IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 30 January 2024, accepted 19 February 2024, date of publication 23 February 2024, date of current version 11 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3369233

RESEARCH ARTICLE

Abandoned Object Detection and Classification Using Deep Embedded Vision

ARBAB MUHAMMAD QASIM[1], NAVEED ABBAS[1], AMJID ALI[1], AND BANDAR ALI AL-RAMI AL-GHAMDI[2]

[1]Department of Computer Science, Islamia College University Peshawar, Peshawar, Khyber Pakhtunkhwa 25120, Pakistan
[2]Faculty of Computer Studies, Arab Open University, Riyadh 11681, Saudi Arabia

Corresponding author: Naveed Abbas (naveed.abbas@icp.edu.pk)

This work was supported by the Faculty of Computer Studies, Arab Open University, Riyadh, Saudi Arabia, under Grant AOURG-2023-010.

ABSTRACT One indispensable element within security systems deployed at public venues such as airports, bus stops, train stations, and marketplaces is video surveillance. The evolution of more robust and efficient automated technological solutions for video surveillance is imperative. In light of the escalating global threat of terrorist attacks in recent years, any unattended object in public areas is treated as potentially suspicious. Ensuring the protection of individuals in these public spaces necessitates the implementation of safety measures. The intricacies of surveillance recordings introduce challenges when it comes to identifying abandoned or removed objects, owing to factors like occlusion, abrupt lighting changes, and other variables. This paper proposes a novel two-stage method for identifying and locating stationary objects in public settings. The first stage uses a sequential model to capture temporal features and detect potential abandoned objects within the monitored area. When the sequential model detects such an object, it triggers a subsequent phase. The second stage uses the YOLOv8l model to precisely locate the detected objects. YOLOv8l is renowned for its ability to accurately pinpoint object locations within the surveillance scene. The proposed method achieves remarkable accuracy rates of 99.20% and 99.70% on combined PETS 2006 and ABODA datasets, respectively, effectively localizing the target object. This achievement not only underscores the model's precision in accurately pinpointing the object's position within the given context but also establishes its superiority over other existing models. By integrating these two stages, our method provides an effective solution for enhancing the detection of abandoned objects in public spaces, contributing to improved security and safety measures.

INDEX TERMS Abandoned object localization, stationary object detection, embedded vision, abandoned object, video-surveillance.

I. INTRODUCTION

In the various applications of computer vision, video surveillance is attracting the attention of researchers, and actively search for detecting and tracking the objects in the videos [1]. In real-time applications, intelligent video surveillance systems are drastically developed to automate surveillance [2]. The smart video surveillance system autonomously identifies specific occurrences like trespassing, lingering, and abandoned objects [3], [4], [5], [6]. Researchers have explored object detection in videos by approaching it as the recognition of objects within each frame, essentially treating each frame as a standalone image [7]. Video surveillance mainly consists of object detection and tracking such as automotive driving, and intelligent robotic technology [8]. In the realm of video surveillance, there is a focus on dynamic environments to track cars and various real-world objects [9]. In computer vision applications, object tracking and object detection in video surveillance progress hand in hand [10]. Object detection involves classifying and locating objects or instances of interest within a suspicious frame, whereas object tracking is the process of recognizing the trajectory across consecutive

The associate editor coordinating the review of this manuscript and approving it for publication was Zhongyi Guo.

frames [11]. Conversely, the static detection and segmentation of objects in videos remain a challenging and actively researched area [10]. Object segmentation and static object analysis involve the identification, tracking, and assessment of an object's presence [12]. The most recent advancement in video surveillance technology allows it to automatically identify abandoned objects in public areas and illegally parked cars in traffic monitoring systems [13]. Detection of an abandoned object in video surveillance is challenging and essential for maintaining safety [14]. Public safety measures include the automated detection of abandoned items, as manually processing such a huge amount of data seems impossible and time-consuming [15]. An abandoned item is an object that has been left behind by its owner and has not been reclaimed within a predetermined period [5]. Mostly existing abandoned object-detecting algorithms utilize background models for extracting foreground information [16]. The background model serves as an effective method for extracting foreground information from images. However, a notable challenge arises over time, wherein the distinction between the foreground and background becomes less discernible. In other words, as the processing continues, the initial clarity between the foreground and background diminishes, as highlighted by the gradual merging of these elements [17]. This phenomenon creates a hindrance for algorithms that heavily rely on foreground information to detect and track target objects [18]. Additionally, these models are susceptible to variations in lighting, which can alter the image's shape and lead to unstable model outputs. These effects can notably elevate the false alarm rate, resulting in the subpar performance of the surveillance system [19]. One of the core elements of these systems is object detection and tracking, which watches the target over time [20]. Furthermore, the use of video surveillance cameras to identify suspicious situations has significantly increased in the past few years [21]. With recent progress in object detection and facial classification, video surveillance systems incorporating both object detection and facial recognition have become more prevalent [22]. Motion detection-based methods are used for abandoned object detection in surveillance systems [23], the method consists of background subtraction techniques, followed by optical flow analysis techniques and temporal differences techniques [23]. Detecting intentionally discarded or abandoned objects within a scene poses a significant challenge for object detection techniques. Unattended object detection, which identifies unattended items in a series of video frames, aims to address this issue [24].

One of the domains that Artificial Neural Networks (ANNs) have been successfully applied to is computer vision, which is the field of study that enables machines to understand and interpret visual information, such as images, and videos. All ANNs possess a shared capability to extract and learn high-level features from visual information, enhancing machines' ability to comprehend and interpret visual data more effectively. Abandoned object detection is a challenging task in which not only machines are required to locate and segment objects in complex scenes but also classify them. A recent study [1] has used ANN for abandoned object detection in outdoor environments, which detects hand luggage using a deep learning-based detection method. However, their method only focuses on the hand luggage object and does not consider other types of objects that may be abandoned, such as backpacks, and boxes.

In this paper, we propose a novel method for abandoned object detection that can handle both suspicious and non-suspicious scenes. Our method consists scene classification module (SCM) and an object detection module (ODM) which will be discussed in more detail in the methodology section. Our main contributions are as follows:

We develop a novel two-stage method for abandoned object detection that can adapt to different types of scenes and reduce false positives.

Our proposed model achieves an exceptional level of precision, with an impressive accuracy rate of 99.20% for the classifier and 99.70% for the localizer, setting a new standard in abandoned object detection and surpassing existing methods.

Our proposed method outperforms all other existing methods on two public datasets, ABODA and PETS 2006.

The paper is structured to provide a clear progression of the research. It starts with a review of related work, offering context and identifying research gaps. Following this, the proposed methodology introduces the innovative two-stage model for abandoned object detection, addressing surveillance system challenges. The experimental evaluation section is divided into three sub-sections. First, the ''Experimental Setup'' outlines the tools and data sources used. Then, the ''Experimental Results'' demonstrate the model's performance with thorough analysis and evidence of its superiority over existing approaches. This structure ensures a logical flow from the background research to the proposed solution and empirical support for the research's contributions.

II. RELATED WORK

Several methods were used in this area, including object tracking, object identification, and object classification. Diverse techniques were applied to delineate the background and foreground regions of stationary objects over a certain duration. Many prior investigations on abandoned object detection have focused on the analysis of foreground information derived from one or multiple background models. This analysis is conducted to discern the differentiation between stationary and dynamically moving objects. Subsequently, the stationary objects are tracked over a specified duration to ascertain whether they exhibit characteristics indicative of abandonment. In their work, Fan and Pankanti [25] achieved a reduction in the false alarm rate through the utilization of a single background model and a finite state machine (FSM). They accomplished this by modeling objects that exhibit temporary static behavior, such as a car that briefly halts and subsequently resumes movement. Their investigation also encompassed the concept of ''healed'' objects, referring to

