

Received 10 June 2022, accepted 9 July 2022, date of publication 18 July 2022, date of current version 21 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3192034

## RESEARCH ARTICLE

# STBi-YOLO: A Real-Time Object Detection Method for Lung Nodule Recognition

KEHONG LIU<sup>ID</sup>

College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an, Shaanxi 710054, China

e-mail: 1292692089@qq.com

This work was supported in part by the Science and Technology Plan Project of Shaanxi under Grant 2017JM6105, in part by the Ministry of Education Collaborative Education Project with Seclover Corporation, and in part by the Ministry of Education Collaborative Education Project with HUAWEI Corporation.

**ABSTRACT** Lung cancer is the most prevalent and deadly oncological disease in the world, but a timely detection of lung nodules can greatly improve the survival rate of this disease. However, due to the tiny size of lung nodules and inconspicuous edges, lung nodules are not easily distinguished by naked eyes thus medical image diagnosticians are prone to misdiagnosis simply based on their own experiences and subjective judgements. In recent years, the machine-learning-based image processing techniques find their wide applications in the field of medical diagnosis, and have been proved to be an efficient way to aid diagnosticians to accurately identify subtle lesions in images. To accurately recognize lung nodules in CT images, in this paper, we propose an approach, called STBi-YOLO. This approach stems from YOLO-v5, but makes significant improvements from three dimensions—we first use the spatial pyramid pooling network that is based on stochastic-pooling method to modify the basic network structure of YOLO-v5; then apply a bidirectional feature pyramid network to perform multi-scale feature fusion; finally improve the loss function of the YOLO-v5 and adopt the EIoU function to optimize the training model. To evaluate our approach, we compare STBi-YOLO with YOLO-v3, YOLO-v4, YOLO-v5, and multiple leading object detection models, such as Faster R-CNN and SSD. The experiments show that STBi-YOLO achieves an accuracy of 96.1% and a recall rate of 93.3% for the detection of lung nodules, while producing a 4× smaller model size in memory consumption than YOLO-v5 and exhibiting comparable results in terms of mAP and time cost against Faster R-CNN and SSD.

**INDEX TERMS** Lung nodules, object detection, YOLO-v5, bidirectional feature pyramid, stochastic-pooling.

## I. INTRODUCTION

Global cancer statistics for 2020 shows that lung cancer is the most prevalent and deadly oncological disease in worldwide for many years [1]. Clinical studies have shown that the survival rate within five years of the late-stage patients are only between 10% and 16%, but could be increased to 52% if early diagnosis and treatment were provided. As an important sign of early-stage lung cancer, lung nodules mostly appear as focal, round-like lung shadows of no more than 3cm in diameter on CT images. However, due to the tiny size of lung

nodules, their morphology, brightness and other characteristics are similar to those of blood vessels and other tissues in the lung parenchyma, thus doctors need to carefully consider and screen them one by one; this process is inefficient and easily leads to fatigue, increasing the probability of misdiagnosis. Therefore, it is important to develop an automatic detection approach to aid physicians to improve the efficiency and accuracy for detecting lung nodules.

The approaches for detecting lung nodules can be roughly classified into two categories: the traditional *segmentation-based* detection techniques and the *deep-learning-based* detection techniques. The segmentation-based methods mainly use manually extracted features for training, which

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei<sup>ID</sup>.

suffer from the weakness such as cumbersome steps, low accuracy and poor overall performance. With the rapid development of machine learning techniques, the deep learning techniques are widely used in object detection. One kind of these work, such as Faster R-CNN [2] and Mask R-CNN [3], focuses on the *two-stage* detection algorithm based on candidate regions; the other is the *one-stage* detection algorithm based on regression, such as the YOLO algorithms [4]–[6] and SSD [7]. Among them, YOLO-v5, as the latest version of the YOLO algorithms, is used successfully in engineering applications and brings excellent results. However, when applying the YOLO-v5 algorithm to lung CT images, it performs with poor detection accuracy and low processing speed; this prevents it from being workable in the detection of lung nodules.

To solve this problem, we propose a novel approach, called STBi-YOLO, originated from YOLO-v5s, to optimize the detection performance for small targets. We make three significant improvements for YOLO-v5. Firstly, we introduce a multi-scale convolutional layer based on stochastic-pooling in the original Spatial Pyramid Pooling (SPP) Network [8] to improve the recognition accuracy; then apply a Bi-directional Feature Pyramid Network (BiFPN) for multi-scale feature fusions to improve the fusion effect and thus promote the detection capability for small targets; finally we use the EIou loss function to optimize the trained model.

Note that, the name “STBi-YOLO” was coined as an acronym for *Stochastic-pooling-based spatial pyramid pooling network* and *Bidirectional feature pyramid network*, which highlight our improvements for the original YOLO algorithms.

The main contributions of this paper are as follows:

- (1) use stochastic-pooling method to replace max-pooling of YOLO-v5 in SPP Network;
- (2) apply a bidirectional feature pyramid network to perform multi-scale feature fusion;
- (3) improve the loss function of YOLO-v5 and adopt the EIou function to optimize the training model.

This paper is organized as follows. Section 2 introduces related work about lung nodule identification and object detection. In Section 3, we propose STBi-YOLO, and introduce its overall structure. We mainly focus on the three improvements over the YOLO-v5 algorithm: the stochastic-pooling-based multi-convolutional layer SPP-Net, the improved FPN network, and the EIou loss function. Section 4 presents experimental results and discussions, in which the model size, real-time performance measured by Frames Per Second (FPS), recall, and mean Average Precision (mAP) are given. Moreover, we compare STBi-YOLO with other state-of-the-art models in this section. Finally, the conclusion is drawn in Section 5.

## II. RELATED WORK

### A. THE DETECTION OF LUNG NODULES

The detection of lung nodules usually consists of two parts: the first is the detection of candidate lung nodules and the

second is to reduce the false-positive lung nodules. In recent years, many solutions have been proposed, which can be generally divided into traditional detection methods, machine-learning-based methods and deep-learning-based methods.

The traditional computer-aided detection methods mainly use manually extracted features for training to identify whether a patient has lung nodules. Due to the limited computing power of GPU, these methods require a certain amount of manual intervention and user’s assistance. The main weakness of the methods is the poor generalization ability, which makes it difficult to achieve a multi-category, data-intensive, and real-time accurate recognition in practical situations.

The representative works for the traditional computer-aided detection methods are Scale-Invariant Feature Transform (SIFT) [9] and Histogram of Oriented Gradients (HOG) [10]; The problem-solving process of these methods can be roughly summarized as the following three steps:

- (1) region selection, which is mostly based on the sliding window approach;
- (2) feature extraction, which is to design extraction algorithm according to target color and texture;
- (3) classification recognition, which mainly applies Support Vector Machine (SVM) [11] or AdaBoost [12].

