IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 30 July 2023, accepted 28 September 2023, date of publication 5 October 2023, date of current version 11 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3322371

RESEARCH ARTICLE

UCN-YOLOv5: Traffic Sign Object Detection Algorithm Based on Deep Learning

PEILIN LIU, ZHAOYANG XIE, AND TAIJUN LI
Internet Information Retrieval Major Laboratory, Hainan Province School of Information and Communication Engineering, Hainan University, Haikou 570288, China

Corresponding author: Taijun Li (HDLTJ08@126.com)

ABSTRACT Traffic sign detection plays an important role in traffic safety and traffic management. In view of the complex and changeable environment and detection accuracy of traffic sign detection, this paper proposes UCN-YOLOv5 model based on the framework of YOLOv5.This model first replaces a new backbone network, which uses the core module RSU of U2Net to enhance the feature extraction of the network. Then, ConvNeXt-V2 is integrated, and the C3 module of its Block and YOLOv5 network is used to construct the C3_CN2 structure. The utilization of the proposed lightweight receptive field attention module LPFAConv in the Head section represents a potential enhancement for the extraction of receptive field features. Finally, for small targets in traffic signs, Normalized Wasserstein Distance (NWD), which is insensitive to targets of different scales, is added to calculate the position loss function to replace the IoU metric to a certain extent, which further improves the detection ability of our model for traffic signs. Experiments on the TT100K dataset show that UCNYOLOv5 has excellent detection performance. Compared with the baseline model (Y0Lov5s, YOLOV5m, YOLOV5l), it improves the Map.5 index by 5.9 %, 4.9 % and 4.6 %; in the Map.5: .95 index, it is 4.4 %, 3.5 % and 2.8 % better. Moreover, the enhanced algorithm demonstrated favorable performance on the LISA and CCTSDB2021 traffic sign datasets. This research has important value for the accurate detection of traffic sign detection, and has guiding significance for in-depth research in related fields.

INDEX TERMS Object detection, traffic sign detection, YOLO, YOLOv5, U2Net, Convnext, RFAConv.

I. INTRODUCTION

Traffic signs are important signs set up to ensure traffic order and traffic safety. Their main purpose is to provide information on road conditions and auxiliary guidance for pedestrians and vehicles on the road [1], [2]. Traffic sign detection is a very important task when it comes to traffic safety. It is one of the important elements that need to be perceived in the driving environment and is of great significance to ensure the safety of drivers and pedestrians. By identifying traffic signs on the road in real time, traffic sign detection can provide some traffic information required by the driver, such as speed limit, prohibition, warning, etc., to remind the driver to follow the rules, which is conducive to reducing the incidence of traffic violations and accidents. Traffic sign detection can also improve road traffic efficiency and keep intersections open [3], [4]. In addition, traffic sign detection is also an important part of autonomous driving and intelligent transportation systems. By using advanced computer vision technology and deep learning algorithms, traffic signs on the road can be accurately identified and understood, so as to be able to independently decide and control vehicle driving and achieve safer and intelligent autonomous driving. Therefore, traffic sign detection is of great importance in modern transportation systems [5], [6]. The data sets commonly used in current research on traffic sign detection are GTSRB [7] and GTSDB [8]. Among them, the data provided by GTSDB can be used to study the location and classification of traffic signs, but the location and classification information of traffic signs provided by GTSDB is far less than the types of traffic signs in real life. Although GTSRB provides 43 types of traffic signs, these traffic signs can only be classified and cannot be used for positioning and classification at the same time. With the

The associate editor coordinating the review of this manuscript and approving it for publication was Taous Meriem Laleg-Kirati.

development of deep learning, the use of artificial neural networks has made significant progress in the field of target detection. In the real traffic environment, traffic signs often appear on roads, intersections, schools, tunnels, bridges, urban areas and highways. In various places, it is easily affected by weather, light intensity, reflection and background occlusion by trees, buildings and other vehicles. At the same time, there are many types of traffic signs and small sizes. These factors will affect the detection effect of traffic signs. Since the TT100K [9]dataset provides more complete traffic signs and more complete application of the scene, this paper will use this dataset to test and verify the performance of the algorithm.

II. RELATED RESEARCH

At present, algorithms based on deep learning provide a broad research direction for the detection of traffic signs. The algorithms are mainly divided into two categories: algorithms based on candidate region networks and algorithms based on regression. The former first determines the region of interest by retrieving the approximate position of the object, and uses the feature extraction network to judge the coordinates and specific categories of the target. Typical examples of such algorithms include R-CNN [10], Fast R-CNN [11], and Faster R-CNN [12]. Liang et al. [13] used a deep feature pyramid network with horizontal connections to obtain the semantic features of traffic signs, but the feature information extraction is not perfect and the detection accuracy is low. Zhou Su [14] improved the shallow and deep feature extraction layer and HypeNet layer left(multi-layer feature information fusion layer right) in PVANet network to obtain higher detection accuracy, but such algorithms have poor real-time performance and complex network structure. Zhu et al. [9] used the candidate region network to detect and classify traffic signs, but often could not extract deep feature information. The traffic sign detection algorithm based on regression can realize the whole process of traffic sign image input to classification result output in a network, and the detection speed is more advantageous.Typical representatives of this type of algorithm are YOLO series [15], [16] and SSD [17].

Typical researchers are: Wu et al. [18] introduced Dark-Net19 as a classification network in the YOLOv3 algorithm, and enhanced the GTSDB dataset based on traffic sign features, but this algorithm has poor detection effect on smaller traffic signs; rajendran et al. [19] used YOLOv3 as the detection network and equipped with CNN classifier to construct a traffic sign detection algorithm. This algorithm has a significant improvement in detection accuracy, but it has high requirements for the environment and is easily affected by the environment. Zhang et al. [20] proposed MSA_YOLOv3 algorithm, which uses Mixup [21] technology to enhance the dataset image, and introduces pyramid multi-scale pooling layer to make the model learn the deep information of traffic signs more effectively, but the network of the algorithm is more complex and the generalization ability is weak. Liang et al. [22] propose a lightweight traffic sign detection algorithm, which is developed on the YOLOv4 framework. The algorithm incorporates conventional improvement measures by utilizing the lightweight Mobilenetv3 [23] network and the SE attention module. Additionally, a Feature Fusion and Redistribution Module, similar to the weight allocation concept in BiFPN, is applied. The algorithm demonstrates a certain level of improvement in detection speed. However, the efficacy of the lightweight network is not guaranteed for complex scenarios, and the algorithm's generalization capability is not adequately substantiated as it is only validated on the limited GTSDB dataset. Wang J et al. [24] propose an improved feature pyramid model called AF-FPN, which utilizes the Adaptive Attention Module (AAM) and Feature Enhancement Module (FEM) to mitigate information loss during feature map generation and enhance the representation capacity of the feature pyramid. The improvement is to replace the PANet in the original YOLOv5 model with AF-FPN for traffic sign detection and validate it on the TT100K dataset, Although this paper shows promise, it has minor improvements and insufficient experimental validation. The lack of broader experiments, limited to the YOLOv5s framework, casts doubt on its generalizability.

