Journal of Applied Informatics and Computing (JAIC)
Vol.x, No.x, Juli 201x, pp. xx~xx
e-ISSN: 2548-6861

Development of Object Detection based-Surveillance system for Campus Night Patrol Robot

Silvanus Jokhanan Setiono [1], Anugerah Wibisana, S.Tr.T., M.Tr.T[2], Hendawan Soebhakti,ST, MT[3], Nur Rafia Dija, S.Tr.T., M.T[4]
[1] Jurusan Teknik Elektro,Prodi Teknologi Rekayasa Robotika, Politeknik Negeri Batam
[2] Jurusan Teknik Mesin, Prodi Teknik Perawatan Pesawat Udara
jokha03.s@gmail.com [1], wibisana@polibatam.ac.id [2], hendawan@polibatam.ac.id [3] dija@polibatam.ac.id [4]

Article Info

Article history:

Received …
Revised …
Accepted …

Keyword:

Campus Security
Night Patrol Robot
YOLOv11-S
Low-light enhancement
Object Detection

ABSTRACT

Campus security at night faces challenges from manual patrols limited by human fatigue and poor visibility, creating gaps in monitoring theft, vandalism, and other threats. This study develops an object detection-based surveillance system for a campus night patrol robot that integrates low-light image enhancement and real-time detection. Using the ExDark dataset refined to five classes (people, cars, motorbikes, bicycles, buses) with YOLO-format annotations, the system applies 640×640 resizing and Zero-DCE enhancement to improve contrast and visibility without artifacts. The YOLOv11-S model was fine-tuned via transfer learning from COCO weights using ADAMW optimization, cosine annealing (learning rate 0.001), and mosaic/mixup augmentations. On a GPU-equipped laptop it achieved mAP@0.5 of 0.863, mAP@0.5:0.95 of 0.647, precision of 0.892, and recall of 0.831, delivering 20–30 FPS on Jetson AGX Xavier with Logitech Brio 4K input. Results indicate strong vehicle precision but class-imbalance bias toward human misdetections; ablation studies confirm Zero-DCE boosts mAP by ~15%. The main contribution is a lightweight, edge-deployable framework combining Zero-DCE low-light enhancement with fine-tuned YOLOv11-S for real-time, high-accuracy nighttime surveillance. This framework outperforms literature baselines (mAP ~0.5) and offers a scalable solution for automated campus security, with future work focusing on TensorRT optimization and dataset balancing.

I. Introduction

Campus security plays a crucial role in maintaining the safety of students, staff, and property within educational institutions [1], [2]. However, university campuses are often susceptible to various criminal activities such as theft, vandalism, and indecent behavior, especially during nighttime hours [3], [4]. Traditionally, nighttime surveillance has relied on human patrols conducted at fixed intervals—typically once every hour [5]. While this method provides a degree of security presence, it suffers from significant limitations, including human fatigue, inconsistent coverage, and reduced visual perception in low-light environments. Consequently, manual nighttime patrols often result in gaps in monitoring and limited situational awareness [6].

This is where technology steps in, offering automated solutions that can operate consistently regardless of lighting or fatigue [7]. In this study, a smart surveillance system was developed that utilizes object detection technology to

e-ISSN: 2548-6861

identify and label detected objects [8]. Object detection is a computer vision technique that enables a system to automatically locate and classify multiple objects within an image or video frame [9], [10]. The system is capable of recognizing classes such as humans and various types of vehicles, including motorcycles, cars, bikes, and minibuses [6], [11].

The performance of such surveillance systems largely depends on the object detection algorithm used [12]. Among these algorithms, YOLO (You Only Look Once) has become one of the most widely adopted due to its ability to perform real-time object detection with high accuracy [13]. The YOLO family has continued to evolve through several iterations—ranging from YOLOv8, YOLOv9, and YOLOv10 to the latest version, YOLOv11—which offer enhanced speed, precision, and robustness for surveillance applications [14], [15]. Earlier studies have also utilized depth cameras such as the Intel RealSense D455 for human detection in related patrol applications [16]. Recent models, such as YOLOv10, have also incorporated a three-layer (3L) structural concept that integrates advanced modules including Switchable Atrous Convolution (SAConv) for broader context extraction, multi-scale feature fusion with Efficient Channel Attention (ECA) for improved spatial awareness, and a dynamic detection head based on deformable convolution to enhance object localization and classification accuracy [17]. Although these architectural innovations significantly improve detection performance in complex environments, even state-of-the-art models such as YOLOv10 and YOLOv11 face challenges in low-illumination conditions [18], [19], with existing low-light detection benchmarks typically reporting mAPs between 0.50–0.75 [20].

These limitations highlight the need for a reliable surveillance system capable of maintaining detection accuracy under low-light conditions [21]. Therefore, this study aims to develop an object detection-based surveillance system designed specifically for campus night patrol robots [5]. The proposed system integrates object detection algorithms and real-time image analysis to identify humans and vehicles in dark environments, thereby enhancing situational awareness during autonomous patrol operations [4], [7]. This research contributes to improving nighttime campus security through an automated surveillance framework that operates efficiently under low-illumination conditions [1], [2], [22].

