IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 6, 2019, accepted December 27, 2019, date of publication January 3, 2020, date of current version January 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2963926

Deformation and Refined Features Based Lesion Detection on Chest X-Ray

CE LI[1], DONG ZHANG[1], SHAOYI DU[2], (Member, IEEE), AND ZHIQIANG TIAN[3]
[1]College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China
[2]School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[3]School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Corresponding author: Ce Li (xjtulice@gmail.com)

This work was supported in part by the National Natural Science Foundation (NSFC) of China under Grant 61866022, Grant 61876148, and Grant 61573274, and in part by the Gansu Province Basic Research Innovation Group under Project 1506RJIA031.

ABSTRACT Automatic and accurate detection of chest X-ray lesion is a challenging task. In the chest X-ray image, the lesions are shown with blurred boundary contours, different sizes, variable shapes, uneven density, etc. Besides, the deep convolutional neural network (CNN) consists of traditional convolution units, which has the limitations of rectangular sampling. The CNN extracts difficultly the deformation and refined features of chest X-ray lesions. Because of these factors, the accuracy of the lesion detection algorithm is not high. To deal with problems, we propose the deformation and refined features based lesion detection on the chest X-ray algorithm called DRCXNet. Firstly, the deformable convolution with amplitude modulation (AMDCN) is built to extract the deformation features of the lesions on the chest X-ray. Secondly, to obtain the refined feature, the global features and local features are fused, which can enrich the feature space of the lesion. Thirdly, the pooling layer combines with the AMDCN and region proposal network to establish the deformable pooling layer, which enhances the model's sensitivity to the lesion location. During the training, the model is optimized by the improved regression loss function with a gradient control factor. On the public datasets RSNA and ChestX-ray8, the proposed method outperforms seven popular detection algorithms. The proposed method is a significant performance in both qualitative and quantitative experiments. Its comprehensive evaluation scores, sensitivity, precision, and the mean dice similarity coefficient are 0.866, 0.914, 0.836 and 0.859 respectively. The proposed algorithm achieves a very satisfactory result.

INDEX TERMS Chest X-ray, deformable convolution, deformation feature, refined feature, lesion detection.

I. INTRODUCTION

As an essential part of the respiratory system, the quantitative diagnosis evaluation of the lung is critical. The chest X-ray examination is one of the most common and cost-effective medical imaging techniques, which screen lung diseases and others [1]. Given an X-ray image, radiologists can identify acute and chronic cardiopulmonary disease, or verify that the pacemaker is orthodontic. However, the proportion of radiologists is declining, especially in areas with inadequate medical resources [2], [3]. Therefore, high-precision automated image screening technology can improve the work efficiency of radiologists and allow doctors with more time to focus on diagnosis. Moreover, the technology is also expanded to remote areas and compensates for the lack of local medical services. In the past few decades, although computer-aided diagnosis (CAD) technology assists the radiologist to read and analyze image data [4]–[6]. However, in CAD systems, many lesion detection algorithms are designed by shallow machine learning or manual feature methods [7], e.g., on the qualitative diagnosis of the pulmonary nodule in CT images, the CAD system still has a large number of leaky detection nodule lesions. Therefore, Bergtholdt *et al.* [8] proposed a support vector machine (SVM) classifier, and the sensitivity of classifier is 0.859. Alam *et al.* [9] built a classifier based on PCA, linear SVM, and multi-kernel SVM, which can compare and discriminate between healthy controls (HC) and Alzheimer's disease (AD) patients. This classifier up to about 0.84 stratification accuracy with multi-kernel SVM along with high sensitivity and specificity above 0.85. However, the accuracy of these traditional algorithms in detecting lesions generally not meet the diagnosis's requirements at

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

current medical images analysis. Besides, due to the differences of patient or imaging equipment, the lesion detection of traditional algorithms are weakly robust and generalization, and expand the application in other tasks difficultly. Therefore, the reading and analysis of chest X-ray images are mainly performed by radiologists. In recent years, the risk of lung disease increases dramatically due to environmental factors, and the data of lung X-ray screening also increases dramatically. Due to the characteristics of lesions are an insignificant feature, variable size, morphology, and the subtle changes of lesion texture, it is difficult for doctors to analyze a large number of chest X-ray images manually in long hours of work. Furthermore, manual screening relies on doctors' experience, which leads to some mistakes inevitably. Therefore, for the above issue, we can develop an automatic lesion detection algorithm by computer vision technology in chest X-ray. It will be a high clinical diagnostic value to assist doctors in diagnosis.

In recent years, with the development of artificial intelligence technology, deep learning has been widely applied in the field of medical image analysis and promotes the development of medical image analysis to precision medicine [10], [11]. In 2017, Litjens *et al.* [12] surveyed and analyzed 300 articles of deep learning about medical image analysis. They classified the methods and the applied scenes of the model. The report pointed out that deep learning methods, especially convolutional neural networks, have quickly become a method of usually favorite for medical image analysis. Deep learning is benefited from large-scale label data to learn models. It is used to apply end-to-end learning tasks, and its performance is significant compared with the traditional algorithm [13]. e.g., in works of Pattrapisetwong and Chiracharit [14] and Anavi *et al.* [15], they used deep convolutional neural networks (CNN) to extract different types of lesion features in the chest X-ray image and to classify multiple classes of lesions. Esteva *et al.* [16] proposed a detection algorithm based deep learning on the dermatological grading of skin cancer, which reached the level of dermatologists. Zhu *et al.* [17] designed a DeepLung model, which contained a three-dimensional fast convolutional neural network based on R-CNN and a gradient enhancement machine based on the dual-path network. Its detection performance in pulmonary nodules was comparable with the experienced doctors. Rajpurkar *et al.* [18] built a pneumonia detection CheXNet with a 121-layer DenseNet, which achieved an F1 score of 0.435 on the ChestX-ray8 database [19] and its performance was higher than the radiologist statistically and significantly. Kermany *et al.* [20] used transfer learning and DCNN to identify the lesion categories in retinal images. By transfer learning, the method was also tested on the classification of pneumonia in pediatric chest X-ray. Liang *et al.* [21] proposed a dense network with relative location awareness for thorax disease identification. They used U-Net to segmentation and introduced the location information into the network. After the location of the disease was combined with the incidence, the method achieved the area under the curve (AUC) of 0.820. Besides, He *et al.* [22] proposed a Mask R-CNN method, which integrated three-branch (segmentation, regression and classification) to achieve the segmentation and detection. Mask R-CNN as a baseline used to other segmentation, detection or both task, eg. Chathurika and Wijesinghe [23] proposed an approach based on Mask-RCNN to automate the detection and segmentation of ulcers. The position of the lesion is also crucial to the doctor. However, these methods focus on the classification and few studies care on the detection of lesions. Moreover, these methods are using conventional convolution units. Because of CNN's limitation of rectangular sampling, the model based-CNN is limited to learn the fixed geometric structure when learning the transformation of feature space of lesions. In the chest X-ray image, e.g., bronchial type, the typical image changes are irregular infiltration around the bronchi, and its cross-section and longitudinal section feature an outward-extending image. The interstitial type shows a nearly reticular texture on the image, and the pulmonary angiogram disappears almost. Alveolar type shows ''air bronchus characteristics'' and ''air alveolar characteristics'' in the chest X-ray. Lung nodule shows circular plaques of varying sizes and uneven density, and so on. Therefore, it is difficult to extract the deformation features, refined and reinforced features by the traditional CNN. Those factors lead to lower performance of the lesion detection method with traditional CNN.

For the above problems, we propose the deformation and refined features based lesion detection on the chest X-ray algorithm. It mainly contains dynamic convolution and balance loss function. The method can get the deformation and refined features, and it reinforces the attributes of the lesion feature. Moreover, the feature fusion links the global features and local features of the lesion, and the rich feature information can be obtained. Furthermore, in training, the improved loss function suppresses the imbalance problem between the difficult and easy sample in the multi-classification lesion detection task. Those strategies can improve the accuracy of lesion detection in chest X-rays.

In this paper, the major contributions of our work are as follows:

1) *Reinforced deformation features*. In the proposed method, the residual module is designed by the convolution factorization and the amplitude modulation deformable convolution module (AMDCN). By this module, the sub-network in our framework is built to extract the deformation features of the lesion. It reinforces the attributes of the lesion feature.
2) *Refined features*. The global features and local features are connected by feature fusion, i.e., the detail features in lower layers and semantic features in high layers are combined to enrich the feature space of the lesion.
3) *Deformable pooling layer*. To solve the alignment problem in detection, and introduce the lesion location information for the pooling layer, the pooling layer combines with the AMDCN and region proposal

