

Research on Autonomous Driving Performance Enhancement in Low-Light Scenarios Based on Improved Object Detection Algorithm

Xiye Luo. Shixin Li

Tianjin University of Technology and Education, Tianjin, 300222, China

Email: 841593202@qq.com

How to cite this paper: Luo, X., & Li, S. (2026). Research on Autonomous Driving Performance Enhancement in Low-Light Scenarios Based on Improved Object Detection Algorithm. Academic Journal of Emerging Technologies, 3(3), 58–68. ISSN Print: 3104-4417, ISSN Online: 3104-4425. <https://doi.org/10.63313/AJET.9073>

Published: 2026-09-28

Copyright © 2026 by author(s) and Erytis Publishing Limited.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Abstract

In low-light scenarios, autonomous driving perception systems face problems such as weak contrast, high noise, blurred boundaries, and missing texture details, which pose great challenges to object detection tasks. Existing methods often use image enhancement as a preprocessing step to improve detection performance; however, such methods usually require additional computational overhead and struggle to simultaneously meet the requirements of detection accuracy and real-time performance. To address these issues, this paper takes YOLOv8n as the baseline model and proposes an improved object detection algorithm, BiF-VoV-YOLO, for low-light autonomous driving scenarios. First, a lightweight feature fusion module, VoVGSCSP, is introduced into the backbone network to replace the original SPPF module. Through grouped feature reuse and channel shuffling, it reduces the number of model parameters and computational cost while enhancing low-light feature representation. Second, a BiFormer fusion layer is embedded in the neck, using a content-adaptive sparse attention mechanism to filter low-light background noise and improve global dependency modeling and detection accuracy. Finally, the upsampling node nn.Upsample is reconfigured and added in the neck. To address the lack of content-awareness in original upsampling, which leads to the loss of small-object details, this design improves information fidelity during the upsampling stage. The synergistic effect of the three modules significantly alleviates missed detection and false detection of multi-scale objects in low-light scenarios on the DarkFace dataset while maintaining the model's real-time inference capability. Experimental results show that the improved model achieves an mAP@0.5 of 53.9%, an improvement of 8.9 percentage points over the YOLOv8n baseline, and a reduction of about 0.05MB in parameter count. It achieves the best detection accuracy among various classic models and state-of-the-art models, verifying the effectiveness and generalization of the proposed algorithm in low-light autonomous driving object detection.

Keywords

Object Detection; Autonomous Driving; Low-Light Scenarios; YOLO

1. Introduction

With the rapid development of intelligent transportation systems and autonomous driving technology, environmental perception has become a core component of safe decision-making and path planning for autonomous vehicles. As a fundamental task in environmental perception, object detection is responsible for identifying key targets such as vehicles, pedestrians, traffic signs, and obstacles from images or videos. Its detection accuracy and real-time performance directly affect the safety of autonomous driving systems. However, in real road scenarios, lighting conditions are often complex and variable. Nighttime, tunnels, underground parking garages, rainy weather, and areas with strong shadows can all create low-light environments. Low-light images usually suffer from insufficient brightness, reduced contrast, increased noise, blurred edges, and color distortion, which significantly reduce the distinguishability between targets and background, posing severe challenges to object detection algorithms.

In low-light scenarios, traditional object detection methods often encounter two types of problems. The first is missed detection: because target edge gradients are weakened and texture information is missing, the network struggles to extract effective features, causing small targets such as pedestrians and non-motorized vehicles to be submerged in the background. The second is false detection: due to factors such as sensor noise, road shadows, and light reflections, the network may misclassify non-target regions as targets. For autonomous driving, both missed detection and false detection can lead to serious consequences. Therefore, researching robust and efficient object detection algorithms for low-light scenarios has important theoretical significance and engineering application value.

Existing research usually improves low-light object detection performance from two directions. One is to adopt image enhancement methods, such as histogram equalization, Retinex, and low-light image enhancement networks, to preprocess input images and improve visual quality. However, image enhancement methods often require additional computational resources, and the enhanced results are not necessarily beneficial to the detection network. Over-enhancement may amplify noise and cause the detector to focus on incorrect regions; separately training the enhancement model and the detection model may also lead to task inconsistency. The other direction is to directly improve the object detection network structure so that it has stronger representation capability and robustness on low-light features. Compared with image enhancement, network structure improvement is more likely to achieve a balance between accuracy and speed, and is especially suitable for the real-time inference requirements of autonomous driving.

