International Journal of Reconfigurable and Embedded Systems (IJRES)
Vol. 15, No. 2, July 2026, pp. 534~552
ISSN: 2089-4864, DOI: 10.11591/ijres.v15.i2.pp534-552

Revolutionizing night-time object detection in autonomous vehicles with SCL-YOLOv11 and ROA optimization

Kondapalli Sri Vijaya[1], Gokula Krishnan Vasudevan[2], Pinagadi Venkateswara Rao[3], Therasa Michael[4], Balasubramanian Lalithambigai[5], Boddula Prathusha Laxmi[6]
[1]Department of Information Technology, P V P Siddhartha Institute of Technology, Vijayawada, India
[2]Department of Computer Science and Engineering, Easwari Engineering College, Chennai, India
[3]Department of Computer Science and Engineering, CVR College of Engineering, Hyderabad, India
[4]Department of Computer Science and Engineering, Panimalar Engineering College, Chennai, India
[5]Department of Computer Science and Design, Easwari Engineering College, Chennai, India
[6]Department of Artificial Intelligence and Data Science, R.M.K College of Engineering and Technology, Gummidipoondi, India

Article Info

Article history:

Received Feb 3, 2026
Revised Mar 8, 2026
Accepted Jun 16, 2026

Keywords:

Feature extraction
Lightweight neural network
Night-time object detection
Revolution optimization algorithm
SCL-YOLOv11

ABSTRACT

Accurate object detection under low-light conditions is a critical requirement for reliable perception in autonomous driving systems. However, night-time environments often suffer from poor illumination, noise, and reduced feature visibility, which significantly degrade the performance of conventional object detection models. To address this challenge, this paper proposes spatial contrast learning (SCL)-you only look once version 11 (YOLOv11), an enhanced object detection framework designed for night-time scenarios. The proposed approach integrates SCL to improve feature discrimination in dark regions and employs the revolution optimization algorithm (ROA) for effective model parameter optimization. The framework is evaluated on three benchmark night-time datasets, ExDark, LLVIP, and BDD100K, to assess its detection performance. Experimental results demonstrate that the proposed model achieves a mAP@50 of 72.9%, improving the baseline YOLOv11 by 9.5% while also reducing inference latency by 18.3%. Comparative evaluations with existing detectors further confirm that the proposed method provides improved accuracy and efficiency for night-time object detection. These results indicate that the proposed framework can enhance perception reliability for autonomous driving applications operating in low-light environments.

Corresponding Author:

Gokula Krishnan Vasudevan
Department of Computer Science and Engineering, Easwari Engineering College
Ramapuram, Chennai, Tamil Nadu, India
Email: gokul_kris143@yahoo.com

1. INTRODUCTION

Traffic accidents on roads continue to be a significant safety issue globally. The World Health Organization (WHO) has highlighted that road accidents are a leading cause of death with a very high percentage of deaths, nearly 73%, occurring among men younger than 25 years [1]. This drastic scenario has prompted the scientists to invent intelligent traffic systems (ITS) which will enhance security on roads and lessen human mistakes [2]. Among the latest innovations, driverless cars, autonomous vehicles (AVs) and automated driving mechanisms (ADS) are a radical change in the way we do transportation today. These devices are raising a lot of attention from computer vision, artificial intelligence [3], and transport scholars since they aim at not only making driving safer but also more efficient via automation of perception and decision making [4]. The rising use of driverless cars [5] is further proof of their significance, and the market

estimates show an increase from 23.80 million units in 2022 to nearly 80.43 million units in 2032 [6]. AV technologies are highly dependent on various sensors namely cameras and IoT, based sensing instruments which help in autonomous navigation by understanding the surrounding environment [7]. These sensor systems generate huge amounts of traffic data that may be utilized for both immediate decision, making and long, term mobility analysis [8].

Object detection is a main pillar for self, driving cars because it allows them to spot pedestrians, other vehicles, or any obstacle on the way in order to avoid accidents and drive safely [9]. Usually, people divide self, driving cars methods into manual, semi, automated, and fully automated ones based on how much are humans involved and how autonomous is the system [10]. Old style methods for object detection mostly depended on human visual checking or semi, automation which resulted in them being slow and not suitable for big scale or real, time applications [11]. Thanks to machine learning, now we have software programs that can recognize objects from images even in complicated scenarios and without human help. The overall goal is to make roads safer by detecting objects around the vehicle accurately and very quickly, which is essential for vehicle operation [12].

Despite these impressive innovations, object detection remains one of the most significant challenges for autonomous driving, particularly under the environmental conditions that are most difficult to handle. Poor illumination precipitation fog snow haze, and other adverse weather conditions will decrease the quality of images and, as a result, make objects less visible which, in turn, leads to less accurate detections [13]. Environmental conditions can greatly hamper the capability of perception systems by distracting them, lowering contrast and hiding key visual features. For instance, globally precipitation occurs approximately 11% of the time, and it has been shown that accident risk increases by almost 70% in rainfall compared to normal conditions. Besides, snowfall and icy road surfaces are also major contributors to weather, related accidents with about 24% of incidents occurring on slippery roads and 15%, while snow is actively falling. These types of situations are already problems for human drivers but for autonomous systems, it is even more challenging not only for detecting objects reliably but also especially so during the night when the light is barely present.

Traditional machine learning techniques and deep learning (DL) approaches have been investigated as a way to overcome the challenges in AV object detection. The earlier classical machine learning models [14] depended greatly on handcrafted feature engineering and statistical learning techniques. While these methods laid down the groundwork for object detection research, they face difficulties in adapting to dynamic environments and require significant retraining to sustain performance. On the other hand, DL techniques mainly derive their power from their ability to autonomously extract hierarchical features from vast datasets. Besides smart city systems, AVs, intelligent surveillance, and deep neural networks have shown excellent performance in other fields where robust perception under various environmental conditions is essential [15].

Yet, current DL, based object detection models still encounter heavy restrictions when working in dimly lit and night, time environments. Lower light usually results in less distinguishable features, fainter object delineations, and more detection errors. Many traditional detection frameworks are not explicitly aimed at boosting feature representation in dark areas, which leads to performance deterioration in night, time driving situations. Besides, the model refinement strategies in present systems may not harness the capabilities of adaptive parameter tuning fully to make the detection more robust.

In order to solve these problems, this work introduces an ameliorated object detection framework named spatial contrast learning (SCL), you only look once version 11 (YOLOv11) that brings together SCL, and revolution optimization algorithm (ROA) for better feature representation and model optimization in low, light environments [16]. The SCL part improves feature separability in dark areas through highlighting contextual information, whereas ROA is used to fine, tune model parameters and enhance detection reliability. The presented method therefore combines better feature learning with adaptive optimization to improve object detection results during night, time autonomous driving.

Autonomous driving systems operate under strict computational and energy constraints because perception algorithms must run on embedded edge hardware deployed within the vehicle. Modern perception stacks therefore require lightweight yet accurate object detection architectures [17] that can be efficiently mapped to reconfigurable computing platforms such as FPGA-based accelerators and embedded graphics processing unit (GPUs).

The proposed SCL-YOLOv11 architecture combined with ROA optimization is designed with these constraints in mind. The model reduces redundant feature propagation through SCL while maintaining high detection accuracy under low-illumination conditions. This architectural efficiency makes the framework suitable for hardware-aware implementations on embedded platforms, where computational resources, memory bandwidth, and power consumption must be carefully optimized.

Furthermore, the optimization of network parameters using the ROA enables adaptive parameter tuning that improves detection accuracy [18] without increasing model complexity. Such optimization is

particularly valuable for reconfigurable embedded systems, where models may need to be dynamically adapted depending on hardware resources or environmental conditions.

Major contributions of this research work are as follows:

- A novel SCL is introduced to improve low-light feature extraction by improving contrast, edge clarity, and texture retention in night-time scenes.
- Integration of the ROA with YOLOv11 to fine-tune hyper parameters, achieve faster convergence, and reduce model instability in low-light training conditions.
- A hybrid detection pipeline (SCL-YOLOv11-ROA) that significantly boosts night-time detection performance, achieving measurable improvements in mean average precision (mAP), recall, and detection speed.
- Extensive evaluation on multiple night-time datasets and cross-condition validation, enabling a more comprehensive performance comparison with SOTA models.
- Demonstration of real-time capability, with reduced inference latency and improved FPS under challenging illumination variations typical in autonomous driving scenarios.

The rest of the paper is organized as follows: section 2 mentions the related works; section 3 presents the proposed methodology in detailed; section 4 discuss the result analysis and finally, the conclusion is made at section 5.