Therefore, aiming at the problems of detection accuracy and generalization ability of computer vision target detection tasks due to environmental complexity, this paper proposes an improved model UCN-YOLOv5 to strengthen the detection of traffic signs in various traffic scenarios. In the backbone network, we use the ' U ' -shaped structure ReSidual-Ublock left(RSU right) in U2Net [25], [26] to achieve the fusion of multi-scale features and the retention of detailed information, which improves the accuracy and perception of image detection. Through downsampling and upsampling operations, the effective receptive field of the network can be improved while maintaining the resolution. Then, in the Neck part of the network, while retaining PANet, the Block module of ConvNeXt-V2 [27] high-performance CNN network is used to form the C3_CN2 module to replace the C3 structure of the original network to explore better prediction potential. In order to strengthen the ability of the model to detect small targets and make up for the defect that the commonly used IOU is too sensitive to small targets, NWDLoss is used as part of the calculation of position loss.Finally,The lightweight receptive field attention module LRFAConv is used in Network Head to better extract spatial receptive field features. Compared with the basic network YOLOv5, the improved UCN-YOLOv5 can better detect images of traffic signs in various scenes.

III. TRAFFIC SIGN DETECTION ALGORITHM BASED ON DEEP LEARNING

A. THE OVERALL FRAMEWORK OF UCN-YOLOv5 ALGORITHM

Initially, we introduced the RSU (Recursive Scale-Up) structures from U2Net to enhance the backbone network of

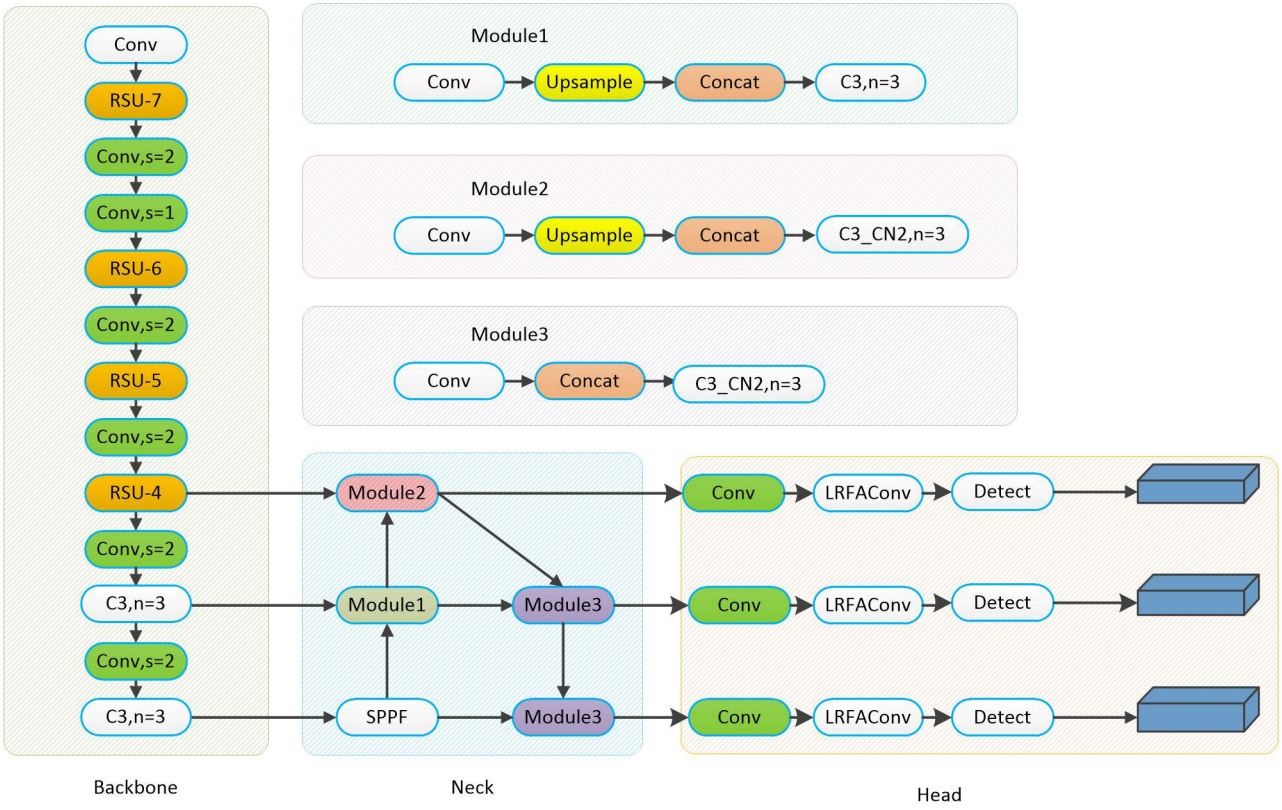


FIGURE 1. Overall network structure.

the original YOLOv5. Subsequently, the Bottleneck of the C3 module in YOLOv5 was replaced with the Block from ConvNeXt-V2, resulting in a more efficient C3_CN2 module, which was integrated into the Neck section of the network. Moreover, a novel spatial attention module, LRFAConv, was devised to more effectively capture spatial receptive field features. Finally, we improved the computation method for location loss by introducing the scale-insensitive NWDLoss, thereby enhancing the detection performance for smaller traffic sign targets. The network architecture is visually depicted in Figure 1.

B. RSU

The RSU (ReSidual U-block) module serves as a pivotal component in the U2Net, a salient object detection network designed to enhance feature extraction and information transmission capabilities. The RSU module typically adopts a U-shaped structure, encompassing two branches for downsampling and upsampling operations. Leveraging the concept of residual connections, the RSU module achieves the retention of low-level features and the fusion of high-level features, thus significantly improving the network's ability to express intricate features. Notably, U2Net has demonstrated remarkable efficacy in the task of salient object detection (SOD). The salient object detection task bears certain similarities to image segmentation and object detection tasks, as they all entail distinguishing foreground objects from the background within an image. U-shaped networks, renowned for their robust feature fusion capabilities, are extensively employed in image segmentation tasks. The RSU structure, short for ReSidual U-block, assumes a 'U' shape, involving multiple upsampling and downsampling operations alongside dilated convolution. The RSU family encompasses different depths, such as RSU-7, RSU-6, RSU-5, and RSU-4, enabling adaptability to varying network complexities. Figure 2 visually illustrates the RSU network architecture. In this paper, we leverage the RSU module primarily to augment the original YOLOv5 backbone network, aiming to enhance its feature extraction capacity and facilitate more effective detection, particularly for small targets. The incorporation of the RSU module is expected to bolster the overall performance and robustness of the YOLOv5-based traffic sign detection system.

