

Article

Object Detection in Hazy Environments, Based on an All-in-One Dehazing Network and the YOLOv5 Algorithm

Aijuan Li ^{1,*}, Guangpeng Xu ¹, Wenpeng Yue ^{2,*}, Chuanyan Xu ¹, Chunpeng Gong ¹ and Jiaping Cao ¹

¹ School of Automotive Engineering, Shan Dong Jiaotong University, Jinan 250357, China; cyxu225@163.com (G.X.); xcy@sdu.edu.cn (C.X.); gongcp19@163.com (C.G.); c161229@163.com (J.C.)

² Operation Management Department, Sinotruk Jining Commercial Vehicle Co., Ltd., Jining 272073, China

* Correspondence: liaijuan@sdu.edu.cn (A.L.); yuewp@sinotruk.com (W.Y.); Tel.: +86-15053180992 (A.L.); +86-13791761137 (W.Y.)

Abstract: This study introduces an advanced algorithm for intelligent vehicle target detection in hazy conditions, aiming to bolster the environmental perception capabilities of autonomous vehicles. The proposed approach integrates a hybrid convolutional module (HDC) into an all-in-one dehazing network, AOD-Net, to expand the perceptual domain for image feature extraction and refine the clarity of dehazed images. To accelerate model convergence and enhance generalization, the loss function has been optimized. For practical deployment in intelligent vehicle systems, the ShuffleNetv2 lightweight network module is incorporated into the YOLOv5s network backbone, and the feature pyramid network (FPN) within the neck network has been refined. Additionally, the network employs a global shuffle convolution (GSconv) to balance accuracy with parameter count. To further focus on the target, a convolutional block attention module (CBAM) is introduced, which helps in reducing the network's parameter count without compromising accuracy. A comparative experiment was conducted, and the results indicated that our algorithm achieved an impressive mean average precision (mAP) of 76.8% at an intersection-over-union (IoU) threshold of 0.5 in hazy conditions, outperforming YOLOv5 by 7.4 percentage points.

Keywords: haze environment; object detection; YOLOv5s; AOD-Net



Citation: Li, A.; Xu, G.; Yue, W.; Xu, C.; Gong, C.; Cao, J. Object Detection in Hazy Environments, Based on an All-in-One Dehazing Network and the YOLOv5 Algorithm. *Electronics* **2024**, *13*, 1862. <https://doi.org/10.3390/electronics13101862>

Academic Editor: Felipe Jiménez

Received: 31 March 2024

Revised: 1 May 2024

Accepted: 4 May 2024

Published: 10 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The environmental perception system is paramount and indispensable for enabling autonomous driving capabilities in intelligent vehicles [1]. However, in conditions of reduced visibility such as haze, visual sensors encounter limitations imposed by both external environmental factors and their own inherent constraints. These limitations lead to the acquisition of ambiguous information regarding the vehicle's external environment, characterized by degraded image quality and diminished contrast. Collectively, these factors significantly impair the precision of target detection [2]. Therefore, to ensure safe navigation and mitigate the risk of traffic accidents, it is imperative that intelligent vehicles are equipped with the capability to accurately perceive their surroundings, even under the challenging conditions presented by haze.

Object detection in hazy environments is a multifaceted challenge that can be broadly divided into two critical stages: image defogging and subsequent target detection [3]. The domain of image defogging research involves two primary categories: model-based and model-free defogging algorithms [4]. Model-based approaches delve into the causal factors of image degradation, examine the atmospheric conditions at the time of image acquisition, and construct mathematical models to emulate this degradation. Prominent among these are the atmospheric scattering model [5,6] and the retinex theory [7,8]. Li et al. [9] introduced an end-to-end defogging network model that integrated atmospheric light values and atmospheric transmittance, derived from the scattering model, employing a neural network to learn and apply the function K(x) for restoring the fog-free image.

Zhang Qian et al. [10] enhanced the AOD-Net neural network defogging algorithm, addressing issues of inferior defogging quality and chromatic aberration in images. Their improvements included normalizing convolutional layers by position to accelerate convergence, utilizing multi-scale convolutional kernels for feature extraction, and adjusting network channel weights for more efficient feature utilization. In contrast, model-free algorithms bypass the complexities of model establishment and parameter estimation, effecting fog removal through direct learning of the mapping between fogged and fog-free imagery. Wang et al. [11] proposed an architecture to address inhomogeneous defogging that incorporated a haze imaging mechanism, establishing high-dimensional, nonlinear mapping. Bharath et al. [12] suggested that an enhanced conditional generative adversarial network (GAN) could be utilized for defogging, with a hybrid weighted loss function to refine output image quality.

Traditional image enhancement methods, including certain model-free defogging algorithms, are often constrained by high computational demands, leading to protracted processing times unsuitable for real-time defogging requirements in vehicular applications. In this study, we adopt the AOD-Net neural network defogging algorithm, a model-based approach that facilitates end-to-end defogging and is amenable to integration with target detection models, thereby harmonizing the processes of defogging and target detection.

In the realm of intelligent vehicle systems, computer vision techniques are pivotal for processing environmental data and facilitating the generation of target detection information. Deep learning-based target detection algorithms are conventionally divided into two primary categories: single-stage algorithms that approach detection as a regression problem, and two-stage algorithms that rely on region proposal (RP) [13]. The R-CNN (Region-based Convolutional Neural Network) series, proposed by Girshick et al. [14–16], epitomizes the two-stage approach. The algorithms bifurcate the detection process into the generation of regions of interest (ROI) for subsequent target classification and bounding box regression. In contrast, single-stage methods, such as the YOLO (You Only Look Once) series [17–19] and SSD (Single-Shot MultiBox Detector) [20], perform target detection as an end-to-end task in real time, eschewing the need for a separate region-proposal step. Among these, the YOLO series, exemplified by YOLOv5, is particularly adept at target detection on mobile platforms. By treating the detection task as a regression problem, YOLOv5 achieves a commendable balance between detection speed and accuracy, surpassing two-stage models in terms of processing velocity. This study leveraged YOLOv5 as the foundational target detection network, with the aim of enhancing its performance to ensure precise target identification in intelligent vehicles. The selection of YOLOv5s was motivated by its suitability for the real-time constraints and computational efficiency required for in-vehicle applications.

Ensuring the safe operation of intelligent vehicles necessitates keen consideration of the vehicle's operational state, the dynamic driving environment, and the precision of real-time target detection. The accuracy and speed of target detection are paramount for the safe navigation of autonomous vehicles, and thus, the enhancements to the YOLOv5s network presented in this study are designed to meet these stringent requirements.

In this research, we address the critical issue of target detection in hazy environments for intelligent vehicles. The improved AOD-Net model defogs the input foggy image and its output image is used as input to the improved YOLOv5s model for target detection. This study was motivated by the need for real-time, accurate perception in autonomous driving systems, which is often hindered by poor visibility conditions.

The AOD-Net model, while capable of meeting real-time requirements for image defogging, has limitations in feature extraction and does not account for human visual perception. Similarly, the YOLO algorithm, despite its strengths in real-time performance and accuracy, lacks specific optimization for defogging and vehicular applications. To overcome these limitations, we introduce the following contributions:

- (1) Enhanced Image Defogging Algorithm: We propose an improved AOD-Net algorithm that incorporates a hybrid convolutional module (HDC) to broaden the receptive field for feature extraction and refine the details of defogged images. Additionally, we refine the loss function to expedite model convergence and improve generalization. These enhancements significantly upgrade the defogging performance of the AOD-Net neural network.
- (2) Optimized YOLOv5s Detection Algorithm: An enhanced YOLOv5s target detection algorithm is introduced, featuring the ShuffleNetv2 lightweight network module to reduce complexity and improve efficiency. We also refine the feature pyramid network (FPN) and introduce global shuffle convolution (GSconv) to balance accuracy with the parameter count. The convolutional block attention module (CBAM) is incorporated to heighten target attention without compromising network accuracy, making it suitable for in-vehicle mobile devices.

The structure of this study is as follows: Section 2 establishes the foundation by detailing the AOD-Net dehazing algorithm's structure and principles, followed by our proposed improvements. Section 3 focuses on enhancing the YOLOv5s model through modifications to the backbone, neck network, and attention mechanism. Section 4 presents experimental validation of our proposed methods and compares them with mainstream algorithms. Finally, Section 5 concludes the study with a summary of findings and future research directions.

