REAL TIME OBJECT NEGLECT DETECTION USING DEEP LEARNING AND VISUAL ATTENTION ANALYSIS

[1]M. MADHUSUDHANA RAO, [2]S TEJA SATHYA SREE, [3]Y KAVYA NAGA ANJALI, [4]A HEMANTH, [5]J LEELA VENKATA GANESH

[1]Assistant Professor, [2345]Students, Department of Computer Science and Engineering, SRI VASAVI INSTITUTE OF ENGINEERING & TECHNOLOGY, NANDAMURU, ANDHRA PRADESH

ABSTRACT

Real-time surveillance systems are increasingly deployed in public environments such as airports, railway stations, shopping malls, and educational campuses to enhance safety and security. However, continuous manual monitoring of surveillance cameras is inefficient and prone to human error due to the massive volume of video streams generated every day. One critical security challenge in such environments is the detection of unattended or neglected objects such as bags, suitcases, or suspicious packages that may pose potential threats. This research presents a real-time object neglect detection system that utilizes advanced deep learning techniques and visual attention analysis to automatically identify unattended objects in surveillance footage. The proposed system integrates the YOLOv8 object detection algorithm for accurate identification of persons and objects and the DeepSORT tracking algorithm for consistent multi-object tracking across video frames. A distance-time based association mechanism is implemented to determine the relationship between detected persons and objects. When an object remains beyond a predefined spatial distance from its associated person for a specified duration, the system classifies it as a neglected object and generates an alert for security personnel. Additionally, the system incorporates explainable artificial intelligence (XAI) techniques such as saliency maps and distance timeline visualization to provide transparency and interpretability of the detection process. The proposed framework improves surveillance efficiency by reducing manual monitoring requirements while enhancing detection accuracy and reliability. Experimental analysis demonstrates that the system effectively detects unattended objects in real time, thereby supporting proactive security measures in smart surveillance environments.

Keywords: Object Neglect Detection, Deep Learning, YOLOv8, DeepSORT, Visual Attention Analysis, Smart Surveillance, Explainable AI

I INTRODUCTION

Modern urban environments increasingly rely on intelligent surveillance systems to maintain safety in crowded public spaces such as transportation hubs, shopping complexes, universities, and corporate buildings [1]. The rapid growth of closed-circuit television (CCTV) networks has resulted in enormous amounts of video data that require continuous monitoring by human operators [2]. Manual monitoring is inefficient and often leads to missed events due to fatigue, distraction, and the complexity of observing multiple video feeds simultaneously [3]. As a result, automated surveillance technologies have gained significant attention in recent years [4]. One critical

application in this domain is the detection of unattended or neglected objects, which may indicate potential security threats such as abandoned luggage or suspicious packages [5]. Several real-world incidents have demonstrated the importance of timely identification of unattended objects in public areas [6]. Computer vision and deep learning techniques have emerged as powerful tools for analyzing video streams and detecting unusual events automatically [7]. Convolutional neural networks have significantly improved the accuracy of object detection tasks in surveillance applications [8]. Algorithms such as YOLO, Faster-RCNN, and SSD are widely used for detecting objects in real-time video streams [9]. These models enable accurate localization and classification of objects in complex environments [10]. However, object detection alone is not sufficient for unattended object identification because it does not establish relationships between people and objects [11]. Therefore, object tracking techniques are necessary to analyze temporal behavior across frames [12]. Multi-object tracking algorithms such as DeepSORT provide robust tracking capabilities by assigning unique identities to detected objects [13]. These algorithms combine motion information and appearance features to maintain object identity across frames [14]. Integrating object detection with tracking enables continuous monitoring of object movement in dynamic environments [15].