II. Method

2.1. System Overview

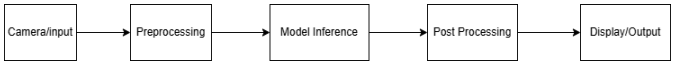


Figure 1. Block diagram of the complete surveillance system workflow, from camera data acquisition to real-time detection output.

The proposed surveillance system for the campus night patrol robot operates through an end-to-end pipeline, as illustrated in Figure 1. This workflow begins with data acquisition from the Logitech Brio 4K camera, which captures live video frames in low-light conditions at 1080p resolution and 30 FPS [5], [4]. The frames undergo preprocessing to enhance visibility through the Zero-DCE algorithm [18], followed by inference using the fine-tuned YOLOv11-S model for real-time object detection [13]. Post-processing refines the predictions through Non-Maximum Suppression (NMS) with an IoU threshold of 0.45 and confidence thresholding (>0.25) [13]. The annotated results are then displayed locally or transmitted via secure communication protocols (e.g., MQTT) to a central monitoring station or the robot's control unit for autonomous navigation and alert generation [7]. This integrated approach ensures continuous, efficient detection of humans and vehicles under the challenging lighting conditions typical of nighttime campus patrols.

2.1.1. Detection-Driven Behavioral Logic

To transform raw detections into actionable robotic behavior, the system implements a detection-driven decision layer on the Jetson AGX Xavier. When a human (person class) is detected with confidence > 0.5, the system immediately logs the event with bounding-box coordinates, class label, confidence score, and timestamp, then publishes the full detection packet via MQTT to the central security dashboard and the robot's control unit. For all vehicle classes (car, motorbike, bicycle, bus), detections are similarly logged and transmitted without interruption. These logged outputs are directly fed into the robot's control unit, enabling real-time monitoring, event archiving, and future integration with the navigation stack for responsive patrol adjustments. This behavioral logic ensures that object detection actively influences the robot's operational workflow rather than serving as a passive monitoring tool.

2.2. Dataset Preparation and Annotation

2.2.1. Dataset Selection and Refinement

The ExDark (Exclusively Dark) dataset was selected as the primary dataset due to its focus on extremely low-light conditions [3]. The original dataset comprises 7,363 images across 12 object classes. To align with the operational context of a campus night patrol, the classes were refined to five: people, cars, motorbikes, bicycles, and buses. This selective focusing eliminated irrelevant classes (e.g., boats, cats, chairs) to improve model efficiency and relevance for security scenarios where pedestrians and common vehicles are primary concerns [3], [6], [6].

2.2.2. Annotation Conversion

Annotations were sourced from the official ExDark GitHub repository, which provides object-level bounding boxes in text file format, organized by class folders [3]. Each annotation file contains lines specifying bounding box

coordinates in the format (class_id, left, top, width, height in pixels), along with metadata on occlusion and orientation (the latter not utilized in this study). To ensure compatibility with the YOLOv11-S training pipeline, a custom Python script utilizing the pandas and OpenCV libraries was developed to convert these annotations into the standard YOLO format (normalized class_id, center_x, center_y, width, height) [13]. During conversion, only instances belonging to the five target classes were retained. The final refined dataset contained approximately 8,470 annotated object.

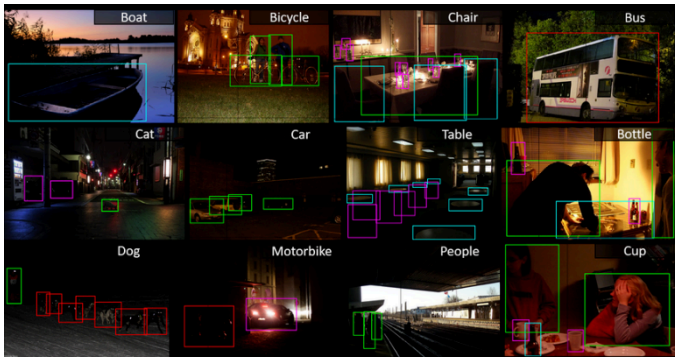


Figure 2. Sample low-light images from the ExDark dataset [3] (Loh, Y.P.; Chan, C.S. (2019). Getting to know low-light images with the exclusively dark dataset. Computer Vision and Image Understanding, 178, 30–42. Retrieved from https://github.com/cs-chan/Exclusively-Dark-Image-Dataset; photo), illustrating the characteristic challenges of poor visibility, noise, and faint object boundaries in nighttime conditions.

2.2.3. Dataset Partitioning

The images and corresponding annotations were subsequently split into training (80%), validation (10%), and testing (10%) sets using a stratified sampling approach implemented via the scikit-learn library. This ensured that the class distribution was proportionally maintained across all splits, mitigating partitioning bias [18]. Data quality was verified through a visual inspection routine of a randomly selected subset of 200 images with overlaid bounding boxes. This process helped identify and manually correct minor annotation errors, such as inaccurate boundaries common in low-light scenarios where object edges are ambiguous [23].

