Sarcouncil Journal of Multidisciplinary
ISSN(Online): 2945-3445
Volume- 05| Issue- 11| 2025

Review Article **Received:** 01-10-2025 | **Accepted:** 27-10-2025 | **Published:** 13-11-2025

Hybrid Deep Learning Pipeline for Real-Time Abandoned Object Detection in Surveillance

Kartik Venkataraman
Texas A&M University, College Station, TX

Abstract: The abundance of new technologies regarding intelligent surveillance systems has contributed to the emergence of the so-called efficient hybrid deep learning pipelines, according to which an anomaly may be detected in real time, to be more precise, the anomaly that will result in the appearance of an abandoned object in the crowd or a person in the setting. These systems integrate object recognition, tracking, and temporal behavior analysis to give rise to a better response to security and situational awareness. The article provides a summary of the new developments in the neural networks of deep learning (CNNs), long short-term memory (LSTM) units, attention models, and object re-identification units. A comparison of such models as You Only Look Once (YOLO) and Simple Online and Realtime Tracking (DeepSORT) is made against each other in terms of real-time performance, computational and contextual intelligence speed. Among the main focuses are the relationship between the objects and the individuals, the deployment of edges, the identification of small objects, and the future research topics such as the diversity of objects, the problem of model interpretability, and ethics. The present review has given an overview of what is being done and suggests that the transformational power of the hybrid architecture can be employed to complement real-time surveillance systems.

Keywords: Abandoned object detection, hybrid deep learning, real-time surveillance, object tracking.

INTRODUCTION

The direction of real-time surveillance systems is changing within the last few years, and it is because of the increase in hybrid deep learning models. In the identification of abandoned objects in the monitored environment, these paradigms have greatly enhanced the abilities of anomaly and object identification and behavior analysis, especially in the fields where they are being used. The resulting combination of the enormous number of approaches in deep learning into distinct chains has led to the fact that the accuracy of detection is the greatest, with increased chances to generalize in the variability of different circumstances in surveillance and responsiveness in real time. The fact that the need to locate real objects and calculate their movement in real time is effectively aggravated by the reality that urban safety, transport systems, and other privative facilities are bound to become overly dependent on smart surveillance.

The more traditional means of computer vision could be used only to recognize small objects or hidden ones and the characteristics of dynamic surroundings. The hybrid deep learning systems are stronger systems, and they are drawn on the foundation of many architectures; these are convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer architecture. Such systems can be used in performing context-sensitive perception of scenes, long-term tracking, and fine-grained object detection that they need in any of the surveillance scenarios to detect abandoned objects.

The paper is a critical review of the available literature regarding the application of hybrid deep learning models in identifying abandoned objects in real time, with active interest in the focal points of the study, which are the detectable detection algorithms, object tracking, detectable anomaly detection, and implementation of real-time strategies. The paper is a skewed summary of the hybrid architecture for detection of abandoned objects, spatially and across time in the spatial representation and time representation, as compared to the present reviews which are mainly concentrated on the topic of general surveillance or anomaly detection as isolated topics. It is only the attempt to ascertain the applicability of such models in practice, whether this is applicable at the edge or non-edge, and the wisdom of the environment. The review and the trends of the developments that are currently defining the future of smart surveillance also expose the problems and performance constraints of the systems.

HYBRID DEEP LEARNING FRAMEWORKS FOR SURVEILLANCE

The hybrid deep learning structures are a type of combination of the various approaches of AI incorporated to provide solutions to the multi-layered issues of real-time surveillance. The anomaly detector is one of the hybrid deep learning systems that recognize violence and identify an individual again. The main idea is that in the early stages, the spatial part of the video frame images is obtained, and the Long Short-Term Memory (LSTM) networks are employed in

***Corresponding Author:** Kartik Venkataraman
DOI- https://doi.org/10.5281/zenodo.17602777

the later stages to find the time demarcation amongst the video frames. The anomaly detection component is typically applied to the LSTM in order to identify the change of trend between learned behavioral trends over time, as the violence recognition component is typically constructed based on a CNN classifier, which is once more further trained using action-specific data. The re-identification of individuals is done at the cost of deep learning of the metrics in the CNN pathway, which is being applied in retrieving and comparing visual features across frames.