In machine-learning-based approaches, researchers combine classification models with advanced features to detect nodules [13]–[17]. For example, Khordehchi *et al.* designed a set of spectral, textural, and shape features to characterize nodules and then used SVM to classify candidate nodules [18]. Nithila and Kumar developed a Computer-Aided Detection (CAD) system for detached lung nodule detections that focused on heuristic search algorithms and particle clustering algorithms for network optimization [19].

Instead of manually designing the advanced features, the deep-learning-based methods apply the deep learning techniques to automatically extract typically features from large amount of labelled data and accurately recognize lung nodules in CT images. These methods commonly used the two-dimensional convolutional neural network (2D-CNN) and three-dimensional convolutional neural network (3D-CNN) as the training models. For example, Setio *et al.* [20] and Lanfredi *et al.* [21] proposed a multi-view-based 2D-CNN and applied it to the detection of lung nodules. They divided this network structure into two parts: the first part consists of three detectors to detect and identify suspicious candidate nodules; the second part consists of two 2D-CNNs—the first 2D-CNN contains a convolutional layer and a maximum pooling layer, and another one contains a full convolutional layer and a softmax layer.

3D-CNN [22], [23] is an improvement of 2D-CNN, which can better acquire the spatial data images and extract much more information-rich features than 2D-CNN. Hamidian *et al.* [24] proposed a computer-aided diagnosis system based on 3D-CNN. The lung nodule detection process is divided into 2 steps: screening and identification. Firstly, a 3D-FCN [25] turns the fully connected layer in the 3D-CNN into a convolutional layer which outputs marked

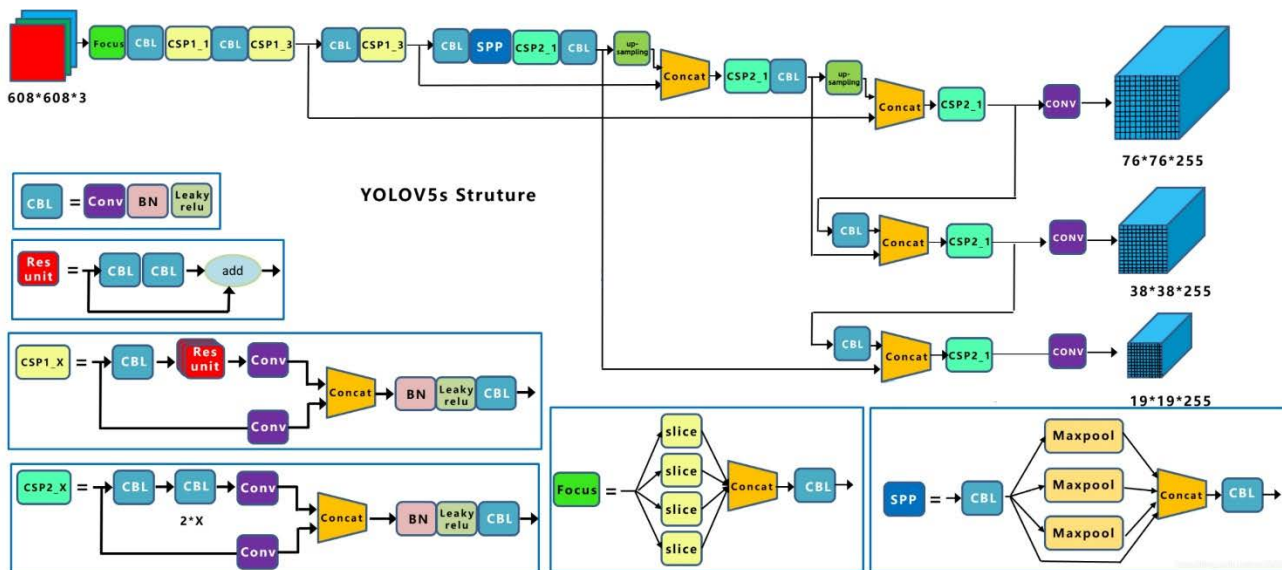


FIGURE 1. Network structure of YOLO-v5s.

images instead of vectors. Then, since the 3D-FCN can accept images of arbitrary size, up-sampling is performed by using the deconvolution layer to restore the image to its original size. Finally, a 3D-FCN is used to mark the candidate regions of lung nodules and apply CNN to classify the suspected lung nodules in the candidate regions to determine the true nodules.

**B. OBJECT DETECTION**

The most typical algorithms for object detection with deep learning techniques are the two-stage detection algorithms based on anchor boxes, and the single-stage detection algorithms based on anchor-free boxes. The former includes R-CNN [26], [27], Fast R-CNN [28], Faster R-CNN [2], [29], R-FCN [30], [31] and Mask R-CNN [3]. This kind of algorithms usually has a high accuracy but spends much time in detection. The latter includes SSD [7], Retina-Net [32] and YOLO algorithms. The single-stage algorithms use a forward inference network to obtain the target location and reach the classification results.

YOLO is a family of algorithm. The original YOLO algorithm is proposed by Yan and Xu [33], it converts an object detection problem into a regression problem by dividing the images into grids that can be used to predict the targeted objects. After then, Redmon and Farhadi [5] improved the original one with the YOLO-v2 algorithm, which used Darknet-19 as the feature extraction network and introduced the anchor box to promote the recall of algorithm. After that, Redmon and Ali [34] additionally introduce the YOLO-v3 algorithm based on the YOLO-v2 algorithm. The YOLO-v3 replaced the feature extraction network in the YOLO-v2 with the Darknet-53; this improvement and greatly upgraded the detection accuracy and accelerated the detection speed. The YOLO-v4 algorithm was proposed by

Bochevskiy *et al.* [35] on the basis of YOLO-v3 in 2020, which combines the CSPNet (Cross Stage Partial Network) with Darknet53 as the backbone named CSPDarknet53. In addition, the feature extraction network of YOLO-v4 was enhanced by SPP (Spatial Pyramid Pooling) and PANet (Path Aggregation Network). The YOLO-v5 object detection algorithm is a lightweight detection model based on the Python framework released by Ultralytics in 2020, which continues to use the CSP structure and adds it to the backbone and Neck to enhance the network feature fusion. YOLO-v5 also adds the Focus structure to the backbone network to slice the feature map, which reduces the computation burden and speeds up the procession. FIGURE 1 shows the network structure of YOLO-v5.