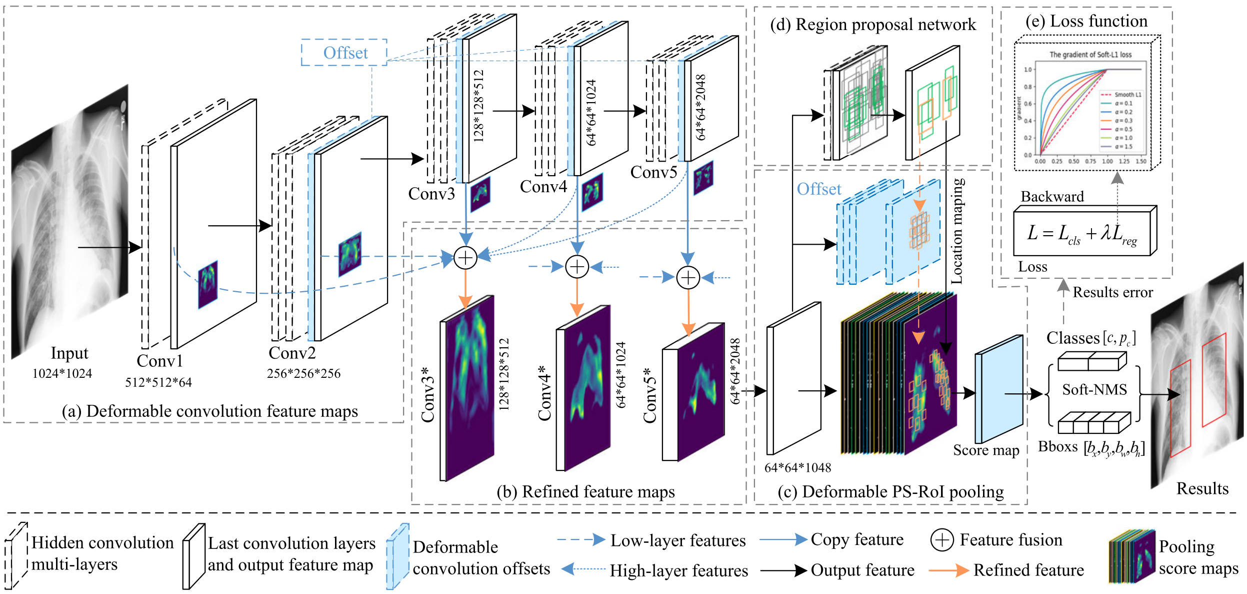


FIGURE 1. The framework of the proposed method. (a) The sub-network extracts deformation features by the deformable convolution with amplitude modulation layers. (b) The special feature fusion structure obtains the refined features. (c) The pooling layer in lesion predictor is built by the deformable convolution with amplitude modulation and the RoI location information from RPN. (d) The region proposal network [24](RPN). (e) The improved loss function optimizes the model in training.

network (RPN) to build the position-sensitive RoI pooling layer.

4) *Reliable loss function*. The imbalance between difficult and easy samples leads to the imbalance of gradient propagation. To avoid the training difficulty, which is caused by the imbalance back-propagation, the model is optimized by the improved regression loss function with a gradient control factor.

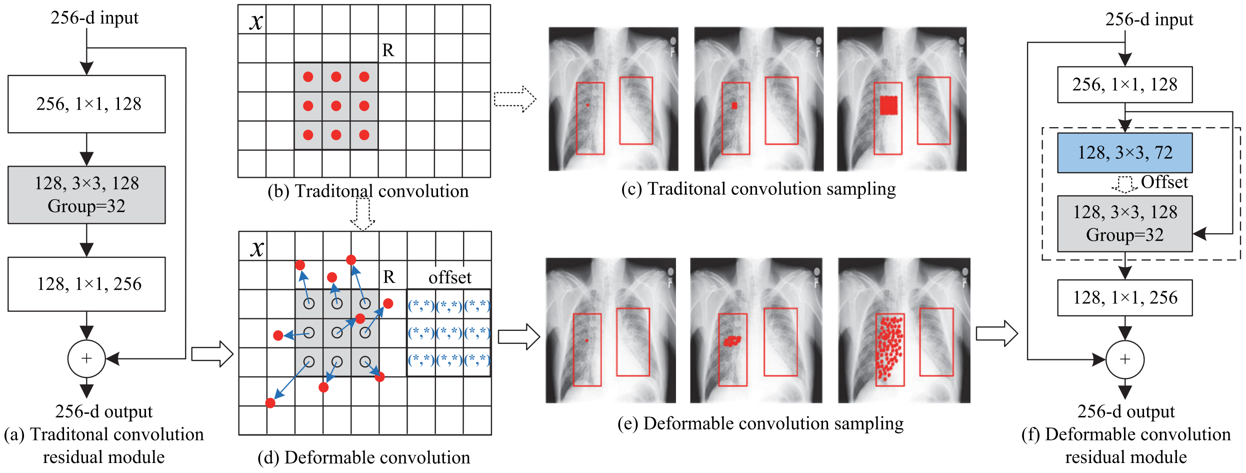
The structure of the paper is as follows. First, the framework of the proposed method is introduced. Second, the framework is further described in detail by the deformation feature extraction method, the refined features extraction method, the deformable position-sensitive RoI pooling layer, and the improved loss function. Third, the proposed method compares with the mainstream method on the public dataset, and also make the ablation test to verify the effectiveness and robustness of the proposed method. Final, we discuss and summarize the advantages, disadvantages, and improvement direction of the proposed method.

II. FRAMEWORK OVERVIEW AND MATERIALS

In this section, we introduce the deformation and refined features based lesion detection on the chest X-ray (DRCXNet) in detail. It mainly includes five parts. (i). framework overview. (ii). deformation feature based deformable convolution. (iii). refined features and deformable pooling layer. (iv). optimized prediction and loss function. (v). dataset.

A. FRAMEWORK OVERVIEW

The framework of the proposed method consists of five modules as shown in Fig. 1. The feature space is enriched by extracting deformation features and refined features, thereby improving the accuracy of lesion detection. Moreover, the model parameters are optimized by the improved loss function. The operating mechanism between modules is shown in Fig. 1(a)-(e). Firstly, in order to obtain the deformation features and reinforced the feature characterization, the deformable CNN sub-network is established for extracting features, as shown in Fig. 1(a). The better result is achieved by the deeper CNN theoretically. However, the deep network also brings the gradient dispersion or gradient explosion, which leads to the network system cannot converge [25]. Although batch normalization can alleviate these issues and make the network deeper, the batch normalization is added excessively, which leads to network degradation, accuracy reduction, or training saturation. Therefore, to avoid the negative effect of depth and batch normalization, the particular residual units is established by convolution factorization and residual method, which make the output change of the network layer to be more sensitive and highlight the slight adjustment. To obtain the deformation feature of lesion, the amplitude-modulated deformable convolution is introduced to improve the feature extractor, which avoids the limitation of rectangular sampling in traditional convolution. The blue blocks are offsets of the deformable convolution layers, which are attached to the traditional convolution. This method makes convolution more free and flexible. Secondly, in order to obtain the refined features and enrich feature space, a special feature fusion structure is established, as shown in Fig. 1(b). As the number of convolutional layers increases, the detail features (such as texture features, shapes, etc.) are transformed into semantic features in the



(a) Traditonal convolution residual module

(b) Traditonal convolution

(c) Traditonal convolution sampling

(d) Deformable convolution

(e) Deformable convolution sampling

(f) Deformable convolution residual module

FIGURE 2. Traditional convolution residual module transforms to deformable convolution residual module. (a) The residual block is built by the traditional convolution. (b) The kernel shape of the traditional convolution. (c) The sampling effect is calculated through three layers of the traditional convolution. (d) The kernel shape of the deformable convolution with amplitude modulation. (e) The sampling effect is calculated through three layers. (f) The residual block is built by the deformable convolution with amplitude modulation.

high dimensional feature space, and the detection operations are performed on high dimensional features. Because of the deeper CNN networks misses the detail feature of lesion easily, the feature fusion layer is built by the Gaussian delocalization algorithm to overcome this issue, i.e., the refined features are obtained by connecting the high-level semantic features and the low-level details globally from the feature extractor. Thirdly, the region proposal network (RPN) [24], a candidate bounding box generation network, is used in the proposed method to obtain the location information of the region of interest (RoI). Fourthly, to build a better lesion predictor, we establish the deformable position-sensitive RoI pooling layer in lesion predictor through the deformable convolution with amplitude modulation and the location information of RoI, as shown in Fig. 1(c) and (d). Because of the traditional pooling process is similar to the convolution process, it reduces the feature space resolution under a fixed scale. To avoid repeated convolution operations and reduce computational complexity, the features of inputting the pooling layer are obtained by once convolution and dimension reduction process on the refined features. Moreover, the lower spatial resolution features make to align between the prediction bounding boxes and the label boxes difficultly, i.e., the prediction bounding boxes are too big or too small. Therefore, the module (c) and (d) suppress this problem by introducing the location information and the deformable ability in the proposed method simultaneously. At last, for optimal model parameters of the proposed method, the Soft-NMS is applied in the proposed method to screen for more accurate results. Then, the prediction error is calculated by the improved loss function to optimize the model.

