ILKOM Jurnal Ilmiah Vol. 18, No. 2, August 2026, pp. 207-220
Accredited 2nd by KEMDIKTISAINTEK No. 10/C/C3/DT.05.00/2025; E-ISSN 2548-7779 | P-ISSN 2087-1716



# An Optimized YOLOv7-Based Object Detection Framework Leveraging Low-Light Image Enhancement to Improve Accuracy

**Rasim[a,1,*]; Farhan Nurzaman[a,2]; Yaya Wihardi [a,3]; Herbert Siregar [a,4]; Samialloi Nusratullo [b,5]**
[a] *Universitas Pendidikan Indonesia, Jalan Dr. Setiabudhi No. 229, Sukasari, Bandung 40154, Indonesia*
[b] *Borough of Manhattan Community College, 199 Chambers St, New York, NY 10007, United States*
[1] *rasim@upi.edu;* [2] *farhannz@upi.edu;* [3] *yayawihardi@upi.edu;* [4] *herbert@upi.edu;* [5] *nusratullosamiallo@gmail.com*
** Corresponding author*

**Article history:** Received September 18, 2025; Revised November 16, 2025; Accepted April 22, 2026; Available online 10 August 2026.

**Abstract**

Low-light environments remain a persistent challenge in computer vision, often leading to notable degradation in object detection performance. This is primarily caused by reduced contrast, increased noise, and the loss of critical visual details, all of which hinder reliable feature extraction. To address these limitations, this study proposes an integrated framework that combines Zero-Reference Deep Curve Estimation (Zero-DCE) for adaptive image enhancement with the YOLOv7 architecture for efficient and accurate object detection. The study was conducted through a structured pipeline consisting of several key stages: (1) preparation of the ExDark, NOD, and LOD datasets; (2) preprocessing, including annotation and labelling; (3) low-light image enhancement using Zero-DCE; (4) dataset selection, partitioning, and utilization; (5) image resizing to ensure model compatibility; (6) model development, comprising both a baseline YOLOv7 model and an enhanced Zero-DCE + YOLOv7 configuration; and (7) performance analysis and evaluation using mean Average Precision (mAP) as the primary metric. Experimental results demonstrate that the integration of Zero-DCE with YOLOv7 improves detection performance, with mAP@0.5 increasing from 0.785 in the baseline model to 0.794. Although the improvement is modest, it is consistent and indicates the effectiveness of incorporating illumination enhancement into the preprocessing stage. In addition, this study's contribution lies in demonstrating that the proposed framework enhances the robustness of object detection systems under challenging lighting conditions without incurring significant computational overhead.

**Keywords:** Object Detection; Low-light; Zero-DCE; Image Enhancement; YOLOv7.

## Introduction

Object detection constitutes a fundamental component of modern computer vision applications [1], spanning autonomous vehicles [2], AI-driven surveillance systems [3], medical image analysis [4], and human–computer interaction [5]. This technology enables systems to automatically recognize, localize, and classify objects within images or videos [6], [7], thereby facilitating rapid, accurate, and data-driven decision-making [8].

In autonomous vehicles, object detection plays a pivotal role in identifying traffic signs [9], [10], [11], [12], pedestrians [13], [14], [15], and other vehicles to ensure safe navigation [16], [17]. Within the medical domain, object detection techniques have made significant contributions to supporting diagnostic processes and clinical interventions. For example, Kim et al. [18] employed object detection algorithms to identify structural abnormalities in radiological images, whereas Alsallal et al. [19] applied them for tumor and lesion detection, which are critical for early diagnosis and targeted therapeutic planning.

In the industrial sector, this technology has been widely adopted to improve operational efficiency and enhance the accuracy of automated systems. For instance, intelligent traffic monitoring systems [20], [21], [22] leverage object detection to calculate vehicle density, identify violations, and dynamically control traffic signals. Similarly, computer vision-based systems in logistics and manufacturing enable real-time inventory tracking [23], defect detection [24], and optimization of distribution routes [25]. Furthermore, in security surveillance, object detection assists in identifying suspicious activities in public areas [26], critical facilities [27], and restricted zones [28], even under challenging environmental conditions [29].

http://dx.doi.org/10.33096/ilkom.v18i2.3102.207-220

With the increasing demand for intelligent systems capable of performing reliably under diverse environmental conditions, new challenges have emerged—particularly regarding the quality of visual inputs provided to these systems. One of the most significant challenges involves low-light environments [30], which substantially impact the performance of computer vision-based object detection algorithms. Yi et al. [31] demonstrated that under poor illumination, digital images frequently suffer from visual degradation, including reduced contrast, increased noise, color distortions, and the loss of essential semantic details necessary for effective feature extraction. Consequently, detection models that heavily depend on spatial and textural representations often struggle to accurately recognize and localize objects.

This issue is particularly critical in real-world scenarios, such as nighttime surveillance in public spaces, monitoring poorly lit indoor facilities (e.g., warehouses or industrial corridors), and extreme weather conditions like dense fog or heavy rainfall, which significantly reduce visibility. In these contexts, insufficient illumination not only degrades detection accuracy but may also lead to serious safety risks and operational inefficiencies—especially in critical domains such as autonomous driving, AI-powered security systems, and medical image analysis [32], [33], [34].