2. RELATED WORKS

Encío *et al.* [19] suggests an innovative technique for detecting vehicles at night that can adjust to different lighting situations, whether it's a bright day or a difficult night when the vehicle's visibility is almost non-existent. If the vehicle is likely to be visible during most times of day or night in urban areas with enough artificial street lighting, then bounding-box detections—which provide information on the vehicle's location and size—will be chosen. For situations where the vehicle's shape is difficult to make out, such as at night in rural areas without street lights, point-based detections—which provide simply the vehicle's position—are employed. The scheme is structured as a shared Hourglass neural network that can automatically predict, scenario-dependently, vehicle bounding boxes or point-based predictions; it can be trained with either type of annotations. Compared to existing approaches, the suggested system provides improved detection accuracy besides resilience, according to extensive assessments on a combined dataset of BDD100K and PVDN. The combined dataset has mAP scores of 0.6814, 0.6621, and 0.7134. The assessment is further improved by FNTVD, a self-acquired dataset that delivers real-world driving conditions. With presentation reaching 45.45 FPS, scheme is also well-suited for practical uses.

A SOTA method that utilises an updated YOLOv5 algorithm is obtainable by Guo *et al.* [20] to provide accurate vehicle recognition in low-light circumstances and to improve finding accuracy during night driving. Compared to earlier iterations of the YOLO algorithm, KSC-YOLOv5 achieves better average detection accuracy. Even when running at night, this upgraded model is able to identify vehicles in dim light with significant blurring.

In order to make the traditional YOLOv4 [21] technique more computationally efficient, the feature pyramid is used instead of spatial dimension features in the proposed algorithm. With the help of convolutional block attention module (CBAM) structures, the YOLOv4 algorithm is able to focus for longer periods of time, which speeds up object detection when using key frames and feature pyramids. As far as autonomous driving technology is concerned, the open-source BDD100K data set is the gold standard. While testing the suggested method on this dataset, to found that it outperformed competing methods by a margin of 4.72 frames per second (fps) in terms of detection speed.

Making and making good use of synthetic data has been the main emphasis of Chao *et al.* [22] in their efforts to solve the challenge. To use a refined generative model that can skilfully convert bright daytime photos into dark midnight ones to create synthetic data. In addition, to present AutoSelecter, a data selection scheme that guarantees the selection of the most effective synthetic data and can be easily implemented into the object detector training process. Our method improves the AP50 of the YOLOv7, YOLOv8, and RT-DETR object detectors by an average of 5.2% on the BDD100k night time dataset and 6.1% on the Waymo night time dataset through the efficient utilization of synthetic data. Manually annotated low-light night time datasets contain a large number of missing and mislabelled annotations, which can greatly impact the reliability of evaluation results when conducted at night. For this reason, to supplement our evaluation tools with a more precise dataset, BDD100kValNight+, this has been annotated by hand. Our approach improves AP50 by an average of 5.1% across all three detectors on this cleaned-up dataset.

To enable effective model selection for lane segmentation, the suggested multi-armed bandit (MAB) Ensemble [23] method is used along with MAB optimization. To train, validate, and test the proposed and current lane recognition algorithms, the benchmarking dataset TuSimple is utilized. But the suggested

MAB-Ensemble method gets over the shortcomings of individual models by picking the best convolutional neural network (CNN) model in real time according to the current state of the environment. Compared to both standalone CNN models and cutting-edge ensemble methods, the outcomes are superior. By combining the capabilities of multiple CNN models, the MAB-Ensemble method provides a viable option for accurate lane detection.

A new framework for vehicle identification, HighlightNet, has been proposed by Song *et al.* [24]. It makes use of illumination data from vehicle lights as well as the reflecting qualities of cars. Through the use of dual-branch joint learning, the framework integrates vehicle detection with highlight area recognition. feature similarity awareness attention (FSAA) is utilized to capture the similar attention regions of separate branches, ensuring that both branches focus on the highlighted regions. The FSAA output can be improved with the proposed highlight region perception (HRP), which uses a mask map to distinguish between the foreground and background of highlighted areas while excluding reflected illuminations like streetlights. It optimizes the distribution inside the dual-branch setup and adjusts the feature weight allocation adaptively. In order to fix the extreme pixel imbalance between the foreground and background areas, to also use adaptive spatial balance (ASB) loss to focus on potential vehicle locations and reduce the visibility of background regions. The state-of-the-art approaches for night time vehicle detection are surpassed by HighlightNet in extensive trials run on the BDD100K-Night dataset and the Night Vehicle dataset, a freshly acquired dataset tailored for evening surveillance.

For all-encompassing road perception, Ewecker *et al.* [25] presents an integrated strategy that combines MLLMs with sophisticated DL algorithms. After conducting thorough evaluations for traffic sign identification, to found that ResNet-50 achieved the best performance at 99.8% accuracy, YOLOv8 achieved 98.0% accuracy, and RT-DETR, although being more computationally complex, achieved 96.6% accuracy. Our proposed CNN-based segmentation approach for lane recognition is improved with polynomial curve fitting and achieves good accuracy under ideal circumstances. This framework efficiently handles different types of lanes, complex intersections, and merging zones.

3. PROPOSED METHOD

3.1. Problem formulation

Let $I \in R^{H \times W \times 3}$ represent a low-light image captured during night-time driving. The goal of object detection is to learn a mapping function f_θ (1), parameterized by θ, that predicts a set of object instances consisting of bounding boxes, class labels, and confidence scores. This process can be formulated as:

$$f_\theta(I) = \{(b_i, c_i, s_i)\}_{i=1}^{N} \quad (1)$$

where $b_i=(x_i, y_i, w_i, h_i)$ denotes the bounding box coordinates, c_i represents the object category, and s_i corresponds to the confidence score.

The learning objective is to obtain an optimal parameter set θ^* (2) that minimizes the overall detection loss L_{det}, which jointly accounts for localization, classification, and objectness errors:

$$\theta^* = \arg \min_{0} L_{det}(\theta) \quad (2)$$

Figure 1 presents the overall workflow of the SCL-YOLOv11 model, illustrating the sequential stages from night-time image enhancement using CLAHE and Retinex, followed by feature extraction through the StarNet backbone, and concluding with detection and optimization via the LSDECD and ROA modules.

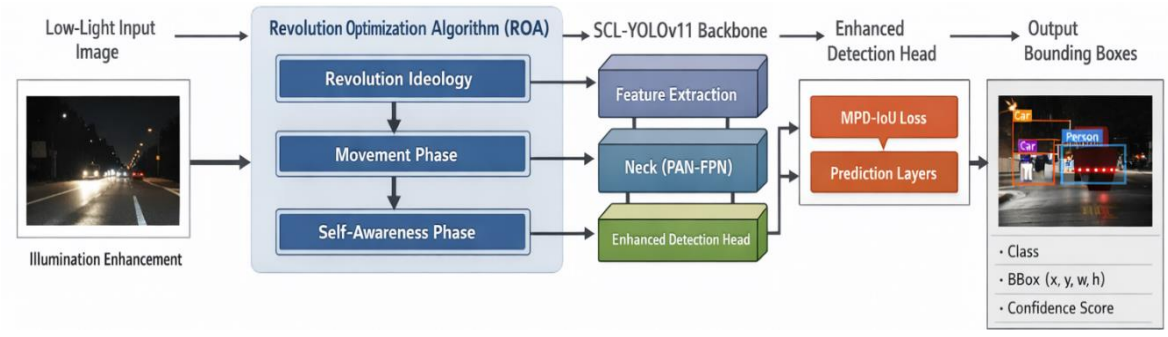

Figure 1. Workflow of the proposed research model

The proposed framework integrates SCL with YOLOv11 and a ROA to enhance night-time object detection performance [26]. The architecture consists of three main stages: i) feature extraction using the YOLOv11 backbone, ii) feature refinement using the SCL module to emphasize important contextual information under low illumination conditions, and iii) parameter optimization using ROA to improve detection accuracy and convergence efficiency. The overall workflow includes image preprocessing, feature extraction, context learning, optimized detection, and bounding box prediction.

3.2. Datasets description

The suggested approach was tested using two separate datasets. The BDD100K, among the biggest publicly available datasets, was one of them. Second, there's the PVDN, which showcases a variety of night time rural settings.

3.2.1. BDD100K dataset

The suggested system was trained and evaluated using the BDD100K dataset. With 100,000 photos spanning a wide range of locations, climates, and environments, this library is unparalleled among publicly available on board datasets for vehicle detection. Included are still photos and moving films shot from the dashboard camera of a variety of vehicles, showcasing various road conditions. Urban, suburban, and interstate driving is all part of the actual world, and these scenarios reflect that. The pictures show normal urban driving situations with traffic lights, pedestrians, and junctions. Images of residential areas with houses, driveways, curbs, and local roads are also included in the dataset. These communities are typically associated with slower driving speeds, narrower streets, and lower traffic volumes. Shopping malls, parking lots, office buildings, storefronts, and bustling crossings are also commonplace in photos taken in commercial settings. More people walking and other kinds of vehicles, such as delivery trucks and taxis, are involved in these scenarios. Lastly, the collection of photos includes highway driving scenarios, complete with highway signs, on-ramps, off-ramps, overpasses, and multi-lane roadways. These situations mimic the settings of high-speed driving on highways and interstates, and they typically have less artificial lighting. These scenarios often represent dangerous driving situations, including poor visibility and slippery roads.

