IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 May 2023, accepted 24 May 2023, date of publication 31 May 2023, date of current version 7 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3281551

RESEARCH ARTICLE

An Improved Traffic Sign Detection and Recognition Deep Model Based on YOLOv5

QIANYING WANG[1], XIANGYU LI[1], AND MING LU[2]
[1]College of Mathematics and Statistics, Hebei University of Economics and Business, Shijiazhuang 050061, China
[2]School of Mathematical Sciences, Hebei Normal University, Shijiazhuang 050024, China

Corresponding author: Ming Lu (luming5@mail2.sysu.edu.cn)

This work was supported in part by the Science and Technology Project of the Hebei Education Department under Grant ZD2021043, in part by the Hebei University of Economics and Business Foundation under Grant 2020YB13 and Grant 2020YB01, and in part by the Hebei Provincial Department of Human Resources and Social Security '333 Talent Project' under Grant A202101014.

ABSTRACT In this paper, we aim at the traffic sign detection and recognition in complex road conditions. We proposed a deep model for traffic sign detection and recognition. There are a few difficulties in traffic sign detection task, such as, less recognizable, small target size, easily leading to detected failure and so on. First, for the failure detection, we introduce Coordinate Attention (CA); second, to accelerate the regression of prediction box, we introduce the angle loss into our objective function; third, for the overlapping and occlusion phenomenon of ground truth, a dynamic label assignment strategy- simple Optimal Transport Assignment (SimOTA) is utilized during label assignment; the last and the most important, for the target size problem, we propose a feature fusion network, named hierarchical-path feature fusion network (H-PFANet). Experiments were conducted on two public data sets, the results show that our improved model performed better than YOLOv5s which is the base model and other popular algorithms on precision, recall and mAP. For the difficult samples in data set CCTSDB-2021, the results show that compared to YOLOv5s, the mAP@0.5 is improved by 6.3%, the mAP@0.5 : 0.95 is improved by 5.3%, and our method achieved a detection speed of 91 FPS, with better robustness to changes in various traffic scenes, while maintaining the volume of the original YOLOv5s model. On the whole CCTSDB-2021 data set, the precision of our model reached 98.1%, the recall of our model reached 97.6% and the mAP@0.5 reached 98.8% with a speed of 91 FPS. We also compared our method with other current detection algorithms on TT100K data set, the results show that our proposed method performed better, and show the effectiveness of our method.

INDEX TERMS Attention mechanism, dynamic label assignment strategy, feature fusion, traffic sign detection, YOLOv5.

I. INTRODUCTION

With the rapid development of social economy and information technology, unmanned driving technology has also made rapid progress. The core of the driverless system can be summarized into three parts: perception, planning and control. Perception is the first and the most important part, it mainly refers to driverless system gathering information from the environment, and then extracting related knowledge. During the process, environmental awareness refers to the ability to understand the environment, such as, obstacle identification, traffic sign detection and traffic signal recognition, pedestrian and vehicle detection and so on.

As an important part of traffic system, traffic signs play an important role in regulating, directing and controlling the flow and direction of vehicles. They also prompt road conditions, prevent traffic accidents, and play an important role in ensuring people's travel and vehicle driving safety. As an important part of driverless system, traffic sign recognition attracts more and more attention from researchers. At present, traffic sign detection methods are mainly divided into two categories. One is the traditional detection method, which mainly extracts information through color, edge information, picture shape, etc., and then combines machine learning

The associate editor coordinating the review of this manuscript and approving it for publication was Jon Atli Benediktsson.

methods to detect. The other is the popular deep learning detection method. Loy and Barnes [1] established possible centroid position through image symmetry and edge information to detect traffic signs. Its disadvantage is that it is not applicable to all shapes and has poor generalization. Bascon et al. [2] first segmented the pixel color in the image, and then used the support vector machine to classify and detect the shape. Although it can cover all colors and shapes, it is inefficient. Those two methods belong to the first category. In recent years, with the emergence of deep learning, target detection task has made a major breakthrough. The research is mainly divided into two directions, one is the two-stage method, such as, R-CNN [3], Fast R-CNN [4], Faster R-CNN [5] and so on. Those series methods are with high recognition precision, but are slow for detection. Another category is the one-stage method. Single Shot MultiBox Detector (SSD) [6] and You Only Look Once (YOLO) series [7], [8] [9], [10] belong to this category. Those methods are fast, but their recognition accuracy is a little lower than the two-stage methods. With the development of deep model, more efficient methods are proposed, such as YOLOv5. Those methods are not only fast but also with good recognition results. Wang et al. [11] proposed a cascade mask generation framework. The proposed framework takes multi-scale images as input and processes them in ascending order of the scale to deal with the detection of small objects with low resolution.

For traffic sign detection, most of the research deals with simple traffic scenes at present, which cannot meet the practical requirements. To our best knowledge, only a little research is on the complex scene recognition, but it is for special scene, and the ability of generalization is weak. Dong et al. [12] first used the wavelet decomposition to reduce the influence of rain and snow on recognition task, and then used the improved YOLOv3 to recognize the traffic signs. In order to enhance the detection effect in weak light scene, Zhao et al. [13] first enhanced the brightness and contrast ratio of the original image to increase the difference between the traffic signs and the background, and then used the deep learning algorithm for recognition. The detection effect is improved, but the time complexity is a little high. Obviously, the methods mentioned above cannot well adapt to various changes in natural scenes, and the recognition accuracy and detection speed still need to improve.

Complicate environment, for example, the bad weather, illumination, occlusion, small target and so on, will cause low recognition or undetection problems in traffic signs recognition. In order to deal with those problems, we proposed a traffic sign detection and recognition deep model based on YOLOv5. The contribution of this manuscript mainly includes four aspects:

- We introduced the Coordinate Attention (CA) at the end of the backbone. This can deal with the other interferences in complex background and increase the attention of the model to important features.
- We improved the regression loss function with the introduction of the vector angle between the regression boxes. In this way, it will reduce the freedom of the prediction box in the process of convergence, accelerate the network convergence, and improve the detection effect.
- We also improved the label assignment strategy. The dynamic label assignment strategy was used in positive and negative sample selection. This can accelerate the network optimization, alleviate the problems of dense object distribution and serious occlusion in complex environment.
- For the target size problem, we proposed a feature fusion network named hierarchical-path feature aggregation network (H-PFANet). Two different information fusion strategies were designed and added between the backbone network and the deep network to alleviate the pixel loss caused by small targets as the convolution deepens, and increase the recognition ability of small targets in complex background.

Section II shows the construction of the based YOLOv5 network; section III introduces our proposed network; section IV shows the experiments and section V concludes the paper.

II. INTRODUCTION OF OUR BASED YOLOv5

YOLOv5 is a new single-stage target detector opened by Ultralytics in May 2020. It integrates many advanced achievements. What we used in this manuscript is the lasted vision 6.0. It consists of four parts, from the smallest to largest are: YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x. YOLOv5s is the smallest and YOLOv5x is the largest, the larger is the model, the more are the parameters, the more complex is the model and the higher is the precision. In this manuscript, our goal is to develop a light model, so our model is based on YOLOv5s.