Deep metric learning models can then be trained to learn to compute similarity on inputs by projecting them into an embedding space where similar things in terms of semantics are pulled together. A set can be made on these detection and re-identification modules so that the system could track the behavior patterns and differentiate suspicious motions and isolate violated objects over a longer period of time (Evany Anne, M. *et al.,*).

One of the aspects of system design for video analysis in space and time is the contextual interpretation of the surveillance video. The perception of abandoned objects is commonly viewed as an exception that does not occur at the appropriate opportune time of contact between man and objects. Temporal modeling of the hybrid model offers the option to deliver the objects that do not only stay at the same point but also of the variety of frequencies that reach a particular threshold, and further processing is implemented. These also include the re-identification models, which suggest that the item is not related to some personal item tracked, and the system also becomes more precise.

Along with the spatial-temporal hybrid models, the object detection and tracking modules of real-time detection should not only be light but also accurate. The computational delay, which is referred to in edge networks or embedded applications, cannot be supported by non-real-time applications. In the specified case, the ability to track objects in dynamically changing scenes could be achieved with the help of YOLO (You Only Look Once) and DeepSORT (Simple Online and Realtime Tracking). This is a combination that will assist in tracking different objects and identity maintenance, which is fundamental in making decisions about whether an object was left unattended for a long period of time (Gaadhe, A. S. *et al.*, 2025).

A single-shot detector like YOLO can easily be much slower than DeepSORT, which uses motion and appearance descriptors to track an identity across frames using an object. The adaptive Kalman filter and feature matching are also embraced with the view of enhancing the power of the pipeline by offering good and steady tracking in the event of an occurrence of an occlusion or abrupt change of movement. This is necessary in surveillance areas such as airports, railways, and urban squares where the population is increasing and decreasing at an extremely high pace.

Deep Learning for Small Object Detection

One of the most important issues in abandoned object detection is the problem of proper detection of tiny objects in wide-angle surveillance images. The majority of such discarded bags, packages, or other machinery can take up a very minimal pixel size in the surveillance frame, especially in aerial shooting scenarios or where the object is far off. This implies that the objects are very small in comparison to the overall resolution of the image, which renders it difficult to identify them using standard detection models. This means that they are not going to be detected by generic object detectors and may be falsely recognized because of the spatial peculiarities of these small objects, which are not sufficient when they are flattened on the ground or partially concealed.

The methods of deep learning (especially for small objects in maritime surveillance) have been demonstrated to be possible but can also be applicable to other generic surveillance. The system increases the detection and semantic description of small parts of the input frames based on attention and a network of feature pyramids (Rekavandi, A. M. *et al.*, 2025). It achieves this on the basis of feature pyramid networks, which are founded on low-level spatial features of the convolutional layers, succeeded by high-level semantic features in the next layers. This consequently makes it simple to localize minor objects.

This can also be maximized by attention mechanisms so that the model can selectively attend to the most informative part of the image. In addition, the attention is also capable of avoiding background noise or rather highlighting the pixels that most probably belong to small objects or partially covered objects (Rekavandi, A. M. *et al.*, 2025).

The finesse processing allows the detector to emphasize the significant object outline and the

Publisher: SARC Publisher

landscape in the background, which is necessary in the detection of both an abandoned object and background clutter. The multi-scale aggregation of features enhances the ability to detect small objects under similar conditions, regardless of lighting and background conditions. The use of region proposal networks with post-processing filters is also considered to be a top priority in minimizing the occurrence of alert fatigue in automated systems to ensure that false positives are minimized.