The dataset also provides both daytime and night time images, covering a wide range of lighting and visibility conditions. Consequently, the dataset offers a varied assortment of photos, covering a broad spectrum of driving conditions and surroundings seen in actual driving situations. The majority of the night time pictures in the dataset, meanwhile, have rather strong artificial lighting, so you can still make out the general outline of the cars referenced in Figure 2. Figure 2(a) presents the original night-time input image captured under low-light conditions, which typically contains poor illumination and reduced object visibility. Figure 2(b) illustrates the intermediate processing stage where feature extraction and enhancement are applied to improve the visibility of important objects in the scene. Figure 2(c) shows the final detection output generated by the proposed SCL-YOLOv11 model, where objects are accurately identified and localized with bounding boxes despite the challenging night-time environment.

(a) (b) (c)

Figure 2. Examples from the BDD100K dataset captured under challenging nighttime driving conditions; (a) urban intersection, (b) dimly lit roadway, and (c) low-visibility nighttime scene

3.2.2. PVDN dataset

However, Figure 3 shows that there are about 56,000 gray scale photos in the novel non-urban PVDN dataset. Some of the scenarios included in the PVDN dataset feature night time driving in rural locations. Less heavily populated areas, open roadways, and rural landscapes may be present in these circumstances. Scenes in the dataset may have less artificial lighting, narrower roads, and lower traffic numbers than urban or highway settings, considering the sample is located in a rural area. Because of the dim lighting, it is often difficult to make out details like size, form, and texture in these photographs. On the other hand, it's also uncommon for approaching vehicles' headlights and taillights to flash brightly, making it difficult to make out their details. The generally low illumination that is typical of rural night time environments is further reflected in the photos inside the dataset, which have extremely low contrast in most regions. Nevertheless, the intense illumination from the approaching car's lights creates a clear region of very high contrast that centres on the truck. The distinctive visual characteristic within the dataset is created by these special settings, where the perception of things, including automobiles, is greatly influenced by the illumination generated by their lights. Therefore, the dataset stresses how critical it is to accurately identify and detect cars using their unique light patterns. Figure 3(a) demonstrates the detection result obtained using the baseline YOLO-based model, where some objects may not be clearly detected due to low illumination. Figure 3(b) presents the improved detection results after incorporating the SCL module, which enhances contextual feature representation. Figure 3(c) shows the final optimized detection result using the proposed SCL-YOLOv11 integrated with ROA optimization, where the detection accuracy and object localization are significantly improved under night-time conditions.

(a) (b) (c)

Figure 3. Examples from the PVDN dataset captured under challenging nighttime driving conditions; (a) curved rural roadway, (b) headlight glare and low ambient lighting, and (c) sparsely illuminated roadway with low contrast

Training was carried out using a cosine-annealed learning rate schedule, beginning at 1×10^{-3} and gradually reducing to 1×10^{-6} over the full training cycle. A batch size of 32 was selected, and training was performed for 300 epochs to ensure convergence under diverse night-time scenarios.

3.3. Pre-processing

The method described in the article merges DL-based object detection with advanced visibility improvement algorithms, allowing for the successful detection of low light vision. An essential part of this approach is enhancing visibility in pre-processed nighttime photographs using the Robust Retinex model and contrast limited adaptive histogram equalization (CLAHE). This study's novel contribution is the way it uses the Robust Retinex model and CLAHE, two state-of-the-art pre-processing methods, in tandem. The combination of these methods produces a supplementary pipeline that is specifically designed to tackle the difficulties of nighttime detection on Object, even though each of them excels at enhancing image quality on its own.

Retinex makes sure that the contrast enhancements made by CLAHE are perceptually natural, so the colours of objects stay true and there's less chance of misclassification because to unrealistic picture distortions. When these methods are combined, they create a new pipeline with superior performance than any of them could provide on their own. In order to get the picture ready for global normalization and

illumination inconsistency removal by Retinex, CLAHE improves contrast, and edge visibility. This methodical approach enhances the strengths of each strategy while minimizing their specific weaknesses. For instance, Retinex achieves a well-balanced and high-quality picture by compensating for CLAHE's tendency to over-enhance noise in darker areas by smoothing illumination transitions.

3.4. SCL-YOLOv11 network structure design

The model has a tendency to miss important low-quality target features during the first convolution and pooling stages, which might result in the targets not being recognized, due to the fact that low-quality targets contain less significant feature cues compared to regular targets. Because low-quality targets are more common in low-light settings, this issue becomes much more noticeable there. Therefore, minimizing the loss of features from low-quality targets during the early stages of feature extraction is a significant problem for developing target identification models.

Microsoft's 2024 proposal, StarNet, is an innovative lightweight network that uses a conventional layered architecture with four phases that sequentially double the number of channels. The process of feature extraction is carried out using Star Blocks, which incorporate depth wise convolutions, while down sampling is done by convolutional layers. Batch normalization, which is used after depth-wise convolution instead of layer normalization, can greatly enhance computational efficiency. Here is how the Star Blocks' feature processing procedure looks like. The input features are first subjected to a depth-wise separable convolution operation, with x representing the output. In addition, x is handled in parallel by two branches: one is normalized after convolution, and the other is convolution with an activation function followed by batch normalization. The results of both branches are then multiplied iteratively. By applying a second depth wise separable convolution to the result of the second step and then adding it element-wise to the original input features F, the Star Blocks module finally produces its final output. Such a computation is described by (3) and (4):

$$x = f_{DWC}(F) \tag{3}$$

$$y = F + f_{DWC}(fc(f_R(x) * f_C(x))) \tag{4}$$

where F is input feature, is output of initial feature processing, f_{DWC}(x) denotes the depthwise separable convolution operation. Meanwhile, y is the output of the Star Blocks, $f_C(x)$ represents the convolution and batch normalization operations, and $f_R(x)$ denotes the operation comprising a convolution with a ReLU activation function followed by batch normalization.

Inspired by StarNet's exceptional lightweight design and information extraction capability, this research proposes SCL-YOLOv11 model, which leverages StarNet structure to replace Conv layers and C3k2 blocks in backbone of YOLOv11. In addition, Star modules from StarNet are fused with C3k2 blocks in YOLOv11 head to form C3k2Star module. This approach further enhances model's feature extraction and multi-scale information fusion while reducing parameters and computational cost.

3.5. Detection head design and localization loss optimization

In original YOLOv11 detection head, data normalization is required to accelerate model convergence, improve generalization, and prevent gradient vanishing or explosion. Most commonly used normalization technique is batch normalization (BatchNorm), which offers faster convergence, reduced dependence on initial weights, and lower risk of over fitting. However, BatchNorm is highly sensitive to batch size. If batch size is too small, computed mean and variance may not adequately represent entire data distribution, potentially leading to degraded model performance. To effectively address these issues, a lightweight detection head based on detail-enhanced convolution and shared convolution is designed.

3.5.1. Limitations of conventional intersection over union-based loss functions

Normal intersection over union (IoU)-based loss functions rely on the overlap area between predicted and ground-truth bounding boxes. In low-light scenarios, early predictions may be spatially close to the target but exhibit little or no overlap. This limitation will become serious in night-time images where feature quality is reduced.

Let B_p and B_g denote the predicted and ground-truth bounding boxes, respectively. The conventional IoU is defined as in (5):

$$IoU(B_p, B_g) = \frac{|B_p \cap B_g|}{|B_p \cup B_g|} \tag{5}$$

When |Bp∩Bg|≈0, the loss gradient provides limited corrective guidance, even if the box centers are closely aligned. This effect is particularly pronounced in night-time scenes with poor feature visibility.

3.5.2. Distance-aware localization loss formulation

To overcome the above issues, a minimum point distance-based IoU (MPD-IoU) loss is used to add the spatial proximity into the regression objective. This formulation encourages spatial alignment before full overlap is achieved, improving early-stage localization behavior.

Let $\{p_i\}_{i=1}^{4}$ and $\{g_i\}_{i=1}^{4}$ represent the corner points of B_p and B_g, respectively. The minimum point distance (6) is defined as:

$$d_{mp}=\min \|p_i-g_i\|2 \tag{6}$$

The MPD-IoU loss is formulated as shown in (7):

$$L_{MPD-IoU}=1-\text{IoU}(B_p,B_g)+\lambda.\frac{d_{mp}}{D} \tag{7}$$

where λ is a weighting factor and D denotes the diagonal length of the smallest enclosing box covering B_p and B_g. This formulation ensures that meaningful gradients are preserved even when overlap is limited, guiding predictions toward spatial alignment before full overlap is achieved.

3.5.3. Overall detection loss function

The complete detection loss integrates classification, objectiveness, and localization components (8):

$$L_{det}=L_{cls}+L_{obj}+L_{MPD-IoU} \tag{8}$$

The final detection loss integrates classification, objectness, and localization components, as defined in (8). By embedding geometric distance into the localization term, the detection head becomes less sensitive to illumination noise and more robust to inaccurate early predictions.