Various approaches have been developed to mitigate these challenges. Classical methods, such as Histogram Equalization [35] and Retinex-based enhancement [36], suffer from limited adaptability and often produce visual artifacts or undesirable distortions. In contrast, deep learning-based techniques—including LLNet [37], EnlightenGAN [38], and Zero-DCE [39]—achieve more adaptive image enhancement. Among them, Zero-DCE has demonstrated competitive performance while maintaining a lightweight architecture through its curve-estimation mechanism, enabling effective illumination adjustment without requiring paired training data. Recently, Wang et al. [40] proposed a domain adaptation strategy for low-light object detection by integrating a Laplacian pyramid-based enhancement module with multi-scale feature alignment, enabling the extraction of more robust features under varying illumination conditions. Similarly, Liu et al. [41] introduced Dark-YOLO, an object detection algorithm incorporating multiple attention mechanisms and adaptive enhancement modules, which demonstrated significant performance improvements on the ExDark and DarkFace [42] datasets. Although Dark-YOLO achieved promising results, its reliance on additional enhancement and attention components increases architectural complexity, suggesting that further improvements may be achieved through more efficient and modular integration strategies. Moreover, Anoop and Deivanathan [43] conducted a comparative analysis of enhancement techniques and detection models, emphasizing the importance of selecting algorithms that preserve semantic features rather than optimizing solely for visual quality.

Despite these advancements, there remains a substantial gap in developing efficient and lightweight pipelines suitable for real-world deployment, where detection systems are often required to operate under varying illumination conditions and limited computational resources. In practical scenarios such as surveillance, traffic monitoring, and autonomous systems, enhancement methods must not only improve visibility but also preserve structural consistency without imposing significant computational overhead. In this context, Zero-DCE provides an effective, end-to-end, reference-free solution for low-light image enhancement through its lightweight curve-estimation mechanism, which adaptively adjusts pixel-wise illumination without requiring paired training data. On the detection side, YOLOv7 [44] delivers fast and accurate object detection through its optimized one-stage architecture, efficient feature aggregation, and real-time inference capability, making it suitable for deployment in time-sensitive applications.

Given these complementary characteristics, integrating Zero-DCE and YOLOv7 becomes particularly appealing for real-world low-light detection tasks, as the enhancement module can improve input visibility while the detector leverages its efficient architecture to maintain high-speed and accurate predictions. However, studies explicitly integrating Zero-DCE as a pre-processing stage with YOLOv7 for object detection remain limited. Therefore, this study aims to evaluate the effectiveness of the Zero-DCE+YOLOv7 pipeline in improving object detection performance under low-light imaging conditions. Specifically, it investigates the contribution of illumination enhancement to detection accuracy and assesses the robustness of the proposed model across diverse scenarios with extreme lighting variations.

## Method

The methodological framework of this study follows a systematic and reproducible process designed to enhance object detection under low-light conditions. The dataset was obtained from publicly available benchmark sources and

underwent a series of preprocessing steps, including resizing, labeling, and annotation, to prepare it for model training. Image enhancement was subsequently performed using the Zero-DCE algorithm to improve visual quality and object visibility in low-light environments. The processed dataset was then divided into training, validation, and testing subsets to ensure unbiased model evaluation, and all images were standardized to a resolution of 416 × 416 pixels for consistency across experiments. Two model configurations were developed and compared: the baseline YOLOv7 and the Zero-DCE-enhanced YOLOv7. The performance of both models was evaluated using mean Average Precision (mAP) as the primary evaluation metric. An overview of the complete research workflow is presented in Figure 1, providing a comprehensive illustration of the proposed methodology.

All experiments were conducted on a workstation equipped with an NVIDIA GeForce GTX 1080 Ti (11GB), an Intel Core i7-6900K processor, and 64GB of RAM. Considering the available computational resources, the enhancement and detection modules were trained sequentially rather than within a fully end-to-end joint optimization framework. This setup reflects a practical single-GPU configuration representative of real-world deployment and applied research environments.

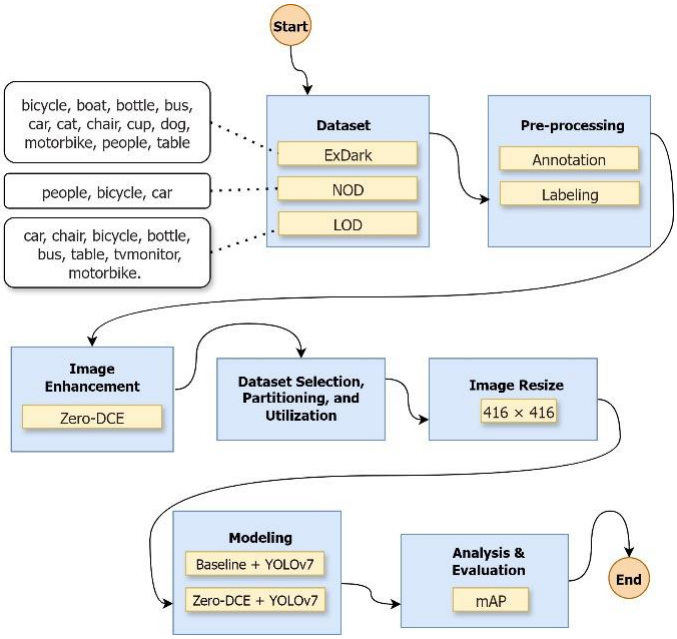


**Figure 1.** Research Workflow

## Result and Discussion

### *A. Result*

*1. Dataset*

This study employs three publicly available datasets to facilitate model development and performance evaluation in low-light object detection. Among them, the ExDark dataset serves as the primary source for training, validation, and internal testing, whereas the NOD and LOD datasets are used exclusively for external testing. This separation ensures an objective assessment of the model's generalizability under diverse illumination conditions and object distributions.