The structure of YOLOv5s is shown in figure 1, it mainly consists of four parts: input, backbone, neck and output. Input mainly refers to the reprocessing, including Mosaic4 data enhancement, K-Means clustering to generate anchor frame and image scaling and so on. Compared to the former vision, there are some changes in the backbone in vision 6.0. First, 6×6 convolution layer replaces the previous Focus module on the first layer of the network for down-sampling operation. Second, the previous SPP layer was replaced by SPPF layer. The previous SPP layer consists of three pooling layers with size $5 \times 5, 9 \times 9, 13 \times 13$ separately, which are organized in parallel. The SPPF is organized by three pooling layers with the same size 5×5, which are connected in series. In this way, the SPPF can achieve the same effect as SPP, but its speed is the twice of SPP. In addition, the CBS module and C3_1 module are included in the backbone, where CBS module encapsulates the convolution layer, batch processing layer and activation function. Neck module is constructed based on PANnet feature fusion net in

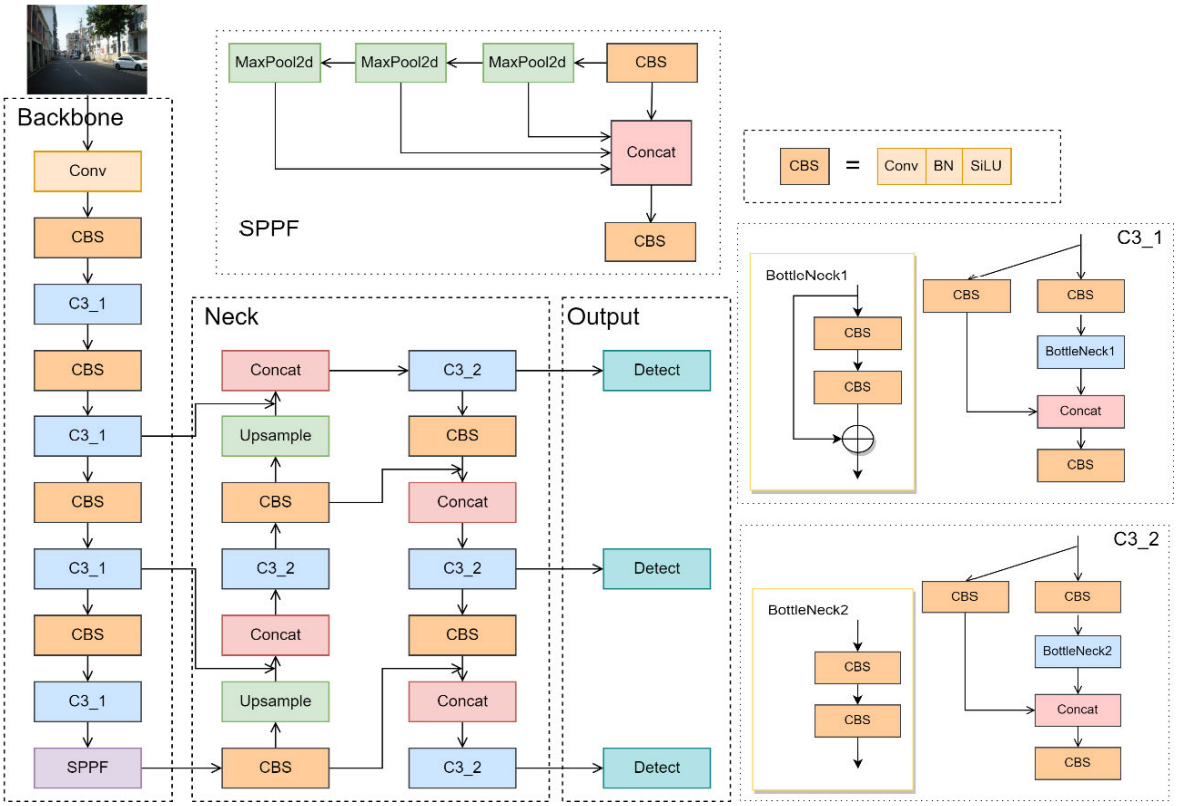


FIGURE 1. The construction of YOLOv5s.

FPN, which can strengthen information dissemination. The difference between C3_2 module and C3_1 module is in the Bottleneck. C3_1 module contains a residual link while C3_2 does not contain. The difference can be found in figure 1. Finally, the output uses CIoU to compute regression loss of bounding box, and predict for the three images' feature with different sizes.

III. THE LIGHTWEIGHT TRAFFIC SIGN DETECTION AND RECOGNITION DEEP MODEL

In this section, we will introduce our traffic sign detection and recognition deep model. First, subsection III-A will state the coordinate attention (CA); subsection III-B describes the loss function; subsection III-C shows the dynamic label assignment strategy-simple Optimal Transport Assignment (SimOTA); and the last subsection III-D introduces the hierarchical-path feature fusion network (H-PFANet).

A. COORDINATE ATTENTION (CA)

Attention mechanism is a data processing method in machine learning, which can significantly improve the feature extraction ability of neural network, and is widely used in various types of machine learning tasks such as natural language processing, computer vision and speech recognition.

Presently, the widely used attention are Squeeze-and-Excitation attention (SE) [14] and Convolutional Block Attention Module (CBAM) [15]. In SE, the channel weights are only determined by the relationship between the channels, but the spatial structure and location information are not considered. CBAM connected the channel attention and spatial attention by series, and tried to decrease the number of channels to extract the location information, but the local information extracted by convolution does not offer long-range dependence.

In order to deal with the weakness mentioned above, coordinate attention (CA) is proposed in [16]. Because 2-dimension global pooling causes the loss of location information, CA decomposes the channel attention into two 1-dimension feature code aggregated in different directions. Then one direction can be used to retain the long term information, and the other direction is used to capture the location information. Then coding the image feature to obtain direction and location sensitive image feature. In this way, those information can be embedded into image feature to enhance the interesting target representation.

As figure 2 shows, the process of coordination attention can be completed in two steps: coordinate information embedding and coordinate attention generating.

(1) Coordination information embedding

In order to capture the remote spatial interaction with accurate location information, the 2-dimension global pooling is decomposed to equation (1), and converted to one to one

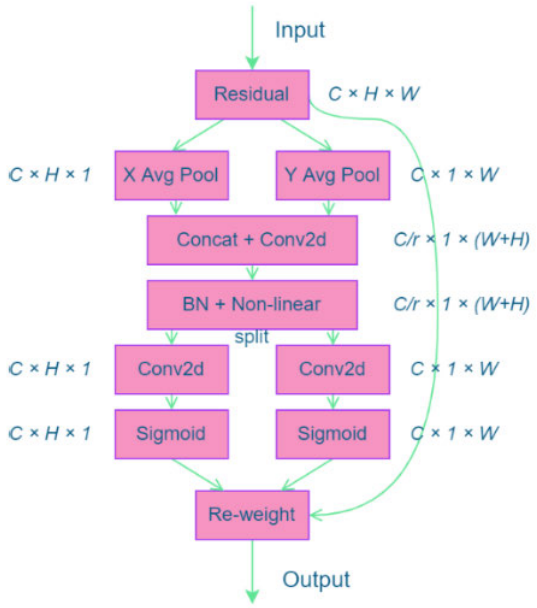


FIGURE 2. The construction of coordinate attention.

coding operation.

$$Z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1} W x_c(i, j) \tag{1}$$

For a given input, use the pooling kernel of size $(H, 1)$ or $(1, W)$ to coding every channel by horizontal and vertical coordinates separately. Then capture the image feature of height and width by equation (2) and (3).

$$Z_c^h(h) = \frac{1}{W} \sum_{0 \le i \le W} X_c(h, i) \tag{2}$$

$$Z_c^w(w) = \frac{1}{H} \sum_{0 \le j \le H} X_c(j, w) \tag{3}$$

The two transformations aggregate features along two spatial directions, so the module can obtain long-range dependence and accurate location information.