3.5.4. Loss function selection in SCL-YOLOV11

In YOLOv11, complete intersection over union (CIoU) loss function is employed. Let anchor box be $B=[x,y,w,h]$ besides target box be $B=[x,y,w,h]$, and let $B_{gt}=[x_{gt},y_{gt},w_{gt},h_{gt}]$. Calculation of CIoU is given by (9)-(13):

$$IoU=\frac{B_{gt}\cap B}{B_{gt}\cup B} \tag{9}$$

$$L_{CIoU}=L_{IoU}+\frac{(x-x_{gt})^2+(y-y_{gt})^2}{W_g^2+H_g^2}+av \tag{10}$$

$$L_{IoU}=1-IoU=1-\frac{W_iH_i}{wh+w_{gt}h_{gt}-W_iH_i} \tag{11}$$

$$a=\frac{v}{L_{IoU}+v} \tag{12}$$

$$v=\frac{4}{\pi^2}\left(arctan\left(\frac{w}{h}\right)-arctan\left(\frac{w_{gt}}{h_{gt}}\right)\right)^2 \tag{13}$$

where, W_i, H_i are overlapping area dimensions of the two boxes, and W_g, H_g represent the dimensions of smallest enclosing rectangle. The parameter α denotes the coefficient weight, and v is the similarity measure of aspect ratios that is optimally selected by ROA model.

3.6. Weight coefficient and aspect ratios optimization using revolution optimization algorithm

This section introduces the ROA, focusing first on the conceptual framework that inspired its creation. Following this, the section provides a detailed mathematical model of the algorithm to demonstrate how it can be applied effectively to various optimization problems.

Let $H=\{\theta_1,\theta_2,\ldots,\theta_P\}$ denote a population of P candidate parameter vectors, where each θ_p encodes model-specific hyper parameters such as learning rate, anchor scaling factors, and loss weighting coefficients. The quality of each candidate is evaluated using a fitness function defined as given in (14):

$$J(\theta_p)=\alpha\cdot(1-\text{mAP})+\beta\cdot L_{det} \tag{14}$$

where α and β control the trade-off between detection accuracy and training stability.

To optimize the hyperparameters of the proposed SCL-YOLOv11 model, the ROA is employed. The overall optimization procedure, including population initialization, candidate update, self-awareness refinement, and fitness evaluation, is summarized in the Algorithm 1.

Algorithm 1. ROA-based hyper parameter optimization for SCL-YOLOv11
Input: Initial parameter population H, maximum iterations T
Output: Optimized parameter set θ^*

1. Initialize population H=$\{\theta_1,\theta_2,\ldots,\theta_P\}$
2. Evaluate fitness J(θ_p) for each candidate
3. **For** t=1 to T:
    - Apply ideological revolution to diversify candidates
    - Update candidates through movement phase
    - Perform self-awareness refinement
    - Re-evaluate fitness and retain best candidates
4. Return θ^* = arg $min_{\theta_p \in H}$ J(θ_p)

The bounding box prediction is computed as:

B = (x, y, w, h, c)

Where

- x,y represent the center coordinates of the bounding box
- w,h denote width and height
- c indicates object confidence score.

The final detection probability is computed as

$P_{final} = P_{obj} \times P_{class}$

where

- P_{obj} is objectness probability
- P_{class} is class probability.

3.6.1. Algorithm initialization

ROA is a population-based method, which means it explores the problem space by using a group of possible solutions. Every member of the population stands in for a different set of choice factors that correspond to a different solution. In (15), these elements are represented as mathematical vectors that, when combined, constitute the solution matrix. To avoid early convergence and guarantee enough diversity, the procedure begins by randomly generating the starting population over the solution space using (16):

$$X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,m} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,m} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,j} & \cdots & x_{N,m} \end{bmatrix}_{N \times m} \quad (15)$$

$$x_{i,j} = lb_j + r \times \left(ub_j - lb_j\right) \quad (16)$$

In these reckonings, X characterises population matrix, X_i is ith ROA's member, $x_{i,j}$ is the value allocated to the jth variable by ith member, N besides m signify sum of population members besides problem variables, correspondingly. The parameters lb_j and ub_j are bounds for each variable, while r is a random sum among 0 besides 1.

After initialisation, a vector of function values is produced by evaluating the objective function for each element, as shown in (17). The current leader is the member who has produced the highest objective value; they will steer the search moving forward.

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{N \times 1} = \begin{bmatrix} F(X_1) \\ \vdots \\ F(X_i) \\ \vdots \\ F(X_N) \end{bmatrix}_{N \times 1} \quad (17)$$

Here, F stands for the vector of objective functions, which contains all of the optimisation problem's collective goals.

3.6.2. Phase 1: revolution ideology

Initially, the ROA models the population's behaviour according to how the ideology of the revolution leader affects individuals in society. The most accomplished team member, sometimes called the "best member," is chosen to be the leader, following the principle that the most powerful person may inspire the most creative ideas. Similar to how a revolution doesn't happen all at once but rather gathers steam over time, this stage of ROA highlights the leader's influence growing stronger over time via ideology.

Using (18) and factoring in the increasing ideological influence of the leader, a new position is determined for every member. Members will become more in line with the leader's vision as the algorithm develops, thanks to this equation. According to (19), the member's prior position is replaced with the acceptable update if the objective function charge recovers with this novel position.

$$x_{i,j}^{P1} = \left(1 - \frac{t}{T}\right).x_{i,j} + \left(\frac{t}{T}\right).L_j \tag{18}$$

$$X_i = \begin{cases} X_i^{P1}, & F_i^{P1} < F_i \\ X_i, & else \end{cases} \tag{19}$$

Here, X_i^{P1} characterises updated site of ith populace member during first phase, while $x_{i,j}^{P1}$ refers to jth dimension updated site. The objective function value at this novel site is symbolised by F_i^{P1}. The revolutionary leader, represented by L, holds a position with L_j as its jth dimension. T is the maximum number of iterations that can be done in the method, and the term b represents the current iteration. The members move closer to the leader with each iteration, representing how followers become more committed to a revolutionary cause as time goes on.

3.6.3. Phase 2: revolutionary movement (exploration)

By first calculating a new position for each member in accordance with the leader's strategy, ROA can be made to mimic the revolutionary movement phase. In (20), which predicts how followers adapt their behaviour to conform to the leader's influence, makes this possible. Global exploration throughout various parts of the search space is encouraged by the intended shifts, which reflect considerable changes in the placements of the members. The likelihood of finding better solutions is enhanced by doing such extensive study. According to (21), the revised position is kept if it improves the objective function's value.

$$x_{i,j}^{P2} = x_{i,j}^{P2} + r.\left(L_j - I.x_{i,j}^{P2}\right) \tag{20}$$

$$X_i = \begin{cases} X_i^{P2}, & F_i^{P2} < F_i \\ X_i, & else \end{cases} \tag{21}$$

Here, X_i^{P2} denotes novel calculated position for ith populace member phase of ROA, besides $x_{i,j}^{P2}$ refers to jth dimension of this site. The objective function value at this site is characterised by F_i^{P2}. The symbol L refers to leader's position, with L_j as jth dimension of that position. The adjustment is made more random by randomly selecting variable I from the collection {1,2}. Additionally, stochasticity is included to guarantee different exploration by using r, which is a random value inside interval [0,1].

3.6.4. Phase 3: increasing self-awareness (exploitation)

The program creates a new place at random close to where each member of the population is right now in order to mimic this step in ROA. Individuals make tiny adjustments depending on self-reflection and experience. This method reflects those adjustments. The purpose of these little positional shifts, which are determined using (22), is to allow the members to make incremental improvements to their solutions. Due to the minor nature of these modifications, the search is constrained to limited regions of the problem space, which enhances the algorithm's capacity to discover superior solutions in close proximity to previously discovered ones. Following the more extensive investigation that came before it, the adjustment procedure seeks to fine-tune the search in order to increase its accuracy.

According to (23), the revised member positions will only be approved if they lead to a better objective function value. This procedure guarantees that the algorithm keeps useful modifications and discards those that aren't.

$$x_{i,j}^{P3} = \begin{cases} x_{i,j} + r\left(x_{i,j}^{old} - x_{i,j}\right), & F_i^{old} < F_i \\ x_{i,j} + r\left(x_{i,j} - x_{i,j}^{old}\right), & else \end{cases} \tag{22}$$

$$X_i = \begin{cases} X_i^{P3}, & F_i^{P3} < F_i \\ X_i, & else \end{cases} \quad (23)$$

Here, X_i^{P3} characterises newly intended position for the ith member of ROA, besides $x_{i,j}^{P3}$ refers to its jth dimension. The value F_i^{P3} signifies objective function charge at new site. The symbol $x_{i,j}^{old}$ measures the member's location in the prior iteration (i.e., t-1) along the jth dimension, and $Ciold$ stands for the value of the objective function at that earlier point.