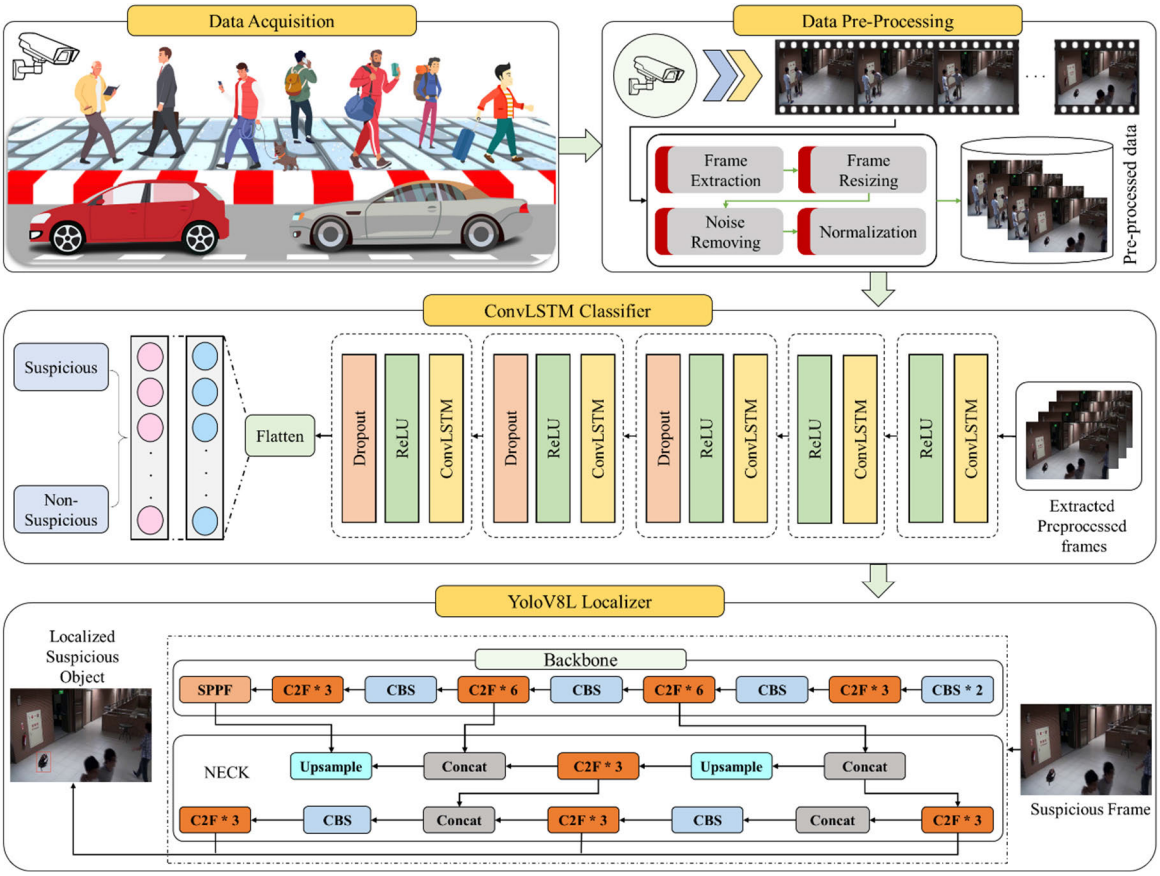

FIGURE 1. The 3-tier main framework of the proposed method: The first tier depicts the data acquisition and data pre-processing stage. The second tier involves ConvLSTM classification of objects, while the third tier integrates the YOLOv8l localizer for object detection.

those objects that have already assimilated into the background. However, it's noteworthy that their study did not address the issue of illumination changes. Omrani et al. [26] introduced a system using stereovision for detecting and tracking objects in maritime environments. An autonomous surface vehicle (ASV) with a stereo camera was used to test the system. In the first stage, the ASV approached stationary objects to identify both static and dynamic things. After that, the ASV tracked a target boat using RTK-GPS to determine its absolute and relative positions. To differentiate between moving and static objects, image processing and stereo vision techniques were applied.

Narwal and Mishra [27] presented a system designed specifically for real-time unattended baggage detection, using frame captures from video surveillance cameras. The system encompasses several phases. It initiates with background subtraction, followed by subsequent steps that leverage the background-subtracted frames for recognizing static regions in the foreground, identifying object types, and validating thresholds. Based on the outcomes of these earlier steps, if an object is determined to be unattended, the system triggers an alarm to alert the relevant authorities. Mahalingam and Subramoniam [28] presented an effective method for tracking and detecting objects in videos that are divided into three separate stages: tracking, evaluating, and detecting. The MoAG model (Mixture of Adaptive Gaussian) was introduced to improve the efficiency of foreground segmentation during the detection phase, which includes noise reduction and foreground segmentation. Din et al. [29] have introduced a framework designed for the detection of abandoned luggage within public areas. The framework begins by utilizing the initial frames to model the background scene. To detect and track moving objects, including both the luggage owner and the luggage itself, the framework employs a model. Significantly, this method sustains its effectiveness in the face of challenges like affine distortion, noise, and occlusion. Moreover, for establishing the history of luggage, the model under consideration records the positional history of mobile objects and utilizes frame differencing as a fundamental technique. Hassan et al. [30] presented a real-time system to detect and track moving objects that become stationary in the restricted zone a pixel classification method based on a segmentation history image is used to identify stationary objects. Park et al. [19] introduced an algorithm specifically crafted to reliably detect abandoned objects, even when faced with changes in illumination. This algorithm demonstrates the ability to quickly adjust to a range of illumination variations.

It includes a presence authentication mechanism relying on the largest contour, allowing accurate tracking of target objects and the identification of abandonment, regardless of

foreground information presence and the impact of illumination variations. Park et al. [31] presented a novel algorithm that accurately differentiates between stolen items, ghost regions, and abandoned objects by fusing conventional image processing methods with artificial intelligence technologies. This method uses two main strategies: segmenting objects using CNN-featured mask regions (Mask R-CNN) to provide comprehensive object mask information and using a dual background model to detect possible stationary objects. Lwin and Tun [32] proposed a method using YOLOv4 to identify abandoned objects in video surveillance. This study developed a framework for detecting abandoned items by expanding the capabilities of YOLOv4. To support this research, a self-assembled dataset was used for training. The neural network was trained to predict six specific objects, including people, backpacks, handbags, books, hats, and backpacks. Wahyono et al. [33] introduced a dual Gaussian mixture model-based cumulative dual foreground difference for stationary object detection. An SVM-based classifier is then integrated to verify the region candidates whether they are vehicles, humans, or other objects. Palivela and Ramachandran [34] introduced a system for abandoned object detection, employing a hash-based approach and incorporating an SVM classifier. Their main area of interest is the identification of abandoned or unattended objects in indoor and outdoor environments. This is accomplished by taking video frames, extracting their features, and using machine learning algorithms to analyze them. Using hash value descriptors from the training phase, the performance of a binary SVM classifier is assessed, and a confusion matrix is used to compare the results with those from other classifiers.

Samaila et al. [35] have introduced a real-time vision-based abandoned object detection system. This system utilizes the Gaussian Mixture Model (GMM) and is capable of detecting objects as small as a teacup. It outperforms the Self-organizing Background Subtraction (SOBS) technique in handling background obstacles. Additionally, this system can classify abandoned items into two categories: non-human and human, a capability not found in other existing techniques. Smitha and Palanisamy [36] proposed a mathematical technique called running average for video sequences of traffic taken from a static camera the background image is a static image that represents the scene without any moving objects. Chen et al. [37] have developed a systematic model pruning approach that evaluates the balance between accuracy and efficiency across diverse structured model pruning techniques and datasets, including CIFAR-10 and ImageNet. They used the VGG-16 model on Tensor Processing Units (TPUs) as a representative example. Additionally, they have introduced a structured model pruning package for TensorFlow2, allowing for in-place modification of models to evaluate their real-world performance. In their work, Palivela and Ramachandran [34] introduced a hash-based approach for abandoned object detection using an SVM classifier. They evaluated its performance in identifying and classifying unattended objects, comparing it to other classifiers through a confusion matrix. Table 1 provides a concise summary of key literature, offering insights into the main findings and methodologies employed by various studies in the field.

TABLE 1. Key findings and methodologies from relevant studies in the field.

Author	Methodology	Main Findings
Wahyono et al. [33]	- Dual background modeling - SVM classifier	- Detects and tracks stationary objects.
Hassan et al. [30]	- Pixel classification - Adaptive tracking	- Real-time system achieving high detection success rate for stationary objects.
Palanisamy et al. [36]	- Background modeling	- Detects stationary foreground objects in traffic videos based on their stationary time.
Park et al [19]	- Dual Background model - Mask R-CNN segmentation	- Presence authentication mechanism for accurate tracking and identification of abandonment.
Lwin et al. [32]	- YOLOv4	- Framework for detecting abandoned objects, extending YOLOv4 capabilities. Trained on a self-assembled dataset predicting six specific object classes.
Palivela et al. [34]	- Hash-based approach - SVM classifier	- Detection of unattended items in public and indoor spaces. Evaluation using hash value descriptors and comparison with other classifiers.
Samaila et al. [35]	- Gaussian Mixture Model (GMM)	- Real-time vision-based abandoned object detection, capable of classifying abandoned items into non-human and human categories.