B. DEFORMATION FEATURE BASED DEFORMABLE CONVOLUTION

In chest X-ray, the lesion is characterized by insignificant features, diverse lesions, and variable scales. Among them, the contour is blurred due to the similarity of the lesion characteristics to the background tissue characteristics. It is difficult for a general model to extract key information from these lesion characteristics. Due to the increased depth of the convolutional network, its feature extraction capability is greatly degraded after being negatively affected by depth [26], In order to fully learn the characteristics of the lesion, the feature extraction sub-network of the proposed method references to [26]. In the proposed method, the residual method and convolution factorization are built to establish residual block, as shown in Fig. 2(a), i.e., the volume integral group calculation is implemented by the convolution factor decomposition, and the feed-forward link is fitted to map the residual feature. At the same time, due to the limitations of traditional convolution rectangle sampling, it results in traditional CNN being limited to fixed geometries when learning the spatial transformation of lesion features. Besides, since the lesion label is also a rectangular bounding box, the background noise will be introduced into the feature space while learning by traditional CNN. Under the influence of the above factors, the feature representation ability will be weakened by traditional CNN, as shown in Fig. 2 (b) and (c). In order to overcome this limitation of traditional convolution, the dilated convolution [27] and the dynamic convolution [28], [29] are designed by the special sampling of the convolution kernel, which improves the feature spatial resolution and rich features effectively. However, for the spatial representation of lesion feature, there is still a problem that the convolution structure design is single, and the feature extraction effect is not good. Therefore, the learning ability of spatial geometric deformation is introduced into the convolution in the proposed method by referring to the deformation mechanism [29]. The intensity of the deformation is controlled by the amplitude modulation mechanism, so that the convolutional model can obtain more features of the lesion and suppress the background noise. In this way, the proposed method can obtain the enriched deformation features, thereby suppressing the influence of background noise on the network, as shown in Fig. 2(d) and (e). Therefore, the offsets $\{\Delta p_k | k = 1, 2, 3 \ldots N, N = |R|\}$, which are a two-dimensional fraction, are attached to the traditional convolutional kernel. Then, the use of amplitude modulation Δm_k controls the sampling points of the convolutional kernel to select the foreground and background. The deformable convolution with amplitude modulation module (AMDCN) can be got, as in

$$y(p) = \sum_{p_k \in R} \omega(p_k) \cdot x(p + p_k + \Delta p_k) \cdot \Delta m_k, \tag{1}$$

where $y(p)$ denotes the output feature of convolution calculation at a sampling point p on the input feature map. p_k is the sampling point of the convolution kernel, R denotes the receptive field size of the convolution kernel. Therefore, the established residual module with AMDCN is shown in Fig. 2(f). Considering the computational efficiency of the model and referencing to the ResNeXt-101 [26] structure, the group convolution is used to build the backbone of the proposed method. The group convolution can decrease the computation complexity and obtain the model with fewer parameters. At the 101st layer of backbone, it takes a 1 × 1 × 1024 fully convolution for dimensionality reduction. All convolutional blocks are reconstructed by a topology of convolution factorization. Then those blocks are connected in parallel according to the concept of group in the proposed method. The capacity of the group is limited by a cardinality, which also makes the network design more convenient. Therefore, the feature extraction sub-network with AMDCN is established to extract the deformation features. Besides, the feature set $\{C_l | l \in [2, 3, 4, 5]\}$ is copied from the output of convolutional layers: Conv2, Conv3, Conv4, and Conv5, which are used as input to the feature fusion module.

C. REFINED FEATURES AND DEFORMABLE POOLING LAYER

In the convolution process of CNN, the output neurons are only related to the small portion of input image or feature map, which is determined by the operational characteristics of the convolution kernel. To solve this problem, the proposed method fuses high-level semantic features and low-level details to enrich the feature space, which can improve the prediction accuracy of the network for the target of the lesion. The strategy is to fuse the twice sampling features of the C_{l+1} layer and the C_l layer. Then multi-scale features of the lesion are obtained by analogy. However, this way is the kind of adjacent feature space information supplement, and it often dilutes and ignores the information of non-adjacent feature layers. In order to solve the above problem, the refined feature fusion is proposed in the proposed method, as shown in Fig. 1 (b). The lateral output features are obtained from some network layers of the proposed method, and the scale of the output feature is reshaped to fit the scale of the target feature fusion layer. Combined with the Gaussian non-localized feature fusion mechanism, the weighted feature fusion is performed according to the correlation between the features. Therefore, the refined features are obtained, as shown in

$$\phi_t = \frac{1}{N_l} \sum_{l_{\min}}^{l_{\max}} f(C_l, C_t \times s_l(C_l | C_t)) \times r_t(C_l), \tag{2}$$

where ϕ_t denotes the refined feature with the same dimension generated on the l-th feature space C_l. N_l is the number of feature layers of the fusion. $f(C_l, C_t)$ denotes the correlation coefficient matrix of the l-th and t-th layer feature, which is calculated by Embedded Gaussian. $r_t(C_l)$ denotes the feature of the input feature C_l after reshape. $s_l(C_l | C_t)$ denotes the scale factor of the l-th layer feature C_l and the t-th layer feature C_t. Then, the refined features are entered into the pooling layer and fully convolution layer, which can obtain preliminary prediction results of the lesions in the X-ray image.

Because of the deepening of the number of layers of the network, the invariance of the network model to the target's translation and rotation is stronger. This property is of positive significance to ensure the robustness of the lesion detection model. However, in the problem of lesion detection, the layers of the network are over-deepen to lead a significant reduction in the network's ability to perceive the lesion location. Therefore, to solve this problem, the proposed method uses the RPN to introduce the lesion position information into the pooling process and adopts the same structure as the deformable convolution to construct the deformable position-sensitive RoI pooling layer with amplitude modulation, as shown in Fig. 1 (c) and (d). In this way, the quantization operation in the feature aggregation process is canceled during the pooling process. The pixel coordinates of the float number are obtained by using the bilinear interpolation method, which causes the feature aggregation process to be converted into a continuous amount of operations. Therefore, the deformable position-sensitive RoI pooling response model can be obtained by referring to (1), as shown in (3). First, the RoI region generated by the RPN is mapped to the feature map and the offset receptive field, which is generated by the fully convolution layer. The RoI area is divided into $m \times m$ bins (m is an adjustable parameter, the default setting is 7). Therefore, the RoI pooling offset $\{\Delta p_{ij} | 0 \leq i, j \leq m\}$ of the (i, j)-th bin can be generated by a fully convolutional layer. The output of the (i, j)-th bin in the model is calculated from its feature score map.

$$y_c(p_k | \Theta) = \sum_{p_{ij} \in bin(i,j)} x(p_k + p_{ij} + \Delta p_{ij} | \Theta) \cdot \Delta m_{ij} / n_k, \tag{3}$$

where p_k denotes the upper left corner bin of the k-th RoI feature. p_{ij} denotes the (i, j)-th bin in the RoI. Δm_{ij} denotes the amplitude modulation, which can suppress the diffusion phenomenon of the sample beyond the region of interest. Therefore, the sampling of the region of interest is concentrated around the lesion. In the actual calculation, the features of $m \times m$ bins are normalized by the fully convolution layer to obtain $m \times m$ offsets of $\Delta \hat{p}_{ij}$. However, the size of the RoI is inconsistent, and the height and width of the input feature map are also inconsistent, these offsets cannot be directly used. Therefore, it is corrected by a gain γ (default γ is 0.1), and by dot (w, h) with $\gamma \Delta \hat{p}_{ij}$, the truth value is $\Delta p_{ij} = \gamma \Delta \hat{p}_{ij} \cdot (w, h)$. After the pooling layers and post-processing layers, the preliminary prediction results are obtained for the position information of the lesions in the image. Then, the prediction error is used by loss function to optimize the model of the proposed method.

D. OPTIMIZED PREDICTION AND LOSS FUNCTION

Since the preliminary prediction results contain many noise targets, the results are needed to optimize the bounding boxes of the lesion and remove the repeated boxes. However, the non-maximum suppression method (NMS) uses the highest confidence coefficient, the confidence level of the adjacent lesion bounding boxes are forced to 0. Therefore, the missed detection occurs for the lesion with large regional overlap, resulting in the detected area of the lesion is incomplete, such as pneumonia and pulmonary nodules simultaneously. In order to solve such problems, the proposed method uses Soft-NMS to improve the NMS using linear weighting and establishes a soft non-maximum suppression model, as in (4).

$$s_i = \begin{cases} s_i, & IoU \ < N_t \\ s_i(1 - IoU(b^*, b_i)), & IoU \geq N_t \end{cases}, \tag{4}$$

where s_i is the probability on predicted lesion, N_t is the threshold of suppression, and $IoU(b^*, b_i)$ denotes intersection over union (IoU) between the predicted lesion bounding boxes b_i and the ground-truth b^*. Using Soft-NMS to optimize the preliminary prediction results can reduce the missed rate of lesions and increase the detection rate of lesions.

In the training, the loss function makes the predicted value to gradually approach the true value, the loss reaches a minimum when the predicted value is equal to the true value. In order to realize the end-to-end and efficient learning of the proposed method, the proposed method establishes the loss function, as in (5) and (6).

$$S_c(\Theta) = \frac{e^{y_c(\Theta)}}{\sum\limits_{c_i=0}^{c} e^{y_{c_i}(\Theta)}}, \tag{5}$$

$$L(S_c, b_{(x,y,w,h)}|\Theta) = L_{cls}(S_c, S_{c^*}) + \lambda[c^*]L_{reg}(b, b^*), \tag{6}$$

where S_c is the predicted classification response. c denotes the predicted lesion category ($c = 0$ is the background) and c^* indicates the labeled value. $y_c(\Theta)$ is the response of the deformable position-sensitive RoI pooling. Θ denotes all learned parameters of the proposed method. $b_{(x,y,w,h)}$ is the predicted lesion bounding box and b^* is the ground-truth value. L_{cls} denotes the cross entropy loss function of classification and $L_{cls}(S_c, S_{c^*}) = -\ln_{cls}(S_c|S_{c^*})$. λ is the weighted average parameter of the loss L. L_{reg} denotes the Soft-L1 regression loss function of bounding box. In the training, the proposed method uses the OHEM [30] method to mine difficult samples. However, the method is sensitive to noise samples, and a large gradient is generated for difficult samples, resulting in simple samples are ignored, i.e., the small gradient is submerged. The proposed method improves the regression loss function Smooth-L1 to overcome this problem. The loss function is properly modulated on the gradient response of the sample to balance the training process, i.e., when the sample error response value is in the $|\hat{x}_{(b,b^*)}| \leq 1$ range, a slightly larger gradient value is generated. Therefore, the Smooth-L1 loss function is remodeled into the Soft-L1 boundary regression loss function model, as in (7).