C. ConvNeXt-V2

Inspired by the Convnext network combined with self-supervised MAE, ConvNeXt-V2 is a high-performance model. The main core structure of the network is Block, a depthwise convolution with a convolution kernel size of 7, a pointwise Convolution with an elevated dimension and a piontwise Convolution with a reduced dimension. This paper uses its core Block module, combined with the C3 structure of YOLOv5, to construct the C3_CN2 structure. And apply it to the Neck part of the detection network.The network structure is shown in Figure.3.

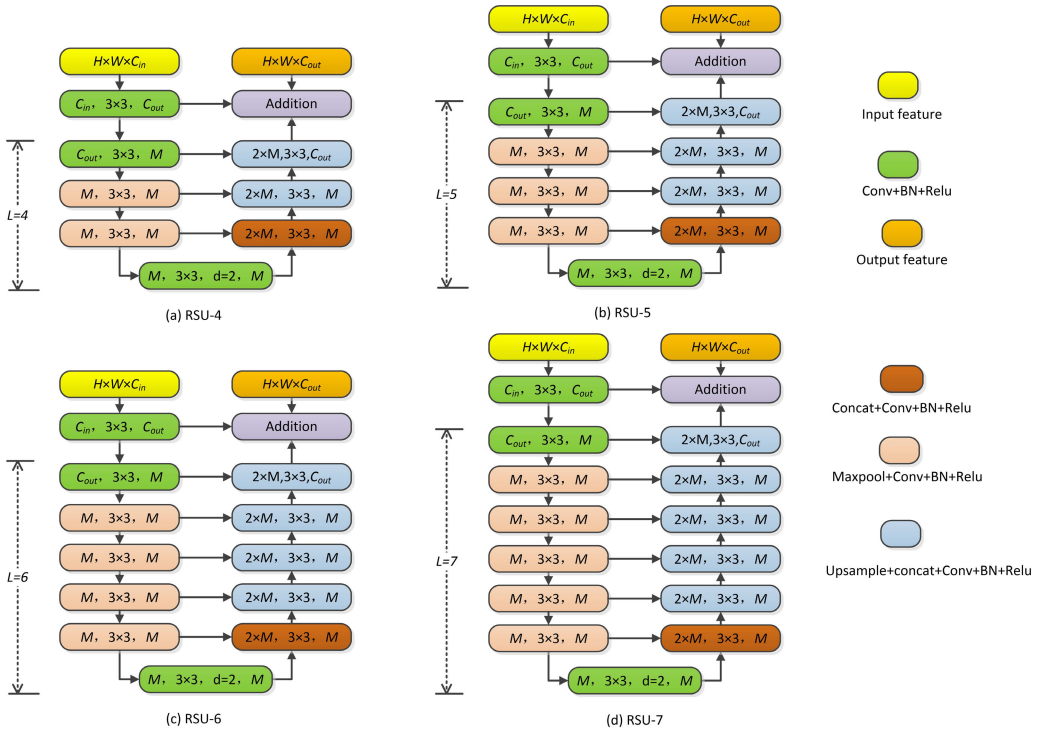


FIGURE 2. RSU.

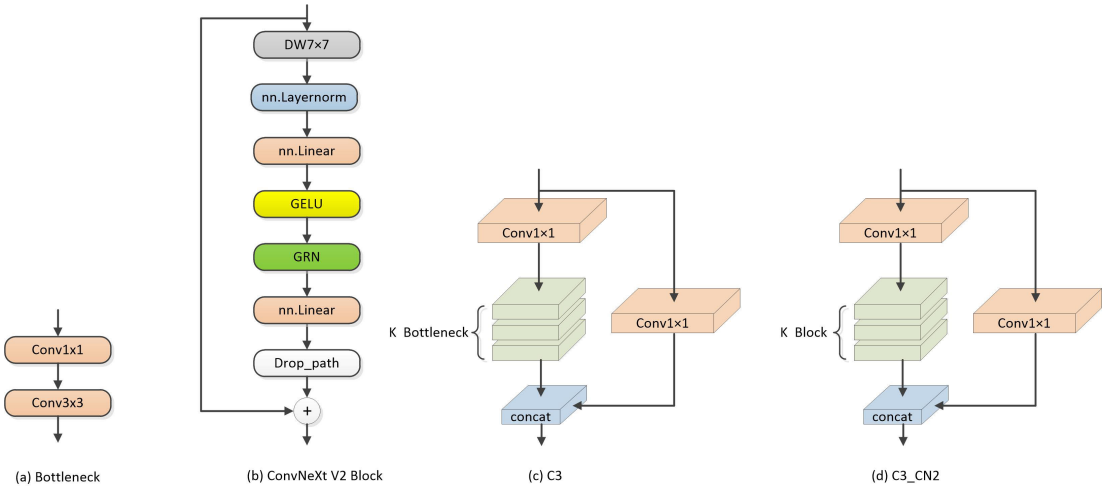
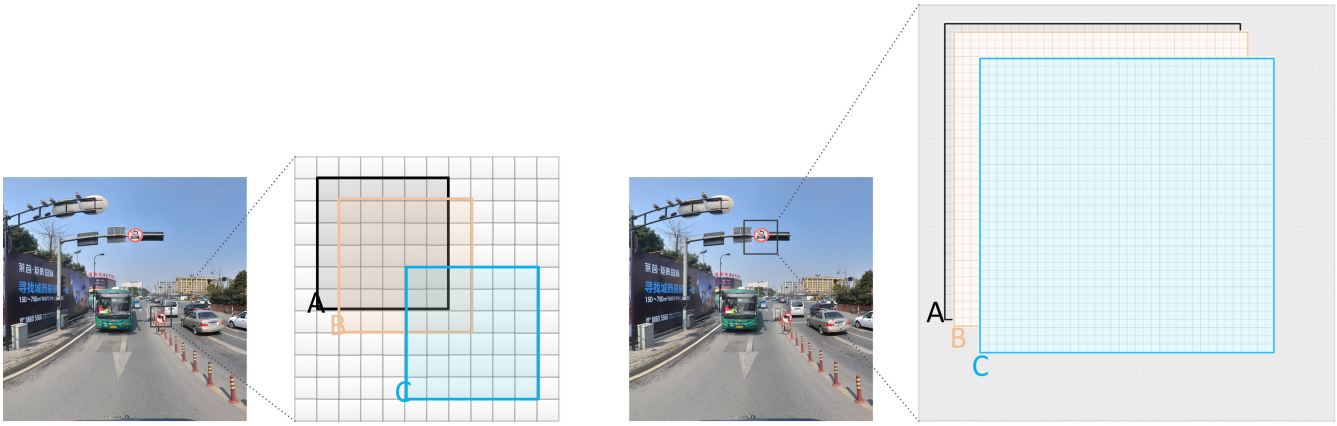


FIGURE 3. Using the block(b) module of Convnext-v2 network to replace Bottleneck (a) in C3 (c), construct C3_ CN2 (d) module.