III. PROPOSED METHOD

The proposed method comprises three components: (1) an enhanced pre-processing step that enhances data quality through the use of advanced and refined techniques; (2) a ConvLSTM layer that captures both temporal and spatial information from the frames; and (3) the state-of-the-art YOLOv8, which identifies stationary abandoned objects within the frames. The complete workflow of the proposed method is depicted in Figure (1). The proposed method for abandoned object detection can handle both suspicious and non-suspicious scenes. Firstly, SCM is based on a sequential

model that can analyze the input image and classify it as a suspicious or non-suspicious scene. Secondly, ODM is based on the YOLOv8 architecture, which can locate and classify various types of objects in the input image. If the scene is classified as suspicious, the ODM will detect the objects that may be abandoned, such as luggage, backpacks, and boxes. Conversely, if the scene is classified as non-suspicious, the object localization will not be attempted by ODM. The ODM can detect various objects that may include abandoned items, such as luggage, backpacks, boxes, etc. Our method can not only detect abandoned objects in real time but also distinguish between different types of objects that may have different levels of risk or importance.

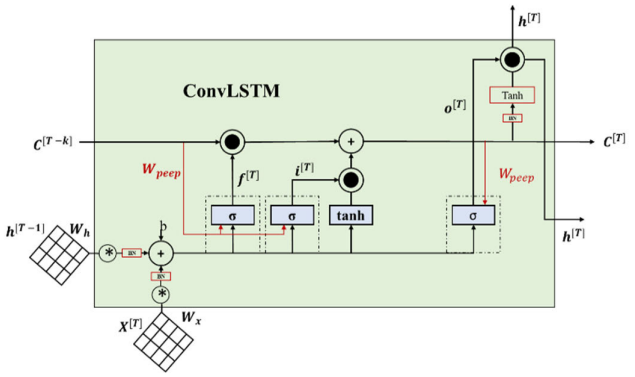

FIGURE 2. ConvLSTM structure.

A. DATASETS DESCRIPTION

The dataset is a critical element for assessing the performance of any system. Evaluating the proposed algorithm using a well-established dataset presents a notable challenge in the domain of visual surveillance systems. In recent years, the availability of standard datasets for abandoned object detection has been limited.

1) PETS 2006 DATASET

PETS 2006 is a publically available dataset for abandoned object detection, which we have used for our experiments. The dataset consists of seven videos with 25 frames per Second (FPS) and a standard resolution of 768 × 576 for evaluating the performance of object detection and tracking systems. The videos show different scenarios of left-luggage events in an outdoor parking lot, captured from multiple cameras and different angles. The dataset videos are annotated with bounding boxes and event types for abandoned object detection, such as left luggage, removed objects, or vehicle movement.

2) ABODA DATASET

ABODA is a public dataset for abandoned object detection. The dataset consists of 11 video sequences captured from different CCTV footage, showing various rea-application scenarios that are challenging for abandoned object detection, such as crowded scenes, lighting changes, night-time detection, and indoor and outdoor environments. The videos are annotated with bounding boxes and ownership information of the abandoned object, indicating whether they belong to a person who is still present in the scene or not.

B. DATA PREPROCESSING

Pre-processing is a vital and indispensable step for achieving better performance in modeling. This process encompasses all the steps that enhance the quality of data, enabling the model to extract and interpret the data effectively. Consequently, the foundation of superior model performance is the pre-processing steps applied to the data before model feeding.

1) FRAME EXTRACTION

Our data preprocessing steps involve frame extraction from and convert it into a batch of frames for models. We extract 15 frames per second from the video and discard irrelevant frames that have no significance for the model. This reduces the computational burden and likelihood of overfitting. We specifically retain the frames with totally distinct features relevant to the model, discarding irrelevant frames.

2) FRAME NORMALIZATION

The pre-processing steps involve the normalization of the characteristics of individual frames, this ensures consistency and comparability. Normalization frames can enhance the performance and reliability of the model. The formula of normalization is represented in equation (1):

$$I_{normalized}(x, y) = \frac{I(x, y) - I_{min}}{I_{max} - I_{min}} \tag{1}$$

Where I (x, y) represents the intensity (brightness) value of the pixel. I_{min} is the minimum intensity value found in the frame. I_{max} is the maximum intensity value found in the frame.

3) FRAME CROPPING

In the pre-processing steps, we cropped the frame to size 512,512 size before feeding it to the model. This allowed us to extract specific regions of interest from a larger frame, which is beneficial for various reasons. Firstly, cropping helps the model to focus on key details, enabling it to concentrate its attention on the most relevant information for the task at hand. Secondly, it reduces the data size, making it more manageable for subsequent processing and lower computational requirements. The mathematical form of the cropping can be represented as in the equation (2).

$$C = I[x_1 : x_2, y_1 : y_2] \tag{2}$$

where I [x_1:x_2, y_1: y_2] represents a subarray or sub region of the original image or frame I. x1 and 1y1 specify the starting coordinates (top-left corner) of the crop region. x_2 and y_2 specify the ending coordinates (bottom-right corner) of the crop region.

4) FRAME AUGMENTATION

In our preprocessing data, we augmented frames by applying various methods such as rotation, scaling, translation,

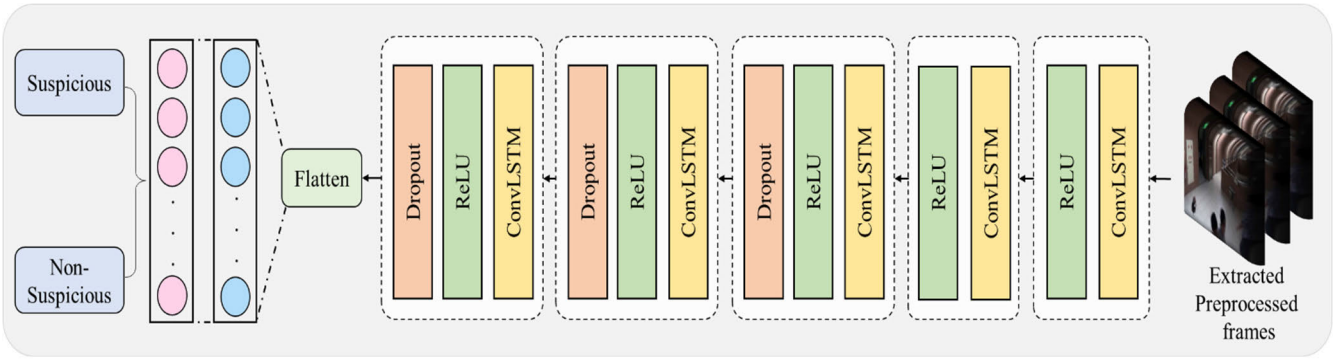

FIGURE 3. Propose ConvLSTM model architecture featuring five ConvLSTM layers, two dense layers with 1000 units each, and final classifier layer with one unit for classifying suspicious and non-suspicious objects.

flipping, and brightness adjustment. The process helps the model become resilient to variation in real-world scenarios such as illumination changes in viewpoint and object orientations. Moreover, frame augmentation can help in overcoming overfitting by expanding the volumetric dataset, thereby improving the model's ability to generalize to unseen data. The general formula of augmentation is given in equation (3).

$$A(I) = T(I, p_1, p_2, \ldots, p_n) \tag{3}$$

where A (I) represents the augmented version of the input data I.T denotes a data transformation function. p_1, $p_2 \ldots$, p_n are parameters that control the specific augmentation techniques applied, such as rotation angles, scaling factors, translation distances, brightness adjustments, etc.

C. CONVLSTM ARCHITECTURE

ConvLSTM is a convolutional neural network with an LSTM network. This is similar to LSTM with the additional functionality of convolutional operation performed on the tensor, the structure of ConvLSTM is illustrated in Figure (2).