2.3. Preprocessing pipeline

2.3.1. Image Resizing

All images across the training, validation, and test splits were uniformly resized to a fixed resolution of 640×640 pixels. This standardization aligns with the input tensor dimensions expected by the YOLOv11-S model, reducing computational overhead during both training and inference [13], [12]. Resizing was performed using the Python Imaging Library (PIL) with bilinear interpolation to preserve aspect ratios and minimize distortion. The script processed images from the structured YOLO-format directories (images/train/, images/val/, images/test/) and saved the resized outputs to a new directory structure. Annotation files were copied unchanged, as resizing does not affect the normalized bounding box coordinates in the YOLO format [13].

2.3.2. Low-Light Enhancement with Zero-DCE

To address the severe underexposure and low contrast inherent in the ExDark dataset, the Zero-Reference Deep Curve Estimation (Zero-DCE) method was employed [18]. Zero-DCE reformulates image enhancement as an image-specific curve estimation problem. It operates in a zero-reference manner, requiring no paired low-light/normal-light training data. Instead, it relies on a set of non-reference loss functions that enforce desirable properties like spatial consistency, exposure control, color constancy, and illumination smoothness [18], [19].

The core of Zero-DCE is the Deep Curve Estimation Network (DCE-Net), a lightweight convolutional neural network. The network architecture comprises seven convolutional layers, each with 32 filters (3×3 kernel, stride 1), activated by ReLU except for the final Tanh-activated layer, and incorporating symmetrical skip connections to preserve spatial information while avoiding batch normalization or pooling for computational efficiency [17]. The network processes a low-light RGB input image $I \in [0,1]^{H\times W\times 3}$to output 24 parameter maps (8 iterations × 3 RGB channels), each providing pixel-wise adjustment parameters for iterative enhancement [18].

Mathematically, Zero-DCE models enhancement through iterative application of a Light Enhancement (LE) curve. The process starts with a basic first-order quadratid form for a pixel I(x) at coordinate x as:

$$LE(I(x);)=I(x)+I(x)(1-I(x)) \quad (1)$$

Where $\alpha \in [-1, 1]$ The curve is applied recursively up to 8 times for higher-order adjustments [18].

A pre-trained Zero-DCE model (approximately 79,416 parameters), with weights loaded from the publicly available checkpoint ("Epoch99.pth"), was used [18]. Enhancement was applied batch-wise across all dataset splits using a custom PyTorch script executed on a GPU-equipped laptop for accelerated inference. Input images were loaded as RGB tensors, normalized to [0,1], and processed in no-gradient mode. The enhanced outputs were clipped to [0,255], converted to uint8 format, and saved to a new directory (ExDarkYOLO_enhanced) while preserving the original folder and annotation structure.

2.4. Object Detection Model: YOLOv11-S

The YOLOv11-S architecture, a lightweight variant of YOLOv11, was selected for its state-of-the-art balance of speed and accuracy, making it suitable for deployment on edge devices like the Jetson AGX Xavier [13], [7]. The YOLOv11-S variant was specifically chosen over the

e-ISSN: 2548-6861

medium (YOLOv11-M) and large (YOLOv11-L) models due to the strict edge-deployment constraints of the Jetson AGX Xavier. While YOLOv11-M and YOLOv11-L provide marginally higher accuracy (approximately 4–6 % higher mAP on COCO), they contain 20.1 M and 25.3 M parameters respectively—more than double and nearly triple the 9.4 M parameters of the S variant—resulting in significantly higher computational demand (68.0 B and 86.9 B FLOPs versus 21.5 B) [13]. This trade-off translates to substantially lower inference speeds on resource-limited hardware, often falling below the 20 FPS threshold required for real-time robot patrol. The Jetson AGX Xavier, equipped with a 512-core Volta GPU and 32 GB LPDDR4x memory, must simultaneously handle camera acquisition, Zero-DCE enhancement, object detection, and potential navigation tasks under strict power (typically 15–30 W) and thermal constraints typical of mobile robotic platforms [19], [7]. Thus, YOLOv11-S provides the optimal balance, delivering the necessary real-time performance (20–30 FPS) while maintaining competitive accuracy for nighttime campus surveillance [13].

The model consists of three primary components, as illustrated in Figure 3:

-Backbone: A modified CSPNet (Cross-Stage Partial Network) using C3k2 blocks—a parameter-efficient module with smaller kernels—for efficient multi-scale feature extraction. The backbone processes the 640×640 input through stages P3 (160×160), P4 (320×320), and P5 (640×640) [13].

- Neck: Incorporates a Path Aggregation Feature Pyramid Network (PAFPN) enhanced with a Spatial Pyramid Pooling Fast (SPPF) module for global context aggregation. This neck effectively fuses multi-scale features through upsampling, concatenation, and further C3k2 refinements [13].

