Research on The License Plate Detection Algorithm Based on YOLOv5

Guanlin Du*, **Zhipeng Lin***, **Jingling Huang***, **Haidi Tang***, **Zhenxiang Xu***,

Xiamen University, Xiamen 361005, China
{23020231154179, 23020231154208, 23020231154139, 23020231154225, 23020231154240}@stu.xmu.edu.cn

Abstract

With the advancement of technology and the development of smart cities, computer vision has found wide application in various fields, including traffic management. License plate detection is one of the key technologies in smart transportation systems. License plate detection can be applied in assisting traffic accident processing and offers advantages such as safety, efficiency, and convenience.

Due to the need for real-time performance in most traffic scenarios, this dissertation selects the YOLOv5 algorithm among various real-time algorithms. Based on the YOLOv5 model, this dissertation proposes improvements specifically for the task of small object license plate detection to meet the real-time and accuracy requirements in traffic scenarios.

Firstly, to address the issue of poor detection performance for small objects like license plates, this dissertation introduces a small object detection layer. Compared to the original feature layers used for detecting small objects, this new layer has higher resolution and provides better detection performance for small objects. Secondly, the loss function is improved by using the SIoU(Gevorgyan 2022) loss function, which enhances accuracy.

Introduction

With the advancement of technology and the development of smart cities, intelligent transportation, including license plate detection, has made rapid progress and found wide applications in modern intelligent traffic systems. In China, with its large population, the number of motor vehicles has reached 408 million. During holidays and festivals, traffic congestion is a common phenomenon on highways in both small and large cities. The increasing number of vehicles has brought significant pressure to urban traffic. Therefore, the development and optimization of license plate detection technology have important implications for smart transportation systems and traffic management.

License plate detection can be used for traffic violation monitoring, traffic accident investigation, and vehicle tracking, helping law enforcement agencies and traffic management departments effectively monitor road traffic conditions, promptly identify and address traffic violations, and improve traffic safety(Ren et al. 2015). As an important component of intelligent transportation systems, license plate detection enables functionalities such as automatic vehicle recognition, automated tolling, and intelligent navigation. By applying license plate detection technology, traffic efficiency can be improved, traffic congestion reduced, and the urban transportation system enhanced.

In criminal investigations, license plate detection can be used to track and identify vehicles involved in crimes, providing crucial leads and evidence. This is of great significance in combating crime and maintaining social security. License plate detection can also be applied to parking lot management systems, enabling automatic entry and exit of vehicles and intelligent allocation of parking spaces. This improves the efficiency and convenience of parking management while reducing the waste of parking resources.

Finally, license plate detection is an important research direction in the fields of computer vision and deep learning. Studying license plate detection algorithms and methods can drive the development and innovation of technologies such as object detection, image recognition, and deep learning.

Related Work

Foreign research on license plate recognition began earlier. Kanayama (Cagnoni et al. 2020)et al. used the color difference between license plates and the background environment, performed edge detection using Sobel operator. As the detection effect is affected by the surrounding environment, vertical edge detection was performed and candidate frames were selected based on the aspect ratio threshold. However, this method is sensitive to lines, resulting in reduced accuracy in complex background scenes. Lee et al. used the color feature of license plates and employed the HSL color space for license plate positioning. This method improved the detection effect for distorted license plates, but was susceptible to the impact of license plate illumination and noise. Matas (Huang et al. 2022)et al. used character detection to determine the position of the license plate, which had good detection effect for distorted license plates. However, it was time-consuming and easily interfered with by background characters. Ashtari et al. utilized image segmentation algorithms to detect license plates. By analyzing pixels in the image and setting templates, the corresponding license plate frames were drawn based on color. Yuan et al. divided li-

*These authors contributed equally.

cense plate detection into three processes: downsampling images, converting them to grayscale, computing their edge features, and selecting several candidate frames before classifying them using a classifier to determine the license plate area.