2. Image Defogging Algorithm and Model

In a hazy environment, the propagation of light is influenced by a large number of minute suspended particles floating in the air, causing scattering and refraction phenomena when the light passes through the suspended particles and preventing the visual sensor from gathering accurate and useful image data. The researchers employed image processing algorithms to analyze photographs acquired in a hazy environment, in accordance with the mechanism of haze creation, recovering crystal-clear, haze-free photographs in the process.

2.1. AOD-Net Network

2.1.1. Atmospheric Scattering Model

The degradation of acquired images is primarily attributed to two main factors: the absorption and scattering effects of suspended particles in the air, which impact the light reflected from objects, and the formation of a background light that surpasses the light emitted by the object surface during the propagation of light. Narasimhan [21] established a mathematical model for the above two factors to analyze the imaging process and the influencing factors in a foggy environment.

The atmospheric scattering model [5] developed from this physical model is formulated as follows:

$$L(x, y) = L_0(x, y)e^{-kd(x, y)} + A(1 - e^{-kd(x, y)}) \quad (1)$$

where $L(x, y)$ is the image with fog, $L_0(x, y)$ is the image without fog, $e^{-kd(x, y)}$ is the transmittance, k is the atmospheric scattering coefficient, $d(x, y)$ is the scene depth, and A is the atmospheric light coefficient.

The atmospheric scattering model is simplified according to Equation (1) as:

$$I(x) = J(x)t(x) + A(1 - t(x)) \quad (2)$$

where $I(x)$ is the fogged image acquired by the imaging device, $J(x)$ is the fog-free image, and $t(x)$ is the transmittance.

Based on the above mathematical model, $t(x)$ or $A(1 - t(x))$ is estimated from the haze image, and the estimated number of parameters is brought into the mathematical model equation to find the image $J(x)$ after defogging:

$$J(x) = \frac{1}{t(x)} I(x) - A \frac{1}{t(x)} + A \quad (3)$$

2.1.2. Principle of the AOD-Net Architecture

The AOD-Net defogging algorithm is a method based on estimating the parameters of atmospheric scattering. Unlike other algorithms that estimate the transmittance and atmospheric light coefficient separately, AOD-Net integrates the relationship between these two variables. It achieves end-to-end image defogging through integrating the two parameters into a single variable, denoted as $K(x)$, which is learned and trained through the algorithm. The transformed and integrated equation is as follows:

$$J(x) = K(x)I(x) - K(x) + b \quad (4)$$

The transformation leads to $K(x)$:

$$K(x) = \frac{\frac{1}{t(x)}(I(x) - A) + (A - b)}{I(x) - 1} \quad (5)$$

where b is the constant deviation.

The AOD-Net network structure comprises two modules, the K-estimating module and the defogging image generation module, as depicted in Figure 1 of the AOD-Net network model. The K-estimating module is responsible for obtaining the estimated parameters $K(x)$, and the AOD-Net convolutional layer structure is detailed in Table 1. The defogging image generation module utilizes the generated $K(x)$ parameters to produce the defogged image, denoted as $J(x)$.

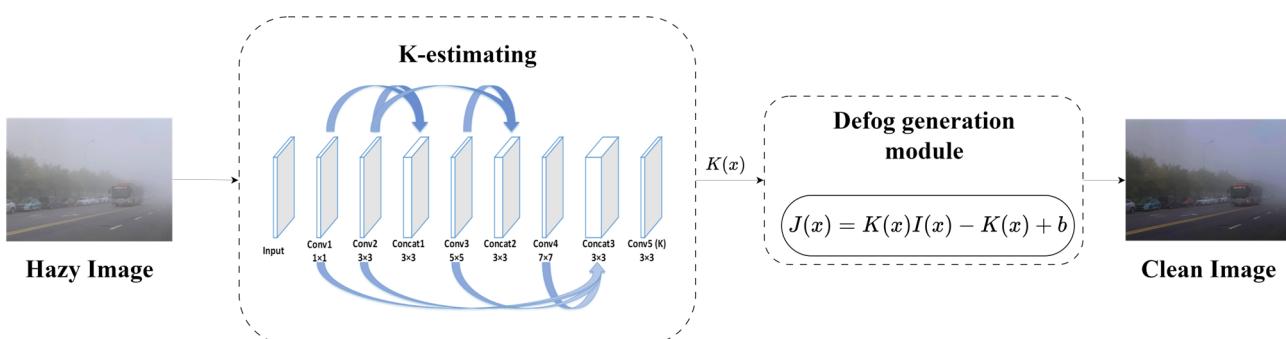


Figure 1. AOD-Net network model.

Table 1. AOD-Net convolutional layer structure.

| Convolutional Layer | 1 | 2 | 3 | 4 | 5 |
|---------------------------|--------------|--------------|--------------|--------------|--------------|
| convolution kernel | 1×1 | 3×3 | 5×5 | 7×7 | 3×3 |
| void rate | 1 | 1 | 1 | 1 | 1 |
| receptive field size | 1 | 3 | 7 | 13 | 15 |
| number of output channels | 3 | 3 | 3 | 3 | 3 |

The AOD-Net convolutional layer structure presented in Table 1 demonstrates the simplicity and suitability of the AOD-Net neural network topology for mobile devices. The application of the AOD-Net neural network algorithm for defogging purposes leads to a significant reduction in image fog, resulting in improved visibility of vehicle information

within the image. These enhancements provide a valuable advantage for subsequent target detection tasks, as illustrated in Figure 2 showcasing the AOD-Net defogging comparison results.

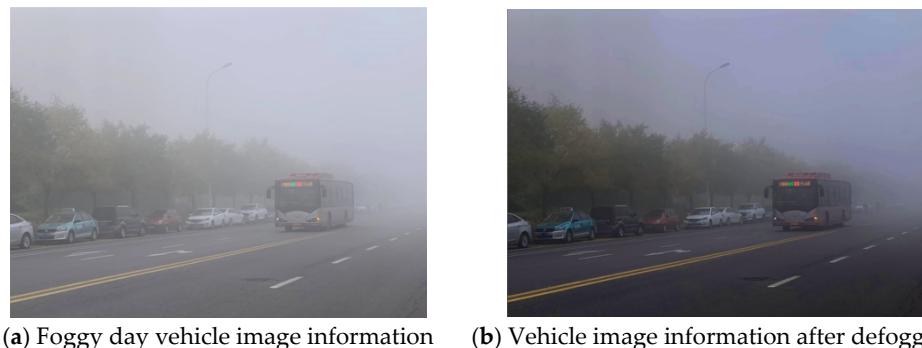


Figure 2. AOD-Net defogging comparison results.

AOD-Net uses mean squared error (MSE) as the loss function and refers to it as L_2 . The L_2 loss function facilitates rapid convergence of the model through calculating the mean of the squared difference between the predicted and true values in the neural network model, and a model with this loss function is expressed as follows:

$$L^{l_2} = \frac{1}{N} \sum_{p \in P} (x(p) - y(p))^2 \quad (6)$$

where $x(p)$ denotes the predicted value, $y(p)$ denotes the actual true value, N denotes the number of pixels in the region, P denotes the image block, and p denotes the pixel.

2.2. Improved AOD-Net's Defogging Algorithm

2.2.1. Improvement of Convolution Module

To address the issue of large-scale feature information being ignored and the resulting inaccurate $K(x)$ values in the AOD-Net network module, this paper introduces the hybrid dilated convolution (HDC) module [22]. The AOD-Net network structure is thereby enhanced, as depicted in Figure 3, to ensure more accurate and effective defogging.

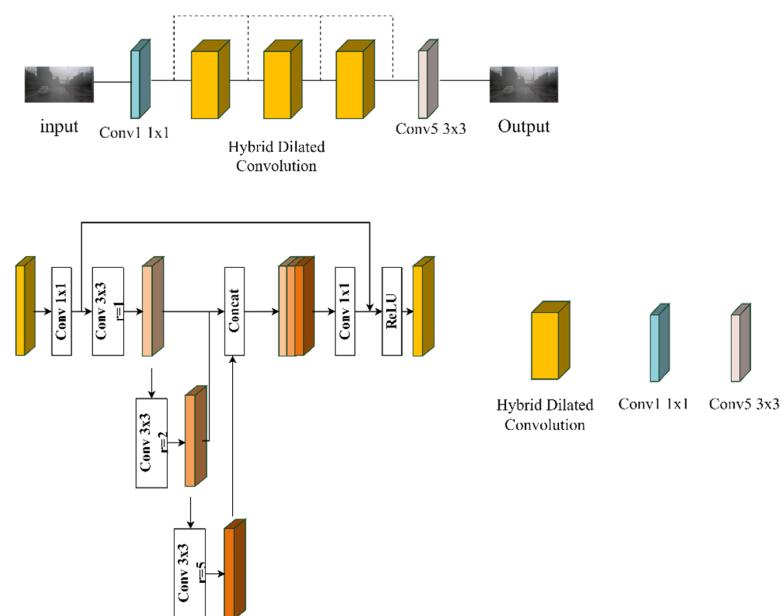


Figure 3. Improved AOD-Net network structure.