In what follows, we first present the overall framework of STBi-YOLO, and then focus on the three improvements one by one in much detail.

**III. STBi-YOLO**

There are four versions of YOLO-v5; they are YOLO-v5s, YOLO-v5m, YOLO-v5l and YOLO-v5x, with the weights, width, depth of each model being sequentially increased. The STBi-YOLO model proposed in this paper is based on YOLO-v5s algorithm, but makes significant improvements from the following three aspects: we first use a spatial pyramid pooling network that is based on the stochastic-pooling method to modify the basic network structure of YOLO-v5; then apply a bidirectional feature pyramid network for multi-scale feature fusion; finally improve the loss function of the YOLO-v5 and adopt the EIou function to optimize the training model.

**A. THE FRAMEWORK OF STBi-YOLO**

The overall framework of STBi-YOLO is shown in FIGURE 2. First an input image with flexible size enters the

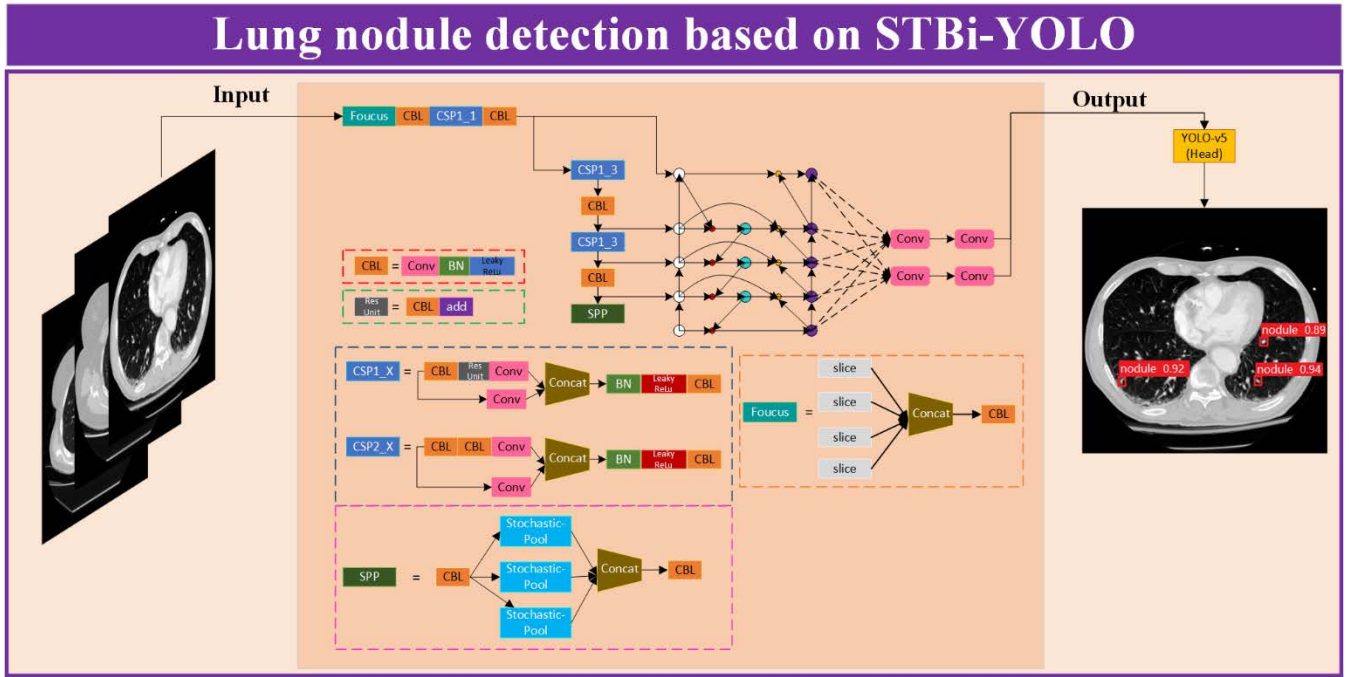


FIGURE 2. Overall structure of the algorithm.

Focus model after being unified to  $640 \times 640$  through adaptive scaling. Next, with the multi-convolutional operations, the input image is processed and enters the stochastic-pooling-based SPP-Net for image sub-sampling so as to reduce the dimension and network parameters while increasing the local receptive field of the convolutional kernel. After then, the input image goes to the BiFPN structure for feature fusion at different dimensions, in order to further reduce redundant calculations and improve detection accuracy. At the final stage, the EIou loss function is applied to suppress noise, speed up convergence and improve the robustness of the model.

**B. STOCHASTIC-POOLING-BASED SPP NETWORK**

YOLO-v5 uses Max-Pooling in the SPP structure for calculations. The purpose of pooling is to compress the information of a certain region, so as to receive the extraction and abstraction of information. The pooling can achieve the benefits of data dimensionality reduction and feature compression, as well as expand the receptive field and accomplish invariances (including translation, rotation and scale invariances). Therefore, when designing pooling operations, the loss of information in the feature mapping should be minimized on the basis of simplifying operations.

The most commonly used pooling methods are Average-Pooling [36] and Max-Pooling [37]—the former can output the mean value of feature values in a subregion and retain more background information; the latter gives the maximum value of feature values in a subregion, emphasizing the strongest part of the image. But for the case where the

difference is not obvious, the Average-Pooling and Max-Pooling are easy to cause feature information loss.

To this end, we introduce the Stochastic-Pooling [38] to balance Average-Pooling and Max-Pooling. The idea of Stochastic-Pooling is to assign a probability value to each pixel point according to its pixel value; A large value will have a higher probability of being selected. Such a design strategy is a compromise between Average-Pooling and Max-Pooling—it is similar to Average-Pooling in average cases, but still respects the rules of Maximum-Pooling in local information calculations [39], FIGURE 3 shows the definition graph of the three types of pooling.

