ISSN 2616-5775 Vol. 6, Issue 7: 101-113, DOI: 10.25236/AJCIS.2023.060714

The Research of Spectrum Signal Recognition Based on Improved YOLOv5 Algorithm

Qinjun Li[1,2,*], Yan Zhao[1,2], Tianwei Cui[1,2], Yuying Wu[1,2]

[1]Shaanxi Joint Laboratory of Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, 710021, China
[2]College of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, 710021, China
**Corresponding author*

Abstract: *To enhance the rapid detection and accurate identification of radio signals, we propose using a deep learning method called YOLOv5s-CWMDSQ for signal identification. This method involves sending the waterfall map of a radio signal spectrum to an improved target detection network for classification. Our algorithm builds upon YOLOv5 by introducing improvements such as the CA attention mechanism to enhance the model's accuracy in locating and identifying targets, and the boundary loss function WIOU to improve overall detector performance. We also replaced the MESwish activation function, utilized Decoupled Head to speed up network convergence, and incorporated the SPPFCSPC module to elevate the model's receptive field and feature expression ability through multi-scale spatial pyramid pooling. Lastly, we integrated data enhancement to improve the diversity, robustness, and generalization ability of the model, achieving higher accuracy and performance. Experimental results exhibited an increase in the mAP value from 82.2% to 90.1% and detection speed of 44.488FPS in the dataset with 3,000 samples of 6 signal types, proving the model's superior accuracy and real-time capabilities.*

Keywords: *Deep Learning, Signal Recognition, Signal Classification, Spectrum Waterfall Map*

1. Introduction

Early electrical signal detection was originally developed for military purposes and later applied to commercial communication [1-3]. Various methods can be used to achieve signal detection, including traditional tools such as oscilloscopes and spectrum analyzers as well as more advanced technologies like FPGA (Field-Programmable Gate Array) and DSP [4-5]. Additionally, incorporating artificial intelligence algorithms can improve the speed and accuracy of signal identification [6-12].

For instance, O'Shea T J et al. applied unsupervised and semi-supervised learning for radio signal recognition [13-16], while Kulin M et al. utilized an end-to-end neural network to detect signal spectrums [17].Other researchers proposed convolutional neural networks for signal identification, such as Selim A et al. for radar signals [18] and Maglogiannis V et al. for WiFi signals [19]. Image detection algorithms like R-CNN [20], YOLO [21] and SSD [22] also have applicable uses, with R-CNN using region extraction and detection classification and YOLO and SSD using end-to-end direct detection for location and identification. Li Xun et al. used the TensorFlow framework for broadcast signal identification using AI algorithms [23]. Meanwhile, Zhou Xin et al. utilized YOLOv2 algorithm with waterfall maps of radio signal spectrums as a dataset to identify signals [24], and Zhou Yuhang et al. employed the YOLO algorithm for signal spectrum detection and identification [25].

Based on these sophisticated approaches, we find the end-to-end direct detection method to be the most effective. In particular, YOLOv5 algorithm offers many advantages like small model size, high flexibility, and fast detection speeds. Therefore, we propose an improved YOLOv5s-CWMDSQ algorithm, which incorporates enhancements to the network structure and loss function for the identification and classification of six distinct signals.

2. YOLOv5 Network Structure

The YOLO algorithm can input signal spectrum waterfalls into the network and simultaneously

ISSN 2616-5775 Vol. 6, Issue 7: 101-113, DOI: 10.25236/AJCIS.2023.060714

identify both target judgment and target recognition. The output consists of the position and category of the regression prediction frame. Based on its depth and width, YOLOv5 has four variations: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. YOLOv5s has the smallest network parameters and the fastest training speed, making it the ideal choice as our basic network.

The YOLOv5s model comprises four parts: Input Terminal, Backbone Network, Neck, and Prediction Terminal. The Input Terminal functions to preprocess the input data by applying adaptive image scaling, data enhancement, and adaptive anchor frame techniques. This step boosts the network's generalization ability and robustness, making it more reliable for detecting signals. The Backbone Network extracts the feature information, which consists of Focus structure, C3 structure, CBL structure, and SPPF structure. Focusing on these structures enhances signal detection and improves target classification accuracy. The Neck employs the FPN+PAN (Feature Pyramid Network, Path Aggregation Network) feature extraction method. This operation significantly improves the model's target detection capabilities, particularly in detecting small-sized target objects. Enhancing the network's ability to fuse different feature maps. The Head is the YOLOv5's head structure and outputs the prediction results of the network. There are three different types of feature maps (76×76, 38×38, and 19×19) generated by the Head. These maps are used to detect small, medium, and large objects, respectively. The detection results contain crucial information such as target positioning coordinates (x, y, w, h), object information (obj), and category information (cls). The YOLOv5 algorithm generates prediction frames based on the feature maps' different sizes, performs non-maximum suppression on the prediction frames, and screens target frames. The algorithm deletes candidate frames with low confidence scores while retaining the candidate frames that have high scores as the final results. This detailed process effectively enhances the signal detection capabilities of the YOLOv5s-CWMDSQ algorithm.

3. Improved Target Detection Algorithm

By integrating deep learning with signal recognition, a neural network algorithm can recognize various signal types. Using the lightweight YOLOv5s as a basis, signal detection and recognition is achieved, resulting in the creation of the YOLOv5s-CWMDSQ model. The updated network structure is displayed in Figure 1, showcasing the new enhancements made to the network.