The YOLO series algorithms are widely used in autonomous driving object detection due to their simple structure, fast detection speed, and convenient deployment. As a lightweight model, YOLOv8n has a small number of parameters and high inference efficiency, making it suitable for deployment on edge devices. However, in low-light

scenarios, YOLOv8n still suffers from insufficient feature extraction capability, severe background noise interference, and loss of small-object details. To this end, this paper takes YOLOv8n as the baseline model and proposes the BiF-VoV-YOLO algorithm for autonomous driving object detection in low-light environments. The specific improvements of this paper are as follows:

- (1) A lightweight feature fusion module, VoVGSCSP, is used to replace the original SPPF module at the end of the backbone network. Based on lightweight convolution and a cross-stage partial connection strategy, VoVGSCSP employs grouped feature reuse and channel shuffling to reduce computational cost while preserving a richer spatial response distribution, enabling the network to focus more on the overall semantic structure of targets under low-light conditions.
- (2) A BiFormer fusion layer is embedded into the low-resolution branches of the neck. By constructing content-adaptive sparse attention paths, BiFormer dynamically selects key regions at different spatial locations for computation. This captures global dependencies while maintaining linear computational complexity, thereby filtering low-light background noise and improving detection accuracy.
- (3) The upsampling node `nn.Upsample` is reconfigured in the neck. To address the lack of content awareness in original upsampling, which leads to the loss of small-object details, a deterministic nearest-neighbor interpolation strategy is adopted. Without introducing additional trainable parameters, this strategy preserves deep feature semantics to the greatest extent and uses shallow features to supplement spatial details.

2. Dataset

2.1. DarkFace Dataset

This low-light face detection dataset contains approximately 6,000 low-light face images captured in real nighttime environments, covering various complex scenes such as teaching buildings, overpasses, parks, and roadsides. The dataset additionally provides 9,000 unlabeled low-light images from the same scenes, and 4,000 annotated images are reserved as an independent test set. The image acquisition environments truly reflect problems commonly encountered in nighttime surveillance and nighttime perception for autonomous driving, such as complex illumination, motion blur, sensor noise, and image shadows. Therefore, this dataset has become one of the important benchmark datasets in the field of low-light computer vision. Figure 1 and Figure 2 show some sample images from the dataset.



Figure 1. DarkFace Dataset



Figure 2. DarkFace Dataset

From the perspective of data distribution, DarkFace has the following characteristics. First, the illumination is extremely uneven, with some regions nearly completely dark, so target contours can only be identified by weak reflected light or local light sources. Second, the noise types are complex, including not only Gaussian noise generated by low-light imaging but also mixed degradations such as compression and motion blur. Third, the target scale varies significantly, including both large targets at close range and small targets at long range. Fourth, background interference is strong, as streetlights, vehicle lights, glass reflections, and road shadows are easily confused with real targets. These characteristics make DarkFace not only suitable for face detection but also

capable of effectively evaluating the robustness of object detection algorithms under low-light conditions.

3. Model Improvements

3.1. Lightweight Feature Fusion Module VoVGSCSP

In the baseline backbone network of YOLOv8, the SPPF module expands the receptive field by cascading max-pooling layers of different scales, thereby enhancing the network's perception capability. However, in nighttime environments, the edge gradients of targets are weakened and texture contrast decreases, causing a large amount of critical detail information to be ignored. In addition, low-light images suffer from sensor noise and road shadows. To address these problems, this paper proposes replacing the original SPPF module at the end of the backbone network with the VoVGSCSP module. This module is based on lightweight convolution and a cross-stage partial connection strategy. Through grouped feature reuse and channel shuffling, it reduces computational cost while preserving a richer spatial response distribution. Grouped convolution can reduce the number of parameters and floating-point operations, while channel shuffling promotes information interaction between different groups and avoids the information isolation caused by grouped convolution. The cross-stage partial connection divides the feature map into two parts: one part is processed by convolution, and the other part is directly concatenated with the processed result, thereby reducing computational complexity while maintaining gradient flow.

This module can also aggregate local context in a more flexible manner, enabling the network to focus more on the overall semantic structure of targets under low-light conditions. Thus, while reducing the number of model parameters and floating-point operations, it provides more robust feature inputs for the subsequent detection head, fundamentally alleviating the detection performance degradation problem in low-light scenarios.

