TLDM: An Enhanced Traffic Lights Detection Model Based on YOLOv5

Jun Song , Tong Hu , Zhengwei Gong , Youcheng Zhang * , Mengchao Cui *

Posted Date: 2 July 2024

doi: 10.20944/preprints202407.0228.v1

Keywords: Traffic lights, unmanned systems, YOLOv5, mosaic-9, Squeezed-and-Excitation(SE), EIoU_loss

Article

TLDM: An Enhanced Traffic Lights Detection Model Based on YOLOv5

Jun Song [1], Tong Hu [1], Zhengwei Gong [1], Youcheng Zhang [1] and Mengchao Cui [2,*]

[1] College of Information Science and Technology, Nanjing Forestry University, 159 Longpan Road, Nanjing 210037, China; songjun@njfu.edu.c.n (J.S.); 64928773@qq.com (T.H.); georgecumt@163.com (Z.G.) ; 3088105971@qq.com (Y.Z.)

[2] School of Foreign Languages, China University of Political Science and Law, 25 West Tu Cheng Road, Haidian District, Beijing 100088, China.

* Correspondence: cu008564@cupl.edu.cn

Abstract: The traffic light detection and recognition are crucial for enhancing the security of unmanned systems. This study proposes a YOLOv5-based traffic light detection algorithm to tackle the challenges posed by small targets and complex urban backgrounds. Initially, the mosaic-9 method is employed to enhance the training dataset, thereby boosting the network's ability to generalize and adapt to real-world scenarios. Furthermore, the network incorporates the Squeezed-and-Excitation (SE) attention mechanism to improve. Moreover, the YOLOv5 algorithm's loss function has been optimized by substituting it with EIoU_loss, which addresses issues like missed detection and false alarms. Experimental results demonstrate that the model , trained with this enhanced network, achieves a 99.4% mAP on a custom dataset, which is 6.3% higher than the original YOLOv5, while maintaining a detection speed of 74 f/s. Therefore, this algorithm offers higher detection accuracy and effectively meets real-time operational requirements.

Keywords: Traffic lights; unmanned systems; YOLOv5; mosaic-9; Squeezed-and-Excitation(SE); EIoU_loss

0. Introduction

In recent years, advancements in science, technology, and artificial intelligence have led to the development of autonomous and assisted driving technologies. These technologies, which utilize deep learning and computer vision, are progressively supplanting traditional target detection algorithms in road traffic scenarios. Traffic lights are vital for road safety. Efficient and accurate detection of their status allows intelligent vehicles to gather crucial intersection information beforehand, thereby preventing accidents and enhancing passenger safety. Early traffic light detection algorithms primarily relied on traditional image processing methods which involved manual feature extraction using sliding windows and the application of machine learning classifiers for detection and recognition [1]. Omachi et al.[2] employed the Hough transform to convert RGB images into standardized forms, accurately pinpointing traffic light positions within candidate regions. Zhu Yongzhen et al.[3] transformed the image's color space to HSV, applied the H color threshold to segment candidate areas, and used the Hough transform to predict suspected regions after performing grayscale morphological operations on the original image. The fusion and filtering of these methods facilitated the recognition of traffic light information.. These traditional image processing algorithms depend heavily on task-specific, hand-designed model features, which leads to low robustness and inadequate generalization, thus failing to meet the real-time requirements of complex traffic scenarios.

