Traffic & Transportation

DarkDet – Object Detection Method Based on Image Enhancement and Channel Fusion for Autonomous Driving in Low-Light Conditions

Huiyong LI[1], Yujing WANG[2]

Original Scientific Paper
Submitted: 7 Apr 2025
Accepted: 12 Sep 2025
Published: 27 May 2026

[1] lihuiyong@jssc.edu.cn, School of Intelligent Manufacturing and Information, Jiangsu Shipping College, Nantong, China
[2] Corresponding author, wangyujing@abtu.edu.cn, Faculty of Physics and Electrical-Electronic Engineering, Aba Teachers University, Aba, China; Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia

Publisher:
Faculty of Transport and Traffic Sciences, University of Zagreb

ABSTRACT
As a significant innovation in modern transportation, autonomous driving technology offers unprecedented opportunities for improving traffic safety and travel convenience. However, under low-light conditions, insufficient illumination can cause object blurring, indistinct edges and confusion between the foreground and background, thereby compromising the reliability and safety of autonomous driving systems. Consequently, achieving robust object detection under low-light conditions has become a critical challenge in autonomous driving research. To address this issue, we propose DarkDet, a framework specifically designed for object detection in low-light environments. DarkDet consists of a Low-Light Enhancement Module (LLEM) and a Channel Fusion Module (CFM). The LLEM enhances features in dark regions of an image, while the CFM employs an attention mechanism to fuse multi-channel features, which are then fed into an object detector for final predictions. Furthermore, to expand object detection datasets tailored for low-light autonomous driving scenarios, we augment the original BDD100K dataset to construct Dark-BDD100K and collect an additional dataset, DarkCity, comprising 5,000 annotated nighttime urban road images. Finally, we evaluate the proposed DarkDet framework on Dark-BDD100K and DarkCity. Experimental results demonstrate that DarkDet achieves significant performance improvements under low-light conditions, effectively enhancing the accuracy and stability of object detection.

KEYWORDS
object detection; low-light scenarios; low-light object detection datasets; YOLOv9; autonomous driving.

1. INTRODUCTION

With the continuous advancement of science and technology, autonomous driving technology is becoming a revolutionary change in the field of modern transportation [1]. Autonomous driving systems have the capacity to enhance driving convenience and safety, thereby exerting a transformative influence on urban transportation planning [2]. Moreover, they contribute to the reduction of traffic accidents and present intelligent mobility solutions [3].

Autonomous driving technology, often referred to as autonomous driving systems or autonomous vehicles, encompasses the technological framework that enables self-sufficient navigation and operation of vehicles, devoid of human driver intervention [4]. This autonomy is facilitated through intricate perception, recognition, decision-making and control systems [5]. At the heart of automated driving technology lies the process of perception and recognition, which harnesses data acquired from the perception system for the identification and analysis of the surroundings [6, 7]. This includes, but is not limited to, road markings, traffic signs, traffic signals, pedestrians and other vehicles.

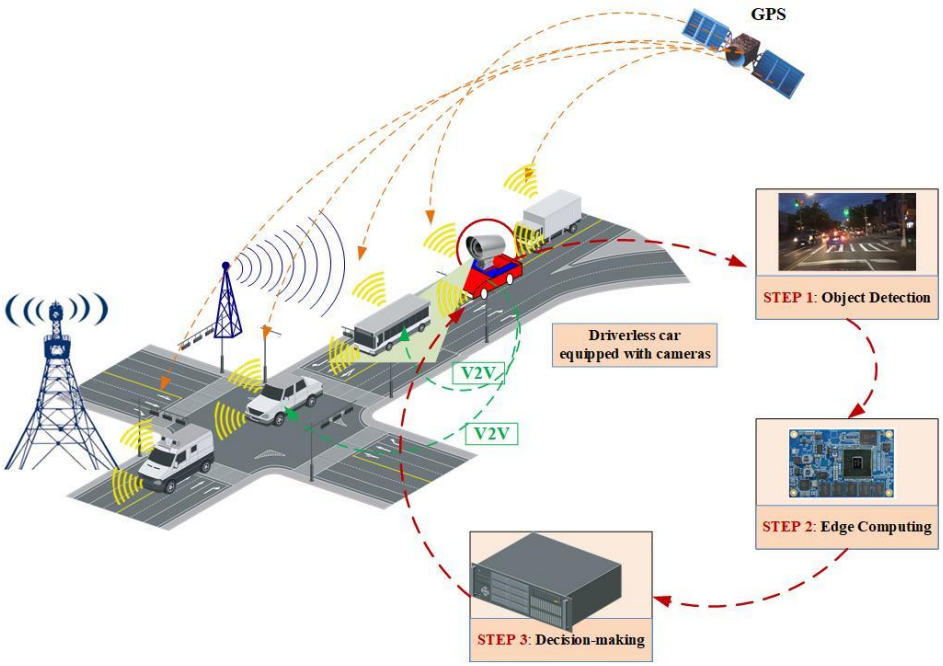


Figure 1 – Schematic diagram of vision-based autonomous driving system. The autonomous driving system on nocturnal urban roads comprises three key components: vision-based target recognition and perception, edge computing and driving planning and decision-making.

The primary constituents of perception devices employed in the domain of autonomous driving encompass an array of sensory technologies, such as cameras, LIDAR, millimetre-wave radar, infrared sensors, thermal imaging cameras and more [8,9]. Among these, camera equipment is particularly noteworthy due to its cost-effectiveness, ease of implementation and high precision. These attributes render it a well-suited component for vision-based autonomous vehicle technology, thereby presenting substantial prospects for expansion and development [10].

In recent years, the continuous advancement of deep learning technology has facilitated remarkable progress in vision-based driverless technology [11]. Nevertheless, vision-based autonomous driving technology confronts several severe challenges in practical applications, including the challenging issue of detecting driving targets in low light conditions, which is especially noteworthy [12]. Statistics reveal a high incidence of traffic accidents at night in urban areas. Despite a 60% reduction in traffic flow at night compared to daytime, 40% of fatal road accidents occur during this time. Furthermore, nighttime traffic accident fatalities make up as much as 50% [13]. Therefore, autonomous driving systems must urgently improve their perception capabilities in low-light conditions in order to address this unavoidable technical challenge and improve road safety [14]. Improving the efficacy of target detection algorithms in low light and complex lighting conditions is a crucial aspect in the widespread adoption of vision-based autonomous technology. The diagram in *Figure 1* illustrates the design of a vision-based autonomous driving system for urban settings during night time.

As illustrated in *Figure 1*, the autonomous driving system on nocturnal urban roads comprises three key components: vision-based target recognition and perception, edge computing and driving planning and decision-making. Initially, a real-time video stream captured by the vehicle's camera is processed using the target recognition algorithm to determine the position and shape of objects surrounding the vehicle. The edge computing device utilises the computing resources of a vehicle-mounted device and cloud computing technology to analyse and process target recognition outcomes of the surrounding environment. It conducts an all-encompassing evaluation based on real-time road conditions and vehicle status information. The computing device responsible for decision-making then determines the most appropriate driving actions, including acceleration, braking and steering, according to the thorough evaluation results gathered via edge computing, in order to maintain vehicle safety.

Under low-light conditions, inadequate lighting results in indistinct target object images, with unclear edges and reduced contrast between the background and foreground [15]. As a result, conventional computer vision algorithms may be unable to correctly identify and detect road obstacles, such as pedestrians and vehicles, thus compromising the dependability and safety of autonomous driving systems [16]. The task of detecting objects for autonomous driving in low light conditions is vital to ensure driving safety and increase the reliability of

autonomous driving systems [17]. Enhancing the performance of object detection algorithms in low-light environments, particularly in the autonomous driving sector, not only has a significant impact on enabling intelligent travel but also offers valuable insights for other computer vision tasks.

The aim of this study is to enhance the YOLOv9 algorithm [18] for detecting objects in autonomously driven vehicles in low-light conditions, thereby improving detection performance in such a demanding environment. Although YOLOv9 algorithm has end-to-end training capabilities, its ability to function under low-light conditions requires improvement [19, 20]. Improving the adaptability and performance of the YOLOv9 algorithm under low-light conditions could enhance the reliability and safety of autonomous driving systems in those situations.