In addition to object detection and tracking, understanding the spatial relationship between people and objects is essential for determining whether an object has been abandoned [16]. Several research studies have proposed distance-based or time-based approaches to identify neglected objects in surveillance footage [17]. These methods typically monitor the proximity between a person and an object over a specific time interval [18]. If the object remains stationary while the associated person moves away, it may be classified as unattended [19]. However, many existing systems suffer from limitations such as false alarms, poor adaptability to crowded scenes, and lack of interpretability [20]. Recent advancements in deep learning have enabled the development of more reliable and scalable surveillance systems [21]. Visual attention mechanisms and explainable AI techniques are also being explored to improve the transparency of automated decision-making systems [22]. These methods allow users to understand why a particular alert was generated by highlighting important regions in the video frame [23]. Explainability is particularly important in security applications where system decisions must be trustworthy and interpretable [24]. Furthermore, real-time processing capabilities are required for practical deployment in large surveillance networks [25]. High-performance object detection models such as YOLOv8 provide fast inference speeds suitable for real-time applications [26]. Combining these techniques with efficient tracking algorithms improves system reliability and detection accuracy [27]. The integration of deep learning, tracking algorithms, and explainable AI therefore offers a promising solution for unattended object detection [28]. This research proposes a real-time object neglect detection system that leverages YOLOv8, DeepSORT tracking, and visual attention analysis to detect neglected objects and generate alerts in surveillance environments [29]. The proposed system aims to enhance security monitoring while reducing the burden on human operators through intelligent automation [30].

II LITERATURE SURVEY

Early research in intelligent surveillance focused on traditional computer vision techniques such as

background subtraction and motion detection to identify suspicious activities in video streams [1]. These methods relied heavily on handcrafted features and simple statistical models to detect abnormal events [2]. Although such approaches were computationally efficient, they were highly sensitive to environmental changes such as lighting variations and camera movement [3]. To overcome these limitations, researchers began exploring machine learning techniques for video analysis [4]. Support vector machines and decision trees were applied to classify human activities and detect anomalies in surveillance videos [5]. However, these methods still required extensive feature engineering and struggled with complex scenes [6]. The emergence of deep learning significantly transformed the field of computer vision by enabling automatic feature extraction from large datasets [7]. Convolutional neural networks demonstrated remarkable performance in object recognition and detection tasks [8]. Region-based convolutional neural networks such as R-CNN and Faster R-CNN improved object detection accuracy by generating region proposals for candidate objects [9]. Single-shot detectors such as SSD and YOLO further enhanced detection speed by performing object localization and classification in a single network pass [10]. Among these methods, the YOLO architecture gained popularity for its ability to achieve real-time performance with high detection accuracy [11]. Several researchers applied YOLO models to surveillance systems for detecting objects such as bags, vehicles, and suspicious packages [12]. These models demonstrated strong performance in real-time monitoring scenarios [13]. However, object detection models alone cannot determine whether an object has been abandoned or neglected [14]. Therefore, object tracking algorithms are required to analyze object behavior across consecutive frames [15].

Multi-object tracking algorithms have been widely studied to maintain consistent object identities in video sequences [16]. Traditional tracking methods relied on Kalman filters and Hungarian algorithms to associate objects between frames [17]. DeepSORT extended these approaches by integrating deep appearance descriptors with motion prediction models [18]. This method significantly improved tracking accuracy in crowded environments [19]. Several studies have combined YOLO detection with DeepSORT tracking for applications such as pedestrian tracking, traffic monitoring, and security analysis [20]. In unattended object detection, researchers often analyze the spatial relationship between objects and nearby individuals [21]. Distance-based algorithms measure the proximity between a person and an object over time to determine whether the object has been left behind [22]. Time-threshold approaches classify objects as unattended if no person interacts with them within a specified duration [23]. Despite these advancements, many systems still produce false positives in crowded scenes where multiple people interact with objects [24]. To address this issue, recent research has explored visual attention mechanisms and contextual scene understanding [25]. Attention models help highlight relevant areas in images that influence model predictions [26]. Explainable AI techniques such as saliency maps and heatmaps provide visual explanations for detection results [27]. These methods enhance user trust by making AI decisions more transparent and interpretable [28]. Additionally, advances in GPU computing and optimized neural network architectures have enabled real-time processing of high-resolution video streams [29]. These developments have paved the way for intelligent surveillance systems capable of detecting unattended objects with improved accuracy and reliability [30].