The ExDark dataset was obtained from https://github.com/cs-chan/Exclusively-Dark-Image-Dataset and comprises 7,363 low-light images distributed across 12 object categories: bicycle, boat, bottle, bus, car, cat, chair, cup, dog, motorbike, people, and table. Each image is accompanied by a .txt annotation file specifying object class labels and bounding-box coordinates in the format (x_min, y_min, width, height). These annotations were utilized during both training and evaluation stages to enable accurate object localization. In total, the dataset provides 23,711 annotated instances, with people, car, and chair being the most frequently represented categories.

The NOD dataset, retrieved from https://github.com/igor-morawski/NOD, contains three subsets of low-light images captured using different imaging devices: Nikon D750, Sony RX100 VII, and a hybrid combination of both.

To ensure consistency in data distribution, only the Nikon D750 subset was selected for this study. This subset comprises 4,006 images with 28,046 annotated instances spanning three object categories: people, bicycle, and car.

The LOD dataset, obtained from https://github.com/yanghong7410/LODDataset, consists of four distinct data types—RGB-Normal, RGB-Dark, RAW-Normal, and RAW-Dark—classified according to image format (RGB vs. RAW) and illumination level (Normal vs. Dark). Specifically, RGB images represent standard pixel-based formats containing red, green, and blue channel values, whereas RAW images consist of unprocessed sensor data captured using a Bayer filter configuration. For this study, the RGB-Dark subset was selected to evaluate the proposed model's performance in low-light environments. This subset comprises 2,230 images distributed across eight object categories: car, chair, bicycle, bottle, bus, table, tvmonitor, and motorbike. Due to its limited representation, the tvmonitor category was excluded, resulting in a final evaluation set of 1,881 images.

By integrating these three datasets, this study establishes a comprehensive experimental framework that enables both robust training and unbiased cross-dataset evaluation. This design ensures a fair comparison of model performance and provides insights into the adaptability of low-light object detection systems under diverse imaging conditions.

*2. Pre-Processing*

The three datasets employed in this study utilize heterogeneous annotation formats, necessitating a standardized representation to ensure consistency throughout the data processing pipeline. All annotations were therefore converted to the YOLO format, which encodes object coordinates using x_midpoint, y_midpoint, width, and height, with all values normalized to a pixel range of [0, 1]. The conversion process commenced by loading each annotation file and assessing its structural compatibility with the YOLO specification. For annotations initially defined using x_min and y_min, the midpoint coordinates were subsequently recalculated using Equations (1) and (2).

$$x_{\text{midpoint}} = x_{min} + \left(\frac{\text{width}}{2}\right) \quad (1)$$

$$y_{\text{midpoint}} = y_{min} + \left(\frac{\text{width}}{2}\right) \quad (2)$$

In addition to spatial normalization, categorical labels were numerically encoded to align with YOLO's classification schema. Specifically, the object classes bicycle, boat, bottle, bus, car, cat, chair, cup, dog, motorbike, people, and table were mapped to integer identifiers ranging sequentially from 0 to 11. A representative sample of the reformatted annotations is presented in **Table 1**, illustrating the final structure adopted for model training and evaluation.

**Table 1.** Sample of Processed Annotations in YOLO Format

| Filename | Label | $x_{center}$ | $y_{center}$ | width | height |
|---|---|---|---|---|---|
| 2015_00001.png | 0 | 0.67900 | 0.33200 | 0.54200 | 0.51467 |
| 2015_00002.png | 0 | 0.35100 | 0.71701 | 0.15800 | 0.31965 |
| | 0 | 0.50100 | 0.69648 | 0.12600 | 0.38416 |
| | 0 | 0.63000 | 0.73314 | 0.15200 | 0.36364 |
| | 0 | 0.75300 | 0.65543 | 0.11400 | 0.23754 |
| 2017_07358.jpg | 6 | 0.84717 | 0.21303 | 0.10645 | 0.13324 |
| | 6 | 0.59424 | 0.39678 | 0.12988 | 0.46267 |
| | 6 | 0.56104 | 0.51830 | 0.20215 | 0.60029 |
| | 6 | 0.74854 | 0.64715 | 0.26465 | 0.49780 |
| | 11 | 0.79199 | 0.48390 | 0.34570 | 0.49048 |

*3. Image Enhancement*

Prior to the training phase, the ExDark dataset was subjected to low-light image enhancement using the Zero-DCE model. This enhancement step was essential to establish a clear distinction between the original dataset (without enhancement) and the modified dataset (with enhancement), thereby enabling comparative analysis of model performance under varying illumination conditions. The Zero-DCE model was implemented using the PyTorch framework and comprises a series of Conv2D layers integrated with ReLU and Tanh activation functions. The architectural structure of the enhancement model is illustrated in **Figure 2**.