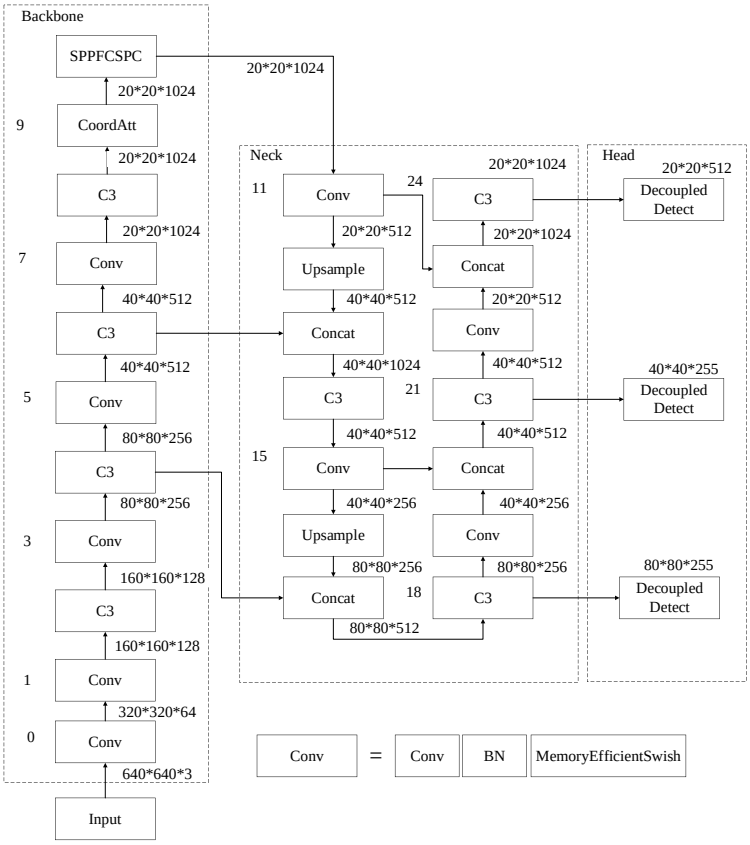


Figure 1: Improved network structure diagram

ISSN 2616-5775 Vol. 6, Issue 7: 101-113, DOI: 10.25236/AJCIS.2023.060714

The YOLOv5s-CWMDSQ model incorporates improvements applied to the YOLOv5s basic model. These enhancements include:

(1) The addition of a CA attention mechanism to the Backbone network. This mechanism reutilizes useful feature information, enhances location and channel information extraction, and improves target positioning accuracy.

(2) The replacement of the original CIOU loss function with the WIOU loss function. In WIOU calculations, different relative positions of overlapping areas lead to variable weights of overlapping areas, and the pixel number is weighted based on frequency of occurrence to reduce the influence of incorrect samples on the dataset. The dynamic non-monotonic focusing mechanism improves the regression of low-quality samples, enabling WIOU to focus on ordinary quality anchor frames and significantly improving the overall performance of the detector.

(3) The use of MESwish activation function over SiLU activation function. MESwish has a higher memory efficiency and computing speed than SiLU.

(4) Decoupled Head is applied to enhance the accuracy of classification and positioning. This decoupling head realizes the coordinate regression of bounding box and object classification, which enhances the convergence speed and detection accuracy.

(5) The SPPFCSPC module improves the model's receptive field and feature expression via multivariate spatial pyramid pooling. This technique significantly enhances the model's target detection capabilities.

(6) Data enhancement techniques including image scaling, translation, flipping, mosaic, and cropping improve data diversity, enhance the generator's robustness and generalization abilities, and significantly improve the model's accuracy and performance.

3.1. Improvement of Attention Mechanism

The introduction of the CA (Coordinated Attention) module to the network structure comprises two parts: global information embedding part and coordinate attention generation part. The global information embedding part advances the network's receptive field and enables the network to focus on larger areas, solving the issue of large-scale variation in datasets. The coordinate attention generation part extracts critical features from the image, enhancing the feature extraction abilities of small targets. It improves the focus on essential features and facilitates improved target positioning, resulting in better model accuracy. Figure 2 illustrates the realization process of the CA attention mechanism.

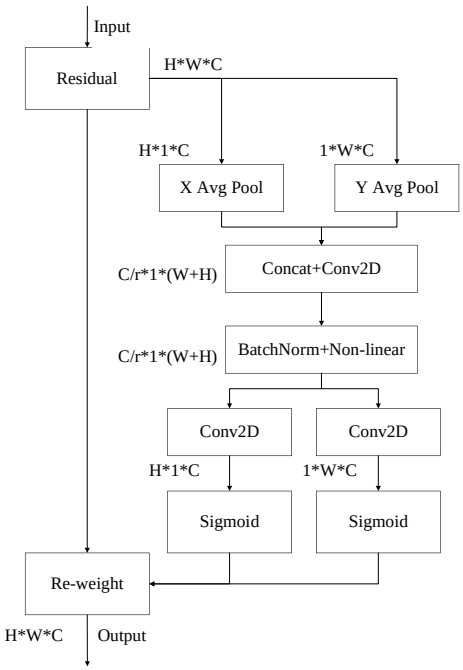


Figure 2: Block diagram of CA attention mechanism implementation

ISSN 2616-5775 Vol. 6, Issue 7: 101-113, DOI: 10.25236/AJCIS.2023.060714

The CA's attention mechanism fully concentrates on channel and spatial information. The feature map of the global receptive field is acquired by utilizing global average pooling in both width and height directions. Equations (1) and (2) represent the output formulas in the width and height directions, respectively. In Equation (1), x_c denotes the input of the c-th channel, and the output of the c-th channel with height h is represented. In Equation (2), x_c represents the input of the c-th channel, while the output of the c-th channel with width w is demonstrated.

$$z_c^h(h) = \frac{1}{W}\sum_{0\le i\le W} x_c(h,i) \tag{1}$$

$$z_c^w(h) = \frac{1}{H}\sum_{0\le i\le W} x_c(j,w) \tag{2}$$

Subsequently, the characteristic images of width and height are concatenated, followed by a convolution operation utilizing a convolution kernel of 1×1. This process results in the channel width being reduced to C/R. After batch normalization, the resulting feature map, F_1, undergoes a Sigmoid activation function. Formula (3) demonstrates the production of the final feature map, f. The symbols [,] and δ denote the concat and activated function operations, respectively, while F_1 represents the output of the 1×1 convolution operation, and f represents the feature map.

$$f = \delta(F_1([z^h, z^w])) \tag{3}$$

Afterward, the feature map F undergoes two 1×1 convolution operations, namely, f^h and f^w, to produce F_h (the height feature map) and F_w (the width feature map). Next, g^h (the height attention weight) and g^w (the width attention weight) are generated through activation function calculations, as indicated in Equations (4) and (5).

$$g^h = \sigma(F_h(f^h)) \tag{4}$$

$$g^w = \sigma(F_w(f^w)) \tag{5}$$

Ultimately, the characteristic map of width and height is achieved through multiplication and weight calculations, as illustrated in Formula (6).

$$y_c(i,j) = x_c(i,j) * g_c^h(i) * g_c^w(j) \tag{6}$$

To enhance the effectiveness of location information extraction, the CA attention mechanism is integrated between the C3 layer and the SPPF layer in the backbone network of YOLOv5s. The introduction of the CA coordination attention mechanism facilitates efficient utilization of limited visual information processing resources. This mechanism emphasizes relevant attention candidate regions, while automatically filtering out unessential background and redundant information. As a result, it helps the model to better identify and locate targets. Experimental results demonstrate that of the four attention mechanisms tested, the CA attention mechanism shows the most significant improvement. Therefore, making it more advantageous for the CA attention mechanism.

