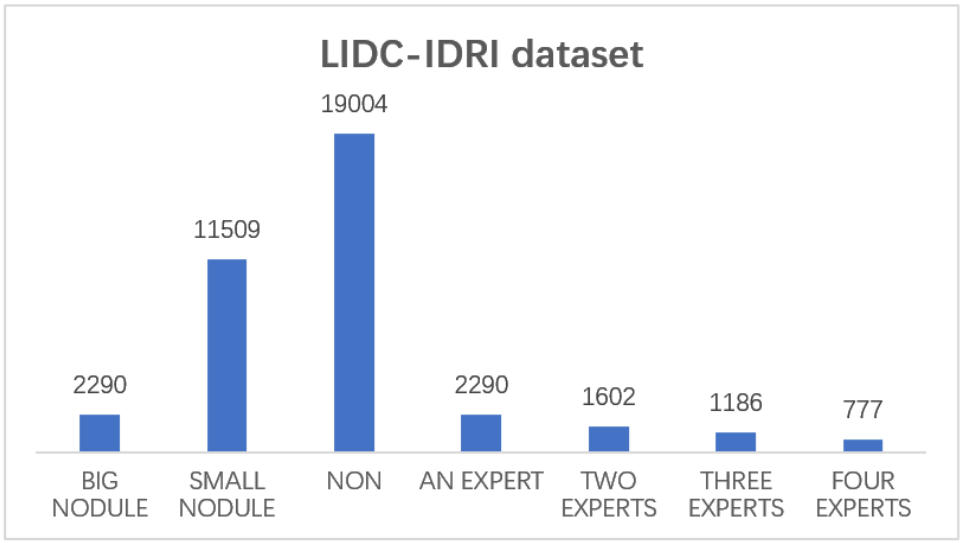

FIGURE 5. The picture shows the statistical analysis of the LIDC-IDRI datasets. Three types are marked: large nodules with diameter ≥3 mm, small nodules with diameter < 3 mm and non-nodules (but with pulmonary aberration). Nodules are independently labeled by experts in two stages. According to this rule, there are 2,290, 1,602, 1,186, 777 nodules labeled by at least 1, 2, 3, and 4 experts, respectively.

TABLE 1. The table shows the results of using different levels of feature maps as the prediction map. At the same time, our results also reflect the information contained in each level of feature map from the side, as well as the contribution to the task of pulmonary nodule detection. The higher the accuracy of the selected feature map, the more abundant the information contained in the feature map of this level.

$Conv3\_3$	$Conv4\_3$	$Conv5\_3$	$Fc7(conv)$	$Conv6\_2$	MAP
✓				✓	62%
	✓			✓	72.58%
		✓		✓	70.85%
			✓	✓	72.91%
				✓	71.82%
✓	✓	✓	✓	✓	74.53%
	✓		✓	✓	74.93%

TABLE 2. Results of feature fusion.

Method	MAP
SUM	79.8%
PROD	82.1%
CONCATENATE	82.7%

number of deformable convolution in different locations of the network. We first try to use the deformable convolution module for the whole network model, and the network model does not converge. Then use $Conv4\_1$ as the boundary and divide it into shallow and deep networks, deformable convolutions are used for all shallow and deep networks respectively. It is found that in shallow networks, even if the learning rate is adjusted to a very small level, the network not converge. After that, the number of deformable convolution modules is gradually decreased, and found that three

TABLE 3. Results of using deformable convolution with different numbers of such layers in different locations in the network. We use horizontal lines to represent non-convergence. In this paper, the former network and the latter network of*fc*6 and *fc*7 have experimented as groups respectively.

Deformable convolution position	MAP
C1($Conv1\_1$, $Conv1\_2$)	-
C2($Conv2\_1$, $Conv2\_2$)	-
C3($Conv3\_1$, $Conv3\_2$)	-
C4($Conv4\_1$, $Conv4\_2$,$Conv4\_3$)	64.5%
C5($Conv5\_1$, $Conv5\_2$,$Conv5\_3$)	80.6%
C6($Conv5\_1$, $fc6$, $fc7$)	81.8%
C7($fc6$,$fc7$, $Conv6\_1$)	79.5%

layers are the best. Therefore, deformable convolution from shallow to high level in 3units are added. According to the experimental results, it is determined that the best result can be obtained by adding deformable convolution to the deep network. Detailed experimental results are shown in Table. 3.