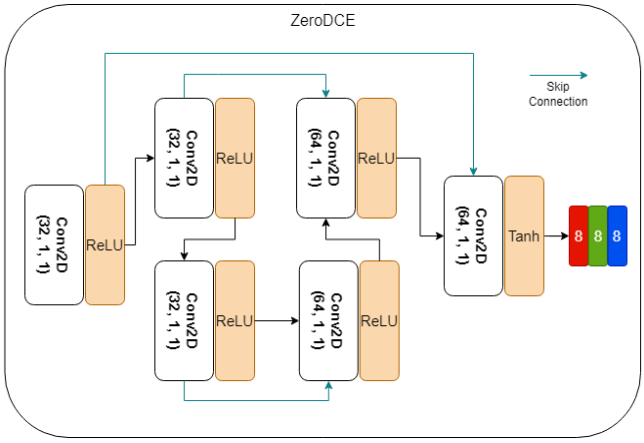


**Figure 2.** Zero-DCE Architecture

Subsequently, the pixel values of the input image must be normalized prior to processing. Initially ranging from [0, 255], these values are rescaled to the interval [0, 1] using min-max normalization to ensure compatibility with the model's input requirements. This normalization procedure is illustrated in **Figure 3**.

$$normalize\left(\underbrace{\begin{bmatrix} 2 & 1 & 1 & \dots & 1 \\ 2 & 1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 4 & 2 & 3 & \dots & 4 \end{bmatrix}}_{height \times width}\right) = \underbrace{\begin{bmatrix} 0.00784 & 0.00392 & 0.00392 & \dots & 0.00392 \\ 0.00784 & 0.00392 & 0.00392 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0.0157 & 0.00784 & 0.0118 & \dots & 0.00157 \end{bmatrix}}_{height \times width}$$

**Figure 3.** Pixel Normalization

Each image in the ExDark dataset was processed using the Zero-DCE model to enhance illumination quality under low-light conditions. The resulting images exhibit improved lighting compared to their original counterparts, thereby facilitating more effective object detection. A comparative visualization of pre- and post-enhancement outputs is presented in **Figure 4**.

**Figure 4.** Pre- and Post-Enhancement Using Zero-DCE

*4. Dataset Selection, Partitioning, and Utilization*

Among the three datasets mentioned earlier, the ExDark dataset was selected as the primary source for model training and validation. This dataset was divided into three subsets: training, validation, and testing, with respective proportions of 70%, 15%, and 15%. In contrast, the NOD and LOD datasets were used exclusively for external testing to evaluate the model's generalization capability under varying illumination conditions (see **Table 2**). The partitioning of the ExDark dataset was performed through an iterative sampling process across all image categories to ensure proportional allocation into the designated subsets. The resulting samples from each class were stored in separate lists to maintain categorical consistency and facilitate subsequent processing. This procedure resulted in a total of 5,148 images for training, 1,111 images for testing, and 1,104 images for validation.

**Table 2.** Testing Categories

| Test | Dataset | Class |
|---|---|---|
| *Internal* | ExDark | *Bicycle, Boat, Bottle, Bus, Car, Cat, Chair, Cup, Dog, Motorbike, People, Table* |
| *External* | ExDark | *Boat, Cat, Cup, Dog* |
| | LOD | *Car, Chair, Bicycle, Bottle, Bus, Table, Motorbike* |
| | NOD | *People, Bicycle, Car* |

*5. Image Resize*

Prior to performing object detection using the YOLOv7 model, all input images are resized to a fixed resolution of 416 × 416 pixels. This preprocessing step is crucial for reducing computational complexity and ensuring compatibility with the model's input specifications. The resizing operation is automatically executed during the image-loading process by leveraging the resize function provided by the OpenCV library. However, since many input images do not inherently conform to a 1:1 aspect ratio, direct resizing can introduce geometric distortions and produce unnatural object proportions, as illustrated in **Figure 5**. Such distortions may negatively affect detection accuracy, particularly in scenarios where precise spatial localization is required. To overcome this issue, a padding strategy is employed to maintain the original aspect ratio. By adding border pixels around the image, this technique preserves the structural integrity of the visual content while simultaneously conforming to the model's fixed input dimension requirements.

**Figure 5.** Resizing Non-Square Images

*6. Modeling*

Two modeling scenarios were implemented to evaluate the performance of the object detection system under varying lighting conditions, namely Baseline+YOLOv7 and Zero-DCE+YOLOv7. The Baseline+YOLOv7 model served as the baseline approach, in which object detection was performed directly on the original images without applying any image enhancement techniques. In contrast, the Zero-DCE+YOLOv7 model integrates the Zero-DCE algorithm to improve image brightness and contrast prior to the detection stage. The integration of Zero-DCE was designed to enhance detection performance in low-light environments without relying on additional reference data. Both models were trained and evaluated using four different configurations to enable a comprehensive and robust comparative analysis of the system's accuracy and efficiency, as summarized in **Table 3**.

**Table 3.** Training Configuration

| Configuration | Optimizer | Learning rate | Epochs | Batch | Freezed |
|---|---|---|---|---|---|
| A | SGD | 0.01 | 150 | 24 | No |
| B | SGD | 0.01 | 50 | 32 | Yes |
| C | ADAM | 0.001 | 150 | 24 | No |
| D | ADAM | 0.001 | 75 | 32 | Yes |

**Table 4** presents the performance results of the Baseline+YOLOv7 scenario model. In this experiment, four object detection configurations (A, B, C, and D) were evaluated using the mAP@0.5 metric across multiple object classes. The findings indicate that configuration B achieved the highest overall performance, with an average mAP@0.5 score of 0.785, outperforming D (0.766), A (0.758), and C (0.731). These results suggest that the detection strategies or parameter settings applied in configuration B were more effective in identifying objects within the evaluated dataset.

From a class-level perspective, objects such as Bus, Bicycle, and Motorbike demonstrated consistently high detection accuracy, with mAP scores approaching or exceeding 0.85 in several configurations. Conversely, classes such as Table, Bottle, and Chair exhibited relatively lower performance, with the lowest score recorded for Table in configuration C (0.531).