This paper presents the introduction of the HDC module in the study of the AOD-Net neural network-based defogging algorithm. The aim is to expand the perceptual field of image features and enhance the extraction capacity of depth image feature information, while preserving resolution and detailed feature information.

2.2.2. Selection of the Loss Function

The loss function employed in the AOD-Net defogging algorithm does not consider human visual perception and sensitivity to regional color changes, luminance in texture-free regions, etc. In order to address this limitation, this paper utilizes a hybrid loss function combining $MS-SSIM + L_2$ to serve as the loss function for the AOD-Net defogging algorithm.

The L_2 loss function cannot express the human visual perception system's intuition of the picture, whereas the $SSIM$ loss function starts from three key features of the image: brightness, contrast, and structure to mimic the human perception system; its function expression is:

$$S(x, y) = f(l(x, y), c(x, y), s(x, y)) \quad (7)$$

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (8)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (9)$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (10)$$

$$SSIM(P) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} = l(p) \cdot cs(p) \quad (11)$$

where $l(x, y)$ is the image brightness similarity, $c(x, y)$ is the image contrast, and $s(x, y)$ is the image structure comparison function.

$MS-SSIM$ is based on the $SSIM$ loss function, introducing multiple scales, taking into account the subjective factor of image resolution; its function is defined as:

$$MS-SSIM(p) = l_M^\alpha(p) \cdot \prod_{j=1}^M cS_j^{\beta_j}(p) \quad (12)$$

$$L^{MS-SSIM}(P) = 1 - MS-SSIM(\tilde{p}) \quad (13)$$

where c denotes contrast, s denotes structural comparison, and \tilde{p} is the central pixel in the image block P . Although this function effectively preserves image edge and detail information, it tends to lead to changes in image brightness and color deviations.

Considering the drawbacks of the above loss function, the $MS-SSIM + L_2$ loss function can effectively keep the brightness and color of the image unchanged [23]. Its functional expression is:

$$L^{Mix} = \alpha \cdot L^{MS-SSIM} + (1 - \alpha) \cdot G_{\alpha_G^M} \cdot L^{l_2} \quad (14)$$

where G is the Gaussian distribution parameter.

Through analysis and comparison of the image processing challenges encountered with the L_2 loss function utilized by AOD-Net, it is evident that a hybrid loss function is required to address the issues related to human visual perception and the enhancement of brightness and color changes in the processed images. Therefore, in this paper, the improved defogging algorithm model incorporates the hybrid loss function to mitigate these challenges.

3. Improved YOLOv5s Target Detection Model

3.1. YOLOv5s Network

The YOLOv5s target detection model is specifically designed to be lightweight, considering the detection environment and network structure requirements for smart vehicle operations. YOLOv5 is a popular object detection model, and its architecture can be divided into three main parts: the backbone, the neck, and the head. The backbone network captures low-level and mid-level features such as edges, textures, and shapes. As the network deepens, these features become more abstract, forming high-level semantic information. The neck usually consists of a multi-scale feature pyramid network (FPN) and a path aggregation network (PAN). The FPN combines features of different scales through upsampling and downsampling operations to capture targets of various sizes. The PAN enhances feature propagation through top-down and bottom-up paths, allowing the model to better understand the context of targets. The head network performs two main tasks: one is to predict the category of the target, and the other is to predict the bounding boxes of the targets.

The improved YOLOv5s network structure, as depicted in Figure 4, uses a lightweight network structure ShuffleNetv2 Block, a BiFPN module that increases the fusion of feature information, a lightweight GSConv module, and a CBAM attention mechanism that improves the attention to target information, enabling better applicability of the YOLOv5s target detection model for in-vehicle end device platforms.

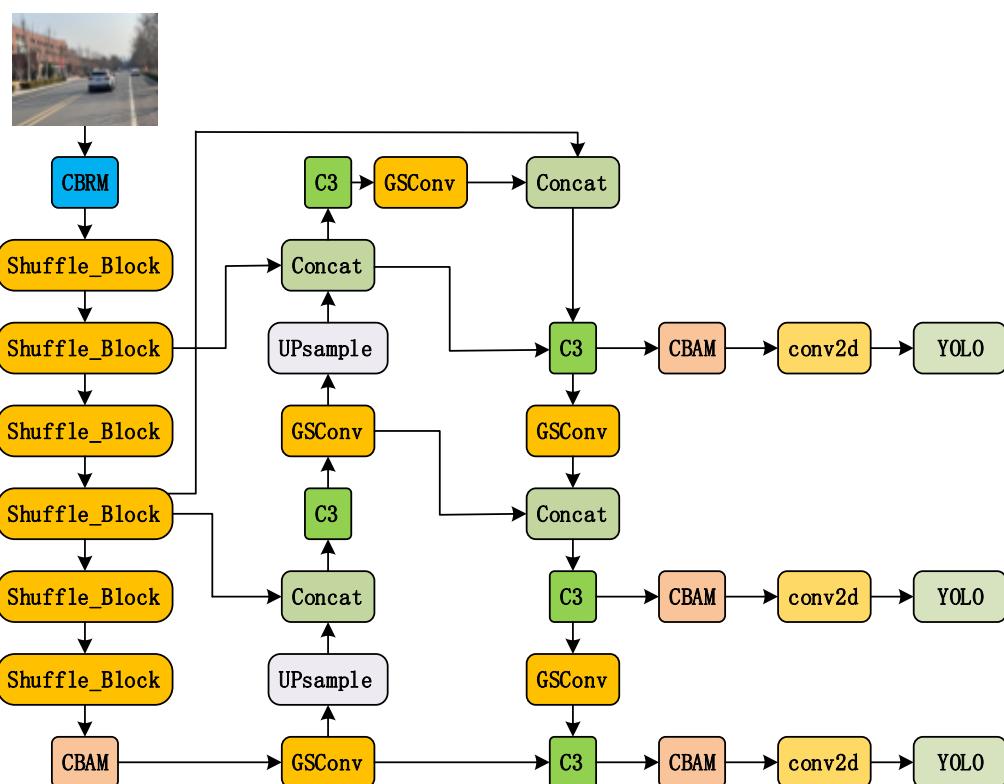


Figure 4. Improved YOLOv5s network structure.

3.2. Lightweighting of Backbone Network

In this study, the ShuffleNetv2 module was adopted due to its real-time image processing capabilities and lower performance requirements on intelligent vehicle application platforms. As shown in Figure 5, this module combines deep separable convolution and group convolution in its network structure, leading to a substantial reduction in network parameters compared with other residual networks. To optimize the model's size and characteristics, the ShuffleNetv2 Block was employed to restructure the network's backbone.

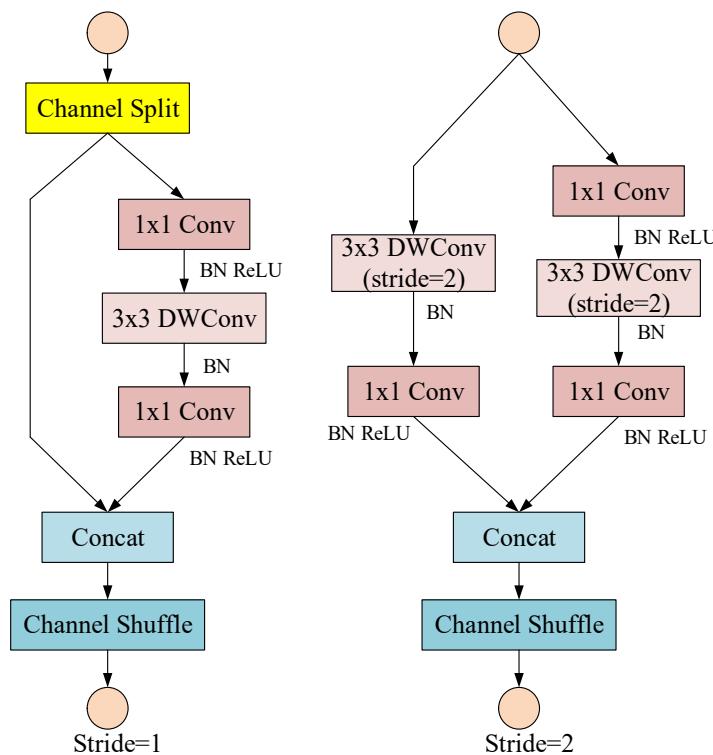


Figure 5. ShuffleNetv2 module.