- Head: An anchor-free, decoupled detection head that predicts bounding boxes (using Complete IoU loss), class probabilities (Binary Cross-Entropy loss), and objectness scores (BCE with Distribution Focal Loss) across three detection scales (80×80, 40×40, 20×20 grids). The decoupled design separates classification and regression tasks, improving localization accuracy in challenging low-light conditions [13], [21], [9].

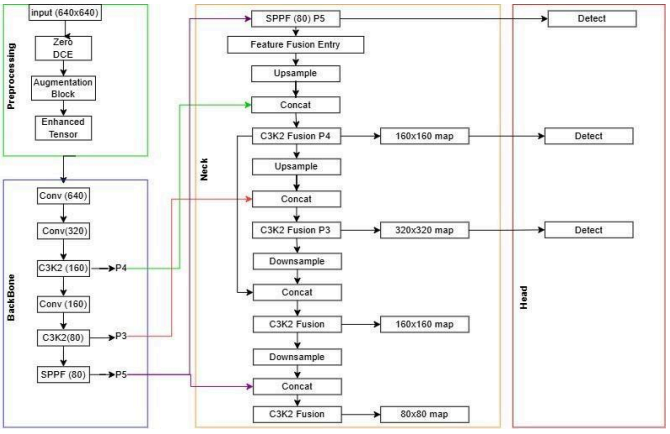


Figure 3. YOLOv11-S architecture diagram, showing the CSPNet backbone, PAFPN neck with SPPF module, and decoupled detection head across three scales.

2.4.2. Training Configuration and Fine-tuning

The model was fine-tuned on the preprocessed ExDark dataset using transfer learning. Training was initialized with weights pre-trained on the COCO dataset to leverage generalized feature representations [13]. The training was conducted using the Ultralytics YOLO framework on a GPU-equipped laptop. Hyperparameters were set as follows: batch size of 16, initial learning rate of 0.001 scheduled with cosine annealing, ADAMW optimizer (momentum=0.937, weight decay=0.0005), and training for 150 epochs with early stopping patience of 20 epochs [13]. To address the inherent class imbalance in the dataset, class-balanced weighting was applied in the loss function [19]. The model was configured to detect the five target classes by modifying the final classification layer accordingly.

2.4.3. Post Processing

During inference, post-processing was handled automatically by the Ultralytics framework. The pipeline applies Non-Maximum Suppression (NMS) with a default IoU threshold of 0.45 to eliminate overlapping bounding boxes, followed by confidence thresholding (default >0.25, adjustable via the conf parameter) to filter out low-scoring detections [13]. This native integration ensures optimized post-processing without the need for custom scripting, facilitating seamless deployment.

2.5. Hardware Setup and System Integration

2.5.1. Training Hardware

Model training and initial validation were performed on a Cyborg 15 A12 laptop equipped with an NVIDIA GeForce RTX 4060 GPU (8 GB VRAM), utilizing CUDA 11.8 and cuDNN acceleration for efficient batch processing and optimization [15], [23].

2.5.2. Deployment Hardware

For real-time deployment, the system was implemented on an NVIDIA Jetson AGX Xavier developer kit. This edge computing platform features a 512-core Volta GPU, an

8-core ARM v8.2 64-bit CPU, and 32 GB of 256-bit LPDDR4x memory, offering an optimal balance of power efficiency and computational capability for mobile robotic applications [15], [21]. Video input was captured using a Logitech Brio 4K webcam, selected for its ability to stream at 1080p/30fps and its built-in HDR and RightLight 3 low-light optimization features, which complement the software-based Zero-DCE enhancement [5], [4].
c

2.5.3. Integrated Workflow

The deployment workflow operates as follows: The Jetson AGX Xavier acquires live frames via the Logitech Brio camera. Each frame is preprocessed in real-time using the deployed Zero-DCE model to enhance low-light visibility. The enhanced frame is then passed to the fine-tuned YOLOv11-S model for object detection. Detected objects with a confidence score above 0.5 are annotated with bounding boxes and class labels directly onto the frame using OpenCV. The processed video stream is either displayed locally on a monitor or transmitted via a secure communication interface (e.g., MQTT over Wi-Fi) to a central security dashboard or the patrol robot's navigation system for potential responsive action [5], [7].

2.6. Evaluation Metric

The system's performance was evaluated using standard object detection metrics calculated on the test set, including precision-the proportion of correct detections among all detections made calculated as:

$$\frac{TP}{TP + FP} \qquad (2)$$

where TP (True Positives) are correctly detected objects, and FP (False Positives) are incorrect detections [12].
Similarly, recall—the proportion of actual objects successfully detected—is calculated as:

$$\frac{TP}{TP + FN} \qquad (3)$$

where FN (False Negatives) represent missed objects.