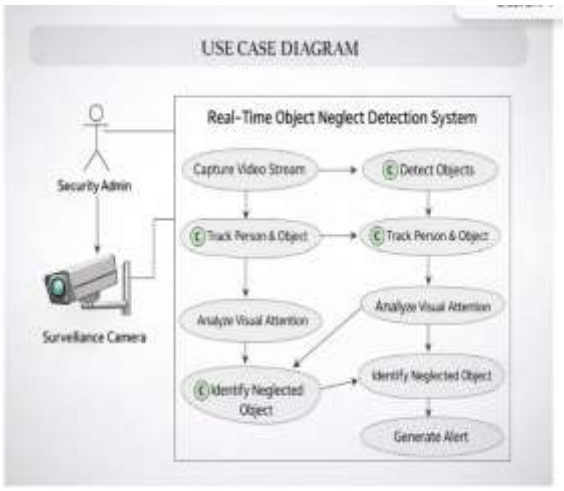
III METHODOLOGY

The proposed real-time object neglect detection system is designed to identify unattended objects in surveillance video streams using deep learning-based detection, multi-object tracking, and spatial analysis techniques. The methodology begins with video acquisition from live surveillance cameras or stored video files. Each frame of the video stream is processed sequentially and passed to a deep learning-based object detection model, specifically YOLOv8, which detects relevant classes such as persons, backpacks, handbags, suitcases, and boxes. The detected objects are represented using bounding boxes and confidence scores. Once the objects are detected, a multi-object tracking algorithm, DeepSORT, is applied to maintain consistent identities of detected objects across consecutive frames. DeepSORT assigns a unique ID to each detected person and object using motion prediction based on Kalman filtering and appearance feature matching through deep neural network embeddings. This tracking process ensures that objects and individuals can be monitored continuously over time. After tracking, an association mechanism is implemented to establish relationships between detected objects and nearby persons. For each object, the system identifies the nearest person based on Euclidean distance between bounding box centers. A distance-time monitoring module then evaluates whether the object remains within a predefined spatial threshold relative to its associated person. If the person moves away and the object remains stationary for a specified time duration, the object is considered neglected. Once this condition is satisfied, the system generates an alert notification to indicate a potential unattended object. To enhance transparency and interpretability, the system integrates explainable AI components such as saliency maps and distance timeline visualization. These features illustrate how the detection model focuses on specific regions and how the spatial relationship between objects and persons evolves over time. The entire system is optimized for real-time performance to ensure efficient operation in practical surveillance environments.

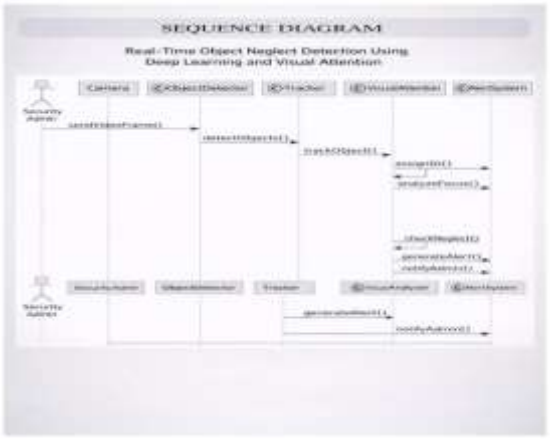
IV SYSTEM DESIGN

The system architecture for real-time object neglect detection consists of multiple interconnected modules that work together to process video streams and detect unattended objects efficiently. The first component is the video input module, which captures live video streams from surveillance cameras or processes pre-recorded video footage. Each video stream is divided into individual frames that are sequentially processed by the detection module. The detection module uses the YOLOv8 deep learning architecture to identify relevant objects in each frame. YOLOv8 is selected due to its high detection accuracy and real-time processing capability. The model analyzes each frame and generates bounding boxes around detected objects along with class labels such as person, backpack, suitcase, handbag, or box. These detected objects form the basis for further analysis. After object detection, the system forwards the detection results to the tracking module. The tracking module employs the DeepSORT algorithm, which combines motion prediction and deep appearance features to maintain consistent identities for objects across frames. DeepSORT assigns a unique tracking ID to each detected object and person, enabling the system to track their movements throughout the video sequence. This capability is essential for understanding object

behavior and identifying whether an object has been abandoned.