The primary contributions of this paper are as follows:

1) To address the significant performance degradation of existing object detection methods in low-light environments, we propose DarkDet, a novel object detection framework designed specifically for low-light conditions. DarkDet consists of a Low-Light Enhancement Module (LLEM) and a Channel Fusion Module (CFM). The LLEM enhances exposure in dark regions of an image, while the CFM module leverages an attention mechanism to further refine and enhance meaningful features.
2) Given the lack of dedicated object detection datasets for low-light environments, we constructed Dark-BDD100K[21] and DarkCity. Dark-BDD100K comprises real nighttime images from the original dataset as well as synthetically generated nighttime images. DarkCity is collected using an in-vehicle camera on urban roads in Nantong, Jiangsu Province, and contains 5,000 annotated nighttime images.
3) Extensive comparative and ablation experiments were conducted on Dark-BDD100K and DarkCity. The results demonstrate that the proposed DarkDet framework significantly improves object detection performance in low-light conditions while exhibiting strong generalisation capabilities.

The article is structured as follows. Section 2 outlines the related research work, covering target detection algorithms in vision-based autonomous driving technology and target detection methods in low-light conditions. Section 3 delves into the specific research methods employed in this article, encompassing the overall framework, low-light enhancement module, attention collaboration mechanism and detection head. Section 4 presents the methodology and findings of this study, encompassing the dataset, software and hardware settings, results and analysis and ablation experiments. The concluding Section 5 outlines the research findings.

2. RELATED WORKS

2.1 Object detection tasks in autonomous driving

With the continuous advancement of autonomous driving technology, it has become imperative for self-driving vehicles to accurately perceive their surroundings and promptly identify various objects on the road, including pedestrians, vehicles and traffic signs. Object detection is a crucial component of the autonomous driving system, serving not only as the core module for environmental perception, but also directly affecting the driving safety and efficiency of the vehicle. The objective of object detection is to forecast a series of bounding boxes and classification tags for each previously specified item.

Currently, several autonomous driving technologies have introduced advanced object detection methods. The traditionally used engineering-based methods are now being substituted by deep learning approaches, such as Fast R-CNN [22], FCOS [23], DeformableDETR [24] and CenterNet [25]. These techniques enhance the precision and efficacy of object detection by incorporating anchor boxes and prior boxes [26]. They are versatile and adapted to various situations and requirements.

The YOLOv9 algorithm consists of three parts: the backbone network, the feature fusion network and the detection head. The backbone network consists of modules such as CBL, C2f and SPPF, among which C2f is the main module for feature learning. This module imitates the ELAN structure of YOLOv8 [27], extracts features through the residual module, and splices and fuses the features to form a module with stronger feature representation capabilities; CBL represents ordinary feature extraction operations, mainly composed of convolution, batch normalisation and activation; SPPF mainly performs serial maximum pooling calculations to achieve the fusion of local features and global features. The feature fusion network adopts the PAN (Path Aggregation Network) structure, which can enhance the network's feature fusion ability for objects of different scaling scales. The function of the detection head is to infer the target category and location of the features extracted by the network, screen the results for positive and negative samples and calculate the loss. YOLO

positive and negative samples are screened using the Task Aligned Assigner [28] method. The loss calculation includes two branches: classification and regression. BCE loss is used for classification, and Distribution Focal Loss [29] and CIOU (Complete Intersection Over Union) loss are used for regression. In multi-target detection applications, YOLO has always been the most mainstream target detection method, and YOLOv9 is the latest SOTA algorithm in the YOLO series.

2.2 Object detection method under low light conditions

The challenge of detecting objects under low-light circumstances is a significant issue in autonomous driving technology. Reduced visibility of target objects in images occurs due to insufficient illumination, which consequently affects target recognition and positioning. Moreover, low-light conditions may generate amplified image noise, resulting in blurred edges of the targeted objects.

Currently, there are ongoing studies focused on resolving the issue of object detection in situations of poor lighting. Yuhua Chen and colleagues have proposed a technique called "Domain Adaptive Faster R-CNN" to reduce both image-level domain offsets (such as image size, style and lighting) and instance-level domain offsets (such as target appearance and size). In this approach, each module learns a domain classification and domain-invariant features through adversarial training, with the addition of consistency regularisation of the classifier to ensure that the RPN learns domain-invariant proposals [30]. Lintao Xu and colleagues proposed an unbalanced points-guided multi-scale transformer-based conditional normalising flow (UPT-Flow) for low-light image enhancement [31]. Yu-Jhe Li and colleagues introduced an adaptive teacher (AT) teacher-student framework that leverages domain adversarial learning and weak-strong data augmentation to address domain discrepancies [32]. The algorithms for object detection under low-light conditions also include Nerf in the dark algorithm [33], Retinexformer algorithm [34], NLE-YOLO algorithm [35], Pe-YOLO algorithm [36] and Dark-YOLO algorithm [37].

In recent years, Zero-DCE [38] has achieved state-of-the-art performance in the field of enhancing low-light and low-contrast images. Zero-DCE is an algorithm designed for image enhancement which can automatically adjust the curve of the image to improve contrast and brightness, resulting in a more visually appealing image. The acronym stands for Zero-Reference Deep Curve Estimation [39, 40]. The term "Zero-Reference" in the name Zero-DCE signifies that the algorithm operates without a reference image or curve, utilising self-supervised learning for image enhancement [41, 42]. Consequently, it is more versatile, since it does not rely on prior knowledge or user input. The primary concept behind the algorithm involves using a deep convolutional neural network to learn the mapping function of the image curve [43]. This system has been designed to enhance low-contrast or low-light images, mapping them to those with improved contrast and brightness. Through comparison of the differences between the input and generated image during the training process, the system learns how to automatically adjust the curve of the image to achieve enhancement.

3. THE PROPOSED METHOD

3.1 Framework

The DarkDet algorithm's overall framework is illustrated in *Figure 2*. DarkDet consists of a Low-Light Enhancement Module (LLEM) and a Channel Fusion Module (CFM). The LLEM enhances features in dark regions of an image, while the CFM employs an attention mechanism to fuse multi-channel features, which are then fed into the object detector for final predictions.

3.2 Image enhancement mechanism

Data enhancement is a crucial method of improving the model's robustness and plays a significant role in detecting objects under low-light conditions. To tackle the challenges posed by low-light images, we investigated two data augmentation approaches.

Firstly, we present an adversarial training data generation method, using a generative adversarial network (GAN) [44] to produce a series of virtual images in low-light conditions, which are then combined with real images for training. Adversarial training may enhance the model's ability to withstand low-light images and boost its practical performance.

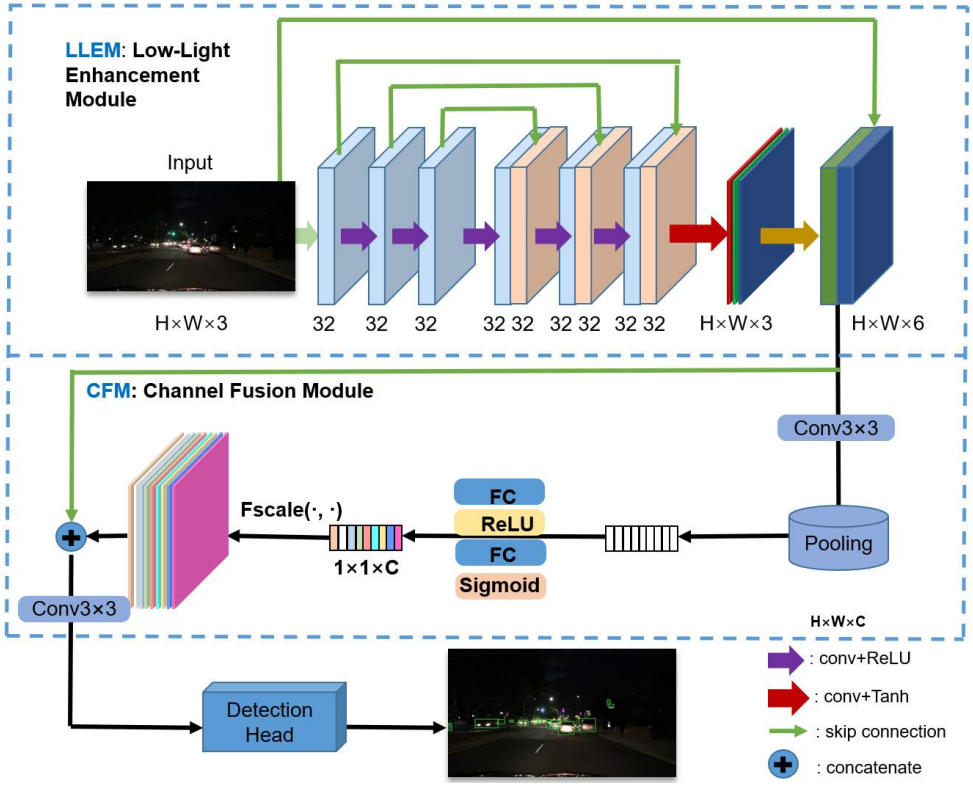


Figure 2 – The DarkDet architecture. DarkDet consists of a Low-Light Enhancement Module (LLEM) and a Channel Fusion Module (CFM).