(2) Coordinate attention generating

Concatenate the result of equation (2) and (3), then transform with a convolution function F_1 and activate by δ, where δ is a nonlinear activation function: $f = \delta(F_1([z^h, z^w]))$. After batch normalization and nonlinear transformation, decompose f into two independent tensors along the spatial dimension to get f^h and f^w, then use two convolutions of size 1×1 to sample them up r times and get:

$$g^h = \sigma(F_h(f^h))$$
$$g^w = \sigma(F_w(f^w))$$

Finally, the output of CA (figure 2) can be represented: $y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j)$.

Experiments show that the introduced CA module enhances the network's ability to accurately locate the target, increases the model's attention to important features, and significantly improves the model detection effect. In addition, the CA module has only a small number of parameters, which is very lightweight, and hardly brings extra computing overhead.

B. LOSS FUNCTION

The effectiveness of target detection depends largely on the loss function. From IoU in the first generation of YOLO to CIoU [17] in the latest version of YOLOv5, the frame regression loss has always been an important part of YOLO series loss function. Based on IoU, GIoU [18] added non-crossing area proportion and got the deviation trend measurement capability. DIoU [17] added a penalty term of center point distance proportion based on GIoU, so DIoU can better measure the distance between the center point of the predicted border and the ground truth. CIoU added a penalty term of aspect ratio on the basis of DIoU. When the center point of the prediction border coincides with the ground truth, it has a better effect on aspect fitting. To our best knowledge, the present methods did not take into account the direction of the mismatch between the required ground truth and the prediction box. This weakness will lead to slower convergence and lower efficiency, because the prediction box may 'wander around' during the training process and eventually produce a worse model.

In order to deal with the weakness, this manuscript replaces CIoU in YOLOv5 with SIoU [19]. SIoU considers the vector angle between the regression boxes and redefines the penalty. The SIoU loss function consists of four parts: angle loss, distance loss, shape loss and IoU loss.

1) ANGLE LOSS

The angle loss component attempts to bring the prediction to X axis or Y axis (whichever is closer), so that the center point of the prediction box and the ground truth are in the same horizontal or vertical direction, and then continues to converge along the relevant axis. To achieve this, the convergence process needs to minimize α first if $\alpha < \frac{\pi}{4}$, otherwise minimize $\beta = \frac{\pi}{2} - \alpha$, where α is the angle between the prediction center and the ground truth center (please refer to figure 3). As figure 3 shows, σ is the Euclidean distance between the prediction box center and the ground truth center, c_h is the absolute value of the difference between the two centers along the vertical axis coordinates. The angle loss will minimize the angle between the prediction box center and the ground truth center, and the angle loss is defined as follows:

$$\Lambda = 1 - 2sin^2(arcsin(x) - \frac{\pi}{4})$$

where

$$x = \frac{c_h}{\sigma} = sin(\alpha)$$
$$\sigma = \sqrt{(b_{c_x}^{gt} - b_{c_x})^2 + (b_{c_y}^{gt} - b_{c_y})^2}$$
$$c_h = max(b_{c_y}^{gt}, b_{c_y}) - min(b_{c_y}^{gt}, b_{c_y})$$

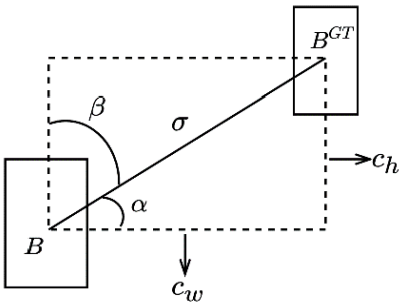


FIGURE 3. Angle Loss calculation.

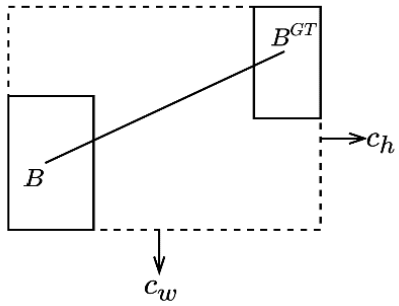


FIGURE 4. Distance Loss calculation.

2) DISTANCE LOSS

Considering the introduced angle loss, the distance loss is redefined as follows:

$$\Delta = \Sigma_{t=x,y}(1 - e^{\gamma\rho_t})$$

where $\rho_x = (\frac{b^{gt}_{c_x} - b_{c_x}}{c_w})^2$, $\rho_y = (\frac{b^{gt}_{c_y} - b_{c_y}}{c_h})^2$, $\gamma = 2 - \Lambda$. As figure 4 shows, c_h and c_w are the height and width of the minimum circumscribed rectangle of the ground truth and the prediction box. Δ will be smaller when α goes to 0, then the contribution of distance loss is greatly reduced. Otherwise, when α goes to $\frac{\pi}{4}$, Δ will be larger.

3) SHAPE LOSS

The shape loss is defined as: $\Omega = \Sigma_{t=w,h}(1 - e^{-\omega_t})^{\theta}$, where $\omega_w = \frac{|w - w^{gt}|}{max(w, w^{gt})}$, $\omega_h = \frac{|h - h^{gt}|}{max(h, h^{gt})}$; θ is a hyper-parameter, which is the weight of shape loss; w, w^{gt}, h, h^{gt} are the width and height of the prediction box and ground truth respectively. Compared to EIoU [20], SIoU does not need to calculate the width and height of the minimum circumscribed rectangle of the ground truth and the prediction box, and only needs the width and height of the ground truth and the prediction box.

Finally, the loss of SIoU is defined as follows:

$$L_{box} = 1 - IoU + \frac{\Delta + \Omega}{2}$$

where $IoU = \frac{|B \bigcap B^{GT}|}{|B \bigcup B^{GT}|}$.

The angle loss component added in SIoU effectively reduces the degree of freedom of frame regression in the training process, accelerates the network convergence, and further improves the regression accuracy.

C. DYNAMIC LABEL ASSIGNMENT STRATEGY–SIMPLE OPTIMAL TRANSPORT ASSIGNMENT (SimOTA)

In recent years, the search for more advanced tag allocation strategies has become an important research direction in the field of detection. In previous studies, researchers mostly decide the fate of the prediction box based on a fixed threshold (the intersection and combination ratio of the prediction box and the ground truth), and discard all those below the fixed threshold. This static label assignment strategy may cause some useful prediction boxes to be unable to be assigned to truth labels in some scenarios. For example, when two positive samples overlap and occlude seriously, they may only be assigned to one positive label. Optimal Transport Assignment (OTA) algorithm published by Kuangsi Technology Team in CVPR in 2021 provides a new perspective for optimizing label assignment [22]. In OTA, the authors re-examined the label allocation from a global perspective, creatively transformed it into an optimal transportation (OT) problem in a linear programming problem, calculated the transportation cost between ground truth and all prediction boxes, and minimized the transportation cost by finding an appropriate mapping relationship. In OTA, authors defined the unit transportation cost between each demander (anchor box) and supplier (ground truth) pair as the weighted sum of their classification and regression losses, and converted the search for the optimal allocation solution to the solution of the optimal transportation plan with the minimum transportation cost.