In addition, VoVGSCSP continues the OSA design concept of VoVNet. It concatenates and aggregates all branch features of the intermediate layers only once at the end of the module, avoiding the problem of excessively high channel expansion caused by DenseNet-style multi-layer connections, and improving the deficiency of long-range capability in object detection networks. Let the set of output feature maps of N GSBottlenecks in the module be $\{F_1, F_2, \dots, F_N\}$. Then the OSA aggregation operation can be expressed as shown in Equation (1).

$$F_{out} = Conv_{1 \times 1}(Concat(F_1, F_2, \dots, F_N, F_{shortcut})) \quad (1)$$

To verify the effectiveness of the VoVGSCSP module structure for autonomous driving object detection algorithms in low-light environments, GSConv, PConv,

and ODConv are respectively introduced into YOLOv8n, and object detection results are validated on the DarkFace dataset. The experimental results are shown in Table 1.

Table 1. Comparative Experiment of the VoVGSCSP Module

Models	Parameters(MB)	mAP50(%)	mAP50-95(%)	GFLOPs
YOLOv8n-GSConv	2.7	48.9	19.5	8.0
YOLOv8n-PConv	3.2	49.4	19.7	8.6
YOLOv8n-ODConv	3.5	49.6	20.2	9.4
YOLOv8n-VoVGSCSP	2.6	50.1	20.6	8.6

3.2. BiFormer Fusion Layer

In low-light autonomous driving scenarios, target features are often difficult to extract effectively due to low contrast and strong noise interference, which places higher demands on the feature fusion layer of the network. Traditional convolutional neural networks mainly rely on local convolution kernels to extract features. Although they can capture local textures and edges, their receptive field is limited. Under weak-light conditions, long-range contextual information, overall contours, and scene relationships of targets are particularly important for distinguishing real targets from background noise. If only ordinary 3×3 convolutions are used for feature transmission, the network is prone to falling into local optima and struggles to make full use of global information, resulting in missed detection of small targets or false detection of shadows.

Zhu et al. proposed the BiFormer mechanism. By constructing content-adaptive sparse attention paths, this mechanism can dynamically select key regions involved in computation at different spatial locations, effectively capturing global dependencies while maintaining linear computational complexity. Compared with standard self-attention, BiFormer does not perform dense computation over all spatial positions. Instead, it first uses a routing mechanism to select the regions most relevant to the current query, and then performs attention computation on these regions. This reduces computational complexity while retaining the content-adaptive capability of dynamic sparse attention. For low-light images, BiFormer can suppress irrelevant background regions and concentrate computational resources on regions that may contain targets, thereby improving the pertinence and robustness of feature fusion. In response to this characteristic, this paper improves the original structure of the neck by embedding the module into the 80 and 40 branches, replacing the paths that originally relied only on ordinary 3×3 convolutions for feature transmission. At the concatenation point of each branch and after the C2f feature extraction module, the module is serially connected to compensate for the deficiency of convolutional local receptive fields in utilizing long-range contextual information under weak-light environments, while ensuring real-time inference efficiency and the purity of shallow texture features.

The BRA mechanism is summarized from the BiFormer multi-scale fusion layer. Its

main idea is to preserve the global advantages of Transformers while using a content-aware strategy to reduce computational cost and complexity. Standard multi-head self-attention computes affinity for tokens at all spatial positions, and its computational complexity is $O(N^2)$, where $N=H \times W$ is the total number of tokens. The BRA mechanism implements adaptive attention through two-level routing. First, it divides the input into $S \times S$ non-overlapping regions, each containing $\frac{HW}{S^2}$ tokens. The query, key, and value tensors obtained by linear mapping are expressed as shown in Equations (2)–(4).

$$Q = X^r W^q \quad (2)$$

$$K = X^r W^k \quad (3)$$

$$V = X^r W^v \quad (4)$$

Where W^q, W^k , and W^v are learnable projection matrices.

To verify the effectiveness of the BiFormer multi-scale fusion layer structure for autonomous driving object detection algorithms in low-light environments, FPN, ASPP, and CA-PAN are respectively introduced into YOLOv8n, and object detection results are validated on the DarkFace dataset. The experimental results are shown in Table 2.