D. NWDLoss

In the TT100k traffic light dataset, there are many small target objects, because of the small number of pixels and the lack of effective information. At the same time, YOLOov5 is a target detection network based on Anchor, and its IOU-based measurement method is too sensitive to the position change of small targets. That is, the sensitivity of IoU to objects of different scales varies greatly. For a small target area of 6 × 6 size, let it move down and right one unit, the value of IOU changes from 1 to 0.532, and then move right and down three units. At this time, the IOU is 0.059.However, for a medium-sized target of 36 × 36, its two movements lead to IOU changes of 0.896 and 0.653, as shown in Figure.4.

It can be seen from the above results that a slight position deviation leads to a significant decrease in IoU, which is not conducive to the detection of small targets. Therefore, this paper introduces a new index Normalized Gaussian Wasserstein Distance (NWD), which is insensitive to different scale targets and measures the coincidence degree



(a) IOU(A,B)=0.532, IOU(A,C)=0.059

(b) IOU(A,B)=0.896, IOU(A,C)=0.653

FIGURE 4. Moving change diagram.

of two boxes. The calculation is as follows equation 1:

$$\begin{aligned} &NWD \\ &= EXP\left(\frac{\sqrt[2]{(x-x^{gt})^2+(y-y^{gt})^2+\frac{(w-w^{gt})^2+(h-h^{gt})^2}{4}}}{Constant}\right) \\ &NWDLoss \\ &= 1-NWD \end{aligned} \tag{1}$$

NWD can partially replace the IOU metric in evaluating small targets. We need to combine both metrics to compute our localization loss function, as shown in Equation 2.

$$location_{loss} = \alpha NWDLoss + (1-\alpha)CIOULoss \tag{2}$$

After conducting multiple experiments, we have determined that the hyperparameter *Constant* should be set to 3, and the parameter α should be set to 0.3. These specific values have been chosen based on empirical observations and extensive trial and error.

E. LRFAConv

Spatial features have proven to be instrumental in improving network performance, as demonstrated by the effectiveness of attention mechanisms like Channel Attention (CA) [28] and Convolutional Block Attention Module (CBAM) [29]. Among spatial features, receptive field features play a critical role in small target detection algorithms, as they contain vital information required for accurate detection.

Inspired by a recently proposed spatial attention module, RFAConv [30], this paper introduces a more lightweight and efficient Receptive Field Attention module, named LRFAConv (Lightweight Receptive Field Attention Convolution). The primary focus of LRFAConv is to enhance the extraction of spatial receptive field features, thereby improving the network's ability to discern intricate details, especially in the context of detecting small targets.

The LRFAConv module leverages pooling operations to gather extensive local information and incorporates a 3×3 Depthwise convolution to facilitate meaningful feature interactions. Additionally, a softmax function is employed to emphasize the importance of relevant features. Implemented within the Head section of the networks, LRFAConv is complemented by a 1×1 convolutional layer, effectively reducing the number of feature channels and alleviating excessive computational burden, thereby enhancing network efficiency. The structural design of the LRFAConv module, as illustrated in Figure 5.

IV. EXPERIMENTAL ANALYSIS

A. EXPERIMENTAL ENVIRONMENT

The experiment in this paper is carried out under the Linux operating system. The experimental environment configuration is shown in Table.1:

TABLE 1. Experimental environment configuration.

Name	Version
CPU	Intel Xeon(R) E5-2683 V4
GPU	RTX 3090(24GB)
CUDA	11.3
Frame	PyTorch1.11.0
Language	Python3.8.10
Software	PyCharm

B. EVALUATION INDICATORS

In the target detection tasks, evaluation indicators are divided into two categories: one is under PACSALVOC evaluation indicators, and the other is under COCO evaluation indicators. This article will select mAP50 in the former and mAP50:95 in the latter as the corresponding evaluation indicators to detect the proposed algorithm on the traffic sign dataset.

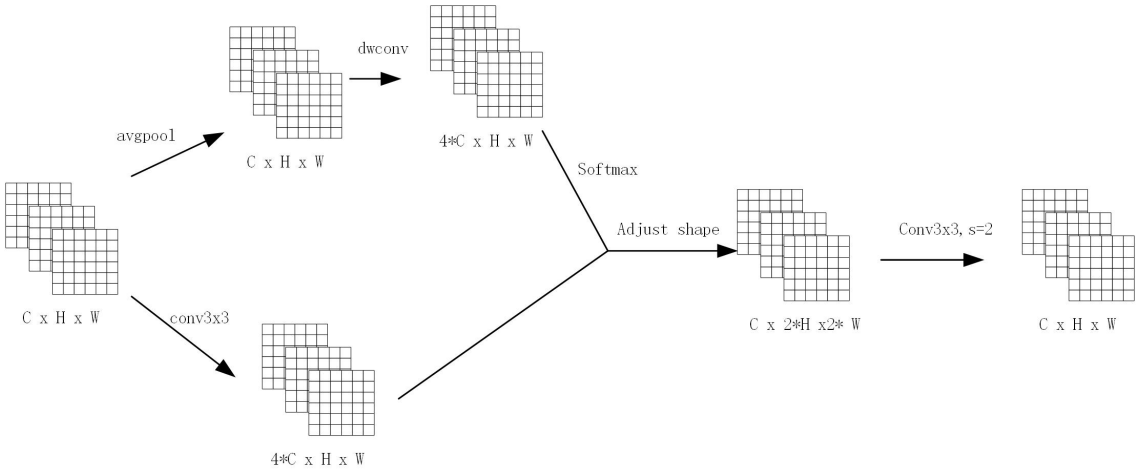


FIGURE 5. LRFAConv.

