

An Efficient and Robust Night-Time Surveillance Object Detection System Using YOLOv8 and High-Performance Computing



Monish Sai Krishna Namana^{*}, Budidi Udaya Kumar^{*}

Department of Electrical Electronics and Communication Engineering, Gandhi Institute of Technology and Management, Visakhapatnam 530045, India

Corresponding Author Email: mnamana@gitam.in

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijse.140611>

ABSTRACT

Received: 9 September 2024

Revised: 26 October 2024

Accepted: 26 November 2024

Available online: 31 December 2024

Keywords:

object detection, YOLOv8, high-performance computing, deep learning, surveillance

Surveillance is critical in ensuring security and safety in the modern world, with daytime surveillance systems operating under favorable conditions. In contrast, nighttime surveillance presents a far more complex challenge due to reduced visibility and varying lighting conditions. Performing object detection at night poses a distinct challenge due to limited visibility and low-light conditions. Most conventional object detection methodologies struggle to perform accurately for small and distant objects in low light conditions, leading to deep learning techniques like You Only Look Once (YOLO), which gained significant attention due to its ability to achieve higher detection accuracies at incredible speeds. Advancements in deep learning methods have led to substantial improvements, but object detection in low-light scenarios remains a formidable challenge. This article proposes a method based on YOLOv8, a computer vision and deep learning-based technology on the ExDark dataset, which is a collection of images in low-light conditions using the maximum resources of high-performance computing (HPC) for an enhanced object detection model. The study demonstrates the effectiveness of YOLOv8 in overcoming limitations seen in other object detection frameworks, including Fast R-CNN, Faster R-CNN, and SSD, by utilizing a real-time processing approach that maintains high accuracy even under challenging conditions. This study explores various image augmentation techniques, optimization strategies, hyperparameters tuning, and optional techniques to improve the model's detection capabilities, further increasing object detection robustness in surveillance tasks. The results of this study showcase that the object detection effectiveness of YOLOv8 is promising and achieves a significant Precision of 0.908, Recall of 0.819, and accuracy mAP of 0.886, as well as other metrics in low light and nighttime surveillance for small and distant objects. This study contributes to the ongoing development of intelligent surveillance systems by comprehensively evaluating YOLOv8 every model's nano(n), small(s), medium(m), large (l), and extra-large (x) performance in low-light conditions. It offers insights into improving object detection in critical security operations.

1. INTRODUCTION

Surveillance systems are vital in playing a pivotal role in ensuring safety and security across various environments in the present-day world. Surveillance technologies help monitor activities, roads, urban centers, private properties, critical infrastructures, and many more by responding swiftly to potential threats [1]. Advancements in computer vision technology have significantly impacted the development of surveillance systems, enabling automated and real-time object detection. However, one of the persistent challenges in surveillance systems is maintaining high performance under nighttime and dark conditions, where low light levels and poor visibility will severely impact the detection capabilities. Nighttime object detection is even more essential for security purposes. Object detection at nighttime and in dark scenarios poses unique challenges and requires particular approaches to

ensure reliable performance. Traditional object detection models are not considered as they often struggle in dark and low-light conditions due to factors like occluding objects and shadows. These challenges highlight the need for robust object detection models for nighttime surveillance.

Nighttime surveillance is challenging based on many environmental, technological, and physiological factors. The primary factor is the low-light conditions, but there are many other interrelated factors, such as shadows, glare from artificial lighting, motion blur due to slower camera shutter speeds, and occlusion by objects that are harder to distinguish at night. These factors make the accurate detection and identification of objects much more complex in low-light conditions, leading to higher failure rates in object detection systems. Low-light conditions often introduce visual noise, reducing the accuracy of the object detection algorithm and causing the system to miss a few objects or create false

positives, where non-objects are misidentified or misclassified as significant objects. Objects in motion at night are challenging to detect due to the need for more prolonged camera exposure, leading to motion blur. Objects may become completely distorted or unrecognizable due to this visual blur. Similarly, occlusion due to darkness can hide objects, creating blind spots in the surveillance footage.

Research on nighttime object detection models by Wang et al. [2] showcases that traditional object detection models like Faster R-CNN exhibit a significant decline in accuracy and other metrics in low-light conditions. Moch and Supriyanto [3] studied thermal image-based object detection and showcased that, despite using advanced deep learning models such as YOLOv4, detection accuracy decreased when distinguishing between similar heat signatures at night. It is especially problematic in detecting industrial surveillance systems, where objects like machinery generate overlapping heat patterns. Liu et al. [4] studied YOLOv4 object detection was used for low-light surveillance tasks using the ExDark dataset. The model has showcased difficulty detecting objects that blended with dark backgrounds or were occluded by poor lighting conditions. Luo et al. [5] investigated the performance of the Single Shot Detector (SSD) model on nighttime traffic monitoring. This study indicated that glare from headlights and poor visibility have caused misclassifications in object categories like vehicles and pedestrians.

The ExDark dataset [6] is an exclusively dark dataset consisting of extremely low-light nighttime images that provide and serve as a critical resource for developing and evaluating object detection algorithms in low-light, dark, and nighttime conditions. ExDark dataset comprises standard RGB images with detailed texture information, which is crucial for accurately identifying and differentiating objects at nighttime. Alternative to RGB images are thermal images, also considered for nighttime object detection. They provide an alternative method of computing heat signatures and effectively detecting objects based on temperature variations [3]. Infrared thermal data images are essential in object detection when visual light cameras may not meet performance standards. Still, RGB images have several advantages over thermal images in surveillance applications as they enable the detection of a wide variety of objects with much greater detail and context, as thermal images often lack fine-grained details and color information present in RGB images. RGB cameras are generally more affordable and widely available than thermal imaging devices.

YOLO is one of the critical advancements in object detection, which uses a convolutional neural network (CNN), a deep learning model that revolutionized object detection into a single regression problem by directly predicting an image's bounding boxes and class probabilities in one stage. It is hence referred to as a one-stage detector [5]. This methodology enables YOLO to achieve high-speed performance without compromising accuracy, making it highly suitable for real-time applications. In real-time object detection scenarios, the YOLO algorithm has outperformed other detection methods like Fast R-CNN, Faster R-CNN, and SSD regarding precision, recall, accuracy, and other metrics [7]. Every iteration of the YOLO model improves its precision, accuracy, and object recognition capabilities [2]. Surveillance systems require real-time monitoring and analyzing video feeds, detecting and tracking objects such as people, vehicles, and potential threats. The rapid processing of video frames by the YOLO model's optimized design promotes increased

situational awareness and swift responses to suspicious activities [8].

This study presents an overview of implementing the YOLOv8 object detection model using the ExDark dataset for nighttime surveillance. Section 2 provides an outline of relevant research, i.e., a literature review. Section 3 explains the usage of high-performance computing systems and its benefits. Section 4 explains the implementation of our methods and their outcomes. Section 5 discusses the results and outcomes of the research study. Finally, Section 6 concludes this research study with a summary of our contributions and suggestions for future research studies.

