Application of YOLOv8 in Monocular Downward Multiple Car Target Detection

Shijie Lyu[1,a,*]

[1]*Georgia institute of technology, North Avenue Atlanta, GA, 30332, United States of America*
a. slyu41@gatech.edu
*corresponding author

Abstract: Autonomous driving technology is progressively transforming traditional car driving methods, marking a significant milestone in modern transportation. Object detection serves as a cornerstone of autonomous systems, playing a vital role in enhancing driving safety, enabling autonomous functionality, improving traffic efficiency, and facilitating effective emergency responses. However, current technologies such as radar for environmental perception, cameras for road perception, and vehicle sensor networks face notable challenges, including high costs, vulnerability to weather and lighting conditions, and limited resolution.To address these limitations, this paper presents an improved autonomous target detection network based on YOLOv8. By integrating structural reparameterization technology, a bidirectional pyramid structure network model, and a novel detection pipeline into the YOLOv8 framework, the proposed approach achieves highly efficient and precise detection of multi-scale, small, and remote objects. Experimental results demonstrate that the enhanced model can effectively detect both large and small objects with a detection accuracy of 65%, showcasing significant advancements over traditional methods.This improved model holds substantial potential for real-world applications and is well-suited for autonomous driving competitions, such as the Formula Student Autonomous China (FSAC), particularly excelling in scenarios involving single-target and small-object detection.

Keywords: YOLOv8, autonomous, Security, object detection, Multiscale, FSAC

1.   Introduction

Autonomous driving technology has advanced rapidly, offering benefits such as improved safety, traffic efficiency, comfort, and environmental friendliness [1][2]. It reduces human errors like fatigue and distraction, enhances safety, optimizes driving paths, alleviates congestion, and allows passengers to focus on other activities [3]. However, achieving high-precision detection of road targets remains a critical challenge [4].

Object detection is essential for autonomous driving, directly impacting safety, autonomy, traffic efficiency, and user experience [5]. It enables vehicles to identify and track pedestrians, vehicles, and obstacles, ensuring timely actions to avoid collisions and improve safety. Accurate detection supports autonomous decision-making, adherence to traffic rules, and seamless integration into traffic flow, reducing congestion and travel time. In emergencies, it helps vehicles respond swiftly to hazards, protecting passengers and pedestrians. For autonomous racing, object detection is crucial to advancing driverless technology and achieving full autonomy in driverless formula cars.

Current autonomous driving detection technologies [6], including radar, cameras [7], and on-board sensor networks [8], face challenges such as high costs, weather susceptibility, and resolution limitations. Radar accuracy decreases in adverse weather and on reflective surfaces, while cameras, though effective for detecting road signs and lane lines, struggle with poor lighting and weather conditions like rain or haze [9]. On-board sensor networks provide comprehensive environmental perception but involve complex data processing [10].

2. YOLOv8 Improves the Frame Structure of the Model

The original YOLOv8 model faces challenges such as limited adaptability, inadequate feature extraction, and suboptimal information flow when dealing with multi-scale, small, and distant object detection.

This paper enhances the YOLOv8-based autonomous target detection network by incorporating structural reparameterization, a two-way pyramid structure, and a new detection pipeline. These improvements aim to achieve high-efficiency and high-precision detection of multi-scale, small, and distant objects.