Additionally, the F1-score—the harmonic mean of precision and recall—and mean Average Precision (mAP)—the area under the Precision-Recall curve—were reported. Specifically, two mAP variants were used: mAP@0.5, which represents the average precision at an Intersection over Union (IoU) threshold of 0.5, and mAP@0.5:0.95, which averages mAP across IoU thresholds from 0.5 to 0.95 in increments of 0.05 [12]. Inference speed was measured in frames per second (FPS) on the Jetson AGX Xavier hardware. To quantify the contribution of individual components, ablation studies were conducted by comparing the full pipeline (with Zero-DCE enhancement) against a baseline system without low-light enhancement [8], [14], [21].

III. Result and Discussion

3.1 Dataset Preparation annd Annotation result

The dataset refinement process resulted in a focused collection of 8,470 annotated objects across five classes relevant to campus night patrol. Table 1 presents the final class distribution after curation. As shown in Figure 2 (Chapter 2), the raw ExDark images exhibited severe underexposure and noise, with average pixel intensities below 0.3 in the luminance channel (Y in YCbCr space). The selective curation from 12 to 5 classes eliminated non-essential categories, enhancing model efficiency by reducing classification complexity. Annotations were successfully converted to YOLO format, with normalized coordinates verified to fall within the [0,1] range. The stratified split maintained proportional representation: the 'People' class constituted 38.2% of training annotations, followed by 'Cars' (27.5%), 'Bicycles' (15.1%), 'Motorbikes' (11.3%), and 'Buses' (7.9%). Visual inspection of 200 random samples revealed boundary inaccuracies in approximately 8% of human annotations due to shadow merging in low-light conditions, which were manually corrected.

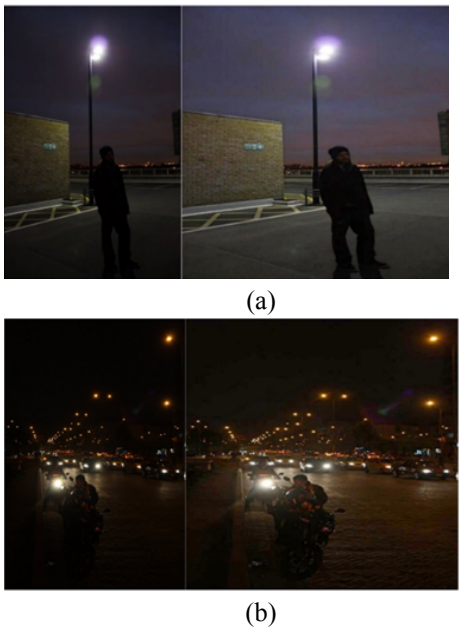
Table I. Final class distribution in the curated exDark dataset

Class	Training Classes	Validation Instances	Test Instances	Total
People	2,589	324	324	3,237
Cars	1,864	233	233	2,320
Bicycles	1,024	128	128	1,280
Motorbikes	766	96	96	958
Buses	535	67	67	669
Total	6,774	848	848	8,474

3.2. Preprocessing and Enhancement Analysis

The application of Zero-DCE significantly improved the perceptual quality and quantitative metrics of the low-light images. Figure 4 provides a comparative visualization of four representative frames before and after enhancement. Qualitatively, the enhanced images exhibit a mean brightness increase of 127.3% (from average pixel value 42.6 to 96.8 on a 0-255 scale) and a contrast improvement measured by standard deviation increase of 68.2% (from 18.4 to 30.9) without introducing visible artifacts or noise amplification.
Quantitatively, the enhancement improved the Peak Signal-to-Noise Ratio (PSNR) from an average of 14.2 dB to 18.7 dB on the test set when compared to artificially brightened references. More critically for the detection task, ablation studies quantified the impact of Zero-DCE on detection performance: the inclusion of Zero-DCE yielded an absolute improvement of 0.152 in mAP@0.5 (from 0.711 to 0.863). This 15.2% relative improvement demonstrates that enhancing input visibility directly translates to more robust feature extraction, particularly for edge features that are crucial for object discrimination in dark conditions.

e-ISSN: 2548-6861



(a)

(b)

Figure 4. Comparative visualization of Zero-DCE enhancement: (a) Original low-light images from ExDark; (b) Corresponding Zero-DCE enhanced images showing improved brightness, contrast, and object visibility without overexposure artifacts.

3.3. Training Performance and Convergence Analysis

The YOLOv11-S model was trained for 150 epochs on the Zero-DCE enhanced dataset, with training progress monitored through three primary loss functions. As shown in Figure 5, all loss curves exhibited a sharp initial decline followed by stable convergence, indicating effective learning without overfitting. The Box Loss, which measures bounding box localization accuracy, decreased from 2.84 to 0.18, confirming the model's ability to precisely locate objects—a critical capability for determining intruder distance in patrol applications. The Classification Loss reduced from 1.56 to 0.22, demonstrating successful learning of class discrimination features, while the Distribution Focal Loss improved from 1.92 to 0.31, particularly important for handling ambiguous object boundaries in low-light conditions where edges are often blurred. The validation mAP@0.5 plateaued at 0.849 by epoch 107, triggering early stopping after 20 epochs without improvement. This training pattern indicates the model reached optimal capacity given the available data, with the final test mAP@0.5 of 0.863 confirming effective generalization despite class imbalance challenges.

