International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhanced Surveillance: Triple Background Subtraction with YOLO V8

Tabiya Manzoor Beigh*[1], V. Prasanna Venkatesan[2], J. Arumugam[3], S. Geetha[4]

Submitted: 29/01/2024 **Revised**: 07/03/2024 **Accepted**: 15/03/2024

Abstract: In congested areas like malls, airports, train stations, etc., video surveillance facilitates monitoring and provides a sense of security. There is a need for advancements in video surveillance technology to be more robust and efficient. Due to increasing terrorist and criminal activities, addressing the unattended static artefacts on public premises has become a high-priority task. To mitigate human and financial loss, abandoned objects should be dealt with the utmost priority. Identifying abandoned or removed objects in surveillance footage proves challenging due to its complexity, driven by occlusion and sudden alterations in lighting. This paper proposes a novel technique for detecting and classifying abandoned objects, particularly bags. The work aims to automatically detect abandoned objects. The method involves a robust triple background subtraction technique that extracts background using three sub-models. A Convolutional Neural Network (CNN) -based classifier is used to classify abandoned artefacts. You Only Look Once YOLO V8 is used as the classification algorithm. After the foreground is extracted, graph-based segmentation is used to extract candidate static objects. Final static objects are extracted using the stability rank calculation method. The suggested approach is validated on three benchmark datasets: PET 2006, PET 2007, and i-LIDS AVSS. Performance parameters include precision, recall, and accuracy. In realistic environments and factual situations like poor illumination and occlusion, the proposed solution outperforms the existing methods. The proposed methods help in the reduction of false positives, reducing the false alarm rate. The proposed method reaches an accuracy of 99.5%, precision of 93%, and recall of 90%, much higher than earlier proposed systems.

Keywords: *Abandoned objects; Background subtraction; CNN; Segmentation; Video surveillance; YOLO V8.*

1. Introduction

Computer vision has achieved magnificent results in analysing videos and images, contributing to a more comprehensive understanding of the visual environment. Video surveillance is one of the areas that focuses on understanding the environment to support better decision-making processes in various applications. In real-time video surveillance, object detection and tracking go hand in hand to gather the motion information of the objects in the scene. Object detection and tracking are used in many fields, which include traffic management [1], healthcare [2], sports analytics [3], robotics [4], autonomous vehicles [5], and surveillance and security [6] [7]. In traffic management, vehicle travel is examined and tracked to report offenses, if any, and improve traffic. In the healthcare industry, it is used to identify and follow abnormalities in X-ray and MRI images and monitor the movement of surgical tools as they are being used. In sports analytics, athletes, balls, and other items are tracked while they are in play, which are used to generate statistics for broadcast usage as well as for performance analysis.

[1,3]*Research Scholar Department of Computer Science, Pondicherry University, 605014- India*
[1] *ORCID ID : 0000-0001-6358-8161*
[2] *Professor, Department of Banking Technology, Pondicherry University, 605014-605014*
[2] *ORCID ID 0000-0002-1444-0918*
[3]*ORCID ID 0000-0002-4374-019*[4]
[4] *Assistant Professor, Department of Banking Technology, Pondicherry University, 605014- India*
[4]*ORCID ID : 0000-0002-1345-8403*

Depending on the domain, the objects detected can be either moving or stationary. Moving-object detection has much research work attributed to it. As compared to dynamic object detection, there is a lack of attention in the static object detection domain. Static object detection includes the processes of locating, tracking, and confirming the presence of an object. Objects of interest may be surrounded by unimportant things with similar visual appearances in complicated backgrounds. Detecting static objects and observing in cluttered and crowded situations is still an open problem. For security experts, abandoned objects pose a serious dilemma. These items might include explosive devices or dangerous materials, putting people's safety and well-being at risk when left unattended in public areas. Due to the increasing number of terrorist attacks, it has become mandatory to address every unattended object at the earliest possible time. This could help mitigate the hazards and causalities. As a result, it is important to improvise when looking for abandoned luggage or bags because they may contain explosives or other dangerous items that should not be present in public areas like bus stops, airports, and train stations. Abandoned object detection can be done either by a tracking approach or background subtraction. The main challenge in static object detection lies in subtracting intricate backgrounds where the main subjects may be flanked by less important or seemingly similar elements. It is a difficult task that calls for efficient background suppression strategies to distinguish between objects and

background noise. Depending on the specific requirements, a particular method should be employed.

In this study, the detection of abandoned static objects, especially bags, in airports and railway stations is done. In the suggested study, three foreground sub-models are used to effectively segregate the foreground from the background. To the best of our knowledge, abandoned object detection research has often relied on dual background models. This is the first work that formulated triple background subtraction for addressing background subtraction complexities such as occlusion, various illumination changes, etc. The main contributions are: -

- Background subtraction is carried out using the triple background subtraction model. Using the three sub-models, static, semi-static, and dynamic foregrounds are extracted. Candidate region identification is done by finding the difference between the extracted foregrounds.
- Candidate static objects are identified using graph-based segmentation. Spectral clustering is applied to get the disjoint regions. Validation of the candidate static objects is done using the thresholding method. Threshold value calculation is done using the stability rank calculation method.
- YOLO V8 is used for the final classification of stationary objects.