3.3. Improvements to Neck

The bidirectional feature pyramid network (BiFPN) incorporates a bidirectional cross-scale connection mechanism. It introduces weight values to assign different weight coefficients to different features, enabling effective utilization of information from each feature layer and enhancing feature fusion across different layers. This partitioning of feature layers facilitates improved connectivity and fusion between layers. The fusion mechanism is depicted in Figure 6 for visual comparison.

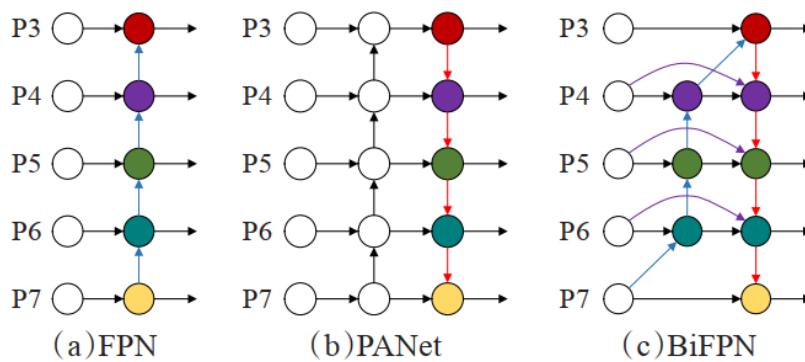


Figure 6. Fusion structure comparison diagram.

GSConv is a compact network module that aims to balance model accuracy and network simplicity [24]. The network structure, depicted in Figure 7, maintains model accuracy while simplifying the network architecture. This structure leverages a combination of depth-separable convolution and standard convolution to effectively utilize both types of convolutional features. By doing so, the model weight is reduced without compromising accuracy. This approach addresses the issue of channel information loss observed in the CSP convolution module.

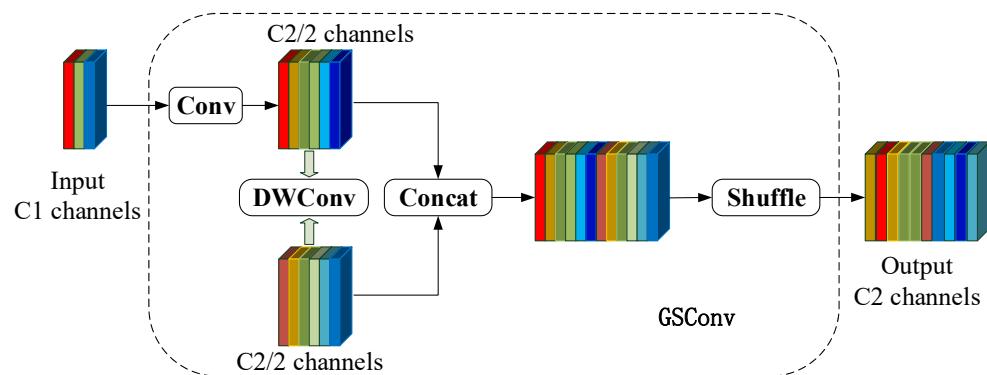


Figure 7. Structure of GSConv.

3.4. Attention Mechanism

The attention mechanism (AM) [25] relies on a technique that mimics the selective attention mechanism observed in human visual perception, enabling the network model to prioritize relevant target categories during training and learning. In this research, a convolutional block attention module (CBAM) [26] has been introduced to address the specific requirements of the target detection algorithm model for intelligent vehicles in complex driving environments.

The CBAM attention mechanism module consists of two components: the channel attention module (CAM) and the spatial attention module (SAM). Two attention techniques in the input image feature layer help the network structure model pay better attention to the major feature image information. Figure 8 depicts the structure of the CBAM module.

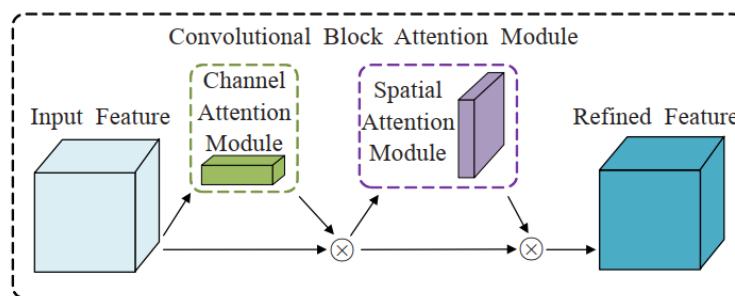


Figure 8. Structure of CBAM.

4. Experimental Results and Analysis

4.1. Improved AOD-Net Image Defogging Comparison Experimental Analysis

4.1.1. Datasets and Experimental Platforms

In this study paper, we used a Windows 10 operating system, Intel(R) Core(TM) i7-10700F CPU@2.90 GHz processor, 24 GB RAM and other hardware facilities to build the defogging experiment testbed, using Pycharm (Version 2023.1.2) as the compilation environment. The training parameters of the defogging algorithm were set as follows: epochs 20, batch_size 16, and learning rate 0.0001.

4.1.2. Defogging Quality Evaluation Index

Image defogging quality evaluation reflects the quality of the generated images after the fogged images are subjected to the defogging algorithm. Since the synthetic fogged image dataset was converted from clear images by the algorithm, the clear fog-free images were used as a reference basis to evaluate the defogging quality by peak signal-to-noise ratio and structural similarity.

(1) Peak Signal-to-Noise Ratio

Peak Signal-to-Noise Ratio (PSNR) is a commonly used error-sensitive image quality evaluation index obtained through comparing the error values of the corresponding pixel points between two images. In this study, higher *PSNR* values indicated reduced image distortion and superior defogging performance. The *PSNR* formula is given by the following expression:

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_L^2}{MSE} \right) \quad (15)$$

where L denotes the maximum value of image pixels.

(2) Structural similarity

Structural similarity was assessed to measure the similarity between the fogged and defogged images based on three key features: image brightness, contrast, and structure. A higher SSIM value indicated less distortion and higher quality in the defogged image. The equations are shown in Equations (7)–(11).

4.1.3. Analysis of Defogging Results

In this study, we compared the performance of the enhanced AOD-Net neural network defogging algorithm with the widely used DCP, MbE [27], and AOD-Net defogging algorithms. The SOTS outdoor synthetic foggy day environment dataset provided the foggy day data photos, and Figure 9 compares the defogging results in the synthetic foggy day environment.

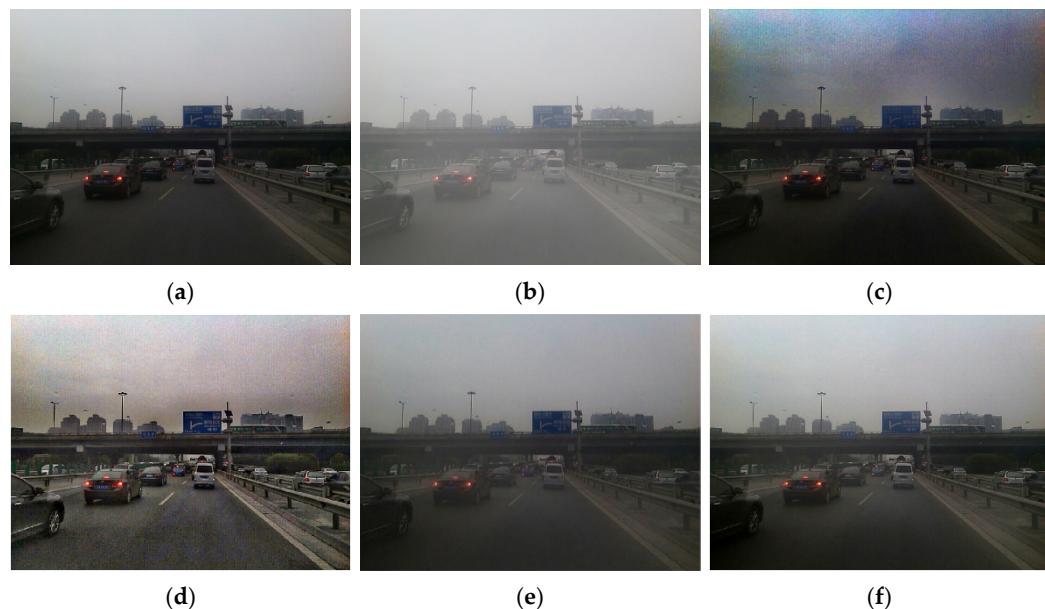


Figure 9. Comparison of the defogging effect on synthetic fog: (a) original fog-free image; (b) synthetic foggy images; (c) DCP algorithm for defogging images; (d) MbE algorithm for defogging images; (e) AOD-Net algorithm for defogging images; (f) improved AOD-Net defogged images.

