

Citation: Dubey Prati, Mittan Rakesh. Enhanced real-time suspicious object detection with YOLO in diverse environmental conditions. *Journal of Harbin Institute of Technology (New Series)*, DOI:10.11916/j.issn.1005-9113.2025019

Enhanced real-time Suspicious Object Detection with YOLO in Diverse Environmental Conditions

Dubey Prati* and Mittan Rakesh

(Department of Computer Science & Engineering, Rabindranath Tagore University, Bhopal 462042, India)

Abstract: Various imaging techniques have been employed to detect suspicious objects and activities to help preserve peace and order. However, challenges such as illumination changes, occlusion, noise, and low-resolution imagery significantly hinder the effectiveness of automated detection methods. To address these issues, machine learning techniques have been applied, but they often struggle with the multiple activities detection simultaneously, leading to ambiguity and reduced accuracy. To mitigate these issues, the proposed methodology presented the YOLOv8 model for detection of suspicious objects and activities in images and video frames. To remove the environmental noise, adaptive sliding window based bilateral filter is used to remove local and global noise in input noisy images, then YOLOv8 model is trained to identify the suspicious and non-suspicious objects and activities. The performance was evaluated on suspicious object and activity dataset collected from publicly available resources such as Roboflow. Performance was measured using mean average precision (mAP) and compared to existing state-of-the-art models. The proposed model achieved an average mAP of 74.5%, which represents approximately a 13% improvement over current leading methods. Therefore, the study shows the efficacy of the proposed model to enhance the surveillance system to handle environmental complexities.

Keywords: suspicious object, suspicious activities, deep learning, YOLOv8, environmental complexities

CLC number: TP391

Document code: A

Article ID: 1005-9113(2025)00-0000-12

0 Introduction

The field of detection of suspicious human behavior using imaging or video surveillance is rapidly advancing with image analysis and computer vision technologies. The main objective is to distinguish between normal and abnormal activities of people in public places. Activities such as walking or waving hands are considered to be normal activities, whereas theft or potential attacks are considered to be abnormal activities^[1]. Therefore, this increases the demand for image/video monitoring in public domains such as financial institutions, government properties, transportation centers, etc. The conventional monitoring approaches often involve constant human monitoring that is quite expensive and also less efficient. With the growing need for smart and self-reliant monitoring systems that are capable of

autonomously identifying unusual behaviors^[2-4], researchers are increasingly focusing on developing advanced artificial intelligence-based frameworks to enhance reliability, adaptability, and real-time decision-making. These smart approaches could identify abandoned packages, violent activities, etc. With the increasing risk of attacks in public places, real-time smart systems that can quickly identify such activities/objects and promptly notify the security staff are needed^[5-8]. To achieve this, conventional systems follow a series of operations such as identifying foreground objects, detecting specific objects, extracting features, classifying objects, and analyzing them. Number of machine learning techniques are explored by researchers, such as support vector machines, Bayesian methods, K-Nearest neighbours, etc. These learning models are utilized for classifying objects and recognizing activities^[9]. However, there are still obstacles to

Received 2025-02-07.

* Corresponding author. Dubey Prati, Ph.D. Email: pratidubey.245@gmail.com.

overcome, such as varying lighting conditions, object overlaps, background noise, and low-quality resolution. These need to be overcome. Additionally, existing machine learning models have limitations in accurately identifying multiple activities concurrently, which affects their overall effectiveness. Hence, while smart monitoring systems present a more proficient way to keep public and sensitive spaces secure, there are still areas that require further innovation.

Therefore, motivated by these limitations, the paper contributed the following:

- This study presented a collective suspicious object/activity database for recognition of human suspicious activity in real-time videos.
- Then we designed a real-time suspicious activity recognition model for recognition of multiple activities in real-time images/video.
- The study presented a hybrid model that can increase the probability of detection of suspicious objects under low resolution and noisy conditions.

The proposed method effectively improves detection accuracy and addresses environmental complexity by combining an adaptive sliding window-based bilateral filter with the YOLOv8 object detection model. The adaptive filter enhances image quality by locally adjusting its parameters to remove various types of noise while preserving critical edge and texture information. This preprocessing step ensures that the input to YOLOv8 is cleaner and more feature-rich, allowing the model to focus on meaningful objects rather than being distracted by noise or distortions. As a result, YOLOv8 can more accurately detect suspicious objects and activities, even in challenging conditions, which leads to higher recognition accuracy and greater robustness in real-world surveillance scenarios.

The rest of the paper is organized as follows: Section 1 illustrates the existing approaches and contributions. Section 2 presents the overview of YOLOv8 architecture. Section 3 presents the proposed methodology with algorithms. Section 4 presents the result analysis. Section 5 presents the conclusion of the implemented model and its future scope.

1 Literature Review

Gawande et al.^[1] developed a surveillance system for monitoring student activities such as cheating, theft, and physical confrontations within educational