D. RESULT

This paper adopts the method of ten-fold cross-validation. Data sets are divided into ten parts, eight parts of them for training data and two parts for validating data and testing data. Repeat 10 times and take the average of the results. The Caffe framework is used and the learning rate strategy is multistep, with 40,000,60,000 and 120,000 for each stage respectively and the initial learning rate is 0.01. This paper focuses on the task of lung nodule detection based on CT images, that is, a CT image is given to determine the probability of the suspicious part belonging to the lung nodule and its location information. Finally, the model needs to evaluate the performance of classification and location regression at

TABLE 4. Model comparison.

Method	MAP
SSD [18]	77.8%
Faster r-cnn [13]	81.5%
Model [25]	81.2%
Ours	82.7%

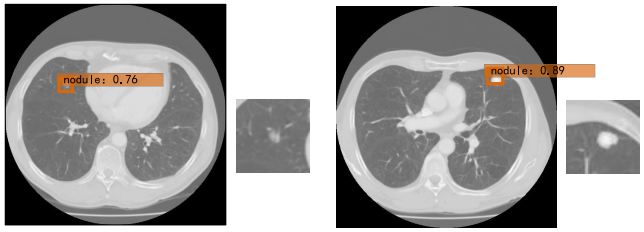

FIGURE 6. The results of two CT images of different sizes, next to these are the enlarged image corresponding to the nodule position in the CT image.

the same time. So this paper choose the commonly used mAP (mean average precision) in object detection as the evaluation criterion. Due to the migration method used, We mainly compared our approach with Faster R-CNN [14] and Single Shot MultiBox Detector (SSD) [18] which are the most advanced object detection methods in the field of computer vision. Based on the same data sets, three methods are trained and tested simultaneously. The experimental results show that our model has better performance. The experimental results are shown in Table. 4. In addition, Fig. 6. shows the results of our model on the test set. We selected two CT images of different sizes. The orange rectangle represents the location of the nodule, the text ''nodule'' is the label name, and the number is the probability of belonging to the pulmonary nodule. We see that although the nodes are very small, the model can still be detected well, which reflects that our model has good performance.

V. CONCLUSION

The convolution neural network has obvious advantages in the field of image classification and recognition. It can directly input images into the network model, automatically extract features, and output category probability and location information. In view of the small sizes and different shapes of pulmonary nodules, two innovative points was added on the basic convolution network, feature fusion prediction layer and deformable convolution structure. The experimental results show that the methods are feasible for the task, what makes sense is further combined with other methods to improve the detection performance. Deformable convolution can be applied to the research of medical problems with similar characteristics, or further extended to 3D medical image task, which can be added to the three-dimensional convolution neural network after being converted to three-dimensional. That may be our next step.