With the continuous development of convolutional neural network architecture, object detection algorithms are experiencing significant growth. Currently, deep learning-based object detection algorithms are categorized into two types: two-stage algorithms, exemplified by Faster R-CNN [4], which generate preliminary candidate bounding boxes and subsequently refine classifications, offering high accuracy but slower detection speeds. One-stage algorithms, such as Single Shot Multibox Detector (SSD)[5] and You Only Look Once (YOLO)[6], bypass the candidate region

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 July 2024 doi:10.20944/preprints202407.0228.v1

generation, enabling faster detection speeds. However, these one-stage algorithms generally exhibit reduced accuracy compared to their two-stage counterparts. Pan Weiguo et al.[7] enhanced domestic traffic signal data, applied the Faster R-CNN algorithm to a custom dataset, and identified the best feature extraction network through experimental analysis to detect and recognize traffic signals, though the detection efficiency remained low. Literature [8] describes a multi-task convolutional neural network, based on CIFAR-10, designed to detect traffic lights in complex environments; however, the model's generalization and robustness were found lacking. Literature[9] proposes the SplitCS-Yolo algorithm for the rapid detection and recognition of traffic lights, noted for its robustness, although it requires improvements in detecting yellow and digital traffic lights accurately. Literature[10] improves YOLOv3 by integrating two additional residual units into the second residual block of Darknet53, boosting the network's ability to detect small targets with high accuracy, though it falls short of real-time processing requirements. Yan et al.[11] enhanced YOLOv5 using K-means clustering, which sped up traffic light detection in the BDD100K dataset, albeit at the cost of increased model complexity. While current deep learning-based traffic light detection algorithms circumvent issues like manual feature extraction and task-specific dependence seen in traditional methods, they still face challenges including complex network architectures, extensive parameterization, reduced detection efficiency, and elevated training cost. The detection performance of these enhanced methods on traffic lights fails to meet practical requirements, struggling to balance speed and accuracy effectively.

To address issues in traditional image processing methods and daily target detection algorithms, a traffic signal light detection model is designed based on the YOLOv5 target detection algorithm. Mosaic-9 method is utilized to enhance the training set for better adaptation to the application scenario. Additionally, the SE attention mechanism is incorporated into the network to emphasize the target features in the image and enhance the feature extraction ability of the algorithm. Finally, the EIoU_loss is employed to optimize the training model and address missing and false detection. The experimental results show the effectiveness of the proposed traffic signal detection algorithm in the urban road traffic scene using a self-made domestic urban traffic signal dataset, yielding favorable detection outcomes.

1. YOLOv5 Network Structure

YOLOv1 to YOLOv5, YOLO series[12], represent single-stage object detection algorithms, and has integrated many advantages of deep learning object detection networks. For this study,, YOLOv5s, with the smallest depth and the fastest speed in YOLOv5 model[13], was chosen as the base network for the experiment. Its network structure is illustrated in Figure 1 and consists of four parts: Input, Backbone, Neck and Prediction.

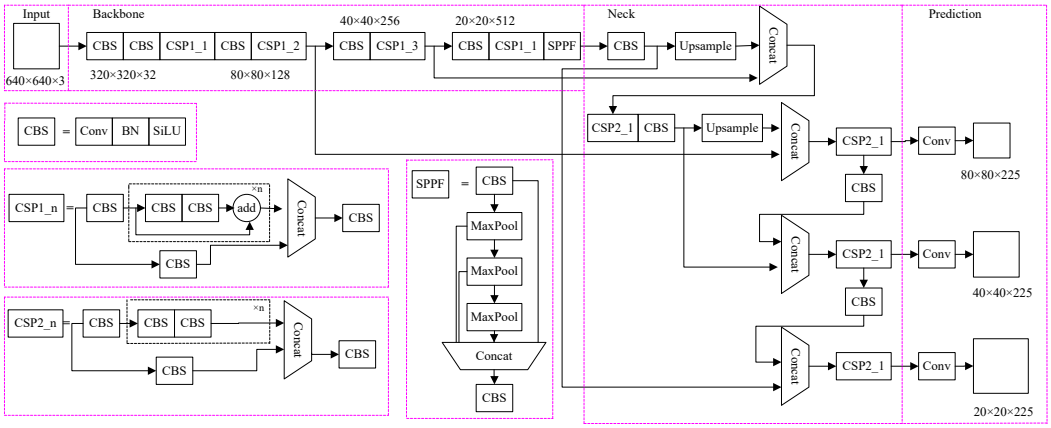

Figure 1. YOLOv5 network structure.

