Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 15, No. 1, 2025

Research on Multi object Detection in Foggy Sky Based on Fog Mod YOLOV8n

Jiangfeng Sun[1, a, *], Xilong Zhang[1, b], Yanyang Zeng[1, c]

[1]Henan Polytechnic University, College of Computer Science and Technology, Jiaozuo 454000, China

[a]Corresponding author:sunjiangfeng@hpu.edu.cn, [b] 2069065992@qq.com, [c] zengyy@hpu.edu.cn

Abstract: With the rapid development of smart cities and intelligent transportation, the level of informatization and intelligence in traffic systems is continuously improving. Studies have shown that the accident rate in foggy conditions is more than 30% higher than under normal weather. Fog significantly affects drivers' visibility, reducing safety and increasing driving risks. To address this issue, this paper proposes an improved Transformer-based model, Retinexformer, aimed at enhancing multi-object detection performance in foggy environments. First, to enhance the model's adaptability in complex environments, the Retinexformer model is optimized by incorporating a multi-branch Transformer module. This module enables the model to handle diverse input conditions and process information from different types of scenes more effectively. In addition, an adaptive parameter estimation module is designed to automatically adjust dehazing and image enhancement strategies based on the image's brightness and contrast. By dynamically setting different dehazing parameters, this module significantly improves the robustness of Retinexformer under varying foggy conditions.In the model's Decoder section, a multi-level feature enhancement module is introduced. After the IGT (Image Gradient Transformation) output, this module fuses multi-scale features, with a particular focus on enhancing details and small object detection capabilities, ensuring accurate identification of targets such as pedestrians and vehicles.The experimental findings indicate that the improved Retinexformer model demonstrates strong robustness and superior detection accuracy in foggy multi-object detection tasks, significantly enhancing traffic safety in foggy environments.

Keywords: Fog Environment, Multi-Object Detection, Transformer, Retinexformer, Dehazing, Image Enhancement, Perceptual Loss Function.

1. Introduction

As the need for smart cities and intelligent transportation systems continues to rise, tackling the difficulties of object detection in harsh weather conditions, such as fog, has become an essential task for improving traffic safety. Fog significantly impairs visibility [1], posing a considerable challenge for both drivers and autonomous systems in object recognition. In foggy conditions, traditional object detection algorithms often fail to maintain high performance, resulting in a higher likelihood of traffic accidents and diminished driving safety. In order to tackle this problem, this paper introduces a Foggy Multi-Object Detection Approach utilizing Retinex Theory and a Multi-Branch Transformer [2].

Retinex Theory[3], a classical image enhancement method, is particularly suitable for image processing under low-visibility conditions. It effectively recovers image details obscured by fog, enhancing the contrast and visibility of object. By incorporating Retinex Theory, the proposed algorithm mitigates the impact of fog on images, improving the clarity of important objects such as vehicles and pedestrians, while preserving the natural appearance of the scene. In this study, we extend the Transformer architecture by incorporating a multi-branch structure, allowing the model to process different levels of features and contextual information simultaneously[4]. This multi-branch structure enables the model to handle the complexity of foggy scenes, where different objects (e.g, distant objects, small targets, or occluded items) require separate handling. Each branch of the Transformer focuses on different aspects of the scene, such as foreground objects, background structures, and contextual information, ensuring that the model can effectively address occlusion issues and variations in object scales.The combination of Retinex Theory and Multi-Branch Transformer enables the proposed algorithm to better tackle the challenges posed by foggy environments [6].

The model is able to efficiently and accurately detect multiple objects in complex foggy scenes.This proposed foggy multi-object detection algorithm provides a reliable and efficient solution for object recognition in adverse weather conditions, making it especially suitable for real-time applications such as autonomous driving and intelligent transportation systems, where object detection in foggy environments is crucial for ensuring system safety and functionality[5].

2. YOLOv8 Network Model

Since its launch, the YOLO series has been widely used in object detection, video analysis, and autonomous driving with its efficient detection speed and real-time performance. YOLOv8[6] Continue the core design concept of YOLO and achieves goal detection through a single forward delivery. Its backbone network is based on the CSPDarknet[7] architecture, using more convolutional modules C2f[8], and SPPF[9] components, and residual connections to enhance the feature extraction capability. The neck network combines feature pyramid network (FPN) and path aggregation network (PAN) to effectively integrate multi-scale features and improve the detection accuracy of small objects.