First, we calculate the statistical sum of the pooling regions, i.e.,

$$\sum_{k=R_j} a_k$$

and then every feature value  $a_i$  is divided by this statistical sum to calculate the probability value

$$p_i = \frac{a_i}{\sum_{k \in R_j} a_k}$$

of each feature. Afterwards, we apply Random-Sampling [40] according to these probability values to achieve Stochastic-Pooling, as in (1)

$$s_j = a_l, \quad l \sim P(p_1, \dots, p_{|R_j|}) \tag{1}$$

where the  $R_j$  is the window size for sampling;  $a_i$  is the feature value being sampled;  $l$  is the value randomly

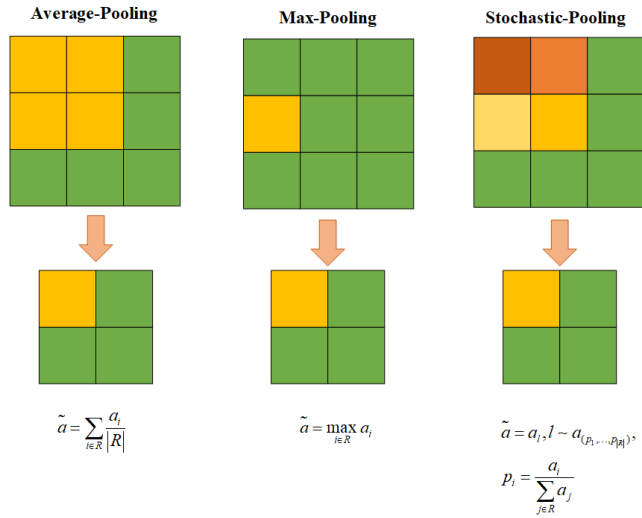


FIGURE 3. Definition graph of the algorithm.

chosen according to  $p_i$ . The improved SPP structure is shown in FIGURE 4.

### C. IMPROVED FPN WITH BiFPN

Traditionally, YOLO algorithms use FPN (Feature Pyramid Network) and PAN (Path Aggregation Network) as the network for multi-scale feature fusion. To improve the accuracy of object detection, in STBi-YOLO, we use BiFPN (Bidirectional Feature Pyramid Network) [41] for multi-scale feature fusion. BiFPN is an improved structure from PAN, which combines two directions, top-down and bottom-up, of the feature fusions together. We use the structures of FPN, PAN and BiFPN, as illustrated in FIGURE 5, to show the advantage of BiFPN.

FIGURE 5(a) shows the FPN structure, which establishes a top-down pathway for feature fusion, followed by prediction using the fused features with higher semantic information. As this structure is limited by the one-way information flow, Liu *et al.* [42] proposed the PAN structure, as shown in FIGURE 5(b), which establishes a bottom-up pathway on the basis of FPN to send the location information from the bottom layer to the prediction feature layer, therefore the prediction feature layer has both semantic information of the top layer and location information of the bottom layer, greatly improving the accuracy of target detection.

BiFPN is an improved structure from PAN, as shown in FIGURE 5(c). For bidirectional cross-scale connections, we first delete the node with only one input edge to simplify bidirectional network, because it has little contribution to feature fusion; then add an edge between the original input and the output node in order to fuse more features with less cost; finally, the top-down and bottom-up paths are fused into a module so that they can be stacked repeatedly for higher level feature fusion. For weighted feature fusion, BiFPN uses the fast normalized fusion, which achieves normalization directly by dividing the weight of each node into the sum of every node value, such that all weights are normalized

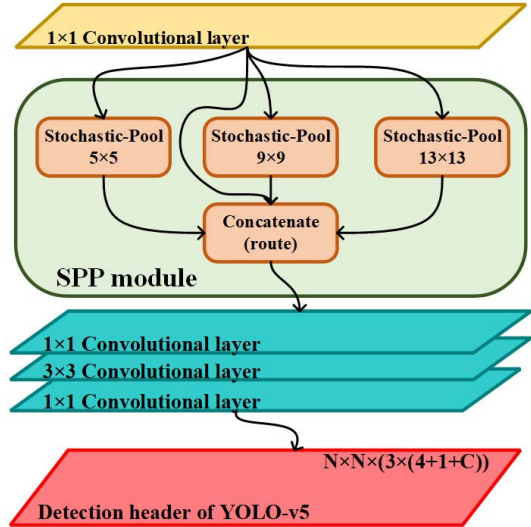


FIGURE 4. Structure of stochastic-pooling SPP.

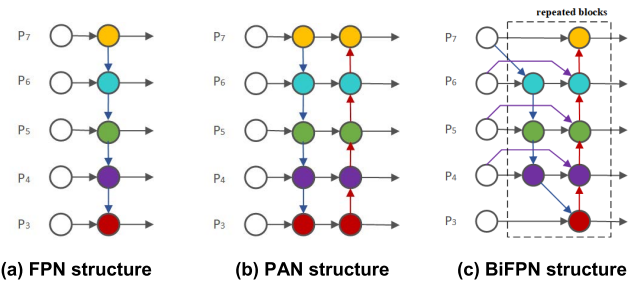


FIGURE 5. FPN, PANet and BiFPN structure.

to  $[0,1]$ , increasing the computing speed, as in (2).

$$O = \sum_i \frac{w_i}{\varepsilon + \sum_j w_j} \times I_i \quad (2)$$

The activation function ReLU is applied to ensure each  $w_i \geq 0$ . The values of  $w_i$  are obtained from network training, and  $I_i$  represent the input features. However, the values of scalar weights may be infinite and lead to training instability, we therefore apply softmax to proceed numerical normalization. To perform cross-scale connection and weighted feature fusion, BiFPN takes three different scales of features  $P_3$ ,  $P_4$  and  $P_7$ , extracted from the backbone as the input of BiFPN.

Finally,  $20 \times 20$ ,  $40 \times 40$ , and  $80 \times 80$  are set as the prediction branches of three different scaled feature resolutions. Some of the weights are  $([0.49998, 0.49998])$ ,  $([0.33332, 0.33332, 0.33332])$ ,  $([0.49998, 0.49998])$ ,  $([0.50000, 0.50000])$ ,  $([0.33325, 0.33325])$ ,  $([0.50000, 0.50000])$ . Taking the node  $P_6$  as an example, performing two fusion features is shown as follows.

$$P_6^{td} = Conv \left[ \frac{w_1 \cdot P_6 + w_2 \cdot Resize(P_7^{in})}{w_1 + w_2 + \varepsilon} \right] \quad (3)$$

$$P_6^{out} = Conv \left[ \frac{w_1' \cdot P_6^{in} + w_2' \cdot P_6^{td} + w_3' \cdot Resize(P_5^{out})}{w_1' + w_2' + w_3' + \varepsilon} \right] \quad (4)$$

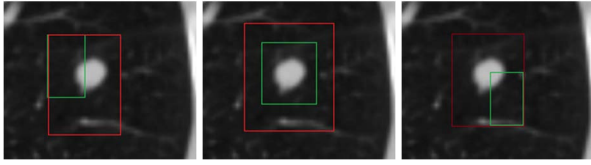


FIGURE 6. Inclusion of prediction box and ground truth box.

where  $P_6^{td}$  is the top-down middle feature (the middle blue circle in FIGURE 5(c));  $P_6^{out}$  is the bottom-up output feature (the rightmost blue circle in FIGURE 5(c)); *Resize* is the up-sampling or sub-sampling and *Conv* is the convolution operation.