$$L_{reg}(b, b^*) = \begin{cases} \dfrac{\alpha}{\beta}(\hat{\beta})\ln(\hat{\beta}) - \alpha|\hat{x}|, & if\,|\hat{x}| \leq 1 \\ \ln(\beta + 1)|\hat{x}| + C, & others, \end{cases} \tag{7}$$

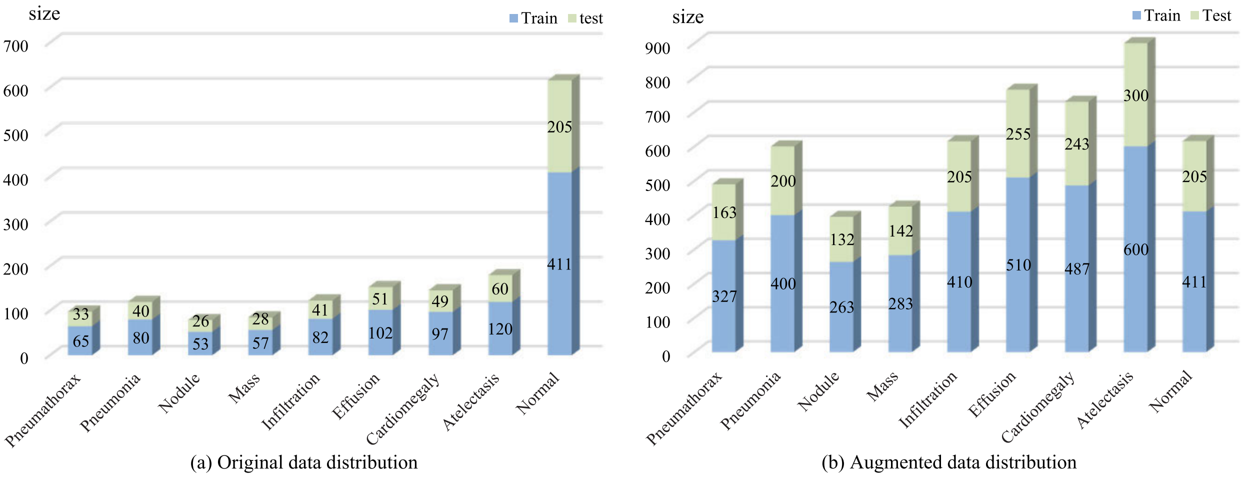
where, $\hat{\beta} = \beta|\hat{x}_{(b,b^*)}| + 1$. $\hat{x}$ represents the regression error value of (b, b^*). α represents the gradient control factor and default value is 0.5. β represents the upper bound factor of the adjustment regression error, the value is $e^{2/3} - 1$ and. C indicates that the connection factor ensures that L_{reg} is continuously steerable, the value is $1 - 0.75\beta^{-1}$.

In order to optimize the loss function, the weight and bias parameters need to be constantly updated by the back-propagation. Because the deformable convolution is used to improve the proposed method, the gradient of the offset in the back-propagation of the proposed method is calculated as in (8) and (9).

$$\begin{aligned} &\frac{\partial y(p)}{\partial \Delta p_k \partial \Delta m_k} \\ &= \sum_{p_k \in R} \omega(p_k) \cdot \frac{\partial x(p + p_k + \Delta p_k) \cdot \Delta m_k}{\partial \Delta p_k \partial \Delta m_k} \\ &= \sum_{p_k \in R} \left[\omega(p_k) \cdot \frac{\partial G(q, p + p_k + \Delta p_k) \cdot \Delta m_k}{\partial \Delta p_k \partial \Delta m_k} x(q) \right], \end{aligned} \tag{8}$$

$$\begin{aligned} &\frac{\partial y(p_k)}{\partial \Delta p_{ij} \partial \Delta m_{ij}} \\ &= \frac{1}{n_k} \sum_{p_{ij} \in bin(i,j)} \frac{\partial x(p_k + p_{ij} + \Delta p_{ij}) \cdot \Delta m_{ij}}{\partial \Delta p_{ij} \partial \Delta m_{ij}} \\ &= \frac{1}{n_k} \sum_{p_{i,j} \in bin(i,j)} \sum_{q} \left[\frac{\partial G(q, p_k + p_{ij} + \Delta p_{ij}) \cdot \Delta m_{ij}}{\partial \Delta p_{ij} \partial \Delta m_{ij}} x(q) \right], \end{aligned} \tag{9}$$

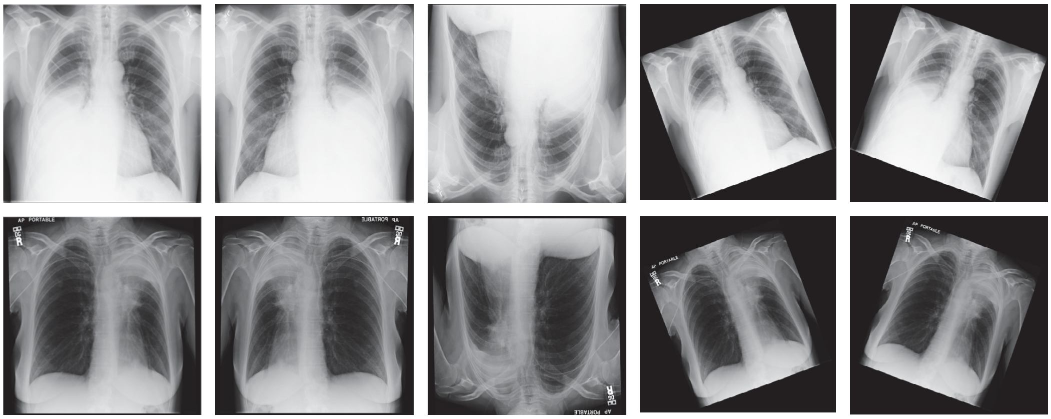
where Δp_k and Δp_{ij} are both a two-dimensional vector, then $\partial \Delta p_k$ denotes $\partial \Delta p_k^x$ and $\partial \Delta p_k^y$, and Δp_{ij} denotes $\partial \Delta p_{ij}^x$ and $\partial \Delta p_{ij}^y$. G represents a bilinear difference function. Because G is a two-dimensional function, the proposed method simplify it by dimensionality reduction, $\hat{p} = p + p_k + \Delta p_k$ denotes an arbitrary fractional position, and q enumerate any position in the feature map x. For the convenience of calculation, $G(q, \hat{p}) = g(q_x, \hat{p}_x) \cdot g(q_y, \hat{p}_y)$ can map the offset to the x-axis and y-axis.



(a) Original data distribution

(b) Augmented data distribution

FIGURE 3. The statistical distribution of lesions. (a) indicates the statistical distribution of lesions in BBP dataset. (b) indicates the statistical distribution of lesion after the preprocessed BBP_x dataset.



Original

horizontal flip

diagonal flip

contrarotate

clockwise

FIGURE 4. The ways of preprocessing data are horizontal flip, diagonal flip, contrarotate and clockwise rotation in BBP dataset.

E. DATASETS

In this paper, the proposed method is verified by two public chest X-ray image datasets. The detailed introduction of the datasets is as follows.

RSNA Pneumonia dataset [31]. Radiological Society of North America (RSNA) collaborated with the US National Institutes of Health, Society of Thoracic Radiology, and MD.ai to develop a rich dataset as the RSNA pneumonia dataset. Radiologists reviewed all pneumonia data labels for chest X-ray images and confirmed by clinical history, vital signs, and radiology. RSNA dataset contains 29,684 DICOM files with a resolution of 26,684 training dataset and 3,000 test dataset. The annotation information includes patient ID, lesion category, and lesion position.

ChestX-ray8 dataset [19]. Dr. Lule's team develops this dataset. ChestX-ray8 dataset contains 32,727 cases of patients, a total of 108,948 frontal view X-ray images. Since the category information is obtained from the NLP [19] method, which is linked to one or more keywords for each image radiographic report, the data have not the position information of each lesion. For obtaining the accurate position information of lesion, radiologists manually annotate a part of the dataset to form the bounding boxes for pathologies (BBP) dataset. The BBP dataset contains 983 images, and each of the eight classifications consisted of 200 case lesions, with a total of 1,600 case lesions. In the experiment, the multi-classification performance of the proposed method is tested by the BBP dataset.

Data preprocessing. In the BBP dataset, the use of expended data avoids overfitting on the training model of the proposed method. The numbers of each category images are divided into the training data and testing data according to 2:1. The training dataset is 1,067, and the testing dataset is 533, as shown in Fig. 3(a), showing the statistical distribution of lesions of type 8+1 (1 is the normal case). Then, the BBP dataset is preprocessed through the three ways, which are horizontal flip, diagonal flip, and rotation of ±20°, as shown in Fig. 4. Such operations keep the features of lesion structure and texture consistent with the original data and ensure that the images have predictive invariance of scaling, rotation, and translation. The preprocessed data is named BBP_x data, among which the training set is 3,691, and the test data set is 1,845. The statistical distribution of the preprocessed data, as shown in Fig. 3(b).