YOLOv8 Using three decoupled detection heads to handle detection and classification tasks separately, enhancing model accuracy and flexibility. At the same time, YOLOv8 introduces the anchor-free detection method, which improves the positioning and classification ability in complex scenes. To meet computational resource limitations, YOLOv8 also provides a lightweight version (YOLOv8n), which provides good detection accuracy while maintaining low computational overhead. YOLOv8 Has been widely used in autonomous driving, security monitoring, intelligent traffic safety and other fields, and is the preferred model for many real-time visual tasks[10].

3. Fog-Mod-YOLOv8n Algorithm for Robust Object Detection in Foggy Environment

While the original YOLOv8n algorithm demonstrates outstanding performance in both accuracy and speed, its performance in foggy environments may degrade due to the low-quality images affected by fog, which can lead to a decrease in detection precision, especially in cases of occlusion, blurring, and uneven lighting between objects[11]. To overcome these issues, this paper presents an innovative architectural design—Fog-Mod-YOLOv8n.This model applies the Retinex theory for fog image enhancement and utilizes a multi-branch Transformer network to model interactions between multiple objects at a granular level. This method significantly improves both detection accuracy and robustness in foggy environments. When compared to the conventional YOLOv8, the upgraded version excels in identifying small, overlapping objects in low-contrast, fog-impacted scenes, while maintaining a balance between high efficiency and a lightweight framework. The model's performance has been confirmed through experiments on various fog-related datasets. The architecture of the Fog-Mod-YOLOv8n network is illustrated in the figure.

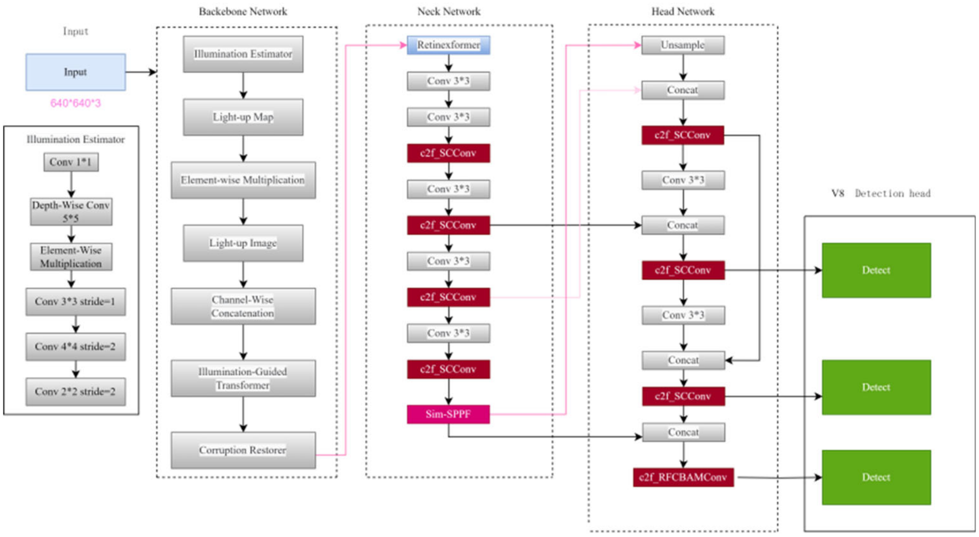

Figure 1. Structure of Fog-Mod-YOLOv8n

1) Introduction of Retinexformer model in this study consists of a single-stage Retinex framework and an Illumination-guided Transformer. The ORF module enhances the brightness and contrast of the input image by generating brightness-enhanced images, providing high-quality input for the IGT. The IGT module, with its multi-branch Transformer structure, processes the highlights, shadows, and foggy areas in the image separately, further enhancing the details and small target information. Additionally, the adaptive parameter estimation module dynamically adjusts the dehazing strategy, improving the detection robustness of Retinexformer across different foggy scenarios[12].

2) Under dense fog, identifying vehicles and pedestrians presents distinct difficulties, including the degradation of spatial details and alterations in color representation. By replacing the neck structure of YOLOv8n with C2f_ScConv[13], we can better extract features of vehicles and pedestrians from the foggy environment while reducing computational load, thereby improving detection accuracy and performance.

3) The attention mechanism is implemented through the use of channel and spatial attention modules to enhance the features of targets in foggy environments. To improve the information representation capability of the feature maps and suppress the interference from foggy environment information, this study introduces dual attention by drawing on the CBAM[14] (Convolutional Block Attention Module).