The remainder of this article is organized as follows: The literature review is described in depth in Section 2. A comprehensive explanation of the suggested technique is included in Section 3. Results and performance analysis of the suggested strategy are provided in Section 4. The task is finally concluded in Section 5.

2. Related Works

The study used a variety of methods, including item tracking, object identification, and object categorization. To extract the objects of interest that are static, many strategies were used. The methods used in the literature are explained here. An unsupervised approach is proposed for developing a scene-specific pedestrian detector that is flexible enough to be trained across many target domains without the need for human-annotated target samples [8]. The process entails moving a general detector from a labelled source domain dataset to various target domains and then expanding it to a multi-level classifier to gather samples from both domains. This method outperformed available scene-specific pedestrian recognition techniques and demonstrated promising results in enhancing pedestrian detection in diverse domain shift conditions. In [9], a novel background removal method based on deep neural networks for recognizing foreground objects in video surveillance systems is suggested Major challenges such as camouflage, abrupt illumination changes, and shadows are addressed. It leverages optical-flow details to add temporal information. It inputs a set of stacked images (input, background, and optical) to the convolutional neural network to extract the foreground. The proposed method is trained using ground truth images drawn from a CDnet-2014 dataset and randomly selected training photos. To identify surface defects in the electronics and manufacturing industries, an object detection technique based on different variants of YOLO has been used [10]. The model detects defects in electronic object surfaces with real-time inference speed. It extracts features in multiple scales with one self-attention module in it. In [11], the detection of firearms in CCTV images is emphasized. It identifies perilous situations within the images. The system is designed in such a way that it promptly notifies the human operator upon detection of any weapon. The results obtained in terms of sensitivity and specificity are better than those obtained by the existing techniques. For identifying static abnormalities, a novel intelligent anomaly detection system that relies on self-organising maps to supplement the ineffectiveness of surveillance systems is suggested [12]. Optimised SOM connection weights are used to analyse the cluster distribution of the neurons and categorize anomalies. The irregular object is found using the shortest path technique. The suggested method is applied to the CD Net 2014 dataset. A stereo-vision-based system is developed to detect objects in marine habitats [13]. It takes into account static as well as dynamic objects in the environment. Image processing techniques have been used to detect the objects. For tracking selected boats, the EKF algorithm has been used. The results obtained from the tracking process are compared with RTK-GPS data. The results produced by the system were robust and better than the results of existing techniques. An intelligent video surveillance system is proposed that works in both static and dynamic environments [14]. The proposed system employs several cameras to gather various detected items through various channels and combine them. Objects are monitored based on their attributes, which include their structure, appearance, and resemblance. Dense block-based CNN is utilized to enhance the accuracy of accurately detecting the objects. Shelf background subtraction method-based object detection is used. Foreground items and background information are subtracted from the sequence of frames using this shelf background approach, which also learns online information. A system in which the background region is found using the averaging approach is suggested [15]. Utilizing a fuzzy integral technique, identification of the foreground is done. Candidate static objects are found using finite state machines at three levels. YOLO V5 is used for the classification of abandoned objects. A method of detecting stolen and abandoned objects is proposed [16]. By using a two-level detection technique in a spatial

and temporal context, an object that is suspected of being abandoned might be categorized as either stolen or abandoned. Benchmark datasets have been used to demonstrate the suggested techniques. A useful technique for locating abandoned bags in CCTV footage is suggested [17]. To separate the foreground, short and long-term models are blended. It is possible to identify abandoned static items by following down the owners of the luggage and exploiting temporal transitions in the video. Benchmark datasets PETS 2006 and AVSS 2007 were utilized to evaluate the proposed approach. A heat map-based approach for abandoned object detection is proposed [18]. Updation of background is based on the current change as well as the previous history of spatiotemporal information, which is implicitly encoded into a heat map. The suggested method has undergone two rounds of datasets that are common and extensively used, such as the CAVIAR Dataset (CAVIAR, 2003) and the Imagery Abandoned Baggage Dataset Library for Intelligent Detection Systems (i LIDS, 2007). The efficacy and reliability of our technology have been confirmed by comparison of the acquired findings with cutting-edge methodologies. An efficient bank surveillance system is proposed in [19]. It uses two object detection models, namely YOLO V4 and YOLO V5. Both models were used to detect people and weapons on the bank premises. The dataset consisted of four classes of weapons. It was obvious from the results that YOLOV4 and YOLO5 achieved better results than the existing techniques. A unique method of extracting static objects utilizing point feature matching and an enhanced hashing methodology, along with the integration of spatial and temporal relationships, is implemented [20]. It uses a hash-based model to detect the objects. It uses an SVM-based classifier, which worked very well in these specific environment settings. A probabilistic neural network along with CNN is employed to analyse surveillance video, and a CNN is used. [21]. The simulation outcome demonstrates that the CNN-PNN achieved enhanced simulation results for the best object detection in video streaming. A technique for real-time detection of abandoned objects is proposed, in which background subtraction is the first stage of the system [22]. It follows a generic set of steps that include foreground static area recognition and identification of the class of object. A robust method for the identification of objects in surveillance videos is proposed [23]. It uses an adaptive Gaussian- based approach to segment the foreground. A filter model based on fuzzy logic is put into practice to get rid of the noise in the foreground segmented frames. It has been observed that the results obtained are very accurate and robust. To detect abandoned objects in surveillance videos, CNN-based YOLO V4 is used [24]. The proposed method was implemented on the custom-made dataset. The dataset is comprised of six different classes. The experiment's findings revealed superior performance for abandoned item recognition. Techniques available in the literature in this research domain are given in Table 1.