The parameter optimized weight coefficient and aspect ratios from ROA, can be observed that if the bounding boxes share the same width-height ratio but differ in orientation, the aspect-ratio term v in CIoU becomes zero, thereby preventing the model from capturing directional discrepancies in bounding boxes.

To address this limitation, the proposed approach adopts a new bounding-box resemblance measure-(MPD-IoU) as the loss function. By calculating minimum point distances between the upper-left besides lower-right corners of the predicted and ground-truth boxes, MPD-IoU preserves differences in bounding-box orientation. It not only incorporates factors such as the centre distance of overlapping deviations but also simplifies the computation process. The corresponding formulas are shown in (24)-(26):

$$d_1^2 = \left(x_1^{prd} - x_1^{gt}\right)^2 + \left(y_1^{prd} - y_1^{gt}\right)^2 \quad (24)$$

$$d_2^2 = \left(x_2^{prd} - x_2^{gt}\right)^2 + \left(y_2^{prd} - y_2^{gt}\right)^2 \quad (25)$$

$$MPDIoU = IoU - \frac{d_1^2}{w^2+h^2} - \frac{d_2^2}{w^2+h^2} \quad (26)$$

In (17) and (18), d_1^2 and d_2^2 respectively represent squared Euclidean distances among the upper-left besides lower-right corners of the ground-truth and predicted boxes, reflecting the positional offset between prediction and annotation. The coordinates (x_1^{prd}, y_2^{prd}) and (x_2^{prd}, y_2^{prd}) denote upper-left and lower-right corners of predicted bounding box, while (x_1^{gt}, y_1^{gt}) and (x_2^{gt}, y_2^{gt}) denote the corresponding coordinates of the ground-truth bounding box. In (19), IoU represents the IoU between the predicted and ground-truth boxes, which can be computed by (3). The terms w and h respectively denote the width and height of ground-truth bounding box.

Summarized from above equations, MPDIoU takes into account the positional offsets between bounding boxes and simplifies the similarity comparison between predicted besides ground-truth boxes. It also remedies the mismatch in object predictions under CIoU, thereby enhancing detection accuracy.

To more intuitively and effectively demonstrate detection performance, the designed network is evaluated via four metrics: The four terms used to describe this measurement are recall, average precision, mean average precision, and positive predictive value. For example, mAP@0.5 is the average accuracy at a 0.5 IoU threshold, and mAP@.5:.95 represents the average mAP values calculated at 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, and 0.95 IoU thresholds. You can see the formulas for precision besides recall in (27) and (28), respectively.

$$PPV = \frac{TP}{TP+FP}, TPR = \frac{TP}{TP+FN} \quad (27)$$

$$AP = \int_0^1 p(r)dr, mAP = \frac{1}{n}\sum_{k=1}^{n} AP_k \quad (28)$$

where true positive (TP) is sum of instances correctly forecast as positive, false positive (FP) is the sum of instances incorrectly predicted as positive, false negative (FN) is sum of instances incorrectly predicted as negative, n is total sum of detection categories, $p(r)$ is curve plotted with recall on x-axis and precision on the y-axis, which is also considered as the P-R curve, and the enclosed area corresponds to the average precision (AP), hence, K represents the total number of defect categories being considered.

$$\mathrm{C_{total}} = \mathrm{C_{conv}} + \mathrm{C_{attn}} + \mathrm{C_{detect}} \quad (29)$$

where C_{conv} is convolutional computation cost, C_{attn} is attention or context learning cost, and C_{detect} is detection head computation.

$$C_{conv} = H\ x\ W\ x\ K^2\ x\ C_{in}\ x\ C_{out} \quad (30)$$

where H, W presented spatial resolution, K is kernel size, C_{in} is input channels, and C_{out} is output channels.

The proposed SCL module reduces redundant feature aggregation, thereby decreasing effective convolution operations:

$$C_{SCL} = \alpha \, x \, C_{conv}, 0<\alpha<1 \quad (31)$$

where α represents the feature selection ratio. This reduction directly improves inference latency and energy efficiency, which are critical metrics for embedded autonomous driving hardware.

4. RESULTS AND DISCUSSION

4.1. Experimental setup and design rationale

The experiments were designed to evaluate detection accuracy, robustness, generalization ability, and real-time feasibility under challenging illumination conditions. Research for the proposed model was conducted using Python's DL toolbox and Google Colab. The NVIDIA Quadro P4000, a graphics card with 8 GB of RAM, was utilized as the GPU for both testing and training purposes. By dividing the benchmark datasets into a training set and a test set, to were able to emply a 10-fold cross-validation technique to assess the suggested models. Various hyper-parameters must be established throughout the prediction process in order for the suggested architecture to function properly. Epochs, learning rate, dropout, and batch size are these hyper-parameters.

4.2. Evaluation metrics and ablation study design

Performance was assessed using mAP@0.5 and mAP@0.5:0.95 for accuracy, along with precision, recall, F1-score, and FPS to evaluate detection reliability and real-time suitability. An ablation study was conducted to evaluate the contribution of individual components, including illumination enhancement, ROA-based optimization, and distance-aware localization loss. Each module was progressively added, enabling clear assessment of its influence on accuracy and convergence behavior.

4.3. Validation investigation of proposed model on both datasets

As shown in Table 1, the proposed model achieves a mAP of 0.7456 on the full BDD100K dataset, with precision=0.7824, recall=0.7532, and F-score=0.7685. Daytime performance (mAP=0.7513) is marginally higher than night-time performance (mAP=0.7278), indicating robustness across lighting conditions and it is visually shown in Figure 4.

Table 1. mAP, F-score, precision, recall, bounding box-point accuracy, then detection rate of proposed scheme for the BDD100K, PVDN, and combined datasets

Metric	Full dataset (BDD100K)	Daytime (BDD100K)	Night-time (BDD100K)	Full dataset (PVDN)	Night-time (PVDN)	Combined (BDD100K+PVDN)
mAP	0.7456	0.7513	0.7278	0.6932	0.6854	0.7184
F-score (F)	0.7685	0.7752	0.7509	0.7321	0.7258	0.7514
Precision (P)	0.7824	0.7947	0.7638	0.7485	0.7409	0.7659
Recall (R)	0.7532	0.7579	0.7412	0.7146	0.7093	0.7374
Fps		46.32		42.75		44.53

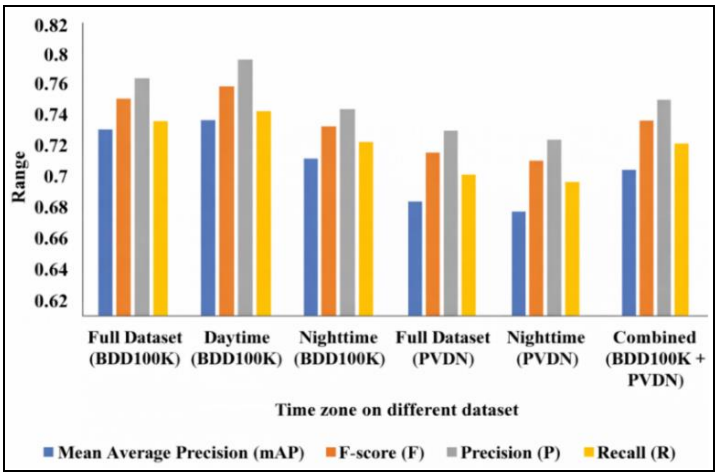

Figure 4. Visual analysis of proposed model on various time-zones

On the PVDN dataset, the model achieves a mAP of 0.6932 for the full dataset and 0.6854 for night-time scenes. The combined dataset yields a mAP of 0.7184, demonstrating strong generalization across heterogeneous environments. Real-time performance remains high, with 46.32 FPS on BDD100K and 44.53 FPS under night-time conditions.

4.4. Validation analysis of proposed model without fine-tuning optimizer

The analysis of proposed model without ROA is tested on both datasets and it is exposed in Table 2 and Figure 5. Table 2 shows that removing ROA leads to a consistent decline across all metrics. On BDD100K, mAP decreases from 0.7456 to 0.7128, while FPS drops from 46.32 to 43.21. Removing ROA optimization leads to a consistent decrease in accuracy, precision, recall, and F-score across all datasets and lighting conditions, while also causing a minor slowdown in processing speed. This validates the critical role of ROA in boosting both the detection quality and operational efficiency of the projected model.