**Table 4.** Internal Testing Results of Baseline+YOLOv7 on the ExDark Dataset

| Class | Configuration | | | |
|---|---|---|---|---|
| | A | B | C | D |
| Bicycle | 0.828 | 0.875 | 0.828 | 0.857 |
| Boat | 0.744 | 0.784 | 0.766 | 0.776 |
| Bottle | 0.664 | 0.650 | 0.647 | 0.639 |
| Bus | 0.900 | 0.919 | 0.866 | 0.910 |
| Car | 0.805 | 0.829 | 0.799 | 0.813 |
| Cat | 0.780 | 0.767 | 0.709 | 0.746 |
| Chair | 0.644 | 0.734 | 0.635 | 0.715 |
| Cup | 0.701 | 0.747 | 0.686 | 0.735 |
| Dog | 0.809 | 0.809 | 0.727 | 0.801 |
| Motorbike | 0.822 | 0.848 | 0.791 | 0.814 |
| People | 0.822 | 0.833 | 0.788 | 0.823 |
| Table | 0.571 | 0.620 | 0.531 | 0.569 |
| **mAP@.5** | **0.758** | **0.785** | **0.731** | **0.766** |

Table 5 presents a comprehensive summary of the Zero-DCE+YOLOv7 model's performance across four different configurations for 13 object classes. Among these, configuration B achieves the highest detection accuracy (mAP@0.5 = 0.794), demonstrating superior precision for Bus (0.921), Dog (0.853), and Cat (0.808). In contrast, configuration C records the lowest overall performance (mAP@0.5 = 0.724), particularly struggling with Table and Chair, which are more challenging to detect due to weak object boundaries and high background similarity.

Configurations A and D deliver comparable overall accuracy (mAP@0.5 = 0.758); however, their performance varies across specific classes. Configuration A performs better on Bicycle and Motorbike, whereas configuration D achieves higher precision for Car and People. Overall, larger objects with distinct visual features, such as Bus and Motorbike, are detected more effectively, while smaller or visually ambiguous objects, such as Bottle and Table, remain more difficult to identify accurately. A visual comparison between sample ground-truth images and the corresponding prediction results is provided in Figure 6 and Figure 7 to illustrate the model's detection performance.

**Table 5.** Internal Testing Results of Zero-DCE+YOLOv7 on the ExDark Dataset

| Class | Configuration | | | |
|---|---|---|---|---|
| | **A** | **B** | **C** | **D** |
| Bicycle | 0.870 | 0.878 | 0.836 | 0.839 |
| Boat | 0.749 | 0.803 | 0.728 | 0.719 |
| Bottle | 0.637 | 0.656 | 0.656 | 0.623 |
| Bus | 0.896 | 0.921 | 0.901 | 0.913 |
| Car | 0.802 | 0.828 | 0.783 | 0.816 |
| Cat | 0.746 | 0.808 | 0.682 | 0.783 |
| Chair | 0.661 | 0.713 | 0.645 | 0.694 |
| Cup | 0.725 | 0.759 | 0.664 | 0.704 |
| Dog | 0.788 | 0.853 | 0.726 | 0.798 |
| Motorbike | 0.832 | 0.857 | 0.752 | 0.829 |
| People | 0.812 | 0.836 | 0.785 | 0.821 |
| Table | 0.575 | 0.611 | 0.537 | 0.560 |
| **mAP@.5** | **0.758** | **0.794** | **0.724** | **0.758** |

The comparative evaluation between SGD and ADAM optimizers reveals that configurations using SGD (particularly Configuration B) consistently achieved superior performance compared to ADAM-based configurations. This observation may be attributed to SGD's stable convergence behavior and its tendency to generalize better in object detection tasks, especially when trained over extended epochs. In contrast, ADAM, while offering faster initial convergence due to adaptive learning rates, may lead to suboptimal generalization under certain dataset distributions, particularly in complex low-light scenarios. These findings suggest that optimizer selection plays a critical role in balancing convergence stability and detection accuracy within low-light detection pipelines.


**Figure 6.** Sample Ground Truth Images from ZeroDCE+YOLOv7 Configuration B


**Figure 7.** Sample Predicted Images from ZeroDCE+YOLOv7 Configuration B

*5. Analysis and Evaluation*

As previously described, configuration B demonstrated the highest performance among the evaluated models. To assess its generalization capability, configuration B was subsequently tested using the External-test dataset, a composite dataset characterized by a distinct image distribution compared to the ExDark dataset. This evaluation aimed to determine the extent to which the best-performing configuration could detect objects under conditions involving distributional shifts from the training data. The results of the External-test evaluation are presented in **Table 6**, providing insight into the model's robustness when confronted with broader visual variability.

**Table 6.** External Testing Results Using the Optimal Configuration

| Class | Baseline | Zero-DCE |
|---|---|---|
| Bicycle | 0.404 | 0.403 |
| Boat | 0.763 | 0.799 |
| Bottle | 0.405 | 0.529 |
| Bus | 0.373 | 0.497 |
| Car | 0.613 | 0.643 |
| Cat | 0.718 | 0.767 |
| Chair | 0.445 | 0.514 |
| Cup | 0.791 | 0.748 |

| Class | Baseline | Zero-DCE |
|---|---|---|
| Dog | 0.721 | 0.754 |
| Motorbike | 0.279 | 0.338 |
| People | 0.530 | 0.520 |
| Table | 0.522 | 0.547 |
| **mAP@.5** | **0.547** | **0.588** |
| **time** | **4.2ms** | **10.3ms** |

### *B. Discussion*

The external testing results presented in Table 6 indicate that the integration of the Zero-DCE-based image enhancement technique yields varying impacts on object detection performance under low-light conditions. Overall, there is an improvement in the mean Average Precision at an IoU threshold of 0.5 (mAP@0.5), increasing from 0.547 in the baseline configuration to 0.588 after applying Zero-DCE. Similarly, in internal testing on the ExDark dataset, the integration of Zero-DCE improves mAP@0.5 from 0.785 to 0.794. These results correspond to relative improvements of approximately 1.15% in internal evaluation and 7.5% in external testing, indicating that illumination enhancement contributes more significantly under distributional shifts. This finding suggests that enhancement plays a more critical role when the model encounters unseen lighting distributions compared to in-domain evaluation, thereby strengthening robustness in cross-dataset scenarios.