**D. IMPROVEMENT OF LOSS FUNCTION**

The loss function [43] can be used to calculate the knot level of the model to the outcome prediction, and determine whether there is a bias between the model and the actual data. Therefore, the loss function is crucial in the process, choosing a proper loss function is beneficial to get a better model and faster convergence in the process of training.

The loss functions of YOLO-v5 include classification loss (cls\_loss), bounding box loss (box\_loss), and objectness loss (obj\_loss). YOLO-v5 applies the BCELogits loss function to calculate the loss of objectness score, BCELoss (binary cross entropy loss function) to calculate class probability loss, and GIoU [44] as the loss function of bounding box. GIoU loss function is shown in (5).

$$L_{GIoU} = 1 - \frac{|B \cap B_i|}{|B \cup B_i|} + \frac{|C - B \cup B_i|}{|C|} \tag{5}$$

where  $C$  is the area of the smallest minimum bounding rectangle. Because of the introduction of the smallest minimum bounding rectangle, the GIoU loss function can still find the descent gradient when the prediction box and the ground truth box do not intersect. Although the GIoU loss function can solve such situation, when one box lies in another, the union of the two boxes is equal to the area of one, then the GIoU loss function is unable to determine the location between the two frames and will occur large inaccuracy. FIGURE 6 is the situation where the prediction box (the green one) lies in the ground truth box (the red one).

In this paper, we adopt the EIou loss function [45] to replace the GIoU loss function used by YOLO-v5. The EIou loss function removes the smallest minimum bounding rectangle added by the GIoU loss function. Moreover, it applies the minimized scalar distance between the centroids of the two boxes, which calculates the length and width of the object box separately. The EIou loss function is shown in (6).

$$L_{EIou} = 1 - \frac{|B \cap B_i|}{|B \cup B_i|} + \frac{\rho^2(b, b^{gt})}{d^2} + \frac{\rho^2(w, w^{gt})}{C_w^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \tag{6}$$

TABLE 1. Model parameter settings.

Parameters	Values
weight decay	0.0005
batch size	4
learning rate	0.01
epoch	300

In (6),  $b$  and  $b^{gt}$  are the centroids of the prediction box and ground truth box respectively,  $\rho$  is the euclidean distances of the two centroids;  $d$  is the diagonal distance between the prediction box and the smallest minimum bounding rectangle of the ground truth box;  $w, w^{gt}, h$  and  $h^{gt}$  are the width and length of the real box and ground truth box;  $C_w$  and  $C_h$  are width and length of the minimum external rectangle covering the two boxes.

Since the EIou loss function calculates the length and width of the object box separately, it solves the problem of large calculation errors in the horizontal and vertical directions of the GIoU loss function, and improves the convergence speed and the accuracy of regression.

**IV. EXPERIMENTAL RESULTS**

In this section we first introduce the experimental settings and dataset, then conduct a set of experiments to compare with other leading object detection models in terms of detection speed, Recall, and mAP. After that, we analyze the experimental results and draw some conclusions.

**A. EXPERIMENTAL SETTINGS**

The STBi-YOLO is based on Tensorflow deep learning network and trained on NVIDIA Tesla K80 in order to save training time. The programming language of this model is Python; GPU is accelerated using CUDAv11.0 and CuDNNv8.0. The model parameter settings are shown in TABLE 1 below.

**B. EXPERIMENTAL DATASET**

We use LUNA16 as the experimental dataset. LUNA16 is a high quality lung nodule CT image dataset launched in 2016. It is the most authoritative and representative dataset among the current lung nodule detection researches. This dataset contains a total of 888 3D lung CT image, 1,186 lung nodules and 36,378 annotated information by four professional radiologists. The dataset consists of four main parts: the original CT images, the annotation files of lung nodule locations, the original CT lung regional segmentation files, and the diagnosis result files.

We choose 70% of lung nodule samples in LUNA16 as the training set, 15% as the test set and 15% as the validation set. FIGURE 8 shows the analytical results of the dataset, where (a) shows the distribution of lung nodules, and (b) the distribution of nodule sizes. The horizontal and vertical ordinates represent the width and height of lung nodules respectively.

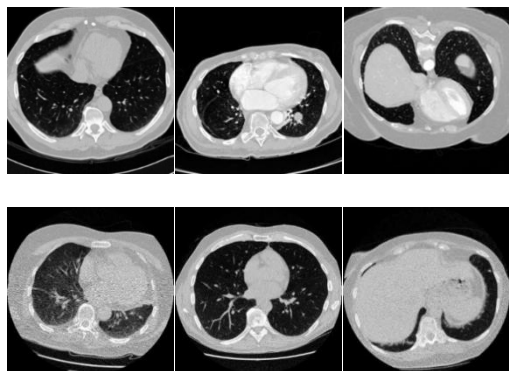


FIGURE 7. Experimental dataset.

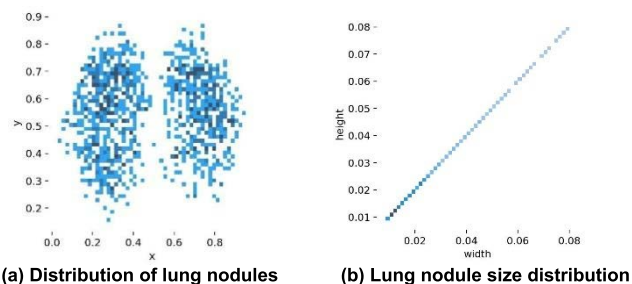


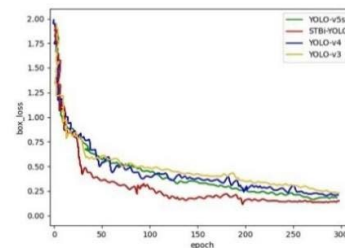
FIGURE 8. Dataset analysis.

TABLE 2. Ablation experimental data.

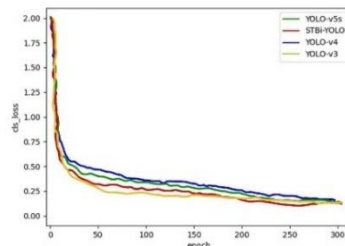
Algorithm	Stochastic-Pooling-based SPP	BiFPN	EIoU	mAP/%	Detection speed/FPS
Proposed method(1)	√			74.26	37.9
Proposed method(2)		√		74.88	33.5
Proposed method(3)			√	73.95	40.1
Proposed method(4)	√	√		74.35	35.5
Proposed method(5)		√	√	75.37	35.8
Proposed method(6)	√		√	75.92	36.8
Proposed method	√	√	√	77.13	43.1

C. MODEL TRAINING

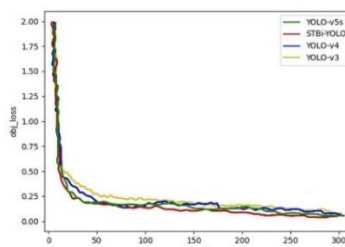
To compare with the YOLO algorithms, we apply the same dataset and parameter settings to YOLO-v3, YOLO-v4, YOLO-v5s, and STBi-YOLO respectively. According to the log files saved during the training process, we plot the loss comparison curves of the four models, as shown in II-B, where (a) represents box bounding loss, (b) the classification loss, and (c) the objectness loss. The horizontal and vertical ordinates indicate the iterations and loss value respectively. From the comparison curves, we see that the loss values of the four models are large at the beginning and experience a sharp



(a) box\_loss



(b) cls\_loss



(c) obj\_loss

FIGURE 9. Loss comparison.

decline before 50 epochs, gradually getting into a decreasing convergence. During the training process, the loss value keeps decreasing and the network keeps fitting. Compared with YOLO-v5 algorithm, the STBi-YOLO is more stable in the training process and the loss value fluctuates in a smaller range. The loss value finally stops at about 0.2 with a better convergence.

D. ABLATION EXPERIMENTS ON STBi-YOLO

To show the contributions of every proposed technique (See the above Formula 1~6) to the performance of STBi-YOLO, we carried out the ablation experiments for each possible combination of the three features (stochastic-pooling, BiFPN, and EIoU). TABLE 2 gives the results of mAP, FPS for the ablation experiments.

As shown in TABLE 2, compared with YOLO-v5s, BiFPN single makes a great contribution to the improvement of mAP; EIoU loss function plays a more important role in increasing detection performance of 7.6FPS without making markable impact on the detection accuracy; stochastic-pooling-based SPP-Net increases the mAP by 2.5% together with BiFPN. Utilizing the EIoU function increases the mAP

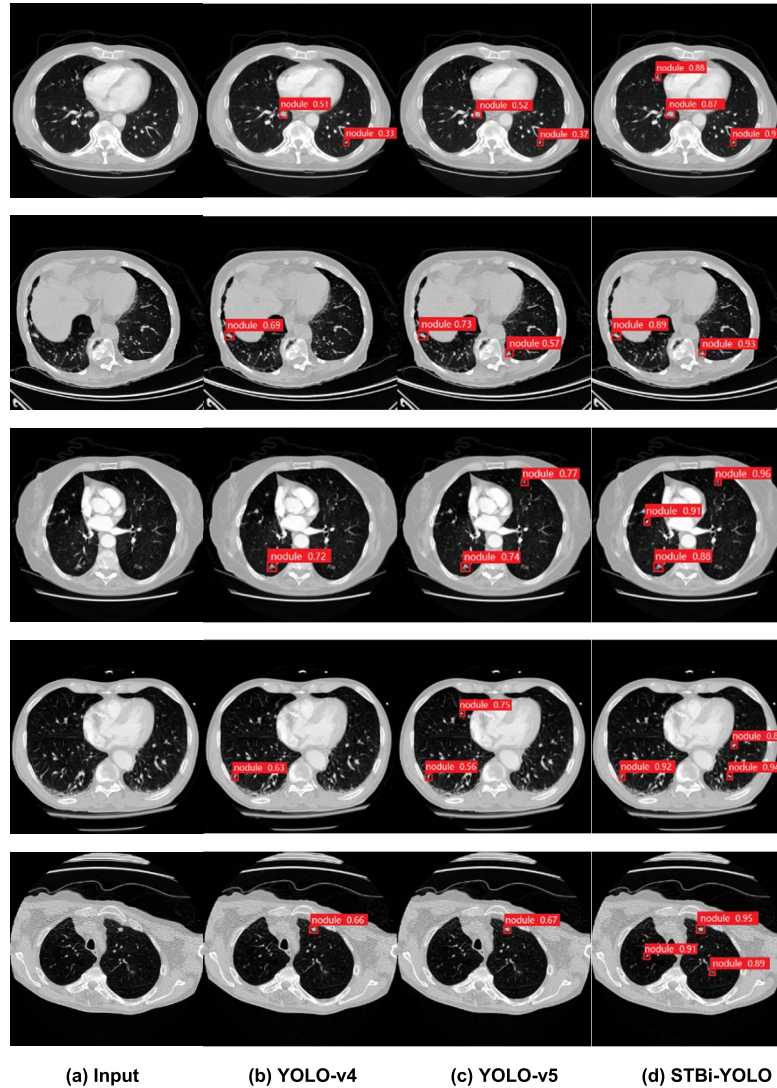


FIGURE 10. Experimental results.

by about 3% in pair of stochastic-pooling-based SPP-Net. The different combinations also positively optimize the overall performance of the STBi-YOLO; the combination of the three improved features has the best effect on the promotion of the detection accuracy and performance.

**E. PERFORMANCE COMPARISON BETWEEN STBi-YOLO AND OTHER DETECTION MODELS**

To verify the effectiveness of the proposed model, we compare it with other typical nodule detection approaches published in recent years. TABLE 3 shows the results in comparison with Faster R-CNN, Mask R-CNN, YOLO-v3, YOLO-v4 and YOLO-v5s.

It can be concluded that, YOLO-v5s is considered as a lightweight network model compared with the two-staged Faster R-CNN, Mask R-CNN, and the one-staged SSD, YOLO-v3 and YOLOv4. Although the proposed STBi-YOLO model is 2FPS slower than YOLO-v5s in

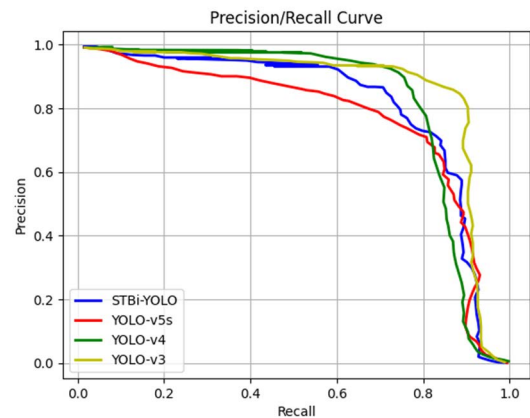


FIGURE 11. PR curves of the four models.

detection speed, the average precision and recall are improved by 5.1% and 5.6% compared with YOLO-v5s and are much



TABLE 3. Model parameter settings.

Model	Weights/MB	mAP/%	Recall/%	Detection speed/FPS
Faster R-CNN	159	91.9	92.5	191
Mask R-CNN	121	73.65	78.3	54
SSD	100.2	75.2	77.0	98
YOLO-v3	235	81.3	82.9	59
YOLO-v4	246	88.4	90.0	41
YOLO-v5s	41.9	90.8	91.1	25
STBi-YOLO	43.6	95.9	96.7	27

higher than YOLO-v3, YOLO-v4, and SSD. Furthermore, the weight of the STBi-YOLO is much lighter than Faster R-CNN and Mask R-CNN without increasing too much detection speed. The average precision and recall of the proposed STBi-YOLO model are not much prominent, but the detection time of each image is 164FPS faster than Faster R-CNN, meeting the requirement of real-time detection.

## V. CONCLUSION

In this study, we proposed an improved lung nodule detection method based on YOLO-v5 algorithm to tackle the challenges confronted when using deep learning methods for lung nodule detection in CT images. We first improve the SPP-Net with multi-convolutional layers based on Stochastic-Pooling to optimize the feature extraction effect, highlighting strong features while retain the less differentiated features. Second, we use BiFPN structure for multi-scale feature fusion to reduce redundant computation. Finally, the original loss function GIoU is replaced by the EIou function, which speeds up the convergence speed and improves the robustness. The algorithm is able to meet the real-time requirements with low computational cost, which improves the localization accuracy of lung nodules. The detection method has been compared with other deep learning detection methods in experiments and obtained accuracy of 96.1% and recall of 93.3%, which can effectively improve the detection performance of lung nodules compared with the traditional YOLO-v5.

Our work still comes in its early stage. In the future, we can try to compress the model, including channel pruning and parameter quantization to reduce the number of parameters and improve the detection speed. We consider to improve the accuracy of the model on the premise of a small amount of parameters and fast inference speed. Ultimately, we will expand the scope of the study and improve the detection model in order to meet the actual requirements of clinical work and further improve the lung nodules detection performance.

## REFERENCES

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, A Cancer J. Clinicians*, vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/caac.21660.

[2] 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, doi: 10.1109/TPAMI.2016.25777031.

[3] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, "Mask R-CNN," in *IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2980–2988, doi: 10.1109/ICCV.2017.322.

[4] J. Redmon, S. Divvala, R. 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, doi: 10.1109/CVPR.2016.91.

[5] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 6517–6525, doi: 10.1109/CVPR.2017.690.

[6] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[7] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis (ECCV)*, Amsterdam, The Netherlands, Sep. 2016, pp. 21–37.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015, doi: 10.1109/TPAMI.2015.2389824.

[9] Z. Dashdorj and M. Song, "An application of convolutional neural networks with salient features for relation classification," *BMC Bioinf.*, vol. 20, no. S10, May 2019, doi: 10.1186/s12859-019-2808-3.

[10] B. K. Chaitanya, A. Yadav, and M. Pazoki, "High impedance fault detection scheme for active distribution network using empirical wavelet transform and support vector machine," in *Proc. 15th Int. Conf. Protection Autom. Power Syst. (IPAPS)*, Shiraz, Iran, Dec. 2020, pp. 149–152, doi: 10.1109/IPAPS52181.2020.9375620.

[11] Q. Li and X. Wang, "Image classification based on SIFT and SVM," in *Proc. IEEE/ACIS 17th Int. Conf. Comput. Inf. Sci. (ICIS)*, Singapore, Jun. 2018, pp. 762–765, doi: 10.1109/ICIS.2018.8466432.

[12] Q. Huang, Y. Chen, L. Liu, D. Tao, and X. Li, "On combining biclustering mining and AdaBoost for breast tumor classification," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 4, pp. 728–738, Apr. 2020, doi: 10.1109/TKDE.2019.2891622.

[13] M. Javaid, M. Javid, M. Z. U. Rehman, and S. I. AliShah, "A novel approach to CAD system for the detection of lung nodules in CT images," *Comput. Meth. Programs Biomed.*, vol. 135, pp. 125–139, Oct. 2016, doi: 10.1016/j.cmpb.2016.07.031.

[14] A. O. de Carvalho Filho, A. C. Silva, A. C. de Paiva, R. A. Nunes, and M. Gattass, "3D shape analysis to reduce false positives for lung nodule detection systems," *Med. Biol. Eng. Comput.*, vol. 55, no. 8, pp. 1199–1213, Aug. 2017, doi: 10.1007/s11517-016-1582-x.

[15] L. Ebner, M. Tall, K. R. Choudhury, D. L. Ly, J. E. Roos, S. Napel, and G. D. Rubin, "Variations in the functional visual field for detection of lung nodules on chest computed tomography: Impact of nodule size, distance, and local lung complexity," *Med. Phys.*, vol. 44, no. 7, pp. 3483–3490, Jul. 2017, doi: 10.1002/mp.12277.

[16] L. Jikui, J. Hongyang, G. Mengdi, H. Chenguang, W. Yu, W. Pu, M. He, and L. Ye, "An assisted diagnosis system for detection of early pulmonary nodule in computed tomography images," *J. Med. Syst.*, vol. 41, p. 30, Feb. 2017, doi: 10.1007/s10916-016-0669-0.

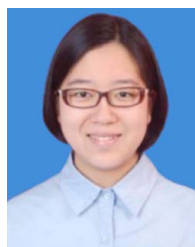
[17] A. Teramoto and H. Fujita, "Automated lung nodule detection using positron emission tomography/computed tomography," in *Artificial Intelligence in Decision Support Systems for Diagnosis in Medical Imaging*. Cham, Switzerland: Springer, 2018, pp. 87–110, doi: 10.1007/978-3-319-68843-5\_4.

[18] E. A. Khordehchi, A. Ayatollahi, and M. R. Daliri, "Automatic lung nodule detection based on statistical region merging and support vector machines," *Image Anal. Stereol.*, vol. 36, no. 2, pp. 65–78, Jun. 2017, doi: 10.5566/ias.1679.

[19] E. E. Nithila and S. S. Kumar, "Automatic detection of solitary pulmonary nodules using swarm intelligence optimized neural networks on CT images," *Eng. Sci. Technol., Int. J.*, vol. 20, no. 3, pp. 1192–1202, Jun. 2017, doi: 10.1016/j.jestch.2016.12.006.

[20] A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. van Riel, M. M. 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 2017, doi: 10.1109/TMI.2016.2536809.

- [21] R. B. Lanfredi, J. D. Schroeder, C. Vachet, and T. Tasdizen, "Adversarial regression training for visualizing the progression of chronic obstructive pulmonary disease with chest X-rays," in *Medical Image Computing and Computer Assisted Intervention*. Shenzhen, China, Oct. 2019, pp. 685–693, doi: [10.1007/978-3-030-32226-7\\_76](https://doi.org/10.1007/978-3-030-32226-7_76).
- [22] M. Ghafoorian, N. Karssemeijer, T. Heskes, M. Bergkamp, J. Wissink, J. Obels, K. Keizer, F. E. de Leeuw, B. V. Ginneken, E. Marchiori, and B. Platel, "Deep multi-scale location-aware 3D convolutional neural networks for automated detection of lacunes of presumed vascular origin," *NeuroImage. Clin.*, vol. 14, pp. 391–399, Feb. 2017, doi: [10.1016/j.nicl.2017.01.033](https://doi.org/10.1016/j.nicl.2017.01.033).
- [23] 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, doi: [10.1109/TBME.2016.2613502](https://doi.org/10.1109/TBME.2016.2613502).
- [24] S. Hamidian, B. Sahiner, N. Petrick, and A. Pezeshk, "3D convolutional neural network for automatic detection of lung nodules in chest CT," *Proc. SPIE*, vol. 10134, pp. 54–59, Mar. 2017, doi: [10.1117/12.2255795](https://doi.org/10.1117/12.2255795).
- [25] H. R. Roth, H. Oda, X. Zhou, N. Shimizu, Y. Yang, Y. Hayashi, M. Oda, M. Fujiwara, K. Misawa, and K. Mori, "An application of cascaded 3D fully convolutional networks for medical image segmentation," *Comput. Med. Imag. Graph.*, vol. 66, pp. 90–99, Mar. 2018, doi: [10.1016/j.compmedimag.2018.03.001](https://doi.org/10.1016/j.compmedimag.2018.03.001).
- [26] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587, doi: [10.1109/CVPR.2014.81](https://doi.org/10.1109/CVPR.2014.81).
- [27] B. Li, W. Jiang, and J. Gu, "Research on target detection algorithm based on deep learning technology," in *Proc. IEEE Int. Conf. Power Electron., Comput. Appl. (ICPECA)*, Jan. 2021, pp. 137–142, doi: [10.1109/ICPECA51329.2021.9362714](https://doi.org/10.1109/ICPECA51329.2021.9362714).
- [28] A. Z. Syaharuddin, Z. Zainuddin, and Andani, "Multi-pole road sign detection based on faster region-based convolutional neural network (faster R-CNN)," in *Proc. Int. Conf. Artif. Intell. Mechatronics Syst. (AIMS)*, Bandung, Indonesia, Apr. 2021, pp. 1–5, doi: [10.1109/AIMS52415.2021.9466014](https://doi.org/10.1109/AIMS52415.2021.9466014).
- [29] X. Chen and A. Gupta, "An implementation of faster R-CNN with study for region sampling," Feb. 2017, *arXiv:1702.02138*.
- [30] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in *Proc. Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Feb. 2017, pp. 379–387.
- [31] B. Singh, H. Li, A. Sharma, and L. S. Davis, "R-FCN-3000 at 30fps: Decoupling detection and classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1081–1090, doi: [10.1109/CVPR.2018.00119](https://doi.org/10.1109/CVPR.2018.00119).
- [32] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007, doi: [10.1109/ICCV.2017.324](https://doi.org/10.1109/ICCV.2017.324).
- [33] F. Yan and Y. Xu, "Improved target detection algorithm based on Yolo," in *Proc. 4th Int. Conf. Robot., Control Autom. Eng. (RCAE)*, Nov. 2021, pp. 21–25, doi: [10.1109/RCAE53607.2021.9638930](https://doi.org/10.1109/RCAE53607.2021.9638930).
- [34] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
- [35] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
- [36] Y. Pang, M. Sun, X. Jiang, and X. Li, "Convolution in convolution for network in network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1587–1597, May 2018, doi: [10.1109/TNNLS.2017.2676130](https://doi.org/10.1109/TNNLS.2017.2676130).
- [37] S. R. Buló, G. Neuhold, and P. Kotschieder, "Loss max-pooling for semantic image segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2126–2135, doi: [10.1109/CVPR.2017.749](https://doi.org/10.1109/CVPR.2017.749).
- [38] Y.-F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan, "Focal and efficient IOU loss for accurate bounding box regression," 2021, *arXiv:2101.08158*.
- [39] Y. Wang, B. Wu, N. Zhang, J. Liu, F. Ren, and L. Zhao, "Research progress of computer aided diagnosis system for pulmonary nodules in CT images," *J. X-Ray Sci. Technol.*, vol. 28, no. 1, pp. 1–16, Feb. 2020, doi: [10.3233/XST-190581](https://doi.org/10.3233/XST-190581).
- [40] A. Painsky and M. Feder, "Robust universal inference," *Entropy*, vol. 23, no. 6, p. 773, Jun. 2021, doi: [10.3390/e23060773](https://doi.org/10.3390/e23060773).
- [41] T. Zhao, D. Gao, J. Wang, and Z. Yin, "Lung segmentation in CT images using a fully convolutional neural network with multi-instance and conditional adversary loss," in *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 505–509, doi: [10.1109/ISBI.2018.8363626](https://doi.org/10.1109/ISBI.2018.8363626).
- [42] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8759–8768, doi: [10.1109/CVPR.2018.00913](https://doi.org/10.1109/CVPR.2018.00913).
- [43] R. Hooda, A. Mittal, and S. Sofat, "Segmentation of lung fields from chest radiographs—A radiomic feature-based approach," *Biomed. Eng. Lett.*, vol. 9, no. 1, pp. 109–117, Feb. 2019, doi: [10.1007/s13534-018-0086-z](https://doi.org/10.1007/s13534-018-0086-z).
- [44] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 658–666, doi: [10.1109/CVPR.2019.00075](https://doi.org/10.1109/CVPR.2019.00075).
- [45] Z. Wang, J. Xin, P. Sun, Z. Lin, Y. Yao, and X. Gao, "Improved lung nodule diagnosis accuracy using lung ct images with uncertain class," *Comput. Methods Programs Biomed.*, vol. 162, pp. 197–209, Aug. 2018, doi: [10.1016/j.cmpb.2018.05.028](https://doi.org/10.1016/j.cmpb.2018.05.028).



**KEHONG LIU** was born in Xi'an, Shaanxi, China, in 1999. She is currently pursuing the bachelor's degree with the Xi'an University of Science and Technology. During the undergraduate studies, she has been engaged in the research project of deep-learning-based lung nodule detection. She has published several journals and conference papers on this topic.

• • •