1) IoU

Intersection over Union (IoU) is used to evaluate the intersection between the target prediction box and the real box. Define as follows Equation 3:

$$IoU = \frac{A \cap B}{A \cup B} \tag{3}$$

2) mAP

A sample is divided into two categories: positive samples and negative samples. The classification results are shown in Table.2:

TABLE 2. Classification results.

sample grading	Positive forecast	Negative forecast
Positive sample	TP	FN
Negative samples	FP	TN

P(precision) is the accuracy rate, which is used to evaluate how many proportions of the positive samples predicted by the model are predicted correctly. The calculation is as follows Equation 4:

$$P = \frac{TP}{TP + FP} \tag{4}$$

R(recall) is the recall rate, which is used to evaluate how many proportions of all positive samples are correctly predicted by the model. The calculation is as follows Equation 5:

$$R = \frac{TP}{TP + FN} \tag{5}$$

AP (Average Precision) is the average accuracy value, calculated based on the curve shape of P and R. Our commonly used AP50 refers to setting the IoU value of the predicted box and the real box greater than 0.5 threshold and the maximum confidence score to TP, setting the excess predicted boxes of the same GTbox to FP, and FN is the number of GTboxes detected as missed. Based on this information, draw the P-R curve of AP50 and set different IoU thresholds to obtain other AP values. For example, AP55 and AP75 set the corresponding IoU thresholds to 0.55 and 0.75. The calculation of AP is shown in the following Equation 6.

$$AP = \int_0^1 precision\, d_{recall} \tag{6}$$

mAP(Mean Average Precision) is used to evaluate the average performance of the model to detect all categories. It is the average of the values of all categories, calculated as follows Equation 7, the number of categories:

$$mAP = \frac{1}{C} \sum_{j=1}^{c} AP_j \tag{7}$$

C. DATASET

1) TT100K

In this paper, the traffic light detection experiment uses the TT100 K public data set launched by Tencent and Tsinghua University. The image resolution of the data set is 2048 × 2048, and the quality is compared. Its training set contains 6105 pictures, and the verification set contains 3071 pictures. The data set has a serious category imbalance, and the number of instances of some traffic signs is very small. Therefore, this paper has carried out some screening of the categories, and requires that there are at least 100 pictures

TABLE 3. TT100K traffic sign class ID and names.

Class Id	Class Name
0-6	pn, pne, i5, p11, pl40, pl50, po
7-13	pl80, pl60, p26, i4, pl30, io, i2
14-20	pl100, p5, w57, il60, pl5, p10, ip
21-27	p23, il80, pl120, pr40, w59, p12, p3
28-34	w55, ph4.5, pl20, pm20, pg, pl70, pm55
35-41	il100, p27, w13, p19, ph4, p6, pm30

TABLE 4. Class ID and corresponding class names for the traffic sign categories in the LISA dataset.

Class Id	Class Name
0-5	stop, pedestrianCrossing, signalAhead, speedLimit35,speedLimit25, keepRight
6-11	addedLane, merge, yield,laneEnds, stopAhead, speedLimit45
12-17	speedLimit30, school, speedLimitUrdbl, schoolSpeedLimit25, turnRight, rightLaneMustTurn
18-23	speedLimit65, speedLimit40, truckSpeedLimit55, yieldAhead, roundabout, curveRight
24-29	speedLimit50, noLeftTurn, curveLeft, dip, slow, turnLeft
30-35	rampSpeedAdvisory45, noRightTurn, doNotEnter, zoneAhead25, zoneAhead45, thruTrafficMergeLeft
36-41	rampSpeedAdvisory50, speedLimit15, rampSpeedAdvisory20, doNotPass, thruMergeRight, thruMergeLeft
42-47	rampSpeedAdvisory35, rampSpeedAdvisory40, rampSpeedAdvisoryUrdbl, speedLimit55, intersection

FIGURE 6. Image display of the dataset.

for each type of traffic signs detected. After this filtering, 42 more common traffic signs are selected for experiments. The 42 types of traffic signs were tested according to the training set and the validation set. The number of pictures of 42 traffic signs selected in the data set is shown in Figure 6. Figure 7 shows some pictures of TT100K traffic signs in the data set,the class id and class name correspond to Table 3.

2) LISA

The TT100K dataset is considered the highest-quality traffic sign detection dataset in the current market, offering high-quality image samples with corresponding annotations. In this research, we made algorithmic improvements based on the TT100K dataset, leading to significant performance gains. To further validate the efficacy of our proposed UCN-YOLOV5 algorithm, we conducted supplementary experiments on the LISA [31] traffic sign dataset, which contains 6618 frames with 7855 annotations across 47 classes. However, we observed a noticeable class imbalance in the LISA dataset, potentially causing overfitting on classes with abundant samples and underperformance on those with limited samples. To address this issue, we conducted training and validation exclusively on classes with more than 100 instances, thereby mitigating the impact of data imbalance. Ultimately, we obtained 5827 frames from the LISA dataset, comprising 16 traffic sign categories, and partitioned them into training and validation sets using a ratio of 7:3. In Figure 8, we present the class distribution of instances in the LISA traffic sign dataset, highlighting the disparities in sample quantities among the different classes. For reference, Table 4 provides the Class ID and corresponding class names for the traffic sign categories in the LISA dataset.

TABLE 5. Model parameter settings.

Parameter	Value
Learning_rate	0.001
epoch	300
momentum	0.937
decay	0.0005

3) CCTSDB2021

In addition to the previously mentioned datasets, we also utilized the CCTSDB2021 [32] traffic sign dataset. Introduced by Changsha University of Science and Technology in 2021, this dataset categorizes all traffic signs into three classes: "mandatory," "prohibitory," and "warning." It comprises 16,356 training images and 1,500 validation images. Incorporating the CCTSDB2021 dataset allowed us to further validate the effectiveness and robustness of our proposed UCN-YOLOV5 algorithm across diverse traffic sign datasets, particularly in the context of Chinese traffic sign scenarios.

D. MODEL TRAINING

The random gradient descent method is used to propagate the input image back, and the training parameters are shown in Table.5.

In order to verify the feasibility of the network model proposed in this paper, 6105 images in the sample data set is used as the training data set to train the model,3071 images in the sample data set is used as the validation set for model training, and the data is filtered before model