F. EVALUATION

The proposed method, which is the particular design framework, can get the classification and position of the lesion simultaneously. Moreover, II-D indicates the classification accuracy and localization accuracy shadow each other. The comprehensive evaluation score is used to illustrate the performance of the proposed method, which is calculated by the accuracy of classification and localization. nTP is the number of correct lesions detected, nFP is the number of wrong lesions detected, and nFN is the number of missed lesions detected. Sensitivity S, Precision P, and Consistency D are calculated respectively, as in (10) and (12).

$$S = nTP/(nTP + nFN) \tag{10}$$

$$P = nTP/(nTP + nFP) \tag{11}$$

$$D = \frac{1}{N}\sum_{k=1}^{N} DSC \tag{12}$$

where DSC (Dice similarity coefficient) is the consistency coefficient between the predicted banding box (BBox) and the real labeling BBox. DSC is calculated by $DSC = 2 \times |A_k \cap B_k| \big/ (|A_k| + |B_k|)$. A_k denotes the bounding boxes of labeling lesions, and B_k denotes the bounding boxes of testing lesion. As in (13), the comprehensive evaluation score is equal to the average weighted sum of $F1$ score and consistency D, where $F1 = 2 \times S \times P/(P + S)$. The higher the CE value is, the better the detection performance of the model.

$$CE = 0.5 \times F1 + 0.5 \times D \tag{13}$$

III. RESULTS

In the experiment, we designed quantitative experiments and qualitative experiments. The experiments of all methods work on the same experimental platform. The graphic card configuration of the platform is GTX2080 8G and CUDA 10.1. The proposed method is achieved by Python 3.6, OpenCV-4.1, MXNet-1.5, Numpy-1.11, Cython-0.25, and so on. Experimental results show that the lesion detection performance of the proposed method is consistent on both systems of windows10 and Ubuntu 16.04. The model can predict the lesions of 5-6 images per second. The detailed parameters used for training the model follow: Epochs (15); learning rate (0.5) and step (4.83); Warm up for learning rate (0.0005) and step (4000); Frequent (100); Image batch size (ctx*2) and scales (1024, 1024); Momentum (0.9); Optimizer (Nadam); RCNN feat stride (16); Shuffle image (true); Anchor scales (8, 16, 32) and ratios (0.5, 1, 2); ROIs batch size (128); BBOX regression thresh (0.5); RPN Batch size (256); RPN Positive overlap (0.7); NMS thresh (0.7).

A. QUALITATIVE RESULTS

In the results of qualitative experiments, the proposed method compares with some popular detection methods. These comparison methods mainly include YOLOv3 [32], Faster R-CNN [24], R-FCN [33], Mask R-CNN [22], CheXNet [18], and Kaggle-rsna18 [31] models. The Kaggle-rsna18 model comes from the champion free-source model in 2018 Kaggle's RSNA competition. The v1 of the model indicates that it's structure is based on the RetinaNet [34], which is a feature pyramid framework. At the same time, the model uses the Focal loss function to optimize the problem of extremely unbalanced positive and negative samples in the lesion detection task. Faster R-CNN(D) and R-FCN(D) are using the deformable convolution module. Mask R-CNN* are using the detection function without the segmentation. In the test data, the results of the partial images are visualized, as shown in Fig. 5. The bottom in the figure shows the ground-truth (GT) of the X-ray image, and the lesion is masked the green bounding box. The prediction results are marked by a red bounding box to indicate the lesion location and area. At the same time, the category and probability of lesion prediction are marked on the upper left of the BBox. In the visualization results, it shows that there are serious misdetection lesions on ChexNet and above other models. For small target lesions, such as Nodule lesions, the proposed method and Kaggle-RSNA are detecting it, but misdiagnosis of lesions also occurred. We analyzed some of the test results and found that each model performed poorly on the location of small lesions. However, the proposed method has more accurate detection results than the comparison method. Therefore, the results show that the detection effect of DRCXNet is better than the comparison method. The qualitative results prove that the proposed method has a higher performance to lesion detection in chest X-ray.

B. QUANTITATIVE RESULTS

In the quantitative experiments, the lesion detection performance of the method is analyzed by calculating the CE, P, S, and D values of the prediction results. The ability of the method to locate the lesion can be also analyzed by calculating the D value. The quantitative experimental results of the proposed method and the popular depth learning detection algorithms are shown in Table 1. Among these comparison methods, YOLOv3 [32] uses up-sampling and feature fusion to get multi-scale lesion detection. Its structure and feature pyramid are very similar. Faster R-CNN [24] and R-FCN [33] are divided into two stages. First, the residual units are used to extract features. Second, the prediction of classification and position is performed by the features. CheXNet [18] is a 121-layer DenseNet structure that captures the details of the lesion. Kaggle-rsna18 v1 model is based on the RetinaNet [34] structure, which is a feature pyramid model [37]. It also uses the focal loss function to optimize the difficult training problem, which is caused by a sample imbalance in the dataset. Kaggle-rsna18 v2 uses a deformable convolution [29] without amplitude modulation. Faster R-CNN (D)

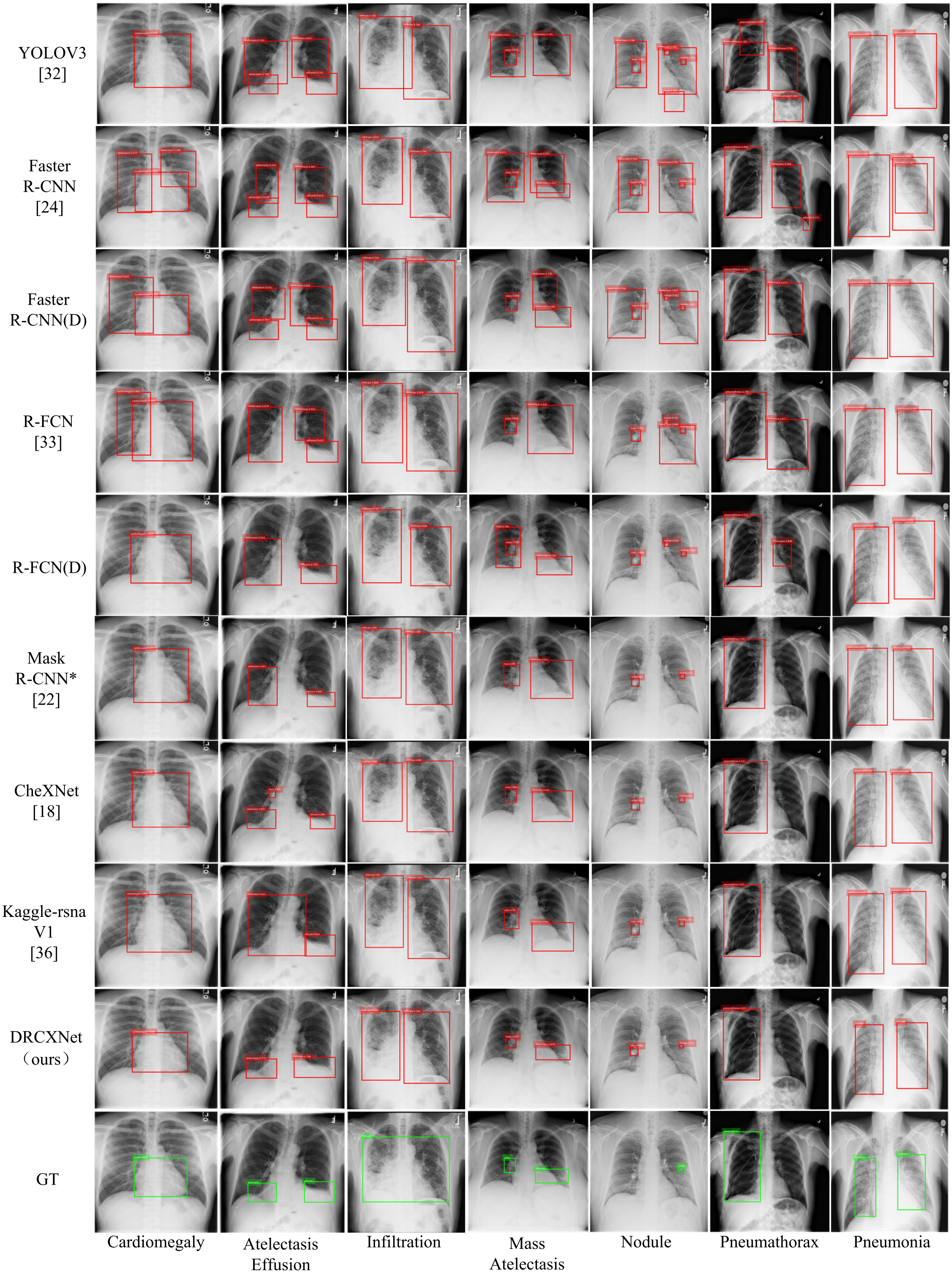


FIGURE 5. The qualitative results of the proposed method and comparison methods. Faster R-CNN (D) and R-FCN (D) indicate that those models use the deformable convolution with amplitude modulation.

TABLE 1. The experimental results of the quantitative between the proposed method and comparison methods on RSNA and BBP_x datasets. Faster R-CNN (D) and R-FCN (D) indicate that those models use the deformable convolution with amplitude modulation.

Method	backbone	RSNA test dataset				BBP_X test dataset				Speed
		CE	D	S	P	CE	D	S	P	pic/s
YOLOv3 [32]	Darknet53	0.806	0.788	0.862	0.789	0.773	0.746	0.831	0.772	**0.089**
Faster R-CNN [24]	ResNet-101	0.716	0.689	0.776	0.711	0.687	0.651	0.746	0.702	0.189
Faster R-CNN (D)	ResNet-101	0.738	0.711	0.798	0.732	0.714	0.694	0.777	0.696	0.195
R-FCN [33]	ResNet-101	0.789	0.773	0.848	0.766	0.754	0.728	0.814	0.750	0.180
R-FCN (D)	ResNet-101	0.830	0.834	0.890	0.772	0.796	0.765	0.849	0.805	0.185
Mask-RCNN (*)	ResNet-101	0.832	0.834	0.891	0.776	0.789	0.759	0.843	0.797	0.177
CheXNet [18]	DenseNet-121	0.845	0.846	0.901	0.793	0.817	0.784	0.866	0.834	0.232
kaggle-rsna-v2 [35]	ResNet-101	0.819	0.811	0.876	0.782	0.780	0.774	0.843	0.735	0.182
kaggle-rsna-v1 [36]	ResNet-101	0.836	0.824	0.889	0.808	0.812	**0.814**	0.874	0.756	0.190
DRCXNet(Ours)	ResNeXt-101	**0.866**	**0.859**	**0.914**	**0.836**	**0.841**	0.792	**0.879**	**0.900**	0.201

TABLE 2. The experimental results of the number of layers of deformable convolution with amplitude modulation in the proposed method, e.g., AMDCN-5/4 indicates that Conv5 and Conv4 of the proposed method are the deformable convolutional layers.