3.1. Improved Retinex Theory for Foggy Multi-object Detection

To address the issue of fog that diminishes image contrast, visibility, and overall sharpness, the Retinex theory has been introduced. This approach simulates the human visual system's capability to distinguish between reflectance (the inherent characteristics of a scene) and illumination (the fluctuating lighting conditions in the surrounding environment). By applying Retinex theory to foggy images, one can enhance the visibility of objects by improving the contrast and visibility of important features while suppressing the foggy effects[15].The following are detailed improvements based on the Retinex theory:

(1) Adaptive Parameter Estimation

In traditional Retinex theory, fixed parameters are used for separating reflectance and illumination. However, this approach does not perform well under varying fog intensity, lighting conditions, or different object types. To overcome this, an adaptive parameter estimation module is introduced, which dynamically adjusts the dehazing and image enhancement strategies.

(2) Illumination Guided Attention Blocks

To better handle varying light conditions in foggy environments, the model incorporates Illumination Guided

Attention Blocks[16]. These attention blocks help guide the network's focus towards regions with significant illumination variations, effectively capturing features that are crucial for object recognition. By using the illumination information as a guide, the model can enhance the recovery of details in shadowed or overexposed regions, improving the overall detection performance.

(3) Multi-Branch Transformer Integration

A multi-branch Transformer architecture is employed to further enhance feature extraction. Each branch of the Transformer focuses on different aspects of the image, such as spatial information, texture, and color contrast[17]. By applying multiple branches in parallel, the model is capable of processing different scales and contextual information simultaneously, making it more robust to occlusion, overlapping objects, and varying fog intensity.

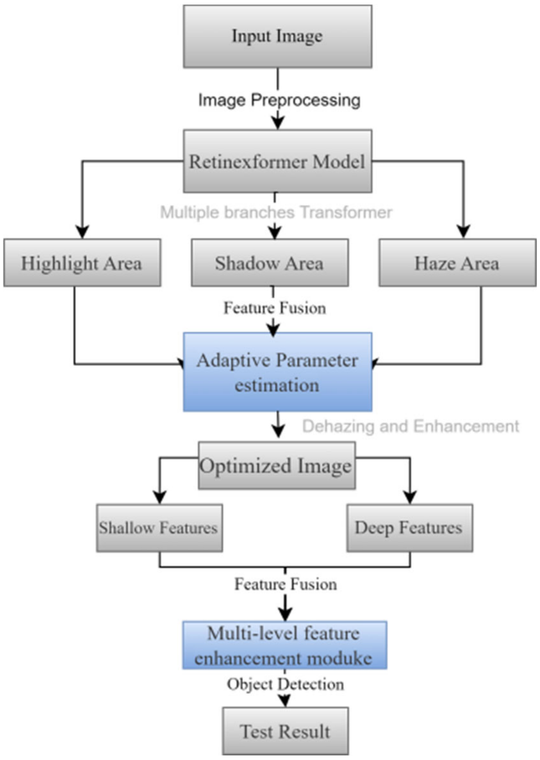

Figure 2. Structure of the Retinexformer improvement module

3.2. Design of the C2f-ScConv Module

In order to solve the color distortion caused by the scattering and attenuation of light in the atmosphere and the local information in the target, the YOLOv8 neck structure has been modified to incorporate C2f_ScConv. This modification enhances the model's ability to extract meaningful features from the foggy environment while also improving computational efficiency, ultimately leading to better detection accuracy and performance.

The C2f module ensures that attention is paid to multiple scales simultaneously, which is critical when dealing with foggy environments where objects may appear blurred or distorted at different levels[18].The ScConv module allows the neck to adapt dynamically to the conditions of the input image, ensuring that the features extracted from foggy areas are still meaningful and relevant for detection.With this modified neck structure, YOLOv8 is capable of detecting vehicles and pedestrians with greater precision, even in difficult foggy environments, while simultaneously minimizing the likelihood of overfitting and enhancing computational performance.