Figure 9a illustrates the original image without any fog, while Figure 9b represents the image with artificially added fog. It can be observed that Figure 9b demonstrates increased brightness, reduced contrast, and obscured details due to the presence of fog. The defogged image obtained using the DCP method is displayed in Figure 9c, revealing a decrease in brightness, blurring of the background sky, and significant loss of color information. The image in Figure 9d is one that the MbE algorithm successfully defogged. As shown in Figure 9e, the image processed by the AOD-Net algorithm exhibited a slight reduction in brightness and a mild fog effect; notably, the fogged image produced by this method

demonstrated superior results and maintained a closer resemblance to the original image, as depicted in Figure 9f. To objectively assess and analyze the quality of the synthetic fog map defogging, two evaluation indices, namely, PSNR and SSIM, were utilized. Table 2 compares the two evaluation indexes for the defogging effect.

Table 2. Comparison of evaluation indexes of defogging effect.

| Evaluation Indicators | DCP | MbE | AOD-Net | Improved AOD-Net |
|-----------------------|-------|-------|---------|------------------|
| PSNR | 15.22 | 18.44 | 20.28 | 21.02 |
| SSIM | 0.79 | 0.67 | 0.82 | 0.85 |

The proposed method achieved a PSNR value of 21.02, outperforming the AOD-Net defogging algorithm by 0.74 in PSNR, the MbE algorithm by 2.58, and the DCP algorithm by 5.8. In terms of SSIM, this method achieved a value of 0.85, surpassing the AOD-Net defogging algorithm by 0.03 in SSIM, the MbE algorithm by 0.18, and the DCP algorithm by 0.06. These results demonstrate the enhanced advantages of the upgraded AOD-Net defogging algorithm.

To evaluate the effectiveness of the improved AOD-Net defogging algorithm in the real foggy environment, three road driving scenarios of light fog, medium fog, and dense fog were chosen from the RTTS dataset for verification analysis. The real scenario defogging effect comparison chart is shown in Figure 10.

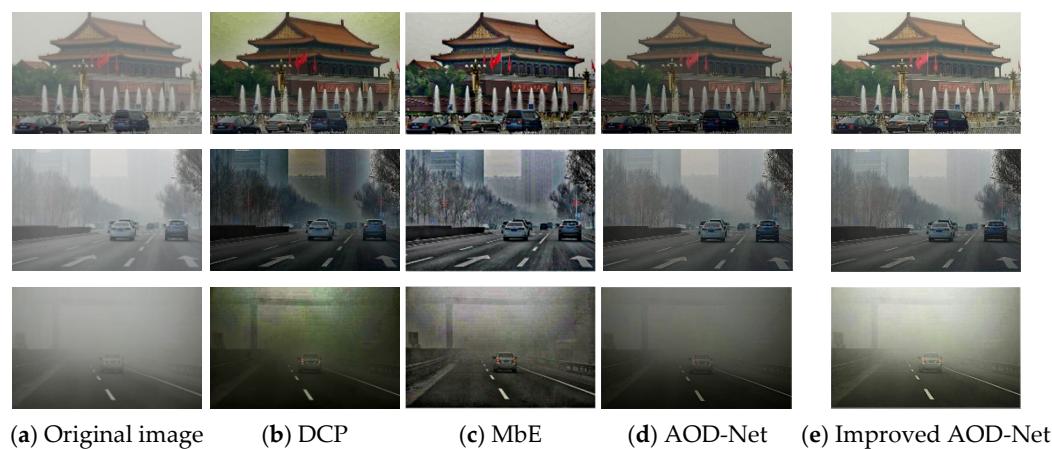


Figure 10. Comparison of the fogging effect on the real scene.

Figure 10b depicts the defogging graph of the DCP algorithm. It can be observed that the algorithm performed well in defogged light fog environments, but the sky background information was significantly distorted, and images in medium and dense fog environments suffered from brightness loss, loss of information detail, and low contrast. The defogging graph of the MbE algorithm is shown in Figure 10c. The algorithm demonstrated a satisfactory defogging impact in light fog conditions, boosting the contrast and visual appearance of the image information. However, in medium and heavy fog environments, the algorithm exhibited certain limitations; the background was badly blurred and distant information seriously lost. Figure 10d depicts the AOD-Net algorithm's fog-removal map. The algorithm effectively enhanced the color information in the image after fog removal, resulting in richer color representation. However, it was observed that the image appeared darker after the fog removal process, which may pose challenges for subsequent tasks such as target detection. Figure 10e depicts the improved AOD-Net algorithm's fog-removal map, which successfully improved the image color information while maintaining the target detection ability. The enhanced AOD-Net defogging map, which successfully eliminated the fog from the picture while preserving the color information of the original image, is shown in Figure 10e. It also enhanced the color brightness of the defogged image.

The comparison experiments demonstrated that the enhanced AOD-Net neural network defogging algorithm beat the competing defogging algorithms in terms of enhancing image contrast, color recovery, and brightness.

4.2. Improved YOLOv5s Target Detection Algorithm

4.2.1. Datasets and Experimental Platforms

A total of 18,000 photos with five different types of detection targets—pedestrians, buses, vehicles, bicycles, and motorcycles—were chosen from the Pascal VOC dataset and the KITTI dataset and were used as the dataset to test the performance of the upgraded YOLOv5s target detection algorithm. Table 3 lists the parameter settings for the model training platform, and Table 4 lists the parameters for the network model.

Table 3. Model training platform configuration parameters.

| Parameter | Configuration |
|------------------|--|
| CPU | Intel(R) Xeon(R) Platinum 8358P CPU @ 2.60 GHz |
| RAM | 40 |
| GPU | RTX 3080 (10 GB) |
| Operating System | Ubuntu18.04 |

Table 4. Network model training parameters.

| Parameter Type | Configuration |
|----------------|---------------|
| Epoch | 100 |
| Batch size | 32 |
| Learning rate | 0.0001 |
| Image size | 640 × 640 |

4.2.2. Analysis of Test Results

Using the training platform described above and setting the training parameters of the model, the improved YOLOv5s target detection network model was trained. The improved YOLOv5s confusion matrix is shown in Figure 11, and the PR curve of the improved YOLOv5s is shown in Figure 12.

The enhanced mAP@0.5 reached 87.3%, as shown by the training model data, and had decent recognition accuracy for larger targets including vehicles, bicycles, and pedestrians.

Table 5 presents a comprehensive analysis of the comparative experimental study, highlighting the advantages of the network model proposed in this paper over the baseline YOLOv5s model. The proposed model demonstrates a significant reduction in complexity, with a 41.7% decrease in the number of parameters, leading to a model size that is merely 51.5% of the original. Despite this reduction, the detection accuracy was only marginally lower, by 3.9%, suggesting that the proposed method offers enhanced adaptability and mobility, making it highly suitable for practical vehicular applications.

Compared with the other popular lightweight network models, the enhanced YOLOv5s algorithm stands out in terms of parameter efficiency and detection accuracy. For instance, the model size is one-fifth that of the YOLOv3-tiny, yet it achieved a 10.4% higher detection accuracy. Against the YOLOv4-tiny, the model exhibited a 5.9% higher accuracy, despite having fewer parameters and a smaller model size. Furthermore, the proposed model surpassed YOLOv5n, with a 9.4% higher mean average precision (mAP). In comparison with YOLOv7-Tiny and YOLOv8s, introduced in 2023, the method presented here is superior in terms of parameter count, GFLOPS, and model size.