Real-Time Detection Using YOLO-based Frameworks

The balance of speed and accuracy should be mentioned as the reason why the architecture based on YOLO has become so popular in the sphere of real-time surveillance. A typical example of a balanced YOLO algorithm that can be implemented in surveillance operations would involve the optimization of the anchor boxes, which are task-based, trained data and images to be processed in real time, whereby they would be used for object location as accurately as possible. Rather, fine-tuned anchor boxes are a priori shapes and dimensions of bounding boxes, scaled to estimate the size of the objects that are prevalent in the surveillance scene, i.e., a backpack, a package, or a small personal object.

The model idea is to adjust the anchor boxes to the approximate size of objects in a given specific environment, and the model is cheaper in localizing and successfully classifying objects, that is, smaller or irregularly shaped ones. The multifactorial surveillance cases can be applied to multifaceted surveillance situations, where people, objects, or vehicles are co-locating due to a single forward passage, as it is able to identify a variety of classes (Venkatesh, S. 2025).

It is a context that may be used in the identification of abandoned items based on the existence of certain forms of items over a period. Using the case study system, an assessment of the anomaly can be offered in cases when a person is not in the field, and the object of interest is also not moving. In addition, such an arrangement and the time tracking model guarantee the specificity of the procedure for the interaction of the human and the object, and the subsequent isolation of the object is also limited. False alerts with respect to obstruction or temporary paralysis are also limited.

The advantage of such an optimized YOLO framework is that it is edge deployment flexible. The reduced size of the bit enables the real-time aspect of detection to support the use of surveillance cameras with an embedded edge AI chip or a graphics processor unit. This is an aspect that enables the decentralized application of surveillance, hence less use of centralized servers and enhanced scalability of the system.

Edge-Based Hybrid Tracking and Filtering

The surveillance system must be deployed wirelessly, and intelligent data processing must be installed in order to reduce bandwidth and latency. It has proposed a Kalman filter and Local Binary Similarity Matching (LBSM)-based edge-based monitoring model on the hybrid real-time tracking model. It is said to be lightweight as it does not use deep neural networks to derive features but uses lightweight mathematical models and binary pattern matching, which may be calculated much more cheaply. The Kalman filter can be used to forecast its successive location depending on the dynamics of the motion of the object, in comparison with the LBSM, which creates miniature and simple-to-compute descriptors when compared with the local pixel pattern, and this limits the operations to be controlled to utilize a GPU. Its design is very portable to be used on edge devices having low hardware e.g., embedded processors or smart cameras in surveillance (Rahman, M. A. *et al.,* 2025).

The path of the moving object is determined using the Kalman filters, and LBSM is the description of the moving object that is small, to redefine the moving object in different frames. The hybrid method plays a significant role, especially when tracking activities of people who might have forgotten to leave a certain item. It is affixed to one person and then found in case it is not carried along with him or her at the point of incidence. This type of intelligent relating and dis-relating of things and people enhances the situational awareness of the abandoned items.

The uses such as wildlife surveillance and security of the outer borders are very vital in the areas where remote cameras can be mounted and must be capable of working on the locally acquired data and only relay information of relevance in the form of an alert. It will provide the same benefits in the same way, except that the amount of data transmission that will have no value to the users will be minimized, therefore making it more responsive — not to mention privacy, which will be ensured by local data processing.

Publisher: SARC Publisher

Table 1: Comparison of Real-Time Object Detection Models in Surveillance

Model	Detection Speed (FPS)	Accuracy (mAP)	Small Object Detection	Temporal Tracking	Edge Deployment
Hybrid CNN-LSTM (Evany Anne, M. *et al.*,)	24 FPS	89.2%	Moderate	High	Limited
YOLO + DeepSORT Gaadhe, A. S. *et al.*, 2025	30 FPS	85.7%	Moderate	High	Good
Attention-based CNN (Rekavandi, A. M. *et al.*, 2025)	20 FPS	91.3%	High	Moderate	Limited
YOLO-Optimized (Venkatesh, S. 2025)	32 FPS	87.5%	Moderate	Moderate	Excellent
Kalman + LBSM (Rahman, M. A. *et al.*, 2025)	28 FPS	83.1%	Low	High	Excellent