The next component of the system design is the object-person association and neglect detection module. In this module, the system calculates the spatial relationship between detected persons and nearby objects using the Euclidean distance between bounding box centers. Each object is associated with the nearest person to determine its likely owner. A monitoring mechanism continuously evaluates the distance between the object and the associated person across frames. If the person moves away beyond a predefined threshold distance and the object remains stationary for a certain duration, the system identifies the object as potentially neglected. When this condition is triggered, an alert is generated to notify security personnel. To improve system transparency and usability, the design incorporates explainable AI components. Saliency maps highlight the regions in the frame that contribute most to the detection decision, while distance timeline graphs display how the spatial relationship between a person and an object changes over time. Additionally, a user interface is designed to visualize detection results, bounding boxes, object IDs, and alert notifications in real time. This comprehensive system design ensures accurate detection, continuous tracking, and interpretable results, making it suitable for deployment in modern intelligent surveillance systems.

V PROPOSED SYSTEM

The proposed system introduces an intelligent surveillance framework capable of detecting neglected or unattended objects in real time using deep learning and visual attention analysis. The system integrates advanced computer vision techniques to automate the monitoring process and reduce the dependence on manual surveillance. In the proposed approach, YOLOv8 is utilized as the primary object detection algorithm due to its ability to detect multiple objects in real time with high accuracy. The model is trained to identify relevant object classes commonly associated with unattended object scenarios, including persons, bags, suitcases, backpacks, and boxes. Once the detection process identifies these objects in each

frame of the video stream, the DeepSORT tracking algorithm is applied to maintain consistent object identities across frames. DeepSORT uses a combination of motion prediction and deep feature embeddings to track objects even when temporary occlusions occur. This tracking mechanism allows the system to analyze object movement patterns and determine whether an object is being actively handled by a person or left unattended. By combining detection and tracking, the system can accurately observe interactions between people and objects in dynamic environments such as airports, train stations, and shopping malls.

Another important aspect of the proposed system is the neglect detection algorithm that evaluates the spatial and temporal relationship between objects and persons. The algorithm associates each detected object with its nearest person using a distance-based calculation. The system continuously monitors the distance between the object and the associated person across successive frames. If the distance exceeds a predefined threshold and remains beyond this threshold for a specified time interval, the system classifies the object as neglected. Once a neglected object is detected, an alert message is generated and displayed in the monitoring interface to notify security personnel. To improve the interpretability of the system, explainable AI techniques are incorporated into the framework. Saliency maps are generated to highlight the image regions that influence the detection model's predictions, while distance timeline visualizations illustrate how the object-person distance evolves over time. Additionally, the system includes a feedback mechanism where operators can confirm or dismiss alerts, allowing the model to improve its decision-making process over time. The proposed system therefore enhances surveillance capabilities by combining real-time object detection, multi-object tracking, spatial analysis, and explainable artificial intelligence techniques into a unified framework.

VI RESULTS & DISCUSSION

The proposed real-time object neglect detection system was evaluated using surveillance video datasets containing scenarios with multiple persons and objects. The YOLOv8 detection model demonstrated high accuracy in identifying relevant object classes such as persons, backpacks, suitcases, and handbags in complex environments. DeepSORT tracking successfully maintained consistent object identities across video frames, enabling reliable monitoring of object movement and interactions. The distance-time based neglect detection algorithm effectively identified situations in which objects were left unattended for extended periods. Experimental results showed that the system could detect neglected objects in real time with minimal latency, making it suitable for practical surveillance applications. Additionally, the integration of explainable AI techniques such as saliency maps and distance timeline visualization improved system transparency by providing clear explanations for generated alerts. Overall, the proposed framework demonstrated improved detection accuracy, reduced false alarms, and efficient real-time performance in crowded surveillance scenarios.