In the 21st century, artificial intelligence and deep learning technologies have become more powerful. Neural networks have also been applied to license plate detection systems. Silva et al. conducted license plate recognition experiments in Brazil using the Convolutional Neural Network (CNN)(Girshick 2015) algorithm based on YOLO, sequentially detecting vehicles, license plates, and characters. This method's accuracy far exceeded that of traditional license plate detection methods. However, this algorithm only detects license plates from the front view, narrowing its application range. YOLO(Redmon et al. 2016), or You Only Look Once, is a novel object detection method proposed by Joseph Redmon et al. in 2016. The main innovation of YOLO algorithm is that it solves the object detection task as an end-to-end regression problem, directly predicting the location and category information of objects from images, avoiding the complex processes of generating candidate boxes and extracting features in traditional object detection methods. YOLO can quickly recognize objects in images or videos and has the fastest recognition speed among methods with the same recognition category range and accuracy. This makes YOLO suitable for real-time license plate detection.(Fang, Wang, and Ren 2019) Although YOLO has high practical value in license plate detection and other fields, it also has unique limitations. First, YOLO's detection effect on small objects is poor. Second, because YOLO requires a large amount of image data sets for calculation during training, compared with other traditional object detection methods, YOLO relies on a large amount of annotated data sets to obtain high-precision models. Furthermore, training with large-scale data sets can also make YOLO have good generalization ability and detect various license plate shapes, colors, and styles.(Terven and Cordova-Esparza 2023) YOLOv1 is the first version of YOLO, which divides the input image into a fixed number of grid units and predicts bounding boxes and categories on each unit using the convolutional neural network. YOLOv2(Redmon and Farhadi 2017) is the second version of YOLO, which uses a series of improvements, including a deeper convolutional neural network (Darknet-19), anchor boxes to predict bounding boxes of different scales and aspect ratios, and multi-scale prediction and feature fusion(Sang et al. 2018). YOLOv3(Redmon and Farhadi 2018) is the third version of YOLO, which further improves YOLOv2 by introducing a deeper Darknet-53 convolutional network as a feature extractor and using different scale feature layers for object detection.(Zhao and Li 2020),(Xu and Wu 2020),(Gong, Ma, and Ouyang 2020) YOLOv4(Bochkovskiy, Wang, and Liao 2020) is the fourth version of YOLO, which uses the CSP-Darknet53 network, spatial attention module (SAM) to extract features, and PANet Path Aggregation module (PPM) to realize multi-scale feature fusion. YOLOv5 (Zhang et al. 2022)is the fifth version of YOLO, which uses a lightweight CSPDarknet53 to reduce parameter volume and computational complexity. It also introduces an automatic scaling function to adaptively adjust network scale based on the input image resolution, making YOLOv5 more flexible in processing different resolution images.(Qi et al. 2022),(Wu, Wang, and Liu 2021),(Wu et al. 2021)Additionally, its introduction of the Generalized Intersection over Union (GIoU) loss function can better measure the overlap of target boxes, thereby improving the detection accuracy.

Proposed Solution

The proposed license plate detection algorithm for small targets in this paper mainly improves the algorithm from two different directions: 1)The method of adding a small target detection layer and modifying the network structure. 2)Using the SIoU loss function.

Adding a small target detection layer. Improving the performance of small object license plate detection: In the dataset used in this study, the scale of faces and license plates is extremely small, some are hidden, and the low resolution of the target in the image makes license plate detection more challenging. This study primarily focuses on the YOLOv5 algorithm. However, the downsampling factor of YOLOv5 is relatively large, making it difficult for deeper feature maps to capture the characteristics of small objects. Therefore, this study proposes a method to add a small object detection layer to the backbone network of YOLOv5, allowing the network to pay more attention to the detection of small objects and improve the detection performance of small license plates.Although YOLOv5s is a multi-scale object detection algorithm, its detection performance for datasets with many small targets is still not ideal. This is because in multi-scale object detection, the input images are downscaled through adaptive image resizing. As the convolutional layers deepen, the resolution of feature maps decreases, leading to the loss of low-level feature information such as color, contours, and shapes, further deteriorating the detection performance for small targets. On the other hand, when the convolutional layers are shallow, the resolution of feature maps is higher, but the high-level semantic features may be difficult to obtain due to the interference from low-level information.