Structural reparameterization technology optimizes the network structure, enhancing its suitability for multi-scale and small target detection without sacrificing accuracy. This adaptation improves the detection of targets of varying sizes. The bidirectional pyramid structure processes multi-scale feature information, capturing spatial and semantic details more effectively, which aids in detecting distant and small targets. Additionally, the new detection pipeline structure optimizes information flow, enhancing inspection efficiency and accuracy through modules like feature fusion and information transfer. Figure 1 illustrates the framework of the improved YOLOv8 model, with detailed explanations of the improvements provided below.
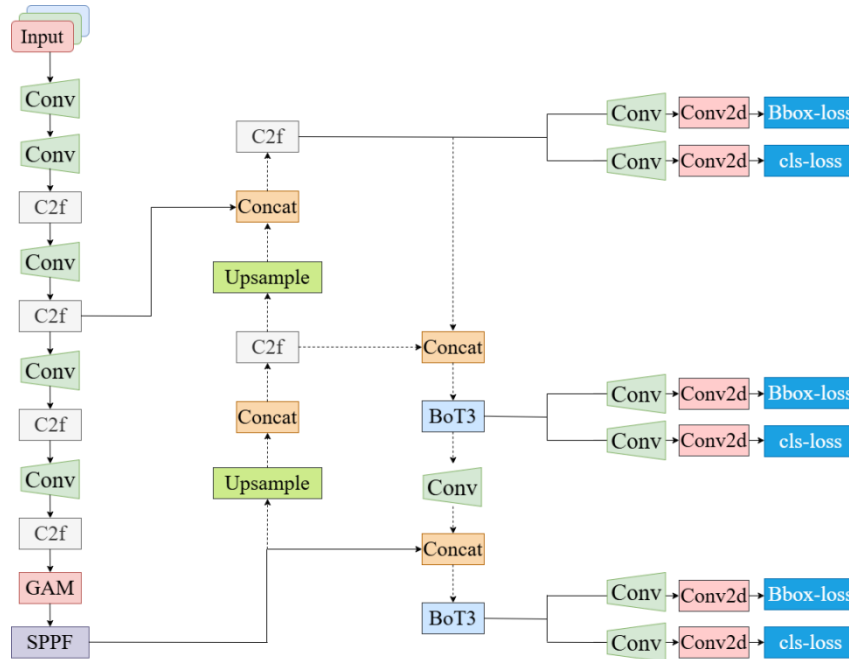

Figure 1: Improved YOLOv8 model structure.

2.1. The introduction of backbone network introduces

The DBB (Different Branch Blocks) model enhances the backbone network's feature extraction by integrating branches that focus on different scales, semantics, or aspects of an image. By augmenting

the original backbone with these branching blocks, the network effectively captures multi-scale and multi-semantic information, significantly improving the detection of distant and small targets.

In autonomous driving, vehicles must promptly detect and recognize obstacles and traffic signs, including distant and small objects like traffic signs, vehicles, and pedestrians. Due to their size and distance, these objects may appear blurry or subtle, requiring accurate detection and identification. By incorporating different branching block models, the network can effectively integrate multi-scale and semantic information, enhancing its understanding of image content.

This paper combines multiple DBB modules to enhance the backbone network for detecting distant and small-sized objects. To complement the improved backbone, structural reparameterization technology is integrated into the C2f-DBB module, boosting detection speed and accuracy, as illustrated in Figure 2.
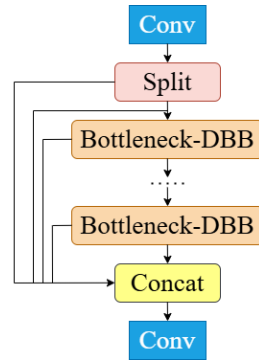

Figure 2: Structural reparameterization technology is introduced into the C2F-DBB module.

2.2. The neck structure introduces a bidirectional pyramid structure network model

The YOLOv8 neck structure transforms feature maps from the backbone into optimized representations for object detection, performing fusion, compression, enhancement, and adjustment to boost network performance and efficiency.

A key limitation of the original YOLOv8 network's Path Aggregation Feature Pyramid Network (PAFPN) as the neck structure is its unidirectionality, which can result in incomplete feature extraction. This restricts the effective integration and utilization of multi-scale and semantic features, potentially impacting object detection performance.

To address this limitation, this paper introduces a bidirectional pyramid network model into the neck structure, maintaining a transmission mode similar to the original YOLOv8 model, as shown in Figure 3. Compared to the unidirectional pyramid structure, the bidirectional design offers enhanced feature integration and flexibility, enabling more comprehensive information capture and improving the performance and accuracy of autonomous target detection.
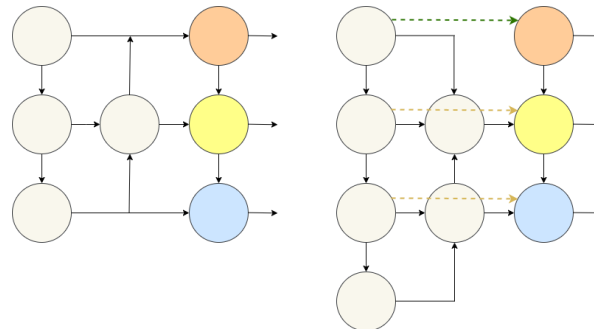

Figure 3: Comparison between the pyramid network of path aggregation features and the bidirectional pyramid structure network model.