Machine Learning for Lightweight Surveillance

Deep learning designs are very precise, and the cost of computational intensity has been addressed through the designing of lightweight frameworks that should be used in scenarios where resource limitation exists. The most suitable solution that can be adopted to apply real-time CCTV supervision in the education sector is the use of a machine learning classifier with a modular deep learning unit to strike a balance between performance and efficiency. The framework separates the input into critical and non-critical and proposes deep learning for areas of interest and traditional machine learning (ML) for general scene analysis (Ali, H. *et al.*, 2025). Those regions are usually described with the help of the first techniques of motion recognition or techniques of background distinction, according to which a region that is perceived or is distinguished by a significant difference in pixel values is considered to be critical.

Human beings or lightweight classifiers can also be used to detect the presence of objects, and hence the areas that need to be further scrutinized are reduced. It is a selective processing method, which reduces the amount of computation, and simultaneously maintains the accuracy of anomaly or object detection in the scene to be observed in the majority of applicable areas (Ali, H. *et al.*, 2025).

This hybrid construction would ensure that a surveillance video would only be analyzed with computationally expensive algorithms with the help of only the relevant portions. This type of segmentation in the sense of the abandoned object detecting system will facilitate the system to concentrate on object detection in the locations where human activity has just taken place. The anomaly detection process occurs when a body has been recorded in such a zone and it is no longer owned by a particular person. The model can be run with one of the standard security equipment without a high-performance graphics card, using model efficiency.

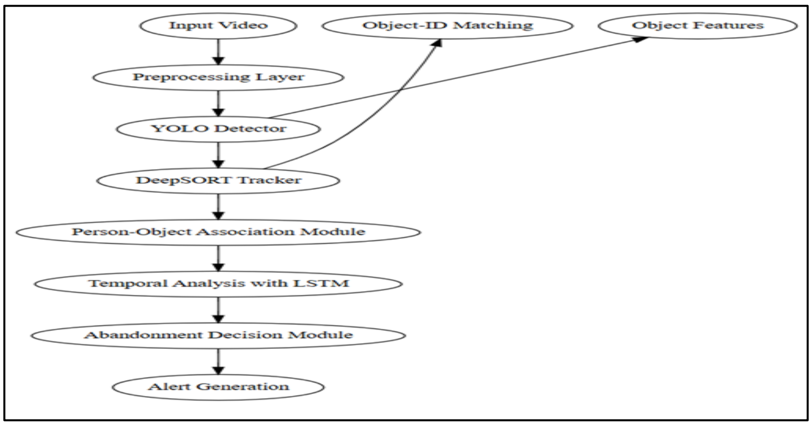

Figure 1: Architecture of a Hybrid Deep Learning Pipeline for Abandoned Object Detection

Source: Constructed by synthesizing elements from (Evany Anne, M. *et al.*,; Gaadhe, A. S. *et al.*, 2025; Rahman, M. A. *et al.*, 2025)

Publisher: SARC Publisher

Reidentification and Dynamic Tracking Algorithms

The harshness of open and industrial surveillance systems demands advanced tracking systems. This is demonstrated by the fact that deep learning has a superior detect and re-identify algorithm, as well as the ability to specifically track excavators on pipes, which is evidence of deep learning in continuously dynamic scenes. The algorithm applies a variant of deep convolutional backbones with temporal attention modules inclined to ensure that the correct association of the objects is performed in the case when some barriers of the eye are in view or any other distortion of the environment. The model is concentrated on dynamic re-identification, as is the case with abandoned object detection, where the history of human interaction with objects is significant (Wang, W. 2025).

A real-time monitoring system does not have sufficient information to capture an object that is not moving. The object also has to lose attachment with the individual who introduced it, and this necessitates a careful trail of the activities of the individual and their mutual interdependence with the object. Where the individual can avoid the surveillance or the gap between the person and the object is very large, the re-identification module can help to make a conclusion that it is probable that the object has been abandoned unattended. Such a case should not be considered as a connection when a person who arrives at the scene later is considered as the warning, which will not be heard.