2. RELATED WORK

Luo et al. [9] introduced a technique called Localization-aware Logit Mimicking (LaLM) based on knowledge distillation. The method aims to enhance object detection accuracy in challenging weather situations by minimizing the disparity between predictions from degraded photos and clean images. The method performs outstanding on three datasets, RTTS, ExDark, and RID, in extreme weather conditions. The methodology employed in this is a knowledge distillation technique. Kou et al. [10] proposed a compact two-stage transformer for improving the quality of low-light images. It also presents a framework for detecting dark objects by combining the enhancement model with YOLO. It achieves excellent low-light enhancement and dark object detection on datasets like Exdark while maintaining a small model size and real-time performance. The methodology applied is a two-stage transformer architecture. Secondly, it suggested the implementation of an FFT-Guidance Block (FGB) to extract frequency components and facilitate the retrieval of picture potential information. Almujally et al. [11] proposed a new framework for efficient and lightweight vehicle detection and tracking in nighttime aerial surveillance systems and for low-illumination conditions, using the YOLOv5 algorithm to achieve high precision scores in VisDrone and UAVDT datasets. Defogging and image enhancement using MIRNet are used for vehicle detection systems, and IDs are assigned to detected vehicles using SIFT features.

Abdusalomov et al. [12] proposed an improved YOLOv3 algorithm to enhance the accuracy of their fire detection method, address blurring issues at night, and develop a lightweight model that can run on embedded devices with low computational capabilities on massive datasets—employed data augmentation and contrast enhancement techniques on the images running the YOLOv3 network with pre-trained weights and evaluating the accuracy and predicting fire occurrence. Sharma et al. [13] presented a study using the YOLOv5 deep learning model to detect and classify vehicles, pedestrians, and traffic signals in real-time video under different weather conditions to improve object detection for autonomous vehicles. The dataset used in this study is the Roboflow self-driving car dataset. Liu et al. [4] presented a technique for enhancing object detection in visually degraded environments by incorporating a feature-guided module and an image enhancement branch which elevates the performance of networks on hazy, night-time, and underwater datasets like URPC2021, ChinaMM, RTTS, ExDark, and LLVIP. The methodology employed in this paper is a parallel architecture using a feature grid module for the enhancement branch and detection branch without additional computational costs

during testing. Xu et al. [14] proposed a new efficient object detection model FL-YOLO in coal mines using a cloud-edge collaboration framework for real-time intelligent video surveillance using the multi-scene and single-scene pedestrian datasets. The edge-cloud collaboration ensures low latency and high accuracy, enabling real-time response to video events and continuous model optimization to ensure production safety.

Song et al. [15] presented a hybrid deep learning framework for ship detection in remote sensing images, consisting of RSI dehaze, RSI enhancement, an image enhancement stage, a SplitShuffle module, and an improved YOLOv5s network. It is implemented on the HRSC2016 and DIOR datasets, achieving state-of-the-art performance. The Dehaze network maintains the colour balance. The hazy RSI Image enhancement network learns illumination mapping and recovers the RSI SplitShuffle module in the YOLOv5 backbone to reduce redundant features. Hui et al. [16] proposed a novel algorithm called WSA-YOLO for object detection in low-light environments, which decomposes the image, predicts image parameters, and adaptively enhances the image to improve detection performance. The WSA-YOLO algorithm demonstrates good generalization ability and performs well on the RTTS dataset, which contains low-light and foggy conditions. Xiao and Liao [17] proposed that LIDA-YOLO is an unsupervised domain adaptation method that improves upon YOLOv3 object detection by using local and global feature alignment modules to address data bias and achieves the highest mAP score of 56.65% on the ExDark dataset compared to other unsupervised methods while requiring fewer samples and having more vital generalization ability. It uses two main components: the multi-scale local feature alignment (MSLA) module and the multi-scale global feature alignment (MSGa) module. The MSLA module aligns perceptual fields. It reduces low-level feature differences between source and target domains. The MSGa module aligns overall image attributes and reduces feature biases like background, scene, and target layout.

Peng et al. [18] presented a new innovative object detection model, NLE-YOLO, incorporating a C2fLEFEM feature extraction module for detecting objects under low light conditions using the ExDark dataset. This paper presents a new Attentional Receptive Field Block (AMRFB) designed to increase the receptive field and improve feature extraction by substituting the SPPF module with the SimSPPF module to achieve quicker inference speed and improved feature representation. Chen et al. [19] proposed a target detection algorithm based on YOLOv7 for underwater resources called Underwater-YCC. It uses data augmentation, attention mechanisms, Conv2Former, and Wise-IoU techniques on the URPC2020 dataset. It has incorporated the CBAM attention mechanism into the Neck component to improve feature extraction and replaced the ELAN-F convolution block in the Neck with the Conv2Former module to better handle blurred underwater images. Bose et al. [20] introduce a novel real-time two-wheeler traffic rule violation detection system that can operate effectively in low-light conditions. The significant contributions in this study are the development of a comprehensive low-light dataset using the latest YOLOv8 model, a low-light enhancement module, and an integrated network of devices for efficient anomaly detection. The dataset used in the study is a custom dataset of two-wheeler traffic rule violations. The low-light videos are enhanced with OTSU's thresholding and CLAHE. The traffic violation

detection was performed using the YOLOv8 object detection algorithm.

Nie et al. [21] introduced a new dataset called TDND datasets, which comprises nighttime driving images from different countries and regions with various challenging visual conditions and evaluates the performance of six typical object detection methods on this dataset. Wang et al. [2] proposed a new object detection model called DK_YOLOv5 using the ExDark dataset and the expanded Mine_ExDark dataset to address the challenge of low detection accuracy in low-light environments, such as underground mines. Rahim et al. [22] proposed an efficient deep learning-based solution for real-time social distance monitoring in low-light environments using the YOLOv4 algorithm and a motionless time-of-flight camera. The YOLOv4 model was trained on the ExDARK low-light dataset to detect people. The model was trained and evaluated on two network sizes. Wang et al. [23] presented a vision-based system that efficiently detects crashes in mixed-traffic scenarios under very low-visibility conditions. The retina image enhancement technique was implemented to boost the image quality under low light conditions. The YOLOv3 model was designed to accurately identify and classify items from the dataset images, such as vehicles, pedestrians, and cyclists. A model was trained on features extracted from the YOLOv3 outputs to detect crashes. The dataset consists of many CCTV video clips collected online. Zhao et al. [24] proposed a real-time small object detection method for power line inspection in low-illumination environments using an adaptive transformer-based image enhancement module and a lightweight object detection model using YOLOv7.

These studies on low-light surveillance and object detection have significantly improved accuracy under challenging conditions. However, several limitations persist, such as the inability to detect small and occluded objects in shallow light conditions and the inability to detect distant objects. Many traditional object detection models, such as Faster R-CNN, Mask R-CNN, and SSD, struggle to maintain high performance in low-light environments, often due to inadequate feature extraction mechanisms for smaller objects. These models typically lack robust techniques for handling the extreme lighting variability encountered at night, which leads to low precision and recall metrics for detecting small or distant objects. In this study, in the subsequent sections, we discuss the implementation of YOLOv8, one of the finest object detection algorithms implemented on the ExDark dataset using the high computational resource HPC for small and distant object detection in low light surveillance applications.