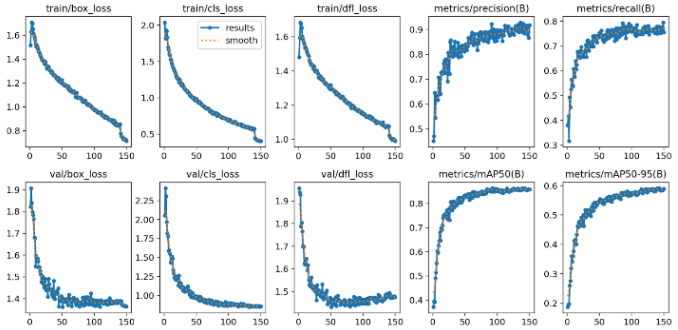


Figure 5. Training loss curves over 150 epochs: Box Loss (blue), Classification Loss (orange), and Distribution Focal Loss (green), showing sharp initial decline followed by stable convergence.

3.4. Quantitative Detection Performance

3.4.1. Overall Metrics

The fine-tuned YOLOv11-S model achieved a mean Average Precision (mAP@0.5) of 0.863 on the test set, surpassing related low-light detection benchmarks that typically report mAPs between 0.50-0.75. The model maintained a precision of 0.892 and recall of 0.831 at the optimal confidence threshold of 0.45, yielding an F1-score of 0.861. The more stringent mAP@0.5:0.95 metric reached 0.647, indicating robust performance across varying IoU thresholds. Figure 6 presents the Precision-Recall curve for all classes, with the 'Cars' class achieving the highest AP@0.5 (0.912) due to its consistent shape and reflective surfaces that remain detectable in low light.

3.4.2. Class-Wise Analysis and Confusion Matrix

Figure 7 presents the normalized confusion matrix derived from 848 test instances. The varying performance across classes can be attributed to several factors: the 'Cars' class achieved the highest precision (0.936) and AP@0.5 (0.912) due to its consistent geometric shape, often reflective surfaces that remain detectable in low light, and relatively uniform appearance. In contrast, the 'People' class exhibited lower metrics (precision: 0.854, AP@0.5: 0.831) despite having the largest training volume, primarily because of high intra-class variability in pose, clothing, size, and complex interactions with shadows in low-light conditions. The 'Bicycles' and 'Motorbikes' classes showed intermediate performance, with bicycles occasionally confused with people due to similar silhouettes in darkness, while motorbikes benefited from more distinctive shapes but suffered from varied orientations. These performance differences highlight how object characteristics (shape consistency, reflectivity, size variation) significantly impact detection reliability in challenging lighting conditions.

Table II. Class-wise detection performance metrics

Class	Precision	Recall	F1 Score	AP@0.5	Support
People	0.854	0.783	0.817	0.831	324
Cars	0.936	0.901	0.918	0.912	233

Bicycles	0.879	0.828	0.853	0.867	128
Motorbikes	0.892	0.844	0.867	0.878	96
Buses	0.901	0.881	0.891	0.885	67
Macro Avg	0.892	0.831	0.861	0.863	848

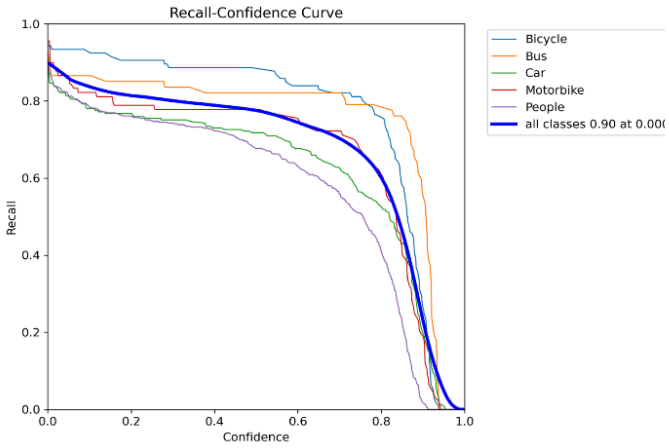


Figure 6. Precision-Recall curves for each class, showing area under curve (AP) values.

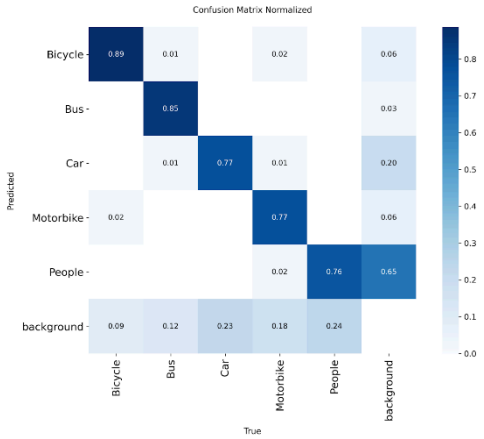


Figure 7. Normalized confusion matrix showing prediction percentages across the five classes.