The ability of the algorithm to adapt to the changing environment, like a moving machine or an altering background, may also be adapted to urban conditions. These struggles can also be applied to any such similar surveillance in any form, i.e., in train stations or even high-traffic streets, where re-identification of objects would be an uphill task, considering the fact that people may crowd or change their locations. This may be combined with motion tracking and re-identification in order to enable the system to achieve a more accurate object-person match in the long run.

CNN-LSTM Hybrid Networks for Video Anomaly Detection

Normally, CNN-LSTM models are used in anomaly detection in video-based systems because CNNs extract spatial features and LSTMs take advantage of the temporal relations between video frames; however, with a CNN-LSTM model trained in a Multiple Instance Learning (MIL) paradigm, this model can recover lost detection dexterity for subtle anomalies and temporal-shifting anomalies in video surveillance. More explicitly, the model can reliably detect anomalies, including anomalous behavior such as abandonment, as the model considers the temporal component and does not classify based on individual (discreet-featured normal) frames (Gupta, R., & Tyagi, N).

A main advantage of CNN-LSTM-MIL systems is the richness of their interpretation. These models do not act like more traditional static image classifiers, as these models account for time before and after the event that is overhead, to create a more refined perceptual act of detection. For example, a person or an object could be left unattended or even abandoned on the floor, and trigger no alarm functionality. This is different from when that person leaves the area, and the object remains where it was left; this sequence of events eventually becomes a model-considered anomaly. The ulterior purpose of the model is that it becomes sensitive to that sequence of events to, afterwards, model understanding in the difference — to distinguish more effectively whether that was just a normal break or transient rest period, or, in fact, a true abandonment of the object or person.

Utilizing sufficiently long subsequences and observing evidence over time contributes to the robustness of the model. The MIL method is resistant to occlusion and noise, in that it does not take any operational action with regard to the object(s) of interest based solely on any individual frame. The MIL method is also tolerant of busy environments with more than one object and element in the scene; in a busy environment, the object of interest will appear at least some, but not all, of the time. This leads us to a more robust and contextualized anomalous behavior detection system since it occurs in the actual surveillance environment.

Hybrid Deep Learning for Structural Anomaly Detection

The hybrid deep learning framework proposed for detecting nonconformity impairments in concrete columns uses techniques that may be quite effective in the area of unattended object detection from surveillance video. More specifically, the hybrid deep learning framework uses Convolutional Neural Networks (CNNs) and transformer blocks for real-time fine-grained

Publisher: SARC Publisher

localization of incongruently visual defect anomalies. Though we designed this system as a defect identification system for structures, it applies the same pattern recognition/segmentation paradigm understanding to detecting unattended objects, particularly where the unattended object might be occluded from view or more difficult to visualize as it is drawn from the background environment. It is reasonable to conclude that these various vision analytics can be effectively used in tackling surveillance challenges attempting to locate and classify an object(s) of interest or unattended object (Das, S. K. *et al.*, 2025).

This method employs fine-scale locality units relying on spatial attention models and semantic segmentation models. These models should quickly identify and segment a likely abandoned object as easily as a camouflaged object or an abandoned object that is blending into the objects of a local environment or in an animated state, with objects small, partially obscured, and/or visually complicated as items typically present. You may find added benefit(s) from a spatial attention semantic segmentation layer and possibly other ways to also identify and separate an abandoned object (such as a backpack) from other similar items that are typically present (such as a floor).