General linear programming problems can be solved in polynomial time. However, in the problem of detection, the generated linear programming is very large, involving the characteristic size square of anchor points in all scales, and the cost of training is high. So we adopt a simplified version of OTA, named SimOTA (Simplify OTA). SimOTA omitted the iterative solution process and instead simplified it to a dynamic top-k strategy to obtain an approximate solution. SimOTA algorithm can dynamically match the optimal label through the prediction information output from the current network, which can not only reduce the training time, but also avoid additional hyper-parametric problems, and has the similar accuracy as OTA. Please refer to table 1 for SimOTA algorithm details.

D. HIERARCHICAL-PATH FEATURE FUSION NETWORK (H-PFANet)

How to alleviate the degradation of detection effect caused by the change of target scale has always been an important research direction in the field of target detection. Early detectors usually perform prediction directly based on the pyramid feature hierarchy extracted from the backbone network. Feature pyramid network (FPN) [21] proposed a classic top-down fusion strategy to combine multi-scale features, i.e. the VGG and other linear networks are further expanded from the deep layer through convolution, up-sampling and other operations, and then the features of the two paths are fused

TABLE 1. SimOTA algorithm.

Algorithm1: Simplify Optimal Transport Assignment (SimOTA)

Inputs :
I is an input image
A is a set of anchors
G is the ground truth annotations for objects in image I
m is the number of iterations
λ is the weighting coefficient

Outputs :
get candidate anchors as positive samples of $gt_i(i = 1, 2, ...,)$
1: identify candidate regions by center prior
2: screen out n prediction boxes as candidate boxes $a_j(j = 1, 2, ...,)$
3: P^{cls}; $G^{cls} \leftarrow$ Forward(I, A)
4: calculate class loss : $c_{ij}^{cls} = \text{BCEWithLogitsLoss}\ (P_j^{cls}, G_i^{cls})$
5: calculate regression loss : $c_{ij}^{reg} = \text{SIoULoss}(P_j^{box}, G_i^{box})$
6: calculate cost: $c_{ij} = c_{ij}^{cls} + \lambda c_{ij}^{reg}$
7: N=min{10, n}
8: select the top N candidate boxes with the highest IoU for each gt_i
9: sum these N IoU and round down to get the top-k for each gt_i
10: for i=1 to m do:
11: select the top-k candidate boxes with the least cost within a fixed center region for gt_i
12: if a candidate box a_j matches multiple ground truths then select the least cost ground truth matching a_j
13: else it is selected as a positive sample of gt_i

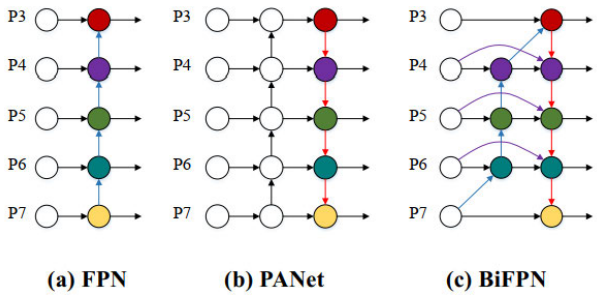


FIGURE 5. Feature fusion network structure.

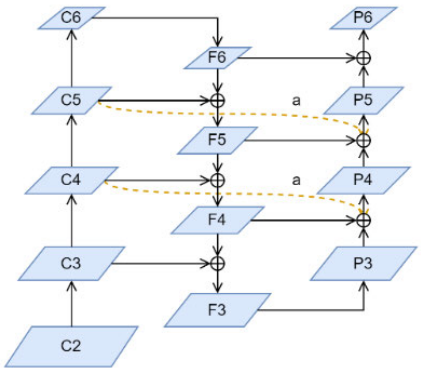


FIGURE 6. The structure of improved PANet.

together through pairwise addition, please refer to figure 5(a) for details. PANet [23] has implemented path enhancement and aggregation to improve performance. On the basis of FPN, PANet added a bottom-up path to make it easier to spread low-level information, please refer to figure 5(b). The feature fusion network in Yolov5 is the PAN structure based on FPN.

Recently, on the basis of PANet, Google team proposed a weighted bidirectional feature pyramid network (BiFPN) [24], please refer to figure 5(c) for details. A major contribution of BiFPN is that it added a shortcut between the same levels except the top and bottom levels, realized the fusion of the original feature information of the backbone feature extraction network and the deep information, and reduced the deviation that may occur in deep learning of the network.

Inspired by the shortcut on BiFPN, Qiu Tianheng, Wang Pengfei et al. [25], [26] improved the PANet in YOLOv5. However, restricted by the YOLOv5 structure, this connection can only be added between C4 and P4 levels, and one connection often cannot improve the model performance. So the two references both added a detection head to the original YOLOv5 structure, which deepens the original feature extraction network, so that the model can add two shortcuts between C4 to P4 and C5 to P5 levels, please refer to figure 6.

Experiments show that the added detector head and C6, F6 and P6 layers in the above improvements will greatly increase the complexity and calculation of the model. From the point of view of keeping the model lightweight, aiming at the existing YOLOv5 structure, we propose a new feature fusion structure, named Hierarchical-Path Feature Aggregation Network (H-PFANet). H-PFANet contains two connection strategies. While retaining a C4 to P4 BiFPN normal form connection (curve *a* in figure 7), it designs a special connection between C3 and P3 (curve *b* in figure 7). A feature fusion structure is added between C3_2 and CBS layer. C3 layer is the middle layer of the backbone network, which contains important target location information and semantic information. The new connection *b* in H-PFANet enables the model to properly integrate the location signal and semantic

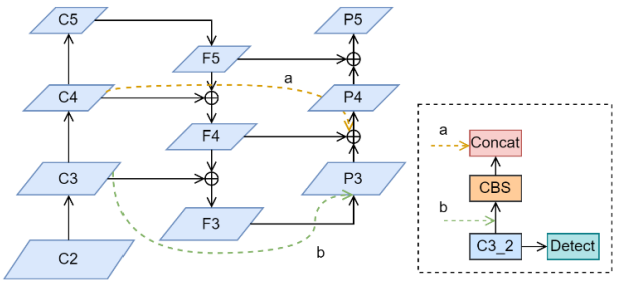


FIGURE 7. The structure of improved PANet.

TABLE 2. Experimental environment.

Parameter	Configuration
CPU	Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50GHz 45GB
GPU	RTX 2080 Ti 11GB
operating system	Ubuntu
language	Python 3.8.10
accelerated environment	CUDA 11.2, cuDNN 8.2
frame	PyTorch 1.10.0

information of the middle layer of the backbone network in the learning process, which can greatly improve the model detection effect. We conduct experiments to show the performance of H-PFANet in subsection IV-E.