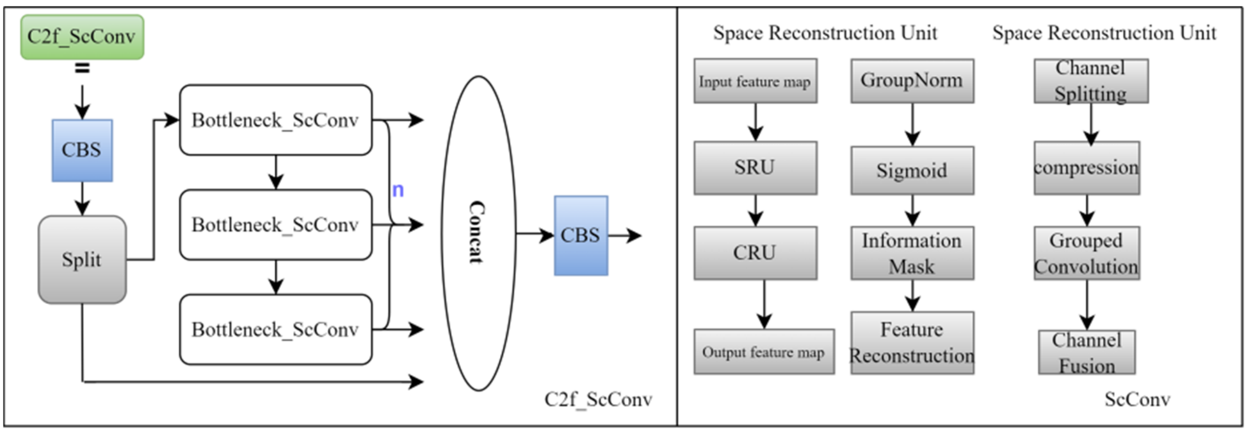

Figure 3. Structure of the C2f-ScConv Module

3.3. CBAM Attention mechanism

To improve the information representation capability of the feature maps and suppress the interference from foggy environment information, this study introduces dual attention by drawing on the CBAM module. The attention mechanism, consisting of both channel and spatial attention modules, enhances target features effectively. Following a structure to Retinexformer[19], the CBAM module is applied after the convolution layers to enhance small target features, enhancing the model's capacity to identify objects in foggy conditions.

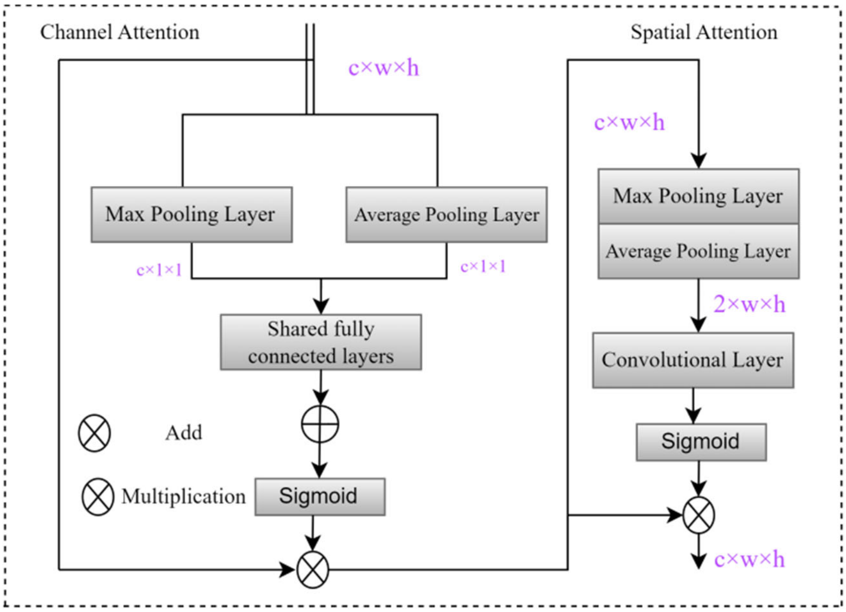

Figure 4. Structure of the CBAM Module

4. Experiments

The RTTS dataset employed in this research is based on the RESIDE[20] dataset. which is designed for foggy image processing and computer vision research. It contains 4,322 real foggy images, serving as the training set for the project. An additional 100 real scene images are used as the validation set, covering a variety of foggy conditions, which helps to train and test the model's performance under different foggy environments.