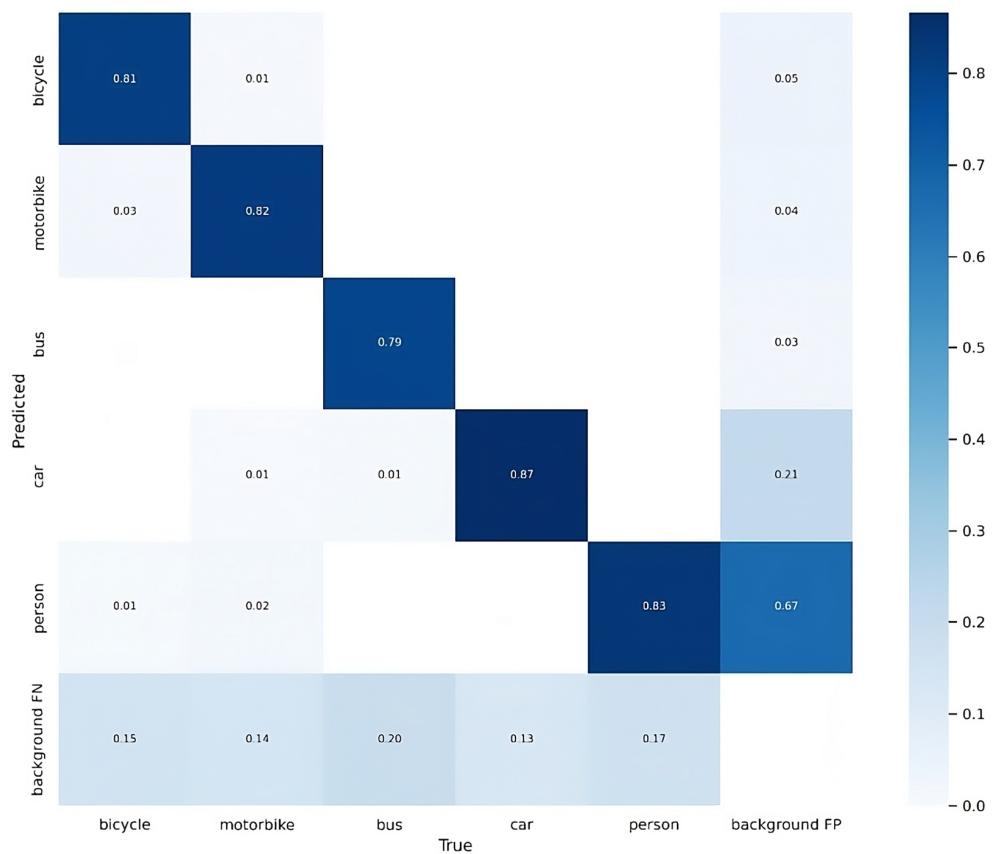


Figure 11. Confusion matrix of the improved YOLOv5s.

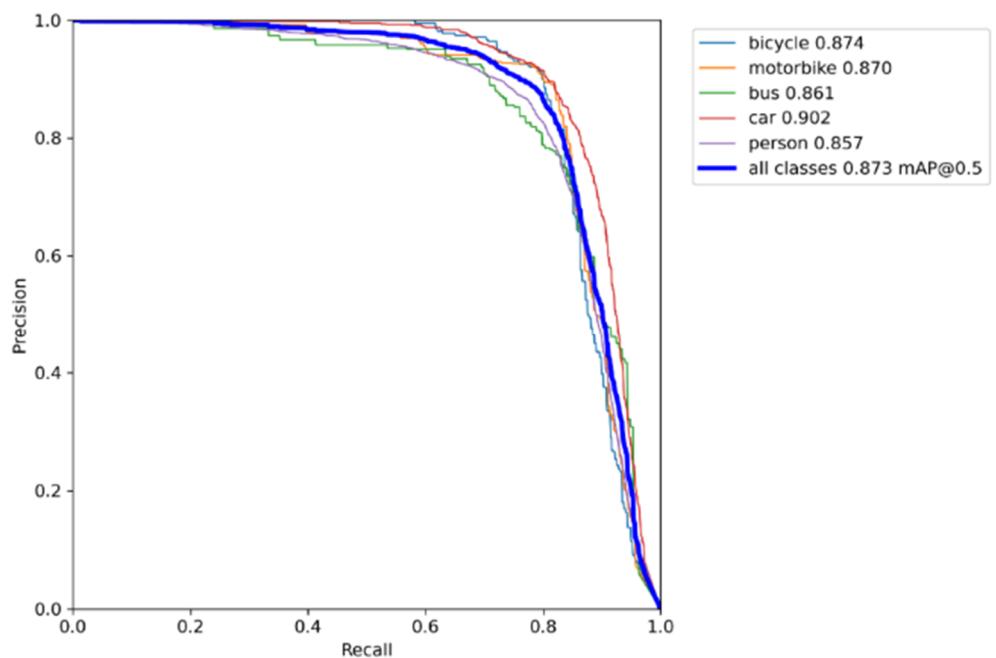


Figure 12. PR curve of the improved YOLOv5s.

Table 5. Comparison of experimental analysis data. ‘Ours’ denotes YOLOv5s + ShuffleNetv2 + BiFPN + GSConv + CBAM.

| Models | Parameters (M) | GFLOPS | Model Size (Mb) | mAP@0.5(%) | FPS |
|-----------------------|----------------|--------|-----------------|------------|------|
| YOLOv3-tiny | 8.7 | 5.6 | 33.4 | 76.9 | 94.3 |
| YOLOv4-tiny | 6.1 | 6.9 | 23.1 | 81.4 | 97.2 |
| YOLOv5s | 7.2 | 16.5 | 14.2 | 91.2 | 87.6 |
| YOLOv5n | 1.8 | 4.2 | 3.7 | 77.7 | 86.3 |
| YOLOv7-tiny | 6.1 | 13.1 | 12.6 | 79.8 | 98.1 |
| YOLOv8s | 11.2 | 28.8 | 13.8 | 83.4 | 90.0 |
| YOLOv5s-MobileNet | 2.8 | 5.6 | 7.4 | 80.2 | 82.1 |
| YOLOv5s-ShuffleNet | 3.8 | 8.1 | 7.6 | 81.4 | 82.7 |
| v5s-ShuffleNet-GSConv | 3.4 | 7.5 | 6.8 | 80.7 | 81.2 |
| v5s-v2-GSConv + BiFPN | 3.9 | 8.1 | 7.1 | 83.1 | 80.7 |
| Ours | 4.5 | 9.4 | 7.3 | 87.3 | 80.4 |

The detection accuracy of our method was notably 7.1% higher than the YOLOv5s-MobilNet network and 5.9% higher than the YOLOv5s-ShuffleNet, without compromising on model size. The comparative analysis, as depicted in Table 5, illustrates the incremental improvements made to the YOLOv5s target detection algorithm. The introduction of GSConv module resulted in a 0.4% reduction in network parameters and a 0.6% decrease in GFLOPS. The incorporation of the BiFPN module increased GFLOPS by 0.8% and boosted model accuracy by 2.4%. After the completion of these enhancements and optimizations, the accuracy of the proposed approach reached an impressive 87.3%, demonstrating a substantial impact on target detection capabilities.

Three different types of scene photos were chosen for evaluation of the detection and recognition achieved with YOLOv5s and the approach in this paper, in order to test the effectiveness of the method. The target detection pairings are displayed in Figure 13.