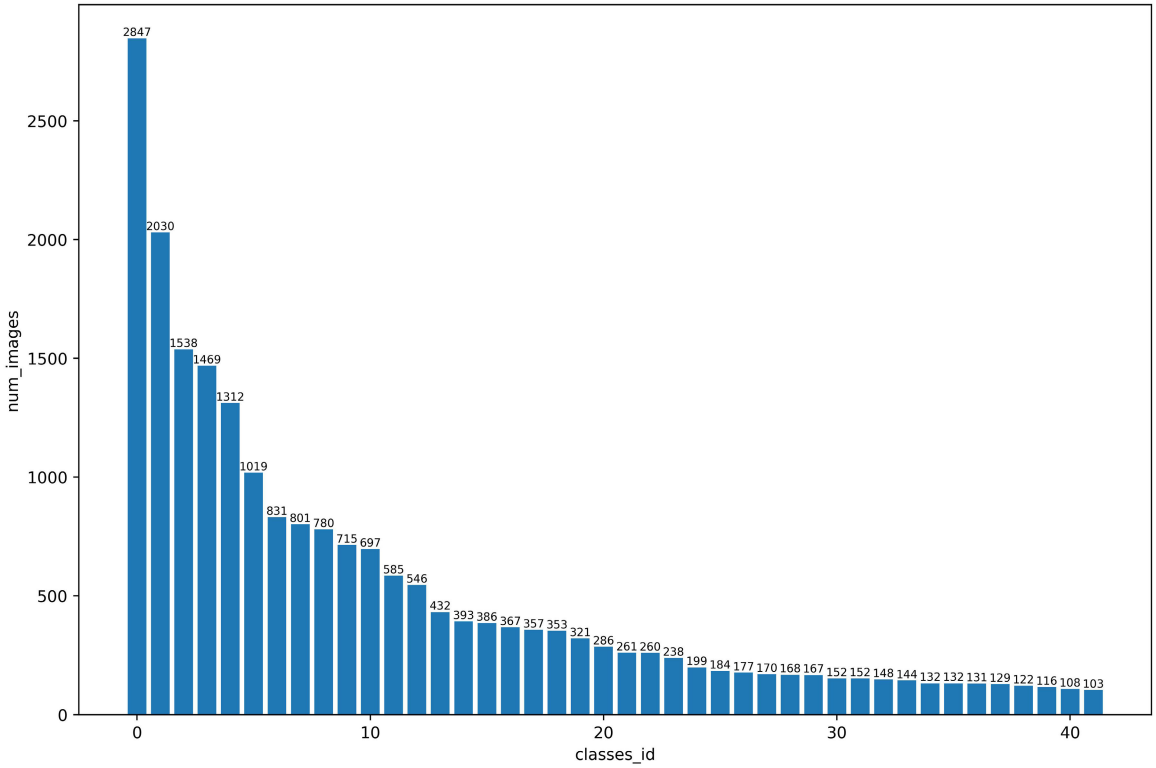


FIGURE 7. The class distribution of instances in the TT100K traffic sign dataset.

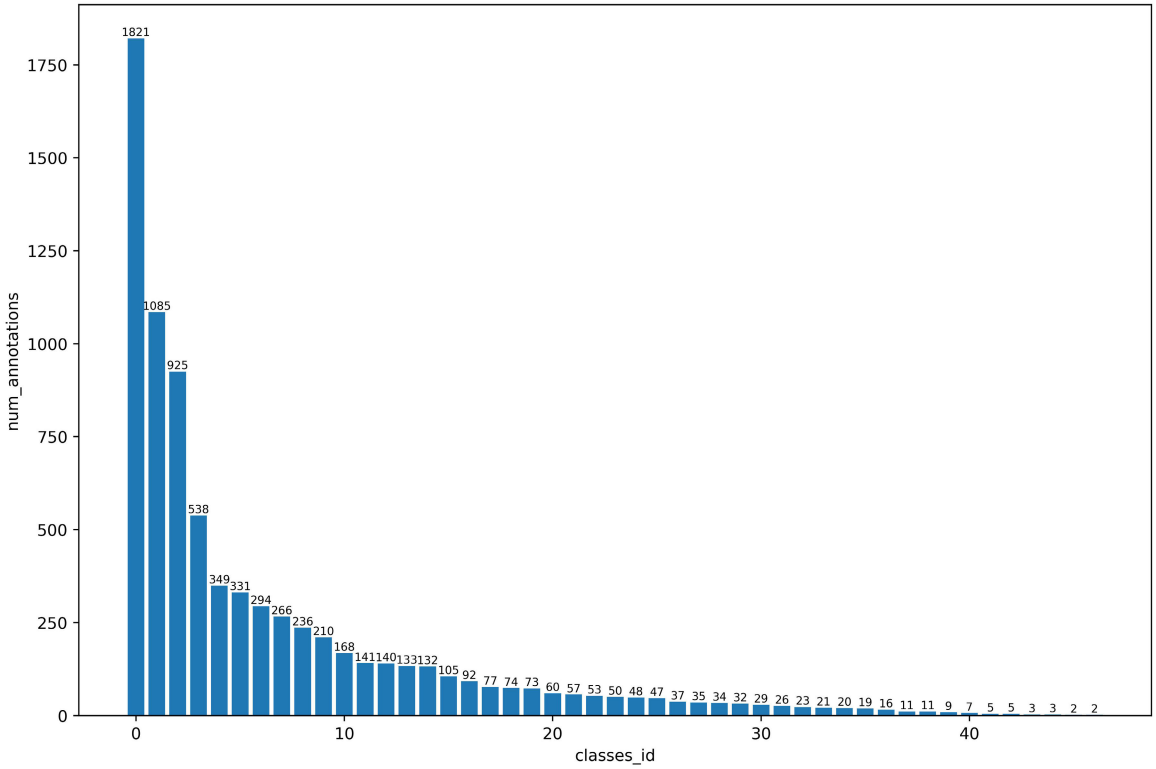


FIGURE 8. The class distribution of instances in the LISA traffic sign dataset.

TABLE 6. Performance comparison of different models.

Modles	Params	GFLOPs	Map.5	Map.5:.95	FPS	Imgsz
YOLOv5S	7.12M	16.2	79.7%	60.7%	84.7	640
UCN-YOLOv5-S	7.44M	40.2	**85.6%**	**65.1%**	44.8	640
YOLOv5M	21.02M	48.5	84.9%	66.5%	60.6	640
UCN-YOLOv5-M	19.68M	81.3	**89.8%**	**69%**	33.2	640
YOLOv3	61.72M	155.2	84.6%	63.9%	34.7	640
YOLOv4	52.69M	119.7	82.6%	62.2%	35.7	640
YOLOv5L	46.33M	108.5	86.6%	68.5%	38.9	640
YOLOv6L	59.57M	150.6	88.3%	69.6%	35.0	640
YOLOv8L	43.64M	165.0	90.0%	71.1%	38.9	640
UCN-YOLOv5-L	46.67M	160.7	**91.2%**	**71.3%**	20.9	640

TABLE 7. Comparative analysis of UCN-YOLOv5 and YOLOv5 algorithms on LISA traffic sign dataset.

Modles	Params	GFLOPs	Map.5	Map.5:.95	FPS	Imgsz
YOLOv5S	7.05M	15.9	96.7%	87.8%	77.5	640
UCN-YOLOv5-S	7.37M	39.9	**97.0%**	**92.3%**	44.8	640
YOLOv5M	20.91M	48.1	97.1%	93.3%	58.1	640
UCN-YOLOv5-M	19.6M	80.8	**97.2%**	**94.4%**	38.8	640
YOLOv5L	46.19M	108.0	97.4%	95.0%	49.5	640
UCN-YOLOv5-L	46.62M	160.5	**97.4%**	**95.3%**	26.6	640

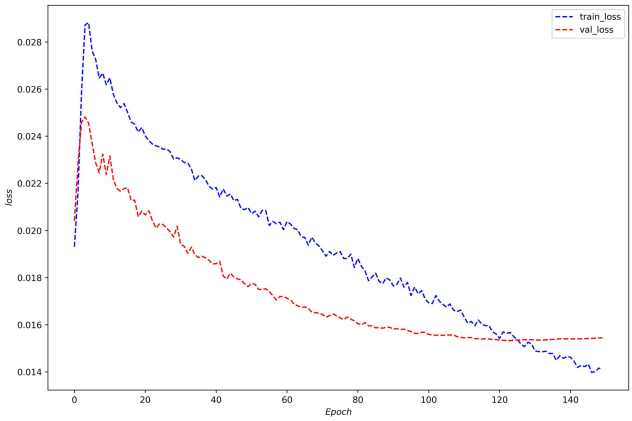


FIGURE 9. Loss analysis.