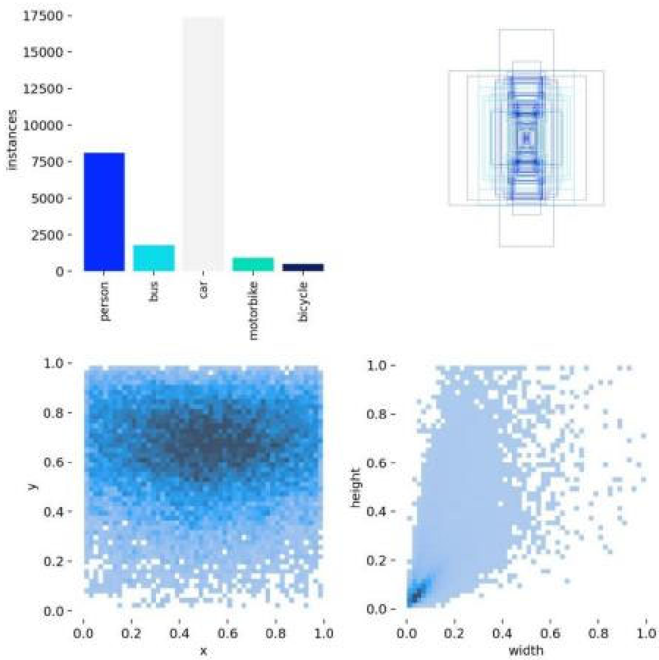

Figure 5. DataSet instance photo

4.1. Evaluation Metrics

In this study, the performance of the model is evaluated using the widely recognized metric of mean Average Precision (mAP)[21], which is commonly employed in object detection tasks. Additionally, a comprehensive comparison is made across several key performance indicators, including precision, recall, F1-score, and model complexity, particularly focusing on the number of parameters. mAP is computed by assessing the average precision at varying confidence thresholds, offering a robust evaluation of the model's ability to detect objects at different levels of certainty. Specifically, it calculates the precision-recall curve for each class, integrating over these curves to derive the mean value.

4.2. Comparative Experiments

In the comparative experiments, the performance of five established object detection algorithms, along with the newly proposed Fog-Mod-YOLOv8, was thoroughly evaluated. The algorithms included in the study are RTDETR, EfficientVit, YOLOv5n, YOLOv6n, YOLOv8n, and Fog-Mod-YOLOv8. The evaluation was conducted using a range of key performance metrics, such as mAP50, precision, recall, and the number of parameters. The results of these experiments are presented in Table 1, offering a comprehensive comparison of the detection accuracy, efficiency, and model complexity across the different methods[22].

Table 1. Comparative Experimental Results on the Custom Dataset.

Method	Dataset				
	mAP50	Precision	Recall	FLOPs(G)	Parameters
RTDETR	0.694	0.679	0.691	108.3	3.28×107
EfficientVit	0.651	0.718	0.706	2.391	3.83×106
YOLOv5n	0.687	0.647	0.663	12.0	7.04×106
YOLOv6n	0.648	0.648	0.623	10.9	4.24×106
YOLOv8n	0.713	0.726	0.718	8.2	3.01×106
Fog-Mod-YOLOV8n	0.744	0.736	0.736	14.8	8.64×106

Fog-Mod-YOLOv8 achieves an mAP50 of 0.744, which is approximately 4.3% higher than the baseline YOLOv8n (0.713).Compared to RTDETR (0.694), it shows a 7.2% improvement, demonstrating superior detection accuracy.These results highlight that Fog-Mod-YOLOv8 outperforms other models in terms of overall detection accuracy under foggy conditions.The Precision of Fog-Mod-YOLOv8 reaches 0.736, representing an improvement of 1.4% over YOLOv8n (0.726) and a significant increase of 8.4% compared to RTDETR (0.679).While the Precision of Fog-Mod-YOLOv8n is close to EfficientVit (0.718), it surpasses EfficientVit in both mAP50 and Recall metrics.Fog-Mod-YOLOv8 achieves a Recall of 0.736, which is 2.5% higher than YOLOv8n (0.718) and 6.5% higher than RTDETR (0.691).6.5% higher than RTDETR (0.691).This indicates that Fog-Mod-YOLOv8n has a significant advantage in ensuring comprehensive detection of targets in challenging environments.