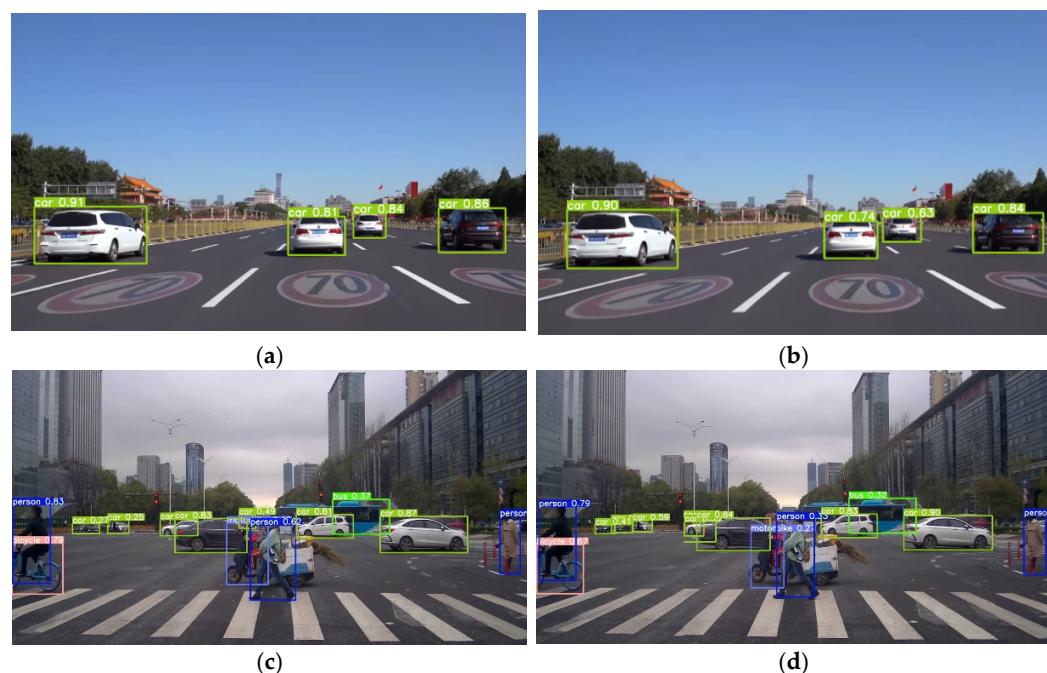


Figure 13. Comparison of target detection effects: (a) YOLOv5s vehicle detection effect on straight road; (b) improved YOLOv5s vehicle detection effect on straight road; (c) YOLOv5s vehicle detection effect at intersection; (d) improved YOLOv5s vehicle detection at intersection.

The comparison of target detection illustrated in Figure 13 shows that the upgraded YOLOv5s can detect and identify the obstacle target information in front of the vehicle and complete the target detection task in both straight road and complex intersection environments.

4.3. Analysis of Fog Target Detection Results

The RTTS dataset was chosen as the test set to examine the performance of the target detection algorithm in a foggy environment, which was necessary to assess the efficacy of the method in this work. Figure 14 displays the comparison of target detection accuracy in a foggy environment, and Table 6 describes the accuracy comparison of the detection models in a foggy environment.

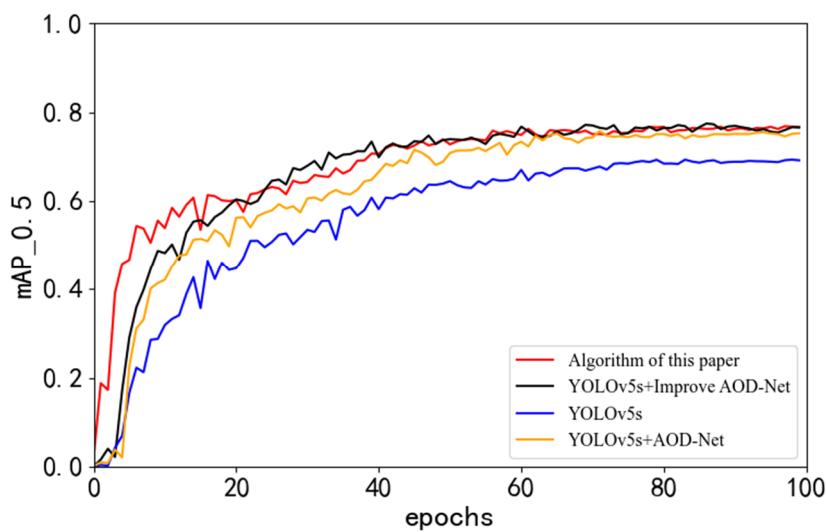


Figure 14. Comparison of accuracy of target detection algorithms in foggy environment.

Table 6. Comparison of accuracy of detection models in a foggy environment.

| Algorithm | Bicycle | Car | Motorbike | Person | mAP@0.5(%) |
|----------------------------|---------|------|-----------|--------|------------|
| YOLOv5s | 39.3 | 82.9 | 73.7 | 81.5 | 69.4 |
| AOD-Net + YOLOv5s | 52.0 | 86.4 | 78.7 | 85.0 | 75.5 |
| Improved AOD-Net + YOLOv5s | 56.4 | 84.7 | 79.8 | 86.2 | 77.4 |
| Algorithm in this paper | 59.0 | 86.1 | 79.7 | 85.0 | 76.8 |

We can observe that the algorithm proposed in this paper exhibited improved target detection accuracy in foggy environments. According to the comparative analysis of the target detection models in foggy environments presented in Figures 15 and 16 and Table 6, our algorithm achieved an accuracy of 76.8%, which was 7.4% higher than that of YOLOv5s. According to a comparison of detection model accuracy, AOD-Net + YOLOv5s can increase the accuracy of target detection in a foggy environment, suggesting that the AOD-Net defogging algorithm has some bearing on target recognition. The improved AOD-Net increased the visibility of the target and improved the image defogging effect. The combination of the improved AOD-Net and YOLOv5s enhanced the precision of target recognition. The accuracy of target detection in foggy environments using the improved AOD-Net algorithm combined with YOLOv5s and the improved YOLOv5s was 77.4% and 76.8%, respectively, with a difference of 0.6% in detection accuracy. The enhanced YOLOv5s target detection model exhibits a reduced parameter count and a lighter model structure, making it better suited for vehicle target detection tasks.

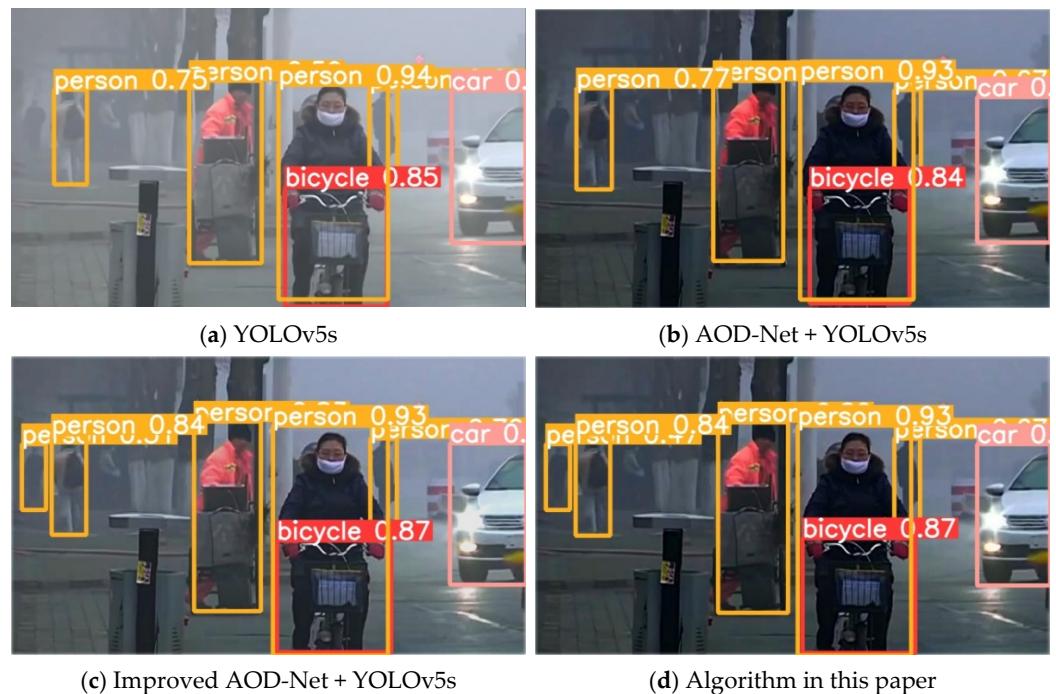


Figure 15. Target detection comparison in a light fog environment.