3. HIGH PERFORMANCE COMPUTING

3.1 Introduction to high-performance computing

High-performance computing (HPC) is a powerful processor with parallel processing, usually called supercomputer, used to solve complex computational problems. These can analyze data and execute complex computations at high speeds, significantly above standard computer capabilities. HPC uses the technology of parallel multiprocessing, which mainly works on the power of multiple processors working in parallel to deliver exceptional processing speed and efficiency. The applications that require

extensive computation, large-scale data analysis, and high-speed processing that are crucial for various scientific, engineering, and business applications use these HPCs. HPC comprises Master Nodes and Compute Nodes. Master nodes function as the central control unit [25]. The master node is responsible for managing the distribution of computational tasks to the compute nodes, scheduling jobs, and monitoring the overall system performance. Compute nodes are the primary components of the HPC system, and they are equipped with powerful core CPUs and GPUs with ample memory to handle complex, extensive, and rigorous tasks. Compute nodes are responsible for performing complex computations and calculations. High-speed interconnect networks like Ethernet, InfiniBand, and Omni-path communication technologies are most commonly used for interconnection. These high-speed interconnect networks connect these nodes for rapid data exchange and communication. HPC is integrated with robust storage systems. These storage systems mainly store large datasets, complex computational tasks, and high-end software. Parallel file systems like Lustre and GPFS manage the enormous amount of generated data. Software like SLURM and PBS are utilized to manage task queues, and Linux-based Operating systems are used for effective resource management and job scheduling in the HPC system. Figure 1 describes the basic block diagram of high-performance computing.

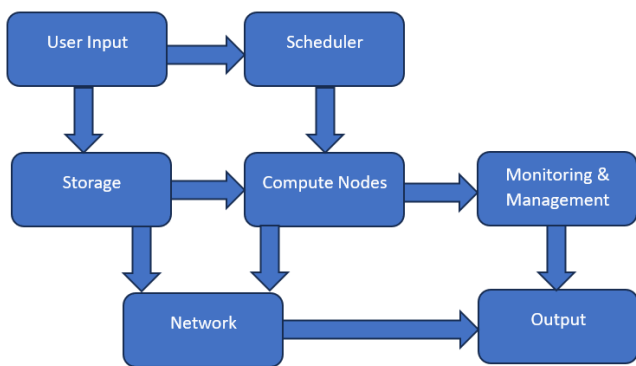


Figure 1. Block diagram of high-performance computing

3.2 Applications of high-performance computing

HPC systems are utilized in applications like artificial intelligence, machine learning, quantum mechanics, and other significant applications requiring complex mechanisms and heavy computational resources. The main advantage of an HPC is that it reduces computational time by providing unparalleled speeds and enabling the rapid processing of large datasets and complex computations, optimizing resource usage through parallel processing and high-speed interconnects, and ensuring efficient performance.

One of HPC's benefits is that it allows the addition of extra compute nodes to meet the growing computational demands. Scalability is an added advantage of HPC. HPC facilitates technological advancements that are not achievable using traditional computing capabilities.

High-performance computing offers the computational power and efficiency needed to tackle some of the most challenging problems. It is combined with the capabilities of master nodes, compute nodes, high-speed interconnects, large storage systems, and software, which, in return, deliver

exceptional performance, driving advancements and fostering innovation across various fields.

4. MODEL AND PERFORMANCE METRICS

4.1 YOLOv8 model

YOLO is an object detection framework developed by Ultralytics, known for its exceptional speed and accuracy. YOLO has revolutionized the field of computer vision by enabling real-time object detection. YOLO divides the input image into an SxS grid. Each grid from that image predicts a set number of bounding boxes and confidence scores, and thus, those confidence scores of the bounding boxes indicate the probability that the bounding box contains an object. YOLO object detection mechanism is efficient because it simultaneously predicts multiple bounding boxes and class probabilities for every grid cell [26]. During inference, these predictions are processed using non-maximal suppression (NMS) to eliminate the redundant boxes and retain only the most accurate and precise ones.

4.2 Architecture of YOLOv8 model

The architecture of YOLO is divided into three main blocks: the backbone, neck, and head model illustrated in Figure 2. The backbone of YOLOv8 is used for feature extraction, which extracts features from the input images using a convolution neural network (CNN) structure with multiple layers of convolutions, batch normalization, and activation functions. YOLOv8 uses Cross-Stage Partial Darknet (CSPDarknet), the optimized version of Darknet-53, as its backbone [27]. The YOLOv8 neck uses the Path Aggregation Network (PANet) to aggregate features from different scales, enabling the detection of objects of various sizes. The neck comprises Spatial Pyramid Pooling (SPP) and Feature Pyramid Networks (FPN) [28]. The head of YOLOv8 processes the aggregated features from the neck and produces prediction maps. These maps include the coordinates of bounding boxes, confidence scores indicating the object's presence, and each object's classification scores.

4.3 Performance metrics

Following the training, validation, and testing, a series of well-defined metrics are considered to calculate all the aspects of the model's accuracy and effectiveness. Understanding these metrics is essential to the model's evaluation as it is crucial for interpreting the results. The primary performance metrics in the YOLO models are precision, recall, mean average precision (mAP), and F1 score. These metrics are defined as

$$Precision = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Positive(FP)} \quad (1)$$

$$Recall = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Negatives(FN)} \quad (2)$$

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (3)$$

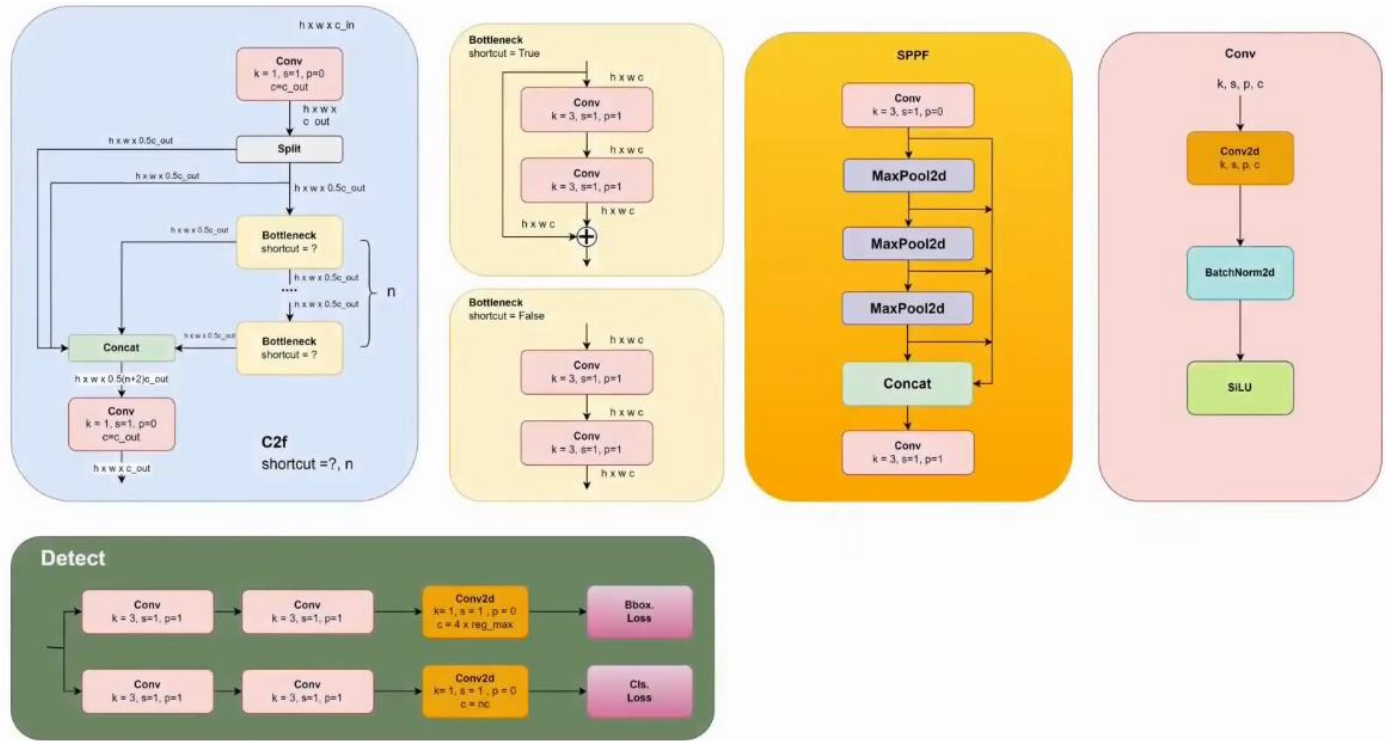


Figure 2. Architecture of YOLOv8 [29]