Table 1. Assorted techniques and scrutiny of parameters in prior research.

References	Dataset	Active field of Study in Video Surveillance	Detection Algorithms	Evaluating Parameters
[8]	PNNL-Parking Lot-2/Pizza PETS2009 Town Center CUHK Square	Pedestrian scene-specific detector	YOLO CycleGAN CSTN	Overlap metric PR curves F-measure Average Precision
[9]	CD net-2014	Foreground Segmentation	CNN	F-measure Precision Recall False positive rate (FPR) False Negative Rate (FNR)
[10]	CD net-2014	Operationally cost-effective intelligent surveillance system	Self-Organising Maps (SOM)	Percentage of wrong classifications (PWC) False Alarm Rate (FAR) Missing Rate (MR)
[13]	AVSS i LIDS 2007	Abandoned Object Detection	YOLO V5	Precision Recall Accuracy
[15]	PETS 2006 AVSS 2007	Abandoned Bag Detection	Temporal location-based trajectory	F – measure Precision Recall

[16]	CAVIAR 2003 I LIDS 2007	Abandoned Object Detection	Spatiotemporal information encoded in heat map	True Detected False Detected

3. Proposed Method

In this section, abandoned objects in surveillance videos are identified through the triple background subtraction model. The suggested approach finds static objects in an area monitored by CCTV cameras. The core aim of the suggested study is to identify objects or items that have been left unattended for a stipulated amount of time to prevent potentially dangerous incidents in civic spaces. The process initiates after capturing the input of CCTV footage from the potential camera. Videos are segregated into a set of frames. A frame-by-frame division of the input video is done. The triple background subtraction method is used to extract the foreground by using three sub-models with variations in the rate of their learning. To gather the motion of the objects in the foreground, pixel information is exploited. The static foreground regions are identified by using the difference operation between the foregrounds. The two sub-models use three-frame differencing with higher learning rates to extract the foreground. The third model uses Adaptive Gaussian Mixture Models (AGMM) for foreground learning. Candidate static regions are segmented using graph-based image segmentation. The outcome of this step will be a list of candidate static objects. For each object in the candidate static object list, the stability rank is calculated. The final static object list is composed based on the static scores of objects. The validated candidates are given as input to the classifier to detect abandoned static objects. The suggested system's overall flow is depicted in Fig 1.

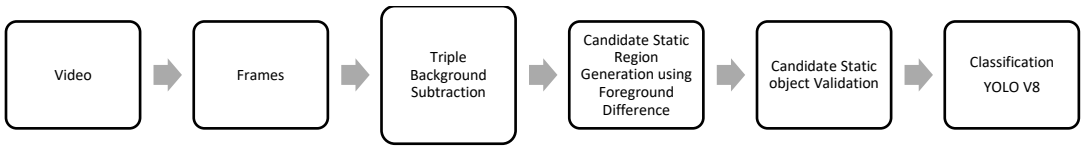

Fig 1: Workflow of the proposed methodology

3.1. Triple Background Subtraction

The triple background model is used to identify abandoned objects in the surveillance videos. The videos are converted into a set of frames. Pixel values among all the frames are normalized for smoother processing. An abandoned object is not by nature an integral part of the video scene. It emerges in the scene at a timestamp t_i, after which it is regarded as part of the background due to its stationary nature. To address the issue of detecting objects blended in the background, a triple background subtraction model that comprises three sub-models (static, semi-static, and dynamic) is used. Each of the sub-models has different learning rates. The dynamic background model learns to extract the moving foreground objects in the current frame. The dynamic background model has a higher learning rate as it has to capture the moving objects (temporary static objects) in the changing environment. The semi-static background model learns to extract those moving items in the foreground until the last frame and stops in the current frame. The static background model extracts the static as well as the moving foreground objects. The static background model has the lowest learning rate among the three sub-models as it has to capture prolonged static objects.