3.2. Improvement of Loss Function

The loss function defines IOU as the intersection ratio between the predicted frame and the actual frame, which is depicted in Formula (7). Moreover, Formula (8) outlines the definition of the loss function.

$$IOU(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{7}$$

$$L_{IOU} = 1 - IOU \tag{8}$$

The original algorithm of YOLOv5 adopted the CIOU (Complete Intersection Over Union) loss function. CIOU incorporates the influence factor αv to DIOU (Distance Intersection Over Union), which augments the scale loss and the loss of length and width of the detection frame. This effectively results in the prediction frame aligning more closely with the actual frame. The loss function of CIOU is defined in Equations (9)-(12).

$$L_{CIOU} = 1 - IOU + R_{CIOU} \tag{9}$$

$$R_{CIOU} = \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{10}$$

ISSN 2616-5775 Vol. 6, Issue 7: 101-113, DOI: 10.25236/AJCIS.2023.060714

$$\alpha = \frac{v}{(1 - IOU) + v} \tag{11}$$

$$v = \frac{4}{\Pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h})^2 \tag{12}$$

However, CIOU has certain drawbacks and fails to consider the issue of balancing easy and difficult samples. To address this, the WIOU (Wise Intersection Over Union) loss function has been further developed. WIOU assigns different weights to overlapping regions based on their relative positions. Overlapping regions in close proximity to both the center of the predicted model frame and the actual target frame are provided with greater weighting, while those in other areas receive lesser weighting. By addressing the issue of low-quality examples overly emphasizing the bounding box regression, the focus can be shifted towards anchor boxes with ordinary quality to improve overall performance. Equations (13)-(15) provide the definition of the WIOU loss function.

$$L_{WIOU_{v1}} = R_{WIOU} \cdot L_{IOU} \tag{13}$$

$$R_{WIOU} = \exp(\frac{\left(x - x_{gt}\right)^2 + \left(y - y_{gt}\right)^2}{(W_g^2 - H_g^2)}) \tag{14}$$

$$L_{IOU} = 1 - IOU \tag{15}$$

The WIOU loss function has replaced the original CIOU loss function, thereby reducing the impact of incorrect samples on the overall dataset and improving the performance of the object detector. Experimental results demonstrate that compared to other loss functions, the WIOU loss function offers the most significant improvement. Therefore, the integration of the WIOU loss function is considered to be more advantageous.

3.3. Improvement of Activation Function

Activation functions in neural networks play a significant role in determining whether signals should be transmitted and to what extent to the next neuron. Nonlinear activation functions are commonly used in neural networks, including Sigmoid, Tanh, ReLU, LReLU, PReLU, Swish and the Mish functions.

The original YOLOv5 algorithm adopts the SiLU (Sigmoid-Weighted Linear Unit) activation function, expressed as Formula (16), where $\sigma(x)$ represents the Sigmoid activation function displayed in Formula (17). This function enables the network to balance the trade-off between expressiveness and computation efforts.

$$SiLU(x) = x \cdot \sigma(x) = x \cdot sigmoid(x) \tag{16}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{17}$$

MESwish (Memory Efficient Self-Gated Activation Function) is an improved version of the Swish activation function, designed for use in neural networks. MESwish augments the Swish function by adding forward propagation and backward propagation, resulting in reduced redundant computations, enhanced calculation speed, and reduced memory usage. Formula (18) depicts the definition of the Swish activation function, while Formula (19) portrays the MESwish activation function. The hyperparameter β is typically set to 1.15.

$$Swish(x) = x \cdot \sigma(\beta x) = x \cdot sigmoid(\beta x) \tag{18}$$

$$MESwish(x) = x \cdot \frac{1}{1 + e^{-\beta x}} \tag{19}$$

The original YOLOv5 algorithm relied on the SiLU activation function; however, it has been replaced with the MESwish activation function. Compared to SiLU, the gradient of MESwish is continuous, and it addresses the issue of vanishing gradients better while requiring minimal memory usage with high computational efficiency. In experimental studies, the MESwish activation function has been demonstrated to offer the most significant improvement, when compared to other activation functions. Thus, integrating MESwish activation function is considered more advantageous than other alternatives.

ISSN 2616-5775 Vol. 6, Issue 7: 101-113, DOI: 10.25236/AJCIS.2023.060714

3.4. Improvement of Head

In target detection, the original YOLOv5 algorithm employs a Coupled Head to output the classification results and frame position of the targets. The Coupled Head directly forwards the feature map output from the convolution layer to multiple fully connected or convolution layers, resulting in target position and category generation. However, this approach possesses certain drawbacks. It ultimately requires excessive parameters and computational resources, leading to over-fitting issues.

To address these limitations, the Decoupled Head has been introduced to YOLOv5. The Decoupled Head processes target position and category information separately, learns from different network branches, and eventually fuses them to decouple a separate feature channel for positioning and classification tasks. This approach dramatically reduces the parameter quantity and computational complexity, accelerates the convergence speed, improves the detection accuracy, and enhances generalization ability and robustness of the model. The structure diagrams of both Coupled Head and Decoupled Head are illustrated in Figure 3.