Figure 6. Yolov8 Benchmark model data

Figure 7. Fog-Mod-Yolov8n improvement model data

4.3. Ablation Experiments

To evaluate the effectiveness of the algorithm, this section uses the following configurations: YOLOv8n, YOLOv8n_C2f, YOLOv8n_ScConv, YOLOv8n_C2f_ScConv, YOLOv8n_CBATM, and YOLOv8n_Fog_Mod. YOLOv8n_C2f_ScConv: By changing the neck structure of YOLOv8 to C2f_ScConv, we can better extract the characteristics of vehicles and pedestrians from the fog environment, while reducing the computational amount, and thus improve the detection accuracy and performance. It reduces the risk of overfitting and also enhances the ability of feature extraction. YOLOv8n-

CBAM: This configuration introduces the CBAM module after the YOLOv8 convolution layer to enhance the small target features, enhance the information expression ability of the feature graph, and inhibit the interference of environmental information in fog. In this comparative analysis, the performance of five well-established object detection algorithms, alongside the newly introduced Fog-Mod-YOLOv8, was systematically assessed. The algorithms evaluated in this study include RTDETR, EfficientVit, YOLOv5n, YOLOv6n, YOLOv8n, and Fog-Mod-YOLOv8. A variety of key evaluation metrics were employed, including mAP50, precision, recall, and model parameter count, to provide a comprehensive assessment. The results, which are summarized in Table 1, offer an in-depth comparison of detection accuracy, computational efficiency, and model complexity across the various approaches.

Table 2. Ablation Experiment Results

Method	C2f	ScConv	CBAM	mAP50	mAP50-90	Precision	Recall	Parameters
yolov8n	×	×	×	0.713	0.695	0.706	0.718	3.01×106
yolov8n_ C2f	√	×	×	0.728	0.711	0.727	0.726	3.02×106
yolov8n_ScConv	×	√	×	0.678	0.664	0.659	0.654	3.01×106
yolov8n_C2f_ScConv	√	√	×	0.736	0.715	0.737	0.726	4.24×106
yolov8n_CBAM	×	×	√	0.718	0.706	0.715	0.663	3.01×106
Yolov8n_Fog_Mod	√	√	√	0.744	0.724	0.736	0.736	8.64×106

As detailed in Table 2, replacing the original C2f module with the enhanced C2f_ScConv module yields improvements of 3.22% and 2.87% in mAP50 and mAP50-90, respectively, when compared to the baseline YOLOv8n model. Additionally, the integration of the CBAM attention mechanism further elevates the mAP50 and mAP50-90 scores by 0.7% and 1.58%, underscoring its contribution to refining the model's capacity to detect objects across diverse scales and to extract relevant features from cluttered or ambiguous contexts. When both the C2f_ScConv and CBAM modules are incorporated into the YOLOv8n_Fog_Mod configuration, the model achieves a notable increase of 4.35% in both mAP50 and mAP50-90. These improvements highlight the combined effect of these modules in addressing class imbalance issues between simpler and more challenging examples within the dataset, thereby enhancing overall detection accuracy.

4.4. Visualization Analysis

Figure 9 illustrates the visualization results of various detection methods, highlighting a substantial enhancement in the performance of the Fog-Mod-YOLOv8n model compared to the baseline YOLOv8 model. Specifically, the proposed Fog-Mod-YOLOv8 algorithm outperforms the baseline by achieving improvements of 4.35% and 0.417% in mAP50 and mAP50-90, respectively. Additionally, it demonstrates notable gains in accuracy and recall, with increases of 4.25% and 2.51%, respectively. Collectively, these results underscore the effectiveness of the augmented algorithm in accurately detecting vehicles and pedestrians in foggy environments, confirming its robustness and practical applicability in real-world scenarios[23].

Figure 9. Comparison of vehicle and pedestrian detection results

5. Conclusion

This study presents the Fog-Mod-YOLOv8n algorithm to effectively perform multi-object detection of vehicles and pedestrians under fog conditions. The model integrates the C2f_ScConv module, replacing the bottleneck in C2F, and adopts the CBATM two-channel attention mechanism to improve the feature extraction in fog conditions to enhance the small target features in the convolution layer. The experimental results show that Fog-Mod-YOLOv8n has strong performance in multi-object detection in fog environment and significant advantages in model efficiency and lightweight design

To enhance detection accuracy, efficiency, and generalization, high-quality datasets and an optimized network structure are essential. Future work will focus on curating high-quality datasets and refining the network architecture to improve both model accuracy and its lightweight design. This study not only introduces an innovative model for vehicle and pedestrian recognition in complex environments but also provides practical solutions for traffic monitoring applications, highlighting the potential of deep learning for recognition tasks in foggy conditions.