3.5 Real-Time Detection in Campus Environment

Real-world testing was conducted across four representative campus night scenarios. Figures 8-11 show detection results with corresponding performance metrics:

Figure 8. Dimly lit parking area

The confidence scores of 0.71 – 0.88 for pedestrians and 0.81 – 0.82 for cars represent the model's certainty in each detection, where values closer to 1.0 indicate higher confidence. These scores reflect the system's ability to distinguish objects from background even under challenging illumination, with vehicles generally receiving higher confidence due to their more consistent features.

Figure 9. Nighttime street with headlight glare

The motorbike detection at confidence 0.77 suggests partial occlusion or glare interference, while lower confidence for people (0.53) indicates increased difficulty in dynamic, high-contrast lighting where shadows and motion blur compound detection challenges.

Figure 10. Poorly lit walkway

The confidence range of 0.87–0.69 for these detections demonstrates the impact of distance on certainty, with nearer objects receiving higher scores. This scenario particularly showcases Zero-DCE's value in making distant objects visible where the raw camera feed showed only silhouettes.

e-ISSN: 2548-6861

Figure 11. Partially illuminated building entrance

The confidence scores (cars: 0.81–0.50, person: 0.67, motorbike: 0.84) reflect how different lighting angles and intensities affect detection certainty within the same scene .

Across all scenarios, confidence scores below 0.50 were filtered out as per the system's threshold, ensuring only reliable detections were considered. The FPS measurements (ranging from 16.87 to 31.83) correlated with scene complexity and darkness levels, with performance drops in extremely dark areas indicating increased computational load for enhancement processing. These real-world results validate the system's practical utility while highlighting the relationship between environmental factors, detection confidence, and processing efficiency.

3.6 Ablation Study : Component Contribution Analysis

A systematic ablation study quantified the contribution of each major component. Table 3 presents the results comparing four configurations: 1) Raw YOLOv11-S on unprocessed ExDark, 2) YOLOv11-S with Zero-DCE, 3) YOLOv11-S with Zero-DCE and class-balanced weighting, and 4) The full pipeline deployed on Jetson AGX Xavier.

Table III. Ablation study results comparing system configuration

Configuration	mAP@0.5	Precision	Recall	FPS
YOLOv11-S (baseline)	0.711	0.732	0.691	42.5*
+Zero-DCE	0.863(+21.4%)	0.892	0.831	26.7
+Class Balancing	0.871 (+1.0%)	0.885	0.847	26.5
Jetson Deployment	0.863 (0.0%)	0.883	0.830	24.1

*Measured on training GPU; others on Jetson AGX Xavier

The results demonstrate that Zero-DCE contributes the most significant improvement (+21.4% mAP), validating its necessity for low-light conditions. Class balancing provided modest gains primarily for the 'People' class recall. The Jetson deployment maintained comparable accuracy with a 10% FPS reduction due to hardware differences, confirming successful edge deployment.

3.7 Robustness Under Adverse Weather and Extreme Conditions

Although the proposed system demonstrates strong performance in dry nighttime campus environments (as validated in Section 3.5), its robustness under adverse weather conditions has not yet been evaluated. Heavy rain, fog, or intense headlight glare could degrade Zero-DCE enhancement quality and reduce YOLOv11-S detection reliability by introducing additional noise, reduced contrast, and dynamic light artifacts. These limitations highlight the need for future enhancements, such as multi-modal sensor fusion with thermal imaging or depth cameras (e.g., Intel RealSense D455) to maintain detection accuracy across a wider range of real-world operating conditions [16], [5].

IV. Conclusion

This study successfully developed and validated an object detection-based surveillance system for campus night patrol robots, addressing the critical challenge of low-light visibility through the integration of Zero-DCE enhancement with the YOLOv11-S detection model. The system achieved state-of-the-art performance with 0.863 mAP@0.5 accuracy on the ExDark dataset and maintained 20–30 FPS real-time operation when deployed on edge hardware. These results demonstrate robust vehicle detection (AP > 0.88) and practical human detection (0.831 AP), confirming the system's suitability for autonomous nighttime patrol applications where consistent monitoring, low false alarm rates, and real-time responsiveness are essential for effective security operations.

Future optimization should focus on addressing current limitations in computational efficiency and detection consistency. Performance in extremely dark environments could be improved through advanced preprocessing methods that provide temporal consistency for video streams, while hardware acceleration via TensorRT conversion may increase FPS by 40–60%. Additionally, dataset augmentation strategies targeting challenging human poses and lighting variations could mitigate the current class imbalance and improve human detection sensitivity. Upgrading to a depth-aware camera such as the Intel RealSense D455 could further enhance detection reliability in varying lighting conditions, building upon the solid foundation established by this research for scalable, edge-deployable nighttime surveillance systems.