In our work, dynamic and semi-static background models work similarly with different learning rates. Both models use a frame differencing technique to calculate the region of motion. Initially, the first frame is considered the background frame. The number of iterations ("N") is defined. Image differences between the current frame and the previous frame are calculated. The difference is calculated for the previous two frames and shown in eq. (1) and eq. (2).

$$MR_1\left(F_{t,t-1}\right) = \begin{cases} 0 & if\ |F_t - F_{t-1}| > T_0 \\ 1 & else \end{cases} \quad (1)$$

$$MR_2\left(F_{t-1,t-2}\right) = \begin{cases} 0 & if\ |F_{t-1} - F_{t-2}| > T_1 \\ 1 & else \end{cases} \quad (2)$$

Where F_t is the current frame, F_{t-1} is the previous frame, and F_{t-2} is the previous frame to F_{t-1}. T_0 *and* T_1 are predefined thresholds for binarization. Background updation is done using the eq. (3)

$$Bg_{(t)} = \alpha\,(Bg_t) + (1-\alpha)Bg_{t-1} \quad (3)$$

where α is the learning rate. This algorithm has a higher learning rate, which makes it suitable for capturing backgrounds blended with static objects. The static background model uses the Adaptive Gaussian Mixture Model (AGMM) to capture long-term static objects [25]. In a given frame F_t, if an object enters the scene and remains stationary for a specific time, it is temporarily allotted a different cluster. The weight of the recently added cluster will increase because of the occlusion

occurrence in the previous background. If the weight of the object is higher than the total weight of foreground objects, it will be classified as background, as shown in eq. (4).

$$Bg = (arg\ min_b\ (w_i > (1 - C_f)) \quad (4)$$

Where, w_b is the weight of a foreground cluster and C_f is a measure of the maximum data proportion that foreground objects can encompass without altering the background mode.

3.2. Candidate Generation using Foreground Difference

The difference (DF1) between the extracted semi-static foreground and dynamic background is calculated. Similarly, the difference (DF2) between the static and dynamic backgrounds is calculated. The difference between (DF1) and (DF2) is calculated to detect the stationary foreground. The static foreground region is segregated from the static foreground image and dynamic foreground image. After the identification of candidate static regions, the identification of static objects is done. This task is achieved by applying graph-based segmentation. To alleviate the disturbance caused by noise and other irregularities, the segmentation process is preceded by preprocessing techniques. Two morphological operations, erosion and dilation, are applied. After the eradication of noise, segmentation is carried out. In graph-based image segmentation, pixels represent the nodes, and the weights on edges represent the similarity among pixels in terms of color and other intensity values. A graph partitioning algorithm known as spectral clustering is applied to get the disjoint regions. The refinement of segments is carried out by merging based on criteria such as color in our case. The segmented output will represent the list of candidate static objects.

3.3. Candidate Static Object Validation

Initially, the stability rank of each candidate static object is initialised to 1. The stability ranking for each candidate static object is determined by finding the intersection between the current and previous frames for a specific time, like 60 seconds. If there is a commonality of objects among the frames, the stability rank of an object is incremented. The procedure is followed for the list of static objects in the frame set. The object is labelled as static if the final static score is greater than threshold value of 700. This step will produce a set of final static objects. The pseudocode for final static objects is given.

Pseudocode for Final Static Object determination

Input: List of Candidate Static Objects {$CS0_1$, $CS0_2$, $CS0_3$, ….., $CS0_n$)
Output: List of Final Static Objects {FSO_1, FSO_2, FSO_3, ….., FSO_n}

Initialize the list of Final Static Objects FSO list = {}
Set the first frame as the Background frame
Set the ceiling on the number of frames equal to 1000
From 2nd frame till 1000 (maximum allowable frame quantity)
  Present Frame = Read the frame
  Find the intersection of the Present Frame and Background Frame
    For each Candidate Static Object $CSO_{j\ in}$ the list
      If the intersection consists of CSO_j
        Add 1 to the stability rank of CSO_j
  Background Frame =Present frame
  Present Frame = Read the next frame
End
For each candidate Static Object (CSO_j)
  If the stability rank of CSO_j is greater than the threshold
    Add candidate Static Object to the Final Static Object List
List the items of the Final Static Object list

3.4. Classification

The basic purpose of classification modules is to distinguish luggage from various classes of items, and this classification may be carried out in a variety of ways. In the suggested study, we used YOLO V8 to classify the items and identify them. YOLO is a popular one-stage strategy object detector model. The term "YOLO" stands for "you only live once." It pertains to the fact that a single neural network evaluation is needed to predict object classes. The recent object detection model called YOLO V8, is also capable of classifying images and segmenting instances. Its lightweight design, increased speed, and improved accuracy make it the ideal model for use in real-time applications like CCTV monitoring. We employed the YOLO V8n model (also known as the YOLO V8 nano variant) in this particular investigation. Fig 2 shows the design of the YOLO V8 model [26]. The YOLO V8 model consists of three components, namely the backbone, split head, and loss metrics.

3.4.1. Backbone: This layer has several convolutional layers that aid in visualizing the picture at various sizes and resolutions. The addition of new convolution layers C2f aids in the extraction of high-level characteristics. It takes into consideration the contextual aspects by incorporating a cross-stage partial bottleneck with two convolutions.

3.4.2. Split head: This is the decoupled neck and head combination that was previously utilised in YOLO V5. The extracted high-level features are all combined by the neck. It has an anchor-free detection model that accurately anticipates an object's center. Since there is no offset

mechanism in anchor-free detection, it can be used in real-time scenarios where the dataset objects may have varying resolutions and magnitudes. To calculate the object detection score, sigmoid function is used. The class probabilities are represented by the objects' likelihood of belonging to each potential class using the SoftMax function.