Table 2. Comparative Experiment of the BiFormer Multi-Scale Fusion Layer Module

Models	Parameters(MB)	mAP50(%)	mAP50-95(%)	GFLOPs
YOLOv8n-FPN	3.2	47.2	17.1	8.8
YOLOv8n-ASPP	3.5	48.6	18.9	9.3
YOLOv8n-CA-PAN	3.3	48.0	17.8	9.0
YOLOv8n-BiFormer	3.7	49.7	20.5	10.2

3.3. Upsampling Operator

In object detection networks, upsampling is responsible for restoring deep low-resolution features to a higher resolution so that they can be fused with shallow features. The neck of YOLOv8 usually employs upsampling operations to achieve multi-scale feature fusion. However, the original upsampling method often lacks content-awareness and easily leads to the loss of small-object details. In low-light scenarios, small objects themselves have few pixels and low contrast. If the upsampling process cannot well preserve semantic and spatial information, the subsequent fused features will find it difficult to recover object contours, resulting in missed detection.

The nn.Upsample module is a deterministic sizing tool based on interpolation algorithms. Its main function is to introduce no additional trainable parameters, and its computational overhead is extremely low. Compared with bilinear interpolation or transposed convolution, nearest-neighbor upsampling does not have an

DOI:10.63313/AJET.9073

DOI:10.63313/AJET.9073

DOI:10.63313/AJET.9073

DOI:10.63313/AJET.9073

DOI:10.63313/AJET.9073



DOI:10.63313/AJET.9073

4.1. Ablation Study

To verify the effectiveness of the multi-scale BiF-VoV-YOLO model improvement method for the experiments, a series of ablation studies were designed. For the DarkFace dataset used in this paper, ablation experiments were designed to examine the degree of optimization of the baseline YOLOv8n network model by the VoVGSCSP backbone network, the BiFormer multi-scale fusion layer, and the nn.Upsample upsampling operator. Evaluating different improvement methods can provide a reference for further optimization of the algorithm. The specific ablation study results are shown in Table 3, where “√” indicates that the method is introduced.

Table 3. Comparison of Ablation Study Results

Models	VoVGSCSP	BiFormer	nn.Upsample	Parameters	mAP50
YOLOv8n	×	×	×	3.1	45.0
YOLOv8n-vov	√	×	×	2.6	50.1
YOLOv8n-bif	×	√	×	3.7	49.7
YOLOv8n-up	×	×	√	3.3	51.3
YOLOv8n-vb	√	√	×	3.1	51.2
YOLOv8n-vu	√	×	√	2.9	52.9
YOLOv8n-bu	×	√	√	3.5	51.8
BiF-VoV-YOLO	√	√	√	3.1	53.9

As can be seen from Table 3, the YOLOv8n baseline model achieves an average precision of 45.0% on the dataset, with a parameter count of 3.1 MB. When only the VoVGSCSP backbone network is added, the YOLOv8n-vov model improves mAP@0.5 by 5.1% over the baseline model and reduces the parameter count by 0.5 MB, indicating that the VoVGSCSP backbone network reduces the model parameter count and computational cost while enhancing low-light feature representation capability. After adding the BiFormer multi-scale fusion layer, the YOLOv8n-bif model improves mAP@0.5 by 4.7% over the baseline model, which can filter low-light background noise and improve detection accuracy. When the nn.Upsample upsampling operator is added, the YOLOv8n-up model improves mAP@0.5 by 6.3% over the baseline model. This module addresses the problem that original upsampling lacks content-awareness and leads to the loss of small-object details, thereby improving information fidelity during the upsampling stage. YOLOv8n-vb, YOLOv8n-vu, and YOLOv8n-bu models are cumulative ablation models, and the final model is the detection model that integrates the three innovations. The YOLOv8n-vb model improves mAP@0.5 by 6.2% over the baseline model, the YOLOv8n-vu model improves mAP@0.5 by 7.9% over the baseline model, and the YOLOv8n-bu model improves mAP@0.5 by 6.8% over the baseline model. All of them show improvements over the YOLOv8n baseline model, and the parameter count does not increase significantly. The final BiF-VoV-YOLO model achieves the best mAP@0.5 of

53.9%, an improvement of 8.9% over the baseline model, while the parameter count is reduced by 0.05 MB. It does not simply increase detection accuracy at the cost of a large increase in parameters, which verifies the effectiveness of autonomous driving object detection in low-light environments.

4.2. Comparative Experiments with Classic and SOTA Models

To further verify the effectiveness of the proposed multi-scale BiF-VoV-YOLO model for autonomous driving object detection in low-light environments, the multi-scale BiF-VoV-YOLO model is compared with six different classic models and state-of-the-art (SOTA) models, namely YOLOv5n, YOLOv8n, YOLOv10n, EDE-YOLO, 3L-YOLO, and ALIEN-YOLO, trained on the DarkFace dataset. The experimental results are shown in Table 4.

