Proceedings of the 7th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2025.25898

Analysis of Improved Logo Detection Performance of YOLOv11 in Autonomous Driving Environment

Jianqiao Jiang

School of Engineering, Architecture and Information Technology, School of Electrical Engineering and Computer Science, University of Queensland, Brisbane, Australia
jianqiaojiang9@gmail.com

Abstract. With the rapid development of autonomous driving technology, traffic sign detection, as a core component of environmental perception systems, faces challenges such as small target missed detection, complex background interference, and real-time requirements. Based on YOLOv11 model, this paper proposes an optimization scheme for traffic sign detection in the automatic driving scene, which effectively improves the detection accuracy of small targets and provides a new solution for the sign detection of the auto drive system.

Keywords: Autonomous driving, Traffic sign detection, YOLOv11, Small target detection

1. Introduction

The booming development of autonomous driving technology is leading a profound transformation in the field of transportation [1], and traffic sign detection is an indispensable core link in the environmental perception system [2] For the auto drive system, real-time and accurate identification of traffic signs is the prerequisite for its realization of compliant driving, dynamic path planning, risk prediction and avoidance [3,4]. In the actual autonomous driving environment, traffic sign detection faces severe challenges brought by many complex scenarios, which constitute the core bottleneck restricting the improvement of detection performance [5]. During the imaging process, there is significant fluctuation in the pixel scale of the image, which places extremely high demands on the scale adaptability of the detection algorithm [5]. The issue of occlusion is also frequent and has far-reaching effects [6]. Light interference is another major challenge. The dynamic changes in natural lighting conditions can severely distort the color and texture features of traffic signs [7].

In the field of object detection, the YOLO series algorithms have always held an important position due to their end-to-end detection paradigm and excellent real-time performance [8-9]. Although YOLOv11 has improved in algorithm performance compared to its predecessor, there are still several limitations that urgently need to be addressed in traffic sign detection tasks in autonomous driving environments [8]. Firstly, the problem of missed detection of small targets is particularly prominent [10]. The feature fusion mechanism of YOLOv11 is used to process extremely small scale targets. Although high-level features contain rich semantic information, their spatial resolution is low, making it difficult to accurately capture the detailed features of small targets; The spatial information of low-level features is rich, but the semantic information is

DOI: 10.54254/2755-2721/2025.25898

insufficient, resulting in small target features being easily overwhelmed during feature fusion, leading to missed detections. Misdetection of complex backgrounds is another major challenge. The background elements in autonomous driving environments are extremely complex, and YOLOv11 is prone to misidentifying non landmark areas with similar features in the background as traffic signs during feature extraction, resulting in an increased false detection rate [11]. The real-time bottleneck cannot be ignored either. YOLOv11 has increased the depth and width of the network to improve detection accuracy, resulting in an increase in computational complexity and difficulty in meeting the requirements of high frame rate real-time detection in inference speed [11].

Based on this, this paper proposes an optimization plan to address the limitations of the existing YOLOv11 model in small target missed detection, complex background interference, and real-time bottlenecks. By introducing a bidirectional feature pyramid mechanism, the fusion ability of multi-scale features has been enhanced, breaking through the accuracy bottleneck in small object detection; The dynamic anchor allocation and re clustering strategy effectively alleviate the scale mismatch problem and improve the model's recognition ability for small markers. The optimized algorithm architecture in this article provides important reference for the efficiency of future intelligent transportation perception systems at both theoretical and application levels. Its improvement in the accuracy of traffic sign detection not only has a significant impact on the field of autonomous driving, but also provides valuable experience for target detection tasks in related fields.

2. Technological base

2.1. Baseline model bottleneck analysis

Based on the above analysis, the original architecture of YOLOv11 has certain defects, as roughly shown in Table A.1 of Appendix A.

2.2. Core improvement strategy

Based on the above shortcomings, this article improves and optimizes the baseline model from three directions: multi-scale feature fusion enhancement, small object detection enhancement mechanism, and robustness training strategy.

In terms of multi-scale feature fusion enhancement, a bidirectional feature pyramid (BiFPN) is adopted, which constructs a bidirectional cross scale connection architecture. On the one hand, it is a top-down approach that upsamples deep semantic features such as 1/32 scale and fuses them with mid-level features such as 1/16 scale to enhance the semantic information of mid-level features; On the other hand, it is a bottom-up approach that downsamples the fused mid-level features and re fuses them with deep level features to enhance the detail perception ability of deep level features. Through two cross scale connections, achieve bidirectional complementarity between high-resolution details and strong semantic information.