1.1. Input

 doi:10.20944/preprints202407.0228.v1

Firstly, he input terminal of the YOLOv5 network can perform mosaic data enhancement on the data set. This data enhancement method concatenates four images using random scaling, random cropping and random arrangement to enrich the data set. In particular, the random scaling function adds numerous small targets, improving the detection ability of small objects and enhancing the network robustness. Specifically, the YOLOv5 network can adaptively generate various prediction boxes during model training and use NMS(Non Maximum Suppression) to select the prediction box that is closest to the real box. Finally, to accommodate input images of different sizes, the YOLOv5 network's adaptive scaling image function allows the image to be scaled to the appropriate size before input to the network, preventing problems such as mismatching between the feature tensor and the fully connected layer.

1.2. Backbone

Secondly, the backbone consists of CBS, CSP and SPPF modules. The CBS module performs conventional convolution, Batch Normalization (BN) and SiLU activation function, primarily handling downsampling. The first CBS module in the Backbone has a 6×6 convolution kernel size, suitable for large input image resolution to capture global features effectively. Subsequent CBS modules use 3×3 convolution kernel size. The CSP module focuses on feature extraction and integrates feature information from different levels through a cross-stage structure to minimize gradient information repetition. The SPPF module is an improved version of Spatial Pyramid Pooling (SPP)[14] module. It processes input features through three 5×5 maximum pooling operations, retaining the advantages of the SPP module in reducing repeated feature extraction and calculation cost.

1.3. Neck

Thirdly, the neck layer of YOLOv5 network primarily combines Feature pyramid network (FPN)[15] and path aggregation network (PAN)[16]. The feature maps from different layers are concatenated, as illustrated in Figure 2. High-level features contain rich semantic information, but have low the resolution, leading to inaccurate target location or even partial disappearance of the target. In contrast, low-level features provide accurate target location with high resolution but limited semantic information. FPN transmits deep semantic features from top to bottom, while PAN transmits target location information from bottom to top. Through fusing top-down and bottom-up feature information, the model transmits the feature information of objects of different sizes, addresses the problem of multi-scale object detection, shortens the information propagation path and obtains richer feature information.

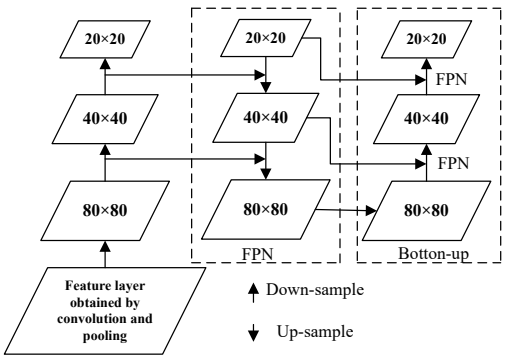

Figure 2. Feature pyramid network(PFN) and path aggregation network(PAN).

1.4. Prediction

Finally, The prediction component of the YOLOv5 network encompasses bounding box regression loss, confidence loss and classification loss. bounding box regression loss function utilizes CIoU_Loss, which comprehensively considers three important geometric factors: overlapping area, center point distance and aspect ratio. This approach enhances the prediction box regression speed

 doi:10.20944/preprints202407.0228.v1

and accuracy, particularly for targets with overlapping shading [17]. The YOLOv5 uses a weighted NMS to sift through multiple target anchor boxes and eliminate redundant candidate boxes during post-processing of target detection.

2. Improved YOLOv5 Model

2.1. Data Enhancement

Data annotation is time-consuming in practice. The best approach to reduce annotation time is to create virtual data and add it to the training set. YOLOv5 utilizes the Mosaic data augmentation method, which involves cropping, scaling, and randomly combining 4 images into a new image to increase the number of target samples. During normalization, the network calculates all 4 images simultaneously, thereby boosting training speed.