2.3. A new model of the inspection pipeline structure was introduced

The YOLOv8 neck structure transforms feature maps from the backbone network into representations optimized for object detection, incorporating feature fusion, compression, enhancement, and adjustment to boost performance and efficiency. However, the original YOLOv8's Path Aggregation Feature Pyramid Network (PAFPN) suffers from unidirectionality, limiting the integration and utilization of multi-scale and semantic features, which can impact detection performance. To address this, a bidirectional pyramid network model is introduced, as shown in Figure 3. Compared to the unidirectional structure, the bidirectional design enhances feature integration and flexibility, enabling more comprehensive information capture and improving detection performance and accuracy.
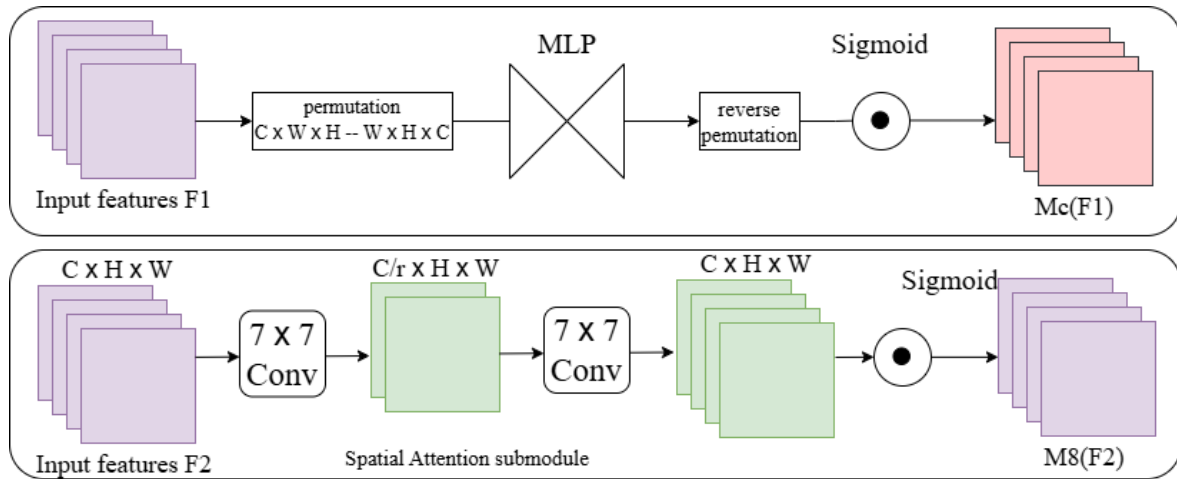

Figure 4: Schematic diagram of the structure of the new inspection pipeline structure model.

3. EXPERIMENTAL RESULTS AND DISCUSSION

3.1. Experimental conditions are set

The images used in this paper have a resolution of 1280. The model was trained for 100 rounds using the SGD optimizer with a batch size of 16 and 64GB of memory.

This paper utilizes the SODA-D and VisDrone datasets for evaluation. SODA-D (Small Object Detection in Aerial Images - Drone) focuses on detecting small objects in UAV aerial images, offering diverse categories that provide abundant data for small object detection research. VisDrone, a large-scale dataset for UAV video analysis and target detection, includes aerial footage from multiple cities worldwide, covering varied scenarios and environmental conditions. Its challenges, such as occlusions, partial visibility, and small-sized targets, demand high algorithmic robustness and performance. Both datasets are well-suited for addressing small target detection needs in FSAC competition scenarios.

3.2. Evaluate the parameters

To demonstrate the reliability of the proposed improved model, evaluation parameters for the SODA-D and VisDrone datasets were compared based on the YOLOv8 model with various block improvements. The results of these comparisons are presented in Table 1 and Table 2, respectively.

The evaluation parameters in this study include precision (P), recall (R), mAP@0.5, GFLOPS, Params, and FPS. Together, these metrics provide a comprehensive assessment of defect detection algorithms, encompassing both accuracy and efficiency.