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (4)$$

5. EXPERIMENT AND RESULTS

5.1 Dataset and HPC configuration

This study uses the Exdark dataset for object detection using YOLOv8. The ExDark dataset comprises of 7344 images divided into 70% training, 20% validation, and 10% testing purposes. The ExDark dataset is primarily not in YOLO format. First, it is converted into a YOLO format, with images and bounding box annotation files, and the class IDs are generated. The training of this dataset is conducted using high-performance computing (HPC), which has GPU Nodes that are equipped with Dual Intel® Xeon® Platinum 8358 processors and four Nvidia A100 GPUs, each with 80 GB of memory with storage of the HPC system with 200 TiB usable capacity. Before the model training, hyperparameters are adjusted to help determine more optimal parameters for training the dataset. Table 1 describes the hyperparameters used to train the ExDark dataset using the YOLOv8 algorithm.

Table 1. Parameter configuration

Parameter Name	Value
Epochs	150
Image size	640
Batch size	32
Learning rate	0.01
Optimizer	AdamW

5.2 Techniques

Several modifications and configurations were implemented to customize the YOLOv8 model for the ExDark

dataset using high-performance computing (HPC) environments to enhance its efficacy in nighttime surveillance applications. The ExDark dataset contains 12 object classes: bicycle, boat, bottle, bus, car, cat, chair, cup, dog, motorbike, people, and table, in which only seven classes relevant to surveillance applications were selected for this study. Those classes include bicycle, bus, car, cat, dog, motorbike, and people, as they represent the key objects typically encountered in real-time monitoring for security and urban surveillance scenarios. The training is executed with the Visual Studio Code (VSCode), remotely connected to high-performance computing (HPC). Python 3.10.7 environment is used, and CUDA 11.8 and cuDNN 11.8 are used to enable GPU-accelerated training. The training was conducted across all YOLOv8 models nano(n), small (s), medium (m), large (l), and extra large(x), with extensive augmentation techniques employed to address the challenges of low-light conditions. The augmentation techniques include horizontal and vertical flips, random brightness contrast adjustments, shift-scale-rotate transformations, coarse dropout, random resized crop, hue saturation value adjustments, RGB shifts, motion blur, Gaussian blur, and CLAHE. Additionally, K-Means clustering was applied to optimize anchor boxes, allowing the model to detect objects more accurately by adjusting bounding box sizes to better fit the dataset's object dimensions.

To further enhance model performance, hyperparameter tuning was conducted using the Optuna framework, refining critical parameters such as learning rate, weight decay, and learning rate factor. AdamW (Adaptive Moment Estimation with Weight Decay) optimization algorithm is used to combine the advantages of AdaGrAD and RMSProp for improved training and generalization capabilities of the model. AdamW optimizer uses a Cosine Annealing learning rate scheduler, which helps in dynamic adjustments of the learning rate through the training process, thereby improving model convergence. The training process is executed with hyperparameters tuned with parameter configuration, with

image size 640, batch size 32, and epochs 150, ensuring comprehensive model learning. The early stopping method is implemented in the training process to automatically halt the training when there is no further improvement in the model's performance, preventing overfitting and ensuring efficient use of high-performance computing resources. After completion of training and validation, the best-performing model is selected for testing, where its performance is rigorously evaluated and metrics are calculated. Using advanced optimization techniques highlights the model's ability to address the unique challenges of low-light object detection, contributing to advancements in nighttime surveillance

systems. Figure 3 illustrates the training outcomes of the YOLOv8m model on the ExDark dataset. The Figure 3 showcases the model's ability and effectiveness in the training process, where the model learned to detect and classify objects across the ExDark dataset's seven classes. In the Exdark dataset, the classes are indexed explicitly as 0 for bicycle, 1 for bus, 2 for car, 3 for cat, 4 for dog, 5 for motorbike, and 6 for person. The results depicted in Figure 3 suggest that the training was adequate, whereby the model could classify the objects according to the designated classes and successfully adapted to the unique challenges of low-light object detection in nighttime surveillance.