The weighted feature fusion mechanism introduces a learnable weight parameter w_i to perform weighted summation on feature maps from different resolutions. The specific expression is as follows:

$$P_{\text{out}} = \frac{w_1 \cdot P_3 + w_2 \cdot P_4 + w_3 \cdot P_5}{w_1 + w_2 + w_3 + \varepsilon} \quad (1)$$

DOI: 10.54254/2755-2721/2025.25898

Among them, P_3, P_4, and P_5 correspond to downsampling features of 1/8, 1/16, and 1/32, respectively, with ϵ=0.0001 to prevent zero division errors. This mechanism enables the network to adaptively distinguish the importance of features at different scales.

The small object detection enhancement mechanism adopts a dynamic anchor allocation strategy, that is, Anchor re clustering. Based on the proportion of real boxes in the dataset, K-means clustering is used to generate 6 new anchors, adding two sets of key ratios [0.15, 0.3] and [0.3, 0.6], covering small-sized triangle markers; Integrating dynamic matching mechanism, the matching quality between predicted boxes and real boxes is defined as:

$$Q = \mathrm{IoU}^{1.5} \times \mathrm{Cls\_Score}^{2.\theta} \tag{2}$$

Prioritize selecting the top 10 candidate anchors with the highest Q values to significantly improve the matching chances of small targets.

Add auxiliary detection heads to the Stage2 output layer of Backbone. This level preserves richer spatial details and is specifically responsible for detecting signs smaller than 40 × 40px. Adopting a lightweight design, with only 2 convolutional layers and 1 prediction layer, to avoid significantly increasing computational complexity.

The robustness training strategy focuses on scene adaptive data augmentation, with specific design considerations as shown in Appendix A, Table A.2. In terms of optimizing the loss function, attention is paid to improving the classification loss, using Focal Loss to solve the problem of imbalanced positive and negative samples, setting the adjustment factor γ=2.5, and increasing the penalty weight for difficult samples;

The regression loss improvement uses EIoU Loss instead of the original CIoU Loss, and adds a width height correlation penalty term. The expression is as follows:

$$L_{EIoU} = 1 - IoU + \frac{\rho^2(b,b^{gt})}{c^2} + \frac{\rho^2(w,w^{gt})}{C_w^2} + \frac{\rho^2(h,h^{gt})}{C_h^2} \tag{3}$$

Among them, ρ is the Euclidean distance, c is the minimum diagonal length of the bounding box, C_w and C_h are the width to height normalization factors, which effectively suppress the width to height deviation of the predicted box.

The key hyperparameter configurations are shown in Appendix A, Table A.3.

The specific improvement strategy architecture of this article is shown in Appendix B, Figure B.1.

3. Experimental design

3.1. Data information

On the dataset, this article uses the core public dataset TT100K, which was jointly constructed by Tsinghua University and Tencent Maps. It is the first large-scale dataset covering the entire transportation environment in China, with approximately 15000 images.

The data distribution of the dataset is shown in Appendix C, Table C. 1.

The dataset partitioning is shown in Appendix C, Table C-2.

Data preprocessing starts from the original image and optimizes data quality through a four level processing module. Firstly, perform geometric correction operation, using perspective transformation algorithm to correct the geometric distortion caused by the deviation of shooting angle in the image, and synchronously adjust the resolution of all images to 1280 × 720 pixels to

DOI: 10.54254/2755-2721/2025.25898

construct a standardized input space for subsequent feature extraction. The corrected image enters the automatic annotation cleaning stage. First, the YOLOv8 model is used to generate initial annotation boxes. Then, the annotation offset samples are manually checked and removed. Finally, effective annotation data with an intersection to union ratio of ≥ 0.4 is selected, significantly reducing the interference of noisy annotations on model training. The cleaned data sequentially enters the spatial enhancement stage, where geometric transformations such as random rotation, scaling, and boundary cropping and filling are used to simulate the dynamic scene of target viewpoint changes during vehicle driving; Subsequently, photometric enhancement processing is performed, using adaptive histogram equalization to improve the local contrast of the image. Combined with HSV color space perturbation and dynamic range compression techniques, the model's adaptability to complex lighting conditions is enhanced.