institutions. YOLOv5 model was used to predict suspicious activities. Tutar et al.^[2] presented a hybrid model of k-NN and SVM for video anomaly detection. In this model level-wise detection was performed. In the first level, pixel-based detection was performed, and in the second level, frame-based detection methods were used. The model was trained and tested on UCF-Crime dataset. Vidya and Selvakumar^[3] presented the human suspicious activity detection model that employed a multi-step process. The authors used the attention based residual network with optimization algorithm for hyper-parameter tuning. Alruwais and Zakariah^[4] used suspicious activity detection for students in online classroom with Convolution Neural Network (CNN). This model was also efficient to identify the mood of the student. Wang et al.^[5] proposed a suspicious activity detection model termed as Quantized Object Recognition Model (QORM). Talib et al.^[6] used the YOLOv8 model for small object detection for suspicious objects and activity detection. The model focused on spatial features of the images to effectively detect the object. Sun et al.^[7] applied the spatio-temporal features from images to detect flying birds in surveillance videos. Gautam et al.^[8] focused on detecting multiple suspicious abandoned objects in public spaces using YOLO model. Pullakandam et al.^[9] used the YOLOv8 model to detect weapons from input images. Jebur et al.^[10] proposed a fusion model using transfer learning such as Xception, Inception, and InceptionResNet. The model was used to detect violence from images and achieved an accuracy of 97%. Huszár et al.^[11] used spatio-temporal features of input images to detect violence using a 3-D CNN model. Jain et al.^[12] proposed a CNN-LSTM model to detect violence in images and classified the violence as weaponized, non-weaponized, and normal, and achieved an accuracy of 98%. Faisal et al.^[13] proposed a driver-assistance system that used YOLOv3 for object detection to enhance road safety. By detecting nearby vehicles and pedestrians and estimating their distance from the height of the objects, the system creates an alert system to prevent collisions. This low-cost solution works effectively even in low-light conditions. Wang et al.^[14] introduced AI-TOD for tiny object detection in aerial images. Mahmmoud et al.^[15] presented a fast overlapped block-processing technique that employed higher-order polynomials with SVM for 3D object recognition.

2 Overview of YOLOv8

In 2016, researchers introduced the YOLO (You Only Look Once) algorithm, a CNN-based method for real-time object detection in images and videos. YOLO divides the input image into a grid, where each cell predicts the objects' locations and categories within its area, leveraging "anchor boxes" for precise detection. The model, trained on extensive annotated datasets, features 24 convolutional layers and 2 fully connected layers, with some 1×1 reduction layers to manage complexity. Its output tensor is reshaped for bounding box predictions, using Intersection over Union (IoU) to select the best bounding box and sum-squared error for loss calculation^[16-19].

Versions of YOLO are:

- YOLOv1: Revolutionized object detection with a single convolutional network, achieving 63.4% mAP on PASCAL VOC (Visual Object Classes) 2012 but struggled with small objects and novel shapes.
- YOLOv2: Enhanced with batch normalization and anchor boxes, achieving 78.6% mAP on COCO.
- YOLOv3: Introduced FPN and a prediction module for small objects, with 57.9% mAP on COCO, offering better speed and accuracy.
- YOLOv4: Added PAN, Mish activation, and SPP, achieving 43.5% mAP on COCO, noted for its speed and accuracy.
- YOLOv5: Released in 2020, focused on

efficiency and speed using EfficientNet and a new loss function, with 50.0% mAP on COCO.

- YOLOv6: Enhanced for industrial applications with an efficient decoupled head, improved training strategy, and a hardware-friendly architecture.

- YOLOv7: Faster and more accurate with the Extended Efficient Layer Aggregation Network (E-ELAN) and a scalable design, integrating features from previous models.

- YOLOv8: The latest version, offering pre-trained models for detection, segmentation, and classification tasks, represents the most advanced iteration of the YOLO series.

The basic architecture of the YOLO model is presented in Fig.1. The two key features of the YOLO model are its pyramidal network and anchor free operation. The pyramidal network is composed of three modules: backbone module, head module, and detect module. The model used the Cross Stage Partial (CSP) for feature extraction with combination of convolution and concatenation layers. The anchor-free detection module is used to predict the distance of the object's center from the boundaries of bounding box. This operation does not need the pre-determined anchor boxes. This anchor-free operation will detect the object dynamically with better efficiency. In this paper, YOLOv8 version is used. Compared with YOLOv5, YOLOv8 has fewer blocks at each stage, reducing computational complexity and potentially improving gradient information extraction.

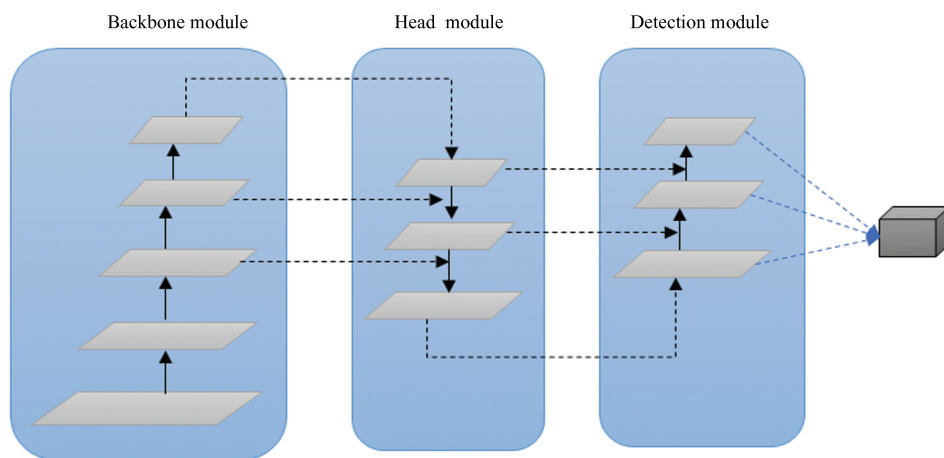


Fig.1 Architecture of YOLO learning model^[19]

Moreover, to improve learning speed, Spatial Pyramid Pooling Fast (SPPF) layer is used. In YOLOv8, the learning complexity is reduced due to

its pyramidal network with aggregation network. The information of the lower layer is preserved as it is passed on to the upper layer in the pyramidal network.

The detection module evaluates the distances between the centres of the objects identified in the bounding box. This module is anchor-free operation. The weighted score for classification of label is equated as:

$$t = s^\alpha \times u^\beta \quad (1)$$

where, the predicted score is represented as s , and the predicted IoU as u .