Table 4. Comparison of Experimental Results

Models	mAP50(%)	mAP50-95(%)	Parameters(MB)
YOLOv5n	44.8	19.5	2.5
YOLOv8n	45.0	19.8	3.1
YOLOv10n	41.3	17.8	2.7
EDE-YOLO	52.0	25.3	15.6
3L-YOLO	47.0	-	-
ALIEN-YOLO	53.8	-	-
BiF-VoV-YOLO	53.9	23.8	3.1

As can be seen from Table 4, on the DarkFace dataset, the multi-scale BiF-VoV-YOLO model achieves the highest mAP@0.5 of 53.9% among all compared models. The algorithms in EDE-YOLO, 3L-YOLO, and ALIEN-YOLO are consistent with the dataset used in this paper. The experimental results show that the algorithm proposed in this paper has superior detection capability. The improved multi-scale BiF-VoV-YOLO model improves recall, precision, and mean average precision in object detection tasks compared with the YOLOv5n, YOLOv8n, and YOLOv10n baseline models. The precision variation of the multi-scale BiF-VoV-YOLO model is more stable, while the downward trend in precision of the traditional models is more obvious. This means that the improved multi-scale BiF-VoV-YOLO model is more lightweight and has more stable performance.

5. Conclusion

This paper addresses the problem of insufficient accuracy in autonomous driving object detection methods under low-light environments and proposes an autonomous driving object detection algorithm based on YOLOv8n, namely the BiF-VoV-YOLO model. The integration of the three modules improves object detection accuracy in low-light environments, mitigates the missed detection and false detection of multi-scale targets in low-light scenarios, while maintaining the

model's real-time inference capability, reducing the model parameter count and computational cost, and enhancing low-light feature representation capability. The accuracy, recall, and mean average precision of object detection on the DarkFace dataset show that, compared with the traditional YOLOv8n model and many SOTA models, the proposed model achieves significant improvements in various aspects of performance and has better effectiveness and generalization.

References

- [1] Abdullah D M, Abdulazeez A M. Machine learning applications based on SVM classification a review[J]. Qubahan Academic Journal, 2021, 1(2): 81-90.
- [2] Al-Refai G, Elmoaqet H, Al-Refai A, et al. Two-stage object detection in low-light environments using deep learning image enhancement[J]. PeerJ Computer Science, 2025, 11: e2799.
- [3] Wei S, Yu S. Improvement of nighttime vehicle detection algorithm based on YOLOv8n[C]//Proceedings of the 2024 International Conference on Computer and Multimedia Technology. 2024: 430-436.
- [4] Zhu Y, Zhu W, Wang L. GLAM-YOLO: An Illumination-Calibrated and Global-Local Attention Enhanced YOLOv11 for Low-Light Object Detection[C]//2025 5th International Conference on Artificial Intelligence, Automation and High Performance Computing (AIAHPC). IEEE, 2025: 267-271.
- [5] Ci W, Hou H, Lu S, et al. Multi-module enhanced low-light object detection: From image enhancement to adaptive feature fusion[J]. Advanced Engineering Informatics, 2026, 69: 104097.
- [6] Zhao D, Shao F, Zhang S, et al. Advanced Object Detection in Low-Light Conditions: Enhancements to YOLOv7 Framework[J]. Remote Sensing, 2024, 16(23).
- [7] Acharya U R, Oh S L, Hagiwara Y, et al. A deep convolutional neural network model to classify heartbeats[J]. Computers in biology and medicine, 2017, 89: 389-396.
- [8] Bharati P, Pramanik A. Deep learning techniques—R-CNN to mask R-CNN: a survey[J]. Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019, 2019: 657-668.
- [9] Ren Y, Zhu C, Xiao S. Object detection based on fast/faster RCNN employing fully convolutional architectures[J]. Mathematical Problems in Engineering, 2018, 2018(1): 3598316.
- [10] Nakahara H, Yonekawa H, Fujii T, et al. A lightweight YOLOv2: A binarized CNN with a parallel support vector regression for an FPGA[C]//Proceedings of the 2018 ACM/SIGDA International Symposium on field-programmable gate arrays. 2018: 31-40.
- [11] Wang C Y, Liao H Y M. YOLOv1 to YOLOv10: The fastest and most accurate real-time object detection systems[J]. arXiv preprint arXiv:2408.09332, 2024.
- [12] Loh Y P, Chan C S. Getting to know low-light images with the exclusively dark dataset[J]. Computer vision and image understanding, 2019, 178: 30-42.