Due to the scarcity of useful traffic signal light images for sampling, this study adopts the Mosaic-9 method as shown in Figure 3. Nine images are randomly selected from the datasets as the original images for Mosaic data augmentation. The process of combining the 9 images involves random cropping. Considering that traffic lights onely occupy a relatively small portion of the original images, the random scaling factor for the 9 original images should not be too high to avoid losing features of small targets. Subsequently, the 9 images are randomly cropped, scaled, and merged to form a new image.

Figure 3. Mosaic-9 data enhancement flowchart.

2.2. Squeeze-and-Excitation Attention Mechanism (SE)

Detecting traffic lights in this experiment presents challenges due to the complex image background, small pixel size, and indistinct features that are susceptible to background interference. The original model's use of convolution for feature extraction may result in information loss and subpar target detection. To address this issue of convolutional neural networks struggling with global feature extraction, employing channel attention SENet[18] can significantly improve the model's performance.

The SE module acts as a channel attention mechanism that enhances the input feature map while maintaining its original size. In this study, the SE module is positioned at the end of the Backbone to reinforce the overall channel features,, enabling the subsequent neck part to effectively consolidate crucial features and ultimately enhance the model's performance. The structure of the SE module,, depicted in Figure 4, comprises three main components: compression, channel feature learning, and excitation.Firstly, the compression phase reduces the input feature map from W×H×C to a compact 1×1×C feature map via global average pooling. Next, in the channel feature learning stage, a 1×1×(C/r) feature map is generated using a 1×1 convolutional kernel with a stride of 1 and the SiLU activation function, where r represents the channel scaling factor. Subsequently, the channel weight coefficients for the 1×1×C feature map are computed through a 1×1 convolutional layer with a stride of 1 and the Sigmoid activation function. Finally, in the excitation step, the original input features are multiplied by the channel weight coefficients to produce a feature map with channel attention, assigning varying significance to channel features based on their respective weight coefficients.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 July 2024 doi:10.20944/preprints202407.0228.v1

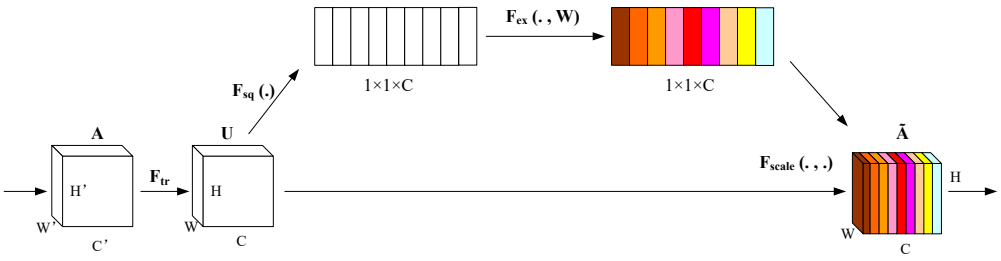

Figure 4. Squeeze-and-Excitation attention mechanism.

2.3. EIoU_loss Function

The formula for calculating the CIoU_loss function used in the original YOLOv5 is presented in formula (1) :

$$loss_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v$$
$$\alpha = \frac{v}{(1 - IoU) + v} \tag{1}$$
$$v = \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h})^2$$

The formula includes the following variables: b and bgt denote the center points of the Predicted Box (PB) and Ground truth (GT), ϱ2() represents the Euclidean distance, α is a positive equilibrium parameter, c represents the shortest diagonal length of PB and GT, and v is the aspect ratio of PB and GT. Additionally, wgt, hgt, w, and h represent the length and width of PB and GT, respectively.