When examining performance across individual object classes, the most significant improvement is observed in the Bottle class, where the mAP increases from 0.405 to 0.529. Substantial gains are also recorded for the Bus and Chair classes, rising from 0.373 to 0.497 and 0.445 to 0.514, respectively. Furthermore, classes such as Boat, Cat, and Dog also exhibit consistent improvements, highlighting the model's sensitivity to objects with distinctive contours and textures in low-light scenarios.

However, not all object classes benefit from the application of Zero-DCE. For instance, the Cup class experiences a slight decrease in performance, with the mAP dropping from 0.791 to 0.748, while People and Bicycle show marginal declines or remain relatively stable. This phenomenon may be attributed to the visual characteristics of these objects, which are less well-defined or become distorted following the image enhancement process, thereby reducing detection accuracy.

From a computational efficiency perspective, the inference time increases from 4.2 ms in the baseline configuration to 10.3 ms after applying Zero-DCE. While this represents a trade-off between accuracy and processing speed, the additional computational cost is still acceptable in applications where real-time performance is not critical and detection accuracy is prioritized.

**Table 7.** Comparison of ExDark Testing Results

| Model | *Test Size* | Baseline | Zero-DCE |
|---|---|---|---|
| YOLOv3 | 608 | 0.764 | 0.769 |
| Faster-RCNN | 800 | 0.621 | - |
| YOLOv7 | 416 | 0.785 | 0.794 |

Comparative analysis with existing studies, as shown in **Table 7**, further highlights the effectiveness of the proposed approach. Despite employing a lower testing image resolution (416×416 pixels), the YOLOv7 algorithm demonstrates superior object detection performance on the ExDark dataset compared to both YOLOv3 proposed by Cui et al. [45] and Faster R-CNN developed by Guo et al. [46]. Under baseline testing conditions, YOLOv7 achieves a notable improvement in mAP@0.5, increasing from 0.764 and 0.621 in YOLOv3 and Faster R-CNN, respectively, to 0.785. Furthermore, incorporating Zero-DCE for low-light enhancement leads to an additional boost in detection performance, where the transition from YOLOv3 to YOLOv7 yields an increase in mAP@0.5 from 0.769 to 0.794.

## Conclusion

This study provides compelling evidence that the integration of Zero-DCE-based image enhancement with the YOLOv7 object detection algorithm substantially improves detection performance under low-light conditions. The consistent increase in mAP@0.5 across the majority of object classes demonstrates that illumination quality plays a decisive role in extracting discriminative visual features, particularly in deep learning-based detection frameworks. These findings reinforce the significance of adaptive image preprocessing as a critical step toward achieving robust and reliable object detection pipelines.

Beyond improving detection accuracy, the results highlight that testing configurations—including input resolution, enhancement strategies, and model architectures—significantly influence overall system performance. Thus, the selection of model configurations should not be perceived as a purely technical choice but rather as a strategic decision that shapes the model's generalization capability in real-world environments, especially within challenging low-illumination contexts.

Despite these advances, the study has several limitations. The training processes for the image enhancement and object detection modules were performed independently, preventing end-to-end parameter optimization. Furthermore, the experiments were conducted under a practical single-GPU computational setting, which limited the exploration of more computationally intensive joint-optimization frameworks that could leverage dynamic feature alignment to further enhance detection performance.

Looking ahead, future research should focus on developing an end-to-end joint training paradigm that tightly couples Zero-DCE and YOLOv7, enabling the network to adaptively learn illumination-aware features within a unified optimization process. In addition, investigating more advanced yet efficient enhancement strategies and feature refinement mechanisms may further improve robustness under extreme lighting conditions.

In conclusion, this work contributes to the advancement of low-light object detection systems by demonstrating the value of combining illumination enhancement with state-of-the-art deep learning architectures. The findings establish a strong foundation for adaptive, high-performance detection pipelines capable of addressing the demands of real-world visual perception under extreme and dynamically varying lighting conditions.

## References

[1] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object Detection in 20 Years: A Survey," Proc. IEEE, vol. 111, no. 3, pp. 257–276, 2023, doi: 10.1109/JPROC.2023.3238524.

[2] J. Gao, H. Li, Z. Li, C. Xie, X. Ji, and Y. Zhang, "An algorithm for road target detection of autonomous vehicles based on improved YOLOv8," Sci. Rep., vol. 15, no. 1, p. 21061, Jul. 2025, doi: 10.1038/s41598-025-06831-y.

[3] Y. Zheng, L. Qian, J. Cao, H. Huang, W. Hou, and Y. Liu, "A reliable enhanced learning algorithm for ship detection in inland water maritime surveillance system," Eng. Appl. Artif. Intell., vol. 159, 2025, doi: 10.1016/j.engappai.2025.111820.