The loss is equated as^[19]:

$$\text{Loss}_{\text{fun}} = \text{Loss}_n + \text{Loss}_{\text{CloU}} \quad (2)$$

$$\text{Loss}_n = -w [y_n \log x_n + (1 - y_n) \log(1 - x_n)] \quad (3)$$

$$\text{Loss}_{\text{CloU}} = 1 - \text{IoU} + \frac{\text{Distance}_2^2}{\text{Distance}_c^2} + \frac{v^2}{(1 - \text{IoU}) + v} \quad (4)$$

where, Loss_n and $\text{Loss}_{\text{CloU}}$ represent the classification loss and IoU loss of predicted value x_n . The actual value is termed as y_n , with aspect ratio of v equated as:

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w^p}{h^p} \right)^2 \quad (5)$$

where, weight of the bounding box is represented as w , and height as h .

3 Proposed Methodology

In this study, the entire methodology is divided into three steps according to the presented research objectives. The proposed methodology flowchart is presented in Fig.2. First, a suspicious object or activity was detected in both images and videos. In this step, suspicious object/activity is detected using YOLOv8 model. If the result achieved is satisfactory or not. The result is evaluated based on whether it can identify objects in real images under environmental complexities. This will be measured in terms of performance parameters. If the result is not satisfactory, the methodology will identify the limitations of the current approach. Then it was found that background complexities are the most contributing factor that degrades performance. Then a digital filter with adaptive sliding bilateral filter was designed to detect suspicious object/activity under environmental complexities. In this step, the object/activities are detected with different environmental complexities such as noise, blur, camera motions, etc.

3.1 Suspicious Object/Activity Detection in Images and Videos

Fig.3 presents the flowchart for object detection

using the YOLOv8 model in images. The YOLOv8 model is applied to the input image to detect objects. The Regions of Interest (RoIs) within the image are identified based on the objects detected by YOLOv8. Features of the objects identified within the RoIs are extracted. This step involves analyzing the detected objects to obtain relevant features for further processing. The YOLOv8 model is trained using the input object/activity dataset. This training process helps the model learn to accurately detect objects in images. The identified objects are further analyzed to determine if any of them are suspicious or require special attention.

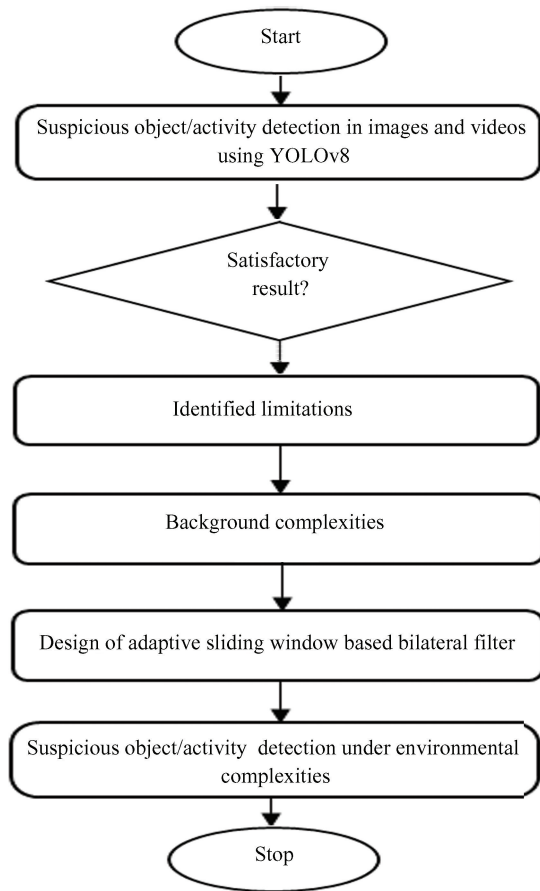


Fig.2 Proposed methodology

In Fig.4, machine learning will be used to design a suspicious object detection in real-time video and determination of suspicious activity. In this model, a suspicious object/activity dataset is prepared, and machine is trained accordingly. Then based on these trained rules, suspicion probability of frames in video will be determined using image processing tool. Detailed steps of the methodology are presented in Fig.4.

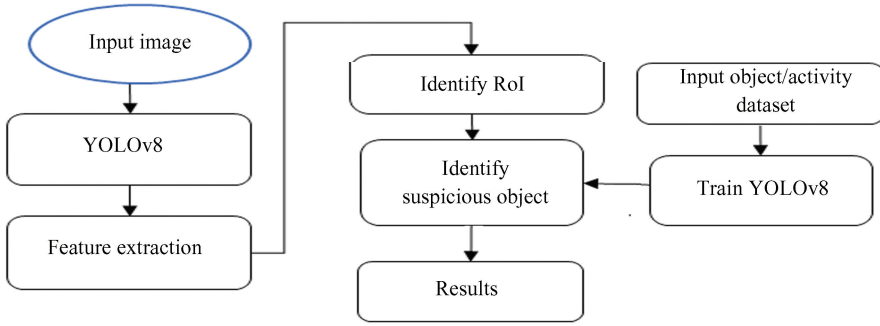


Fig.3 Working model for object detection in images

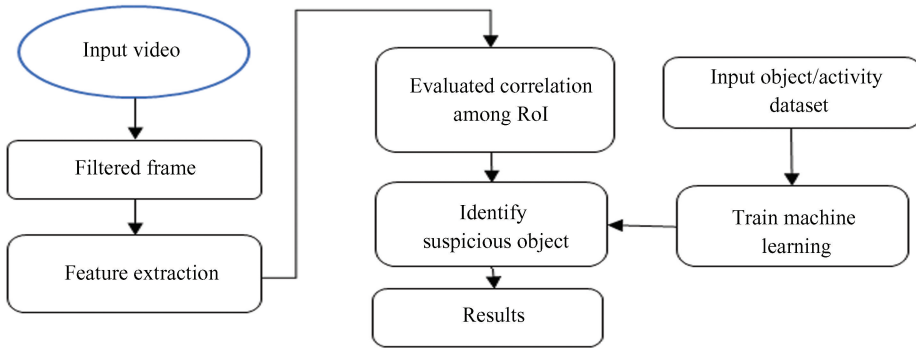


Fig.4 Working model for object detection in videos

3.2 Design of Adaptive Sliding Window based Bilateral Filter for Suspicious Object Detection

The presented adaptive filter is a modification of bilateral filter with spatial-adaptation for noise removal. This will preserve the edge and texture characteristics of the input images. In a conventional bilateral filter, it combines the domain and range kernels to preserve edge and texture information. Mathematically, it is expressed as:

$$I_{\text{filt}}(n) = \frac{1}{N_f} \sum_{m \in P} I(m) \cdot f(\|n - m\|) \cdot g(|I(n) - I(m)|) \quad (6)$$

where, filtered image is considered as I_{filt} for n pixels. The neighboring pixel's (m) intensity is represented as $I(m)$ within spatial domain P with n pixels. Spatial kernel is represented as $f(\|n - m\|)$, which reduces the kernel distance between m and n . The range kernel is represented as $g(|I(n) - I(m)|)$. But in the adaptive bilateral filter, the step uses the sliding window approach to identify local adaptation features, and thus combining local spatial features to generate global feature for noise removal. Mathematically, in sliding window, s_p might be described as:

$$I_{\text{local}}(n) = \frac{1}{N_{\text{flocal } m \in P_{\text{local}}}} \sum I(m) \cdot f_{\text{local}}(\|n - m\|) \cdot g_{\text{local}}(|I(n) - I(m)|) \quad (7)$$

where, filtered image in each sliding window output is considered as $I_{\text{local}}(n)$ for n pixels. The neighboring pixel's (m) intensity is represented as $I(m)$ within local spatial domain P_{local} with n pixels. Local spatial kernel is represented as $f_{\text{local}}(\|n - m\|)$ that reduces the kernel distance between m and n . The local range kernel is represented as $g_{\text{local}}(|I(n) - I(m)|)$. By combining all these local filtration parameters, global parameters are identified to filter out the image.

Fig.5 presents the design of suspicious object/activity detection under environmental complexities. For this, an adaptive sliding window based bilateral filter is presented here. The entire network architecture is presented in Fig. 6. Algorithm is presented below.

Algorithm

Input: Inp_i , images;

Training data, $\text{Tr}_n = \{\text{Inp}_i^n, \text{Leb}_i^n\}$, where $n \in \text{size}(\text{Tr}_n)$

Output: PR_i^n , suspicious region.

1. Initialization
2. $I_n = \text{Inp}_i + \text{Noise}$

4 Results and Discussion

This section presents the training details of the proposed model for suspicious object and activity detection from images. The entire model is simulated on the Python platform over Google Colab with facility of Tesla P100-PCIE GPU. For training the model, hyperparameters were optimized through iterative experimentation. The initial learning rate was set to 0.001 using the Adam optimizer with cosine annealing to gradually decrease the rate over time. Batch size was set to 16 for 100 epochs. After the training, the testing performance of the proposed model is evaluated. To compare the performance, following parameters are evaluated:

Mean average Precision (mAP): mAP is mathematically represented as:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (8)$$

$$\text{Precision} = \frac{(\text{True_Positive})}{(\text{True_Positive} + \text{False_Positive})} \quad (9)$$

$$\text{Recall} = \frac{(\text{True_Positive})}{(\text{True_Positive} + \text{False_Negative})} \quad (10)$$

where AP measures the area under the Precision-Recall (PR) curve and summarizes how well a model

balances precision and recall across different confidence thresholds.

Table 1 shows the detailed dataset description used for the proposed model. The dataset was collected from publicly available resources such as Roboflow. The dataset consists of a number of samples of suspicious object and activity images with their annotations. The sample descriptions are presented in Table 1.

Table 1 Dataset description

Dataset	Classes	Training samples	Testing samples
DS-1 ^[20]	normal, suspicious activity, victim, weapon	1800	172
DS-2 ^[21]	handgun, knife, long-guns, missile, sword, tank	7780	505

In this paper, the implementation of the proposed model is done using Python3 in Google Colab. For model implementation, TensorFlow is used. The entire dataset is divided into 70% training dataset and 30% testing dataset. The learning was done using Adam optimizer. The entire implementation is done using GPU services for 100 epochs. Figs. 7 and 8 present the training graphs of the proposed model on DS-1 and on DS-2, respectively.