While CIoU_Loss considers the center point distance, aspect ratio, and overlap area of bounding box regression, further analysis of the formula reveals that it solely reflects the difference in aspect ratio without capturing the actual disparity between width and height and their respective confidence levels. Consequently, literature [19] introduced the EIoU loss function by isolating the aspect ratio from CIoU, and its calculation formula is shown in formula (2).

$$loss_{EIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{c_w^2} + \frac{\rho^2(h, h^{gt})}{c_h^2} \tag{2}$$

In the formula, cw and ch represent the width and height, respectively, of the smallest bounding Box that encompasses both boxes simultaneously. EIoU_Loss not only accelerates the convergence of the prediction frame compared with CIoU_Loss, but also improves the regression accuracy. Therefore, this paper opts for EIoU_Loss over CIoU_Loss.

3. Experimental Results and Analysis

3.1. Collection of Datasets

Currently, existing open source traffic signal datasets, such as LISA and LARA datasets, are predominantly collected from foreign roads with a uniform background, high repetition rate, and significant regional variations both domestically and internationally. Conversely, domestic open source traffic sign datasets like TT100K and CCTSDB contain limited signal data, making them unsuitable for traffic signal detection. Therefore, this paper adopts the approach of creating custom datasets.

Table 1. Distribution of categories in traffic signal datasets.

Label category	Red	Green	Yellow
number	758	740	689

 doi:10.20944/preprints202407.0228.v1

In complex urban traffic environments in China, traffic signal data was collected using two methods: network screening and real-world photography, resulting in approximately 2000 images. Subsequently, the LabelImg labeling software was employed for manual labeling, resulting in three distinct label categories, as illustrated in Table 1. The dataset was then partitioned into a training set and a test set in a 3:1 ratio. Figure 5 presents a partial sample of the custom traffic signal dataset.

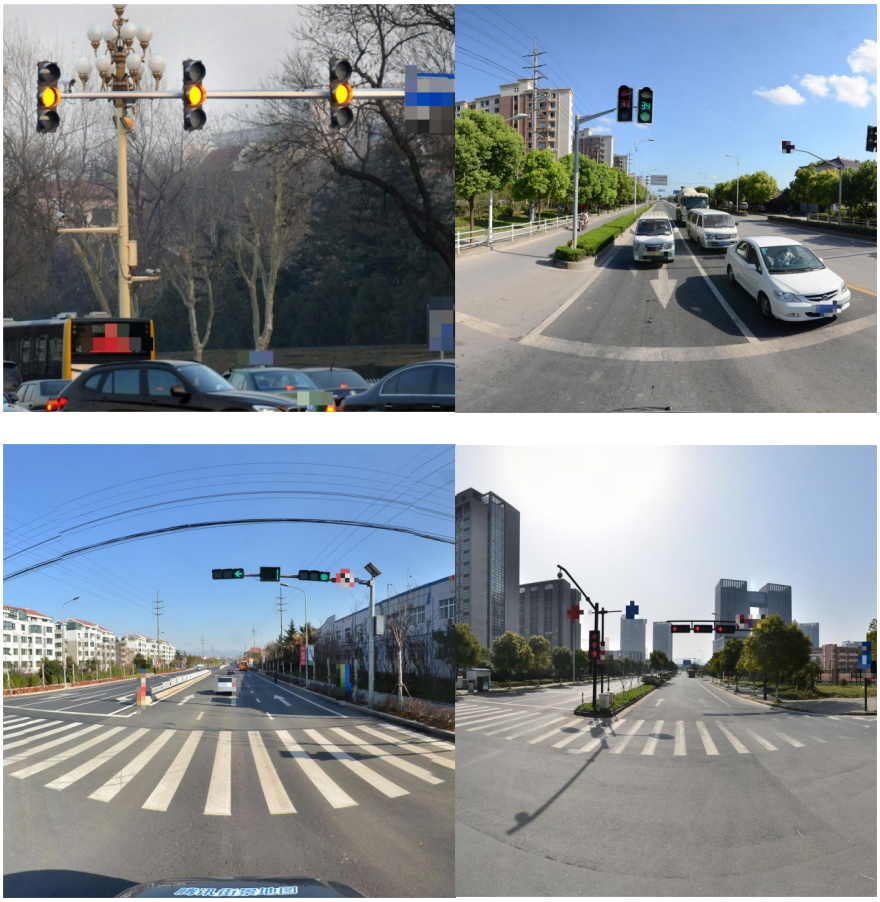

Figure 5. Sample traffic signal datasets.