3.4.3. Loss metrics: The detection results produced by the head are dependent on the loss metrics. Loss metrics run for a series of iterations to get the desired result. Losses include object loss, box loss, and confidence loss. Bounding box loss is handled by YOLOv8 using the CIoU and DFL loss functions, while classification loss is handled using binary cross-entropy.

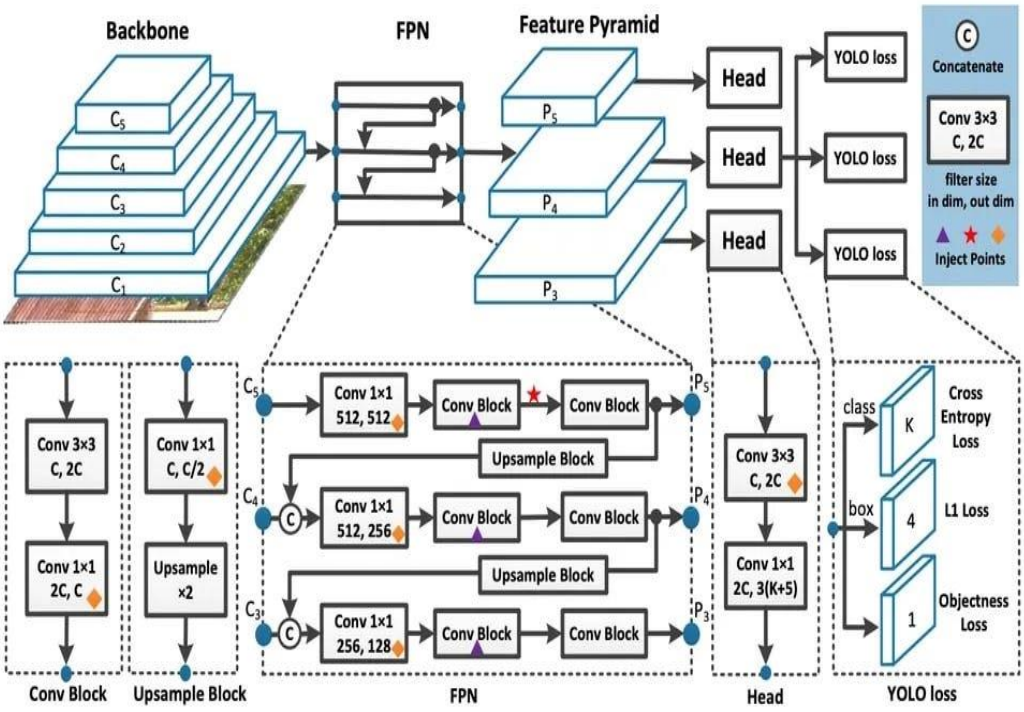

Fig 2: Model design of YOLO V8

4. Results and Discussions

The performance assessment of the suggested system is provided in this section. With an i7 CPU, 16 GB of RAM, and an NVIDIA RTX GPU 3050, the system performance is implemented using the Python deep learning library PyTorch. To demonstrate the effectiveness of our suggested methodology, a comparison of this approach with the existing systems is conducted.

4.1 Dataset Description

The suggested strategy has been used on three publicly accessible benchmark datasets. Real-time CCTV footage from numerous sites, including airports and train stations, is included in a publicly accessible dataset. The following provides comprehensive information on the publicly available datasets:

PETS 2006: This dataset comprises a variety of seven types of events in videos. These videos feature a variety of increasingly challenging unattended baggage scenarios that were recorded in train stations. [27].

PETS 2007: This dataset has four events related to unattended baggage or the exchange of baggage. Each scenario in this movie has two occurrences that are recorded from four distinct perspectives. The abandoned object that has been used in our work [28].

i-LIDS AVSS: This dataset is a subset of the i- LIDS dataset. It contains CCTV footage of two situations, including abandoned luggage and illegally parked cars. In this work, only abandoned luggage videos have been used. [29]

4.2 Quantitative metrics

The proficiency of the proposed strategy. is assessed using the following performance metrics:

True Positive (TP): Events that identify “bag” as “bag”.

True Negative (TN): Events that identify a “non-bag” object as “non-bag”.

False Positive (FP): Events that identify a “non-bag” object as “bag.” These incidents contribute to false alarms.

False Negative (FN): Events that identify “bag” as a “non-bag” object. These instances are missed detections.