The network is suitable for sequential problems like videos where time is dependent. Spatial feature extraction in the model is accomplished through the utilization of convolutional layers. These layers apply filters to individual frames, allowing the capture of crucial spatial information such as object shapes and textures. On the other hand, the ConvLSTM component handles temporal feature extraction by maintaining hidden states, enabling it to grasp the evolution of video frames as they progress over time. This capability enables the model to understand motion, object interactions, and alterations in the video sequence. The key operation of ConvLSTM is given by the equation (4). Where $*$ shows the convolution operator and $\circ$ shows the Hadamard metric.

$$\left\{\begin{array}{l} i_t = \sigma\left(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci}^{\circ} C_{t-1} + b_i\right), \\ f_t = \sigma\left(W_{xi} * X_t + W_{hf} * H_{t-1} + W_{cf}^{\circ} C_{t-1} + b_f\right), \\ C_t = f_t C_{t-1} + i_t \tanh\left(W_{xc} * X_t + W_{hc} * H_{t-1} + b_f\right), \\ O_t = \sigma\left(W_{xo} * X_t + W_{ho} * H_t + W_{co}^{\circ} C_{t-1} + b_o\right), \\ H_t = O_t \tanh\left(C_t\right) \end{array}\right\} \tag{4}$$

In this particular context, the variables are defined as follows: X_t represents the input to the cell, C_t corresponds to the cell's output, and the hidden state of the cell is denoted as H_t. Furthermore, i_t, f_t, and o_t are indicative of the input gate, and the sigmoidal function is represented by σ. The convolution kernels are denoted as W, as previously mentioned in the reference [20].

D. PROPOSED SEQUENTIAL CLASSIFIER MODEL

In the proposed method we have harnessed a sequential model which extracts spatial-temporal features from the video. Recognizing the effectiveness of ConvLSTM, a widely acknowledged method for extracting features from sequential data, we integrate it into our framework to distinguish between static abandoned and non-abandoned objects. The architecture of our ConvLSTM model comprises five layers, with the initial layer serving as the input layer, accommodating sizes 15 × 512×512 × 3. Here '15' Signifies the sequence duration. While the subsequent dimensions correspond to the frame's spatial dimensions and color channels. The first layer incorporates 512, with a stride of 1, and padding set to 'same'. For the subsequent layers, the overall structure remains consistent, except for the number of filters. Specifically, the second employs 256 filters, the Third layer uses 128, the fourth layer employs 64, and the final layer involves 32 filters. After the third, fourth, and last layers, we employed the dropout layers' rate of 0.5 values. Following this, a flattened layer and a flattening layer are applied to transform the 2D data into a 1D format, rendering it suitable for prediction. Subsequently, two dense layers with 1000 neurons at each, are introduced. Finally, the classification layer classifies static objects within the frames Table 2 presents the hyper-parameters employed in the configuration of the proposed classifier model. Throughout the training, the dataset is divided into training, validation, and test sets with a distribution ratio of 70:20:10. Upon generating predictions, if the model's confidence surpasses a predefined threshold, the output is directed to a subsequent static localization model. This model effectively pinpoints the location of static objects within the frames enhancing the precision of the proposed method. Figure (3) illustrates the proposed classifier model architecture.

TABLE 2. Hyper-parameters of the proposed classifier model.

Hyperparameter	Setting
Learning Rate	0.001
Optimizer	Batch Gradient Descent
Output layer (Classifier) activation function	Sigmoid
Input and hidden layer activation function	ReLU
Number of layers	7
ConvLSTM Layers	5
Dense Layers	2
Dropout rate	0.5

E. PROPOSED YOLOv8 LOCALIZER MODEL

YOLOv8 is the most recent version of the YOLO object detection model. While maintaining the same foundational architecture as its predecessors, YOLOv8 incorporates several notable enhancements over prior versions. These include the adoption of an advanced neural network structure that leverages both Feature Pyramid Network (FPN) and Path Aggregation Network (PAN), alongside the introduction of an improved labeling tool designed to streamline the annotation process. The labeling tool encompasses a range of valuable functionalities, including automated labeling, convenient labeling shortcuts, and the ability to customize hotkeys. The synergy of these capabilities significantly simplifies the process of annotating images for model training purposes. Additionally, it's worth noting that YOLOv8 has several versions, such as YOLOv8-S, YOLOv8-M, YOLOv8-L, and YOLOv8-XL. The Feature Pyramid Network (FPN) operates by progressively decreasing the spatial resolution of the input image while concurrently augmenting the number of feature channels. This process leads to the generation of feature maps with the capacity to identify objects at varying scales and resolutions. In contrast, the Path Aggregation Network (PAN) architecture integrates features from diverse network levels using skip connections. Through this approach, the network becomes more adept at capturing features across various scales and resolutions. This capability is of utmost importance for achieving precise object detection, particularly when dealing with objects of diverse sizes and shapes [18].

1) YOLOv8 LOCALIZER MODEL SPECIFICATIONS

In the concluding phase of our proposed methodology, it detects the static abandoned with higher precision. We deliberately selected Yolov8 as the proposed model for static abandoned object detection, under the premise that exhibits the highest probability of detecting stationary objects. Yolov8 is the latest state-of-the-art method known for its higher Mean Average precision (mAPs) and lower inference speed. The model has been meticulously trained on one of the most well known datasets, the COCO dataset. Our research also entailed a comprehensive evaluation of various state-of-the-art object detection models including Faster-RCNN [38], Fast-RCNN, and SSD [39]. Among these all yolov8 series outperformed its counterparts, consistently

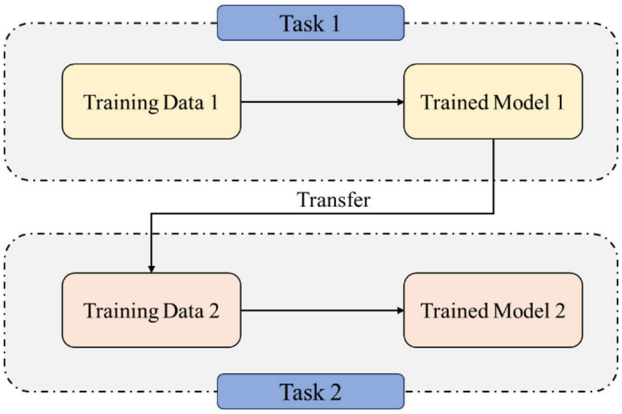

FIGURE 4. Transfer learning mechanism.

achieving either a higher precision or faster inference time. In our proposed method, YOLOv8 utilizes CSPDarknet53 as its backbone. CSPDarknet53 is a deep neural network that excels in extracting features at various resolutions or scales by gradually reducing the size of the input image. The feature maps generated at different resolutions hold essential information about objects present in the image at various scales, offering varying levels of detail and abstraction. Our approach harnesses Yolov8's capability to leverage these diverse feature maps at different features map at different scales to gain insight into the object morphology and texture of objects, thereby enhancing precision in the detection of static abandoned detection objects. Yolov8 backbone consists of four sections, each commencing with a single convolution followed by a c2f module [40]. Notably, our methodology leverages the C2F module introduced by CSP-Darknet53. The module incorporates splits where one branch traverses through a bottleneck module characterized by Two 3×3 convolutions with residual connections. The output of the bottleneck module undergoes further splitting, occurring N times, with N corresponding to the Yolov8 model size. These splits are subsequently concatenated and channeled through a final convolution layer, which serves as the layer responsible for activating the network. This integrated architecture enhances our approach's capability to detect static abandoned objects effectively. The activation map associated with the shallowest c2f module reveals prominent activations corresponding to the object of boards. This module specializes in detecting small objects and identifying their respective classes. Moving to the second activation plays a crucial role in determining the presence of static abandoned objects. As we delved deeper into the model, the third activation started capturing intricate textures associated with static abandoned objects. Ultimately, the model's final C2F module activates to capture exceptionally fine-grained details and outlines within the image under consideration.

2) LOCALIZER OPTIMIZATION USING TRANSFER LEARNING

Yolov8 originally pre-trained on the COCO dataset, encompasses a wide array of object classes. However, our objective

pertains specifically to detecting static objects such as bags. To optimize the model's performance for this specialized task, it must be tailored accordingly. Training for the specific from scratch would be prohibitively expensive and need a more specific dataset. Therefore, we adapted the Transfer leaning [41] technique to leverage the knowledge acquired during pre-training on COCO. This enables us to adapt the model's weights and features to better suit the detection of bags while utilizing a limited dataset. The transfer learning approach is visualized in Figure (4).

F. EVALUATION METRICS

The performance of the proposed methodology is evaluated using specific performance measures including accuracy, precision, and recall.