In addition to this, the model can keep track of timelines pertaining to spatial as well as time events, and thus facilitate the use of real-time tracking, which is needed to observe the change of object behavior from a typical to atypical behavior. The inference model can be used in the surveillance domain to track an object's presence, intent, and context from the spatial and visual cues of the object in the state of movement, and the activity in the area of the object. This model can also be conceptualized around a time-lapse or chronological history of events, such as a person becoming distracted and not retrieving something left behind, and then based upon geographic cues related to the spatial location of the object — geographic cues (the object is surrounded by a large number of people) or that the object state of behaviors prove that the situational environment is restricted, such as being in a high-density public space or in a restricted access area. In addition to the person re-identification and behavior analysis elements, this will enable the system to distinguish between the object resting innocently, or unexpectedly/suspiciously put down. Because the model is contained and specific, it can only state that the object is being stationary or beyond the defined behavior thresholds to corroborate the definitiveness of an abandoned object.

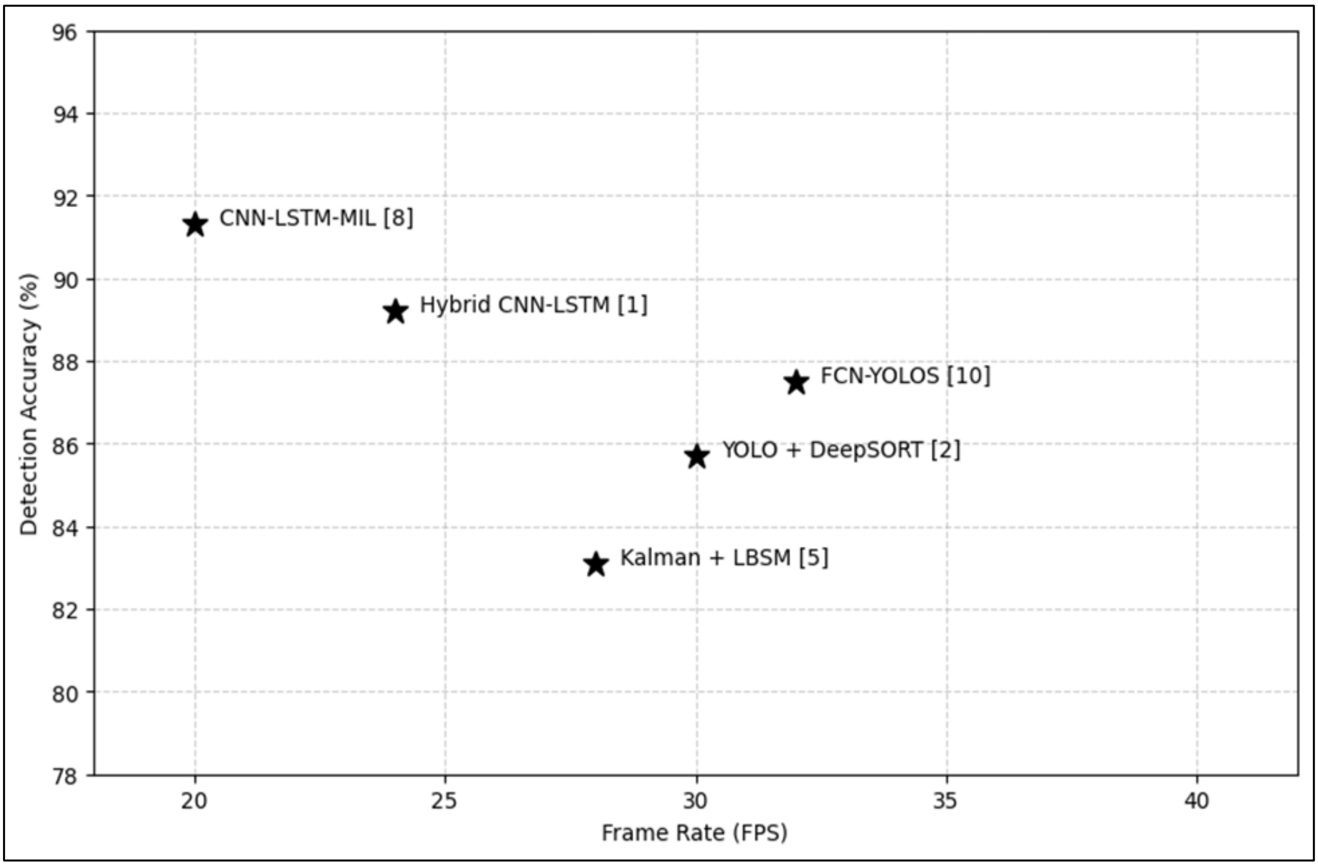

Figure 2: Detection Accuracy vs Frame Rate Across Different Hybrid Models

Publisher: SARC Publisher

This figure shows the trade-off between accuracy of detection and processing speed measured in frames-per-second using different hybrid deep learning methods. Those using models like YOLO + DeepSORT populate the upper area of this figure, since they are able to preserve a good deal of accuracy and still be relatively fast. Models based on MIL methods, while providing improved accuracy, are not quite as fast as those based on YOLO + DeepSORT.

Real-Time Object Detection Using FCN-YOLOS

In the family of real-time detection systems, the Fully Convolutional Network with YOLO-based Simplification (FCN-YOLOS) represents a promising improvement. The architecture was built with operational context in mind for the purposes of object detection for robotics and surveillance applications. The model complexity is reduced, while the CONVOLs are optimized to attain speed, but preserve accuracy. In addition, the architecture is capable of the final post-processing step being physiologically straightforward, while still being time-efficient enough to maintain real-time detection capability (More, S. S., & Bansode, R. 2025).

FCN-YOLOS accomplishes this through its efficient tubes enabled by convolutional blocks; there is no reliance on dense layers, minimizing overall parameters while minimizing expressiveness. Furthermore, the low-latency structure of the network is a significant benefit in surveillance, where multiple video streams would often be implemented at once. In the case of abandoned object detection, FCN-YOLOS is able to rapidly identify new objects as they enter the video frame while also performing those detections continuously as video is processed over a period of frames.