Precision: It is the proportion of correctly identified bags to the sum of correctly identified objects as bags and objects incorrectly identified as bags. A higher precision

value suggests that the algorithm is better at avoiding false positives.

$$\text{Precision} = \frac{TP}{(TP+FP)}$$

Recall: It is the measure of bags unidentified by the system. Higher recall values indicate that the algorithm is better at capturing all positive instances, minimizing false negatives.

$$\text{Recall} = \frac{TP}{(TP+FN)}$$

Accuracy: Accurately identifying and alerting people to the presence of abandoned items in a certain area or on security footage is what the algorithm is capable of. It gauges how effectively the system can distinguish between abandoned objects and other scene elements. It is the ratio of correctly predicted items to the total number of predictions computed by the system. Higher recall values indicate that the algorithm is better at capturing all positive instances, minimizing false negatives.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

4.3 Competitive Analysis

The study comparison of the current and suggested methods is described in this section. Accuracy, precision, and recall performance measures are examined and compared for current algorithms, including YOLO V5, YOLOV8 V4, KNN, RCNN, and SVM. Table 2 provides a comparative analysis of the proposed techniques with the existing ones.

Table 2. Evaluation and Comparison of Current and Proposed Approaches

Algorithm	*Evaluation metrics*		
	Precision	Recall	Accuracy
KNN	0.7	0.71	65
SVM	0.89	0.7	95.15
R- CNN	0.7	0.75	83
YOLO V4	0.8	0.86	84.55
YOLO V5	0.91	0.88	99
Proposed method	**0.93**	**0.90**	**99.5**

Table 2 highlighted the evaluation metrics achieved by different algorithms. It can be clearly seen that KNN achieves moderate precision and recall, but lower accuracy. SVM achieved high precision but lower recall, with good overall accuracy. R-CNN achieved moderate precision and high recall, with relatively lower accuracy. YOLO V4 achieved moderate precision and high recall, with relatively lower accuracy. YOLO V5 attained high precision and recall, with very high accuracy. The proposed method attained highest precision and recall, with very high accuracy.

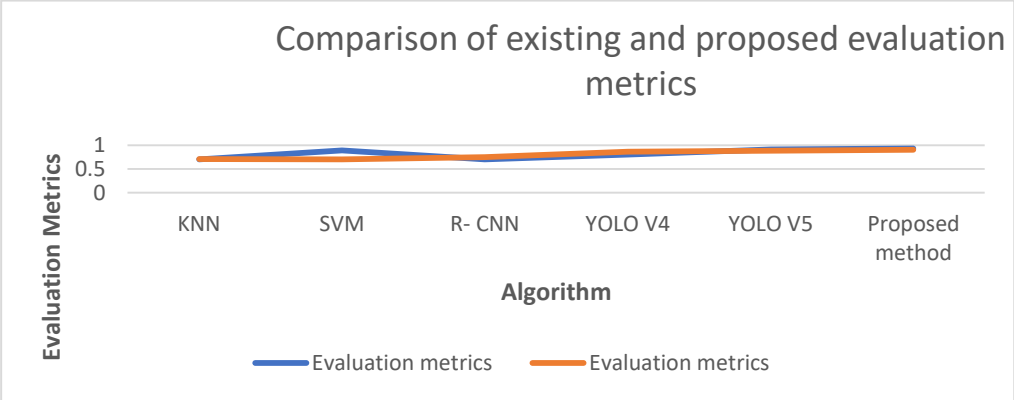

Fig 3: Precision and recall comparison of existing and proposed methods

From Fig 3., it can be seen that the proposed approach has achieved higher precision and recall levels. The precision achieved is 0.93 and recall is 0.90 which is not attained by any of the existing methods. The suggested method is far better than the previously described current approaches in terms of accuracy. The proposed method has achieved 99.5% accuracy, as shown in Fig 4.

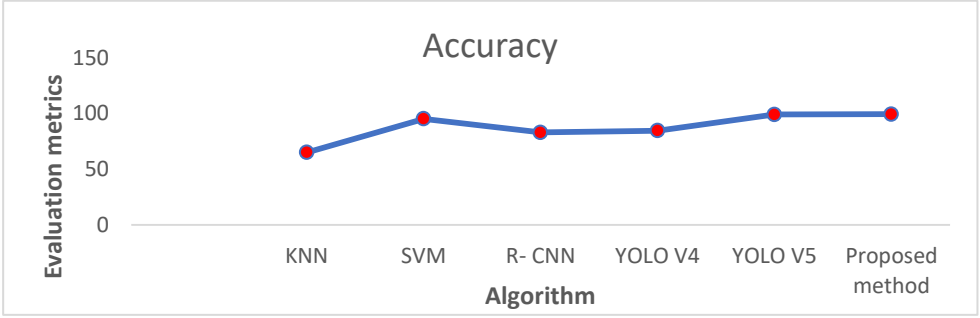

Fig 4: Comparative Accuracy Analysis: Existing Method and Proposed Approach

Table 3 shows the performance concerning the earlier research in the domain. The table presents the accuracy percentages achieved by various authors or studies, along with the accuracy of a proposed method. Mahalingam and M. Subramoniam [23] achieve the lowest accuracy of 65%, indicating comparatively lower correctness in their classifications. Palivela and Ramachandran [20] and Teja [15] also achieve high accuracies of 95.15% and 99%, respectively. The proposed method achieves the highest accuracy of 99.5%, indicating superior performance compared to the referenced studies.