Additionally, our proposed low-light enhancement module, LLEM, draws inspiration from the DCE-Net network of the Zero-DCE algorithm. *Figure 3* demonstrates the network structure of this module, comprising six symmetrical and uncomplicated convolutional layers. To further decrease computational expenses, accelerate inference speed and ameliorate network performance, we recommend substituting ordinary convolutional layers with depthwise separable convolution (DSC) within LLEM. This will reduce network parameters without significantly affecting model performance. Each depthwise separable convolution comprises a depth convolution kernel and a point convolution. The kernel size for the depth convolution is 3×3 with a stride of 1. The point convolution has a kernel size of 1×1 with a stride of 1. The overall length is 1.

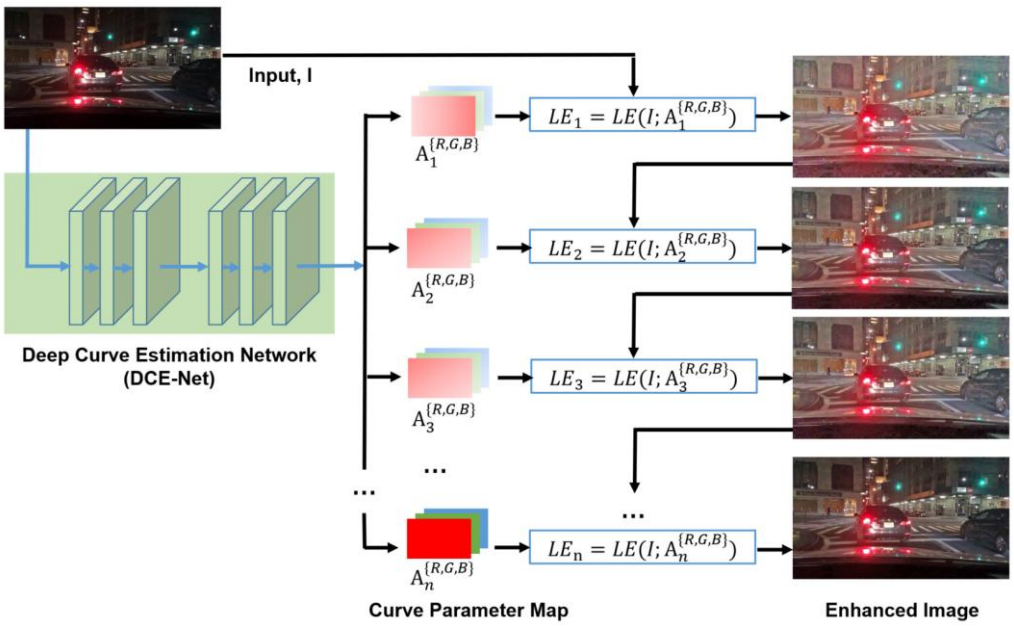


Figure 3 – Low-light image enhancement module architecture. It comprises six symmetrical and uncomplicated convolutional layers.

This method does not necessitate labelled sample data for training, avoiding the issue of model overfitting and significantly enhancing model robustness. To maintain contrast between pixels, the enhancement curve should be kept monotonic. To avoid overflow and truncation of features, the pixel values influenced by the enhancement curve should be normalised. The curve parameters should be differentiable for ease of application in network learning.

Based on Zero-DCE [38], image pixel values are modified using a light enhancement curve. Additionally, the pretrained weights for the LLEM module are derived by mapping this curve. The formulation of the first-order light enhancement curve is provided in *Equation 1*:

$$LE(I(x);\alpha) = I(x) + \alpha I(x)\big(1 - I(x)\big) \tag{1}$$

Here, x represents the pixel location, and $LE(I(x);\alpha)$ indicates the enhanced pixel intensity for the input image $I(x)$. The parameter α ranges from -1 to 1 and is learnable, regulating both the intensity of the enhancement curve and the exposure correction. Each pixel value is scaled to the range [0,1], and all processing is performed independently on each pixel. By applying the enhancement curve to the three colour channels (R, G, B), the method effectively preserves the inherent features of the original image while reducing the risk of oversaturation.

The previously stated curve is a global one, often leading to unnecessary high or low levels of local improvement when implemented across all pixels. This paper will adjust a parameter curve that fits optimally to each individual pixel of the inputted image. The assumption is made that a given small local area holds equivalent intensity levels and adjustment curves, securing a maintained monotonic relationship between adjacent pixels of the output. Hence, *Equation 2* clearly shows the obtained final high-order pixel curve.

$$LE_n(x) = LE_{n-1}(x) + A_n LE_{n-1}(x)\big(1 - LE_{n-1}(x)\big) \tag{2}$$

In this equation, n denotes the number of iterative steps, which determines the degree of curvature of the enhancement. This work adopts n=8 to adequately handle a wide range of scenarios. A signifies the size of the input image, and it is presumed that pixels within a localised region exhibit comparable intensity levels and require consistent adjustment curves.

3.3 Channel fusion based on attention mechanism

To meet the challenge of detecting objects in low-light conditions, it is vital to optimise the attention mechanism. We utilise ResNet's residual connection concept and splice the original input and image-enhanced data into six-channel data. We splice the data using "Concatenate" to merge inside the channel dimension. This approach facilitates the enhancement of the image to confer more comprehensive background details while retaining the original object-relevant information. However, merely concatenating the two data types based on their channel dimension leads to poor information integration and supplementation. Consequently, the resulting auxiliary information lacks sufficiency in supporting subsequent object detection.

To address the aforementioned issues, we employ the attention mechanism to the six-channel spliced data and utilise the attention collaboration mechanism to produce an equitable fusion of two data types in the channel dimension. This process supplies the object detection network with ideal image data. We modify the SE-Net's attention unit as the Channel Fusion Module (CFM) in the algorithm framework, thus improving its efficiency in responding to output from the front layer. The Channel Fusion Module (CFM) is located in the Low-Light Enhancement Module (LLEM) thereafter. Together with LLEM, it constitutes the portion of the algorithm framework dedicated to enhancing low-light images, and it is positioned at the forefront of the object detection module.

Referring to the SE-block proposed in SE-Net, a convolution layer was added to both the front and rear ends. The input is taken from the six-channel data output of the LLEM. Firstly, the data undergoes a global average pooling operation to compress it into 1×1×C data, thereby averaging the information of all points in space into one value. This enables the fusion of global information in a channel feature map, providing a wider sensing area. The calculation formula is as follows (*Equation 3*):

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} u_c(i,j) \tag{3}$$

where the height of the feature map is denoted by H, the width by W and the number of channels by C. The two-dimensional feature matrix of the corresponding input channel is represented by u_c, and the output after global average pooling by z_c. Technical term abbreviations are explained upon first use.

To utilise the combined information in the compressed 1×1×C data and acquire the interdependencies among channels, two fully connected (FC) layers have been incorporated to integrate the feature information of each channel. The process is denoted as FC-ReLU-FC-Sigmoid and can be expressed as *Equation 4*:

$$s = \mathrm{Fex}(z, W) = \sigma\big(g(z, W)\big) = \sigma\big(W_2 \delta(W_1 z)\big) \tag{4}$$

where σ denotes the ReLU activation function, δ denotes the sigmoid activation function. W_1, W_2 and g denote the scaling parameter, while z denotes the output after global average pooling.

The size of s in *Equation 4* is 1×1×C, which signifies the magnitude of the C channel trait maps in the input U. As the weight is learnt through the prior fully connected layer and nonlinear layer, full end-to-end training is attainable. Once these weights are obtained, they must be allocated back to each channel characteristic map in the initial U. The equation for this calculation is as follows (*Equation 5*):

$$\tilde{X}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \tag{5}$$

where s_c denotes the weight of the corresponding channel, while x_c refers to the characteristic two-dimensional matrix that redistributes the weight of each channel.

3.4 Detection head

We use the YOLOv9 method as the detection head. The algorithm implements the target detection task through a series of end-to-end processing steps. First, after the input image undergoes a preprocessing step, features are extracted through a powerful backbone network. Subsequently, through multi-scale feature fusion technology, detection is performed on feature maps at different levels to improve the detection ability of targets of different sizes. YOLOv9 implements target detection and classification in a single forward propagation method, and has the advantages of high real-time performance, excellent accuracy and lightweight model.