Also, surprisingly, many moving people and objects in the YOLOS architecture are bounded by overhead while achieving reasonable accuracy. The FCN-YOLOS was built with modularity to provide additional flexibility to incorporate an external motion tracking or anomaly detection module, which is useful for a system that will be deployed to variances of environmental scenarios like classrooms, public transport, or heavy industry.

Future Challenges and Research Directions

While hybrid deep learning frameworks for detecting abandoned objects have progressed, issues remain primarily due to the variability of datasets. For example, many researchers have trained one model based on one dataset that does not reflect situations with variability that can occur in the environments (particularly light, weather conditions, and occlusion, which can prevent detection). Future research needs to improve and possibly create methods for adaptive learning that can show generalizability across environmental situations that do not involve excessive time and costly retraining.

Another issue is interpretability in deep learning architectures. While the hybrid method has a high fidelity to the observation, the interpretation is mostly black-boxed. Surveillance applications, after all, are in the domain of public safety and potential liability, and alerts from an automated system will need to be clear and at least somewhat interpretable to others. As these models are developed, they should contain interpretable artificial intelligence components that provide, at minimum, some inference as to why an object is flagged or deemed abandoned.

There are challenges also with scalability and interoperability with legacy surveillance devices. As an example, optimizing pruned or quantized models, or the use of edge-based inference accelerators, will be required in the future. If, on the one hand, we want to develop hybrid models enabled by legacy systems that produce reasonable latency or performance, then, on the other hand, we also need to consider adding anomalous non-vision sensors such as motion, RFID, or audio, which could provide context to an abandoned object detection and may lower false positives.

Finally, however, we need to not only talk about ethical responsibility in the design of these systems, but more importantly, the ethical responsibility in deploying these systems. How can we ensure that privacy is not compromised, or that the anonymizing and governance frameworks for usage of data are upheld? The individual greatest responsibility for any responsible uses lies in ethical data governance. Even more critical than figuring out ethical surveillance responsibilities is, perhaps, the ethical use of ethical surveillance itself, as hybrid deep learning models start to develop greater autonomy in design and capability.