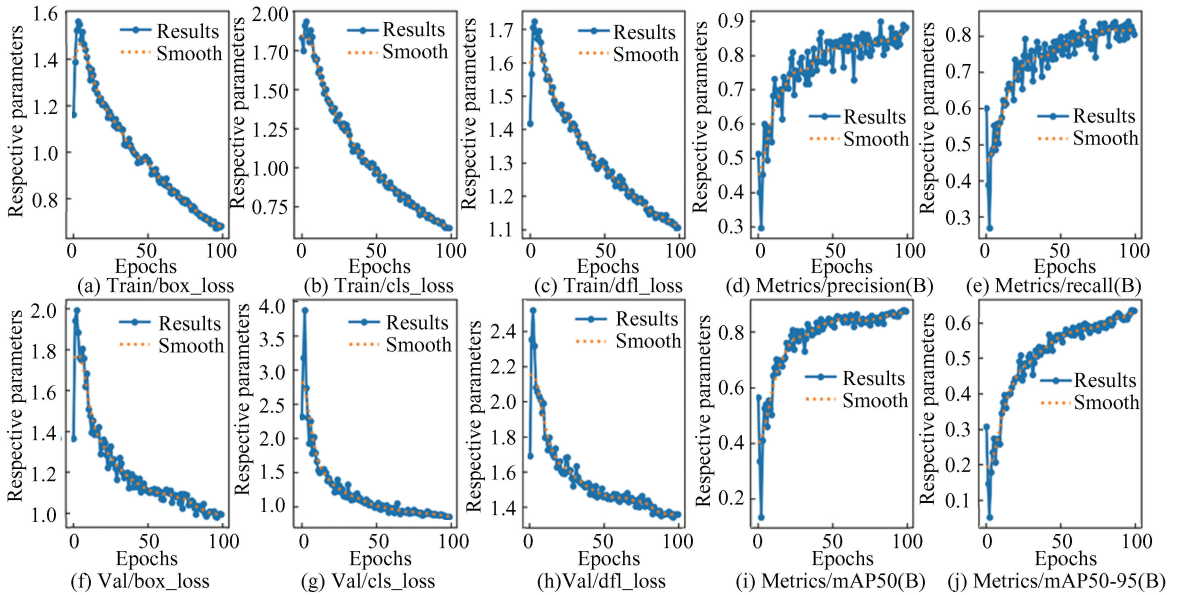


Fig. 7 Training performance on DS-1

4.1 Ablation Study

Fig.9 presents the precision graphs under environmental complexities with and without applying

filters. The graph is presented for 640×640 resolution. The comparison shows that the precision for all classes is approximately 83%, which is an

improvement of approximately 9% compared with the case without the filter condition. Fig.10 presents the recall graphs under environmental complexities with and without applying filters. The graph is presented for 640×640 resolution. The comparison shows that the recall for all classes is approximately 68% , which is an improvement of approximately 1% compared with the case without the filter condition. Fig.11

presents the mAP graphs under environmental complexities with and without applying filters. The graph is presented for 640×640 resolution. The comparison shows that the mAP for all classes is approximately 74.5% , which is an improvement of approximately 4% compared with the case without the filter condition.

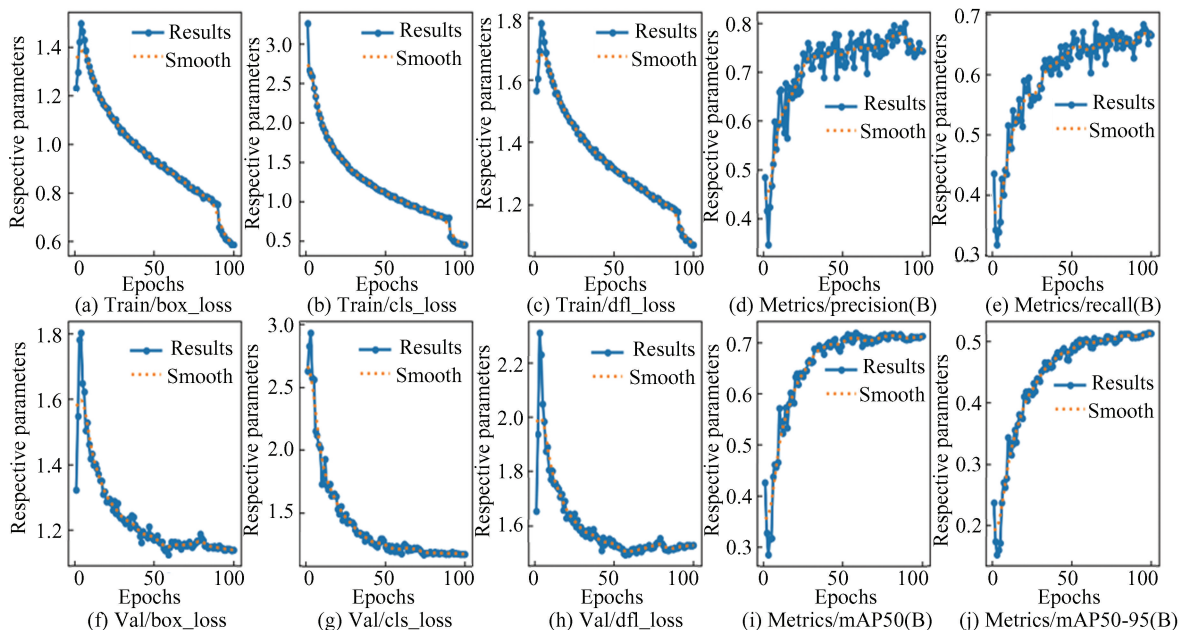


Fig.8 Training performance on DS-2

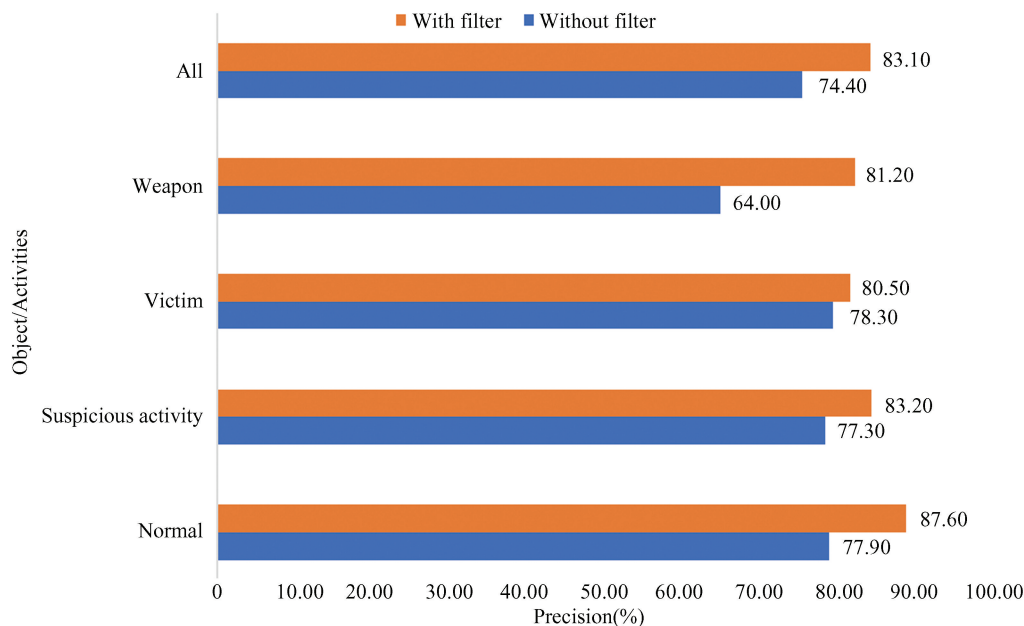


Fig.9 Precision graph for suspicious object/activity detection under environmental complexities with and without filter