Method	RSNA test dataset				BBP_X test dataset				Speed
	CE	D	S	P	CE	D	S	P	pic/s
Null	0.835	0.828	0.890	0.799	0.810	0.772	0.858	0.838	**0.184**
AMDCN-5	0.841	0.828	0.892	0.819	0.817	0.770	0.861	0.865	0.187
AMDCN-5/4	0.857	0.853	0.908	0.817	0.829	0.787	0.872	0.869	0.191
AMDCN-5/4/3	0.862	0.856	0.911	0.829	0.835	0.788	0.876	0.889	0.196
AMDCN-5/4/3/2	**0.866**	**0.859**	**0.914**	**0.836**	**0.841**	**0.792**	**0.879**	**0.900**	0.201

and R-FCN (D) are using deformable convolution with amplitude modulation, which can help the model improve feature extraction capabilities. In the experiment, all model is tested on the BBP_x and RSNA datasets and calculate S, P, and D, and used them to calculate the comprehensive evaluation. As shown in Table. 1, the proposed method shows the outperformance compared with the comparison methods. Moreover, all the indicators of evaluation on the BBP_x dataset are slightly lower than the value on the RSNA dataset. The first reason, there is only pneumonia type in RSNA, and it has a large amount of dataset than BBP_x data. The second reason, these lesions have large differences in shape scale and characteristics, which makes the model shows different detection effects on multi-type lesion detection, such as small lesion is more difficult detection. At the same time, on the BBP_x dataset, YOLOv3, CheXNet, and Kaggle-rsna-v1, they have a higher recall ratio for small target lesions. The reason is that those methods adopt a feature fusion strategy similar to the pyramid model, which can be more scale information is introduced into the model. Besides, the proposed method is more computationally complex, and it takes more time to test one image. However, in medical image analysis, the accuracy and *CE* of lesion detection are more important than the slight difference in computational efficiency.

C. ABLATION STUDY

The ablation experiments of the proposed method are designed by four interesting parts. (i) The effect of different deformable convolution. (ii) The effect of different strategies of refined feature fusion. (iii) The effect under the NMS and Soft-NMS. (iv) The effect of loss function on model training.

1) DEFORMABLE CONVOLUTION WITH AMPLITUDE MODULATION

The two experiments prove that the deformable convolution with amplitude modulation is very helpful to the proposed method, as shown in Table. 2.

a: THE NUMBER OF LAYERS

With the increase in the number of deformable convolution layers, the detection performance of the proposed method is gradually improving. The performance improvement of the proposed method in the BBP_X dataset is slightly better than that on the RSNA dataset. This phenomenon also proves that the deformable convolution is sensitive to the deformation features and scale features of the lesion. The deformable convolution can effectively promote the feature extraction sub-network of the proposed method to learn the rich lesion feature space. As shown in Fig. 6, it visualizes the output feature maps of the different layers in the feature extraction sub-network. This feature visualization further proves the view that as the deformable convolution layer increases, the response of the proposed method to the lesion is also gradually increased, which is presented as a clear high hot spot on the feature map. However, the time of prediction single image is also increasing with the more number of deformable convolutional layers.

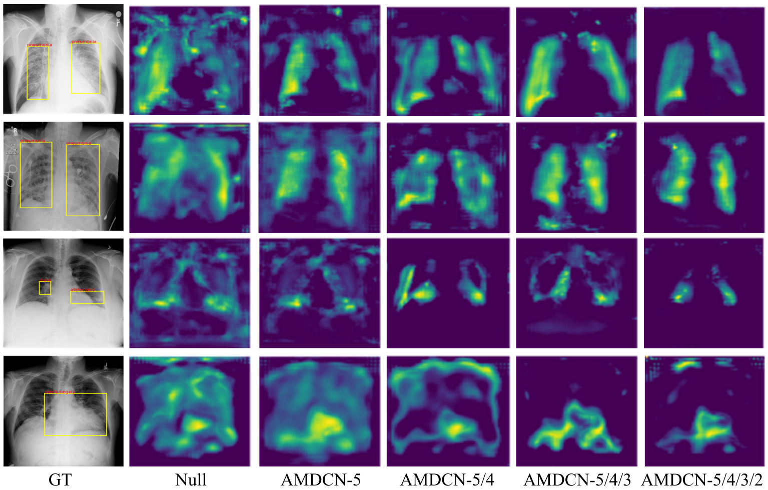


FIGURE 6. The visualization output feature maps of the different layers with the amplitude modulation deformable convolution.

TABLE 3. The experimental results of the deformable convolution with amplitude modulation (AMDCN) and without it (DCN). The (n) is the dilation of deformable convolution and n = 1, 2, 3.

Method	RSNA test dataset				BBP_X test dataset			
	CE	D	S	P	CE	D	S	P
DCN(1)	0.839	0.837	0.895	0.792	0.812	0.787	0.864	0.810
DCN(2)	0.848	0.848	0.903	0.797	0.824	0.791	0.872	0.844
DCN(3)	0.834	0.832	0.891	0.787	0.816	0.802	0.872	0.793
AMDCN(1)	0.858	0.852	0.908	0.824	0.830	0.797	0.876	0.851
AMDCN(2)	**0.866**	**0.859**	**0.914**	**0.836**	**0.841**	0.792	**0.879**	**0.900**
AMDCN(3)	0.852	0.843	0.902	0.822	0.835	**0.810**	0.883	0.840

b: AMPLITUDE MODULATION AND DIFFERENT DILATIONS

The influence of amplitude modulation is verified on the deformable convolution. DCN indicates the deformable convolution without amplitude modulation. AMDCN indicates the deformable convolution with amplitude modulation. In the experiment, the dilation of the deformable convolution is 1, 2, and 3, respectively. As shown in Table. 3, the deformable convolution with amplitude modulation can effectively improve the detection of the proposed method. When the dilation is 2, the DCN and AMDCN have better lesion detection effects, while AMDCN(2) performs better. As shown in Fig. 8, it visualizes the output features of the conv5 in feature extraction sub-network. It shows that AMDCN(2) is relatively outperformance for the feature response of the lesion. The experimental results show that AMDCN can promote the network to better learn the features of the lesion and suppress the background noise.

2) THE REFINED FEATURE FUSION

The experimental results of the different fusion strategies and deformable pooling layer are shown in Table. 4.

a: DIFFERENT FUSION STRATEGIES

The proposed method considers the influence of global features and local features. Four feature refinement strategies are T = 5, T = 4/5, T = 3/4/5, T = 2/3/4/5 for comparison verification. When T is 3/4/5, the proposed method can obtain the best lesion detection effect. Because the more feature layers fused in the network, the more low-level details are introduced into the network, but it also brings in a lot of noise features. Therefore, an appropriate T is obtained to balance the performance of our high-level semantic features and the underlying detail features. As shown in Fig. 8, it visualizes the refined features of different T strategies. The different T strategies have different refined features. The refined feature response performs better as T = 3/4/5. Therefore, the global features combine with local features to enrich the feature space, which enables the network to obtain more detailed feature information.

b: DEFORMABLE POOLING LAYER

As shown in Table. 4, the experimental results are the position-sensitive RoI pooling layer with AMDCN and without it in the last two rows. The D increment in this experimental result is relatively high, which further indicates that the deformable position-sensitive pooling layer improves the regression of lesion location. Furthermore, these experimental results demonstrate that the deformable convolution can improve the performance of the proposed method.

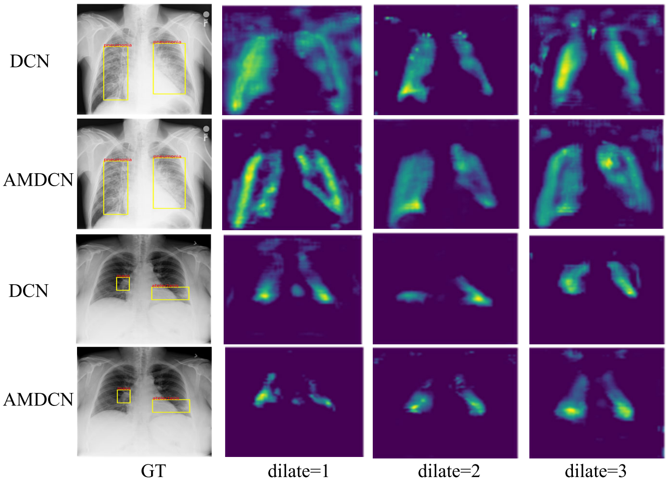


FIGURE 7. The visualization output features of the conv5 in the proposed framework with different dilation.

TABLE 4. The experimental results of refined feature fusion under strategy T, e.g. T = 4/5 indicates that Conv4 and Conv5 are the base-layer for refined feature fusion respectively. In last two rows, the experimental results are the position-sensitive RoI pooling layer with AMDCN and without it.