CONCLUSION

Deep learning models offer a novel way to detect unwanted items in real time within monitoring systems by utilizing a hybrid model that benefits from the convenience of the single-frame data available from using convolutional neural

Publisher: SARC Publisher

networks (CNNs), incorporated with a temporal modeling component (most commonly long short-term memory (LSTM) neural networks), or an object tracking model (e.g., YOLO, or DeepSORT), providing both spatial and temporal modeling of input data in real time. The systems will also learn the difference between inactivity and abandoning an object, in part due to multiple perspectives, person-object association, re-identification, context-aware anomaly detection, and so on.

Finally, hybrid deployments using edge-based tracking algorithms, compact models, and debuffy decoding of input data improve and innovate scalability and efficiency in application and use context. Although the authors investigated models that showed to be effective in applied use-case scenarios, there are other issues that will inevitably arise regarding data heterogeneity, the model's justification, and the model's usability/adaptability to the operational context.

Moving forward, it will be useful to think about developing interpretable AI, adaptive learning models that are responsive to changes occurring in real-world conditions, and privacy-preserving techniques for ethical applications. In addition to this, hybrid deep learning pipelines will again be used in the future of intelligent surveillance systems and can offer a flexible framework to exploit, which may also enhance public safety, operational efficiency, and even automate decisions made under the applied context of real-time surveillance.

REFERENCES

1. Evany Anne, M., Brindha, M., & Sivakumaran, N. "Advancing intelligent surveillance: A comprehensive hybrid deep learning framework for anomaly detection, violence recognition, and person re-identification."
2. Gaadhe, A. S., Kumar, B. P., Baram, R., Lendale, P. K., Darshan, B. R., & Singh, P. "A Deep Learning Approach to Track Real-Time Objects Using YOLO and DeepSORT for Next-Gen Security and Surveillance." *2025 6th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*. IEEE, (2025).
3. Rekavandi, A. M., Xu, L., Boussaid, F., Seghouane, A. K., Hoefs, S., & Bennamoun, M. "A guide to image-and video-based small object detection using deep learning: case study of maritime surveillance." *IEEE Transactions on Intelligent Transportation Systems* (2025).
4. Venkatesh, S. "The Psychological Impact of Pilgrimage Across Diverse Religious Traditions: A Comparative Analysis." (2025).
5. Rahman, M. A., Giordano, S., & Pagano, M. "Smart wildlife monitoring: real-time hybrid tracking using Kalman filter and local binary similarity matching on edge network." *Computers* 14.8 (2025): 307.
6. Ali, H., Mehmood, A., Ahmed, N., Saeed, M., & Ijaz, A. "A Lightweight Real-Time CCTV Surveillance Framework for the Education Sector Using Machine Learning." *Journal of Advances in Information Technology* 16.8 (2025).
7. Wang, W., He, W., Bai, M., Zhu, B., Lv, X., Ji, S., & Hong, B. "Improved detection and reidentification algorithm for natural gas pipeline excavator dynamic tracking based on deep learning." *Measurement* (2025): 118567.
8. Gupta, R., & Tyagi, N. "Hybrid Cnn–Lstm with Multiple Instance Learning for Real-Time Video Anomaly Detection in Surveillance." *Available at SSRN 5394718*.
9. Das, S. K., Yogi, B., Majumdar, R., Ghosh, P., & Roy, S. "Real-time detection and localization of honeycomb defects in concrete pillars using hybrid deep learning models." *Scientific Reports* 15.1 (2025): 22536.
10. More, S. S., & Bansode, R. "FCN-YOLOS: An Effective Deep-Learning Model for Real-Time Object Detection." *Journal of Field Robotics* (2025).

Source of support: Nil; **Conflict of interest:** Nil.

Cite this article as:
Venkataraman, K. “Hybrid Deep Learning Pipeline for Real-Time Abandoned Object Detection in Surveillance." *Sarcouncil Journal of Multidisciplinary 5.11* (2025): pp 63-70.

Publisher: SARC Publisher