To sum up, we mainly improved four parts of the YOLOv5s, that is, the condition mechanism, the loss function, the label assignment strategy and at last we proposed a hierarchical-path feature fusion network. The structure of our proposed method is shown in figure 8, and the red dash line marks the main improved parts.

IV. EXPERIMENTS

A. EXPERIMENTAL ENVIRONMENT AND PARAMETER SETTING

1. Experimental environment is shown in table 2. A regular computer is used for experiment comparison.

2. Parameter setting.

In order to ensure the validity of the control experiment results, all models adopt the same hyper-parameter. The input image size is 640 × 640; the initial learning rate is 0.01; the momentum parameter is 0.937; the final learning rate is 0.1, and the batchSize is 32. At the beginning, three rounds of warm-up were carried out, and then the cosine annealing strategy was used to update the learning rate. A total of 300 epochs were trained.

For SimOTA, in order to improve the matching efficiency in the experiment, before using the SimOTA method to match the positive and negative samples, we first found the sample selection interval through the central prior method, screened out the anchor points and used the boxes generated by these anchor points as candidate boxes, calculated the IoU of this candidate box and the real box and the classification and regression losses respectively, got the IoU matrix and the cost matrix between the anchor points and the ground truth, and then took the smaller value between n and 10, set $N = min\{10, n\}$. The strategy calculated the sum of the first N largest IoU values in the sample selection range and round down, and recorded as k. Finally, for each ground truth, select the first k smallest cost candidate boxes as the positive sample, others are determined as negative samples.

B. DATA SETS AND PREPARATION

We conduct experiments mainly on two data sets, i.e. CCTSDB-2021 [29] and TT100K [30].

First, we conducted experiments with the 4000 difficult samples in the China Traffic Scene Data Set (CCTSDB), which is collected by professor Zhang Jianming's team from Changsha University of Technology [29]. In order to face more realistic and comprehensive traffic scene images, the latest open source CCTSDB-2021 data set added 4000 new difficult samples to replace the simple samples in the previous version. It not only includes a variety of road conditions such as highways, cities and towns, but also has a variety of complex weather and remote shooting pictures such as rain, snow, fog, weak light at night, strong light at night and day, which greatly increases the detection difficulty. The data set has three categories: instruction, prohibition and warning. During the experiment, 4000 pictures were split into training set and test set in a 3:1 ratio, and the data set's format is TXT. The data enhancement used includes translation, left-right rotation, hue, saturation, exposure and Mosaic4. The ratio used of the first five enhancements are 0.5, 0.1, 0.015, 0.7, 0.4 respectively. Mosaic4 refers to randomly selecting four images for splicing during training to enhance the detection effect of small targets.

Second, we conducted experiments on the whole CCTSDB-2021 data set with 17856 images. During the experiment, 17856 pictures were split into training set and testing set in a 8:2 ratio. We compared our model with some popular traffic sign detection algorithms on this data set.

Third, we did experiments on TT100K data set. TT100K is a Chinese traffic sign detection data set jointly collected and organized by Tsinghua University and Tencent. It includes more than 120 subcategories of Chinese traffic signs, including warnings, prohibitions, and instructions. The image scene is rich, covering changes in lighting and weather conditions. In the experiments, the total 9738 images were used. They were split into training set and testing set randomly with the ratio 8:2, that is, it was about 7790 images for training and 1948 images were used for testing.

C. EVALUATION INDEX

In order to comprehensively evaluate the effect of the model from multiple perspectives, this paper selected the number of model parameters (M), the magnitude of calculation (GFLOPs), the model size, the average precision at the threshold 0.5 and 0.5:0.95, i.e. mAP@0.5 and mAP@0.5:0.95, detection speed (frame per second, FPS) as the measurement standard of detection algorithms. mAP (mean Average

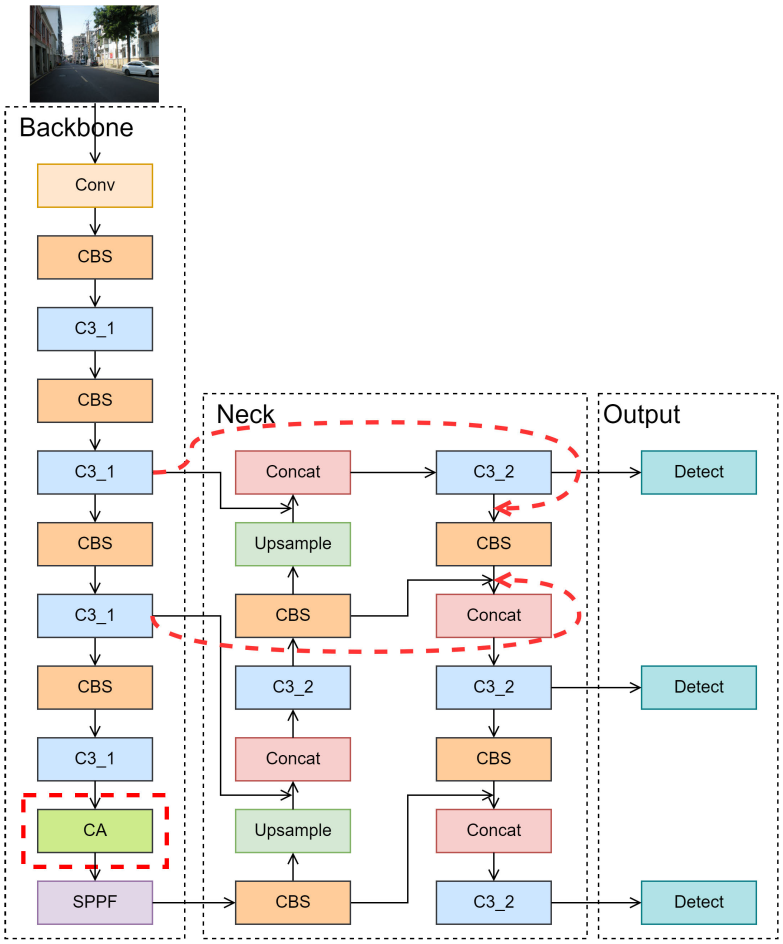


FIGURE 8. The construction of our proposed model, and the red dash line marks the main improved parts.

Precision) is the average precision of all the classes, which can be calculated as follows:

$$mAP = \frac{\sum_{i=1}^{k} AP_i}{k}$$

where k is the number of classes, and AP is the area under the PR (Precision-Recall) curve.

D. THE EFFECT OF CA

In order to show the effect of CA model, we randomly selected two traffic sign images from the CCTSDB-2021 test set, took the Grad-CMA method to generate Class activation thermodynamic diagram for the original YOLOv5 and the YOLOv5 with the CA model. The results are shown in figure 9.

In figure 9, the first row is the original images, the second row is the thermodynamic diagram of YOLOv5 detection process without CA model, and the last row is the result of YOLOv5 with CA model. From the thermodynamic diagram, we can find that after introducing the CA model, model's ability to accurately locate targets is enhanced, the model's attention to important features is increased, and then the confidence of the prediction box is higher than the model without CA. It shows that the CA model can capture more accurate position signals and semantic information, and has stronger detection ability.