Method	RSNA test dataset				BBP_X test dataset			
	CE	D	S	P	CE	D	S	P
Null	0.833	0.840	0.894	0.769	0.813	0.781	0.863	0.828
T=5	0.842	0.840	0.897	0.795	0.820	0.779	0.866	0.859
T=4/5	0.851	0.853	0.906	0.798	0.829	0.790	0.873	0.861
T=3/4/5	**0.866**	**0.859**	**0.914**	**0.836**	0.841	**0.792**	**0.879**	**0.900**
T=2/3/4/5	0.857	0.850	0.907	0.824	0.835	0.790	0.876	0.884
PS RoI pooling	0.849	0.841	**0.916**	0.821	0.822	0.774	**0.913**	0.877
AMDCN pooling	**0.866**	**0.859**	0.914	**0.836**	**0.841**	**0.792**	0.879	**0.90**

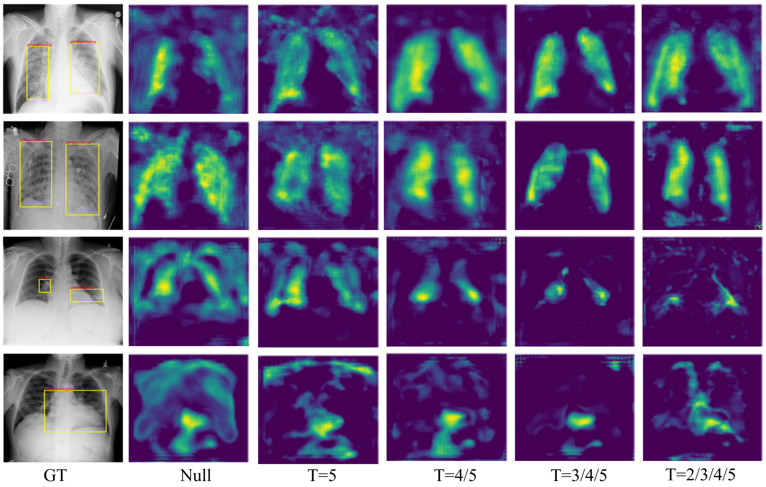


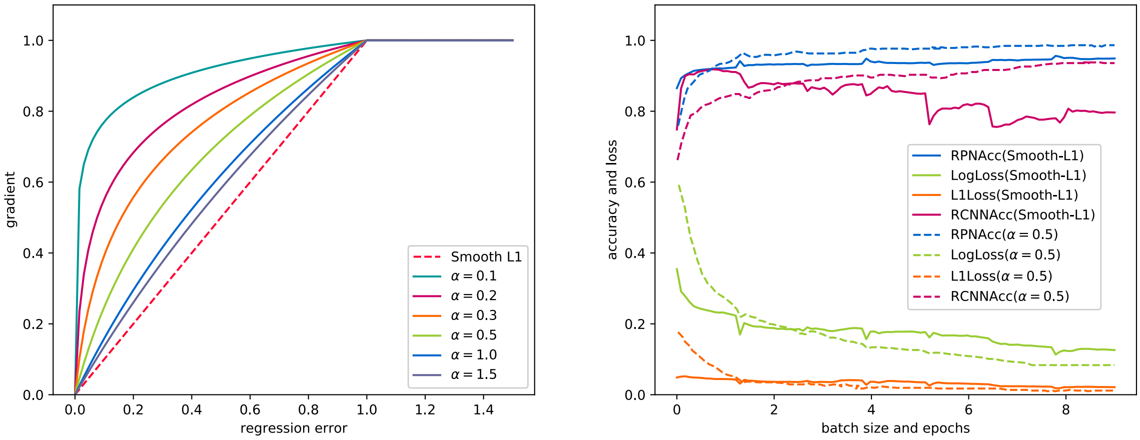
FIGURE 8. The visualization refined features of different T strategies.

3) OPTIMIZATION OF THE PREDICTION RESULTS

We compare the predictive performance of the proposed method by import NMS and soft-NMS. As shown in Table. 5, the experimental results are verified by the impact of the two optimization methods on the model prediction results. It shows that the IoU threshold has a massive impact on the experimental results. On both datasets, our method is a higher performance on lesion detection when IoU is 0.5. In the case

TABLE 5. The experimental results of the proposed method with NMS or soft-NMS on RSNA and BBP_x datasets.

Method	RSNA test dataset				BBP_X test dataset			
	CE	D	S	P	CE	D	S	P
NMS(IoU=0.5)	0.849	0.837	0.886	**0.837**	0.828	0.776	**0.905**	0.858
Soft-NMS(IoU=0.5)	**0.866**	**0.859**	**0.914**	0.836	**0.841**	**0.792**	0.879	**0.900**
NMS(IoU=0.7)	0.757	0.722	0.813	0.771	0.726	0.671	0.780	0.781
Soft-NMS(IoU=0.7)	0.781	0.757	0.839	0.776	0.745	0.700	0.800	0.781



(a) The gradient of Soft-L1 loss

(b) The accuracy and loss in training model

FIGURE 9. The experimental results under the Smooth-L1 and Soft-L1 loss functions. (a) indicates the gradient curves of error include the Smooth-L1 and the different gradient control factors of the Soft-L1. (b) indicates the accuracy curves and loss curves under the Smooth-L1 and Soft-L1 ($\alpha = 0.5$) in training.

of the equivalent IoU, it shows that the Soft-NMS are better than NMS to optimize the proposed method.

4) LOSS FUNCTION

In training, the loss function is used as the connection between the model and the optimization problem. It optimizes our model to achieve the best state through iteration. In order to restrain gradient response imbalance cased the sample differences, the Smooth-L1 regression loss is improved in the proposed method. The loss function properties before and after improvement are shown in Fig. 9(a). The different gradient control factor α is adjustable, which balances the gradient's response strength of the difficult and easy samples during the learning process, and it can make the model get optimal lesion detection performance. In the experiment, some parameters of the proposed method are fixed, such as the learning rate, the number of iterations, and other parameters are set the same. Then, the models with the Smooth-L1 and Soft-L1 loss functions are validated. Their loss error and detection accuracy are recorded during the training process, as shown in Fig. 9(b). At this time, α is an experience value of 0.5. It can be seen from the experimental results that Soft-L1 loss can effectively improve the training accuracy of the model and reduce the loss error of the model.

IV. DISCUSSION

Because of the chest lesions show blurred boundary contours, different sizes, variable shapes, and uneven density, it results in poor performance of most methods for detecting lesions, e.g., pneumonia usually manifests as an area or areas of increased opacity on chest X-ray. However, the diagnosis of pneumonia on chest X-ray is the complex reasoning problem, which often requires careful observation and knowledge of anatomical principles, physiology, and pathology. In this paper, we analyzed the merit and demerit of some methods about deep convolutional neural networks, and some simple characteristics of chest X-Ray images. Then, we propose the deformation and refined features based lesion detection on the chest X-ray algorithm, which can improve the detection accuracy of lung lesions. Moreover, the proposed method can be extended to other tasks of lesion detection. e.g., the deformable convolution with amplitude modulation can be applied to other learning tasks, which can enhance the ability of the model to extract features. However, we also find some points worth pondering in the analysis and verification of the proposed method.

The proposed method achieves better lesion detection performance on two public datasets. Therefore, it also indicates that the proposed method is a better medical image analysis model. However, medical image analysis is a very rigorous task, which needs to be combined with actual clinical applications. Therefore, the model is needed to optimize through continuous exact measurement and analysis by the clinical applications. The proposed method is an end-to-end learning model. However, some hyper-parameter is inserted in the proposed method, e.g., T in the refined feature fusion module, and α in the loss function. For other tasks, it is unknown whether these hyper-parameters allow the proposed method to achieve better detection performance.

The pooling operation in the proposed method is assigned a learning offset, while the pooling process is only the mean pooling or the maximum pooling. It reduces the dimensionality of the feature space roughly. If the pooling operations are learned, we can use a convolution-like process to make the most appropriate pooling output based on the characteristics of the input features. In the study, obtaining large-scale medical image datasets is very difficult and expensive. The transfer learning method provides some theoretical support for the few-shot learning tasks [20], [38]. By searching for common feature spaces or feature hidden spaces between different lesion detection tasks, transfer learning can transform the learned knowledge to new lesion detection tasks. Therefore, the transfer learning method can be applied to the few-shot learning tasks, which can further enhance the generalization of the model. e.g., Muhammed *et al.* [39] proposed an approach that uses deep transfer learning to classify normal and abnormal brain MR images automatically. They achieve higher classification accuracy on MR images by the finder of the optimal learning rate and fine-tuning to train the model. In the study, some lesions have a higher probability of morbidity at a certain location in the organ. One lesion maybe causes another lesion [40]. Therefore, this correlation information can be used to enhance the predictive power of the lesion detection model, e.g., Liang *et al.* [21] used location information of the lesion in the lung to improve the accuracy of the detection of lesions. In the future works, the structure information and some diagnostic information can be used in lesion detection to reinforce reliability and improve the accuracy. We can also use the graph convolution network model [41] or ensemble model [42] to build a better performance fine-grained lesion analysis system. Therefore, there are many ways to improve the model, which requires us to continue to explore the best method for different tasks.

V. CONCLUSION

This paper proposes the deformation and refined features based lesion detection on the chest X-ray algorithm. The model can extract the enhanced deformation features by a deformable convolution with amplitude modulation in the residual module. Furthermore, the enrich feature space of lesions is obtained from the global features and local features by the feature fusion. The lesion position sensitivity of the model is improved by the deformable RoI pooling layer. In training, to avoid the training difficulty caused by the imbalance problem between the difficult and easy sample and the severe imbalance of gradient propagation, the regression loss function is redesigned by gradient control factor, which can balance the feedback gradient in the learning process. In the experiment, the proposed method is evaluated on the RSNA and ChestX-ray8 datasets, The proposed method compared with outstanding methods has a significant performance in both qualitative and quantitative experiments. The experimental results show that the proposed method can reduce false positive and false negative of detection lesions. In the future work, with combining the requirements in medical image analysis, we will continue to optimize the proposed method by the transfer learning and graph network theory.