3.2. Experimental Environment and Evaluation Index

3.2.1. Experimental Environment and Parameter Configuration

The experimental environment utilized the Window 10 operating system, with the deep learning framework being Pytorch, version 1.12.1. The software and hardware platform device parameters are detailed in Table 2, and and calculations were performed using the NVIDIA GeForce GTX 1650 graphics card. CUDA version is 11.4.0, and Python language environment version is 3.9.2.

Table 2. Platform configuration parameters.

Configuration name	Version parameter
Operating system	Window 10
Deep learning framework	Pytorch 1.12.1
GPU	NVIDIA GeForce GTX 1650
CUDA	11.4.0
Programming language	Python 3.9.2

The model's parameter settings are presented in Table 3. The total number of iterations is 400, with a batch size set to 8. The initial learning rate of the model was 0.01.

 doi:10.20944/preprints202407.0228.v1

Table 3. Model parameter settings.

epochs	batch_size	Initial learning rate
400	8	0.01

3.2.2. Evaluation Index

Precision (P), Recall (R), mean average precision (mAP) and Frames Per Second (FPS) were used in this study to evaluate the detection capability of the model [20]. The calculation formula of P and R is shown in formula (3) and (4) :

$$\Pr ecision = \frac{TP}{TP + FP} \times 100\% \tag{3}$$

$$\mathrm{Re}call = \frac{TP}{TP + FN} \times 100\% \tag{4}$$

TP represents the number of correctly detected traffic lights, *FP* represents the number of incorrectly detected traffic lights,, and *FN* represents the number of missed detection. *AP*(average precision) can be regarded as the area under a specific P-R curve, and *mAP* is the average *AP*across all categories. A larger mAP value indicates better detection effectiveness and identification accuracy of the algorithm. The calculation formulas for AP and mAP are shown in (5) and (6) :

$$AP = \int_0^1 P(R)dR \tag{5}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{6}$$

Where N is the total number of classes. FPS measures the number of image frames transmitted per second. A higher value indicates a faster detection speed of the network. In practical applications, an FPS of 25 is required to achieve real-time detection.

3.3. Analysis of Experimental Results

3.3.1. Ablation Experiment

To verify the effectiveness of the proposed model, ablation experiments were conducted in this study using the original YOLOv5 as the base network. The experimental results for different models on the test set are presented in Table 4. Experiment 1 involved training the traffic light dataset with the original YOLOv5, yielding precision (P) of 93.6%, recall ® of 92.2%, mean Average Precision (mAP) of 93.1%, and Frames Per Second (FPS) of 53. In experiment 2, Mosaic-9 was applied to augment the data from experiment 1, resulting in P of 96.6%, R of 97.3%, mAP of 97.2%, and FPS of 65. Subsequently, experiment 3 introduced the SE attention mechanism based on experiment 2 to enhance detection accuracy, achieving P of 97.6%, R of 98.2%, mAP of 98.1%, and FPS of 67. Finally, experiment 4 utilized the EIoU loss function from experiment 3 to mitigate missed and false detections, leading to P of 99.5%, R of 98.9%, mAP of 99.4%, and FPS of 74.

Table 4. Ablation experiment.

Experiment	Model	Precision(P)/ %	Recall(R)/ %	mAP/%	FPS/frame*s^{-1}
Experiment 1	YOLOv5	93.6	92.2	93.1	53
Experiment 2	YOLOv5+Mosaic-9	96.6	97.3	97.2	65

 doi:10.20944/preprints202407.0228.v1

Experiment 3	YOLOv5+Mosaic-9+SE	97.6	98.2	98.1	67
Experiment 4	YOLOv5+Mosaic-9+SE+EIoU	99.5	98.9	99.4	74

In summary, the enhanced algorithm proposed in this paper achieves a detection accuracy of 99.4%, surpassing the original YOLOv5 algorithm by 6.3%. Additionally, the detection speed reaches 74 f/s, meeting the real-time requirements.