TABLE 3. Feature fusion structure ablation experiment.

used module	mAP@0.5(%)	mAP@0.5:0.95(%)
PAN	81.1	50.3
H-PFANet_a	80.5 (−0.6)	50.3
H-PFANet_b	81.8 (+**0.7**)	50.6 (+**0.3**)
H-PFANet_ab	82 (+**0.9**)	51.6 (+**1.3**)

E. THE EFFECT OF H-PFANet

We conducted experiment to show the strategy *ab*'s efficiency on the difficult dataset of CCTSDB-2021. We used the PAN, only strategy *a*, only strategy *b* and strategy *ab* separately in the module, and the results are shown in table 3.

Compared to PAN module, when only using the BiFPN connection, i.e. strategy *a*, the model's mAP@0.5 decreases 0.6%, and mAP@0.5:0.95 is the same as PAN; when only use the strategy *b*, and the model's mAP@0.5 increases

FIGURE 9. The first row are the original images, the second row are the thermodynamic diagram without CA model, and the last row are the results with CA model.

0.7%, mAP@0.5:0.95 increases 0.3%; when using the strategy *ab*, the model's mAP@0.5 increases 0.9%, and mAP@0.5:0.95 increases 1.3%. This experiment shows that our proposed strategy performed well.

F. ABLATION EXPERIMENT

As table 4 shows, in order to verify the detection effect of the proposed algorithm on traffic signs under complex road conditions and the effectiveness of various improvements, four sets of ablation experiments are designed. Method B introduces the CA module into the YOLOv5, and the mAP@0.5 increases 1%, the mAP@0.5:0.95 increases 0.5% with only a small part of the parameter quantity added, and there is almost no additional calculation overhead. Method C continues to improve the loss function on the basis of method B, after introducing SIoU, compared to method B, the mAP@0.5 is improved by 1.7%, and the mAP@0.5:0.9 is improved by 1.3%. While compared to YOLOv5, the mAP@0.5 is improved by 2.7%, and the mAP@0.5:0.9 is improved by 1.8%. Continue to improve the label allocation strategy based on method C, and introduce dynamic label allocation strategy SimOTA, compared to method C, the mAP@0.5 is improved by 2.7%, and the mAP@0.5:0.9 is improved by 2.2%. While compared to YOLOv5, the mAP@0.5 is improved by 5.4%, and the mAP@0.5:0.9 is improved by 4.0%. Method E continues to improve the feature fusion network on the basis of method D. After replacing it with the new feature fusion network (H-PFANet) proposed in this paper, the mAP@0.5 increases 0.9%, and the mAP@0.5:0.9 increases 1.3%. Our final proposed algorithm E gets a significant improvement compared to YOLOv5 with the mAP@0.5 increasing 6.3%, and the mAP@0.5:0.9 increasing 5.3%. While only a small amount of parameters and calculation are added, our proposed algorithm E maintains the lightweight of the model.

Due to the improvement of label allocation strategy, the loss calculation before and after the model are different. The precision has been significantly improved, and the model convergence speed has been accelerated. After using the SimOTA dynamic label allocation strategy, the model detection effect has been improved significantly.

G. COMPARATIVE EXPERIMENT WITH THE DIFFICULT SAMPLES IN CCTSDB-2021 DATA SET

In order to further verify the effectiveness and progressiveness of the proposed algorithm in this manuscript, we conducted experiments compared with other popular detection algorithms: YOLOv5s, YOLOv3, YOLOv4, YOLOv5m, YOLOX-s, YOLOv6s, YOLOv7 [27], YOLOv7-tiny, YOLOv8 [28], where YOLOX [31] won the first place in the 2021 CVPR Streaming Video Challenge, and it is currently recognized as one of the best detectors. YOLOv7 is a new detector designed by the original YOLOv4 team. YOLOv7 has integrated the best research results at present,

TABLE 4. Ablation experiment.

No.	Method	Parems (M)	GFLOPs	mAP@0.5 (%)	mAP@0.5:0.95 (%)
A	Yolov5s	7.02	15.8	75.7	46.3
B	A+CA	7.04	15.8	76.7 (+**1.0**)	46.8 (+**0.5**)
C	B+SIoU	7.04	15.8	78.4 (+**2.7**)	48.1 (+**1.8**)
D	C+SimOTA	7.04	15.8	81.1 (+**5.4**)	50.3 (+**4.0**)
E	**D + H − PFANet**	7.26	16.5	**82** (+**6.3**)	**51.6** (+**5.3**)

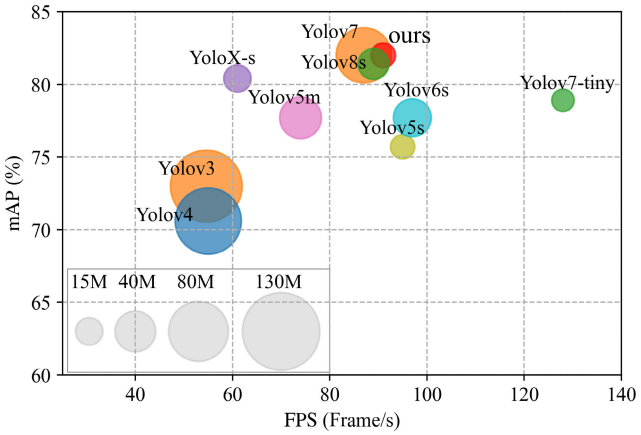


FIGURE 10. Contrast Experiment: the horizontal axis is the detection speed, the vertical axis is the detection accuracy, and the size of the circle represents the size of the model.

and the model structure is redesigned. In its open-source paper in July, the author has demonstrated through a large number of experiments that in the range of 5FPS to 160FPS, the YOLOv7 series are better than the currently known detectors in terms of both speed and accuracy. YOLOv8 was released in January 2023 by Ultralytics, the company that developed YOLOv5.

As shown in Figure 10, we compared the detection effects of ten models from three dimensions: detection speed, detection accuracy and model size. Vertical axis represents the average precision when the IoU threshold is 0.5. The horizontal axis is the detection speed and the size of the circle represents the size of the model. It is obvious that, for the similar model size, our method has a higher precision; for similar precision, our model's size is smaller; for similar speed, our precision is higher.

Table 5 offers the details of the experiments. Compared to YOLOv5m, YOLOX-s, YOLOv6s, our proposed algorithm achieves higher detection accuracy with less parameters and computation, and the detection speed is significantly better than the former two methods, a little slower than YOLOv6s. With its new designed architecture, YOLOv7-tiny has advantages in model size and detection speed, but our precision is higher with mAP@0.5 increasing 3.1% and mAP@0.5:0.95 increasing 6.4%. Compared to YOLOv7, our parameter quantity and calculation quantity are less than one fifth of YOLOv7, but we got the same mAP@0.5 as YOLOv7, a better mAP@0.5:0.95 result, and a higher speed. Compared with YOLOv8s, our model has advantages in the number of parameters, detection accuracy and detection speed.