References

[1] Peng Y, Abdel-Aty M, Lee J, et al. Analysis of the impact of fog-related reduced visibility on traffic parameters[J]. Journal of transportation engineering, Part A: Systems, 2018, 144(2): 04017077.

[2] Han G, Wang Y, Liu J, et al. Low-light images enhancement and denoising network based on unsupervised learning multi-stream feature modeling[J]. Journal of Visual Communication and Image Representation, 2023, 96: 103932.

[3] Xing X, Hu V T, Metzen J H, et al. Retinex-diffusion: On controlling illumination conditions in diffusion models via retinex theory[J]. arXiv preprint arXiv:2407.20785, 2024.

[4] Zhang Y, Balestra G, Zhang K, et al. MultiTrans: Multi-branch transformer network for medical image segmentation[J]. Computer Methods and Programs in Biomedicine, 2024: 108280.

[5] Ding Y, Wu K. A Multi-Task Learning and Knowledge Selection Strategy for Environment-Induced Color-Distorted Image Restoration[J]. Applied Sciences, 2024, 14(5): 1836.

[6] Jocher, Glenn, Ayush Chaurasia, and Jing Qiu. "YOLO by Ultralytics." (2023): 17.

[7] Wang W, Wang Y. Underwater target detection system based on YOLO v4[C]//2021 2nd International Conference on Artificial Intelligence and Information Systems. 2021: 1-5.

[8] Jin Y, Tian X, Zhang Z, et al. C2F: An effective coarse-to-fine network for video summarization[J]. Image and Vision Computing, 2024, 144: 104962.

[9] Zhang G, Wang C, Xiao D. A novel daily behavior recognition model for cage-reared ducks by improving SPPF and C3 of YOLOv5s[J]. Computers and Electronics in Agriculture, 2024, 227: 109580.

[10] Bakirci M. Enhancing vehicle detection in intelligent transportation systems via autonomous UAV platform and YOLOv8 integration[J]. Applied Soft Computing, 2024, 164: 112015.

[11] Huang X, Wang X, Teng Q, et al. Degradation Type-Aware Image Restoration for Effective Object Detection in Adverse Weather[J]. Sensors, 2024, 24(19): 6330.

[12] Li L, Xu W, Gao Y, et al. Attention-oriented residual block for real-time low-light image enhancement in smart ports[J]. Computers and Electrical Engineering, 2024, 120: 109634.

[13] Wang J. Cable Lines Detection of Road Charging Piles Using Improved YOLOv8[C]//2024 6th International Conference on Communications, Information System and Computer Engineering (CISCE). IEEE, 2024: 969-972.

[14] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[15] Xu Y, Wen J, Fei L, et al. Review of video and image defogging algorithms and related studies on image restoration and enhancement[J]. Ieee Access, 2015, 4: 165-188.

[16] Ghari B, Tourani A, Shahbahrami A, et al. Pedestrian detection in low-light conditions: A comprehensive survey[J]. Image and Vision Computing, 2024: 105106.

[17] Qiu Y, Zhang K, Wang C, et al. Mb-taylorformer: Multi-branch efficient transformer expanded by taylor formula for image dehazing[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 12802-12813.

[18] Bian J, Ding S, Deng W, et al. A Fast Foggy Sea Ice Image Lightweight Dehazing Algorithm Using All-in-One Dehazing Network with Convolutional Block Attention Mechanism and Pyramid Pooling Module[J]. Available at SSRN 4976578.

[19] Cai Y, Bian H, Lin J, et al. Retinexformer: One-stage retinex-based transformer for low-light image enhancement [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 12504-12513.

[20] Li B, Ren W, Fu D, et al. Benchmarking single-image dehazing and beyond[J]. IEEE Transactions on Image Processing, 2018, 28(1): 492-505.

[21] Padilla R, Netto S L, Da Silva E A B. A survey on performance metrics for object-detection algorithms[C]//2020 international conference on systems, signals and image processing (IWSSIP). IEEE, 2020: 237-242.

[22] Khan H, Ullah I, Shabaz M, et al. Visionary vigilance: Optimized YOLOV8 for fallen person detection with large-scale benchmark dataset[J]. Image and Vision Computing, 2024, 149: 105195.

[23] Abdelaziz A M, Mahmoud F A. Visual and Inertial Sensors for Robust Autonomous Vehicle Navigation in Urban Environments[J]. AI, IoT and the Fourth Industrial Revolution Review, 2022, 12(12): 15-31.