1) ACCURACY

Accuracy is a metric commonly employed to provide an overall assessment of a model's performance across all classes. This metric is particularly valuable when all classes hold equal significance, quantifying the correctness of predictions by dividing the number of accurate predictions by the total number of predictions, equation (5) demonstrates accuracy.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

2) PRECISION

Precision is determined by dividing the number of Positive samples correctly classified by the total count of samples classified as Positive, whether they were classified correctly or not. It serves as an indicator of the model's accuracy in classifying samples as positive, specifically measuring its ability to correctly identify positive instances. Precision is presented by Equation (6).

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

3) RECALL

Recall is computed by dividing the number of correctly classified Positive samples by the total count of Positive samples. It quantifies the model's capability to identify Positive samples accurately, essentially gauging its sensitivity in detecting such instances. A higher recall value signifies a greater ability to detect Positive samples within the dataset. Recall is illustrated in equation (7).

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

where True Positive (TP): The model correctly identifies something as positive, and it is indeed positive. True Negative (TN): The model correctly recognizes something as negative, and it is indeed negative. False Positive (FP): The model mistakenly identifies something as positive when it's negative. False Negative (FN): The model erroneously identifies something as negative when it's positive.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. EXPERIMENTAL SETUP

In this section, we present a comprehensive overview of the experimental setup of the proposed methodology. The system implementation was executed using Python 3.10.4, PyTorch. 1.12.1, and CUDA 11.7, with both training and inference processes conducted on a powerful 12GB NVIDIA GeForce RTX 3090 GPU. For our primary model, ConvLSTM, we conducted training over a span of 100 epochs. The weights of the model were updated during backward propagation using the Adam optimizer. We configured a batch size of 64 to balance training efficiency and GPU memory usage effectively. Our proposed model tailored specifically for the task of abandoned detection comprises two distinct categories: abandoned and non-abandoned objects. To optimize for the binary classification, we employed binary Cross-Entropy losses during training. It is important to note that yolov8 served as the foundation for our object detection task. The model underwent 100 epochs for training. We fine-tuned it using training learning for our specific use case. This transfer-learning approach allowed us to adapt Yolov8's pre-trained weights to the intricacies of abandoned object localization within the limited dataset. In terms of data preparation, we meticulously annotated the proposed dataset for the model to localize the static object. We annotated 500 images for the model from the different proposed datasets. The annotation process involved defining bounding boxes around the objects of interest, providing essential training data for our models. Our dataset, sourced from diverse environments and scenarios, features variations in lighting conditions, backgrounds, and object poses, reflecting the real-world challenges of abandoned object detection. Ethical considerations guided our data collection, and privacy and bias mitigation measures were considered.

B. RESULTS AND DISCUSSION

In this section, we embark on an empirical evaluation proposed model. This evaluation is structured into three main components. Firstly, in the Architectural Variations Analysis section, we delve into the inner model comparison, where we assess the model's performance using various deep learning architectures for temporal feature extraction. Secondly, is the object detection model evaluation where we conduct a detailed analysis of the object detection model integrated into our proposed framework, aiming to identify strengths and potential areas for improvement Finally, the State-of-the-art model in which, we compare the effectiveness of our proposed model with other state-of-the-art models in the field. This comparative analysis offers valuable insights into the model's performance and its standing in boarder research landscape.

1) ARCHITECTURAL VARIATION ANALYSIS OF CLASSIFIER

To validate the effectiveness of our proposed model in comparison to various sequential models, we conducted a series of empirical experiments. Upon acquiring visual data,

we directed this data sequence to sequential models designed to extract temporal information, a crucial element for precise predictions. This approach capitalizes on the understanding that the quality of features extracted from the frames significantly influences prediction accuracy. During the experiment, our goal was to identify optimal model terms for extracting these enhanced features, which are pivotal for precise temporal predictions. To achieve this, we employed a range of diverse deep learning models for features including the Gated Recurrent Unit (GRU) model, Recurrent Neural Network (RNN), Vanilla Long-Short Term Memory (LSTM) model, and our proposed model. These experiments helped to discern the model that outperforms others in terms of enhancing features critical for precise temporal predictions. We chose the best sequential model based on its superior precision, with the parameter determining our selection. The outcome of the different models along with the proposed is illustrated in Table 3.

TABLE 3. Sequential classifier models performance.

Model	Accuracy (%)	F1-Score (%)
GRU	84.67	80.03
RNN	89.01	88.43
LSTM	95.56	90.40
Bi-Directional LSTM	98.88	95.82
ConvLSTM	**99.20**	**99.05**

The training phase is divided into two segments: training phase for 50 and 100 epochs respectively. In the first, we trained these models for 50 epochs. We compared five sequential models for sequence learning namely (GRU, RNN, LSTM, Bi-Directional LSTM, and ConvLSTM). The vanilla GRU model suffered from overfitting, the RNN from vanishing gradients, and the vanilla LSTM from poor feature extraction. The ConvLSTM outperformed the other models, achieving the highest training and validation accuracy of 80% in the first training phase. In the second phase, we repeated experiments to train these models for 100 epochs. Firstly, the Vanilla GRU model underwent training, during training we kept the sequence of 8 in the data to capture the relation better. The model performance didn't show any significant improvement, rather the model was overfitting on the data. The result was slightly improved but not promising. The model achieved the training and validation accuracy of 55% and 26%. Secondly, in the training phase of RNN, the sequence for the model was 8 to avoid the vanishing gradient problem. During training, we evaluated the model performance was poor on both training and validation data. The model showed a higher under fitting problem. The longer training could not improve the model performance rather than a tinny point improvement. Empirically, the training and validation accuracy of the model was 40% and 30%. Thirdly, the Vanilla LSTM model underwent training with the sequence of 12. Throughout the training,

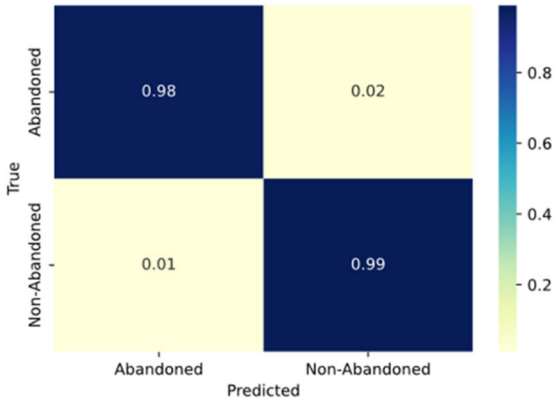

FIGURE 5. Proposed classifier model confusion matrix.

the model performance fluctuated, and the model showed smooth performance. Consequently, in the end, the model could not perform better. Resultantly, the model achieved the training and validation accuracy of 52% and 41%. Finally, we trained the ConvLSTM with 15 sequences in the data. The model showed significant performance throughout the training. The model generalization was remarkably improved as we kept the training. The model outperformed all the remaining models. Empirically we found, the model achieved the training and validation accuracy of 99.20% and 99.50%, respectively, with an impressive F1-score of 99.05%. Based on these evaluated metrics led to the selection of our model for sequential feature extraction.

Figure (4) shows the training and validation accuracy and loss of the proposed model throughout the training. The x-axis of the graphs shows the number of epochs while the y-axis shows the performance of the model. Figure 5(a) indicates at the beginning of training the model started with higher training and validation accuracy. Throughout the training the model showed better generalization however, at the epoch from 20 to 30 model showed some fluctuation but onward there is significant improvement. The graphs showed the model was capturing the relation very smoothly. The fluctuation portion of the model showed the adjustment for learning unseen data, in addition, further training could cause overfitting. Resultantly, the proposed model achieved the highest accuracy among other different sequential models. On the other hand, Figure 5(b) shows the training and validation loss of the proposed model. At the beginning of training, the drastic decrease in the model showed better generalization, throughout the training both losses were smoothly decreasing. The final portion of training showed some overfitting therefore, we stopped the model on 100 epochs. The training and validation loss of the model was 0.01% and 0.02% respectively. Figure (5) shows the confusion matrix of the proposed model, this shows the miss prediction value and the true value for each class. It can be seen from the figure, that the abandoned class has lower accuracy compared to the non-abandoned class, resulting in the model achieving an overall accuracy of 99.20%.