Table 3. Assessment of Performance Metrics of Previous Studies and the Current Study

Author	Accuracy (%)
Mahalingam and M. Subramoniam [23]	65
Kiruthiga and Yuvaraj[21]	83
Lwin and Tun[24]	84.55
Palivela and Ramachandran[20]	95.15
Teja[15]	99
Proposed method	99.5

An accuracy-based comparison based on the available and suggested techniques is plotted in Figure 5. It visualises and compares the works of different authors.

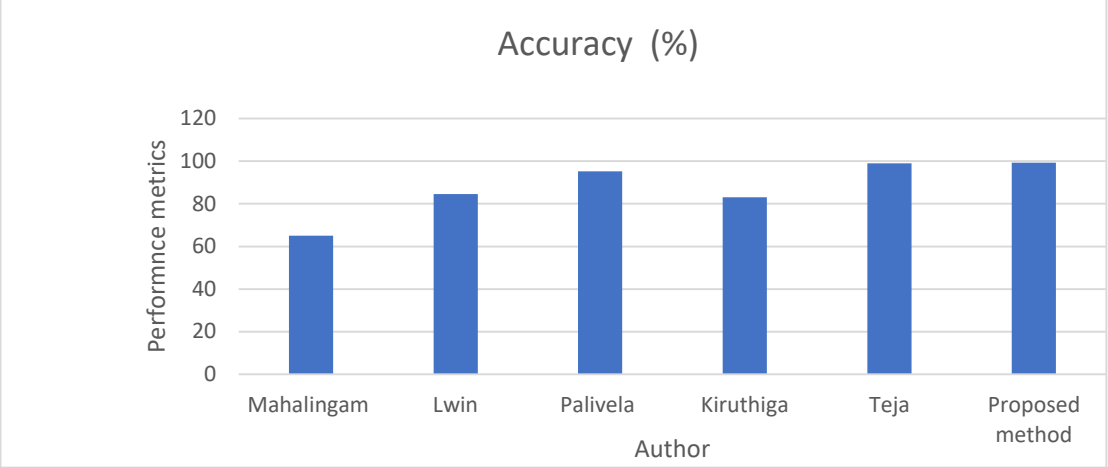

Fig 5: Comparative Graphical Illustration of Proposed Method and Existing Paper

5 Conclusion

In congested areas like malls, airports, train stations, etc., video surveillance facilitates monitoring and provides a sense of security. It is necessary to have a smart surveillance system that can prevent accidents. Abandoned object detection is one of the areas that ensure the security of people and property in highly crowded and sensitive areas. The study aims to identify objects or items that have been left unattended for an extended period to stop potentially dangerous attacks in public areas. The process initiates after capturing the input of CCTV footage. The video is divided into individual frames. The triple background subtraction method extracts the foreground by using three sub-models with different learning rates. The static, dynamic, and semi-static foregrounds are extracted. The difference between the extracted foregrounds is used to identify the static regions in the foreground. For the identification of static objects in the extracted foreground, graph-based segmentation is used. To get the static objects in the form of disjoint clusters, spectral clustering is used. Candidate validation is done by calculating the stability rank for each static object. Finally, the CNN-based classifier YOLO V8 is used to recognise abandoned bags. The proposed method enhances the accuracy in the detection of abandoned objects in less favourable environments such as occluded places, crowded locations, and varied illumination models. The performance of the new approach is then compared to some of the current methods. The proposed method has greatly reduced false positives which happen due to environmental factors and can lead to false alarms resulting in chaos. It is observed that the proposed method performs well when compared with the existing methods in terms of better precision, recall, and accuracy. The future study might incorporate items other than bags.

Data Availability The datasets used in the current study are taken from [27], [28], and [29] respectively.

Conflict of interest The authors declare that we have no conflict of interest.

References

[1] Kamble SJ, Kounte MR (2023) Application of improved you only look once model in road traffic monitoring system. International Journal of Electrical and Computer Engineering 13:4612–4622. https://doi.org/10.11591/ijece.v13i4.pp4612-4622

[2] Aldughayfiq B, Ashfaq F, Jhanjhi NZ, Humayun M (2023) YOLO-Based Deep Learning Model for Pressure Ulcer Detection and Classification. Healthcare (Switzerland) 11:.https://doi.org/10.3390/healthcare11091222

[3] Mavrogiannis P, Maglogiannis I (2022) Amateur football analytics using computer vision. Neural Comput Appl 34:19639–19654. https://doi.org/10.1007/s00521-022-07692-6

[4] Almanzor E, Anvo NR, Thuruthel TG, Iida F (2022) Autonomous detection and sorting of litter using deep learning and soft robotic grippers. Front Robot AI 9:. https://doi.org/10.3389/frobt.2022.1064853

[5] Musunuri YR, Kwon OS, Kung SY (2022) SRODNet: Object Detection Network Based on Super Resolution for Autonomous Vehicles. Remote Sens (Basel) 14:. https://doi.org/10.3390/rs14246270

[6] Shruthi, Pattan P, Arjunagi S (2022) A human behavior analysis model to track object behavior in surveillance videos. Measurement: Sensors 24:. https://doi.org/10.1016/j.measen.2022.100454