Table 1: COMPARISON OF MODEL PARAMETERS BASED ON SODA-D DATABASE

Baseline	DBB	Qb	BFFPN	mAP@0.5	mAP@0.5:0.95	P	R
√				61.8	36.8	70.1	56.1
√	√			62.2	36.3	70.6	56.7
√	√	√		62.8	35.8	71.2	57.4
√	√	√	√	65.2	38.3	72.5	58.9

Table 2: COMPARISON OF MODEL PARAMETERS BASED ON VISDRONE DATABASE

Baseline	DBB	Qb	BFFPN	mAP@0.5	mAP@0.5:0.95	P	R
√				30.5	16.7	42.0	31.7
√	√			31.1	15.8	42.5	31.5
√	√	√		32.6	16.6	43.0	32.4
√	√	√	√	34.5	16.6	44.5	33.9

As shown in the table, the improved model achieves a detection accuracy of 65% for both larger and smaller objects, attributed to enhancements in YOLOv8. Figures 5 and 6 display the PR curves of the YOLOv8 model before and after improvement under identical conditions. The improved model achieves an overall mAP@0.5 approximately 7% higher than the original, with a maximum mAP value of 0.716, significantly surpassing the unimproved structure. These results demonstrate that the proposed improvements to the YOLOv8 model greatly enhance detection accuracy, underscoring the significance of this work.

3.3. Visualize experiments

This paper compares the proposed model with the baseline method based on YOLOv8 technology. Experimental images from the SODA-D database are used, displaying results from the original image, the proposed model, and the baseline method, as shown in Figure 7. The improved model demonstrates higher detection accuracy and more precise object positioning.

Figure 7: The model in this paper is compared with the baseline method.

4. Conclusion

Autonomous technology is rapidly advancing, with object detection playing a crucial role in ensuring driving safety, autonomy, traffic efficiency, and emergency response. Traditional detection technologies face limitations such as high costs, weather susceptibility, and low resolution. This paper proposes an improved autonomous target detection network based on YOLOv8, incorporating structural reparameterization, a bidirectional pyramid structure, and a new detection pipeline to achieve efficient and precise detection of multi-scale, small, and distant objects. Experimental results demonstrate a detection accuracy of 65.2% for larger and smaller objects, highlighting the model's effectiveness. The proposed approach holds promise for real-world applications and autonomous competitions, such as the Formula Student Autonomous China (FSAC), particularly in single-target and small-target detection scenarios.

References

[1] Kaur K, Rampersad G. Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars[J]. Journal of Engineering and Technology Management, 2018, 48: 87-96.

[2] Goodchild M F. GIScience for a driverless age[J]. International Journal of Geographical Information Science, 2018, 32(5): 849-855.

[3] Li T, He Z, Wu Y. An integrated route planning approach for driverless vehicle delivery system[J]. PeerJ Computer Science, 2022, 8: e1170.

[4] Paschalidis E, Hajiseyedjavadi F, Wei C, et al. Deriving metrics of driving comfort for autonomous vehicles: A dynamic latent variable model of speed choice[J]. Analytic methods in accident research, 2020, 28: 100133.

[5] Zhao H, Peng X, Wang S, et al. Improved object detection method for unmanned driving based on Transformers[J]. Frontiers in neurorobotics, 2024, 18: 1342126.

[6] Royo S, Ballesta-Garcia M. An overview of lidar imaging systems for autonomous vehicles[J]. Applied sciences, 2019, 9(19): 4093.

[7] Kuzin V D, Pronin P I, Zarubin D N, et al. Unmanned vehicle LIDAR-camera sensor fusion technology[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2019, 695(1): 012012.

[8] Wu B, Cai Z, Wu W, et al. AoI-aware resource management for smart health via deep reinforcement learning[J]. IEEE Access, 2023.

[9] Du Y, Liu X, Yi Y, et al. Optimizing road safety: advancements in lightweight YOLOv8 models and GhostC2f design for real-time distracted driving detection[J]. Sensors, 2023, 23(21): 8844.

[10] Mecocci A, Grassi C. RTAIAED: a real-time ambulance in an emergency detector with a pyramidal part-based model composed of MFCCs and YOLOv8[J]. Sensors, 2024, 24(7): 2321.