H. COMPARATIVE EXPERIMENT ON THE WHOLE CCTSDB-2021 DATA SET

In this subsection, we conducted experiments on the whole CCTSDB-2021 data set to compare our model with the current traffic sign detection algorithms. The algorithms include: YOLOv5s, Improved MobileNetv2-SSD [32], Faster R-CNN [5], TSR-YOLO [34], M3E-YOLO [35], T-YOLO [36], M-YOLO [33], the method in [37], YOLOV6s, YOLOv7-tiny and YOLOv8s. A short description of some of those algorithms is introduced as follows:

- Improved MobileNetv2-SSD [32] is a method which combined MobileNetv2 and SSD to improve the detection speed and precision.
- Liu et al. proposed M-YOLO [33] based on YOLOv3. M-YOLO's detection speed reached 84 FPS, and its mAP@0.5 reached 97.8% on CCTSDB-2021 data set.
- Song et al. proposed TSR-YOLO [34] by improving YOLOv4-tiny.
- M3E-YOLO [35] is proposed by Guo et al. By introducing a light backbone-MobileNetv3 into YOLOv5, M3E-YOLO maintained the high detection accuracy with the number of the model parameters reduced significantly.
- T-YOLO [36] is proposed by Chen et al. Even though it improved the mAP, its speed is low.
- Yang et al. [37] improved YOLOv5 to proposed an detection algorithm for traffic sign detection.

Table 6 shows the results of the relevant traffic sign detection algorithms. From table 6, it is obvious that our model obtained the best performance with a speed of 91 FPS. YOLOv7-tiny is with the highest speed, but its precision, recall, and mAP@0.5 are not very good. Also M3E-YOLO has the least parameters of these methods, but it is not good on other aspects.

I. COMPARATIVE EXPERIMENT ON TT100K DATA SET

In this subsection, we show experiments on TT100K data set to compare our model with the relevant traffic sign detection algorithms. The algorithms include: YOLOv5s, the method in [38], Faster-RCNN [5], the algorithm in [39], YOLOv5-DH [40], YOLOv5-TDHSA [40], YOLOV6s, YOLOv7-tiny

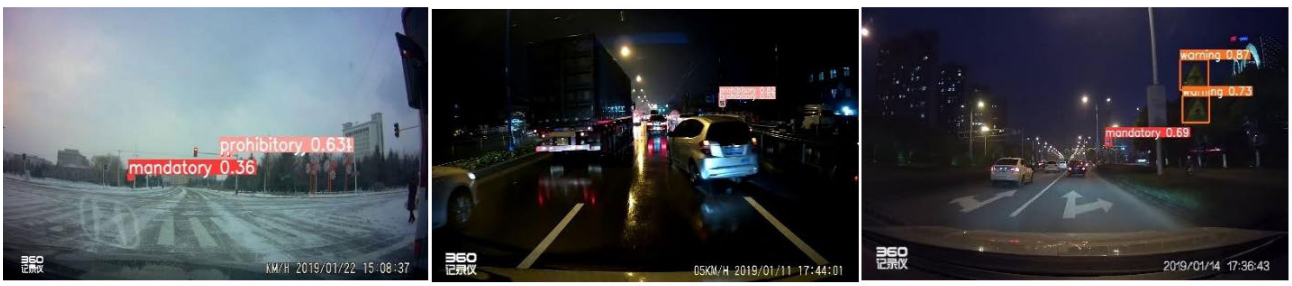
TABLE 5. Contrast experiment with the 4000 difficult samples in CCTSDB-2021 data set.

Method	size	Parems (M)	GFLOPs	model size (M)	mAP@0.5 (%)	mAP@0.5:0.95 (%)	FPS_{b1}
YOLOv5s	640	7.02	15.8	14.4	75.7	46.3	95
YOLOv3	640	61.53	193.89	120.5	73.0	39.9	54
YOLOv4	640	52.5	119.83	100.6	70.6	41.6	55
YOLOv5m	640	20.9	48.0	42.2	77.7	48.3	74
YOLOX-s	640	8.94	26.64	18.5	80.4	48.5	61
YOLOv6s	640	17.19	44.12	36.3	77.7	44.3	97
YOLOv7	640	36.49	103.5	74.8	**82.0**	50.8	87
YOLOv7-tiny	640	6.01	13.1	12.3	78.9	45.2	128
YOLOv8s	640	11.14	28.4	22.5	81.4	51.1	89
Ours	640	7.26	16.5	14.8	**82.0**	**51.6**	91

TABLE 6. Experiment on the whole CCTSDB-2021, where P is the precision, R represents recall, and FPS is the detection speed.

Method	P%	R%	Parems(M)	mAP@0.5%	FPS_{b1}
YOLOv5s	97.4	93.8	7.02	96.5	95
Improved MobileNetv2-SSD [32]	-	-	-	93.2	45
Faster R-CNN [5] [33]	91.6	90.7	153.56	93.5	22
TSR-YOLO [34]	96.6	79.7	20.49	92.7	80
M3E-YOLO [35]	94.5	87.3	3.1	93.6	-
T-YOLO [36]	91.3	-	-	97.3	19
M-YOLO [33]	93.5	96.3	-	97.8	84
Yang et al. [37]	97.6	94.7	-	98.1	88
YOLOv6s	-	-	17.19	97.9	97
YOLOv7-tiny	96.5	93.9	6.01	97.5	112
YOLOv8s	97.1	96.4	11.14	98.3	89
Ours	**98.1**	**97.6**	7.26	**98.8**	91

(a) The detect result of YOLOv5.



(b) The detect result of our method.

FIGURE 11. The detection results of our method and YOLOv5. For the left and middle images, YOLOv5 didn't find the traffic signs, but our method discovered and recognized the signs. For the third image, our method discovered and recognized more signs with a higher confidence.

and YOLOv8s. We compared the algorithms from precision, recall, model parameter, mAP@0.5 and detection speed aspects. The results are shown in table 7.

From table 7, we can find that our model got the best performance on recall with a reasonable speed of 81 FPS. YOLOv8s obtained the best precision and mAP@0.5, but our method got a similar mAP@0.5 result with YOLOv8s and the speed is higher than YOLOv8s. YOLOv5-TDHSA also obtained a similar mAP@0.5 result, but it has a larger parameter than our model. Our based model YOLOv5s has

TABLE 7. Experiment on TT100K data set, where P is the precision, R represents recall, and FPS is the detection speed.

Method	P%	R%	Parems(M)	mAP@0.5%	FPS_{b1}
YOLOv5s	79.6	72.9	7.13	78.5	88
Wang et al. [38]	70.16	64.52	12.04	50.1	87
Faster R-CNN [5] [39]	-	-	-	82.59	24
Wang et al. [39]	-	-	-	81.78	74
YOLOv5-DH [40]	-	-	11.07	79.9	84
YOLOv5-TDHSA [40]	-	-	12.25	83.4	77
YOLOv6s	-	-	17.21	81.5	89
YOLOv7-tiny	58.6	49.1	6.13	45.7	95
YOLOv8s	**84.5**	74.5	11.21	**83.9**	79
Ours	83.3	**77.7**	7.37	83.7	81

the highest speed, but its precision, recall and mAP@0.5 are not good.

J. QUALITATIVE EVALUATION

Figure 11 shows some detection results of YOLOv5 and our method. The image on the left is the small target recognition task of the road beyond the distance after snow. The road in middle column image is with local strong light and reflection on rainy night, it is also with complex road conditions and serious interference. The image on the right is the target recognition of the road with weak light at night, and the visibility is low. In the three groups of experiments, the original YOLOv5 missed all or some traffic signs. Our proposed method not only detected all the targets correctly, but also with a high confidence. It shows that the proposed algorithm captures more accurate position signals and semantic information, and has a better detection performance.