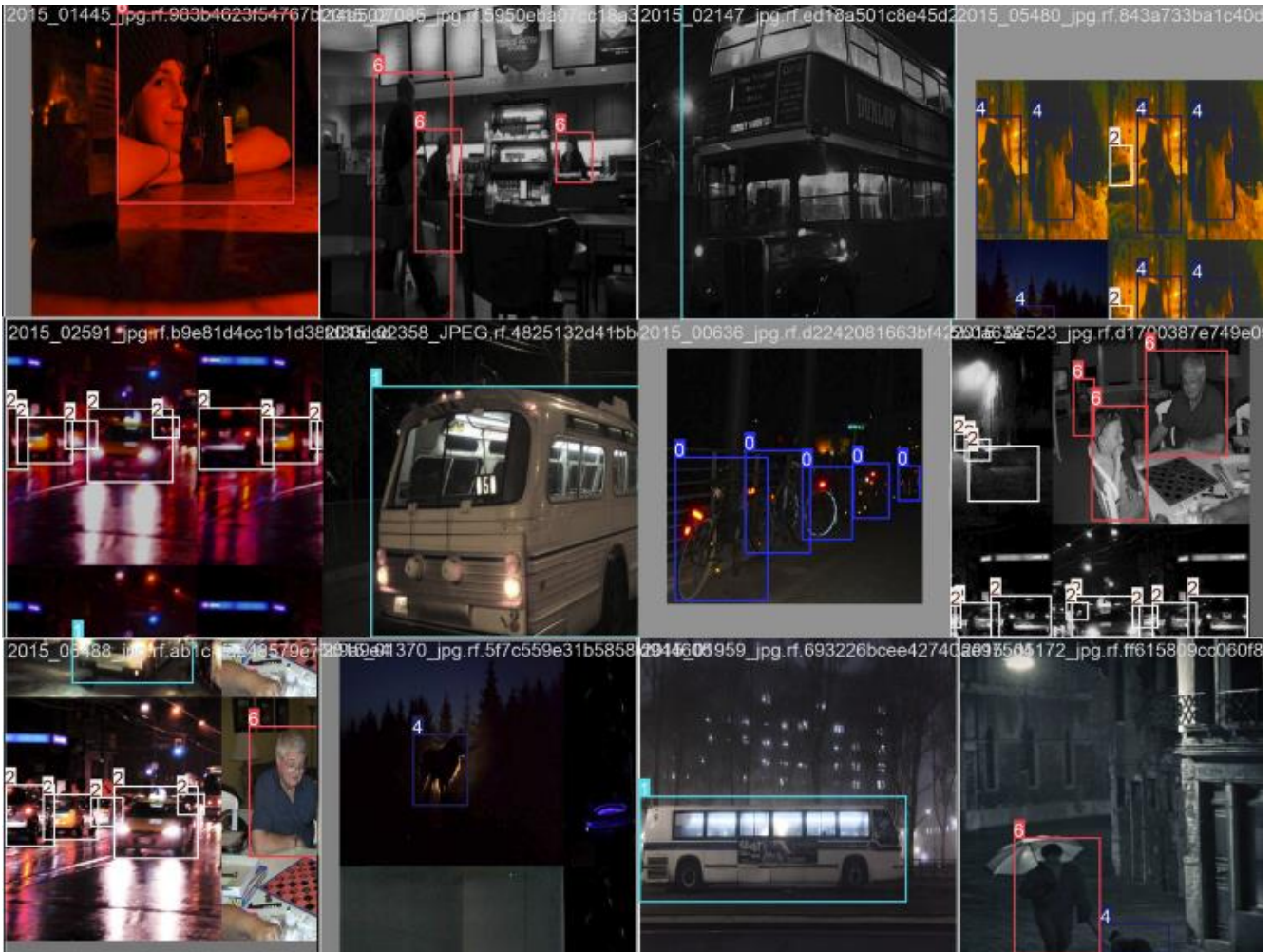


Figure 3. The training batch results of ExDark dataset using YOLOv8m

5.3 Results and discussions

The precision-recall (PR) curve for the YOLOv8m object detection model using the ExDark dataset is illustrated in Figure 4. The mAP@0.5 mean Average Precision for all classes is 0.895, indicating that the model is performing well across all classes, with one or two underperforming classes due to extremely low light conditions. The PR curve remains relatively smooth and stable, indicating that the YOLOv8m model achieves an ideal balance between precision and recall. This balance in the metrics is essential for achieving good object detection in low-light surveillance applications, where missing a critical object can have significant security implications, and false positives can trigger unnecessary alerts. The precision-recall curve remains consistently high for larger objects like Bus and Car, with mAP@0.5 values of 0.971 and

0.878, respectively. These results indicate that the model effectively detects these classes in extremely low light conditions. Smaller or more ambiguous objects like cats and dogs can be detected under very low light conditions. The mAP@0.5 for cats is 0.888, and the dog is 0.956. This showcases that the proposed YOLOv8m model performs well and accurately detects smaller objects with less misclassification.

As illustrated in Figure 5, the relationship between the F1 score and the confidence thresholds is depicted for all the object classes in the ExDark dataset. The blue line indicates the aggregate performance of all the classes, with the highest F1 score being 0.85 at a confidence level of 0.468, thus indicating that the overall model's detection performance is optimal at this confidence threshold.

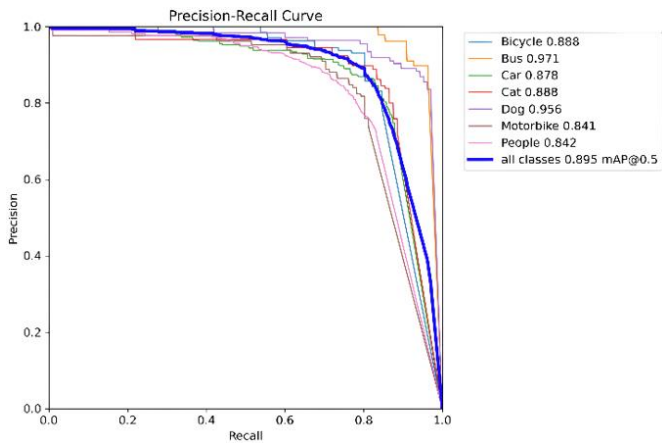


Figure 4. Precision-recall confidence curve of the ExDark dataset using YOLOv8m

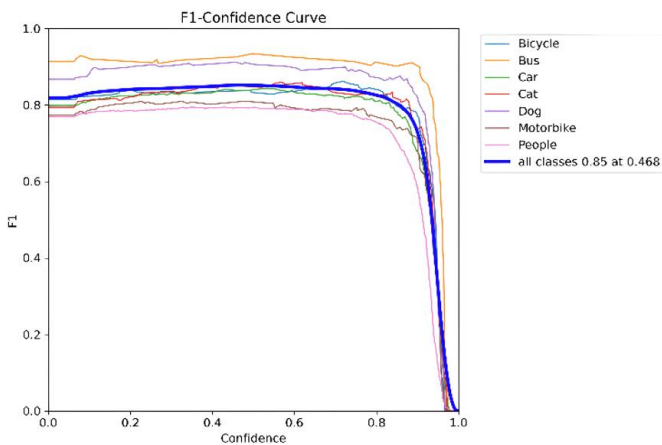


Figure 5. F1-confidence curve of the ExDark dataset using YOLOv8m

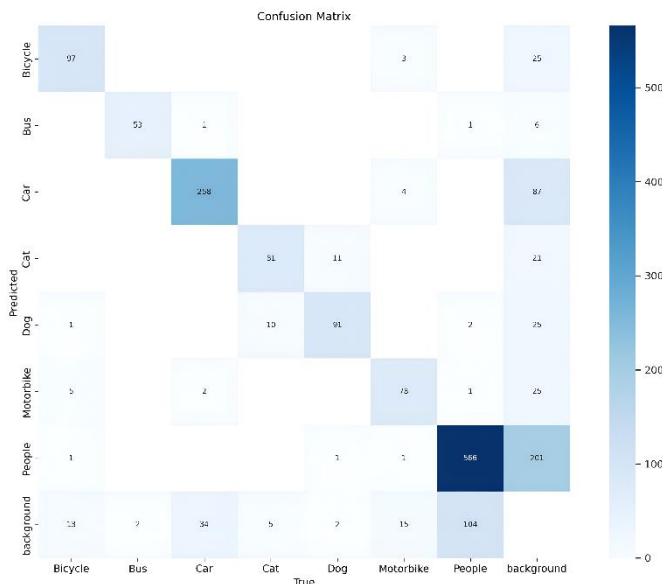


Figure 6. Confusion matrix of the ExDark dataset using YOLOv8m

Figure 6 illustrates the confusion matrix, showcasing how well the model classifies each category individually. The true classes are listed along the vertical axis, and the predicted classes are listed on the horizontal axis. The diagonal elements in the confusion matrix are essential as they are correct

classifications. Off-diagonal elements indicate misclassifications where the model's prediction did not match the actual class. The color intensity in the matrix provides a visual representation of the frequency of predictions, with darker shades indicating higher counts. The 'Car' and 'Bus' classes show high accuracy, with minimal misclassifications. These objects are often prominent in nighttime surveillance. The 'Cat' and 'Dog' show occasional misclassifications, as these objects may blend into dark environments, making them harder to differentiate from one another. Analyzing the confusion matrix, it can be identified which specific class needs more training, and the model can be adjusted to reduce the misclassifications, improving the model's overall accuracy and reliability. The model summary parameters and the training time taken for each model are listed in Table 2.