4. EXPERIMENTS

4.1 Datasets

Dark-BDD100K

In this research, experiments were conducted on the Dark-BDD100K dataset [21]. The Dark-BDD100K dataset consists of two parts – a low-light scene data subset of the BDD100K data and a low-light scene dataset generated based on BDD100K daytime images.

(a) *(b)*

(c) *(d)*

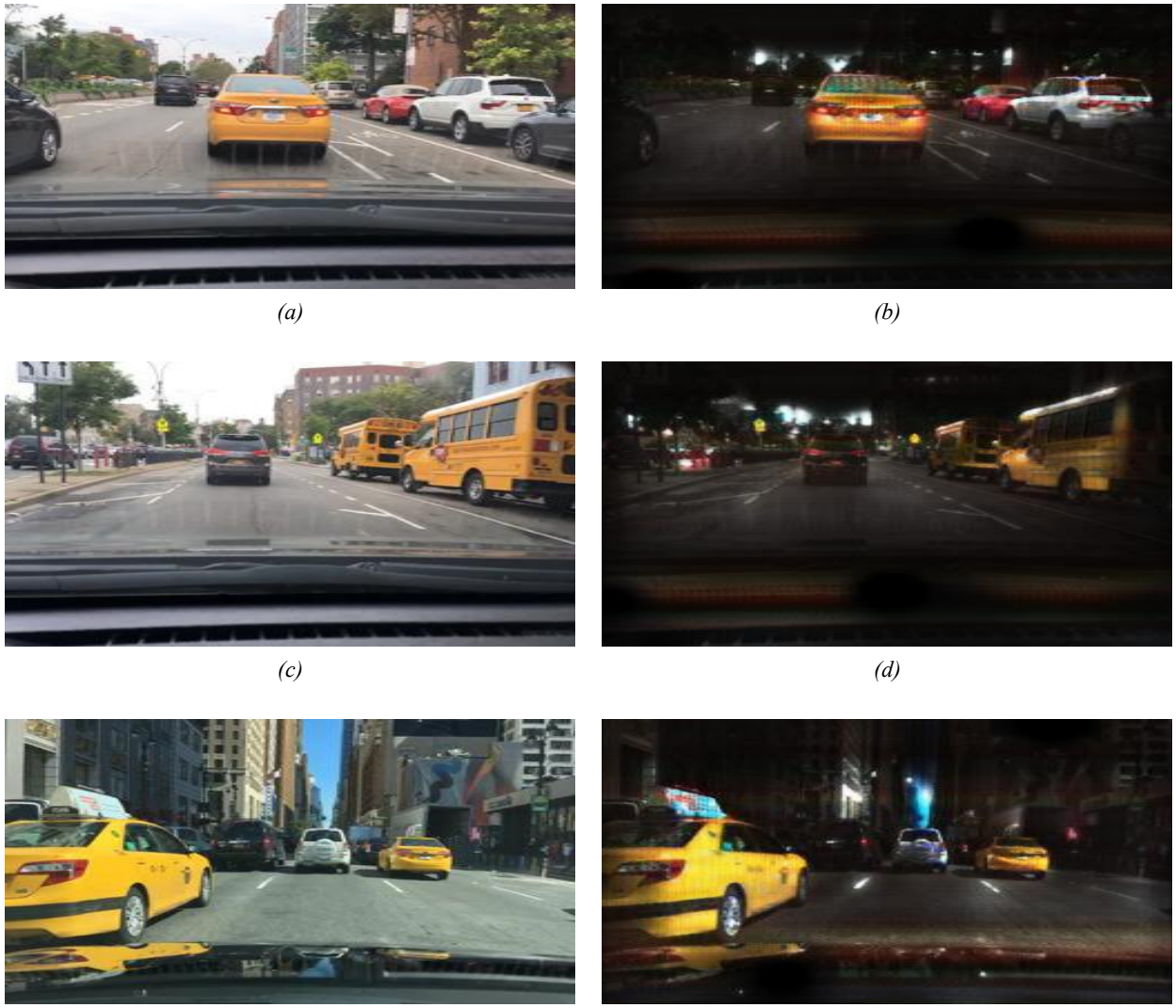
Figure 4 – Real low-light images in BDD100K. a), b), c) and d) are four images randomly selected from the dataset.

Real low-light images. The BDD100K dataset [21], released by the University of California Berkeley AI Laboratory in 2018, is one of the largest and most diverse public driving data sets currently available. BAIR captures key frames at the 10th second of each video, acquires 100,000 images of size 1280×720 and annotates them. The data sets are collected from various locations including New York, Berkeley and San Francisco. The purpose of the data set is to annotate 2D bounding boxes on 100,000 images of vehicles, traffic signals, signs, pedestrians, bicycles and train passengers for road object detection.

Firstly, we excluded 28,028 night-time images out of the 100,000 total images in BDD100k, as illustrated in *Figure 4*.

Generated low-light images. We present an adversarial methodology for generating training data. Our approach involves using a generative adversarial network (CycleGAN [44]) to create 17,972 virtual images experiencing low-light conditions, based on daytime images from the BDD100K dataset as shown in *Figure 5*. Refer to the accompanying figure for a visual representation. Our technique offers a novel means of data augmentation for low-light image recognition tasks.

We blend synthetic and authentic images for model training. This adversarial approach enhances the model's ability to handle low-light images, thereby enhancing performance in practical scenarios. The dataset consists of 10,000 images, from which we selected 7,000 randomly for the training set and the remaining 3,000 for the test set.



(a) *(b)*

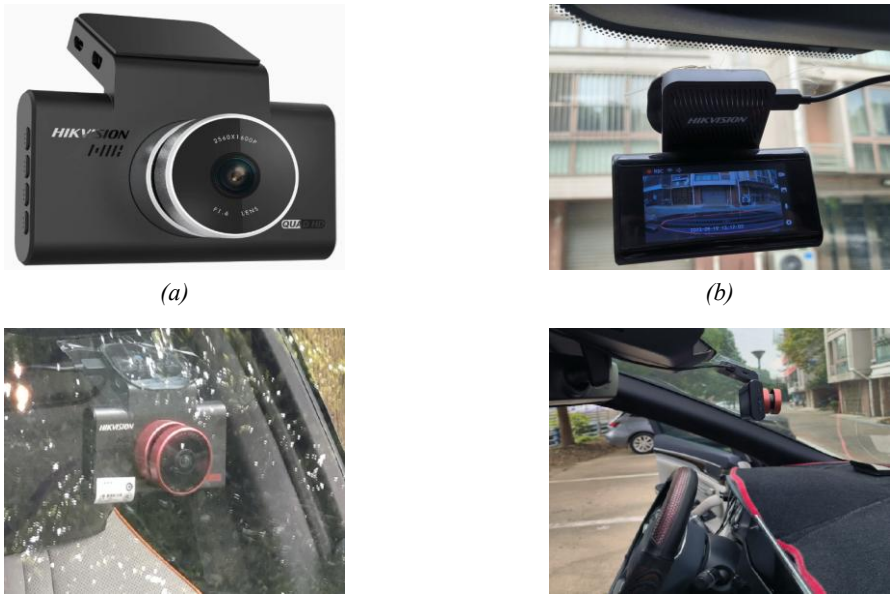
(c) *(d)*

(e) *(f)*

Figure 5 – Low-light images created from daytime driving scene images. a), c) and e) are three images randomly selected from daytime driving scene images dataset; b), d) and f) are the corresponding night scene images generated from these three images by the CycleGAN

DarkCity

We acquired 5,000 real night driving images on a city road in Nantong, Jiangsu Province, China, by using an experimental vehicle. The HIKVISION C6 car recorder served as our collection device and has a frame rate of 25 frames/second and a resolution of 3840×2160. The acquired period was from 19:30 to 21:30 in September 2024. The installation position of the recorder is shown in *Figure 6*.