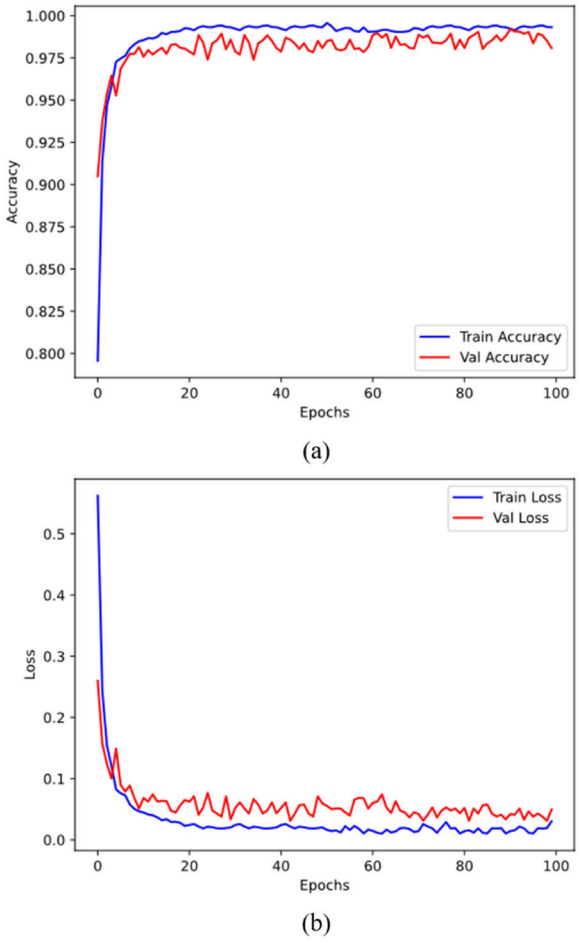

FIGURE 6. Proposed Classifier model learning graphs. (a) Accuracy of the model on the training and validation sets. (b) Loss of the model on training and validation sets.

2) ARCHITECTURAL VARIATION ANALYSIS OF LOCALIZER

In this section, we present our research in three parts. Firstly, we discuss the selection of a robust YOLOv8 variant for abandoned object detection. Secondly, we evaluate our proposed localizer model against various YOLOv8 variants, including YOLOv8-n, YOLOv8-m, YOLOv8-l, YOLOv8-x, and YOLOv8-s. Our experiments involve test images of abandoned objects, and the results are summarized in Table 4, covering precision, recall, F1-score, mAP50, and mAP50-95 metrics. Notably, our proposed YOLOv8l-seg stands out with the highest precision 99.7%, and recall of 99.5% in abandoned object detection. The evaluation highlights the influence of model size and dataset characteristics on performance, with denser models showing fewer promising results. Specifically, YOLOv8n-seg exhibits the lowest precision of 92.4% and recall of 89.1%, YOLOv8x-seg with slightly lower precision of 95.6% and recall of 93.2, YOLOv8m-seg demonstrates precision 98.2% and recall 97.7%, and YOLO8s-seg secures a lower precision score of 94.1% and recall of 96.2%.

The proposed approach for identifying and categorizing stationary objects demonstrates its versatility and

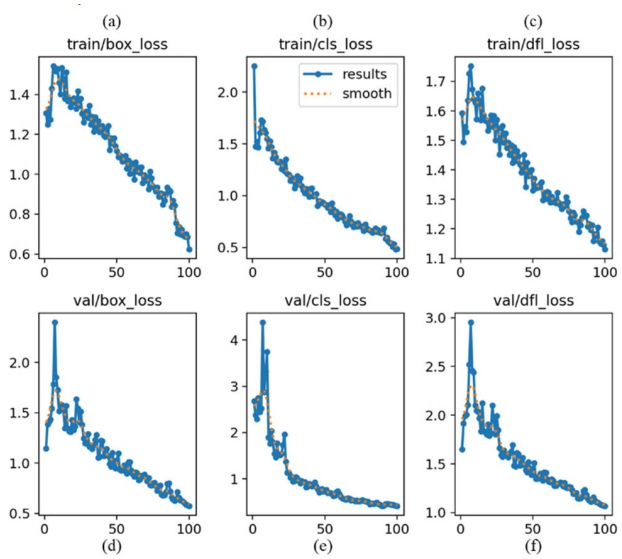

FIGURE 7. Proposed Localizer model training and validation loss graphs. (a) Training box loss. (b) Training classification loss. (c) Training Distribution focal loss. (d) Validation box loss. (e) Validation classification loss. (f) Validation distribution focal loss.

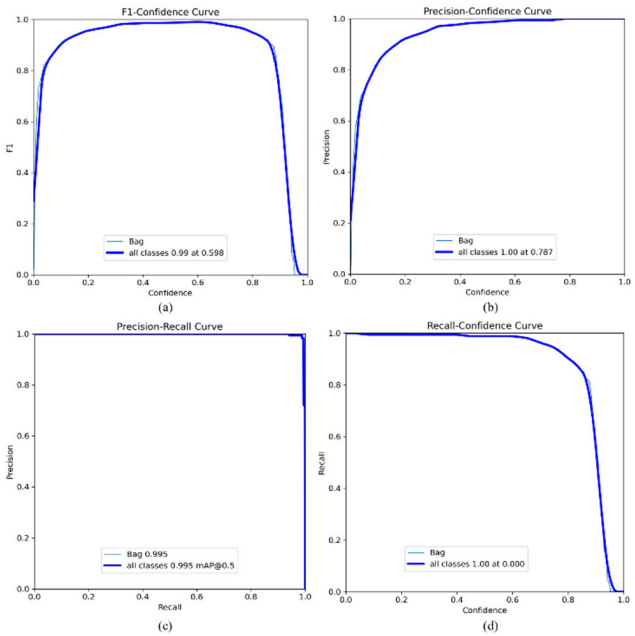

FIGURE 8. Other evaluation graphs of the proposed localizer model. (a) F1-score confidence curve. (b) Precision confidence curve. (c) Precision recall curve. (d) Recall confidence curve.

effectiveness across a spectrum of scenarios, including public transportation hubs, commercial centers, urban streets, public events, smart city infrastructure, residential areas, critical infrastructure sites, and outdoor parks, showcasing its robust applicability in diverse real-world environments, as depicted in Figure (9). This approach significantly enhances precision in localization, a crucial aspect for subsequent classification tasks. Our precision-recall confidence curve achieves an impressive 99.0% mAP for all classes,

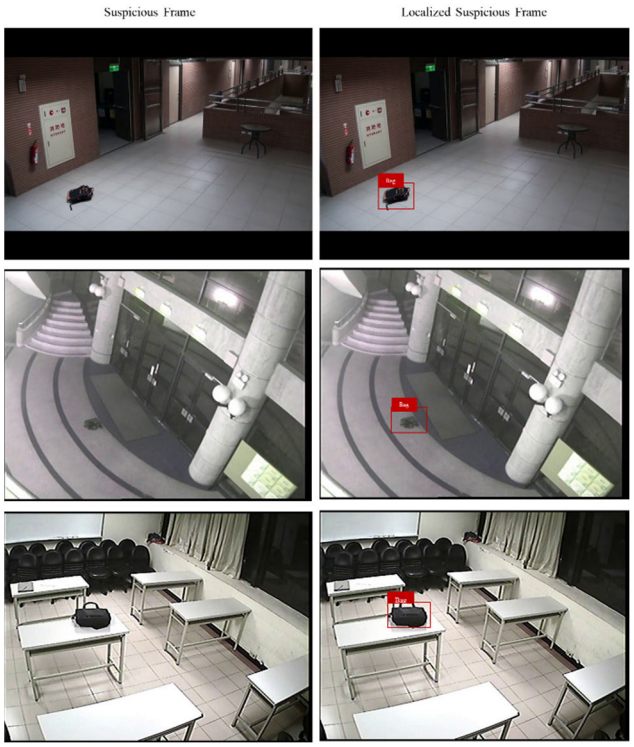

FIGURE 9. Suspicious object localization results on ABODA dataset using YOLOv8l localizer.

illustrated in Figure 8 (c). We conducted a comparative analysis with state-of-the-art models and refine our method for optimal abandoned object detection performance.

The proposed model underwent a comprehensive evaluation during training and validation, assessing key metrics as shown in Figure (6). Notably, the training box loss decreased to 0.06, indicating high confidence in predictions. Validation results, depicted in Figure 7 (d, e, f), revealed significant reductions in box loss (0.5), class loss (0.01), and df1 loss (0.10), showcasing the model's outstanding performance.