[4] C. Jiang, Y. Wang, Q. Yuan, P. Qu, and H. Li, "A 3D medical image segmentation network based on gated attention blocks and dual-scale cross-attention mechanism," Sci. Rep., vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-90339-y.

[5] P. Sonchan, N. Ratchatanantakit, N. O-Larnnithipong, M. Adjouadi, and A. Barreto, "Robust orientation estimation from mems magnetic, angular rate, and gravity (MARG) modules for human–computer interaction," Micromachines, vol. 15, no. 4, 2024, doi: 10.3390/mi15040553.

[6] J. Janai, F. Güney, A. Behl, and A. Geiger, "Computer Vision for Autonomous Vehicles," Found. Trends® Comput. Graph. Vis., vol. 12, no. 1–3, pp. 1–308, Jul. 2020, doi: 10.1561/0600000079.

[7] N. Ismail and O. A. Malik, "Real-time visual inspection system for grading fruits using computer vision and deep learning techniques," Inf. Process. Agric., vol. 9, no. 1, pp. 24–37, Mar. 2022, doi: 10.1016/j.inpa.2021.01.005.

[8] L. Zhou, L. Zhang, and N. Konz, "Computer Vision Techniques in Manufacturing," IEEE Trans. Syst. Man, Cybern. Syst., vol. 53, no. 1, pp. 105–117, Jan. 2023, doi: 10.1109/TSMC.2022.3166397.

[9] Q. Tian and J. Zhang, "STDE-YOLOv5s: A traffic sign detection algorithm based on context information enhancement," Discov. Artif. Intell., vol. 5, no. 1, 2025, doi: 10.1007/s44163-025-00405-7.

[10] D. Zou, H. Xie, and L. Kohnke, "Navigating the future: Establishing a framework for educators' pedagogic artificial intelligence competence," Eur. J. Educ., vol. 60, no. 2, 2025, doi: 10.1111/ejed.70117.

[11] M. Yang and S. Han, "ETS-YOLO: An efficient YOLO-based model for real-time traffic sign recognition," Signal, Image Video Process., vol. 19, no. 8, 2025, doi: 10.1007/s11760-025-04276-4.

[12] J. Qiu, W. Zhang, S. Xu, and H. Zhou, "DP-YOLO: A lightweight traffic sign detection model for small object detection," Digit. Signal Process. A Rev. J., vol. 165, 2025, doi: 10.1016/j.dsp.2025.105311.

[13] X. Y. Wu and T. K. F. Chiu, "Integrating learner characteristics and generative AI affordances to enhance self-regulated learning: A configurational analysis," J. New Approaches Educ. Res., vol. 14, no. 1, 2025, doi: 10.1007/s44322-025-00028-x.

[14] Z. Li, "Enhanced Fault Localization in Energy Systems Using an Improved MVO Algorithm and Multistrategy Optimization," IEEE Access, vol. 13, pp. 50367–50378, 2025, doi: 10.1109/ACCESS.2025.3552761.

[15] S. R. Alotaibi et al., "Harnessing optimization with deep learning approach on intelligent transportation system for anomaly detection in pedestrian walkways," Sci. Rep., vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-99940-7.

[16] N. Nigam, D. P. Singh, J. Choudhary, and S. Solanki, "PVD-GSTPS: Design of an efficient parallel vehicle detection based green signal time prediction system," Discov. Comput., vol. 28, no. 1, 2025, doi: 10.1007/s10791-025-09660-9.

[17] Y. Du, L. Chen, and X. Hao, "RL-Net: A rapid and lightweight network for detecting tiny vehicle targets in remote sensing images," Complex Intell. Syst., vol. 11, no. 8, 2025, doi: 10.1007/s40747-025-01956-z.

[18] M. J. Kim et al., "Comparison of artificial intelligence–derived heart age with chronological age using normal sinus electrocardiograms in patients with no evidence of cardiac disease," J. Clin. Med., vol. 14, no. 15, 2025, doi: 10.3390/jcm14155548.

[19] M. Alsallal et al., "Enhanced lung cancer subtype classification using attention-integrated DeepCNN and radiomic features from CT images: A focus on feature reproducibility," Discov. Oncol., vol. 16, no. 1, 2025, doi: 10.1007/s12672-025-02115-z.

[20] M. F. Fairuz, A. A. Alfariz, R. A. P. Bimantara, M. Irfan, and N. Setyawan, "Traffic jam monitoring system on Malang road junction using YoloV4," in Eighth International Conference of Mathematical Sciences: Icms2024, 2025, p. 050013. doi: 10.1063/5.0260130.

[21] C. N. Reddy, N. U. Kiran, and D. A. Sivasakthi, "Smart traffic monitoring system using YOLO," in AIP Conference Proceedings, 2025. doi: 10.1063/5.0266726.

[22] C. C. Chang, K. H. Huang, T. K. Lau, C. F. Huang, and C. H. Wang, "Using deep learning model integration to build a smart railway traffic safety monitoring system," Sci. Rep., vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-88830-7.

[23] M. I. Zulfiqar, A. Khalid, A. Siddig, M. J. Nawaz, and S. Saay, "Ai-driven smart shopping carts with real-time tracking and inventory forecasting for enhanced retail efficiency," IEEE Access, vol. 13, pp. 55576–55585, 2025, doi: 10.1109/ACCESS.2025.3553854.

[24] M. Wu et al., "Defect intelligent recognition of membrane product based on deep learning," Meas. Control (United Kingdom), vol. 58, no. 8, pp. 1052 – 1066, 2024, doi: 10.1177/00202940241268952.