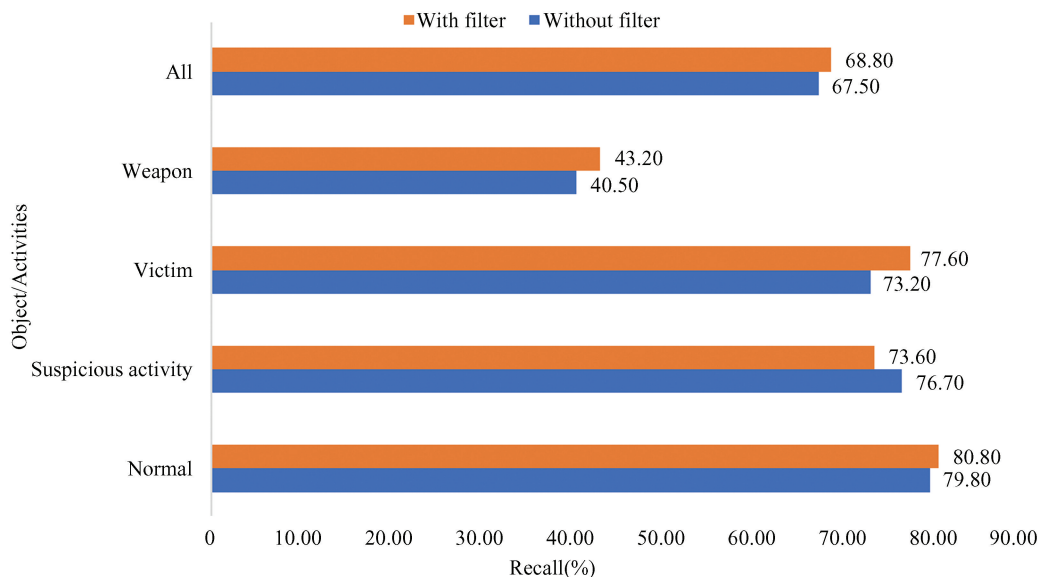


Fig.10 Recall graph for suspicious object/activity detection under environmental complexities with and without filter

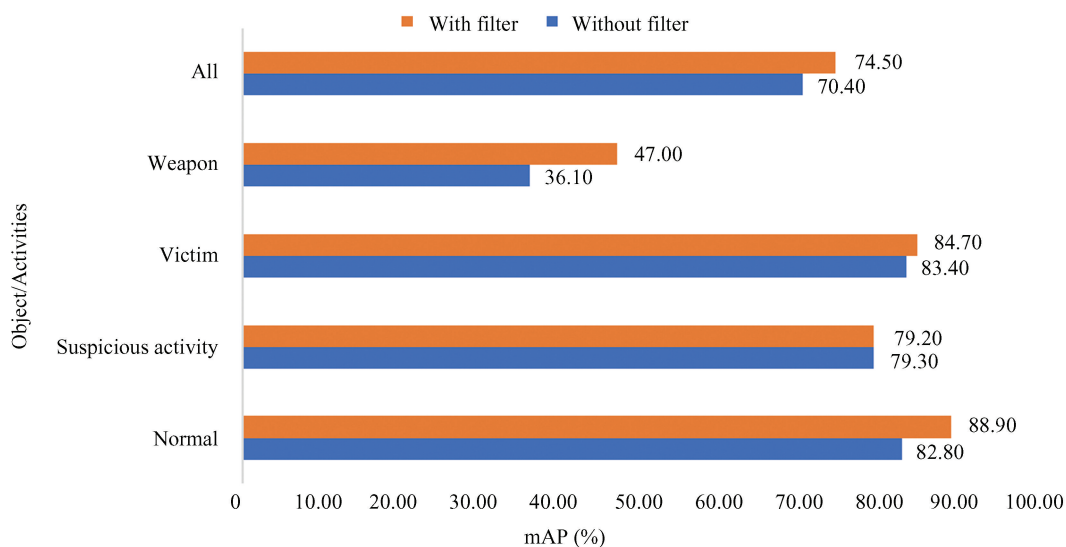


Fig. 11 mAP graph for suspicious object/activity detection under environmental complexities with and without filter

Fig.12 presents the execution time graph comparison under environmental complexities with and without applying filters. The graph is presented for 640×640 resolution. The comparison shows that the execution time for all classes is approximately 30–31 ms for testing samples.

4.2 Comparative Study

In this section, a comparative analysis is presented for suspicious object/activity detection among different state-of-the-art models, including the proposed model. Table 2 presents a comparison of various models used for detecting suspicious objects

and activities, highlighting their strengths and limitations. The deep learning model^[22] focuses solely on activities involving children and detects suspicious activities, but not objects, without considering environmental complexities or image data processing. YOLOv3^[23] is capable of processing image data and detecting objects, with YOLOv4 specifically targeting weapons^[24], but neither detects suspicious activities nor accounts for environmental complexities. The LSTM-CNN^[25] model is designed for crime-specific, detecting suspicious activities but not objects, while improved YOLOv4^[26] processes image data for object

detection without specifying activity types. In contrast, the proposed model stands out by addressing environmental complexities, processing image data, and detecting both suspicious objects and activities, offering a more comprehensive solution.