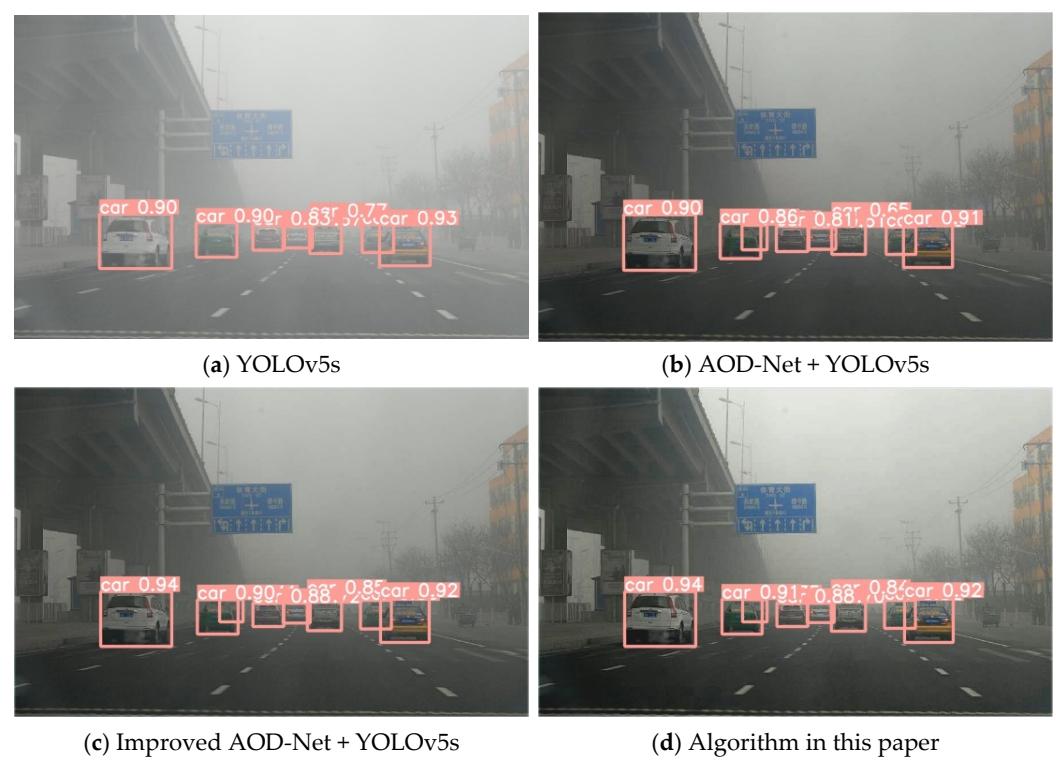


Figure 16. Comparison of target detection in a dense fog environment.

5. Conclusions

- (1) Improved AOD-Net Defogging Technique: To mitigate the challenge of poor image quality in hazy environments, we have developed an enhanced AOD-Net neural network defogging technique. The incorporation of the hybrid convolutional module (HDC) extends the perceptual field within the image feature extraction process, thereby enhancing the capability to extract detailed image feature information. By

employing a hybrid loss function, we address the issues of low contrast between dehazed and original images and the significant discrepancies in detail that arise from utilizing a single mean squared error (MSE) loss function.

- (2) Optimized YOLOv5s for Mobile Devices: To cater to the stringent requirements of target detection tasks on mobile devices used as vehicular terminals, we have developed an enhanced YOLOv5s target identification algorithm. The model parameters have been optimized to reduce complexity and the number of parameters minimized with negligible loss in accuracy. We have made targeted improvements to three key components of the YOLOv5s model: the backbone, the neck, and the detection layer. These enhancements provide significant advantages for application within in-vehicle mobile devices.
- (3) Validation and Analysis: Through rigorous validation and analysis, we have drawn conclusions on the performance of our enhanced AOD-Net + YOLOv5s-based target detection system for hazy environments. The enhanced YOLOv5s algorithm is characterized by a lightweight model architecture, a reduced number of parameters, and improved migration and applicability, all while maintaining equivalent accuracy. Furthermore, our enhanced AOD-Net + YOLOv5s defogging algorithm demonstrates superior performance in enhancing the contrast, color recovery, and overall brightness of dehazed images. The improved target detection approach, which integrates AOD-Net with YOLOv5s, effectively meets the objectives of this research and significantly boosts detection accuracy in hazy conditions.

Author Contributions: Methodology and writing—original draft preparation, A.L. and G.X.; formal analysis and investigation, C.G.; data curation, W.Y.; resources, A.L.; validation, C.X. and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This project is supported by Shandong Province Science and Technology Small and Medium Enterprises Enhancement Project (Grant No. 2023 TSGC0288), Jinan 2023 Talent Development Special Fund Research Leader Studio Project (Grant No. 202333067), Foreign Expert Project (Grant No. G2023023002L), Shandong Provincial Higher Educational Youth Innovation Science and Technology Program (Grant No. 2019KJB019), Major Science and Technology Innovation Project, Shandong Province (Grant No. 2022CXGC020706), Science and Technology Project of Shandong Department of Transportation, China (Grant No. 2021B113). Ministry of Industry and Information Technology Manufacturing High-Quality Development Project, China (Grant No. 2023ZY02002). National Natural Science Foundation of China (Grant No. 51505258), Natural Science Foundation of Shandong Province, China (Grant No. R2015EL019).

Data Availability Statement: Data is contained within the article.

Acknowledgments: We thank all the authors for their contributions to the writing of this article.

Conflicts of Interest: Author Wenpeng Yue was employed by the company Sinotruk Jining Commercial Vehicle Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Shang, H.; Sun, L.; Qin, W. Pedestrian Detection at Night Based on the Fusion of Infrared Camera and Millimeter-Wave Radar. *Chin. J. Sens. Actuators* **2021**, *34*, 1137–1145.
2. Wang, H.; Liu, M.; Cai, Y.; Chen, L. Vehicle target detection algorithm based on fusion of lidar and millimeter wave radar. *J. Jiangsu Univ. Nat. Sci. Ed.* **2021**, *42*, 389–394.
3. Chen, Q.; Ji, J.; Chong, Y.; Gong, M. Vehicle and Pedestrian Detection Based on AOD-Net and SSD Algorithm in Hazy Environment. *J. Chongqing Univ. Technol. Nat. Sci.* **2021**, *35*, 108–117.
4. Wang, K.; Yang, Y.; Fei, S. Review of hazy image sharpening methods. *CAAI Trans. Intell. Syst.* **2023**, *18*, 217–230.
5. Zhang, H.; Patel, V.M. Densely Connected Pyramid Dehazing Network. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3194–3203.
6. Zhao, Q.; Zhang, H.; Cui, J.; Sun, Y.; Duan, S.; Xia, C.; Gao, X. A hybrid of local and global atmospheric scattering model for depth prediction via cross Bayesian model. *Int. J. Comput. Sci. Eng.* **2022**, *25*, 448–459. [CrossRef]
7. Land, E.H. The retinex. *Am. Sci.* **1964**, *52*, 247–264.

8. Petro, A.B.; Sbert, C.; Morel, J.M. Multiscale retinex. *Limage Process. Line* **2014**, *4*, 71–88. [[CrossRef](#)]
9. Li, B.; Peng, X.; Wang, Z.; Xu, J.; Feng, D. AOD-net: All-in-one dehazing network. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4780–4788.
10. Zhang, Q.; Chen, Z.; Jiang, H.; Zhao, J. Lightweight Image Defogging Algorithm Based on Improved AOD-Net. *Res. Explor. Lab.* **2022**, *41*, 18–22.
11. Wang, K.; Yang, Y.; Li, B.; Li, X.; Cui, L. Uneven Image Dehazing by Heterogeneous Twin Network. *IEEE Access* **2020**, *8*, 118485–118496. [[CrossRef](#)]
12. Raj, N.B.; Venkateswaran, N. Single image haze removal using a generative adversarial network. In Proceedings of the 2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), Chennai, India, 4–6 August 2020; pp. 37–42.
13. Sun, G.; Li, C.; Zhang, H. Safety Helmet Wearing Detection Method Fused with Self-Attention Mechanism. *Comput. Eng. Appl.* **2022**, *58*, 300–304.
14. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
15. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision(ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
16. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell. TPAMI* **2016**, *36*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
17. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
18. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
19. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 1981–1990.
20. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision(ECCV), Amsterdam, The Netherlands, 8–16 October 2016; pp. 21–37.
21. Nayar, S.K.; Narasimhan, S.G. Vision in bad weather. In Proceedings of the seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; Volume 2, pp. 820–827.
22. Wang, P.; Chen, P.; Yuan, Y.; Liu, D.; Huang, Z.; Hou, X.; Cottrell, G. Understanding convolution for semantic segmentation. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 1451–1460.
23. Liu, Y.; Zhao, G. Pad-net: A perception-aided single image dehazing network. *arXiv* **2018**, arXiv:1805.03146.
24. Li, H.; Li, J.; Wei, H.; Liu, Z.; Zhan, Z.; Ren, Q. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles. *arXiv* **2022**, arXiv:2206.02424.
25. Niu, Z.; Zhong, G.; Yu, H. A Review on the Attention Mechanism of Deep Learning. *Neurocomputing* **2021**, *452*, 48–62. [[CrossRef](#)]
26. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
27. Cho, Y.; Jeong, J.; Kim, A. Model-assisted multiband fusion for single image enhancement and applications to robot vision. *IEEE Robot. Autom. Lett.* **2018**, *3*, 2822–2829.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.