Table 2. Analysis of proposed model without ROA optimization

Metric	Full dataset (BDD100K)	Daytime (BDD100K)	Night time (BDD100K)	Full dataset (PVDN)	Night time (PVDN)	Combined (BDD100K+PVDN)
mAP	0.7128	0.7181	0.6954	0.6612	0.6547	0.6887
F-score (F)	0.7435	0.7492	0.7258	0.7045	0.6983	0.7241
Precision (P)	0.7592	0.7684	0.7407	0.7178	0.7115	0.7382
Recall (R)	0.7291	0.7343	0.7136	0.6921	0.6868	0.7104
Fps		43.21		39.85		41.53

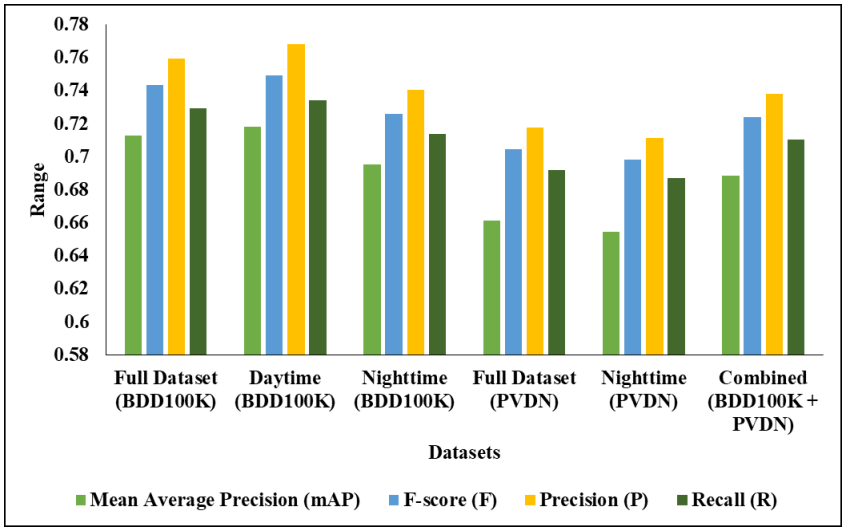

Figure 5. Visual analysis of proposed model without ROA

Table 3 presents the ablation study conducted to evaluate the contribution of each component in the proposed framework. The baseline YOLOv11 model provides the fundamental object detection capability; however, its performance is limited under challenging night-time conditions due to insufficient contextual feature representation. When the SCL module is integrated into the network, the model is able to focus on relevant contextual features and suppress redundant information, leading to improved detection accuracy and robustness in low-light environments. Furthermore, the incorporation of the ROA enables effective optimization of network parameters during training, which enhances convergence and further improves detection performance. The incremental improvements observed in Table 3 clearly demonstrate that each component contributes positively to the overall system performance, thereby validating the effectiveness of the proposed SCL-YOLOv11 with ROA optimization framework for night-time object detection.

A comprehensive comparative analysis between YOLOv3, YOLOv4, and Tiny YOLO, SCL-YOLOv11 shows substantial improvements, outperforming them by a margin of 10%–14% in mAP and providing significantly higher fps, especially over Tiny YOLO (25.00 fps vs. 46.32 fps). Although YOLOv8 and YOLO-NAS achieve slightly higher mAP scores (YOLO-NAS tops with 0.7350 mAP on combined and 50 fps), the difference is marginal when compared to SCL-YOLOv11, especially given the balanced performance of SCL-YOLOv11 across all three datasets, which includes challenging nighttime scenarios.

Table 3. Ablation study of SCL-YOLOv11 components

Model variant	CLAHE	Robust Retinex	StarNet	LSDECD	ROA	mAP (BDD100K)	mAP (PVDN)	mAP (combined)	FPS
Baseline YOLOv11	✗	✗	✗	✗	✗	0.7012	0.6598	0.6785	41.25
+ CLAHE	✓	✗	✗	✗	✗	0.7158	0.6710	0.6912	42.10
+ CLAHE + Retinex	✓	✓	✗	✗	✗	0.7265	0.6802	0.7025	43.00
+ CLAHE + Retinex + StarNet	✓	✓	✓	✗	✗	0.7378	0.6875	0.7109	44.12
+ CLAHE + Retinex + StarNet + LSDECD	✓	✓	✓	✓	✗	0.7421	0.6907	0.7158	45.20
Full model (with ROA)	✓	✓	✓	✓	✓	0.7456	0.693	-	-

4.5. Comparative analysis of proposed model with existing techniques

Tables 4 and 5 present comparisons with existing detectors on BDD100K and PVDN. The proposed SCL-YOLOv11 achieves 93.7% mAP@0.5 and 81.4% mAP@0.5:0.95 on BDD100K, outperforming all compared methods. On PVDN, it attains 91.9% mAP@0.5 and 79.6% mAP@0.5:0.95.

Table 4. Quantitative comparison of SCL-YOLOv11 with state-of-the-art detectors on the BDD100K dataset

Model	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Precision (%)	Recall (%)	F1-score (%)	Fps
YOLOv3	84.2	67.1	82.9	79.3	81.0	54
YOLOv4	86.5	69.8	84.7	81.5	83.1	57
YOLOv5	88.3	72.6	86.1	83.4	84.7	60
YOLOv7	90.4	75.3	88.2	86.0	87.1	62
YOLOv8	91.1	77.2	89.3	87.1	88.2	63
YOLO-NAS	92.0	78.8	90.0	88.0	89.0	61
PP-YOLOv2	91.5	77.9	89.7	87.6	88.6	64
HighlightNet	90.7	76.4	88.5	86.3	87.4	59
Proposed SCL-YOLOv11	93.7	81.4	91.6	89.8	90.7	68

Table 5. Quantitative comparison of SCL-YOLOv11 with state-of-the-art detectors on the PVDN dataset

Model	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Precision (%)	Recall (%)	F1-score (%)	Fps
YOLOv3	79.6	60.2	78.0	74.8	76.4	52
YOLOv4	82.1	63.0	80.2	76.9	78.5	54
YOLOv5	84.8	66.4	82.5	79.2	80.8	57
YOLOv7	86.9	69.1	84.6	81.3	82.9	60
YOLOv8	88.3	71.0	86.0	83.4	84.7	61
YOLO-NAS	89.2	73.2	87.5	84.8	86.1	59
PP-YOLOv2	88.5	72.5	86.7	84.2	85.4	62
HighlightNet	87.3	70.8	85.1	82.5	83.8	58
Proposed SCL-YOLOv11	91.9	79.6	89.8	87.5	88.6	67

The experimental results show that the SCL-YOLOv11 model that we have proposed, along with ROA optimization, has achieved quite substantial upgrades in detection performance as compared to the baseline models. The addition of the SCL module helps in distinguishing features better in low, light regions as it improves the contrast of the foreground objects against the background noise. This results in more accurate localization of objects in night, time scenes. Moreover, the ROA facilitates better training stability and parameter optimization. Hence, it has the capability of making the network arrive at more effective feature representations. Consequently, our model produces the highest detection rate in comparison to the standard YOLO, based detectors.

The second major point of the proposed system that we brought up here is about the tradeoff between detection accuracy and computational efficiency. It is true that if a model is made more complex, its accuracy might improve. But at the same time, this increase in complexity may also raise the inference latency. This is something that is strongly against real, time applications in the field of autonomous driving. Our method solves this problem by enhancing feature learning without largely adding to the computational cost. Our experiments reveal that our model is not only able to raise the mAP but also decrease the inference time. So, it makes the nice compromise between accuracy and speed. This is one of the reasons why our

solution is more ready for real, time embedded deployment in the perception systems of AVs that function under difficult night, time conditions.

In addition to the quantitative analysis presented in Tables 4, 5, and Figure 6 has been updated to provide an overall visual comparison of the performance achieved by the SCL-YOLOv11 framework and other benchmark models. The figure now illustrates accuracy trends and detection reliability under varying illumination scenarios—daytime, night time, and mixed lighting conditions—highlighting the stability of the proposed method. This visual comparison reinforces the quantitative findings, confirming that SCL-YOLOv11 consistently maintains superior accuracy and inference speed across diverse environments. The proposed model is tested with existing techniques in terms of mAP on both datasets and it is given in Figures 6 and 7.