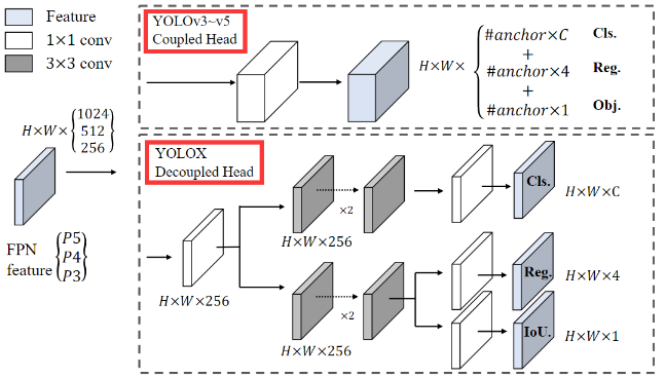


Figure 3: Structure diagram of coupling head and decoupling head

In the figure, it can be observed that the input of the Decoupled Head undergoes a 1x1 convolution to adjust the channel dimension to 256, followed by two parallel 3x3 convolutions for classification and location purposes. Classification is more texture-centric, focusing on the central content area of the target, whereas localization emphasizes the edge information of the target, including predicted bounding box height, width, central coordinates, and confidence score.

Replacing the Coupled Head with the Decoupled Head in YOLOv5's original algorithm has improved classification and location capabilities, avoiding conflicts of feature information, reducing computational overhead, accelerating the network's convergence speed, and achieving faster reasoning speed. Experimental results indicate that the Decoupled Head has the most significant improvement effect compared to other Coupled Heads, making it more advantageous to implement the Decoupled Head.

3.5. Improvement of Spatial Pyramid Pooling

The original algorithm of YOLOv5 uses SPPF (Spatial Pyramid Pooling-Fast) which is an enhancement of SPP (Spatial Pyramid Pooling) and faster than SPP, thus commonly called SPP-Fast. The SPPF structure involves the connection of three 5x5 MaxPool layers in series, equivalent to two 5x5 convolution operations equal to a single 9x9 convolution operation, with three 5x5 convolution operations equal to one 13x13 convolution operation. Structural diagrams of SPP and SPPF are depicted in Figures 4 and 5, respectively.

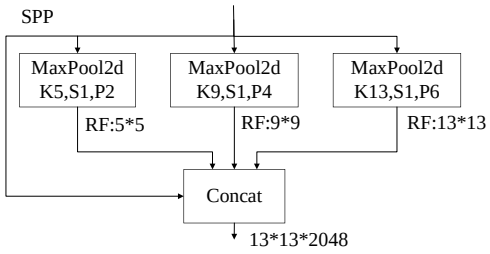


Figure 4: Structure diagram of SPP

ISSN 2616-5775 Vol. 6, Issue 7: 101-113, DOI: 10.25236/AJCIS.2023.060714

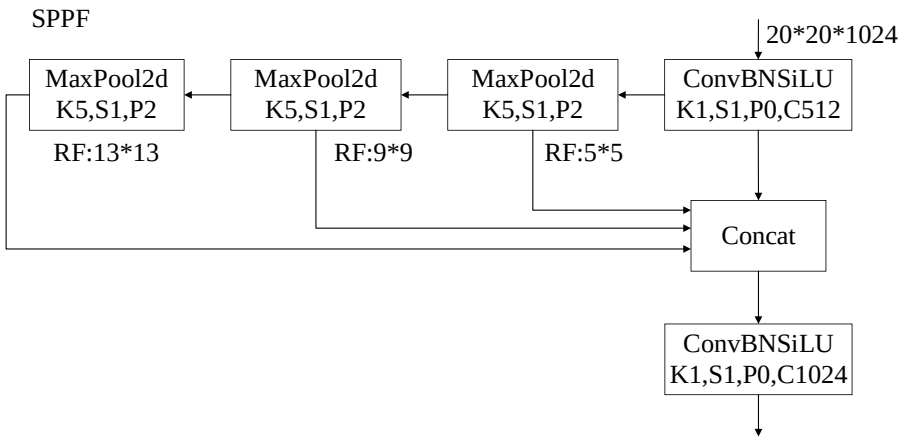


Figure 5: Structure diagram of SPPF

SPPFCSPC (Spatial Pyramid Pooling Fully Connected Spatial Pyramid Convolution) is an improvement over SPPF, which enhances the speed while maintaining the same receptive field. The SPPFCSPC module pools the input feature map into a multi-scale spatial pyramid and enhances model receptive field and feature expression ability. It comprises two sub-modules: SPP module and FCSPC module. The SPP module pools the input feature maps into a multi-scale spatial pyramid, generating several feature maps with various scales. These feature maps capture different target sizes and scene information while improving the model's receptive field and feature expression ability. The FCSPC module convolves the feature map output from the SPP module, further enhancing feature expression capacity. The FCSPC module comprises several convolution kernels of different scales, each with a feature map of distinct scales, capturing target and scene information of various scales.

In YOLOv5's original algorithm, SPPF was used, and it has since been replaced by SPPFCSPC, enhancing the receptive field and feature expression ability of the model through multi-scale spatial pyramid pooling and convolution operations. Experimental results demonstrate that SPPFCSPC is the most effective spatial pyramid pooling technique, making it more advantageous to introduce SPPFCSPC. The structural diagram of SPPFCSPC is depicted in Figure 6.