The initial model of YOLOv5 only performs feature fusion on three feature layers, namely P3, P4, and P5. For the wide-angle dataset used in this paper, the captured images are distorted, and most of them are taken from a considerable distance from the targets, resulting in very small target pixels in most images. Therefore, higher-resolution feature maps are needed, and using low-level features of the backbone network for target detection can improve the performance of small target detection.

The purpose of adding a small target detection layer is to include P2, which refers to the feature map obtained from the second conv layer in the YAML file, in the feature fusion process. At the same time, anchor parameters specific to small targets are added. The P2 layer is located at a lower feature level, where it has strong position information but weak semantic feature information, making it often used for small target detection. The newly added 160×160 detection feature map is concatenated and fused with the second-layer feature map of the backbone network, resulting in a larger feature map that can detect targets larger than 4×4, thus optimizing the performance of small target detection. Some parts of the network structure after adding the small target detection layer are shown in Figure 1.

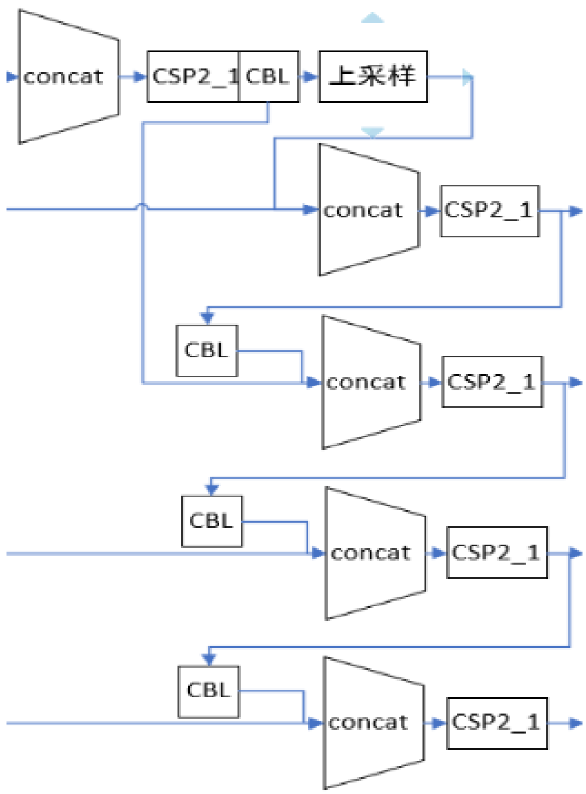

Figure 1: Network structure diagram

Improvement of loss function. This paper modifies the loss function of YOLOv5 and incorporates the SIOU loss function. The SIOU loss function deeply considers the vector angle between the ground truth box and the predicted detection box, and redefines the related loss functions. The SIOU loss includes four components: 1) Angle cost 2) Distance cost 3) Shape cost 4) IoU cost.SIOU's four component losses are defined in Gevorgyan Z.(Gevorgyan 2022)

Using SIoU can improve the accuracy of inference. This paper combines SIoU with the addition of a small target detection layer, resulting in significantly improved accuracy of YOLOv5 on wide-angle datasets.

Experiments

The experimental section provides detailed experimental validation of the proposed improved YOLOv5 object detection method. First, a comparative experiment is conducted between YOLOv3 and YOLOv5. Then, ablation experiments are performed on the license plate detection improvement method, as well as comparison experiments on license plate detection using different datasets. Finally, the experimental results are analyzed and summarized separately.