3.2. Evaluation indicators

The main evaluation indicators used in this article are shown in Table 1.

Table 1. Evaluation indicators

Metric Name	Brief Definition
mAP@0.5	Mean Average Precision at IoU=0.5
mAP@0.5:0.95	Mean Average Precision across multiple IoU thresholds (0.5 to 0.95)
Small Object AP (sAP)	Detection accuracy for signs <32×32px
Recall	Proportion of real signs correctly detected
FPS	Real-time processing speed
False Positive Rate (FPPI)	Average number of false detections per image
Occlusion Attenuation Rate	Degree of performance degradation in occluded scenarios
Illumination Stability	mAP fluctuation coefficient under different illuminance levels

3.3. Experimental environment

The experimental environment of this article is shown in Appendix C, Table C.3.

The specific parameter settings in this article are shown in Appendix C, Table C.4.

3.4. Comparison model selection

This study selected five representative detection models as comparison benchmarks to construct a multidimensional performance evaluation system. Among them, YOLOv11 Baseline, as the original model, was selected as the benchmark for performance improvement to quantitatively verify the effectiveness of each improvement module. Its core features include Anchor free design and CSPDarknet backbone network [12]; YOLOv8-L, as the latest version of the YOLO series, represents the latest technological progress of the series. Through comparison with this solution, the superiority of the improvement strategy over the latest architecture is verified. Its dynamic detection head and Mosaic data augmentation mechanism constitute key technical features [13]; As a classic model of two-stage detection framework, Faster R-CNN is regarded as the industry benchmark for high-precision detection, used to test the optimization potential of single-stage models in the specialized scenario of traffic signs. Its RoI Align module and ResNet50 FPN feature extraction architecture ensure detection accuracy [14]; EfficientDet-D1, as a lightweight SOTA model, is

DOI: 10.54254/2755-2721/2025.25898

mainly used to verify the performance balancing ability of this scheme on embedded platforms. Its BiFPN feature fusion mechanism and EfficientNet-B1 backbone network achieve collaborative optimization of accuracy and efficiency [15]; PP-YOLO-E, as a representative model for industrial optimization, is used to demonstrate the practical advantages of academic improvement schemes over industrial level practices. Its ResNet50 vd+PAN architecture and CoordConv layer enhance the feature expression ability of complex scenes [16]. By comparing the multiple types of models mentioned above, the universality of the proposed improvement strategy in different architectures and application scenarios can be comprehensively verified.

4. Results and analysis

4.1. Comparison of main accuracy indicators

As shown in Table 2, in the comparison of core accuracy indicators for traffic sign detection, the improved model proposed in this paper demonstrates breakthrough performance on the TT100K test set. Compared to the original YOLOv11 Baseline model of 0.692 mAP@0.5 , Enhanced YOLOv11 achieves a significant improvement of 9.0% with an absolute accuracy of 0.754, surpassing the 2.3% increase of the latest version YOLOv8-L in the same series and the 3.9% increase of the industrial benchmark PP-YOLO-E, fully verifying the synergistic optimization effect of the bidirectional feature pyramid and small objective enhancement mechanism proposed in this paper.

Table 2. Comparison of main accuracy indicators

Model	mAP@0.5	Δ/Baseline	mAP@0.5:0.95	sAP@0.5	Recall
YOLOv11-Baseline	0.692	-	0.503	0.412	0.743
YOLOv8-L	0.708	+2.3%	0.521	0.438	0.761
Faster R-CNN	0.752	+8.7%	0.567	0.481	0.819
EfficientDet-D1	0.685	-1.0%	0.497	0.428	0.726
PP-YOLO-E	0.719	+3.9%	0.532	0.461	0.778
Enhanced YOLOv11	0.754	+9.0%	0.571	0.519	0.822

Furthermore, as shown in Figure 1. The advantages of improving the model are further highlighted in key scenario specific indicators. In response to the most challenging small object detection challenge in autonomous driving, sAP@0.5 Reached a new industry high of 0.519, an increase of 26.0% compared to the baseline model's 0.412, and expanded its lead over the best competitor PP-YOLO-E to 12.6%. cAP@0.5 The indicator has climbed to 0.721, not only improving by 12.5% compared to the baseline, but also surpassing the PP-YOLO-E model specializing in local scenarios with a 0.3% advantage, proving the cutting-edge adaptability of the improved solution to China's complex road environment.