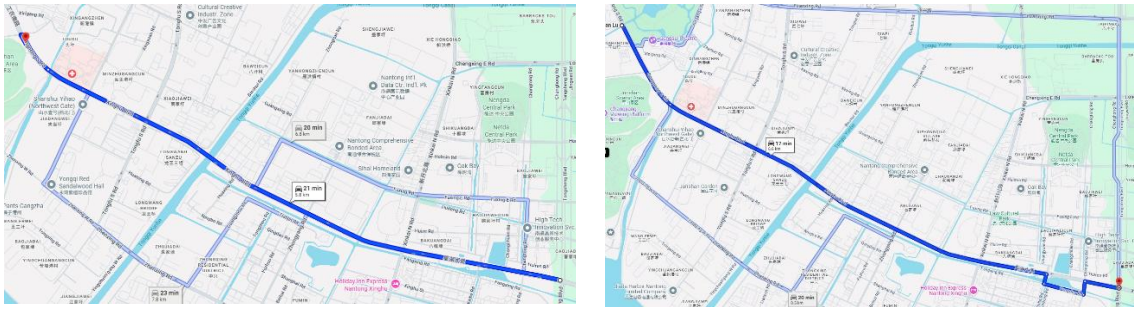
(a) *(b)*

(c) *(d)*

Figure 6 – In-vehicle street view capturing device. a) is a front view of the capturing device. b), c) and d) are photographs of the installation position of the capturing device in the car taken from different angles.

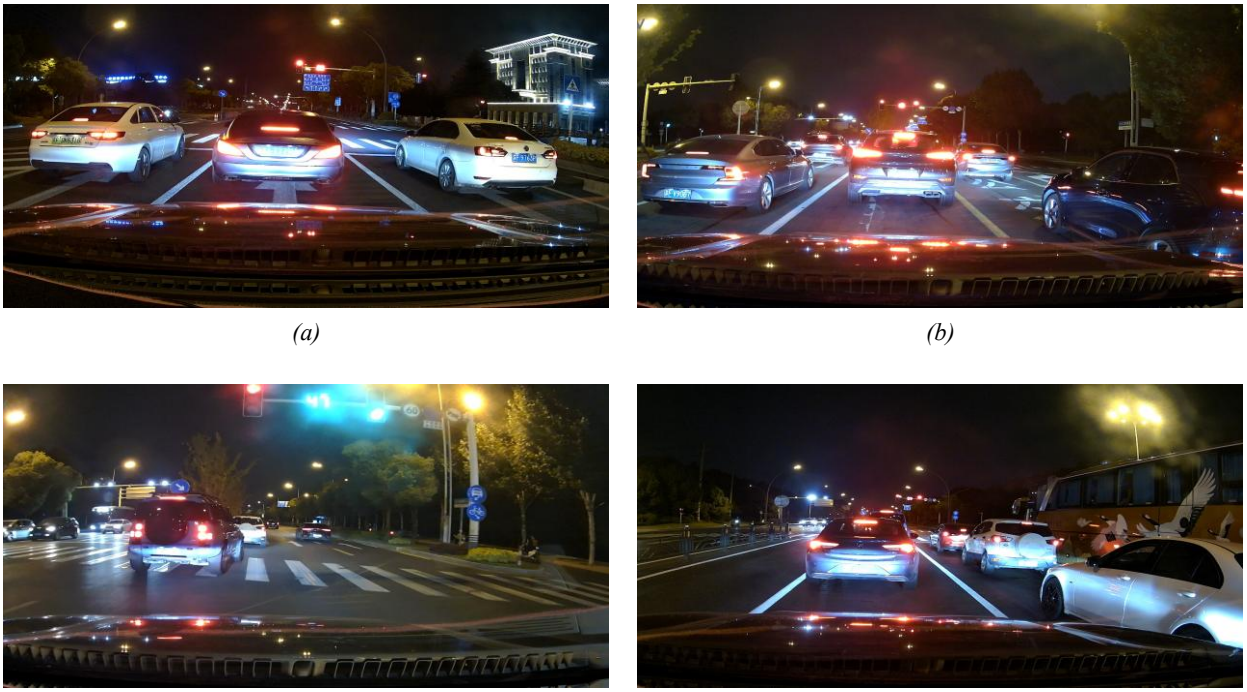
The route map of the street view collection vehicle is shown in *Figure 7*. The route map forms a loop, and the actual driving duration was 41 minutes. The driving route includes two roads, named Gongnong Road and Xinghu Avenue. These two roads are classified as urban expressways. The actual nighttime images captured in the experiment are shown in *Figure 8*. For the purposes of this study, we have named this dataset the DarkCity dataset.

During the data acquisition process for the DarkCity dataset, although the in-vehicle camera (HIKVISION C6) recorded video at a rate of 25 frames per second, we retained only a subset of frames to avoid temporal redundancy and ensure scene diversity. Specifically, we adopted a uniform frame sampling strategy by extracting one frame every 10 frames, which significantly reduced duplicate information between adjacent frames. After automatic frame extraction, a manual filtering process was applied to further enhance data quality by removing blurred images, severely occluded frames and scenes lacking sufficient road target information. The final dataset consists of 5,000 high-quality images, covering a wide range of urban road scenarios, traffic densities and lighting conditions. This frame selection strategy ensures that the collected dataset is both representative and practically valuable for training and evaluating object detection algorithms under low-light conditions.



(a) *(b)*

Figure 7 – The route of the street view capturing vehicle. The route map forms a loop, and the actual driving duration was 41 minutes. a) represents the outbound trip, b) shows the return trip. The route includes two main urban roads: Gongnong Road and Xinghu Avenue.



(a) *(b)*

(c) *(d)*

Figure 8 – DarkCity dataset. a), b), c) and d) are four images randomly selected from the dataset.

4.2 Implementation details

The experimental platform configuration used in all experiments in this article is: CPU: Intel i9-10900K, GPU: NVIDIA GeForce RTX 3090 with 24 GB of video memory, Ubuntu 22.04, development environment: Python 3.7, using PyTorch1.11.0 framework, batch size is 4, epochs are 300, the experimental result is the best value.

4.3 Experimental results and analysis

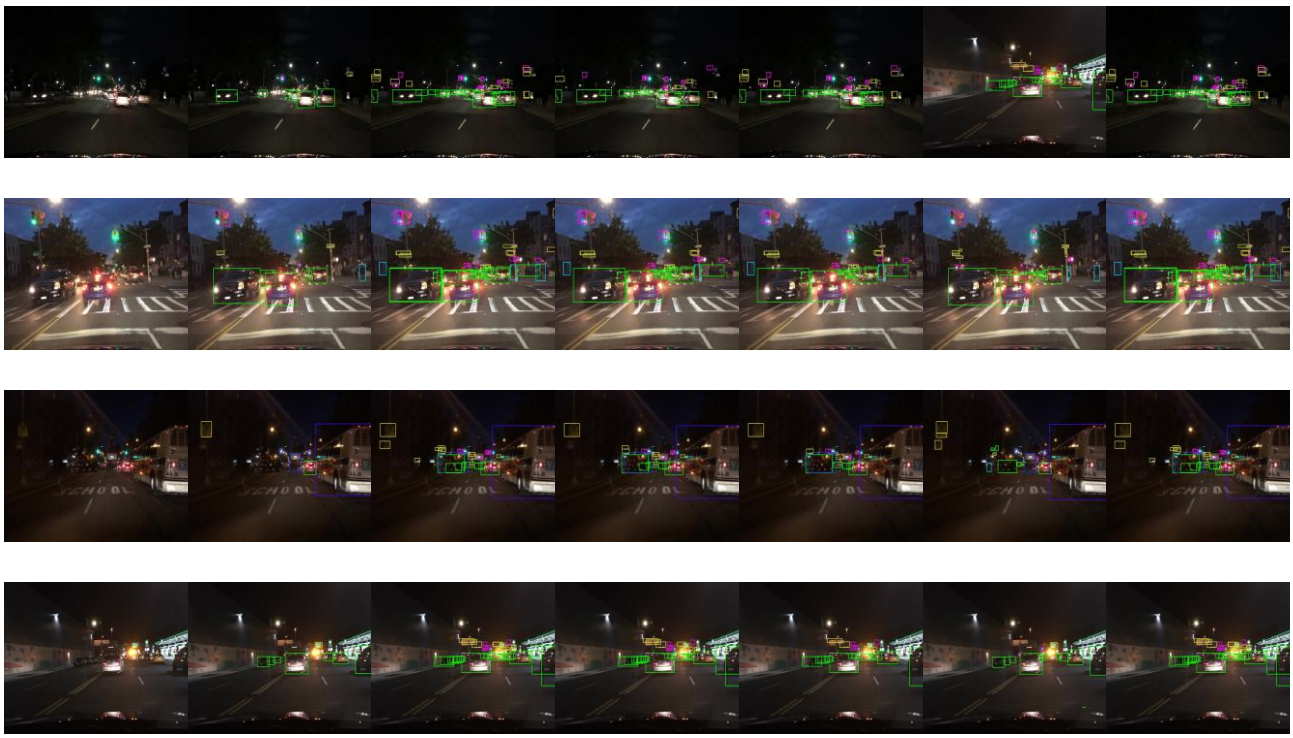
To confirm the effectiveness and benefits of the DarkDet model in comparison to conventional object detection techniques, *Table 1* presents the primary outcomes on the validation dataset. We compare the DarkDet algorithm with the NeRF in the Dark algorithm [33], Retinexformer algorithm [34], NLE-YOLO algorithm [35], Pe-YOLO algorithm [36] and Dark-YOLO algorithm [37]. *Figure 9* illustrates the comparison of object detection results between these algorithms.