REFERENCES

[1] I. Sluimer, A. Schilham, M. Prokop, and B. Van Ginneken, ''Computer analysis of computed tomography scans of the lung: A survey,'' *IEEE Trans. Med. Imag.*, vol. 25, no. 4, pp. 385–405, Apr. 2006.
[2] B. Van Ginneken *et al.*, ''Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The ANODE09 study,'' *Med. Image Anal.*, vol. 14, no. 6, pp. 707–722, Dec. 2010.
[3] A. Jame, ''Reduced lung-cancer mortality with low-dose computed tomographic screening,'' *New England J. Med.*, vol. 365, no. 5, pp. 395–409, 2011.
[4] B. Golosio, G. L. Masala, A. Piccioli, P. Oliva, M. Carpinelli, R. Cataldo, P. Cerello, F. De Carlo, F. Falaschi, M. E. Fantacci, G. Gargano, P. Kasae, and M. Torsello, ''A novel multithreshold method for nodule detection in lung CT,'' *Med. Phys.*, vol. 36, no. 8, pp. 3607–3618, Jul. 2009.
[5] 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.
[6] M. Firmino, A. H. Morais, R. M. Mendoça, M. R. Dantas, H. R. Hekis, and R. Valentim, ''Computer-aided detection system for lung cancer in computed tomography scans: Review and future prospects,'' *Biomed. Eng. OnLineBioMed Eng. OnLine*, vol. 13, no. 1, p. 41, 2014.
[7] Y. Lecun, Y. Bengio, and G. Hinton, ''Deep learning,'' *Nature*, vol. 521, no. 7553, p. 436, 2015.
[8] A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. Van Riel, M. M. W. Wille, M. Naqibullah, C. I. Sanchez, and B. Van Ginneken, ''Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks,'' *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1160–1169, May 2016.
[9] 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.
[10] W. Zhu, Y. S. Vang, Y. Huang, and X. Xie, ''Deepem: Deep 3D convnets with EM for weakly supervised pulmonary nodule detection,'' in *Proc. 21st Int. Conf. Med. Image Comput. Comput. Assist. Intervent. MICCAI*, Granada, Spain, Sep. 2018, pp. 812–820.
[11] P.-P. Ypsilantis and G. Montana, ''Recurrent convolutional networks for pulmonary nodule detection in CT imaging,'' 2016, *arXiv:1609.09143*. [Online]. Available: https://arxiv.org/abs/1609.09143
[12] F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, ''Evaluate the malignancy of pulmonary nodules using the 3-D deep leaky noisy-OR network,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3484–3495, Nov. 2019.
[13] Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng, ''Dual path networks,'' 2017, *arXiv:1707.01629*. [Online]. Available: https://arxiv.org/abs/1707.01629
[14] S. Ren, K. He, R. Girshick, and J. Sun, ''Faster R-CNN: Towards real-time object detection with region proposal networks,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[15] J. Ding, A. Li, Z. Hu, and L. Wang, ''Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks,'' in *Proc. 20th Int. Conf. Med. Image Comput. Comput. Assist. Intervent. MICCAI*, Quebec City, QC, Canada, Sep. 2017, pp. 559–567.
[16] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, ''You only look once: Unified, real-time object detection,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 779–788.
[17] N. Khosravan and U. Bagci, ''S4ND: Single-shot single-scale lung nodule detection,'' in *Proc. 21st Int. Conf. Med. Image Comput. Comput. Assist. Intervent. MICCAI*, Granada, Spain, Sep. 2018, pp. 794–802.
[18] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C. Fu, and A. C. Berg, ''SSD: Single shot multibox detector,'' in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands, Oct. 2016, pp. 21–37.
[19] K. Simonyan and A. Zisserman, ''Very deep convolutional networks for large-scale image recognition,'' in *Proc. 3rd Int. Conf. Learn. Representations (ICLR)*, San Diego, CA, USA, May 2015.
[20] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, ''Spatial transformer networks,'' in *Proc. Adv. Neural Inf. Process. Syst., Annu. Conf. Neural Inf. Process. Syst.*, Montreal, QC, Canada, Dec. 2015, pp. 2017–2025.
[21] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, ''Deformable convolutional networks,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 764–773.

[22] D. Forsyth, "Object detection with discriminatively trained part-based models," *Computer*, vol. 47, no. 2, pp. 6–7, Feb. 2014.
[23] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "DSSD: Deconvolutional single shot detector," 2017, *arXiv:1701.06659*. [Online]. Available: https://arxiv.org/abs/1701.06659
[24] A. P. Reeves, A. M. Biancardi, T. V. Apanasovich, C. R. Meyer, H. Macmahon, E. J. Van Beek, E. A. Kazerooni, D. Yankelevitz, M. F. Mcnitt-Gray, and G. Mclennan, "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, 2007.
[25] R. Gruetzemacher and A. Gupta, "Using deep learning for pulmonary nodule detection & diagnosis," in *Proc. 22nd Americas Conf. Inf. Syst. (AMCIS)*, San Diego, CA, USA, Aug. 2016.

JUNHUA GU received the Ph.D. degree from the Hebei University of Technology, in 1997. His main research interests include data mining, intelligent information processing, and intelligent computing. He is a member of the CCF.

ZEPEI TIAN is currently pursuing the master's degree with the Hebei University of Technology. His main research interests include image processing, deep learning, and medical image detection.

YONGJUN QI is currently pursuing the Ph.D. degree with the Hebei University of Technology, Tianjin, China. He is also a Senior Engineer with the Information Technology Center, North China Institute of Aerospace Engineering, Langfang, China. His main research interests include engineering electromagnetic field comprehensive effect, deep learning, and image retrieval.