[7] Cai H, Song Z, Xu J, et al (2022) CUDM: A Combined UAV Detection Model Based on Video Abnormal Behavior. Sensors 22:. https://doi.org/10.3390/s22239469

[8] Mou Q, Wei L, Wang C, et al (2021) Unsupervised domain-adaptive scene-specific pedestrian detection for static video surveillance. Pattern Recognit 118:. https://doi.org/10.1016/j.patcog.2021.108038

[9] Vijayan M, Mohan R (2020) A Universal Foreground Segmentation Technique using Deep-Neural Network. Multimed Tools Appl 79:34835–34850. https://doi.org/10.1007/s11042-020-08977-5

[10] Wang J, Dai H, Chen T, et al (2023) Toward surface defect detection in electronics manufacturing by an accurate and lightweight YOLO-style object detector. Sci Rep 13:. https://doi.org/10.1038/s41598-023-33804-w

[11] Grega M, Matiolański A, Guzik P, Leszczuk M (2016) Automated detection of firearms and knives in a CCTV image. Sensors (Switzerland) 16:. https://doi.org/10.3390/s16010047

[12] Kim J, Cho J (2019) An online graph-based anomalous change detection strategy for unsupervised video surveillance. EURASIP J Image Video Process 2019:. https://doi.org/10.1186/s13640-019-0478-8

[13] Omrani E, Mousazadeh H, Omid M, et al (2020) Dynamic and static object detection and tracking in an autonomous surface vehicle. Ships and Offshore Structures 15:711–721. https://doi.org/10.1080/17445302.2019.1668642

[14] Adimoolam M, Mohan S, John A, Srivastava G (2022) A Novel Technique to Detect and Track Multiple Objects in Dynamic Video Surveillance Systems. International Journal of Interactive Multimedia and Artificial Intelligence 7:112–120. https://doi.org/10.9781/ijimai.2022.01.002

[15] Teja YD (2023) Static object detection for video surveillance. Multimed Tools Appl. https://doi.org/10.1007/s11042-023-14696-4

[16] Nam Y (2016) Real-time abandoned and stolen object detection based on spatio-temporal features in crowded scenes. Multimed Tools Appl 75:7003–7028. https://doi.org/10.1007/s11042-015-2625-2

[17] Lin K, Chen SC, Chen CS, et al (2015) Abandoned Object Detection via Temporal Consistency Modeling and Back-Tracing Verification for Visual Surveillance. IEEE Transactions on Information Forensics and Security 10:1359–1370. https://doi.org/10.1109/TIFS.2015.2408263

[18] Foggia P, Greco A, Saggese A, Vento M (2015) A method for detecting long term left baggage based on heat map. In: VISAPP 2015 - 10th International Conference on Computer Vision Theory and Applications; VISIGRAPP, Proceedings. SciTePress, pp 385–391

[19] Zahrawi M, Shaalan K (2023) Improving video surveillance systems in banks using deep learning techniques. Sci Rep 13:. https://doi.org/10.1038/s41598-023-35190-9

[20] Palivela LH, Ramachandran S (2018) An enhanced image hashing to detect unattended objects utilizing binary SVM classification. J Comput Theor Nanosci 15:121–132. https://doi.org/10.1166/jctn.2018.7064

[21] Gurusamy K, Yuvaraj N (2021) Improved Object Detection in Video Surveillance Using Deep Convolutional Neural Network Learning. International Journal for Modern Trends in Science and Technology 7:104–108. https://doi.org/10.46501/IJMTST0711018

[22] Narwal P, Mishra R (2019) Real Time System for Unattended Baggage Detection. J Emerg Technol Innov Res

[23] Mahalingam T, Subramoniam M (2017) A robust single and multiple moving object detection, tracking and classification. Applied Computing and Informatics 17:2–18. https://doi.org/10.1016/j.aci.2018.01.001

[24] Lwin SP, Tun T (2022) DEEP CONVONLUTIONAL NEURAL NETWORK FOR ABANDONED OBJECT DETECTION. www.irjmets.com @International Research Journal of Modernization in Engineering

[25] Chen Z, Ellis T (2014) A self-adaptive Gaussian mixture model. Computer Vision and Image

Understanding 122:35–46. https://doi.org/10.1016/j.cviu.2014.01.004
[26] Terven J, Cordova-Esparza D (2023) A Comprehensive Review of YOLO: From YOLOv1 and Beyond. ACM Comput Surv
[27] PETS2006: Performance Evaluation of Tracking and Surveillance 2006, Bench mark Data. http:// www.cvg.reading.ac.uk/PETS2006/data.html. Accessed 18 Jun 2023
[28] PETS2007: Performance Evaluation of Tracking and Surveillance 2007, Bench mark Data. http:// www.cvg.reading.ac.uk/PETS2007/data.html. Accessed 18 Jul 2023
[29] i-Lids: i-Lids Dataset for AVSS 2007. http://www.eecs.qmul.ac.uk/andrea/avss2007_d.html. Accessed 18 Jul 2023