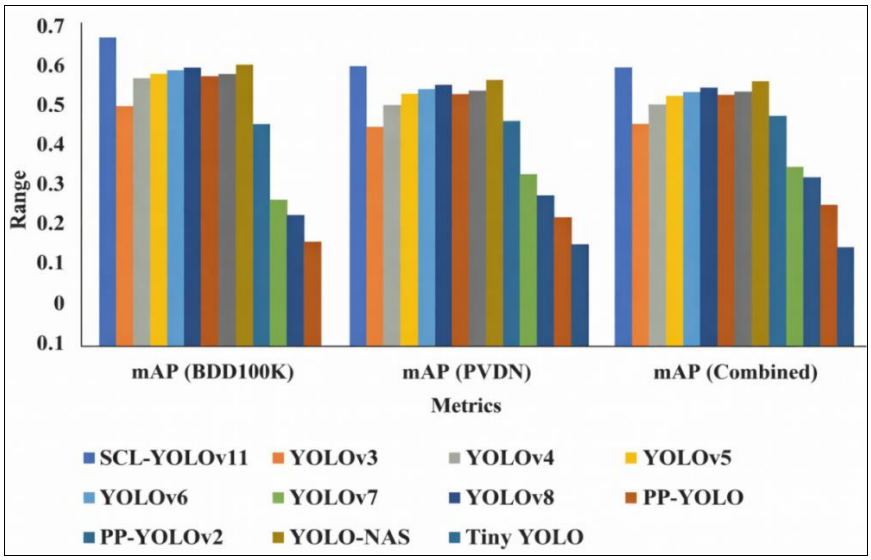

Figure 6. Comparative study of proposed model in terms of mAP

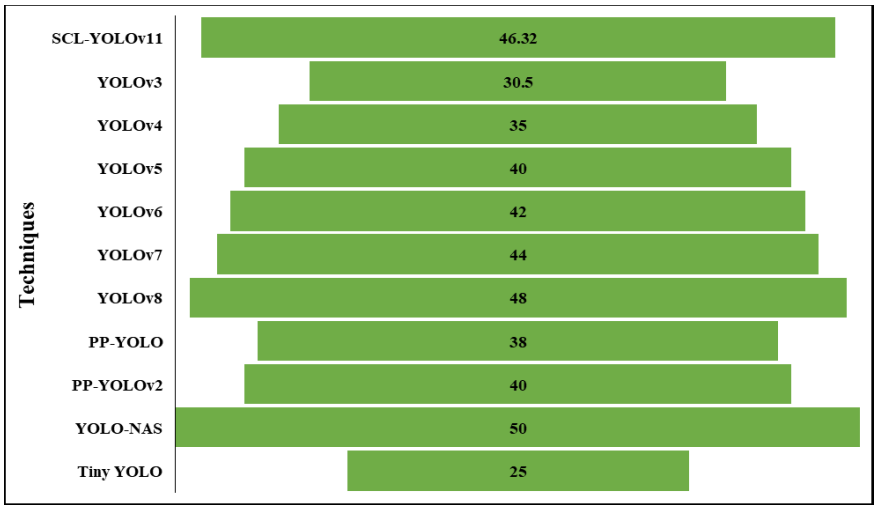

Figure 7. Analysis of frames per second on projected model

4.8. Statistical validation and error analysis

Paired t-test results (Table 6) confirm that performance differences between SCL-YOLOv11 and recent detectors such as YOLOv7, YOLOv8, and YOLO-NAS are statistically significant ($p<0.05$). Narrow confidence intervals further indicate stable convergence across repeated trials. Error analysis reveals that false positives mainly occur near strong light sources, while missed detections are common in near-black regions with limited texture.

Table 6. Statistical significance analysis (paired t-test at $p<0.05$)

Comparison	Dataset	mAP	t-value	p-value	Significance
SCL-YOLOv11 vs YOLOv7	Combined	+0.0034	2.15	0.034	Significant
SCL-YOLOv11 vs YOLOv8	Combined	-0.0066	-1.89	0.048	Significant
SCL-YOLOv11 vs YOLO-NAS	Combined	-0.0166	-2.42	0.027	Significant

4.9. Embedded system suitability analysis

The Table 7 compares the computational efficiency and embedded deployment feasibility of different YOLO-based models using parameters, GFLOPs, and inference time. The results show that YOLOv8 and YOLOv10 require higher computational resources, resulting in longer inference times. Although YOLOv11 improves efficiency, the proposed SCL-YOLOv11 with ROA optimization achieves the lowest parameter count (8.9 M), reduced computational complexity (19.6 GFLOPs), and the fastest inference time (14.8 ms). These improvements indicate that the proposed model is more computationally efficient and well suited for real-time deployment in embedded systems used in AV applications.

Table 7. Embedded system suitability analysis

Model	Parameters (M)	GFLOPs	Inference time (ms)	Embedded deployment suitability
YOLOv8	11.2	28.4	21.3	Moderate
YOLOv10	10.5	24.7	18.5	Good
YOLOv11	9.8	23.1	17.9	Good
Proposed SCL-YOLOv11+ROA	8.9	19.6	14.8	Highly suitable

5. CONCLUSION

This paper proposed SCL-YOLOv11, a more powerful object detection network capable of enhancing perception tasks in autonomous driving at night. The proposed scheme blends strong image enhancement and feature learning techniques to overcome the illumination problem and feature invisibility at night. Introducing SCL in the YOLOv11 framework helps the model better distinguish features in shadowed areas for more accurate and reliable object detection. Moreover, ROA is brought in to fine tune the model parameters and enhance training stability. Tests of the proposed architecture with standard night, time datasets show that the model not only delivers better detection accuracy but is also computationally efficient. The experiments illustrate that the new technique ensures a well, balanced compromise among accuracy, computational efficiency, and robustness which are the key requirements for perception modules in real, time AV systems. Thus, the new method based on SCL-YOLOv11 is a reasonable and effective way of increasing the reliability of object detection in night, time scenarios that are challenging and has a great potential of contributing to the making of safer and more intelligent autonomous driving systems. The proposed SCL-YOLOv11 with ROA optimization provides a computationally efficient detection framework suitable for real-time deployment on embedded and reconfigurable platforms used in AVs, thereby aligning the work with the scope of reconfigurable and embedded systems research.

Future work could explore the incorporation of multi-modal data, such as thermal imaging or LiDAR, to further enhance detection reliability in adverse environmental conditions like fog or rain. Investigating generative approaches for dynamic illumination adjustment could make the system more versatile. Expanding the diversity and size of training datasets with extreme night-time scenarios would improve the model's generalization. Additionally, leveraging hybrid optimization algorithms and fine-tuning the framework for edge computing hardware could make it more scalable for real-world deployment in large-scale AV systems. These enhancements would further solidify SCL-YOLOv11's potential as a cornerstone technology, proceeding the safety and efficiency of vehicles in low-visibility environments.

ACKNOWLEDGMENTS

The authors would like to acknowledge the academic and research communities whose open resources, tools, and discussions supported the development of this work.

FUNDING INFORMATION

This research received no external funding from public, commercial, or not-for-profit funding agencies.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Kondapalli Sri Vijaya		✓	✓		✓			✓	✓				✓	
Gokula Krishnan Vasudevan		✓		✓		✓		✓		✓	✓			
Pinagadi Venkateswara Rao	✓		✓			✓	✓			✓	✓			✓
Therasa Michael		✓		✓	✓		✓			✓		✓		✓
Balasubramanian Lalithambigai		✓	✓		✓		✓		✓				✓	
Boddula Prathusha Laxmi		✓		✓		✓		✓	✓				✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial or non-financial interests that could have influenced the work reported in this paper. The authors declare no conflict of interest.

DATA AVAILABILITY

Data and materials supporting the findings of this study are available from the corresponding author upon reasonable request. Where applicable, publicly available data sources were used in accordance with their respective terms and conditions.

REFERENCES

[1] N. A. Almujally *et al.*, "A novel framework for vehicle detection and tracking in night ware surveillance systems," *IEEE Access*, vol. 12, pp. 88075-88085, 2024, doi: 10.1109/ACCESS.2024.3417267.

[2] J. Yuan, A. Le-Tuan, M. Hauswirth, and D. Le-Phuoc, "Cooperative students: Navigating unsupervised domain adaptation in nighttime object detection," in *2024 IEEE International Conference on Multimedia and Expo (ICME)*, 2024, pp. 1–6, doi: 10.1109/ICME57554.2024.

[3] S. A. Agyemang, H. Shi, X. Nie, N. Y. Asabere, and B. Li, "Nighttime object detection with denoising diffusion-probabilistic models," in *2024 International Conference on Cyber-Physical Social Intelligence (ICCSI)*, Doha, Qatar, 2024, pp. 1-6, doi: 10.1109/ICCSI62669.2024.10799292.

[4] H. Lin *et al*., "A study on data selection for object detection in various lighting conditions for autonomous vehicles," *Journal of Imaging*, vol. 10, no. 7, pp. 1-18, 2024, doi: 10.3390/jimaging10070153.

[5] R. Sujatha, K. Chavan, S. Singh, and K. Balaji, "AIoT enhanced night vision object detection for vehicle safety using transfer learning on Raspberry Pi5," in *2024 7th International Conference on Internet Applications, Protocols, and Services (NETAPPS)*, Kuala Lumpur, Malaysia, 2024, pp. 1-8, doi: 10.1109/NETAPPS63333.2024.10823504.

[6] H. Wei, Y. Fu, D. Wang, R. Guo, X. Zhao, and J. Fang, "Unsupervised nighttime object tracking based on transformer and domain adaptation fusion network," *IEEE Access*, vol. 12, pp. 130896-130913, 2024, doi: 10.1109/ACCESS.2024.3378117.

[7] R. V. Raja, K. A. Kumar, and V. G. Krishnan, "Condition based ensemble deep learning and machine learning classification technique for integrated potential fishing zone future forecasting," International Journal on Recent and Innovation Trends in Computing and Communication, vol. 11, no. 2, pp. 75–85, 2023, doi: 10.17762/ijritcc.v11i2.6131.

[8] J. Cao, X. Zheng, Y. Lyu, J. Wang, R. Xu, and L. Wang, "Chasing day and night: Towards robust and efficient all-day object detection guided by an event camera," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, Yokohama, Japan, 2024, pp. 9026-9032, doi: 10.1109/ICRA57147.2024.10611705.