Table 1 – Experimental results on the Dark-BDD100K dataset

Algorithm	AP	Pedestrian	Rider	Car	Truck	Bus	Motor	Bicycle	Traffic light	Traffic sign
NeRF in the Dark [33]	48.1	40.5	27.7	67.1	52.5	53.9	34.6	43.2	49.0	64.1
Retinexformer [34]	47.7	39.5	26.7	67.5	52.3	53.5	35.1	42.4	48.7	63.8
NLE-YOLO [35]	48.6	41.4	27.6	68.3	53.4	53.9	35.7	44.0	48.5	64.8
Pe-YOLO [36]	48.4	41.9	26.9	67.8	54.1	54.4	33.8	43.5	48.6	64.7
Dark-YOLO [37]	47.9	37.7	28.2	66.9	52.8	53.5	35.1	44.4	47.8	65.0
DarkDet (Ours)	**50.5**	**44.9**	**31.3**	**69.6**	**54.4**	**55.7**	**37.9**	**45.1**	**49.8**	**65.7**

From *Table 1* and *Figure 9*, it is evident that the performance of DarkDet, as proposed in this article, is notably superior to the comparison models. Upon comparison, it can be observed that the DarkDet algorithm has significantly improved performance under low-light conditions. Specifically, this advantage is attributed to the

attention collaboration mechanism and the feature fusion of environmental lighting information. These techniques facilitate a more precise capturing of the target's features, resulting in enhanced accuracy and stability in object detection. Moreover, the utilisation of data augmentation strategies and the generation of adversarial training data further boosts the model's resilience and ability to generalise in low-light scenarios. Moreover, the utilisation of data augmentation strategies and the generation of adversarial training data further boosts the model's resilience and ability to generalise in low-light scenarios.



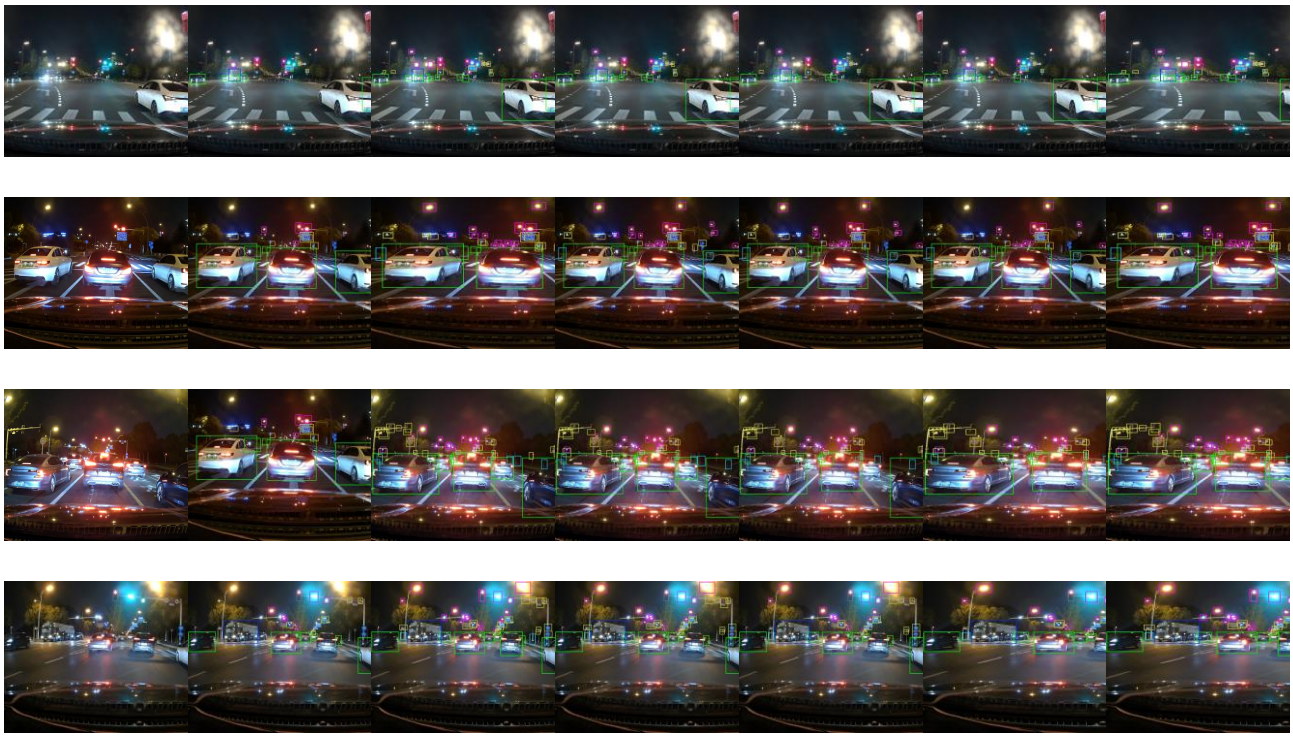
(a) *(b)* *(c)* *(d)* *(e)* *(f)* *(g)*

Figure 9 – Comparison of recognition effects on Dark-BDD100K. Column (a) is the original image. Columns (b) to (f) show the recognition results of the six algorithms: NeRF in the Dark, Retinexformer, NLE-YOLO, Pe-YOLO, Dark-Yolo and DarkDet.

It is evident from *Table 2* and *Figure 10* that the DarkDet algorithm might possess some limitations. Specifically, the enhanced attention mechanism and feature fusion strategy may prove insufficient in complex situations and may require further fine-tuning based on specific circumstances. Furthermore, the upgraded DarkDet algorithm could exhibit a significant reliance on datasets, necessitating appropriate adjustments and training for other low-light datasets. In practical applications, it is necessary to further enhance the algorithm's real-time performance and computational efficiency.

Table 2 – Experimental results on the DarkCity dataset

Algorithm	AP	Pedestrian	Rider	Car	Truck	Bus	Motor	Bicycle	Traffic light	Traffic sign
NeRF in the Dark [33]	43.2	43.6	23.5	60.7	48.9	49.4	28.8	36.5	38.5	59.3
Retinexformer [34]	44.1	42.7	24.1	62.0	50.2	50.7	28.6	36.5	40.4	61.7
NLE-YOLO [35]	43.7	44.2	22.5	62.3	50.6	51.5	25.5	34.8	40.9	60.7
Pe-YOLO [36]	44.7	42.1	25.6	62.9	49.8	51.3	33.4	36.2	40.1	61.5
Dark-YOLO [37]	44.0	43.5	23.2	61.4	49.5	50.7	29.6	35.8	40.4	61.6
DarkDet (Ours)	**47.5**	**49.5**	**28.1**	**65.4**	**51.2**	**52.2**	**34.5**	**41.8**	**42.2**	**62.3**



(a) *(b)* *(c)* *(d)* *(e)* *(f)* *(g)*

Figure 10 – Comparison of recognition effects on the DarkCity dataset. Column (a) is the original image. Columns (b) to (f) show the recognition results of the six algorithms: NeRF in the Dark, Retinexformer, NLE-YOLO, Pe-YOLO, Dark-Yolo and DarkDet.

To validate the generalisation capability of the proposed DarkDet, we conducted additional experiments on the DarkCity dataset to evaluate its performance when integrated with state-of-the-art detectors such as Fast R-CNN under low-light conditions. The detailed experimental results are presented in *Table 3*, where it can be observed that the detection performance of Fast R-CNN [22], FCOS [23], Deformable DETR [24] and CenterNet [25] significantly improves, with Fast R-CNN achieving a notable 6.0 percentage point increase. Furthermore, for the Rider and Traffic light categories, which are typically more challenging to detect, all detectors except CenterNet achieved improvements exceeding 3%, while CenterNet itself improved by 2.2%. These results further demonstrate the effectiveness of DarkDet from another perspective, highlighting its ability to enhance object detection performance in low-light environments.