Dateset. In this paper, the wide-angle dataset used is the PP4AV open-source dataset. PP4AV is the first public dataset that annotates faces and license plates in driving scenes. This dataset utilizes wide-angle images from the WoodScape dataset(Yogamani et al. 2019) and selects 244 images from the camera's perspective as the wide-angle dataset, annotated in YOLO format.Due to the limited number of open-source wide-angle datasets, this paper also used the CCPD license plate dataset. The original format of the CCPD license plate dataset is in VOC format, so it needs to be converted to YOLO format annotations in order to train the model.

Experimental environment. We use Python3.7, Pytorch 1.8.0+cu111, Cuda 11.1 and Cudnn 9.0.CPU is Intel(R) Core(TM) i5-9300H CPU @ 2.40GHz 2.40 GHz.Graphics card is NVIDIA GeForce GTX 1660 Ti 6GB.Memory is 16GB.

Comparative experiment. This paper compares the YOLOv3 and YOLOv5 versions of the YOLO series algorithm. The evaluation metrics for the experimental results mainly consist of precision (P), recall (R), and mean average precision (mAP@0.5, IoU=0.5). Precision (P) refers to the proportion of correctly identified objects among all recognized objects. Recall (R) refers to the probability of correctly identifying all positive samples. Mean Average Precision (mAP@0.5) represents the average precision when the confidence threshold is set to 0.5.

After training on the wide-angle dataset, the mAP@0.5 for human face targets using YOLOv3 was 30.1%, and for license plates, it was 33.3%. The average mAP@0.5 for both was 31.7%. The training results for YOLOv5 are shown in Table 1. The mAP@0.5 for YOLOv3 was 31.7%, while for YOLOv5, it was 39.1%. In comparison, YOLOv5 showed an improvement of 7.4% in accuracy compared to YOLOv3. This paper also tested on the CCPD license plate dataset.

class	P	R	mAP@0.5
all	0.459	0.439	0.391
face	0.437	0.205	0.196
license	0.481	0.673	0.586

Table 1: YOLOv5 result

After converting the VOC dataset into the YOLO dataset format and configuring the relevant configuration files, the experimental results were obtained as shown in Table 2. The accuracy of YOLOv3 reached 99.8%, while that of YOLOv5 was also 99.6%. Therefore, for ordinary and easily recognizable license plate datasets, the experimental results show little difference between the two algorithms. Compared to YOLOv3, YOLOv5 has made improvements in many aspects. The main improvements can be divided into the following four points:

- In terms of network structure, YOLOv3 uses a three-layer convolutional neural network, while YOLOv5, af-

class	P	R	mAP@0.5
YOLOv3-license	1.000	0.776	0.998
YOLOv5-license	1.000	1.000	0.996

Table 2: YOLOv3/v5 result on CCPD

ter being improved, uses the CSPDarknet53 network structure.

- In terms of accuracy, YOLOv5 has added many data augmentation methods and used more convolutional layers, resulting in higher accuracy compared to YOLOv3.
- In terms of speed, YOLOv5 has a more efficient algorithm and a smaller model, making it faster than YOLOv3.
- In terms of scalability, YOLOv5 uses a more flexible network architecture, making it easier to expand compared to YOLOv3.

In summary, the new generation of YOLO after the improvement of YOLOv5 has higher accuracy and speed than the previous generation.

Ablation experiment. This article focuses on the problem of small targets in license plate detection and designs a small target detection module. It also improves the loss function by replacing the original loss function with SIoU to enhance detection accuracy and training speed. In order to verify the effectiveness of these two improvements on license plate detection performance, a ablation experiment was designed. The evaluation metrics for the experimental results consist of precision (P), recall (R), and mean Average Precision (mAP@0.5) with an IoU threshold of 0.5.

The experimental results of comparing the YOLOv5 prototype with YOLOv5 combined with small object detection layer(Sodl), YOLOv5 combined with SIoU loss function, and YOLOv5 combined with small object detection layer and SIoU loss function are shown in Table 3.Its visualization result is shown in Figure 2.