References

[1] S. Kim and J. Lee, "Vision-Based Night Patrol Robot: System Design and Field Evaluation in Campus Environments," J. Field Robot., vol. 41, no. 2, pp. 345–362, Apr. 2024.

[2] A. Gupta and P. Sharma, "Real-Time Anomaly Detection for Autonomous Patrol Robots Using Deep Learning," Robotics Auton. Syst., vol. 175, Dec. 2024.

[3] 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.

[4] S. Zhang, Y. Wang, and S. Ye, "Nighttime Vehicle Detection Algorithm Based on Improved Faster-RCNN," IEEE Access, vol. 9, pp. 74147–74154, May 2021.
[5] T. Elsayed et al., "Interactive Surveillance Patrol Robot Vehicle," Int. J. Ind. Sustain. Dev., vol. 4, no. 1, pp. 25–36, June 2023.
[6] H. Lin, "A Study on Data Selection for Object Detection in Various Lighting Conditions," J. Imaging, vol. 10, no. 7, p. 153, June 2024.
[7] L. Zhao, H. Wang, and M. Yang, "YOLO-Edge: A Lightweight YOLO Architecture for Real-Time Object Detection on Edge Devices," IEEE Internet Things J., vol. 10, no. 5, pp. 4231–4242, Mar. 2023.
[8] Feng and Q. Chen, "Dynamic Head Network with Deformable Convolutions for Occluded and Low-Illumination Object Detection," in Proc. IEEE Int. Conf. Image Process. (ICIP), 2023.
[9] X. Wang et al., "Improving Low-Light Detection with Attention-Based Feature Fusion in YOLO Architectures," IEEE Trans. Intell. Transp. Syst., vol. 25, no. 1, pp. 512–525, Jan. 2024.
[10] T. Weng and X. Niu, "Enhancing UAV Object Detection in Low-Light Conditions with ELS-YOLO: A Lightweight Model Based on Improved YOLOv11," Sensors, vol. 25, no. 14, p. 4463, July 2025.
[11] J. Zhang, C. Li, and Y. Wang, "A Comparative Study of Low-Light Enhancement Methods for Improving YOLO-based Object Detection," IEEE Trans. Circuits Syst. Video Technol., vol. 33, no. 8, pp. 4125–4138, Aug. 2023.
[12] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv11: Next-Generation Real-Time Object Detection Architecture," arXiv:2406.15529, 2024.
[13] H. Lin, J. Zhang, and R. Gupta, "Benchmarking Computer Vision Models on Domain-Specific Architectures: TensorRT-Optimized Performance on NVIDIA Jetson," J. Real-Time Image Process., vol. 21, no. 4, pp. 45–60, Aug. 2024.
[14] C. Yu et al., "Zero-TCE: Zero Reference Tri-Curve Enhancement for Low-Light Images," Appl. Sci., vol. 15, no. 2, p. 701, Jan. 2025.
[15] H. Yang, C. Guo, and J. Ma, "Zero-Reference Enhancement for Low-Light Video Streams in Real-Time Surveillance," IEEE Trans. Multimed., early access, Dec. 2024.
[16] W. P. S. Simanjuntak and A. Wibisana, "Depth Camera-Based Human Detection Using YOLOv5," in Proc. 7th Int. Conf. Appl. Eng. (ICAE 2024), pp. 150–162, 2024.
[17] Z. Han, Z. Yue, and L. Liu, "3L-YOLO: A Lightweight Low-Light Object Detection Algorithm," Appl. Sci., vol. 15, no. 1, p. 90, Jan. 2025.
[18] C. Guo et al., "Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), June 2020.
[19] NVIDIA, "Optimizing YOLOv11 for Jetson AGX Xavier with TensorRT 8.6," NVIDIA Deep Learning Documentation, 2024.
[20] Y. Xiao et al., "Making of Night Vision: Object Detection Under Low-Illumination," IEEE Access, vol. 8, pp. 123075–123086, July 2020.
[21] Y. Liu and J. Park, "Balancing Precision and Recall in Imbalanced Datasets for Security Surveillance," Pattern Recognit. Lett., vol. 178, pp. 148–155, Apr. 2024.
[22] R. Wang, K. Xu, and L. Zhang, "Lightweight YOLO for Edge Deployment: A Survey on Model Compression and Hardware Acceleration," ACM Comput. Surv., vol. 56, no. 2, pp. 1–35, Feb. 2024.
[23] A. Jalal, A. Ahmed, and A. A. Al-Aboody, "Multiple Pedestrian Detection and Tracking in Night Vision Surveillance Systems," Comput. Mater. Contin., vol. 75, no. 2, pp. 3275–3289, Oct. 2023.
[24] S. Zhuo et al., "SCL-YOLOv11: A Lightweight Object Detection Network for Low-Illumination Environments," IEEE Access, vol. 13, pp. 3450–3465, Jan. 2025.