Figure (8) presents evaluations using precision-recall curves, precision-confidence curves, recall-confidence curves, and f1-confidence curves for abandoned object detection. The precision-recall curve consistently yields high values of 99.0%, indicating robust performance in segmenting abandoned objects. The precision-confidence curve affirms the model's accurate identification, while the recall-confidence curve demonstrates the correct identification of all positive instances. The F1-confidence curve shows a balanced trade-off between recall and precision scores, with an F1 score of 1.00%, emphasizing the model's superior performance in accurately segmenting various abandoned objects.

3) COMPUTATIONAL COMPLEXITY ANALYSIS

This section provides a detailed analysis of the computational complexity of our proposed method for abandoned object detection. We have considered various factors, including several parameters, memory usage, training time, and inference speed. Table 5 presents a comprehensive breakdown of the computational complexity analysis for classifier models. Inference time for the classifiers was determined based on a sequence length of 10. Notably, our proposed ConvLSTM model exhibits a lightweight architecture, resulting in faster inference time compared to other models in evaluation.

Table 6 presents a complexity analysis of various YOLOv8 models. Our proposed YOLOv8l model stands out with 102.2 million parameters, consuming 408.8 MB of memory requiring 30 minutes for training, and achieving an inference speed of 29 milliseconds.

4) COMPARATIVE ANALYSIS OF THE PROPOSED METHOD WITH STATE-OF-THE-ART MODELS

This section provides a comprehensive breakdown of the performance analysis and a comparative assessment of both

TABLE 4. Yolov8 models performance.

Model Name	Precision	Recall	F1-Score	mAP50	Map50-95
Yolov8n	0.924	0.891	0.843	0.815	0.799
Yolov8m	0.98	0.977	0.923	0.901	0.852
Yolov8x	0.956	0.932	0.931	0.915	0.932
Yolov8s	0.941	0.962	0.900	0.888	0.799
Yolov8l	**0.997**	**0.995**	**0.991**	**0.965**	**0.953**

TABLE 5. Computational complexity analysis of classifier models.

Model	Parameters	Memory Usage	Training Time	Inference Speed (sec)
GRU	243563	18.53 MB	18 min	1.01 sec
RNN	536540	34.00 MB	29 min	1.30 sec
LSTM	634234	39.00 MB	43 min	1.00 sec
Bi-Directional LSTM	300123	20.00 MB	23 min	0.50 sec
ConvLSTM	25839	8.05 MB	10 min	0.30 sec

TABLE 6. Computational complexity analysis of localizer models.

Model	Parameters (million)	Memory Usage	Training Time	Inference Speed (millisecond)
YOLOv8n	4.3M	4.3 MB	15 min	10 ms
YOLOv8s	12.3M	12.3 MB	20 min	15 ms
YOLOv8m	35.6M	35.6 MB	23 min	20 ms
YOLOv8l	102.2M	408.8 MB	30 min	28 ms
YOLOv8x	235.9M	943.6 MB	35 min	31 ms

TABLE 7. Comparison of the proposed localizer with sota.

Author	Accuracy (%)	Precision (%)	Recall (%)
Mahalingam and M. Subramoniam [26]	65.00	99.00	99.00
Lwin and Tun. [19]	84.55	98.80	97.70
Palivela and Ramachandran [31]	95.15	98.80	98.40
Dr. Kiruthig and Yuvaraj [32]	83	79.00	86.00
Teja, YD [10]	99	91.00	88.00
Proposed Method	**99.70**	**99.98**	**99.50**

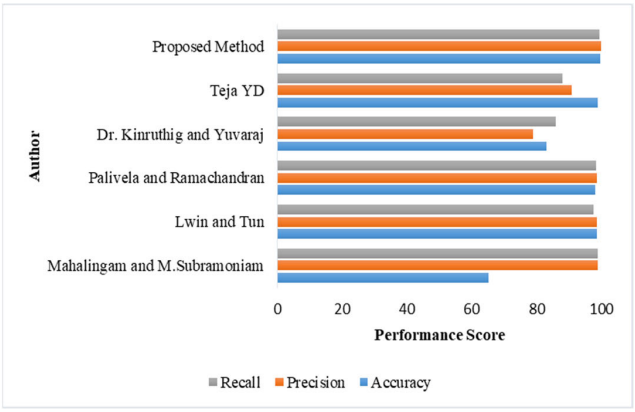

FIGURE 10. Proposed method and SOTA comparison graphical representation.

the established and our proposed methods. An internal localizer comparison, specifically focusing on YOLOv8, has been carried out to enhance the evaluation. Table 2 illustrates the performance of various localizers, where YOLOv8l emerges as the top performer with a precision score of 99.7%, a recall score of 99.5%, and an impressive F1-Score of 99.1%.

For the proposed method and state-of-the-art methods comparison we closely scrutinize key performance metrics, including Accuracy, Precision, and Recall. The proposed model underwent a comprehensive evaluation, comparing it to pre-existing methods. Empirically, the results demonstrated that the proposed model significantly outperformed all existing state-of-the-art models by achieving a substantial increase in accuracy. The detailed comparison findings are outlined in Table 7.

Figure (10) presents a graphical comparison of the performance metrics, including accuracy, precision, and recall between the existing method and the proposed one. Significantly, the proposed method distinctly outperforms the mentioned existing methods by a substantial margin.

V. CONCLUSION

In the realm of video surveillance, a significant yet challenging focus lies on the realm of automatic event detection. Particularly, the detection of abandoned objects (AOD) has garnered substantial attention in recent times, as it plays a critical role in enhancing security in both public and private domains. The global concerns of security and terrorism have escalated to unprecedented levels over the past years, with terrorist attacks claiming innocent lives, often striking crowded locations such as markets, transportation hubs, and airports. To address these pressing security issues effectively, the deployment of automated surveillance technologies in public spaces has become increasingly imperative. In conclusion, our proposed model incorporates a sequential analysis, operating on sequences of 15 frames for initial object detection. This sequential approach facilitates a thorough exploration of temporal patterns and characteristics. Subsequently, the processed data seamlessly advances to the YOLOv8 model, renowned for its exceptional object localization capabilities. By merging the temporal insights derived from the sequential model with YOLOv8's precision in pinpointing object locations, our approach offers a comprehensive and effective solution for object detection and localization tasks. This integration represents a significant stride in enhancing the field of security and surveillance.

REFERENCES

[1] S. Kalli, T. Suresh, A. Prasanth, T. Muthumanickam, and K. Mohanram, "An effective motion object detection using adaptive background modeling mechanism in video surveillance system," *J. Intell. Fuzzy Syst.*, vol. 41, no. 1, pp. 1777–1789, Aug. 2021.

[2] M. Elhoseny, "Multi-object detection and tracking (MODT) machine learning model for real-time video surveillance systems," *Circuits, Syst., Signal Process.*, vol. 39, no. 2, pp. 611–630, Feb. 2020.

[3] N. Bird, S. Atev, N. Caramelli, R. Martin, O. Masoud, and N. Papanikolopoulos, "Real time, online detection of abandoned objects in public areas," in *Proc. IEEE Int. Conf. Robot. Autom.*, Jul. 2006, pp. 3775–3780.

[4] K.-H. Jo, "Cumulative dual foreground differences for illegally parked vehicles detection," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2464–2473, Oct. 2017.

[5] E. Luna, J. San Miguel, D. Ortego, and J. Martínez, "Abandoned object detection in video-surveillance: Survey and comparison," *Sensors*, vol. 18, no. 12, p. 4290, Dec. 2018.

[6] M. A. Mahale, H. Kulkarni, and P. Student, "Survey on abandoned object detection in surveillance video," *Int. J. Eng. Sci. Comput.*, vol. 7, pp. 15595–15599, Jan. 2017.

[7] S. Jha, C. Seo, E. Yang, and G. P. Joshi, "Real time object detection and trackingsystem for video surveillance system," *Multimedia Tools Appl.*, vol. 80, no. 3, pp. 3981–3996, Jan. 2021.

[8] V. Akre, A. Rajan, J. Ahamed, A. A. Amri, and S. A. Daisi, "Smart digital marketing of financial services to millennial generation using emerging technological tools and buyer persona," in *Proc. 6th HCT Inf. Technol. Trends (ITT)*, 2019, pp. 120–125.

[9] B. Qian, Z. Wen, J. Tang, Y. Yuan, A. Y. Zomaya, and R. Ranjan, "Osmotic-Gate: Adaptive edge-based real-time video analytics for the Internet of Things," *IEEE Trans. Comput.*, vol. 72, no. 4, pp. 1178–1193, Apr. 2023.