Sodl	SIoU	P	R	mAP@0.5
N	N	0.459	0.439	0.391
Y	N	0.552	0.494	0.474
N	Y	0.536	0.386	0.403
Y	Y	0.590	0.539	0.501

Table 3: Ablation experiment result

Through this ablation experiment, it can be observed that adding a small object detection layer alone greatly improves the accuracy of object detection, with an increase of 8.3% in mAP@0.5. Similarly, combining SIoU also improves the accuracy, but to a lesser extent, with a 1.2% increase in mAP@0.5. Moreover, when combining SIoU with the addition of the small object detection layer, compared to adding either of them individually, there is a further improvement in mAP@0.5 by 11%. Additionally, both precision and recall are also improved significantly. The detection results of YOLOv5 and YOLOv5 with the small object detection layer

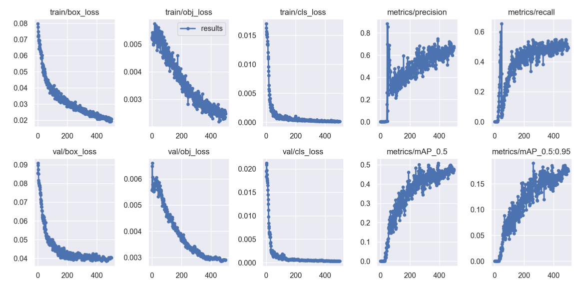

Figure 2: Improved YOLOv5 results

and SIoU integration are shown in Figure 2 and Figure 3. In conclusion, both of these improvement methods contribute to enhancing the overall performance of license plate detection, and the combination of them yields even greater improvements. Furthermore, the larger the scale of the dataset,

Figure 3: YOLOv5 test results

Figure 4: Improved detection result

including different scenarios and variations, the better the model's generalization ability in different situations. The dataset used in this study is relatively small, resulting in

Figure 5: Example of detecting errors

weaker generalization capability of the trained model. When there are changes in the angle, position, or distance of vehicles, the appearance of license plates may alter, leading to inaccurate object detection. Moreover, license plates are typically located within complex backgrounds, which may contain various textures, signs, and texts. This complexity makes it challenging for the object detection algorithm to accurately separate and locate the license plate boundaries. This is illustrated in Figure 4. To address these issues, increasing the training data and making various algorithmic improvements can be explored to assess their impact on performance enhancement, aiming to achieve a more advanced license plate detection algorithm.

Conclusion

The main content of this article is as follows:

- Introducing the current demand for license plate detection and selecting a YOLO-based license plate detection algorithm to address the need for strong real-time performance. This algorithm is a one-stage algorithm that offers advantages in terms of speed and real-time capability, making it highly applicable in scenarios requiring real-time detection.
- Providing a detailed introduction to the principles of YOLOv3 and YOLOv5 algorithms and comparing the two.
- Introducing the improvement method of adding a small object detection layer and the enhanced SIoU loss function used for the dataset. Explaining their principles and highlighting the differences compared to the previous versions.
- Conducting experiments on the proposed methods and comparing the performance of different versions of YOLO algorithms. Performing ablation experiments on the improved license plate detection algorithm, analyzing the experimental results, and comparing the performance improvement before and after the improvements.

There are still several shortcomings in this experiment, and further improvements and extensions can be made in the future:

- Due to the limited availability of open-source wide-angle license plate datasets, the dataset used in this study is relatively small, which may result in insufficient generalization ability of the algorithm when applied to new data. In the future, it would be beneficial to collect wide-angle images captured with fish-eye lenses and create a custom dataset for further model training.
- Due to time constraints, this study only explored limited improvement methods and did not consider other network architecture modifications. In the future, it would be worthwhile to explore alternative improvement methods, such as trying new network architectures and attention modules. Conducting experiments on different datasets to observe the results can help determine whether further enhancements in license plate detection performance can be achieved.
- After license plate detection, it is possible to further expand the functionality to include license plate recognition. This can be done by integrating character recognition to identify the characters on the license plate and recognize the license plate number. For example, a commonly used approach is to combine Paddle OCR and LPRnet for character recognition and annotate the recognized license plate number on the image.