[25] L. Zhu and X. Sheng, "On image-processing-based identification method of express logistics information," Trait. du Signal, vol. 39, no. 3, pp. 1019–1025, 2022, doi: 10.18280/ts.390329.

[26] R. Pravesh and B. C. Sahana, "A dual-stage deep learning framework for simultaneous fire and firearm detection in smart surveillance systems," Results Eng., vol. 27, 2025, doi: 10.1016/j.rineng.2025.106330.

[27] A. Ferone et al., "AiWatch: A distributed video surveillance system using artificial intelligence and digital twins technologies," Technologies, vol. 13, no. 5, 2025, doi: 10.3390/technologies13050195.

[28] H. C. Tsai, Y. W. P. Hong, and J. P. Sheu, "Completion time minimization for UAV-enabled surveillance over multiple restricted regions," IEEE Trans. Mob. Comput., vol. 22, no. 12, pp. 6907–6920, 2023, doi: 10.1109/TMC.2022.3200732.

[29] M. B. V. Bell, A. N. Radford, R. Rose, H. M. Wade, and A. R. Ridley, "The value of constant surveillance in a risky environment," Proc. R. Soc. B Biol. Sci., vol. 276, no. 1669, pp. 2997–3005, 2009, doi: 10.1098/rspb.2009.0276.

[30] K. Wang, Z.-H. Han, K.-S. Zhang, and W.-P. Song, "An efficient geometric constraint handling method for surrogate-based aerodynamic shape optimization," Eng. Appl. Comput. Fluid Mech., vol. 17, no. 1, 2023, doi: 10.1080/19942060.2022.2153173.

[31] A. Yi and N. Anantrasirichai, "A comprehensive study of object tracking in low-light environments," Sensors, vol. 24, no. 13, 2024, doi: 10.3390/s24134359.

[32] H. Tang et al., "Target search for joint local and high-level semantic information based on image preprocessing enhancement in indoor low-light environments," ISPRS Int. J. Geo-Information, vol. 12, no. 10, 2023, doi: 10.3390/ijgi12100400.

[33] J. Zhang, J. Peng, X. Kong, S. Wang, and J. Hu, "Vehicle spatiotemporal distribution identification in low-light environment based on image enhancement and object detection," Adv. Eng. Informatics, vol. 65, 2025, doi: 10.1016/j.aei.2025.103165.

[34] Z. Xiong, M. Wang, R. Kan, and J. Zhang, "YOLO-SAR: An enhanced multi-scale ship detection method in low-light environments," Appl. Sci., vol. 15, no. 13, 2025, doi: 10.3390/app15137288.

[35] S. Agrawal, R. Panda, P. K. Mishro, and A. Abraham, "A novel joint histogram equalization based image contrast enhancement," J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 4, pp. 1172–1182, 2022, doi: 10.1016/j.jksuci.2019.05.010.

[36] L. Liu, Z. F. Pang, and Y. Duan, "Retinex based on exponent-type total variation scheme," Inverse Probl. Imaging, vol. 12, no. 5, pp. 1199–1217, 2018, doi: 10.3934/ipi.2018050.

[37] K. Chang and L. Yan, "LLNet: A fusion classification network for land localization in real-world scenarios," Remote Sens., vol. 14, no. 8, 2022, doi: 10.3390/rs14081876.

[38] R. Ye, G. Shao, Z. Yang, Y. Sun, Q. Gao, and T. Li, Detection model of tea disease severity under low light intensity based on YOLOv8 and EnlightenGAN, vol. 13, no. 10. 2024. doi: 10.3390/plants13101377.

[39] C. Guo et al., "Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Jun. 2020, pp. 1777–1786. doi: 10.1109/CVPR42600.2020.00185.

[40] G. Wang, Y. Guo, and T. Hu, "Enhancing low-light object detection through domain adaptation and image enhancement," Vis. Comput., vol. 41, no. 14, pp. 11947–11958, 2025, doi: 10.1007/s00371-025-04136-9.

[41] Y. Liu, S. Li, L. Zhou, H. Liu, and Z. Li, "Dark-YOLO: A low-light object detection algorithm integrating multiple attention mechanisms," Appl. Sci., vol. 15, no. 9, 2025, doi: 10.3390/app15095170.

[42] Y. P. Loh and C. S. Chan, "Getting to know low-light images with the Exclusively Dark dataset," Comput. Vis. Image Underst., vol. 178, pp. 30–42, Jan. 2019, doi: 10.1016/j.cviu.2018.10.010.

[43] P. P. Anoop and R. Deivanathan, "Advancements in low light image enhancement techniques and recent applications," J. Vis. Commun. Image Represent., vol. 103, p. 104223, Aug. 2024, doi: 10.1016/j.jvcir.2024.104223.

[44] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," in 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Jun. 2023, pp. 7464–7475. doi: 10.1109/CVPR52729.2023.00721.

[45] Z. Cui et al., "You Only Need 90K Parameters to Adapt Light: a Light Weight Transformer for Image Enhancement and Exposure Correction," BMVC 2022 - 33rd Br. Mach. Vis. Conf. Proc., 2022, doi: https://doi.org/10.48550/arXiv.2205.14871.

[46] H. Guo, T. Lu, and Y. Wu, "Dynamic low-light image enhancement for object detection via end-to-end training," in Proceedings - International Conference on Pattern Recognition, 2020, pp. 5611–5618. doi: 10.1109/ICPR48806.2021.9412802.