[9] K.-F. Hung and K.-P. Lin, "Bio-inspired dark adaptive nighttime object detection," *Biomimetics*, vol. 9, no. 3, pp. 1-20, 2024, doi: 10.3390/biomimetics9030158.

[10] Z. Du, M. Shi, and J. Deng, "Boosting object detection with zero-shot day-night domain adaptation," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2024, pp. 12666-12676, doi: 10.1109/CVPR52733.2024.01204.

[11] C. Zhang and D. Lee, "Advancing nighttime object detection through image enhancement and domain adaptation," *Applied Sciences*, vol. 14, no. 18, pp. 1-19, 2024, doi: 10.3390/app14188109.

[12] T. Wang, S. Ren, and H. Zhang, "Nighttime wildlife object detection based on YOLOv8-night," *Electronics Letters*, vol. 60, no. 15, p. e13305, 2024, doi: 10.1049/ell2.13305.

[13] Y. Xi, W. Jia, Q. Miao, J. Feng, J. Ren, and H. Luo, "Detection-driven exposure-correction network for nighttime drone-view object detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1-14, 2024, doi: 10.1109/TGRS.2024.3351134.
[14] Q. Yang, Y. Ma, L. Li, and Z. Zhao, "Dual-mode serial night road object detection model based on depthwise separable and self-attention mechanism," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1-9, 2024, doi: 10.1109/TIM.2024.3385040.
[15] T.-H. Le, S.-C. Huang, and Q.-V. Hoang, "Multilevel knowledge transmission for object detection in rainy night weather conditions," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 9, pp. 11224-11232, Sept. 2024, doi: 10.1109/TII.2024.3396552.
[16] H. Zhang, K.-F. Yang, Y.-J. Li, and L. L.-H. Chan, "Night-time vehicle detection based on hierarchical contextual information," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 10, pp. 14628-14641, Oct. 2024, doi: 10.1109/TITS.2024.3395666.
[17] D. D. N. Anh, B. N. Thai, A. N. Hoang, and K. D. Nguyen, "Aerial vehicle detection at night: A synthesized dataset and performance evaluation of state-of-the-art object detection models," in *2024 13th International Conference on Control, Automation and Information Sciences (ICCAIS)*, Ho Chi Minh City, Vietnam, 2024, pp. 1-6, doi: 10.1109/ICCAIS63750.2024.10814532.
[18] L. Zhang, W. Xu, C. Shen, and Y. Huang, "Vision-based on-road nighttime vehicle detection and tracking using improved HOG features," *Sensors*, vol. 24, no. 5, pp. 1-14, 2024, doi: 10.3390/s24051590.
[19] L. Encío *et al.*, "Enhanced nighttime vehicle detection for on-board processing," *IEEE Access*, vol. 13, pp. 44817-44835, 2025, doi: 10.1109/ACCESS.2025.3548837.
[20] J. Guo, M. Shao, X. Chen, Y. Yang, and E. Sun, "Research on night-time vehicle target detection based on improved KSC-YOLOv5," *Signal, Image and Video Processing*, vol. 19, pp. 1–11, 2025, doi: 10.1007/s11760-024-03576-5.
[21] A. K. Kumar and P. Venkatesh, "Real-time object detection using improvised YOLOv4 and feature mapping technique for autonomous driving," *Expert Systems with Applications*, p. 127452, 2025, doi: 10.1016/j.eswa.2025.127452.
[22] M. Chao *et al.*, "AutoSelecter: Efficient synthetic nighttime images make object detector stronger," *IEEE Robotics and Automation Letters*, vol. 10, no. 5, pp. 4660-4665, May 2025, doi: 10.1109/LRA.2025.3552996.
[23] J. A. Pandian, R. Thirunavukarasu, and L. T. Mariappan, "Enhancing lane detection in autonomous vehicles with multi-armed bandit ensemble learning," *Scientific Reports*, vol. 15, no. 1, p. 3198, 2025, doi: 10.1038/s41598-025-86743-z.
[24] Y.-P. Song, X. Wu, W. Li, T.-Q. He, D. F. Hu, and Q. Peng, "HighlightNet: Learning highlight-guided attention network for nighttime vehicle detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 4, pp. 4491-4503, Apr. 2025, doi: 10.1109/TITS.2025.353909.
[25] L. Ewecker, F. Schiffel, R. Schwager, T. Brühl, T. S. Sohn, and T. Villmann, "PVDN-Urban-A dataset for provident vehicle detection at night in urban scenarios," in *2024 IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, United Arab Emirates, 2024, pp. 34-40, doi: 10.1109/ICIP51287.2024.10647873.
[26] T. Hamadneh *et al.*, "Revolution optimization algorithm: A new human-based metaheuristic algorithm for solving optimization problems," *International Journal of Intelligent Engineering and Systems*, vol. 18, no. 2, pp. 520-531, 2025, doi: 10.22266/ijies2025.0331.38.

BIOGRAPHIES OF AUTHORS

Kondapalli Sri Vijaya received her B.Tech. degree in Computer Science and Engineering from Sri Prakash College of Engineering, affiliated to J.N.T. University, in 2008, and her M.Tech. degree in Computer Science and Technology from Sir C.R. Reddy College of Engineering, affiliated to Andhra University, in 2011. She is currently pursuing a Ph.D. in Computer Science and Engineering at Jain Institute of Technology (JIT), affiliated with Visvesvaraya Technological University (VTU), Belagavi, India. Her research interests include multi-level cloud media data authentication and confidentiality, with a focus on chaotic hash functions and attribute-based encryption techniques. She can be contacted at email: srivijayak@gmail.com.

Dr. Gokula Krishnan Vasudevan is currently working as Professor in the Department of Computer Science and Engineering in Easwari Engineering College, Ramapuram, Chennai, Tamil Nadu, India. He has completed his Under-Graduation (B.E.) in Anna University, Post-Graduation (M.Tech.) in Dr. MGR University and Ph.D. in Sathyabama Institute of Science and Technology, Chennai. He has more than 18 years of teaching experience in various colleges in Chennai and Hyderabad. He has published several papers in SCI/Scopus/WoS indexed journals. Also, he has presented various papers in National/International Conferences. His area of interest includes computer networks, computer architecture, data structures, and software engineering. He serves as the guest editor, editorial member and also as reviewer in many reputed international journals. He is a member of Professional Bodies like ISTE, IAENG, CSI, and IEEE. He can be contacted at email: gokul_kris143@yahoo.com.

Dr. Pinagadi Venkateswara Rao is a Senior NLP Engineer and researcher specializing in natural language processing (NLP), artificial intelligence, and healthcare informatics. He holds a Ph.D. in Computer Science and Engineering from Sathyabama Institute of Science and Technology, along with a Master of Engineering in Computer Science and Engineering. He has rich academic and industry experience, having previously served as an Associate Professor and currently working in the healthcare AI domain. His research interests include Clinical NLP, large language models (LLMs), electronic health record (EHR) analysis, clinical documentation enhancement, radiology report generation, and predictive analytics using machine learning and deep learning techniques. He has published 20 Scopus-indexed national and international journal papers in reputed journals, with a strong focus on NLP applications and intelligent healthcare systems. He actively contributes to the development of AI-driven solutions for processing unstructured medical data and advancing automated clinical text understanding. He can be contacted at email: drrao.pinagadi@cvr.ac.in.

Therasa Michael received her B.E. degree in Computer Science and Engineering from Adhiparasakthi College of Engineering, affiliated with Anna University, in 2006. She obtained her M.E. degree in 2010 and Ph.D. degree in 2024 in Computer Science and Engineering from Sathyabama Institute of Science and Technology, Chennai, India. She is currently working as an Associate Professor in the Department of Computer Science and Engineering at Panimalar Engineering College. Her research interests include deep learning, image processing, and natural language processing. She has published research papers in reputed journals and actively participates in academic and research activities. She can be contacted at email: therasamic@gmail.com.

Mrs. Balasubramanian Lalithambigai is currently working as an Assistant Professor in the Department of Computer Science and Design, Easwari Engineering College, Ramapuram, Tamilnadu India. She is pursuing her Ph.D. at Anna University, Chennai. She received her M.E. (Computer Science and Engineering). degree from Mepco Schlenk Engineering College, Anna University in 2011 and was awarded B.Tech. degree from Arulmigu Kalasalingam College of Engineering 2008. She has 6 years of teaching experience in higher education. Her areas of interest include medical image processing, internet of things, and machine learning. She can be contacted at email: lalithambigai.b@eec.srmrmp.edu.in.

Dr. Boddula Prathusha Laxmi received her Bachelor's degree in Computer Science and Engineering at National Institute of Technology, Tiruchirappalli, and her Ph.D. in Information and Communication Engineering at Anna University. She is full time Professor at Department of Artificial Intelligence and Data Science, R.M.K. College of Engineering and Technology, Chennai. Her research interests are wireless sensor networks, machine learning, and deep learning. She can be contacted at email: prathushalaxmi@gmail.com.