Fig.13 presents a comparative analysis of mAP under varying environmental conditions such as noise, blur, and low illumination. The results indicate that the model performs best under low illumination with mAP of 76.53% that is followed by blurred conditions with 74.83%. The presence of noise leads to the lowest performance at 72.13%. On average, the model achieves a mAP of 74.50%, indicating that further optimization is needed to improve its reliability under such challenging conditions.

Table 2 Feature comparison for suspicious object/activity detection

Model	Activity type	Environmental complexities	Image data processing	Suspicious object	Suspicious activities
Deep learning [22]	Only children	No	No	No	Yes
YOLOv3 [23]	-	No	Yes	Yes	No
LSTM-CNN [25]	Crime	No	Yes	No	Yes
Improved YOLOv4 [26]	-	No	Yes	Yes	No
Proposed method	Not specific	Yes	Yes	Yes	Yes

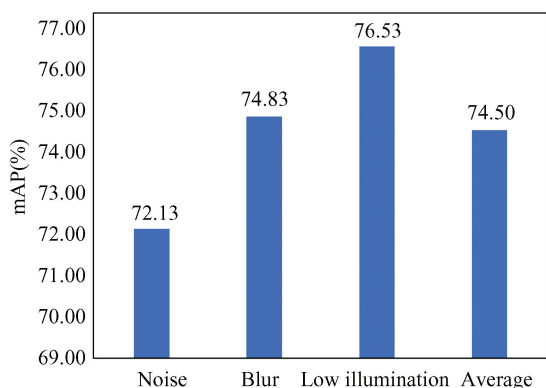


Fig.13 Comparative mAP analysis under different environmental conditions

Fig.14 presents the result comparison of the proposed model with existing works. YOLOv3 [23] model was presented for objected detection and achieved mAP of 65.70%. Whereas the proposed model (YOLOv8) had achieved a mAP of 74.50%. This shows an approximately 13% improvement in mAP compared to the existing models. This improvement is observed due to the integration of a sliding window adaptive filter, which removes unwanted noise factors from image. The integration of YOLOv8 model enhanced training performance due to

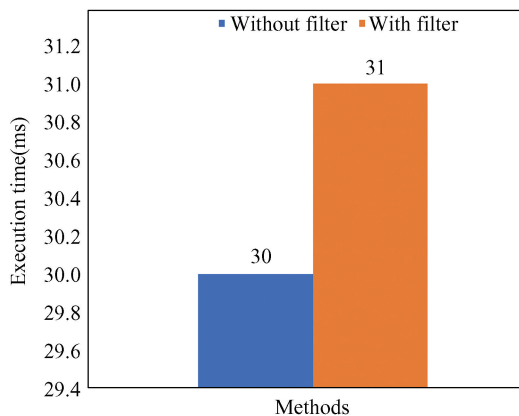


Fig.12 Execution time graph for suspicious object/activity detection under environmental complexities with and without filter

improvements in the loss function, better anchor box selection approach, use of backbone networks for feature extraction, and post-processing enhancements. Integration of all these models makes the proposed model more robust and efficient.

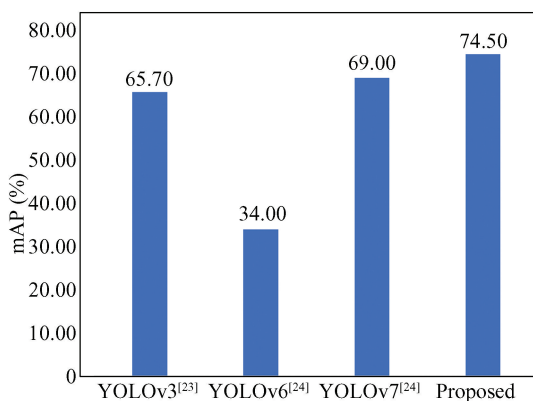


Fig.14 Comparative mAP Analysis

From the methodology and results presented, the primary challenge that the proposed approach directly addresses is environmental complexities. These complexities are related to real-time video/images complexities such as noise, blur, low illumination, etc. For this, the adaptive sliding-window bilateral

filter is presented to handle these issues while preserving edge and texture information, which improves performance compared to applying no filter. The experiment results show improved detection metrics for the filtered images overall but do not isolate each specific challenge that can be considered as limitation of the work.

5 Conclusions

In this paper, deep learning techniques for suspicious object detection are explored. Several challenges are identified in the recognition of suspicious objects and activities in videos, including illumination changes and occlusion, noise, poor resolution. To address these challenges, the YOLO model is adopted for its efficiency and accuracy in object detection. In this initial step of methodology, suspicious objects and activities are detected both in static images and in individual video frames. The YOLOv8 model, known for its improved detection capabilities, is utilized for this purpose. Then the model is advanced to mitigate the impact of noise and other environmental factors that could affect detection accuracy. An adaptive sliding window-based filter is proposed to remove both local and global noise effectively. This filtering technique is then cascaded with the YOLOv8 model, enhancing its ability to detect and classify objects under noisy conditions. The combined approach, using the adaptive sliding window-based filter with the YOLOv8 model, results in a significant improvement in performance. The model achieves an average mAP of 74.5%, which is a 13% improvement compared to existing methods. The proposed model has achieved an mAP of 74.5% which needs to be improved in the future. This can be accomplished by refining the existing model architecture, optimizing hyperparameters, and incorporating more advanced techniques such as attention mechanisms or transformer models.

References

[1] Gawande, Hajari K, Golhar Y. Novel person detection and suspicious activity recognition using enhanced YOLOv5 and motion feature map. *Artificial Intelligence Review*, 2024, 57; article number 16. DOI: 10.1007/s10462-023-10630-0.

[2] Tutar H, Güne A, Zontul M, et al. A hybrid approach to improve the video anomaly detection performance of pixel-

and frame - based techniques using machine learning algorithms. *Computation*, 2024, 12 (2) : 19. DOI: 10.3390/computation12020019.