Table 2. Model summary parameters

Model	Layers	Parameters	GFLOPs	Time
YOLOv8n	168	3,006,428	8.1	1.483
YOLOv8s	192	9,385,763	39.4	1.628
YOLOv8m	218	25,842,076	78.7	1.802
YOLOv8l	268	43,609,692	164.8	2.424
YOLOv8x	297	77,436,893	192.7	2.864

Table 3. Training model results

Method	Precision	Recall	mAP@50	F1-Score
YOLOv8n	0.855	0.728	0.816	0.77
YOLOv8s	0.854	0.767	0.827	0.79
YOLOv8m	0.908	0.819	0.886	0.85
YOLOv8l	0.897	0.790	0.865	0.84
YOLOv8x	0.877	0.777	0.838	0.81

In this study, extensive experiments are conducted on all the models of YOLOv8 on the ExDark dataset for object detection. The detailed comparison and evaluation of all the models of YOLOv8 showcase the detection performance of the YOLOv8 models on the ExDark dataset in nighttime and low-light conditions in Table 3. The results indicate that the YOLOv8m model performs well in all the metrics compared to the larger models. YOLOv8m demonstrated the best overall performance among all the tested models, achieving a precision of 0.908, recall of 0.819, mAP@50 of 0.886, and F1-score of 0.85. Figure 7 illustrates the precision-recall and mAP metrics. It also illustrates the training and validation losses of box loss, class loss, and distribution focal loss (DFL loss). The box loss measures the predicted bounding box's accuracy relative to the ground-truth box. A lower value indicates that the model is correctly learning the locations of objects. Throughout the training, the box loss shows a consistent decline, suggesting that the model is gradually improving in its ability to localize objects accurately in low-light environments. The class loss reflects how well the model is classifying the detected objects. Like the box loss, a steady decline in class loss during training indicates improved performance in distinguishing between different object classes. Distribution Focal Loss (DFL) measures the confidence with which the model assigns objects to predicted classes. The decrease in the DFL value indicates that the YOLOv8m model is performing well and becoming more confident in predicting the objects and their classes correctly, which indicates that the false positives and false negatives are occurring at a minimum. Figure 7 also showcases the validation performance. It shows a steady decrease in

validation loss, indicating that the model is not overfitting and generalizes effectively to unseen data. This is crucial in surveillance applications, where the model must detect objects in real time under various low-light conditions without degrading accuracy. These metrics indicate that YOLOv8m accurately identifies objects in highly dark conditions and

maintains a balanced performance across various evaluation metrics. The metrics achieved with the extremely low-light dataset are promising and are comparably more than the previous studies. The results indicate that our proposed method performs reasonably better than all the previous studies in this area.