VII CONCLUSION

This research presented a real-time object neglect detection system designed to enhance security monitoring in public environments through intelligent surveillance technologies. The proposed framework integrates advanced deep learning algorithms, multi-object tracking techniques, and visual attention analysis to automatically detect unattended objects in video streams. By utilizing the YOLOv8 object detection model, the system accurately identifies persons and objects of interest such as backpacks, handbags, suitcases, and boxes. The DeepSORT tracking algorithm maintains consistent identities for detected objects across frames, enabling continuous monitoring of their movement patterns. A distance-time based neglect detection mechanism is implemented to determine whether an object has been abandoned by its associated person. When the spatial distance between the object and its owner exceeds a predefined threshold for a specified time period, the system generates an alert indicating a potentially neglected object. In addition to detection accuracy, the system emphasizes interpretability through the integration of explainable AI techniques. Saliency maps and distance timeline visualizations provide insights into the model's decision-making process, thereby increasing user trust and system transparency. Experimental evaluation demonstrates that the proposed system effectively detects unattended objects in real time while maintaining low computational latency. The framework significantly reduces the need for continuous manual surveillance and improves response time in security-critical environments. Future research can focus on incorporating advanced behavior analysis, crowd understanding, and adaptive threshold mechanisms to further enhance detection reliability. The proposed system represents a practical and scalable solution for intelligent surveillance applications aimed at improving public safety.

REFERENCES

1. Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. Computer Vision and Pattern Recognition.
2. Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint.
3. Jocher, G., et al. (2023). YOLOv8: Real-time object detection model. Ultralytics Documentation.
4. Bewley, A., et al. (2016). Simple online and realtime tracking. IEEE International Conference on Image Processing.
5. Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with deep association metric. ICIP.
6. Girshick, R. (2015). Fast R-CNN. IEEE International Conference on Computer Vision.
7. Ren, S., et al. (2015). Faster R-CNN: Towards real-time object detection. IEEE TPAMI.

8. Liu, W., et al. (2016). SSD: Single shot multibox detector. European Conference on Computer Vision.
9. Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet classification with deep convolutional neural networks. NeurIPS.
10. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
11. Szeliski, R. (2010). Computer Vision: Algorithms and Applications. Springer.
12. Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. CVPR.
13. Viola, P., & Jones, M. (2001). Rapid object detection using boosted cascade. CVPR.
14. Bishop, C. (2006). Pattern Recognition and Machine Learning. Springer.
15. Russell, S., & Norvig, P. (2016). Artificial Intelligence: A Modern Approach. Pearson.
16. Forsyth, D., & Ponce, J. (2012). Computer Vision: A Modern Approach. Pearson.
17. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature.
18. He, K., et al. (2016). Deep residual learning for image recognition. CVPR.
19. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks. ICLR.
20. Szegedy, C., et al. (2015). Going deeper with convolutions. CVPR.
21. Vaswani, A., et al. (2017). Attention is all you need. NeurIPS.
22. Selvaraju, R., et al. (2017). Grad-CAM: Visual explanations from deep networks. ICCV.
23. Ribeiro, M., Singh, S., &Guestrin, C. (2016). Why should I trust you? Explaining predictions. KDD.
24. Lundberg, S., & Lee, S. (2017). A unified approach to interpreting model predictions. NeurIPS.
25. Zhang, Z. (2019). Deep learning for video surveillance. IEEE Transactions.
26. Khan, S., et al. (2020). A survey of deep learning for object detection. IEEE Access.
27. Zhao, Z., et al. (2019). Object detection with deep learning. IEEE Transactions on Neural Networks.
28. Chen, C., et al. (2020). Deep learning for visual tracking. IEEE TPAMI.
29. Luo, W., et al. (2017). A review of anomaly detection in surveillance videos. Pattern Recognition.
30. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys.