2021 IEEE 6th International Conference on Intelligent Computing and Signal Processing (ICSP 2021)

Blind Detection of Communication Signals Based on Improved YOLO3

Rundong Li[*1, 2], Jianhao Hu[1], Shaoqian Li[1], Shaohe Chen[2], Peng He[2]
[1] National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China
Chengdu, China
[2] Southwest Electronics and Telecommunication Technology Research Institute
Chengdu, China
Email: pikachulee@163.com

2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP) | 978-1-6654-0413-6/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICSP51882.2021.9408998

***Abstract*—Blind detection of communication signals is a challenging task. In this paper, a general and novel blind detection method is proposed based on the similarity between communication signal detection and image object detection. We designed an improved YOLO3 model to detect the communication signals contained in the 2D wide-band spectrograms, the main innovates are as follows: 1) in order to reduce the burden of spectrograms labeling, an ingenious and automatic signal object labeling method is proposed; 2) in view of the fact that the communication signals are long and narrow objects in the spectrograms, the corresponding prior anchors are designed to improve the detection probability; 3) in order to improve the training efficiency and detection accuracy, the CIOU cost function and DIOU-NMS inference algorithm are introduced to achieve high-precision signal detection. The simulation results demonstrate that the proposed method can effectively detect the continuous and burst signals in wide-band communication signal data, and its performance is better than the traditional energy detection method.**

Keywords-communication signals; blind detection; YOLO3

I. INTRODUCTION

In a non-cooperative communication system, blind signal detection refers to determine the existence of signals from the received wide-band signal data which are adversely affected by noise, interference and channel fading without prior knowledge, and estimate the signal parameters such as center frequency, bandwidth, start and end time, signal type and so on. Blind detection of communication signals is very important in civil applications such as spectrum monitoring and management, as well as in military fields such as battlefield electromagnetic spectrum situation awareness.

With the increasing frequency range of wide-band communication signals, the electromagnetic environment is becoming more and more complex, which is mainly manifested in the following aspects: 1) the signals in electromagnetic space are increasing, and the background noise is significantly increased; 2) Various kinds of interference become more serious, and the background noise often presents non-Gaussian characteristics; 3) Under the influence of various burst signals and interference as well as background, the high noise changes dynamically. These complex electromagnetic environments put forward higher requirements and more severe challenges to the existing signal detection techniques : 1) the signal is often submerged in the complex noise, presenting extremely low spectral density; 2) The detection method based on Gaussian noise is difficult to adapt to non-Gaussian background noise; 3) The detection method is susceptible to the influence of high dynamic background environment, and the performance loss is serious; 4) Signal detection technology is difficult to meet the requirements of fast and real-time detection.

Challenged by the above issues, scholars have carried out in-depth researches on the blind detection of communication signals, and proposed a series of solutions, among them, the matched filtering method[1], energy detection method [2], cyclostationary method [3], eigenvalue detection method [4] are the most widely used methods. The matched filtering method has a high detection speed, but it needs prior information of the detection signals, so it is hard to practice. The energy detection method has the same fast detection speed, but its robustness is poor, and it is greatly affected by noise and fading. Cyclostationarity method has good robustness, but its complexity is high, which is not applied to practical application. The eigenvalue detection method is also robust, but it can only be applied to specific signals with known characteristics, so it is difficultly popularized.

In order to achieve robust and reliable blind detection of communication signals, this paper utilize the advanced object detection technology based on deep learning, and transforms the blind signal detection problem into a computer vision problem of detecting the existence of the signal objects and searching the location of objects in the spectrograms of the wide-band signals. Among the lots of object detection algorithms based on deep learning, the most typical ones are RCNN[5], Fast-RCNN[6], Faster-RCNN[7], SSD[8], YOLO[9], YOLO9000 [10], YOLO3[11]. YOLO3 is an improved YOLO object detection network, its detection precision and speed are both excellent, so this paper carried out the blind detection of the wide-band signals by improving YOLO3 network, i.e., the signal objects labeling method, the network structure and the training and inference strategy are all optimized. The test results show this proposed method can achieve high detection precision, and the performance is superior to the traditional detection algorithms.

II. SYSTEM MODEL

Image object detection and recognition is to detect the location area of each object in an image and identify which category each object belongs to. The purpose of blind detection of communication signal is to detect what kind of communication signal appears in which frequency range and which time period within the specified wide-band spectrum range, hence its task objective is very similar to image object detection and recognition. If the received wide-band signal is transformed into a two-dimensional (2D) spectrogram, as

978-1-6654-0413-6/21/$31.00 ©2021 IEEE

shown in Fig. 1, signals will appear as a "strip-like" object with prominent brightness and different texture features in the spectrograms. Therefore, the signal detection problem can be transformed into a problem of detecting and identifying each " strip -like" object in the spectrograms.

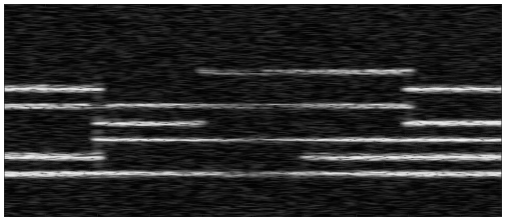

Figure 1. An example of 2D spectrogram of communication signals

Fig. 2 shows the architecture of the proposed signal detection algorithm based on deep neural network (DNN), the implementation steps are as follows: Firstly, spectrogram computation of received communication signals is carried out to obtain a 2D spectrogram data set. Secondly, an automatic labeling algorithm is designed to quickly label the location areas and types of the signals in the data set, so as to obtain enough labeled data samples for training. Thirdly, a DNN model for spectrogram object detection is built. In this paper, the YOLO3 model is selected, and the network structure and training algorithm are designed and optimized. A set of optimal network parameters are obtained by training the model with the labeled data-set. Finally, based on the optimal YOLO3 model, the inference strategy is improved to infer the new spectrograms, and the detection results are obtained, including frequency, bandwidth, start and end time, signal type and other parameters.

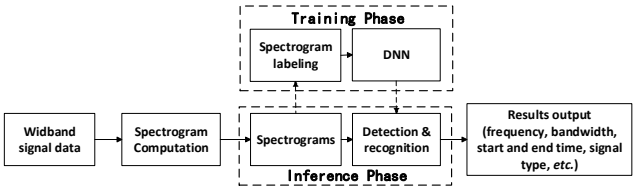

Figure 2. Signal detection and recognition architecture based on DNN.

III. ALGORITHM PRINCIPLE

As shown in Fig. 2, the algorithm mainly includes spectrogram computation, spectrogram labeling, YOLO3 network model construction and improvement, which means improving YOLO3 training and inference phases. The main principles are described as follows.

A. 2D spectrogram computation

Suppose the received time-domain data of the wide-band signals is $x(t)$, then the calculation formula for obtaining 2D spectrogram is defined as follows:

$$STFT_x(t,\Omega)=\int x(\tau)g(\tau-t)e^{-j\Omega\tau}d\tau \qquad (1)$$

In the formula, t,Ω are the time and frequency variables respectively, $g(t)$ is the sliding window function, which control the data length and window type involved in Fourier transform at each time.

B. Automatic spectrogram labeling and sample-set construction

Deep learning is used to carry out object detection, and the quality of sample labeling directly affects the detection and recognition effect. However, the labeling of samples is a laborious task. Generally speaking, in order to obtain high quality labeling data, on the one hand, a large number of reliable original samples should be obtained, and on the other hand, a large amount of manpower and time should be paid to carry out labeling. In order to achieve efficient sample labeling, this paper presents a technical method to quickly and automatically construct a large-scale spectrogram sample-set under the condition that a small number of spectrogram samples are collected. This labeling method generates a large number of new spectrogram samples according to the idea of data enhancement, and automatically constructs labels, thus greatly reducing the workload of labeling. The main steps are as follows.

1) Selecting basic object sets of signals

Several pixel areas corresponding to signal objects in the spectrograms are deducted as spare signal blocks, which were from the acquired signal spectrogram set, through manual cutting or image saliency detection method.

2) Generating background noise images

The background noises does not contain the objects to be detect. It is mainly generated by two methods according to the different channel environments. One is to generate different types of noise signals with Gaussian distribution, Rayleigh distribution, non-Gaussian α-stable distribution and so on, and then transform them to the spectrogram domain. The other is selecting the received noise data without signal, and carrying out spectrogram transformation, and then the background images are generated after superposition. Finally, we need to normalize the power of the background noise images.

3) Synthesizing new spectrograms

The pixel blocks of the signals extracted in step 1 are superimposed with the background noise images obtained in step 2. The operation of superposition is achieved as follows: Firstly, the random seed is generated to randomly place the signal object blocks in the background noise images, and the complex situation of aliasing signals can be simulated at the same time. Secondly，add an appropriate amount of noise and interference to the signal object blocks to simulate a variety of signal spectrograms， which is under the condition of interference and noises.

4) Generating label files

A formatted label file is automatically generated according to the positions and the object types of the signals in the synthetic spectrograms, which mainly includes the signal type, signal position coordinates and other elements.

By using this method, only a small number of actual signal samples and background noise samples should be obtained, which can be enhanced and expanded into a large number of labeled sample-set, which is easy to operate in practice.

Fig. 3 shows the automatic synthesis and labeling results of spectrogram images, and the generated labeling file is shown in Fig. 4, where the information from left to right is the signal category, the starting point of the signal object locations, and the signal width and height respectively.

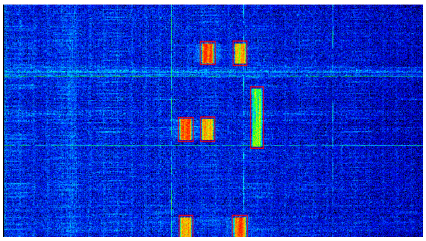

Figure 3. Automatic synthesis and labeling results of signal spectrogram.

```
6 0.4355 0.5363 0.0361 0.1000
6 0.4360 0.9525 0.0332 0.0975
6 0.4873 0.2087 0.0361 0.0950
6 0.4873 0.5350 0.0342 0.1025
6 0.5635 0.9525 0.0361 0.0975
6 0.5640 0.2100 0.0332 0.1025
6 0.6050 0.4850 0.0313 0.2625
```

Figure 4. Generated labeling file.

C. *YOLO3 model construction and improvement*

YOLO3 model is an improved object detection network based on YOLO1 and YOLO2 networks to solve the problem of small object detection. The network structure of YOLO3 is shown in Figure 5, which mainly includes the following two modules:

1) Basic feature extraction network. The basic network of YOLO3 used to extract object features is Darknet-53 with residual structure, which is different from previous versions of YOLO. In previous versions, YOLO adopted a straight-through network such as ordinary multi-layer convolutional network or Darknet-19, which gradient would disappear and was difficult to train effectively. Darknet-53 adopts a cross-layer residual connection similar to ResNet, which greatly reduces the difficulty of training deep network. Therefore, the network can be made to 53 layers or deeper, thus greatly improving the accuracy of feature extraction and providing a nice foundation for high-precision object detection and recognition.

2) Feature pyramid prediction module. Inspired by the structure of the feature pyramid networks (FPN), YOLO3 used multi-scale features of cross layer fusion for detection of different size objects, as shown in the lower part of Fig. 5, three prediction branches output three different sizes of feature maps, which called y1, y2, y3, each grid cell of the feature maps predicts three bounding box, where, each bounding box contains 3 kinds of parameters, which are four box coordinate parameters, one confidence score, and C class probabilities (for example, if there are 80 kinds of object to recognize, then C=80). Therefore, the channel depth of the feature maps y1-y3 is 3× (4+1+80) =255, and their sizes are 8×8, 16×16, and 32×32 respectively (taking input images of 256×256 as an example). The down-sampling rate is higher when the feature map is smaller. Therefore, the 8×8 feature map y1 corresponds to the detection of large objects, the 16×16 feature map y2 corresponds to the detection of medium objects, and the 32×32 feature map y3 corresponds to the detection of small objects. Using the three different size feature maps, a total of (8×8+16×16+32×32)×3=4032 multi-scale bounding boxes were output, far larger than the 98 for YOLO1 and 320 for YOLO2，which was adapted to multiple scale objects. After fusion, the performance of detection of different size objects existed in the figure was improved significantly.

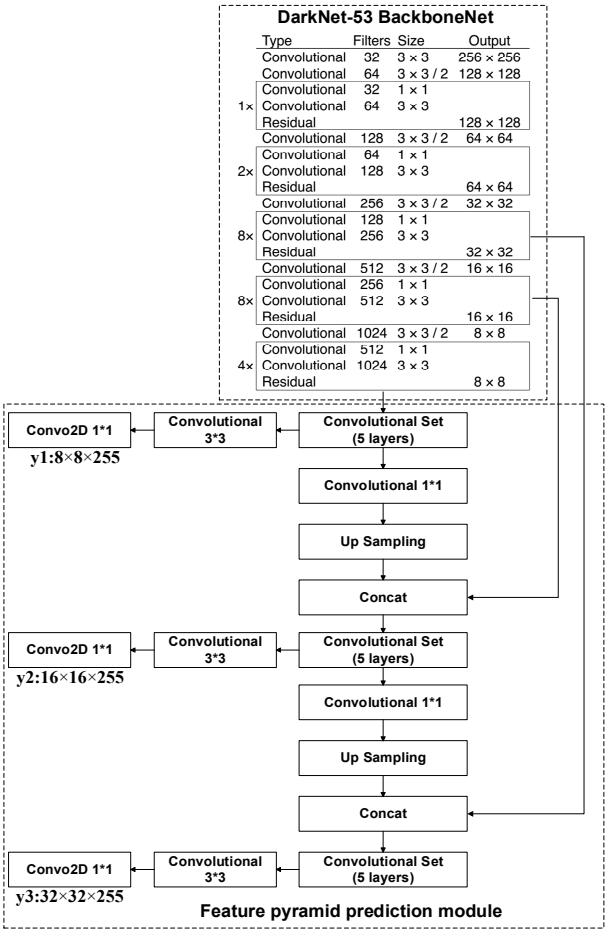

Figure 5. YOLO3 model structure.

In order to further improve the convergence speed of network training, YOLO3 still follows the prior anchors practice in RCCN[5], YOLO2[10] and other models. The prediction bounding box is not randomly initialized, but is obtained by using the border regression algorithm on the basis of a group of optimized priori anchor boxes to ensure that the prediction box quickly converge to the optimal value.

However, the signal objects in the spectrograms is presented as a slender strip (as shown in Figure 1), while the ordinary YOLO3 network is mainly used for natural image processing, and the designed prior anchors rarely has such shape, so it cannot be directly used for signal detection in spectrograms. To help solve the problem, in this paper, based on the characteristics of the communication signal frequency according to using K-Means clustering algorithm to statistical

analysis on the actual labeled borders in the spectrogram dataset. This paper get 9 anchor boxes of different scale ratio, which are supplied to the 3 dimension characteristic diagram, according to the different feature map responsible for detection the different size objects. As shown in Fig. 6, the first three big size anchors are assigned to the 8×8 feature map y1, and is responsible for detecting large objects, in the middle, three medium size anchors are assigned to the 16×16 feature maps y2, which are responsible for detecting medium objects. The last three small anchors are assigned to the 32×32 feature map y3, which are responsible for detecting small objects. The anchor boxes generation and distribution scheme ensures that the multi-scale feature map can detect the multi-scale "slender-strip" spectrogram objects.

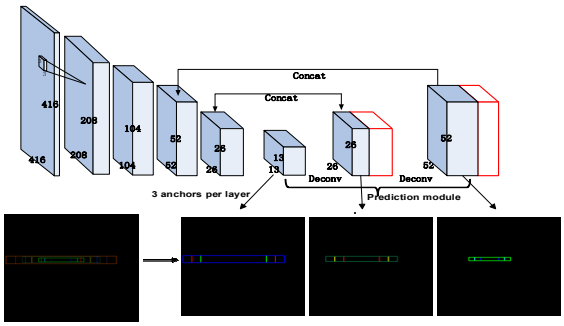

Figure 6. Schematic diagram of prior anchor box allocation.

D. YOLO3 training and inference improvement

After the construction of the network, the training and inference algorithm will be carried out. The cost function used in standard YOLO3 network training is shown in Equation (1) :

$$loss = loss_{coord} + loss_{conf} + loss_{cls} \tag{1}$$

The parameters such as bounding boxes, confidence scores and class probabilities obtained from feature graphs y1-y3 are mainly used to calculate three cost functions of coordinate regression $loss_{coord}$, confidence $loss_{conf}$ and classification $loss_{cls}$ with real label parameters. The cost functions of each part are shown in Equations (2)-(4) :

$$\begin{aligned} loss_{coord} = {} & \lambda_{coord} \sum_{i=0}^{K\times K} \sum_{j=0}^{M} I_{ij}^{obj} \left[\left(x_i - \hat{x}_i\right)^2 + \left(y_i - \hat{y}_i\right)^2 \right] + \\ & \lambda_{coord} \sum_{i=0}^{K\times K} \sum_{j=0}^{M} I_{ij}^{obj} \left(2 - w_i \times h_i\right) \left[\left(w_i - \hat{w}_i\right)^2 + \left(h_i - \hat{h}_i\right)^2 \right] \end{aligned} \tag{2}$$

$$\begin{aligned} loss_{conf} = {} & -\sum_{i=0}^{K\times K} \sum_{j=0}^{M} I_{ij}^{obj} \left[\hat{C}_i \log C_i + \left(1 - \hat{C}_i\right) \log\left(1 - C_i\right) \right] - \\ & \lambda_{noobj} \sum_{i=0}^{K\times K} \sum_{j=0}^{M} I_{ij}^{noobj} \left[\hat{C}_i \log C_i + \left(1 - \hat{C}_i\right) \log\left(1 - C_i\right) \right] \end{aligned} \tag{3}$$

$$loss_{cls} = -\sum_{i=0}^{K\times K} I_{ij}^{obj} \sum_{c \in classes} \left[\begin{array}{l} \hat{p}_i(c) \log\left(p_i(c)\right) + \\ \left(1 - \hat{p}_i(c)\right) \log\left(1 - p_i(c)\right) \end{array} \right] \tag{4}$$

In the formula, $\left(x_i, y_i, w_i, h_i\right)$ is the real bounding box parameters, and $\left(\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i\right)$is the prediction bounding box parameters; C_i is the real confidence scores, $\hat{C}_i$ is the prediction confidence scores; $p_i(c)$ is the real class probabilities, and $\hat{p}_i(c)$ is the prediction class probabilities.

After the network training is completed with the cost function, the new spectrogram is inferred according to the process shown in Fig. 7. which mainly includes YOLO3 features output, boxes and scores calculation and acquisition, non-maximum suppression (NMS) processing and other processes, and finally the detected spectrogram signal object is obtained.

Figure 7. Signal detection process based on YOLO3.

However, there is an important problem in the training and inference of standard YOLO3: in the training phase, calculation of coordinate regression loss $loss_{coord}$ uses the L2 norm of the difference between the prediction boxes and the real boxes, but in the inference phase, the NMS algorithm uses the intersection over union (IOU) parameter, which are not equivalent. If IOU is directly used as the cost function to train the network, there are two problems: 1) If the two boxes do not overlap at all, IOU will always be 0, and the cost function cannot be differentiated, which makes it impossible to learn and train. 2) The coincidence degree of the two boxes cannot be accurately reflected. As shown in Figure 8, IOUs in the three cases are all the same, but in fact their coincidence degrees are different. The regression effect of the left border is the best, while the right border is the worst.

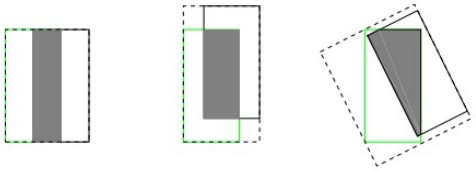

Figure 8. Comparison of overlapping situations on the same situation among three IOUs.

In order to solve this problem, this paper made two improvements:

1) DIOU-NMS inference (DIOU-NMS) was carried out by using Distance IOU(DIOU)[13] to improve the calculation accuracy of IOU and the detection accuracy. Different from the original NMS, DIOU-NMS not only considers the value of IOU, but also considers the distance between the center points of two boxes, which makes IOU calculation more accurate and reliable. Its calculation formula is shown in Equation (5) :

$$IOU\left(\mathbf{b}, \mathbf{b}^{gt}\right) - \frac{\rho^2\left(\mathbf{b}, \mathbf{b}^{gt}\right)}{c^2} \tag{5}$$

Where, $\rho(\cdot)$ is the Euclidean distance between the centers of the two boxes $\mathbf{b}, \mathbf{b}^{gt}$, and c is the diagonal length of the smallest box containing the two boxes.

2) Complete-IOU(CIOU)[13] was adopted as the coordinate regression cost function during training to solve the problem that the original IOU could not be differentiated. The

calculation formula of CIOU as coordinate regression loss is shown in Equation (6) :

$$loss_{CIOU} = 1 - IOU\left(\mathbf{b}, \mathbf{b}_{gt}\right) + \frac{\rho^2\left(\mathbf{b}, \mathbf{b}^{gt}\right)}{c^2} + \alpha v \tag{6}$$

Where, the α is weight function, v used to measure the similarity of aspect ratios,

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{7}$$

The cost function takes into account the overlapping area, distance from the center points and aspect ratios of the boxes. Under the circumstances, faster convergence speed and higher detection accuracy can be achieved when training.

Through the above two improvements, fast training convergence is achieved, and higher detection accuracy is achieved.

IV. Experimental Analysis

To verify the algorithm efficiency, we have generated a simulation spectrogram data-set to carry out experiment, the data-set parameters are as follows: the sampling rate is 250 KHz, signal bandwidth is 120 KHz, each spectrogram contains four object signals at most, signal modulation mode is random among the AM, FM, PSK and FSK, the modulation rate is 3 kBd, add Rayleigh fading in the signal, the signal-to-noise ratio range is 0 to 20 dB, including of continuous and burst signals.

The experimental conditions and result analysis are shown below.

A. Experimental conditions and results

The number of the training spectrograms is 12,500; the number of the validation spectrograms is 2,522; batch size: 32, maximum batches: 70,000, training epoch: 190 epoch.

The IOU threshold of positive and negative samples: 0.3 (the IOU threshold of prediction boxes and real boxes is small, which can ensure that more prediction boxes can fully learned); score threshold =0.5, test IOU threshold =0.45.

Number of test spectrograms: 5,305, including a total of more than 15,097 communication signal objects (8,678 continuous signals and 6,419 burst signals). The detection and recognition results are shown in Table I, and the detection effect is shown in Fig. 9 (the yellow box is the detected continuous signals, and the red box is the burst signals).

TABLE I. The Detects and Recognition Results for Simulation Data

Signal types	*precision*	*recall rate*	*accuracy*	*average accuracy*
continuous	0.947	0.981	0.935	0.960
burst	0.984	0.973	0.969	

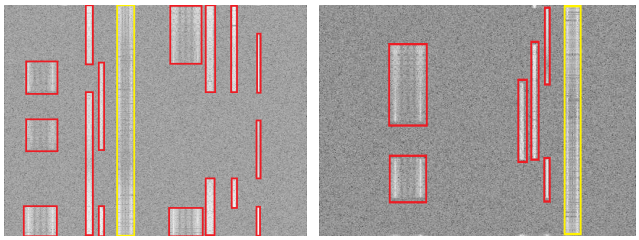

Figure 9. Signal detection results in the spectrograms.

B. Compared with traditional detection methods

In order to verify the efficiency of the algorithm, we adopted the same test data-set to evaluate the traditional energy detection method proposed in literature [2]. The detection results of the continuous and burst signals and the comparison with the proposed method are shown in Fig. 10 and 11. It can be seen that the intelligent detection algorithm proposed in this paper is superior to the traditional energy detection algorithm in terms of precision rate, recall rate, accuracy and other indicators.

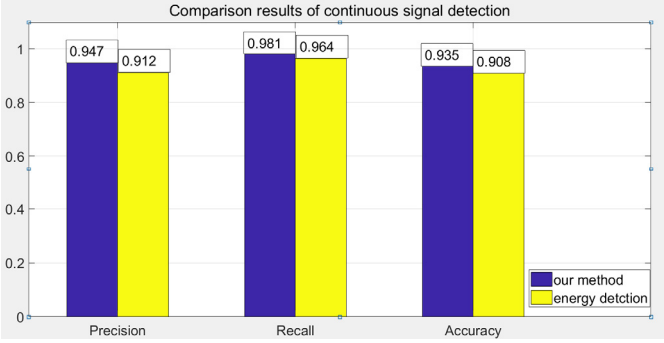

Figure 10. Comparison results of continuous signal detection

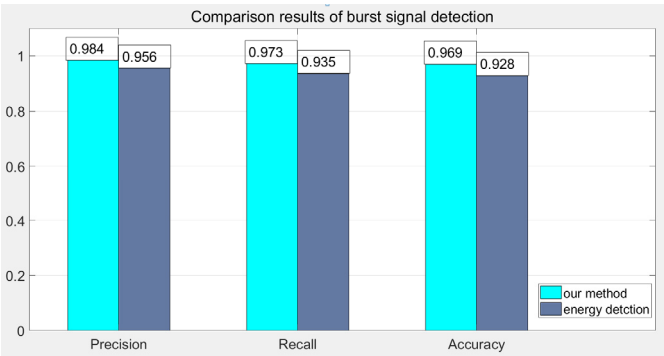

Figure 11. Comparison results of burst signal detection

C. Analysis of experimental results

In the spectrogram of communication signals, the aspect ratio of signal objects is very extreme, and the anchor settings of general detection models, such as R-CNN and RetinaNet detection models [12], has rotation in-variance, which cannot be applied to the detection of spectrogram of communication signals. In this paper, the anchors are set by clustering priori in YOLO3 to make the anchors more consistent with the object distribution, which effectively improved the accuracy of the detection frame.

Otherwise, signal data on the frequency axis belongs to the small object detection, object feature through gradually down-sampling can lead to loss of information, however, the adopted FPN schema can keep multi-scale feature pyramid structure, so as to ease small objects detection problem. In addition, deconvolution can get higher spatial resolution features, which is beneficial to object detection. At the same time, deconvolution can extract stronger semantic information. Cross layer connection can fuse the features reserved by shallow small objects with deep features, which effectively improves the detection effect of signal objects.

V. CONCLUSION

This paper presents a general blind detection method of communication signals, which is implemented based on 2D spectrogram detection by a improved YOLO3 model, the following improvements are made: in order to reduce the burden of signal labeling, an automatic construction method of sample labeling is proposed; in view of the slender shape of communication signals in spectrograms, corresponding prior anchor boxes are designed to improve the detection probability; in order to improve the training efficiency and detection accuracy, CIOU cost function and DIOU-NMS inference algorithm are introduced to realize high-precision signal detection. The proposed detection method is verified by using the simulation data and the results show that the method can effectively detect the continuous and burst signals among the wide-band communication signal, and its performance is better than the traditional energy detection methods, so it has a good promotion significance.

REFERENCES

[1] J. Theiler and B. R. Foy, "Effect of signal contamination in matched-filter detection of the signal on a cluttered background," in IEEE Geoscience and Remote Sensing Letters, vol. 3, no. 1, pp. 98-102, Jan. 2006, doi: 10.1109/LGRS.2005.857619.

[2] Y. Chen, "Improved energy detector for random signals in gaussian noise," in IEEE Transactions on Wireless Communications, vol. 9, no. 2, pp. 558-563, February 2010, doi: 10.1109/TWC.2010.5403535.

[3] J. Lunden, S. A. Kassam and V. Koivunen, "Robust Nonparametric Cyclic Correlation-Based Spectrum Sensing for Cognitive Radio," in IEEE Transactions on Signal Processing, vol. 58, no. 1, pp. 38-52, Jan. 2010, doi: 10.1109/TSP.2009.2029790.

[4] F. Penna, R. Garello, D. Figlioli and M. A. Spirito, "Exact non-asymptotic threshold for eigenvalue-based spectrum sensing," 2009 4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications, Hannover, 2009, pp. 1-5, doi: 10.1109/CROWNCOM.2009.5189008.

[5] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 580-587, doi: 10.1109/CVPR.2014.81.

[6] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.

[7] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

[8] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, and S. E. Reed, "SSD: single shot multibox detector," 2016 European Conference on Computer Vision, Amsterdam, 2016, pp. 21-37.

[9] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

[10] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 6517-6525, doi: 10.1109/CVPR.2017.690.

[11] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement, " arXiv preprint, 2018.

[12] T. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal Loss for Dense Object Detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 2, pp. 318-327, 1 Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.

[13] Z. Zheng, P. Wang, W. Liu, et al, "Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression," 2020 AAAI Conference on Artificial Intelligence, New York, 2020, pp. 1-7.