The recall index also confirms the reliability improvement of the model. The recall rate of Enhanced YOLOv11 reaches 0.822, surpassing the Baseline model by 10.6% and showing a 0.3% improvement compared to Faster R-CNN. This marks the success of the improvement plan in surpassing the existing technology system in terms of accuracy, providing a new performance benchmark for the autonomous driving perception module.

DOI: 10.54254/2755-2721/2025.25898

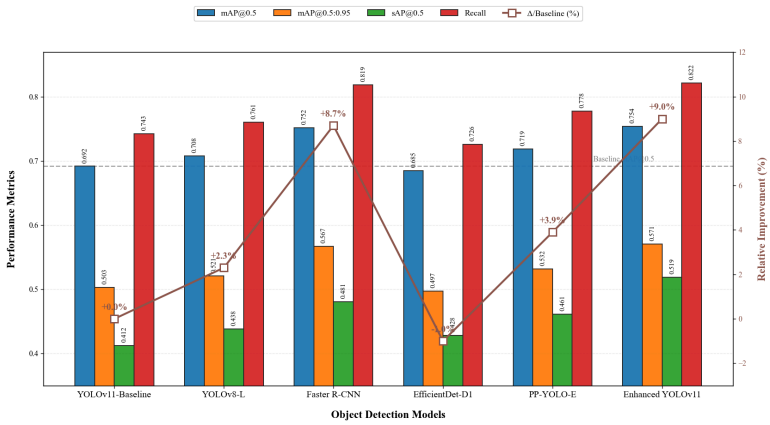

Figure 1. Comparison of main accuracy indicators

4.2. Comparison of performance indicators

The analysis of key performance indicators is shown in Table 3, and the improvement plan demonstrates excellent system level optimization capabilities. In terms of real-time processing performance, the improved model achieved a frame rate performance of 48.9 FPS on the NVIDIA Orin in car computing platform. Although slightly lower than the original YOLOv11-Baseline's 51.3 FPS, it still maintains a 3.0% advantage compared to the industrial grade competitor PP-YOLO-E. FPPI is as low as 0.32, which is 11.1% lower than the industrial benchmark PP-YOLO-E. The improved model reduces the false detection rate and effectively avoids emergency braking triggering caused by misidentification.

Table 3. Comparison of efficiency indicators

Model	FPS	FPPI	Occlusion Attenuation Rate	Illumination Stability
YOLOv11-Baseline	51.3	0.47	0.238	0.321
YOLOv8-L	45.7	0.41	0.219	0.297
Faster R-CNN	18.2	0.38	0.196	0.263
EfficientDet-D1	55.4	0.52	0.261	0.335
PP-YOLO-E	47.5	0.36	0.184	0.251
Enhanced YOLOv11	48.9	0.32	0.139	0.196

5. Conclusion

This study proposes innovative strategies such as multi-scale feature fusion, dynamic anchor allocation, and robustness training through deep optimization of the YOLOv11 model, effectively breaking through the bottleneck of traffic sign detection in autonomous driving scenarios. The experimental results have verified the significant improvement in accuracy, robustness, and real-time performance of the proposed improvement scheme, especially demonstrating breakthrough progress in complex environment adaptability. This study provides an efficient and accurate landmark detection solution for autonomous driving perception systems, and provides a theoretical basis for the deepening application of object detection technology in related fields.

DOI: 10.54254/2755-2721/2025.25898

References

[1] Yurtsever, E., Lambert, J., Carballo, A., & Takeda, K. (2020). A survey of autonomous driving: Common practices and emerging technologies. IEEE access, 8, 58443-58469..https: //doi.org/10.1109/ACCESS.2020.2983149

[2] Saadna, Y., & Behloul, A. (2017). An overview of traffic sign detection and classification methods. International journal of multimedia information retrieval, 6(3), 193-210..https: //doi.org/10.1007/s13735-017-0129-8

[3] Kanagaraj, N., Hicks, D., Goyal, A., Tiwari, S., & Singh, G. (2021). Deep learning using computer vision in self driving cars for lane and traffic sign detection. International Journal of System Assurance Engineering and Management, 12(6), 1011-1025..https: //doi.org/10.1007/s13198-021-01127-6