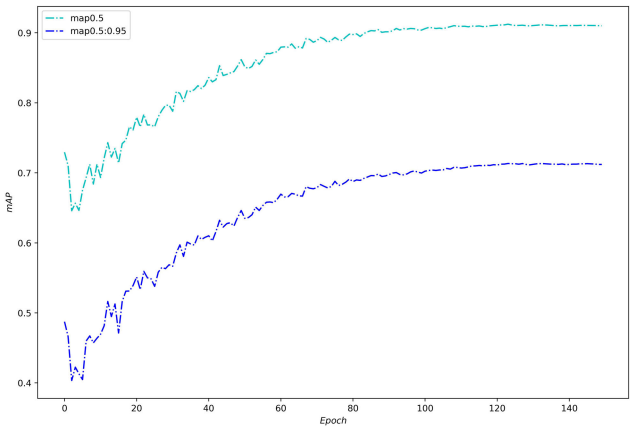


FIGURE 10. Convergence analysis.

training. We initially trained for 150 epochs and after obtaining the training weights, we continued training for an additional 150 epochs. Figure 9 and Figure 10 display the resulting iterative training loss and mAP function values, respectively.The loss function of the model is kept below 0.015 after 300 iterations, while the change of the index mAP curve is kept below 0.002 after 300 iterations. It is proved that the detection accuracy of the proposed method is satisfactory in the test of traffic sign target detection. Predictions are made using the weights obtained after training and their detection effects are evaluated. The detailed results of the detection effect can be seen in Figure 11.

V. MODEL COMPARISON

A. MODEL PERFORMANCE COMPARISON OF DIFFERENT METHODS

To assess the efficacy of UCN-YOLOv5 in traffic sign detection, we conducted a comparative study using the TT100K dataset at three different scales, namely Small (S), Medium (M), and Large (L). In this evaluation, we compared the UCN-YOLOv5 model against YOLOv5 and also included classic YOLO versions, such as YOLOv3, YOLOv4, YOLOv6, and YOLOv8. The experimental results are presented in Table 6.

In addition, we further conducted extensive validation of the algorithmic advancements using the LISA and CCTSDB2021 traffic sign datasets, and compared the performance against the baseline model YOLOv5. The experimental results, as shown in Tables 7 and 8, demonstrate that UCN-YOLOv5 continues to exhibit promising results.

B. ABLATION EXPERIMENT

Based on YOLOv5L, ablation experiments were performed on the TT100K dataset. The RSU structure of U2Net was introduced to change the backbone network of YOLOv5.Referring to the Block structure of ConvNeXt-V2, combined with the C3 network, the C3_CN2 module was constructed and used in the Neck part of YOLOv5.The

FIGURE 11. Visualization of the results of the detection effect.

LRFAConv module with more enhanced spatial Receptive field features is used in the network Head. Next, NWDLoss, which is insensitive to scale information and suitable for small target detection, is introduced. The results of related ablation experiments are shown in Table.9:

Considering the performance requirements of the device, UCN-YOLOv5 is divided into UCN-YOLOv5-S, UCN-YOLOv5-M, UCN-YOLOv5-L according to the width and depth of the network. Their width and depth attribute values are consistent with the original YOLOV5, and their (depth, width) are (0.33, 0.50), (0.67, 0.75), (1.0, 1.0), respectively. Depth affects the number of iterations n of Blocks in C3 and C3_CN2 modules. The calculation is as follows Equation 8:

$$n = Max(round(3 * depth), 1) \tag{8}$$

In the formula, the function of the round ( ) function is rounded, the function of the Max ( ) function is to take the maximum value, and * is the multiplication symbol. Therefore, we obtain that the n value of UCN-YOLOv5-S

TABLE 8. Comparative analysis of UCN-YOLOv5 and YOLOv5 algorithms on CCTSDB2021 traffic sign dataset.

Modles	Params	GFLOPs	Map.5	Map.5:.95	FPS	Imgsz
YOLOv5S	7.02M	15.8	82.5%	54.3%	75.2	640
YOLOv5M	20.86M	48.0	86.7%	58.0%	57.5	640
UCN-YOLOv5-S	7.34M	39.8	**86.3%**	**58.1%**	41.5	640
YOLOv5L	46.12M	107.8	86.3%	59.3%	48.1	640
UCN-YOLOv5-L	46.60M	160.39	**87.5%**	**59.3%**	25.9	640

TABLE 9. Comparison of ablation results.

Model	Module				Params	GFLOPs	Map.5	Map.5:.95	Imgsz
	RSU	Convnext-v2	NWD	LRFAConv					
YOLOv5L	×	×	×	×	46.33M	108.5	86.6%	68.5%	640
	✓	×	×	×	42.97M	146.3	89.0%	69.8%	640
	✓	✓	×	×	41.07M	143.0	89.1%	70.3%	640
	✓	✓	✓	×	41.07M	143.0	90.1%	70.9%	640
	✓	✓	✓	✓	46.67M	160.68	91.2%	71.3%	640

is 1, and the n value of UCN-YOLOv5-M is 2. width affects the number of channels. Its calculation formula is as follows Equation 9:

$$Channel = Ceil(channel * width/divisor) * divisor \quad (9)$$

The function of the Ceil ( ) function is to round up, and the divisor value is 8. Therefore, if Channel = 512, the Channel value obtained in the UCN-YOLOv5-S network is 256, and the Channel value obtained in the UCN-YOLOv5-M network is 384.

VI. CONCLUSION

In this paper, some novel breakthrough technologies are added to YOLOv5. The U-shaped module RSU in U2Net is widely used in the backbone of the network. The Bottleneck in the C3 structure is replaced by Block in the high-performance network ConvNeXt-V2 to construct the C3 _ CN2 module and use it in the Neck part of the network. Using the lightweight receptive field attention module LPFAConv proposed in this paper in the Head section can better extract receptive field features. At the same time, NWDLoss is added to strengthen the detection of small targets in the data set. Finally, an advanced detector UCN-YOLOv5 is formed, which can accurately detect traffic signs in traffic scenes. Experiments on the TT100K, LISA, CCTSDB2021 Traffic Sign Dataset show that compared with the baseline model YOLOv5, UCN-YOLOv5 achieves considerable progress on the two evaluation indicators of Map.5 and Map.5: .95.