V. CONCLUSION

Aiming at the problems of low recognition of traffic signs and serious missing detection under complex road conditions, this paper proposed an improved algorithm based on YOLOv5s. The experiments show that, by introducing coordinate attention, the model can deal with other interference in complex background, and improve feature attention; angle loss components can reduce the degree of freedom of the prediction frame in the process of convergence, the proposed method can fit the real target faster; dynamic label allocation strategy could generate more high-quality positive samples and promote the positive optimization of the network; the proposed H-PFANet could maximize the fusion of backbone network information and improve the detection effect of small targets. Compared with the most advanced model at present to our best knowledge, the model proposed in this paper has higher detection precision under the same volume; smaller volume and faster detection speed under the same precision, and has better robustness to various scene changes. The experiments showed our proposed algorithm's effectiveness and good performance.

ACKNOWLEDGMENT

(Qianying Wang, Xiangyu Li, and Ming Lu are co-first authors.)

REFERENCES

[1] G. Loy and N. Barnes, "Fast shape-based road sign detection for a driver assistance system," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2004, pp. 70–75.
[2] S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno, and F. Lopez-Ferreras, "Road-sign detection and recognition based on support vector machines," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 2, pp. 264–278, Jun. 2007.
[3] 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.
[4] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, p. 28.
[6] W. Liu, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 21–37.
[7] 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)*, Jun. 2016, pp. 779–788.
[8] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.
[9] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[10] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[11] G. Wang, Z. Xiong, D. Liu, and C. Luo, "Cascade mask generation framework for fast small object detection," in *Proc. IEEE Int. Conf. Multimedia Expo. (ICME)*, Jul. 2018, pp. 1–6.
[12] T. Dong and H. Cao, "Research on multi-target recognition method of traffic scene in complex weather," *Inf. Commun.*, vol. 11, pp. 72–74, Jan. 2020.
[13] K. Zhao, "Traffic signs detection and recognition under low-illumination conditions," *Chin. J. Eng.*, vol. 42, no. 8, pp. 1074–1084, Jan. 2020.
[14] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[15] S. Woo, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 3–19.
[16] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13708–13717.
[17] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proc. AAAI Conf. Artif. Intell.*, New York, NY, USA, vol. 34, Feb. 2020, pp. 12993–13000.
[18] 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)*, Jun. 2019, pp. 658–666.
[19] Z. Gevorgyan, "SIoU loss: More powerful learning for bounding box regression," 2022, *arXiv:2205.12740*.
[20] Y. F. Zhang et al., "Focal and efficient IOU loss for accurate bounding box regression," 2021, *arXiv:2101.08158*.

[21] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.
[22] Z. Ge, S. Liu, Z. Li, O. Yoshie, and J. Sun, "OTA: Optimal transport assignment for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 303–312.
[23] 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.
[24] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10778–10787.
[25] Q. Tianheng, W. Ling, W. Peng, and B. Yan'e, "Research on target detection algorithm based on improved YOLOv5," *Comput. Eng. Appl.*, vol. 58, no. 13, pp. 63–73, 2022.
[26] W. Pengfei, H. Hanming, and W. Mengqi, "Improve the complex road target detection algorithm of YOLOv5," *Comput. Eng. Appl.*, vol. 58, no. 17, pp. 81–92, 2022.
[27] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, *arXiv:2207.02696*.
[28] C. Y. Wang, A. Bochkovskiy, and H. Liao. (2023). *YOLO by Ultralytics*. [Online]. Available: https://github.com/ultralytics/ultralytics
[29] J. Zhang, X. Zou, L.-D. Kuang, J. Wang, R. Simon Sherratt, and X. Yu, "CCTSDB 2021: A more comprehensive traffic sign detection benchmark," *Hum.-Centric Comput. Inf. Sci.*, vol. 12, no. 23, pp. 1–19, 2022.
[30] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu, "Traffic-sign detection and classification in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2110–2118.
[31] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.
[32] K. Ren, L. Huang, and C. Fan, "Real-time small traffic sign detection algorithm based on multi-scale pixel feature fusion," *Signal Process*, vol. 36, pp. 1457–1463, Jan. 2020.
[33] Y. Liu, G. Shi, Y. Li, and Z. Zhao, "M-YOLO: Traffic sign detection algorithm applicable to complex scenarios," *Symmetry*, vol. 14, no. 5, p. 952, May 2022.
[34] W. Song and S. A. Suandi, "TSR-YOLO: A Chinese traffic sign recognition algorithm for intelligent vehicles in complex scenes," *Sensors*, vol. 23, no. 2, p. 749, Jan. 2023.
[35] G. Haoran, L. Fan, K. Ping, and X. Gang, "M3E-YOLO: A new lightweight network for traffic sign recognition," in *Proc. 19th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process. (ICCWAMTIP)*, Dec. 2022, pp. 1–6.
[36] Z. Zhe, D. Liang, S. Zhang, X. Huang, and S. Hu, "A depth based traffic sign recognition algorithm," *Telecommun. Technol.*, vol. 61, pp. 76–82, Jan. 2021.
[37] G. Yang et al., "An improved YOLOv5 traffic sign detection algorithm," *Comput. Eng. Appl.*, vol. 59, no. 10, pp. 262–269, 2023.
[38] L. Wang, K. Zhou, A. Chu, G. Wang, and L. Wang, "An improved lightweight traffic sign recognition algorithm based on YOLOv4-tiny," *IEEE Access*, vol. 9, pp. 124963–124971, 2021.
[39] Y. Wang, M. Bai, M. Wang, F. Zhao, and J. Guo, "Multiscale traffic sign detection method in complex environment based on YOLOv4," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–15, Oct. 2022.
[40] W. Bai, J. Zhao, C. Dai, H. Zhang, L. Zhao, Z. Ji, and I. Ganchev, "Two novel models for traffic sign detection based on YOLOv5s," *Axioms*, vol. 12, no. 2, p. 160, Feb. 2023.

QIANYING WANG received the Ph.D. degree in computational science from Sun Yat-sen University, China, in 2014. From October 2010 to July 2011, she was an exchange student in computer science with the Hong Kong Baptist University, Hong Kong. From December 2012 to December 2013, she was a Research Assistant in computer science with the University of Nottingham, U.K. Currently, she is an Associate Professor with the Hebei University of Economics and Business. Her research interests include semi-supervised metric learning, pattern recognition, and deep learning.

XIANGYU LI received the B.S. degree in mathematics and applied mathematics from Hebei Agricultural University, Hebei, China, in 2020. He is currently pursuing the master's degree in applied statistics with the Hebei University of Economics and Business, Hebei. His research interests include machine learning, computer vision, and object detection.

MING LU received the Ph.D. degree in computational science from Sun Yat-sen University, China, in 2012. Currently, he is an Associate Professor with Hebei Normal University. At present, he is going to study the stability of solutions of partial differential equation, including Navier-Stokes equations and Boussinesq equations. His research interests include partial differential equation, fluid mechanics, the well-posedness of fluid solutions, including regularity criteria and the global existence of large initial value solutions, and big data learning.