3.3.2. Contrast Experiment

To further validate the efficiency of the proposed algorithm, we compared the improved algorithm with Faster R-CNN, YOLOv3, YOLOv4, SSD, YOLOv6, and YOLOv7 algorithms, and presented the experimental results in Table 5.

Table 5. Comparative experimental results.

Models	mAP/%	FPS/frame*s^{-1}	Params/MB
Faster R-CNN	91.5	53	74.4
YOLOv3	92.6	58	68.4
YOLOv4	92.1	67	50
SSD	93.5	75	43.5
YOLOv6	98.2	73	41.3
YOLOv7	99.3	73	41.5
Our improved algorithm	99.4	74	42.3

Table 5 shows that our improved algorithm named YOLOv5-MSE exhibits significantly faster speed and a 7.9% increase in mAP compared to Faster R-CNN. It also demonstrates improved mAP values compared to YOLOv3 and YOLOv4. While SSD achieves a detection speed of 75 f/s, its accuracy is lower than the proposed algorithm. When compared to YOLOv6 and YOLOv7, the overall performance gap is minimal. In summary, the enhanced model excels in mAP and parameter quantity, achieving 99.4% and 42.3MB respectively. With a detection speed of 74 f/s, it outperforms other models and is well-suited for deployment on embedded devices to meet real-time detection requirements, making it particularly suitable for traffic light recognition.

3.3.3. Comparison of Experimental Results

To visually demonstrate the recognition performance of the enhanced traffic light model, this study conducted a comparative experiment to detect traffic lights under various scenarios. The left image depicts the original YOLOv5 model, while the right image shows the improved model. The test results are presented in Figure 6, displaying detection outcomes for four scenarios: normal traffic lights, strong light, long distance, and complex background. In Figure 6(a), normal traffic lights are detected with improved accuracy compared to the original YOLOv5 model. In Fig. 6(b), the left image represents the result detected by the original YOLOV5 model under strong light, whereas the right image demonstrates a significant enhancement in detection accuracy. In Fig. 6©, the model accurately identifies and outputs high accuracy over long distances. In Figure 6(d), the proposed algorithm accurately detects the target under complex background, with improved accuracy.

 doi:10.20944/preprints202407.0228.v1

(**a**) Traffic lights with normal background

(**b**) Traffic lights with strong light background

(**c**) Traffic lights in the distance

 doi:10.20944/preprints202407.0228.v1

(**d**) Traffic lights with complex background

Figure 6. Comparison of the model detection results across various traffic scenarios.

4. Conclusions

An improved YOLOv5 traffic signal detection algorithm is proposed to address issues of missing detection, false detection, low accuracy, and excessive model parameters in traffic signal detection. This method enhances the training set data using the Mosaic-9 method, improving network generalization for better adaptation to real-world scenarios. Additionally, the SE attention mechanism is integrated to enhance detection effectiveness, and the EIoU_Loss is introduced to replace the original loss function, addressing missing and false detections while ensuring accuracy. Experimental results show that the improved algorithm achieves an mAP value of 99.4%, 6.3% higher than the original YOLOv5, with a detection speed of 74 f/s, balancing real-time performance and accuracy. Future work will focus on identifying algorithms that are more suitable for detecting traffic lights in complex urban scenes.

Author Contributions: Data curation, T.H. and Y.C.Z; Methodology, J.S.; Software, M.C.; Validation, T.H.; Writing—original draft, T.H.; Writing—review & editing, J.S. All authors have read and agreed to the published version of the manuscript..