[3] Vidya M Q M, Selvakumar S. An effective framework of human abnormal behaviour recognition and tracking using multiscale dilated assisted residual attention network. *Expert Systems with Applications*, 2024, 247: 123264. DOI: 10.1016/j.eswa.2024.123264.

[4] Alruwais N M, Zakariah M. Student recognition and activity monitoring in e - classes using deep learning in higher education. *IEEE Access*, 2024, 12: 66110-66128. DOI: 10.1109/ACCESS.2024.3354981.

[5] Wang J M, Hu F J, Abbas G, et al. Enhancing image categorization with the quantized object recognition model in surveillance systems, *Expert Systems with Applications*, 2024, 238 (Part E) : 122240. DOI: 10.1016/j.eswa.2023.122240.

[6] Talib M, Al-Noori Ahmed H Y, Suad J. YOLOv8-CAB: Improved YOLOv8 for real-time object detection. *Karbala International Journal of Modern Science*, 2024, 10 (1) : Article 5. DOI: 10.33640/2405-609X.3339

[7] Sun Z W, Hua Z X, Li H C, et al. Flying bird object detection algorithm in surveillance video based on motion information. *IEEE Transactions on Instrumentation and Measurement*, 2024, 73: Art No. 5002515. DOI: 10.1109/TIM.2023.3334348.

[8] Gautam D, Gupta H, Shekhar H, et al. Suspicious object tracking with YOLOv3 with Python using Open-CV. *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Greater Noida, 2023: 1772 - 1775. DOI: 10.1109/ICACITE57410.2023.10182703.

[9] Pullakandam M, Loya K, Salota P, et al. Weapon object detection using quantized YOLOv8. *2023 5th International Conference on Energy, Power and Environment: Towards Flexible Green Energy Technologies (ICEPE)*, Shillong, 2023: 1-5. DOI: 10.1109/ICEPE57949.2023.10201506.

[10] Jebur S A, Hussein K A, Hoomod H K, et al. Novel deep feature fusion framework for multi - scenario violence detection. *Computers*, 2023, 12 (9) : 175. DOI: 10.3390/computers12090175.

[11] Huszár V D, Adhikarla V K, Négyesi I, et al. Toward fast and accurate violence detection for automated video surveillance applications. *IEEE Access*, 2023, 11: 18772-18793. DOI: 10.1109/ACCESS.2023.3245521.

[12] Jain N, Gupta V, Tariq U, et al. Fast violence recognition in video surveillance by integrating object detection and Conv - LSTM. *International Journal on Artificial Intelligence Tools*, 2023, 32(3) : 2340018. DOI: 10.1142/S0218213023400183.

[13] Faisal M M, Mohammed M S, Abduljabar A M, et al. Object detection and distance measurement using AI. *2021 14th International Conference on Developments in eSystems Engineering (DeSe)*. Piscataway: IEEE, 2021:

559–565. DOI: 10.1109/DeSE54285.2021.9719469.

- [14] Wang J, Yang W, Guo H, et al. Tiny object detection in aerial images. 2020 25th international conference on pattern recognition (ICPR). Piscataway; IEEE, 2021; 3791–3798. DOI: 10.1109/ICPR48806.2021.9413340.
- [15] Mahmmod B M, Abdulhussain S H, Naser M A, et al. 3D object recognition using fast overlapped block processing technique. *Sensors*, 2022, 22 (23) : 9209. DOI: 10.3390/s22239209.
- [16] Narejo S, Pandey B, Vargas Esenarro D, et al. Weapon detection using YOLOv3 for smart surveillance system. *Mathematical Problems in Engineering*, 2021 (2021) : 1–9. DOI: 10.1155/2021/9975700
- [17] Fang W, Wang L, Ren P M. Tinier-YOLO: A real-time object detection method for constrained environments. *IEEE Access*, 2019, 8 : 1935 – 1944. DOI: 10.1109/ACCESS.2019.2961959.
- [18] Gali M, Sunita D, Suresh K. real-time image based weapon detection using YOLO algorithms. *International Conference on Advances in Computing and Data Sciences*. Cham; Springer International Publishing, 2022; 173–185.
- [19] Wang G, Cheng Y F, An P, et al. UAV – YOLOv8: A small – object – detection model based on improved YOLOv8 for UAV aerial photography scenarios. *Sensors*, 2023, 23 (16) : 7190. DOI: 10.3390/s23167190 .
- [20] GitHub, Inc. Brief summary of YOLOv8 model structure# 189. <https://github.com/ultralytics/ultralytics/issues/189> [Accessed on 12–08–2024].
- [21] Roboflow, Inc. <https://universe.roboflow.com/suspicious-movement/suspicious-detection/dataset/4>.
- [22] Yash Pravesh S. Multi class weapon detection system – Github ZCv24 dataset. <https://universe.roboflow.com/s-yash-pravesh-mu7qu/multi-class-weapon-detection-system-github-zcv24>.
- [23] Vallathan G, John A, Thirumalai C, et al. Suspicious activity detection using deep learning in secure assisted living IoT environments. *The Journal of Supercomputing*, 2021, 77 : 3242–3260. DOI: 10.1007/s11227-020-03387-8.
- [24] Fang W, Wang L, Ren P M. Tinier-YOLO: A real-time object detection method for constrained environments. *IEEE Access*, 2019, 8 : 1935 – 1944. DOI: 10.1109/ACCESS.2019.2961959.
- [25] Ganagavalli K, Santhi V. YOLO – based anomaly activity detection system for human behavior analysis and crime mitigation. *Signal, Image and Video Processing*, 2024, 18 : 417–427. DOI: 10.1007/s11760-024-03164-7.
- [26] Wang G B, Ding H W, Duan M L, et al. Fighting against terrorism: A real-time CCTV autonomous weapons detection based on improved YOLOv4. *Digital Signal Processing*, 2023, 132 : 103790. DOI: 10.1016/j.dsp.2022.103790.