Table 3 – Comparison with the state-of-the-art detectors on DarkCity dataset with IoU threshold being 0.5

Method	LLEM+ CFM	AP	Pedestrian	Rider	Car	Truck	Bus	Motor	Bicycle	Traffic light	Traffic sign
Fast-RCNN [22]	√	43.9 (6.0↑)	44.5 (3.8↑)	22.4 (3.0↑)	60.9 (4.2↑)	49.8 (6.5↑)	50.3 (28.2↑)	29.9 (-2.6↓)	36.2 (3.0↑)	39.5 (3.9↑)	61.7 (3.9↑)
	×	37.9	40.7	19.4	56.7	43.3	22.1	32.5	33.2	35.6	57.8
FCOS [23]	√	45.4 (3.7↑)	44.1 (3.8↑)	25.2 (3.3↑)	62.7 (2.2↑)	50.4 (4.5↑)	49.9 (3.3↑)	32.6 (3.9↑)	40.5 (5.3↑)	41.8 (5.0↑)	62.0 (2.6↑)
	×	41.7	40.3	21.9	60.5	45.9	46.6	28.7	35.2	36.8	59.4
Deformable-DETR [24]	√	44.2 (2.8↑)	39.5 (3.8↑)	24.3 (3.7↑)	61.7 (1.7↑)	49.2 (3.8↑)	48.1 (1.6↑)	33.4 (2.5↑)	41.5 (3.7↑)	39.6 (3.1↑)	60.8 (1.0↑)
	×	41.4	35.7	20.6	60.0	45.4	46.5	30.9	37.8	36.5	59.8
CenterNet [25]	√	44.8 (2.4↑)	42.8 (2.3↑)	25.5 (2.2↑)	62.4 (2.2↑)	50.7 (2.0↑)	49.3 (2.1↑)	31.2 (1.6↑)	40.6 (2.1↑)	38.8 (3.7↑)	61.9 (2.7↑)
	×	42.4	40.5	23.3	60.2	48.7	47.2	29.6	38.5	35.1	59.2

4.4 Ablation study

The image enhancement and attention collaboration components of the DarkDet model used in this study have made distinctive contributions to the model's performance. Our ablation analysis investigates the influence of the two fundamental components on the outcomes of the DarkDet model.

Table 4 – Ablation experiment results on the DarkCity dataset

Methods			AP	APl	APm	APs
LLEM	CFM	YOLOv9				
√	√	√	**47.5**	**48.7**	**28.8**	**9.1**
√		√	43.5	45.6	27.6	8.3
	√	√	44.7	46.1	27.9	8.6
		√	42.3	44.9	28.2	8.2

The proposed image enhancement method enhances object detection performance by analysing disparities between large-scale and small-scale feature maps. *Table 4* shows that the YOLOv9 model's performance improved to some extent upon adding the image enhancement module, confirming its efficacy for the proposed DarkDet model without increasing its complexity.

In this article's attention collaborative fusion module, we utilise the attention mechanism on the six-channel spliced data and implement the attention collaborative mechanism to achieve a balanced fusion of both data types in the channel dimension. This produces optimal image data performance for subsequent object detection networks, resulting in improved object detection performance. As indicated by *Table 4*, the accuracy of the attention collaborative fusion module for YOLOv9 surpasses that of other models, highlighting its significance to the DarkDet model. This underscores the vital nature of the attention collaborative fusion module in this article.

4.5 Discussion

This study proposed DarkDet, an object detection framework specifically designed for autonomous driving in low-light conditions. Experimental results on both the Dark-BDD100K and DarkCity datasets demonstrate that DarkDet achieves superior detection accuracy compared to existing low-light object detection algorithms. The proposed Low-Light Enhancement Module (LLEM) and Channel Fusion Module (CFM) effectively improve the representational capacity of image features, enabling the model to extract meaningful object information even under challenging lighting conditions.

The practical value of this work lies in its potential to significantly enhance the perception capabilities of autonomous driving systems during nighttime or in dim environments. In real-world deployments, improving detection performance under low-light conditions is critical for driving safety, as it helps reduce the risk of missing key road targets such as pedestrians, vehicles and traffic signs. Furthermore, the proposed framework can be extended to other computer vision applications that operate under poor lighting conditions, such as nighttime surveillance or robotic navigation in dark environments.

However, the current study has certain limitations. First, although the proposed method significantly improves detection performance under low-light conditions, the introduction of the additional feature fusion increases computational complexity, which may hinder its deployment on resource-constrained edge devices. Second, the robustness of the algorithm under extreme lighting transitions – such as entering tunnels or encountering high-intensity headlight glare – still requires improvement, and this issue has not been fully addressed in this study. Third, despite the careful construction of the DarkCity dataset, its collection is geographically limited and may not fully represent the diversity of urban road environments across different cities or countries.

Future research will focus on the following directions: (1) further optimising the lightweight design of DarkDet to enhance its real-time performance on edge computing platforms while maintaining detection accuracy; (2) expanding the dataset to include more diverse urban and suburban environments, as well as adverse weather conditions such as rain, fog and snow, to improve the model's generalisation capability; and (3) exploring adaptive exposure correction mechanisms that dynamically adjust during real-time driving to better handle sudden lighting changes.

5. CONCLUSION

This paper aims to enhance the YOLOv9 algorithm to suit the object detection task for automated driving in low light conditions. To achieve this, the proposed DarkDet algorithm introduces an adaptive attention mechanism and feature fusion of environmental lighting information. Additionally, data enhancement strategies and adversarial data generation for training were implemented to enhance the accuracy, robustness and generalisation of the model. The results of this experiment indicate that the suggested algorithm greatly enhances object detection accuracy in low-light situations, thus providing robust backing for the use of autonomous driving technology in intricate environments.

The DarkDet algorithm holds immense potential for use in autonomous driving systems. It enhances object detection in low-light conditions, therefore increasing the adaptability and safety of the autonomous driving system in challenging scenarios such as driving during the night, rain or snow. This has a paramount significance in achieving all-weather and all-scenario autonomous driving.

In future research directions, the robustness of the DarkDet algorithm requires further enhancement. Although the current algorithm effectively detects objects in low-light environments, its performance under varying lighting conditions still requires optimisation. For instance, extreme lighting conditions, such as intense sunlight or very low light during the night, may cause the algorithm's performance to degrade. Moreover, the algorithm should be further developed to accommodate various meteorological conditions, such as rain, haze, and snow. These weather phenomena impose greater demands on the perception capabilities of autonomous driving systems. Future research will focus on enhancing the algorithm's adaptability in complex environments, ensuring its ability to maintain stable and efficient object detection performance under diverse lighting and meteorological conditions.

ACKNOWLEDGMENT

This work was supported in part by the 2023 Natural Science Foundation Program of Nantong, Jiangsu, China (JCZ2023027); in part by the 2023 Natural Science Foundation Program of Nantong, Jiangsu, China (JC2023019); in part by Aba Teachers University Talent Special Project (AS-RCZX2025-07).

DECLARATIONS

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] Zhao J, et al. Autonomous driving system: A comprehensive survey. *Expert Systems with Applications.* 2024;242:122836. DOI: 10.1016/j.eswa.2023.122836.

[2] Xu H, et al. A survey on occupancy perception for autonomous driving: The information fusion perspective. *Information Fusion.* 2025;114:102671. DOI: 10.1016/j.inffus.2024.102671.

[3] Chib PS, Singh P. Recent advancements in end-to-end autonomous driving using deep learning: A survey. *IEEE Transactions on Intelligent Vehicles.* 2023;9(1):103-118. DOI: 10.1109/TIV.2023.3318070.

[4] Reda M, et al. Path planning algorithms in the autonomous driving system: A comprehensive review. *Robotics and Autonomous Systems.* 2024;174:104630. DOI: 10.1016/j.robot.2024.104630.

[5] Chen L, et al. End-to-end autonomous driving: Challenges and frontiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*2024;46(12):10164-10183. DOI: 10.1109/TPAMI.2024.3435937.

[6] Weng X, et al. Obstacle avoidance path planning strategy for autonomous vehicles based on genetic algorithm. *Promet-Traffic& Transportation*. 2024;36(4):733–748. DOI: 10.7307/ptt.v36i4.528.

[7] Wang Y, et al. Driving into the future: Multiview visual forecasting and planning with world model for autonomous driving. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024*, *16-22 June 2024, Seattle, WA, USA.*2024. p. 14749-14759. DOI: 10.1109/cvpr52733.2024.01397.

[8] Kong L, et al. Multi-modal data-efficient 3d scene understanding for autonomous driving. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2025;1-18. DOI: 10.1109/TPAMI.2025.3535625.