Funding: This research was funded by Postgraduate Research & Practice Innovation Program of Jiangsu Province (grant number: SJCX24_0384) and The college student innovation and enterpreneurship training program of Jiangsu Province (grant number: 202410298057Z) .

Data Availability Statement: The data that support the findings of this study are available from the author Jun Song (songjun@njfu.edu.cn) upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chen Yan, Li chungui, Hu bo. An improved feature point extraction algorithm for field navigation[J]. Journal of Guangxi University of Technology, 2018, 29(03): 71-76.
2. Omachi, M (Omachi, Masako) , Omachi, S (Omachi, Shinichiro). Traffic Light Detection with Color and Edge Information[C], IEEE International Conference on Computer Science and Information Technology, 2009: 284-287.
3. Zhu Yongzhen, Meng Qinghu, Pu Jiexin. Automatic traffic light recognition based on HSV color space and shape features[J]. Television technology, 2015, 39(05): 150-154.
4. Zhu Zonghong, Li Chungui, Li wei, et al. The defect detection algorithm of vehicle injector seat based on Faster R-CNN model is improved[J]. Journal of Guangxi University of Technology, 2020, 31(01): 1-10.
5. Tian, Y (Tian, Yan), Gelernter, J (Gelernter, Judith) , Wang, X (Wang, Xun), et al. Lane marking detection via deep convolutional neural network[J]. Neurocomputing, 2018, 280: 46-55.
6. Redmon, J (Redmon, Joseph), Divvala, S (Divvala, Santosh), Girshick, R (Girshick, Ross), et al. You only look once: unified, real-time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2016: 779-788.

 doi:10.20944/preprints202407.0228.v1

7. Pan Weiguo, Chen Yinghao, Liu Bo, Shi Hongying. Traffic light detection and recognition based on Faster-RCNN[J]. Sensors and Micro-systems, 2019, 38(09): 147-149+160.
8. Li Hao. Research on traffic signal detection algorithm based on deep learning in complex environment[D]. Zhengzhou: Zhengzhou University, 2018.
9. Qian Hongyi, Wang Lihua, Mou Honglei. Fast detection and identification of traffic lights based on deep learning[J]. Computer Science, 2019, 46(12): 272-278.
10. Ju Moran, Luo Haibo, Wang Zhongbo, et al. Improved YOLOv3 algorithm and its application in small target detection[J]. Acta Optica, 2019, 39(7):253-260
11. Yan S J, Liu X B, Qian W, et al. An end-to-end traffic light detection algorithm based on deep learning[C]//2021 International Conference on Security, Pattern Analysis, and Cybernetics(SPAC), 2021: 370-373
12. Redmon J, Dvvala S, Girshick R, et al. You only look once: unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition. [S.I.]: IEEE, 2016
13. Chenjun Z, Xiaobing H, Hongchao N. Research on vehicle target detection based on improved YOLOv5. Journal of sichuan university(natural science edition) 2022, 59(05), 79-87
14. HE Kai-ming, ZHANG Xiang-yu, REN Shao-qing, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916
15. LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 936-944
16. LIU Shu, QI Lu, QIN HaiFang, et al. Path aggregation network for instance segmentation[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8759-8768.
17. Wang Pengei, Huang Hanming, Wang Mengqi. Complex road target detection algorithm based on improved YOLOv5[J]. Computer Engineering and Applications, 2022, 58(17): 81-92.
18. HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C] //Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7132-7141.
19. Cira, C.-I.; Díaz-Álvarez, A.; Serradilla, F., e.t.al. Convolutional Neural Networks Adapted for Regression Tasks: Predicting the Orientation of Straight Arrows on Marked Road Pavement Using Deep Learning and Rectified Orthophotography. Electronics 2023, 12, 3980.
20. Guo S, Li L, Guo T, Cao Y, Li Y. Research on Mask-Wearing Detection Algorithm Based on Improved YOLOv5. Sensors (Basel). 2022 Jun 29;22(13):4933.