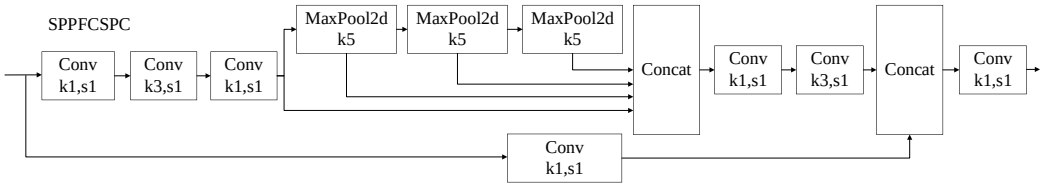


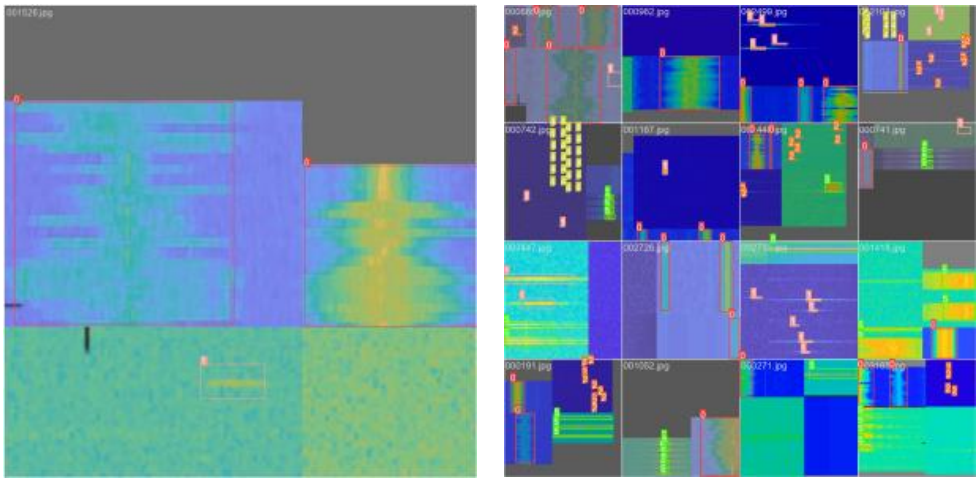
Figure 6: Structure diagram of SPPFCSPC

3.6. Data Enhancement

Data augmentation is an essential technique to enhance network generalization ability, particularly in cases of low data diversity or minimal training sets. Image scaling, translation, flipping, cropping, and mosaicking are common methods of data augmentation. Image scaling and translation improve model recognition ability for objects of varying sizes. Random flipping enhances the model's recognition ability for objects in different orientations. Random cropping enables better recognition of varying sized targets. Mosaic operation randomly selects four images and combines them into one picture, expanding the target dataset, particularly for the small target dataset, significantly enriching the dataset and enhancing network robustness. Mosaic operation also reduces GPU video memory usage, allowing for larger batch processing and faster training speed.

These data augmentation methods enhance data diversity, model robustness, and generalization ability, improving model performance and accuracy. However, YOLOv5's original Mosaic operation randomly crops images, resulting in excessive cropping and the loss of valuable image information. Thus, the method of converting the original 4-picture mosaic to a 16-picture mosaic was adopted to reduce the probability of detection targets being cut out and improve model detection performance. Figure 7 shows a comparison chart before and after data augmentation and improvement.

ISSN 2616-5775 Vol. 6, Issue 7: 101-113, DOI: 10.25236/AJCIS.2023.060714



(a) Original image *(b) Improved image*

Figure 7: Comparison chart before and after data enhancement and improvement

Experimental results indicate that data augmentation can significantly improve performance, demonstrating its advantage for model improvement through data enhancement.

4. Experiment and Result Analysis

4.1. Experimental Platform

This experiment was conducted using the Windows10 operating system and PyCharm software. The experiment used the PyTorch1.8.0 framework and CUDA 11.1 environment and utilized an RTX3060 (8 GB memory) GPU for model training and acceleration. In the experiment, the lightest version of YOLOv5s was employed, and an improvement was made on it. The input size was set to (640,640), with a batch size set to 16, and the iteration number (epoch) was set to 50.

4.2. Spectrum Waterfall Dataset

Deep-learning-based signal detection requires tagged training data, and the spectrum waterfall chart provides essential information for the network to recognize signals. By selecting the target signal's frame in the waterfall chart, the occupation of this signal in both time and frequency domains can be directly given, providing useful information in two dimensions and achieving high processing speed and real-time performance.

In this experiment, various signals were selected and used to construct a training database for the waterfall chart. The signals in the waterfall chart were manually calibrated. The original dataset in this work was obtained from the Spectrum Waterfall Chart Dataset on GitHub. An effective dataset was generated by data augmentation and signal classification, which was then divided into training and test sets at a ratio of 8:2 to provide data support for signal detection. The downloaded spectrum waterfall dataset was preprocessed as follows. Firstly, data augmentation techniques such as random rotation, translation, scaling, cropping, and Mosaic were applied to the existing signal waterfalls to generate the waterfall dataset. Secondly, the labeled waterfall charts were classified into six typical signal spectrum data types, such as Bluetooth radio signals, 2.4GHz wireless telephone signals, 312-512MHz ultrashort wave signals, microwave oven electromagnetic interference signals, Xbox360 remote control signals, FM broadcast long-lasting signals, etc. Finally, the labeled waterfall maps were divided into training and test sets at a ratio of 8:2 to generate an effective dataset.

4.3. Evaluating Indicator

The evaluation metrics used in this experiment include P (Precision rate), R (Recall rate), mAP (mean Average Precision), and the PR (Precision-Recall) curve. Precision, also called positive predictive value, is the proportion of correctly predicted positive samples to all samples predicted as positive. The positive prediction samples consist of TP (True Positive) and FP (False Positive). The formula for P is given in Equation (20), which captures the ratio of true positive samples to the total number of predicted positive

ISSN 2616-5775 Vol. 6, Issue 7: 101-113, DOI: 10.25236/AJCIS.2023.060714

samples.

$$P = \frac{TP}{TP + FP} \tag{20}$$

Recall rate, also referred to as sensitivity, captures the ratio of true positive samples to all samples that are actually positive. The original positive samples include TP (True Positive) and FN (False Negative). The formula for recall rate is defined in Equation (21), which measures the ability to identify the actual positive samples among all positive samples in the dataset.

$$R = \frac{TP}{TP + FN} \tag{21}$$

Mean Average Precision measures the efficacy of a detection algorithm in recognizing various types of targets. A higher mAP value indicates better overall detection performance of the algorithm. mAP0.5 represents the average category precision in detecting targets with an IOU (intersection-over-union) threshold of less than 0.5, while mAP0.5:0.95 represents the average category precision of all targets in the IOU range of 0.5 to 0.95. Here, APC denotes the AP (Average Precision) value of a specific category, N is the total number of categories that are classified, and the definition of mAP is given in Equation (22).

$$mAP = \frac{\sum_0^N APC}{Original\ positive\ class} \tag{22}$$

The vertical axis of the PR curve represents precision, while the horizontal axis represents recall. The closer the curve is to the top-right corner, the better the detection effect of the model and the overall performance of the algorithm. Therefore, the goal is to achieve a PR curve that is as close as possible to the top-right corner in order to achieve optimal detection performance and algorithmic efficiency.

4.4. Analysis of Experimental Results

4.4.1. Experimental Results of Improved Model Before and After

The experiment was conducted in a consistent experimental environment and under the same parameters. The experimental data for both YOLOv5s algorithm and its improved version were analyzed and compared. As presented in Table 1, the improved algorithm outperforms the original algorithm in terms of accuracy, recall, mAP0.5, and mAP0.5:0.95 with improvements of 10%, 21.8%, 7.9%, and 10.1%, respectively. Therefore, the improved YOLOv5s algorithm demonstrates better detection performance compared to its original version.

Table 1: Experimental data comparison table

Algorithm model	P	R	mAP0.5	mAP0.5:0.95
YOLOv5	0.785	0.663	0.822	0.449
YOLOv5- improvement	0.868	0.881	0.901	0.55

Figure 8 depicts the comparison chart of the mAP before and after the improvement.

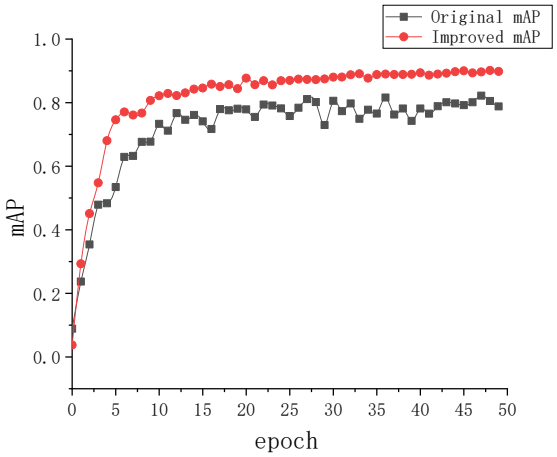
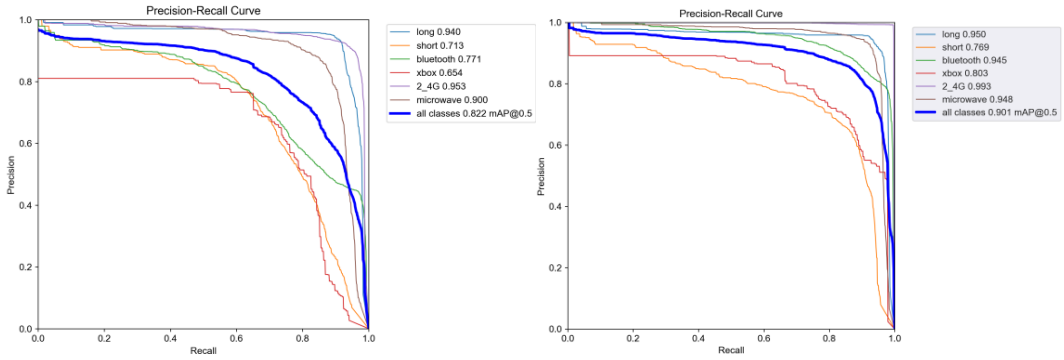


Figure 8: Comparison chart before and after mAP improvement

ISSN 2616-5775 Vol. 6, Issue 7: 101-113, DOI: 10.25236/AJCIS.2023.060714

Figures 9(a) and 9(b) respectively illustrate the PR curves before and after the improvement. The curves of different colors on each PR graph correspond to different signal categories, and the mAP is obtained by calculating the curve and the area below it. Furthermore, Figure 10 displays the detection effect diagram for the model dataset.



(a) Original PR curve *(b) Improved PR curve*

Figure 9: PR curve before and after improvement

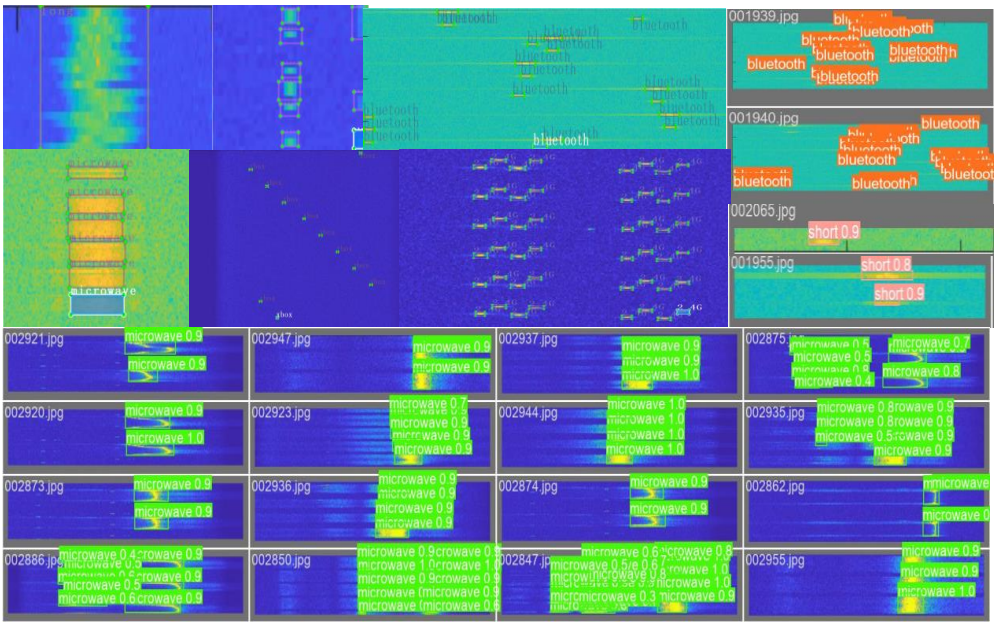


Figure 10: Detection effect diagram of the model dataset

4.4.2. Analysis of Experimental Results of Introducing Attention Mechanism

Table 2 presents a comparison between the CA attention mechanism and three commonly used attention mechanisms. The CA mechanism is introduced and evaluated in this study.

Table 2: Verification experiment of introducing CA attention mechanism

Algorithm model	P	R	mAP0.5	mAP0.5:0.95
YOLO5s	0.785	0.663	0.822	0.449
YOLO5s-CA	0.801	0.814	0.843	0.457
YOLO5s-SE	0.743	0.782	0.819	0.447
YOLO5s-CBAM	0.671	0.833	0.833	0.453
YOLO5s-ECA	0.823	0.807	0.837	0.451

The results in the table indicate that the introduction of the CA mechanism has the most significant promotion effect on the mAP. Thus, introducing the CA attention mechanism is advantageous.

4.4.3. Analysis of Experimental Results with the Introduction of Loss Functions.

Table 3 presents the experimental results obtained by introducing the WIOU loss function and comparing it with other commonly used loss function modules.

ISSN 2616-5775 Vol. 6, Issue 7: 101-113, DOI: 10.25236/AJCIS.2023.060714

Table 3: Verification experiment by introducing WIOU loss function

Algorithm model	P	R	mAP0.5	mAP0.5:0.95
YOLO5s	0.785	0.663	0.822	0.449
YOLO5s-WIOU	0.859	0.865	0.879	0.509
YOLO5s-DIOU	0.845	0.865	0.876	0.506
YOLO5s-EIOU	0.839	0.862	0.875	0.511
YOLO5s- GIOU	0.841	0.863	0.877	0.507

The table indicates that all four loss functions result in a certain improvement in mAP. However, the WIOU loss function shows the most notable effect. Therefore, introducing the WIOU loss function in this model is more advantageous.

4.4.4. Analysis of Experimental Results with the Introduction of Activation Functions.

Table 4 displays the experimental results obtained by introducing the MESwish activation function and comparing it with other commonly used activation functions.

Table 4: Verification experiment by introducing MESwish activation function

Algorithm model	P	R	mAP0.5	mAP0.5:0.95
YOLO5s	0.785	0.663	0.822	0.449
YOLO5s-MESwish	0.836	0.87	0.874	0.51
YOLO5s-FReLU	0.837	0.844	0.859	0.489
YOLO5s-HardSwish	0.85	0.854	0.873	0.503

The table illustrates that all three activation functions result in an enhancement of mAP. However, the MESwish activation function demonstrates the most noticeable effect. Therefore, introducing the MESwish activation function is more advantageous for this model.

4.4.5. Analysis of Experimental Results of Introducing Detection Head

Table 5 displays the experimental results obtained by introducing Decoupled Detect and comparing it with other commonly used detectors.

Table 5: Verification experiment by introducing decoupling head

Algorithm model	P	R	mAP0.5	mAP0.5:0.95
YOLO5s	0.785	0.663	0.822	0.449
YOLO5s-Decoupled Detect	0.845	0.867	0.879	0.515
YOLO5s-ASFF Detect	0.855	0.852	0.873	0.509

The table reveals that both types of header improvements lead to some improvement in mAP. However, the Decoupled Detect header, which decouples the regression and classification tasks, displays the most prominent improvement effect. Thus, introducing Decoupled Detect in this model is more advantageous.

4.4.6. Analysis of Experimental Results by Introducing Spatial Pyramid Pooling

In Table 6, we present the experimental results of comparing SPPFCSPC spatial pyramid pooling with other commonly used spatial pyramid pooling techniques. The introduction of SPPFCSPC has been thoroughly evaluated and analyzed in this study.

Table 6: Introducing SPPFCSPC verification experiment

Algorithm model	P	R	mAP0.5	mAP0.5:0.95
YOLO5s	0.785	0.663	0.822	0.449
YOLO5s- SPPFCSPC	0.845	0.857	0.878	0.507
YOLO5s-SPPCSPC	0.855	0.855	0.874	0.508
YOLO5s-SimSPPF	0.856	0.851	0.876	0.507

The table clearly indicates that all three spatial pyramid pooling approaches lead to some improvement in mAP scores. However, the highest degree of improvement is observed in the case of SPPFCSPC. Thus, incorporating SPPFCSPC in this model provides the most significant advantage over the other methods.

4.4.7. Analysis of Experimental Results by Introducing Data Enhancement

We compared the original YOLO5s model with its variant that incorporates data augmentation. Table 7 presents the experimental results of this comparison.

ISSN 2616-5775 Vol. 6, Issue 7: 101-113, DOI: 10.25236/AJCIS.2023.060714

Table 7: Introducing data enhancement verification experiment

Algorithm model	P	R	mAP0.5	mAP0.5:0.95
YOLO5s	0.785	0.663	0.822	0.449
YOLO5s-Data Enhancement	0.821	0.869	0.874	0.509

Table 7 clearly indicates that data augmentation leads to a noticeable improvement in mAP scores. Thus, incorporating this technique in this model provides a significant advantage.

4.5. Ablation Experiment

This study introduces six proposed improved techniques, namely C (CA attention mechanism), W (WIOU loss function), M (MESwish activation function), D (Decoupled detection), S (SPPFCSPC) and Q (data enhancement). The experiment is conducted in two aspects. Firstly, only one module is added onto the YOLOv5s model at a time to verify its effectiveness. Secondly, one module is added sequentially until all the modules are accumulated and the overall effect of each improved method is evaluated. The experimental results are presented in Table 8.

Table 8: Comparison table of ablation experimental data

Algorithm model	C	W	M	D	S	Q	P	R	mAP0.5	mAP0.5:0.95
YOLOv5							0.785	0.663	0.822	0.449
YOLOv5-C	√						0.801	0.814	0.843	0.457
YOLOv5-W		√					0.859	0.865	0.879	0.509
YOLOv5-M			√				0.836	0.87	0.874	0.51
YOLOv5-D				√			0.845	0.867	0.879	0.515
YOLOv5-S					√		0.845	0.857	0.878	0.507
YOLOv5-Q						√	0.821	0.869	0.874	0.509
YOLOv5-CW	√	√					0.852	0.884	0.885	0.528
YOLOv5-CWM	√	√	√				0.852	0.884	0.891	0.53
YOLOv5-CWMD	√	√	√	√			0.857	0.886	0.9	0.545
YOLOv5-CWMDS	√	√	√	√	√		0.853	0.889	0.901	0.546
YOLOv5-CWMDSQ	√	√	√	√	√	√	0.857	0.889	0.901	0.551

Table 8 reveals that adding each of the six proposed techniques onto the YOLOv5s model leads to an increase in mAP, accuracy P, and recall rate R. Specifically, when only one method is added at a time, there is an observable improvement in the performance metrics. Additionally, when modules are added incrementally, the performance improves incrementally as well.

5. Conclusions

Traditional electric signal detection suffers from slow recognition speed and inaccurate recognition effects. Deep learning-based signal detection and recognition methods can overcome these issues and improve detection accuracy and real-time performance. This study proposes several improvements to the YOLOv5s model to enhance its performance. Firstly, the CA attention mechanism is introduced, followed by the replacement of the original CIOU loss function with the WIOU loss function, and the replacement of the SiLU activation function with the MESwish activation function. Next, the Decoupled Head decoupling detector is used for target prediction, and a multi-scale spatial pyramid pooling SPPFCSPC module is added. Finally, data diversity is enhanced through data augmentation, which increases the model's robustness and generalization ability, leading to improved performance and accuracy. The experimental results demonstrate that the proposed approach, YOLOv5s-CWMDSQ, improves accuracy, recall, mAP0.5, and mAP0.5:0.95 by 10%, 21.8%, 7.9%, and 10.1%, respectively. These improvements confirm the effectiveness and real-time performance of this approach for signal detection and classification.

References

[1] Zhao Q, Mao X, Pang K. Development status and trend of civilian aircraft air-to-ground broadband communication system (in Chinese) [J]. Communications Technology, 2019, 52(10): 2428-2432.
[2] Wang W. Design and implementation of transmit and receive channels for broadband signal processing module (in Chinese) [J]. Communications Technology, 2017, 50(07): 1560-1563.
[3] Dobre, Octavia A. Signal Identification for Emerging Intelligent Radios: Classical Problems and New Challenges [J]. IEEE instrumentation & measurement magazine, 2015, 18(2):11-18.

ISSN 2616-5775 Vol. 6, Issue 7: 101-113, DOI: 10.25236/AJCIS.2023.060714

[4] Tang L. Overview of FPGA-based broadband signal detection design (in Chinese) [J]. Science and Informationization, 2021(12): 13-15.
[5] Wang H, Tao M, Zhao H, Jiang C. Simulation and implementation of multi-antenna signal detection based on FPGA and DSP (in Chinese) [J]. TV Technology, 2014, 38(15): 76-79. DOI: 10. 16280/j. videoe. 2014. 15. 023.
[6] Lei Z , Jiang M , Yang G ,et al. Towards recurrent neural network with multi-path feature fusion for signal modulation recognition[J].Wireless Networks, 2022.DOI:10.1007/s11276-021-02877-8.
[7] Lin Lin, Xuezhi He, Jinbao Xie. Radio Signals Modulation Mode Recognition Based on Semisupervised Deep Learning[C]. //2018 3rd International Conference on Computer Science and Information Engineering (ICCSIE 2018). Proceedings of the 2018 3rd International Conference on Computer Science and Information Engineering (ICCSIE 2018). 2018: 299-305.
[8] Zeng Y , Zhang M , Han F , et al. Spectrum Analysis and Convolutional Neural Network for Automatic Modulation Recognition [J]. IEEE wireless communications letters, 2019, 8(3): 929-932. DOI:10. 1109/LWC. 2019. 2900247.
[9] Qian Mao, Fei Hu, Qi Hao. Deep Learning for Intelligent Wireless Networks: A Comprehensive Survey[J]. Communications surveys & tutorials, 2018, 20(4):2595-2621. DOI: 10. 1109/COMST. 2018. 2846401.
[10] Ya Tu, Yun Lin, Haoran Zha, Ju Zhang, Yu Wang, Guan Gui, Shiwen Mao. Large-scale real-world radio signal recognition with deep learning[J]. Chinese Journal of Aeronautics, 2022, 35(09):35-48.
[11] J. L. Ziegler, R. T. Arn and W. Chambers, "Modulation recognition with GNU radio, keras, and HackRF," 2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Baltimore, MD, USA, 2017.
[12] Parnas, David Lorge. The Real Risks of Artificial Intelligence Incidents from the early days of AI research are instructive in the current AI environment [J]. Communications of the ACM, 2017, 60(10):27-31.
[13] O'Shea T J, Corgan J, Clancy T C. Convolutional Radio Modulation Recognition Networks[J]. Springer, Cham, 2016.
[14] O'Shea T J, Corgan J, Clancy T C. Unsupervised representation learning of structured radio communication signals [J]. IEEE, 2016.
[15] O'Shea T J, West N, Vondal M, et al. Semi-supervised radio signal identification[J]. IEEE, 2017.
[16] O'Shea T J, West N. Radio Machine Learning Dataset Generation with GNU Radio[C]// Proceedings of the GNU Radio Conference. 2016.
[17] Kulin M, Kazaz T, Moerman I, et al. End-to-end Learning from Spectrum Data: A Deep Learning approach for Wireless Signal Identification in Spectrum Monitoring applications [J]. IEEE Access, 2018, 6:18484-18501.
[18] Selim A, Paisana F, Arokkiam J A, et al. Spectrum Monitoring for Radar Bands using Deep Convolutional Neural Networks[J]. arXiv e-prints, 2017.
[19] Maglogiannis V, Shahid A, Naudts D, et al. Enhancing the Coexistence of LTE and Wi-Fi in Unlicensed Spectrum Through Convolutional Neural Networks [J]. IEEE Access, 2019, 7:28464-28477.
[20] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[J]. IEEE Computer Society, 2014.
[21] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger [C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2017:6517-6525.
[22] Wei L, Dragomir A, Dumitru E, et al. SSD: Single Shot MultiBox Detector[J]. Springer, Cham, 2016.
[23] Li X, Chen T. Radio frequency spectrum artificial intelligence recognition technology based on TensorFlow (in Chinese) [J]. China Radio, 2018(9): 55-56, 60. DOI:10. 3969/j. issn. 1672-7797. 2018. 09. 047.
[24] Zhou X, He X, Zheng C. Radio signal identification based on image deep learning (in Chinese) [J]. Journal of Communications, 2019, 40(7): 114-125. DOI:10. 11959/j. issn. 1000-436x. 2019167.
[25] Zhou Y, Hou J, Li J, et al. Spectrum signal identification based on frequency domain stacking and deep learning (in Chinese) [J]. Journal of Computer Applications Research, 2023, 40(3): 874-879. DOI:10. 19734/j. issn. 1001-3695. 2022. 07. 0586.