However, the incorporation of ConvNeXt-V2 network's Block module in UCN-YOLOv5, when compared to the original YOLOv5 algorithm, results in increased training time for the network. Additionally, the LRFAConv module exhibits more challenging convergence behavior, necessitating a greater number of training epochs on the dataset. These aspects impose higher hardware requirements. Nevertheless, achieving better detection results is of utmost importance in certain application scenarios. In the future, further evaluations will be conducted using diverse datasets to assess the proposed methods comprehensively, catering to various application demands and achieving heightened detection accuracy.

REFERENCES

[1] Y. Liu and D. R. Huang, "Color standardization of traffic sign images based on multi-color space cascade classification," *Comput. Eng.*, vol. 46, no. 9, pp. 233–241, 2020.

[2] W. L. Li, X. G. Li, and Y. Qin, "Real-time traffic sign detection algorithm based on dynamic threshold segmentation and SVM," *J. Comput.*, 31, no. 6, pp. 258–273, 2020.

[3] D. Temel, M.-H. Chen, and G. AlRegib, "Traffic sign detection under challenging conditions: A deeper look into performance variations and spectral characteristics," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3663–3673, Sep. 2020.

[4] Y. Xiaoling, J. Weixin, and Y. Haoran, "Recognition and detection of traffic signs based on YOLOV5," *Inf. Technol. Informatization*, no. 4, pp. 28–30, 2021.

[5] D. Tabernik and D. Skocaj, "Deep learning for large-scale traffic-sign detection and recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 4, pp. 1427–1440, Apr. 2020.

[6] J. Cao, J. Zhang, and W. Huang, "Traffic sign detection and recognition using multi-scale fusion and prime sample attention," *IEEE Access*, vol. 9, pp. 3579–3591, 2021.

[7] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German traffic sign recognition benchmark: A multi-class classification competition," in *Proc. Int. Joint Conf. Neural Netw.* Washington, DC, USA: IEEE Press, Jul. 2011, pp. 1453–1460.

[8] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition," *Neural Netw.*, vol. 32, pp. 323–332, Aug. 2012.

[9] 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)*, New York, NY, USA, Jun. 2016, pp. 2110–2118.

[10] R. Girshick, J. Donahue, and T. Darrell, "Rich feature hierarchies for accurte object detection and semantic segmentation," in *Proc. 27th IEEE Conf. Comput. Vis. Pattern Recognit.*, New York, NY, USA, Jul. 2014, pp. 580–587.

[11] R. Girshick, "Fast R-CNN," in *Proc. 28th IEEE Conf. Comput. Vis. Pattern Recognit.*, New York, NY, USA, Dec. 2015, pp. 1440–1448.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[13] Z. Liang, J. Shao, D. Zhang, and L. Gao, "Traffic sign detection and recognition based on pyramidal convolutional networks," *Neural Comput. Appl.*, vol. 32, no. 11, pp. 6533–6543, Jun. 2020.

[14] S. Zhou, X. L. Zhi, and D. Liu, "A convolutional neural network-based method for small traffic sign detection," *J. Tongji Univ. Natural Sci.*, vol. 47, no. 11, pp. 1626–1632, 2019.
[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified real-time object detection," in *Proc. 29th IEEE Conf. Comput. Vis. Pattern Recognit.*, New York, NY, USA, Jun. 2016, pp. 779–788.
[16] J. Redmon and A. Farhadi, *YOLOv3: An Incremental Improvement*. Ithaca, NY, USA: Cornell University, 2018.
[17] W. Liu, D. Anguelov, and D. Erhan, "SSD: Single shot multibox deteector," in *Proc. 14th Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2016, pp. 21–37.
[18] Y. Wu, Z. Li, Y. Chen, K. Nai, and J. Yuan, "Real-time traffic sign detection and classification towards real traffic scene," *Multimedia Tools Appl.*, vol. 79, nos. 25–26, pp. 18201–18219, Jul. 2020.
[19] S. P. Rajendran, L. Shine, R. Pradeep, and S. Vijayaraghavan, "Real-time traffic sign recognition using YOLOv3 based detector," in *Proc. 10th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2019, pp. 1–7.
[20] H. Zhang, L. Qin, J. Li, Y. Guo, Y. Zhou, J. Zhang, and Z. Xu, "Real-time detection method for small traffic signs based on YOLOv3," *IEEE Access*, vol. 8, pp. 64145–64156, 2020.
[21] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," 2022, *arXiv:1710.09412*.
[22] T. Liang, H. Bao, W. Pan, and F. Pan, "Traffic sign detection via improved sparse R-CNN for autonomous vehicles," *J. Adv. Transp.*, vol. 2022, pp. 1–16, Mar. 2022.
[23] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 1314–1324.
[24] J. Wang, Y. Chen, Z. Dong, and M. Gao, "Improved YOLOv5 network for real-time multi-scale traffic sign detection," *Neural Comput. Appl.*, vol. 35, no. 10, pp. 7853–7865, Apr. 2023.
[25] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jagersand, "U2-net: Going deeper with nested U-structure for salient object detection," *Pattern Recognit.*, vol. 106, Oct. 2020, Art. no. 107404.
[26] X. Wu, D. Hong, and J. Chanussot, "UIU-net: U-net in U-net for infrared small object detection," *IEEE Trans. Image Process.*, vol. 32, pp. 364–376, 2023.
[27] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11976–11986.
[28] 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. 13713–13722.
[29] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
[30] X. Zhang, C. Liu, D. Yang, T. Song, Y. Ye, K. Li, and Y. Song, "RFAConv: Innovating spatial attention and standard convolutional operation," 2023, *arXiv:2304.03198*.
[31] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1484–1497, Dec. 2012.
[32] J. Zhang, X. Zou, L.-D. Kuang, J. Wang, R. S. Sherratt, and X. Yu, "CCTSDB 2021: A more comprehensive traffic sign detection benchmark," *Human-Centric Comput. Inf. Sci.*, vol. 12, pp. 1–23, May 2022, doi: 10.22967/HCIS.2022.12.023.

PEILIN LIU was born in Jiangxi, China. He received the bachelor's degree from the East China University of Technology. He is currently pursuing the master's degree in information and communication engineering with Hainan University. His main research interests include deep learning and computer vision.

ZHAOYANG XIE was born in Beijing, China. He is currently pursuing the master's degree with the School of Information and Communication Engineering, Hainan University. His research interests include artificial intelligence and computer vision.

TAIJUN LI received the bachelor's degree in radio engineering and the master's degree in electronic materials and devices from the South China University of Technology, in 1984 and 1987, respectively. After obtaining the master's degree, he taught with the university and has been teaching with the School of Information Science and Technology, Hainan University, where he joined as a Faculty Member, in December 1987. In 2001, he was promoted to a Professor and a Master Supervisor. He has published more than 80 academic papers and authored three academic monographs. His main research interests include multimedia communication, networks and streaming media, and image information processing.