[9] Bi J, et al. Lane detection for autonomous driving: Comprehensive reviews, current challenges, and future predictions. *IEEE Transactions on Intelligent Transportation Systems.* 2025; :1-18. DOI: 10.1109/TITS.2024.3524603.

[10] Mao J, et al. 3D object detection for autonomous driving: A comprehensive survey. *International Journal of Computer Vision.* 2023;131(8):1909-1963. DOI: 10.1007/s11263-023-01790-1.

[11] Sanil N, et al. Deep learning techniques for obstacle detection and avoidance in driverless cars. *Proceedings of the International Conference on Artificial Intelligence and Signal Processing (AISP) 2020, 10-12 Jan. 2020, Amaravati, India.* 2020. p.1-4. DOI: 10.1109/AISP48273.2020.9073155.

[12] Zablocki É, et al. Explainability of deep vision-based autonomous driving systems: Review and challenges. *International Journal of Computer Vision.* 2022;130(10):2425-2452. DOI: 10.1007/s11263-022-01657-x.

[13] Huang R, et al. Analysis of the characteristics and causes of night tourism accidents in China based on SNA and QAP methods. *International Journal of Environmental Research and Public Health.* 2023;20(3):2584. DOI: 10.3390/ijerph20032584.

[14] Li G, et al. A deep learning based image enhancement approach for autonomous driving at night. *Knowledge-Based Systems.* 2021;213:106617. DOI: 10.1016/j.knosys.2020.106617.

[15] Ostankovich V, et al. Application of CycleGAN-based augmentation for autonomous driving at night. *Proceedings of the International Conference Nonlinearity, Information and Robotics (NIR) 2020, 03-06 Dec. 2020, Innopolis, Russia.* 2020. p. 1-5. DOI: 10.1109/NIR50484.2020.9290218.

[16] Wang H, et al. SFNet-N: An improved SFNet algorithm for semantic segmentation of low-light autonomous driving road scenes. *IEEE Transactions on Intelligent Transportation Systems.* 2022;23(11):21405-21417. DOI: 10.1109/TITS.2022.3177615

[17] Li J, et al. Light the night: A multi-condition diffusion framework for unpaired low-light enhancement in autonomous driving. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024*, *16-22 Jun. 2024, Seattle, WA, USA.* 2024. p. 15205-15215. DOI: 10.1109/cvpr52733.2024.01440.

[18] Wang CY, Yeh IH, Mark Liao HY. YOLOv9: Learning what you want to learn using programmable gradient information. *Proceedings of the European Conference on Computer Vision 2024, Sep. 29 2024, Milan, Italy.*2024. p.1-21. DOI: 10.1007/978-3-031-72751-1_1.

[19] Song H, Wang J, Zhang Y. Detection of abandoned objects based on YOLOv9 and background differencing. *Signal, Image and Video Processing.* 2025;19(1):1-8. DOI: 10.1007/s11760-024-03609-z.

[20] Chen J, et al. Multi-scale and dynamic snake convolution-based YOLOv9 for steel surface defect detection. *The Journal of Supercomputing.* 2025;81(4):541. DOI: 10.1007/s11227-025-07036-w.

[21] Yu F, et al. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020, 14 Jun. 2020, Seattle, WA, USA.* 2020. p.2636-2645. DOI: 10.1109/cvpr42600.2020.00271.

[22] Girshick R. Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision 2015, 07 Dec. 2015, Santiago, Chile.* 2015. p.1440-1448). DOI: 10.1109/iccv.2015.169.

[23] Tian Z, et al. FCOS: A simple and strong anchor-free object detector. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2020;44(4):1922-1933. DOI: 10.1109/TPAMI.2020.3032166.

[24] Zhu X, et al. Deformable DETR: Deformable transformers for end-to-end object detection. *Arxiv Preprint* Arxiv:2010.04159. 2020. DOI: 10.48550/arXiv.2010.04159.

[25] Shi T, et al. Feature-enhanced CenterNet for small object detection in remote sensing images. *Remote Sensing.* 2022;14(21):5488. DOI: 10.3390/rs14215488.

[26] Fu G, Chu H, Tu X. Enhancing object detection in low-light conditions with adaptive parallel networks. *Journal of Electronic Imaging.* 2025;34(1):013007. DOI: 10.1117/1.JEI.34.1.013007.

[27] Talaat FM, ZainEldin H. An improved fire detection approach based on YOLO-v8 for smart cities. *Neural Computing and Applications.* 2023;35(28):20939-20954. DOI: 10.1007/s00521-023-08809-1.

[28] Feng C, et al. Tood: Task-aligned one-stage object detection. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) 2021, 10 Oct. 2021, Montreal, QC, Canada.* 2021. p. 3490-3499. DOI: 10.1109/ICCV48922.2021.00349.

[29] Li X, et al. Generalized focal loss: Towards efficient representation learning for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2022;45(3):3139-3153. DOI: 10.1109/TPAMI.2022.3180392.

[30] Chen Y, et al. Domain adaptive faster R-CNN for object detection in the wild. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018, 18 Jun. 2018, Salt Lake City, UT, USA.* 2018. p. 3339-3348. DOI: 10.1109/cvpr.2018.00352.

[31] Xu L, et al. UPT-Flow: Multi-scale transformer-guided normalizing flow for low-light image enhancement. *Pattern Recognition*. 2025; 158: 111076. DOI: 10.1016/j.patcog.2024.111076.

[32] Li YJ, et al. Cross-domain adaptive teacher for object detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022, 18 Jun. 2022, New Orleans, LA, USA.* 2022. p. 7581-7590. DOI: 10.1109/cvpr52688.2022.00743.

[33] Mildenhall B, et al. Nerf in the dark: High dynamic range view synthesis from noisy raw images. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022, 18 Jun. 2022, New Orleans, LA, USA.* 2022. p. 16190-16199. DOI: 10.1109/cvpr52688.2022.01571.

[34] Cai Y, et al. Retinexformer: One-stage Retinex-based transformer for low-light image enhancement. *Proceedings of the IEEE/CVF International Conference on Computer Vision 2023, 01 Oct. 2023, Paris, France.* 2023. p. 12504-12513. DOI: 10.1109/iccv51070.2023.01149.

[35] Peng D, Ding W, Zhen T. A novel low light object detection method based on the YOLOv5 fusion feature enhancement. *Scientific Reports.* 2024;14(1):4486. DOI: 10.1038/s41598-024-54428-8.

[36] Yin X, et al. Pe-yolo: Pyramid enhancement network for dark object detection. *Proceedings of the International Conference on Artificial Neural Networks 2023, 22 Sep. 2023, Heraklion, Crete, Greece,* 2023. p. 163-174. DOI: 10.1007/978-3-031-44195-0_14.

[37] Jiang Z, et al. Low-illumination object detection method based on Dark-YOLO. *Journal of Computer-Aided Design & Computer Graphics.* 2023;35(3):441-451. DOI: 10.3724/SP.J.1089.2023.19354.

[38] Guo C, et al. Zero-reference deep curve estimation for low-light image enhancement. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020, 14 Jun. 2020, Seattle, WA, USA.* 2020. p. 1780-1789. DOI: 10.1109/cvpr42600.2020.00185.

[39] Mi A, et al. Rethinking zero-DCE for low-light image enhancement. *Neural Processing Letters.* 2024;56(2):93. DOI: 10.1007/s11063-024-11565-5.

[40] Li C, Guo C, Loy CC. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2021;44(8):4225-4238. DOI: 10.1109/TPAMI.2021.3063604.

[41] Jie Y, Liu Z. Reversed and fused Zero-DCE. *Proceedings of the International Conference on Computer Vision, Image and Deep Learning (CVIDL) 2024, 19 Apr. 2024, Zhuhai, China.* 2024. p. 887-891. IEEE. DOI: 10.1109/CVIDL62147.2024.10604009.

[42] Du Q, et al. A hybrid zero-reference and dehazing network for joint low-light underground image enhancement. *Scientific Reports.* 2025;15(1):10135. DOI: 10.1038/s41598-025-95366-3.

[43] Pan Q, Zhang Z, Tian N. Zero-reference generative exposure correction and adaptive fusion for low-light image enhancement. *Neurocomputing.* 2025; 636: 129992. DOI: 10.1016/j.neucom.2025.129992.

[44] Engin D, Genç A, Kemal Ekenel H. Cycle-dehaze: Enhanced CycleGAN for single image dehazing. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 18 Jun. 2018, Salt Lake City, UT, USA.* 2018. p. 825-833. DOI: 10.1109/cvprw.2018.00127.