REFERENCES

[1] A. Daoud, A. Laktineh, S. El Zein, and A. O. Soubani, "Unusual presentation of primary lung adenocarcinoma mimicking pneumonia: Case report and literature review," *Respiratory Med. Case Rep.*, vol. 28, Jun. 2019, Art. no. 100881.

[2] C. Qin, D. Yao, Y. Shi, and Z. Song, "Computer-aided detection in chest radiography based on artificial intelligence: A survey," *Biomed. Eng. Online*, vol. 17, pp. 133–155, Dec. 2018.

[3] A. Rodriguez-Ruiz, K. Lang, A. Gubern-Merida, M. Broeders, G. Gennaro, P. Clauser, T. H. Helbich, M. Chevalier, T. Tan, T. Mertelmeier, M. G. Wallis, I. Andersson, S. Zackrisson, R. M. Mann, and I. Sechopoulos, "Stand–alone artificial intelligence for breast cancer detection in mammography: Comparison with 101 radiologists," *J. Nat. Cancer Inst.*, vol. 111, no. 9, pp. 916–922, Sep. 2019.

[4] U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger, "X-ray categorization and retrieval on the organ and pathology level, using patch–based visual words," *IEEE Trans. Med. Imag.*, vol. 30, no. 3, pp. 733–746, Mar. 2011.

[5] S. Jaeger, A. Karargyris, S. Candemir, L. Folio, J. Siegelman, F. Callaghan, Z. Xue, K. Palaniappan, R. K. Singh, S. Antani, G. Thoma, Y.-X. Wang, P.-X. Lu, and C. J. Mcdonald, "Automatic tuberculosis screening using chest radiographs," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 233–245, Feb. 2014.

[6] P. Pattrapisetwong and W. Chiracharit, "Automatic lung segmentation in chest radiographs using shadow filter and multilevel thresholding," in *Proc. Int. Comput. Sci. Eng. Conf. (ICSEC)*, Manchester, U.K., Dec. 2016, pp. 1–6.

[7] A. Criminisi, "Machine learning for medical images analysis," *Med. Image Anal.*, vol. 33, pp. 91–93, Oct. 2016.

[8] M. Bergtholdt, R. Wiemker, and T. Klinder, "Pulmonary nodule detection using a cascaded SVM classifier," *Proc. SPIE, Med. Imag., Comput.-Aided Diagnosis*, vol. 9785, pp. 1–13, Mar. 2016.

[9] S. Alam, M. Kang, J.-Y. Pyun, and G.-R. Kwon, "Performance of classification based on PCA, linear SVM, and multi-kernel SVM," in *Proc. IEEE 8th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Vienna, Austria, Aug. 2016, pp. 987–989.

[10] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology," *Nature Rev. Cancer*, vol. 18, pp. 500–510, May 2018.

[11] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," *Nature Med.*, vol. 25, no. 1, pp. 44–56, Jan. 2019.

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

[13] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018.

[14] P. Pattrapisetwong and W. Chiracharit, "Chest pathology detection using deep learning with non-medical training," in *Proc. IEEE Int. Comput. Sci. Eng. Conf. (ICSEC)*, Chiang Mai, Thailand, Dec. 2016, pp. 294–297.

[15] Y. Anavi, I. Kogan, E. Gelbart, O. Geva, and H. Greenspan, "A comparative study for chest radiograph image retrieval using binary texture and deep learning classification," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Milan, Italy, Nov. 2015, pp. 2940–2943.

[16] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.

[17] W. Zhu, C. Liu, W. Fan, and X. Xie, "Deeplung: Deep 3D dual path nets for automated pulmonary nodule detection and classification," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Lake Tahoe, NV, USA, May 2018, pp. 673–681.

[18] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "Chexnet: Radiologist-level pneumonia detection on chest X-rays with deep learning," Dec. 2017, *arXiv:1711.05225*. [Online]. Available: https://arxiv.org/abs/1711.05225

[19] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recongnit.*, Honolulu, HI, USA, Dec. 2017, pp. 2097–2106.

[20] D. S. Kermany *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, pp. 1122–1131, Feb. 2018.

[21] X. Liang, C. Peng, B. Qiu, and B. Li, "Dense networks with relative location awareness for thorax disease identification," *Med. Phys.*, vol. 46, pp. 2064–2073, May 2019.

[22] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2961–2969.

[23] L. Chathurika and I. S. Wijesinghe, "Instance-based segmentation for boundary detection of neuropathic ulcers through mask-R CNN," in *Proc. Int. Conf. Artif. Neural Netw. Mach. Learn. (ICANN)*, Munich, Germany, Sep. 2019, pp. 511–522.

[24] 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.

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

[26] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recongnit.*, Honolulu, HI, USA, Dec. 2017, pp. 1492–1500.

[27] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recongnit.*, Honolulu, HI, USA, Dec. 2017, pp. 472–480.

[28] Y. Jeon and J. Kim, "Active convolution: Learning the shape of convolution for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recongnit.*, Honolulu, HI, USA, Dec. 2017, pp. 4201–4209.

[29] 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)*, Venice, Italy, Oct. 2017, pp. 764–773.

[30] R. G. Abhinav Shrivastava and A. Gupta, "Training region-based object detectors with online hard example mining," in *Proc. IEEE Conf. Comput. Vis. Pattern Recongnit.*, Las Vegas, NV, USA, Jun. 2016, pp. 761–769.

[31] (Feb. 2019). *RSNA Pneumonia Detection Challenge. Radiological Society of North America and Kaggle's Machine Learning Community*. [Online]. Available: https://www.kaggle.com/c/rsna-pneumonia-detection-challenge

[32] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "YOLOv3: An incremental improvement," Dec. 2017, *arXiv:1804.02767*. [Online]. Available: https://arxiv.org/abs/1804.02767

[33] D. Jifeng, L. Yi, H. Kaiming, and S. Jian, "R-FCN: Object detection via region-based fully convolutional networks," in *Proc. Adv. Neural. Inf. Process. Syst.*, Barcelona, Spain, Dec. 2016, pp. 379–387.

[34] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Dec. 2017, pp. 2980–2988.

[35] *Kaggle-RSNA18 v2:DeformableConvNets*. Accessed: Mar. 10, 2019. [Online]. Available: https://github.com/i-pan/kaggle-rsna18/tree/master/models/DeformableConvNets

[36] *Kaggle-RSNA18 v1:RetinaNet*. Accessed: Mar. 10, 2019. [Online]. Available: https://github.com/i-pan/kaggle-rsna18/tree/master/models/RetinaNet

[37] T. 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 Recongnit.*, Honolulu, HI, USA, Dec. 2017, pp. 2117–2125.

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

[39] M. Talo, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Application of deep transfer learning for automated brain abnormality classification using mr images," *Cogn. Syst. Res.*, vol. 54, pp. 176–188, May 2019.

[40] G. Harerimana, B. Jang, J. W. Kim, and H. K. Park, "Health big data analytics: A technology survey," *IEEE Access*, vol. 6, pp. 65661–65678, 2018.

[41] K. Yan, X. Wang, L. Lu, L. Zhang, A. P. Harrison, M. Bagheri, and R. M. Summers, "Deep lesion graphs in the wild: Relationship learning and organization of significant radiology image findings in a diverse large-scale lesion database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recongnit.*, Salt Lake City, UT, USA, Dec. 2018, pp. 9261–9270.

[42] S. Qummar, F. G. Khan, S. Shah, A. Khan, S. Shamshirband, Z. U. Rehman, I. Ahmed Khan, and W. Jadoon, "A deep learning ensemble approach for diabetic retinopathy detection," *IEEE Access*, vol. 7, pp. 150530–150539, 2019.

CE LI received the Ph.D. degree in pattern recognition and intelligence system from Xi'an Jiaotong University, China, in 2013. He is currently a Professor with the College of Electrical and Information Engineering, Lanzhou University of Technology. His research interests include computer vision and pattern recognition.

DONG ZHANG received the B.S. degree in electronic information science and technology and M.S. degree in pattern recognition and intelligent system from the Lanzhou University of Technology, China, in 2014 and 2019, respectively. His research interests include computer vision, machine learning, and medical image analysis.

SHAOYI DU (Member, IEEE) received the B.S. degrees in computational mathematics and in computer science, the M.S. degree in applied mathematics, and the Ph.D. degree in pattern recognition and intelligence system from Xi'an Jiaotong University, China in 2002, 2005, and 2009, respectively. He worked as a Postdoctoral Fellow in Xi'an Jiaotong University, from 2009 to 2011, and visited the University of North Carolina at Chapel Hill, from 2013 to 2014. He is currently a Professor with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University. His research interests include computer vision, machine learning, and pattern recognition.

ZHIQIANG TIAN received the B.S. degree in automation control from Northeastern University, in 2004, and the M.S. and Ph.D. degrees in control science and engineering from Xi'an Jiaotong University, in 2007 and 2013, respectively. He was a Postdoctoral Fellow with the Department of Radiology and Imaging Sciences, Emory University, from 2014 to 2017. He is currently an Associate Professor with Xi'an Jiaotong University. His research interests are image/video processing, computer vision, multimedia, and medical image analysis.