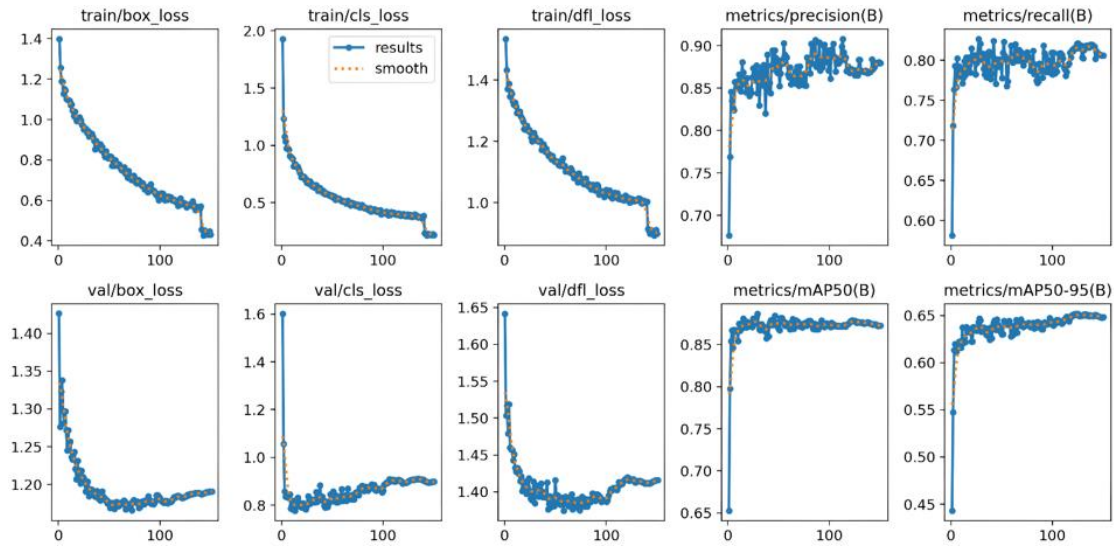


Figure 7. The training and validation losses with metrics

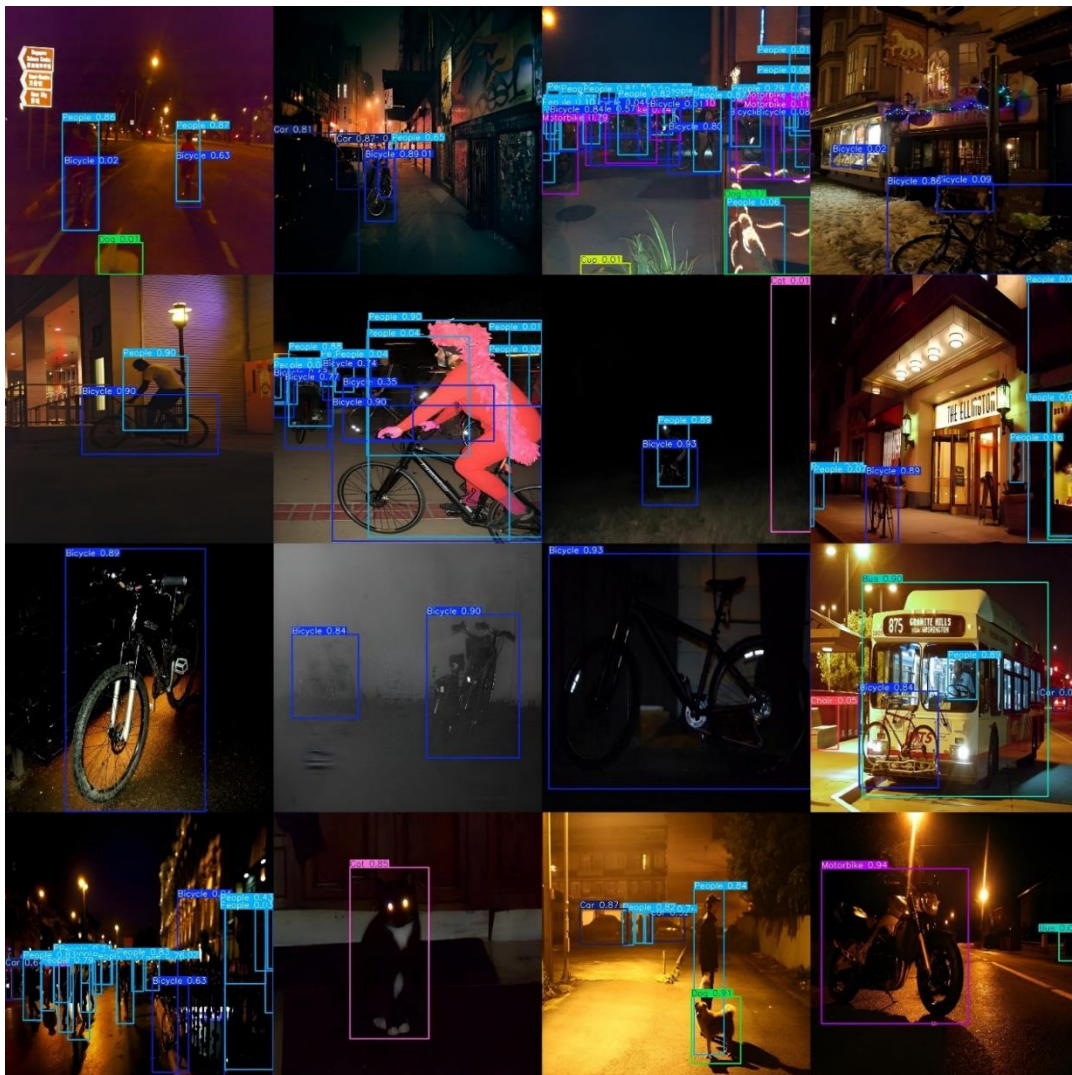


Figure 8. The testing batch results of the ExDark dataset using YOLOv8m

Table 4. Comparison with previous research

Ref.	Method	Dataset	mAP@50
[16]	Faster RCNN	ExDark	0.783
[16]	Mask RCNN	ExDark	0.802
[16]	SSD	ExDark	0.693
[2]	YOLOv5s	ExDark	0.713
[16]	YOLOv7	ExDark	0.798
[2]	YOLOv8s	ExDark	0.683
Proposed Model	YOLOv8m	ExDark	0.886

Table 4 compares metrics from previous research, highlighting the superior performance of our model across all evaluated metrics. This indicates that our model is highly suitable for real-time object detection in surveillance applications, particularly in low-light conditions. Figure 8 showcases the test and prediction results of the trained model.

Table 4 presents a comparison between the proposed YOLOv8m model and various state-of-the-art models, including Faster R-CNN, Mask R-CNN, SSD, YOLOv5s, YOLOv7 and also YOLOv8s based on their mean Average Precision (mAP) on the same ExDark dataset. The proposed YOLOv8m model achieves a significantly higher mAP of 0.886 than the previous studies. Faster R-CNN and Mask R-CNN attain mAP values of 0.783 and 0.802, respectively, demonstrating that while these models perform well, they fall short in precision compared to YOLOv8m. Similarly, YOLOv7, another well-known model, achieves a competitive mAP of 0.798, yet it still needs to be improved compared to the proposed model. SSD lags with an mAP of 0.693, reflecting its limitations in handling complex low-light environments. These results highlight the advancements made by the YOLOv8m model, particularly in the context of nighttime surveillance, where the optimization techniques used in this study, such as data augmentation and anchor box refinement, contribute to a substantial improvement in detection accuracy. The results showcase that our proposed YOLOv8m model is highly suitable for real-time object detection in surveillance applications, particularly in low-light conditions. Figure 8 showcases the test and prediction results of the trained YOLOv8m model on the ExDark dataset. It thus demonstrates the model's ability to detect small and distant objects in low-light environments accurately. The model successfully detects small objects like a cat with an 86% confidence score and distant objects like bicycles with 84%, cars, and people with 87% and 83% confidence levels in extremely low light conditions. These results indicate that the YOLOv8m model performs exceptionally well in challenging low-light conditions, consistently identifying small and distant objects with high precision. The relatively high confidence levels for various object classes reflect the model's robustness and reduced instances of misclassification, further validating its effectiveness in real-time nighttime surveillance tasks.

5.4 Limitations of the study

Hyperparameter sensitivity: The performance of the YOLOv8m model is sensitive to hyperparameter tuning, particularly concerning learning rate and anchor box adjustments. Finding the optimal configuration required extensive experimentation and computing resources. Even slight variations in hyperparameter tuning led to model performance fluctuations.

Model overfitting to specific augmentation techniques: Advanced data augmentation techniques improved the model's

performance in low lighting conditions, but it may have inadvertently caused overfitting due to over-reliance on advanced augmentation techniques. Over-reliance on augmentation techniques may reduce the model's adaptability when exposed to novel lighting conditions not represented in the training set.

Limited dataset diversity: The ExDark dataset images lack diversity regarding environmental conditions, such as varying levels of artificial lighting, and weather conditions, such as fog, rain, and background clutter. These factors can affect the generalizability of the YOLOv8 model when applied to real-world surveillance systems in diverse environments.

Computational demands of HPC: The study relies heavily on HPC resources for training and testing YOLOv8, which may not be accessible to all organizations, especially for real-time, low-cost surveillance systems. The reliance on expensive hardware and resources limits the practicality of the approach for broader applications.

Impact of object motion in low-light conditions: The current study focuses on detecting stationary or slowly moving objects in low-light environments. However, rapidly moving objects, such as vehicles or people in motion, may introduce motion blur, which can significantly degrade detection accuracy. YOLOv8 may have limitations in capturing objects with high-speed movement under such conditions.

6. CONCLUSIONS

In this paper, the study presented an effective methodology for object detection in nighttime surveillance applications on the ExDark dataset based on the YOLOv8 model using high-performance computing (HPC). The study trained and evaluated the performance of various YOLOv8 models nano(n), small(s), medium(m), large(l), and extra-large(x) on the Exdark dataset, and the YOLOv8m model emerged as the best model achieving balanced performance. The YOLOv8m model achieved a precision value of 0.908, a recall of 0.819, and a mAP@50 of 0.886, proving its suitability for real-time object detection in surveillance applications for low light conditions. These results showcase the ability of the YOLOv8m model to address the challenges of low visibility small and distant object detection and make it an ideal choice for low-light surveillance applications. The study focused on the seven most relevant object classes for surveillance: bicycle, bus, car, cat, dog, motorbike, and people from the ExDark dataset, and the model was optimized for object detection in urban surveillance applications by concentrating on these essential objects. High-performance computing (HPC) significantly enabled effective model training by handling intensive computational tasks. It helped to reduce the computational training time by parallelly processing all complex operations, allowing the model to be trained more quickly than conventional hardware.

The YOLOv8m model's performance was enhanced by implementing many argumentation techniques like brightness adjustments and Gaussian blur to improve the model's robustness in various low-light conditions. The performance of the YOLOv8m model was further enhanced by using hyperparameter tuning with the Optuna framework and the usage of the AdamW optimizer. Implementing K-Means clustering for anchor box optimization significantly enhanced the model's detection capabilities in complex, low-light scenarios. Implementing all these techniques helped improve

the YOLOv8m model's performance and perform exceptionally well in object detection tasks, such as detecting small and distant objects in extremely low light conditions. Future work can be focused on expanding the dataset or incorporating more diverse datasets further to test the robustness of the model under varying conditions and optimizing this model on edge computing or deploying lighter versions of YOLOv8 for real-time applications on lower-end systems for cost-sensitive and resource-limited environments. The performance achieved by this model YOLOv8m on the ExDark dataset signifies the model's potential for enhancing security systems and paves the way for further advancements in this critical field.