References

Bochkovskiy, A.; Wang, C.-Y.; and Liao, H.-Y. M. 2020. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.

Cagnoni, S.; Al-Sahaf, H.; Sun, Y.; Xue, B.; Zhang, M.; et al. 2020. Special issue on evolutionary computer vision, image processing and pattern recognition. *Applied Soft Computing*, 97: 106675–106677.

Fang, W.; Wang, L.; and Ren, P. 2019. Tinier-YOLO: A real-time object detection method for constrained environments. *IEEE Access*, 8: 1935–1944.

Gevorgyan, Z. 2022. SIoU loss: More powerful learning for bounding box regression. *arXiv preprint arXiv:2205.12740*.

Girshick, R. 2015. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 1440–1448.

Gong, X.; Ma, L.; and Ouyang, H. 2020. An improved method of Tiny YOLOV3. In *IOP Conference Series: Earth and Environmental Science*, volume 440, 052025. IOP Publishing.

Huang, W.; Xu, X.; Hu, M.; and Huang, W. 2022. A license plate recognition data to estimate and visualise the restriction policy for diesel vehicles on urban air quality: A case study of Shenzhen. *Journal of Cleaner Production*, 338: 130401.

Qi, J.; Liu, X.; Liu, K.; Xu, F.; Guo, H.; Tian, X.; Li, M.; Bao, Z.; and Li, Y. 2022. An improved YOLOv5 model based on visual attention mechanism: Application to recognition of tomato virus disease. *Computers and electronics in agriculture*, 194: 106780.

Redmon, J.; Divvala, S.; Girshick, R.; and Farhadi, A. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788.

Redmon, J.; and Farhadi, A. 2017. YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7263–7271.

Redmon, J.; and Farhadi, A. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Ren, S.; He, K.; Girshick, R.; and Sun, J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.

Sang, J.; Wu, Z.; Guo, P.; Hu, H.; Xiang, H.; Zhang, Q.; and Cai, B. 2018. An improved YOLOv2 for vehicle detection. *Sensors*, 18(12): 4272.

Terven, J.; and Cordova-Esparza, D. 2023. A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond. *arXiv preprint arXiv:2304.00501*.

Wu, T.-H.; Wang, T.-W.; and Liu, Y.-Q. 2021. Real-time vehicle and distance detection based on improved yolo v5 network. In *2021 3rd World Symposium on Artificial Intelligence (WSAI)*, 24–28. IEEE.

Wu, W.; Liu, H.; Li, L.; Long, Y.; Wang, X.; Wang, Z.; Li, J.; and Chang, Y. 2021. Application of local fully Convolutional Neural Network combined with YOLO v5 algorithm in small target detection of remote sensing image. *PloS one*, 16(10): e0259283.

Xu, D.; and Wu, Y. 2020. Improved YOLO-V3 with DenseNet for multi-scale remote sensing target detection. *Sensors*, 20(15): 4276.

Yogamani, S.; Hughes, C.; Horgan, J.; Sistu, G.; Varley, P.; O'Dea, D.; Uricár, M.; Milz, S.; Simon, M.; Amende, K.; et al. 2019. Woodscape: A multi-task, multi-camera fisheye dataset for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9308–9318.

Zhang, Y.; Guo, Z.; Wu, J.; Tian, Y.; Tang, H.; and Guo, X. 2022. Real-time vehicle detection based on improved yolo v5. *Sustainability*, 14(19): 12274.

Zhao, L.; and Li, S. 2020. Object detection algorithm based on improved YOLOv3. *Electronics*, 9(3): 537.