[4] Trappey, A. J., & Shen, O. T. (2024). A universal traffic sign detection system using a novel self-training neural network modeling approach. Advanced Engineering Informatics, 62, 102674.https: //doi.org/10.1016/j.aei.2024.102674

[5] Wali, S. B., Abdullah, M. A., Hannan, M. A., Hussain, A., Samad, S. A., Ker, P. J., & Mansor, M. B. (2019). Vision-based traffic sign detection and recognition systems: Current trends and challenges. Sensors, 19(9), 2093.https: //doi.org/10.3390/s19092093

[6] Rehman, Y., Riaz, I., Fan, X., & Shin, H. (2017). D-patches: effective traffic sign detection with occlusion handling. IET Computer Vision, 11(5), 368-377. https: //doi.org/10.1049/iet-cvi.2016.0303

[7] Temel, D., Alshawi, T., Chen, M. H., & AlRegib, G. (2019). Challenging environments for traffic sign detection: Reliability assessment under inclement conditions. arxiv preprint arxiv: 1902.06857.https: //doi.org/10.48550/arXiv.1902.06857

[8] Kang, S., Hu, Z., Liu, L., Zhang, K., & Cao, Z. (2025). Object detection YOLO algorithms and their industrial applications: Overview and comparative analysis. Electronics, 14(6), 1104.https: //doi.org/10.3390/electronics14061104

[9] Alif, M. A. R. (2024). Yolov11 for vehicle detection: Advancements, performance, and applications in intelligent transportation systems. arxiv preprint arxiv: 2410.22898.https: //doi.org/10.48550/arXiv.2410.22898

[10] Wang, C., Song, X., Wang, J., & Yan, X. (2025). An improved YOLOv11 algorithm for small object detection in UAV images. Signal, Image and Video Processing, 19(6), 1-12.https: //doi.org/10.1007/s11760-025-04157-w

[11] Zhang, Z., Zhang, Z., Li, G., & Xia, C. (2025). ZZ-YOLOv11: A Lightweight Vehicle Detection Model Based on Improved YOLOv11. Sensors, 25(11), 3399.https: //doi.org/10.3390/s25113399

[12] Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. arxiv preprint arxiv: 2410.17725..https: //doi.org/arxiv-2410.17725

[13] Yi, H., Liu, B., Zhao, B., & Liu, E. (2023). Small object detection algorithm based on improved YOLOv8 for remote sensing. IEEE journal of selected topics in applied earth observations and remote sensing, 17, 1734-1747.https: //doi.org/10.3390/rs16163057

[14] Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence, 39(6), 1137-1149.https: //doi.org/10.48550/arXiv.1506.01497

[15] Sharma, S., Patel, D., & Shah, D. (2023, November). Automated Detection and Classification of Defects in Pcb Using Deep Learning Techniques: A Comparison of Resnet50v1 and Efficient Det D1. In 2023 9th IEEE India International Conference on Power Electronics (IICPE) (pp. 1-5). IEEE.https: //doi.org/10.1109/IICPE60303.2023.10475108

[16] Sonsare, P., Gupta, D., Sayyed, J., Nandha, P., & Laira, S. (2024, February). Sign language to text conversion using deep learning techniques. In 2024 2nd International Conference on Computer, Communication and Control (IC4) (pp. 1-6). IEEE.https: //doi.org/10.1109/IC457434.2024.10486568

DOI: 10.54254/2755-2721/2025.25898

Appendix

Appendix A

Table A.1 YOLOv11 original architecture defects

Defect Dimensions	Specific Manifestations	Impact on Sign Detection
Feature Fusion Capability	Unidirectional Feature Pyramid Network (FPN) with severe information attenuation	Loss of small-scale sign features
Anchor Design	Fixed-ratio anchors incompatible with sign aspect ratios	Positioning deviation of slender signs such as speed limit signs
Background Interference Suppression	Lack of spatial attention mechanism	Increased false detection rate in complex scenarios such as tree branch occlusion

Table A.2 Scenario adaptive data enhancement strategy

Enhancement Type	Implementation Principle	Target Scenario
Illumination Disturbance	In the HSV color space: hue shift by ±20°, saturation multiplied by a random coefficient in [0.5, 1.5]	Abrupt light-dark transitions at tunnel entrances and exits
Partial Occlusion	Covering gray rectangles with an area ratio of 0.02~0.2 at random positions, with Gaussian blurred edges applied	Scenarios with occlusion by vehicles/pedestrians
Motion Blur	Randomly generating linear motion blur kernels with directional angles in [-45°, 45°] and kernel sizes of 7~11 pixels	Image motion blur artifacts during high-speed driving