REFERENCES

- [1] Damera, V.K., Vatambeti, R., Mekala, M.S., Pani, A.K., Manjunath, C. (2023). Normalized attention neural network with adaptive feature recalibration for detecting the unusual activities using video surveillance camera. *International Journal of Safety & Security Engineering*, 13(1): 51-58. <https://doi.org/10.18280/ijss.130106>
- [2] Wang, J., Yang, P., Liu, Y., Shang, D., Hui, X., Song, J., Chen, X. (2023). Research on improved YOLOv5 for low-light environment object detection. *Electronics*, 12(14): 3089. <https://doi.org/10.3390/electronics12143089>
- [3] Moch, A.S., Supriyanto, C. (2023). Deep learning model for unmanned aerial vehicle-based object detection on thermal images. *Revue d'Intelligence Artificielle*, 37(6): 1441-1447. <https://doi.org/10.18280/ria.370608>
- [4] Liu, H., Jin, F., Zeng, H., Pu, H., Fan, B. (2023). Image enhancement guided object detection in visually degraded scenes. *IEEE Transactions on Neural Networks and Learning Systems*, 35(10): 14164-14177. <https://doi.org/10.1109/TNNLS.2023.3274926>
- [5] Luo, X., Xiang, S., Wang, Y., Liu, Q., Yang, Y., Wu, K. (2021). Dedark+ detection: A hybrid scheme for object detection under low-light surveillance. In *Proceedings of the 3rd ACM International Conference on Multimedia in Asia, Gold Coast, Australia*, p. 71. <https://doi.org/10.1145/3469877.3497691>
- [6] 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. <https://doi.org/10.1016/j.cviu.2018.10.010>
- [7] Sugiharto, A., Kusumaningrum, R. (2023). Automated detection of driver and passenger without seat belt using YOLOv8. *International Journal of Advanced Computer Science & Applications*, 14(11): 806-813. <https://doi.org/10.14569/IJACSA.2023.0141181>
- [8] Jha, S., Seo, C., Yang, E., Joshi, G.P. (2021). Real time object detection and tracking system for video surveillance system. *Multimedia Tools and Applications*, 80(3): 3981-3996. <https://doi.org/10.1007/s11042-020-09749-x>
- [9] Luo, P., Nie, J., Xie, J., Cao, J., Zhang, X. (2024). Localization-aware logit mimicking for object detection in adverse weather conditions. *Image and Vision Computing*, 146: 105035. <https://doi.org/10.1016/j.imavis.2024.105035>
- [10] Kou, K., Yin, X., Gao, X., Nie, F., Liu, J., Zhang, G. (2024). Lightweight two-stage transformer for low-light image enhancement and object detection. *Digital Signal Processing*, 150: 104521. <https://doi.org/10.1016/j.dsp.2024.104521>
- [11] Almujally, N.A., Qureshi, A.M., Alazeb, A., Rahman, H., et al. (2024). A novel framework for vehicle detection and tracking in night ware surveillance systems. *IEEE Access*, 12: 88075-88085. <https://doi.org/10.1109/ACCESS.2024.3417267>
- [12] Abdusalomov, A., Baratov, N., Kutlimuratov, A., Whangbo, T.K. (2021). An improvement of the fire detection and classification method using YOLOv3 for surveillance systems. *Sensors*, 21(19): 6519. <https://doi.org/10.3390/s21196519>
- [13] Sharma, T., Debaque, B., Duclos, N., Chehri, A., Kinder, B., Fortier, P. (2022). Deep learning-based object detection and scene perception under bad weather conditions. *Electronics*, 11(4): 563. <https://doi.org/10.3390/electronics11040563>
- [14] Xu, Z., Li, J., Zhang, M. (2021). A surveillance video real-time analysis system based on edge-cloud and fly-yolo cooperation in coal mine. *IEEE Access*, 9: 68482-68497. <https://doi.org/10.1109/ACCESS.2021.3077499>
- [15] Song, R., Li, T., Li, T. (2023). Ship detection in haze and low-light remote sensing images via colour balance and DCNN. *Applied Ocean Research*, 139: 103702. <https://doi.org/10.1016/j.apor.2023.103702>
- [16] Hui, Y., Wang, J., Li, B. (2024). WSA-YOLO: Weak-supervised and adaptive object detection in the low-light environment for YOLOv7. *IEEE Transactions on Instrumentation and Measurement*, 73: 2507012. <https://doi.org/10.1109/TIM.2024.3350120>
- [17] Xiao, Y., Liao, H. (2024). LIDA-YOLO: An unsupervised low-illumination object detection based on domain adaptation. *IET Image Processing*, 18(5): 1178-1188. <https://doi.org/10.1049/ipr2.13017>
- [18] Peng, D., Ding, W., Zhen, T. (2024). A novel low light object detection method based on the YOLOv5 fusion feature enhancement. *Scientific Reports*, 14(1): 4486. <https://doi.org/10.1038/s41598-024-54428-8>
- [19] Chen, X., Yuan, M., Yang, Q., Yao, H., Wang, H. (2023). Underwater-YCC: Underwater target detection optimization algorithm based on YOLOv7. *Journal of Marine Science and Engineering*, 11(5): 995. <https://doi.org/10.3390/jmse11050995>
- [20] Bose, S., Kolekar, M.H., Nawale, S., Khut, D. (2023). Loltv: A low light two-wheeler violation dataset with anomaly detection technique. *IEEE Access*, 11: 124951-124961. <https://doi.org/10.1109/ACCESS.2023.3329737>
- [21] Nie, C., Qadar, M.A., Zhou, S., Zhang, H., Shi, Y., Gao, J., Sun, Z. (2023). Transnational image object detection datasets from nighttime driving. *Signal, Image and Video Processing*, 17(4): 1123-1131. <https://doi.org/10.1007/s11760-022-02319-8>
- [22] Rahim, A., Maqbool, A., Rana, T. (2021). Monitoring social distancing under various low light conditions with deep learning and a single motionless time of flight camera. *Plos One*, 16(2): e0247440. <https://doi.org/10.1371/journal.pone.0247440>
- [23] Wang, C., Dai, Y., Zhou, W., Geng, Y. (2020). A vision-based video crash detection framework for mixed traffic flow environment considering low-visibility condition. *Journal of Advanced Transportation*, 2020(1): 9194028. <https://doi.org/10.1155/2020/9194028>

- [24] Zhao, Y., Wu, J., Chen, W., Wang, Z., Tian, Z., Yu, F.R., Leung, V.C. (2024). A small object real-time detection method for power line inspection in low-illuminance environments. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(6): 3936-3950. <https://doi.org/10.1109/TETCI.2024.3378651>
- [25] Rossi, F., Saponara, S. (2024). Edge HPC architectures for AI-based video surveillance applications. *Electronics*, 13(9): 1757. <https://doi.org/10.3390/electronics13091757>
- [26] Gao, X., Zhang, Y. (2023). Detection of fruit using YOLOv8-based single stage detectors. *International Journal of Advanced Computer Science & Applications*, 14(12): 83-91. <https://doi.org/10.14569/IJACSA.2023.0141208>
- [27] Zhang, D. (2024). A Yolo-based approach for fire and smoke detection in IoT surveillance systems. *International Journal of Advanced Computer Science & Applications*, 15(1): 87-94. <https://doi.org/10.14569/IJACSA.2024.0150109>
- [28] Bakirci, M. (2024). Real-time vehicle detection using YOLOv8-nano for intelligent transportation systems. *Traitement du Signal*, 41(4): 1727-1740. <https://doi.org/10.18280/ts.410407>
- [29] Hidayatullah, P. (2024). YOLOv8 architecture detailed explanation—A complete breakdown. <https://www.youtube.com/watch?v=HQXhDO7COj8>.