[10] Y. D. Teja, "Static object detection for video surveillance," *Multimedia Tools Appl.*, vol. 82, no. 14, pp. 21627–21639, Jun. 2023.

[11] S. Khan and L. AlSuwaidan, "Agricultural monitoring system in video surveillance object detection using feature extraction and classification by deep learning techniques," *Comput. Electr. Eng.*, vol. 102, Sep. 2022, Art. no. 108201.

[12] D.-Y. Ge, X.-F. Yao, W.-J. Xiang, and Y.-P. Chen, "Vehicle detection and tracking based on video image processing in intelligent transportation system," *Neural Comput. Appl.*, vol. 35, no. 3, pp. 2197–2209, Jan. 2023.

[13] A. Sathesh and Y. B. Hamdan, "Speedy detection module for abandoned belongings in airport using improved image processing technique," *J. Trends Comput. Sci. Smart Technol.*, vol. 3, no. 4, pp. 251–262, Dec. 2021.

[14] N. Dwivedi, D. K. Singh, and D. S. Kushwaha, "An approach for unattended object detection through contour formation using background subtraction," *Proc. Comput. Sci.*, vol. 171, pp. 1979–1988, Jan. 2020.
[15] B. V. V. Indhuja, V. M. V. Reddy, N. Nikhitha, and P. Pramila, "Suspicious activity detection using LRCN," in *Proc. 5th Int. Conf. Smart Syst. Inventive Technol. (ICSSIT)*, Jan. 2023, pp. 1463–1470.
[16] H. Su, W. Wang, and S. Wang, "A robust all-weather abandoned objects detection algorithm based on dual background and gradient operator," *Multimedia Tools Appl.*, vol. 82, no. 19, pp. 29477–29499, Aug. 2023.
[17] N. Ta, H. Chen, Y. Lyu, X. Wang, Z. Shi, and Z. Liu, "A complementary and contrastive network for stimulus segmentation and generalization," *Image Vis. Comput.*, vol. 135, Jul. 2023, Art. no. 104694.
[18] J. Ju and J. Xing, "Moving object detection based on smoothing three frame difference method fused with RPCA," *Multimedia Tools Appl.*, vol. 78, pp. 29937–29951, Jan. 2019.
[19] H. Park, S. Park, and Y. Joo, "Robust detection of abandoned object for smart video surveillance in illumination changes," *Sensors*, vol. 19, no. 23, p. 5114, Nov. 2019.
[20] H. Lee, J. Yoon, Y. Jeong, and K. Yi, "Moving object detection and tracking based on interaction of static obstacle map and geometric model-free approachfor urban autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 6, pp. 3275–3284, Jun. 2021.
[21] A. Ben Mabrouk and E. Zagrouba, "Abnormal behavior recognition for intelligent video surveillance systems: A review," *Expert Syst. Appl.*, vol. 91, pp. 480–491, Jan. 2018.
[22] V. Tsakanikas and T. Dagiuklas, "Video surveillance systems-current status and future trends," *Comput. Electr. Eng.*, vol. 70, pp. 736–753, Aug. 2018.
[23] S. Ammar, T. Bouwmans, N. Zaghden, and M. Neji, "Moving objects segmentation based on deepsphere in video surveillance," in *Proc. Int. Symp. Vis. Comput.*, Lake Tahoe, NV, USA, 2019, pp. 307–319.
[24] P. Grandhe, P. B. Dhanush, M. Mohammad, A. N. A. A. Lakshmi, and C. V. S. R. Kumar, "An extensive study on unattended object detection in video surveillance," in *Proc. Int. Conf. Intell. Sustain. Syst.*, 2023, pp. 183–193.
[25] Q. Fan and S. Pankanti, "Modeling of temporarily static objects for robust abandoned object detection in urban surveillance," in *Proc. 8th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug. 2011, pp. 36–41.
[26] E. Omrani, H. Mousazadeh, M. Omid, M. T. Masouleh, H. Jafarbiglu, Y. Salmani-Zakaria, A. Makhsoos, F. Monhaseri, and A. Kiapei, "Dynamic and static object detection and tracking in an autonomous surface vehicle," *Ships Offshore Struct.*, vol. 15, no. 7, pp. 711–721, Aug. 2020.
[27] P. Narwal and R. Mishra, "Real time system for unattended Baggag E detection," *Proc. Int. Res. J. Eng. Technol.*, vol. 6, no. 11, p. 3, 2019.
[28] T. Mahalingam and M. Subramoniam, "A robust single and multiple moving object detection, tracking and classification," *Appl. Comput. Informat.*, vol. 17, no. 1, pp. 2–18, Jan. 2021.
[29] M. Din, A. Bashir, A. Basit, and S. Lakho, "Abandoned object detection using frame differencing and background subtraction," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 7, p. 3, 2020.
[30] W. Hassan, P. Birch, B. Mitra, N. Bangalore, R. Young, and C. Chatwin, "Illumination invariant stationary object detection," *IET Comput. Vis.*, vol. 7, no. 1, pp. 1–8, Feb. 2013.
[31] H. Park, S. Park, and Y. Joo, "Detection of abandoned and stolen objects based on dual background model and mask R-CNN," *IEEE Access*, vol. 8, pp. 80010–80019, 2020.
[32] S. P. Lwin and M. T. Tun, "Deep convonlutional neural network for abandoned object detection," *Int. Res. J. Mod. Eng. Technol. Sci.*, vol. 4, pp. 1549–1553, Mar. 2022.
[33] R. Pulungan and K.-H. Jo, "Stationary object detection for vision-based smart monitoring system," in *Proc. Asian Conf. Intell. Inf. Database Syst.*, Dong Hoi City, Vietnam, 2018, pp. 583–593.
[34] L. H. Palivela and S. Ramachandran, "An enhanced image hashing to detect unattended objects utilizing binary SVM classification," *J. Comput. Theor. Nanosci.*, vol. 15, no. 1, pp. 121–132, Jan. 2018.
[35] Y. Samaila, H. Rabiu, and I. Mustapha, "Real-time detection of abandoned object using centroid difference method," *Arid Zone J. Eng., Technol. Environ.*, vol. 16, pp. 48–57, Aug. 2020.
[36] H. Smitha and V. Palanisamy, "Detection of stationary foreground objects in region of interest from traffic video sequences," *Int. J. Comput. Sci. Issues*, vol. 9, p. 194, Dec. 2012.
[37] K. Chen, K. Franko, and R. Sang, "Structured model pruning of convolutional networks on tensor processing units," 2021, *arXiv:2107.04191*.
[38] B. Cheng, Y. Wei, H. Shi, R. Feris, J. Xiong, and T. Huang, "Revisiting RCNN: On awakening the classification power of faster RCNN," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 453–468.
[39] S. Zhai, D. Shang, S. Wang, and S. Dong, "DF-SSD: An improved SSD object detection algorithm based on DenseNet and feature fusion," *IEEE Access*, vol. 8, pp. 24344–24357, 2020.
[40] D. Reis, J. Kupec, J. Hong, and A. Daoudi, "Real-time flying object detection with YOLOv8," 2023, *arXiv:2305.09972*.
[41] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021.

ARBAB MUHAMMAD QASIM received the master's degree in computer science from Iqra National University, Peshawar, Pakistan. He is currently pursuing the Ph.D. degree with the Department of Computer Science, Islamia College University Peshawar, Pakistan.

NAVEED ABBAS received the Ph.D. degree in computer science from Universiti Teknologi Malaysia, in 2016. He is currently an Assistant Professor with the Department of Computer Science, Islamia College University Peshawar, Pakistan.

AMJID ALI received the B.S. degree in computer science from Islamia College University Peshawar, Pakistan, in 2021. He is currently an Assistant Researcher with Islamia College University Peshawar.

BANDAR ALI AL-RAMI AL-GHAMDI received the B.Sc. degree in computer sciences from King Abdulaziz University, Jeddah, Saudi Arabia, in 2003, the M.Sc. degree in information technology from De Montfort University, Leicester, U.K., in 2008, and the Ph.D. degree from Université de Reims Champagne-Ardenne, Reims, France, in 2015. He is currently an Assistant Professor with Arab Open University, Riyadh, Saudi Arabia. His research interests include sensor networks, distributed systems, and eHealth systems.