Table A.3 Key hyperparameter configuration

Module	Parameter Name	Setting Value
Optimizer	Type	AdamW
Learning Rate	LR Schedule	Cosine
Anchor Assignment	SimOTA	top_k=10
Regularization	Label Smoothing	ε=0.05

DOI: 10.54254/2755-2721/2025.25898

Appendix B

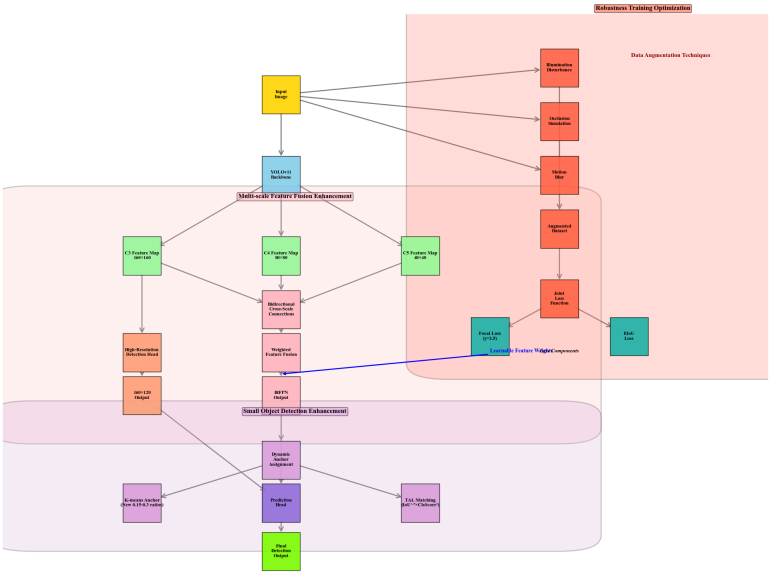

Figure B.1 Schematic diagram of specific improvement strategies

Appendix C

Table C.1 Dataset features

Scale Range	Number of Instances	Proportion	Typical Scenarios
<16×16px	41,327	13.2%	Distant highway guide signs
16-32px	87,461	27.9%	Urban road auxiliary lane signs
32-64px	132,079	42.1%	Main road lane signs
>64px	52,634	16.8%	Large intersection signboards

Table C.2 Dataset division

Subset	Number of Images	Sign Instances	Number of Cities
Training Set	76,800	313,501	32
Validation Set	9,600	39,216	8
Test Set	13,600	55,784	8

DOI: 10.54254/2755-2721/2025.25898

Table C.3 Experimental environment

Category	Configuration Item	Parameter/Model
Computing Platform	Vehicle-mounted Embedded Calculator	NVIDIA DRIVE AGX Orin
	GPU CUDA Cores	2,048
Perception Sensors	Front-facing Camera	OnSemi AR0820AT
	Resolution/FPS	1280×720 @ 30fps
Operating System	Vehicle-mounted System	Ubuntu 20.04 + RT Kernel
Development Framework	Deep Learning Framework	PyTorch 1.13.1
	Compiler	NVCC 11.7
Benchmark Libraries	Vision Processing Library	OpenCV 4.7.0
	Mathematical Operation Library	cuDNN 8.5.0 + cuBLAS 11.10

Table C.4 Parameter settings

Parameter Category	Parameter Name	Setting Value
Data Loading	Batch Size	32
	Input Resolution	640×640
	Warm-up Iterations	500 steps
Optimization Strategy	Optimizer	AdamW
	Base Learning Rate	1e-3
	Learning Rate Scheduler	Cosine Annealing
	Cycle Length	300 epochs
Regularization	Weight Decay	0.025
	Label Smoothing	ε=0.05
	DropPath Probability	0.2
Loss Function	Classification Loss	Focal Loss
	Regression Loss	EIoU Loss
	Positive Sample Threshold	IoU≥0.5
Advanced Settings	EMA Decay Rate	0.9999
	Gradient Clipping	Max Norm=10.0
	Mixed Precision Training	AMP O2 Mode