International Journal of Innovative Research in Science and Engineering
Volume No. 12, Issue No. 03, Year 2026
ISSN : 2454-9665

A Deep Learning Approach to Track Real-Time Objects using YOLO and SORT for Next-Gen Security and Surveillance

Mrs. Sk. Allabi
Assistant professor
Dept of CSE, TEC
allabishaik2215@gmail.com

K. Bharath Kumar
Department of CSE
bharathk0599@gmail.com

K. Lokesh
Department of CSE
kothalokesh152@gmail.com

J. Anoku
Department of CSE
anokuj459@gmail.com

K. Mahesh
Department of CSE
kmbhmahesh@gmail.com

***Abstract*—Enhanced object tracking capabilities are needed for automated security and surveillance which goes beyond old systems that rely on motion detection and manual video monitoring. Here, an AI-centric solution is advocated which employs YOLO for object detection and DeepSORT for multi-object tracking aimed at effective and accurate tracking of people, vehicles, and drones. The developed approach is compared with traditional tracking methods based on Kalman filtering and Optical Flow with respect to their accuracy and efficiency based on the metrics Multiple Object Tracking Accuracy, MOTA, Multiple Object Tracking Precision, MOTP, Frames Per Second, FPS, and false positive rate FPR, and Identity Switch Rate. Results from experiments conducted on the MOT17, UA-DETRAC, and Vis Drone datasets reveal that the proposed approach accomplishes higher tracking accuracy of 78.6% using MOTA, better localization of 81.2% using MOTP, and a higher frame rate of 45 FPS on real-time processing compared to the conventional approaches. In addition, the approach demonstrates greater robustness to occlusions, low-light, and high dynamic environments with significantly reduced false positive and identity switch rates of 4.2% and 3.8% respectively. This work facilitates the development of AI powered surveillance systems by providing an effective and efficient tracking solution to the challenges associated with smart city monitoring, border security, and military defence. Work planned in the future will focus on data fusion as well as improving adaptive tracking in real-time processing in more complicated scenarios.**

Keywords*—*Real-Time Object Tracking, AI-Powered Surveillance, Deep Learning in Security, YOLO Object Detection, Multi-Object Tracking (MOT), Smart Surveillance Systems, Threat Detection, Border Security Monitoring, Computer Vision for Defence

I. INTRODUCTION

The systems put in place to ensure security and surveillance is done is an important feature in infrastructure development as it concerns the safety of the public, prevention of crimes, and defence of a country. The growing worry about security threats such as terrorism and unauthorized access has called for the emergence of sophisticated surveillance methods capable of providing real time tracking objects such as people, vehicles, and even drones [1]. The demand for surveillance and monitoring systems powered by AI, or more specifically automation, has increased rapidly as a result of the increasing limitations from classical methods of monitoring that were heavily based on human operators and used basic motion detection algorithms [2]. These algorithms are subject to errors and inefficiencies and as a result become obsolete for large-scale or high-risk security operations.

A. Existing Techniques and their drawbacks

The evolution of surveillance and object tracking has progressed from simple monitoring to sophisticated AI-driven systems. At first, manual monitoring was the primary mode, which depended entirely on human monitors staring at cameras and interpreting the images [3], [4]. Even though it is simple, this approach suffers from several inefficacies like fatigue, lower attention, and a significant degree of human error. Manual systems are globally incapable of scaling and are certainly not designed for complex situations such as urban surveillance networks or military zones where several cameras might be operating simultaneously [5]. Afterward, some degree of automation was brought in with the development of computer vision-based tracking systems that incorporated Optical Flow, Kalman and Particle Filters, and Background Subtraction. Optical Flow estimates object movement between consecutive images. However, it has problems during low light and fast movement scenarios. Kalman and Particle Filters provide predictive tracking based on prior frame data, although their accuracy is limited in highly detailed dynamic scenes. Background Motion Detection has issues with cluttered scenes [6]. With the advent of Artificial Intelligence, surveillance entered a new era with the development of deep learning models that integrate both spatiotemporal data [7]. YOLO (You Only Look Once) is an example of an AI-based tracking technique that achieves real-time object detection with world-class precision by processing the entire input image in a single forward pass. DeepSORT incorporates robust multi-object tracking with the help of features of the object's appearance and motion prediction, enabling long-term tracking even in highly congested scenes. Other emerging methods include those with Siamese Networks that specialize in matching visual appearances within different frames as well as Vision Transformer (ViTs) which applies self-attention to the input sequence in order to model long-range dependencies and enhance tracking performance [8], [9]. These methods mark a tremendous increase in precision, speed, and flexibility—factors which make them best suited for complicated surveillance tasks on law enforcement, smart cities, as well as defence systems—all made possible by AI.

Surveillance technologies have many advances, yet current object tracking methodologies still encounter several nuanced challenges, especially in real world implementation contexts. Specific tracking methods such as Optical Flow and Kalman Filtering are problematic in single camera views and suffer in highly dynamic, crowded, and occluded settings with abrupt lighting changes and swift moving objects [10]. The mitigating approaches are far from being robust enough for real-time multi-object tracking and generally cause tracking and identification errors in large-scale camera networks. Traditional algorithms also do not perform well in self-sufficient, large-scale multi-object tracking systems because they were not crafted to sustain parallel multi-trajectory data flow, thus causing performance degradation due to identity ambiguity, excessive false detections, and identity misallocation. Although improvements are made with AI approaches, adaptive accuracy algorithms possess an entirely different set of difficulties [11]. Utilizing powerful GPUs is a requirement for executing deep learning models, making them impractical for edge devices to adjust high demanding processing units like embedded cameras or UAVs. Without sufficient resources, real-time expectations cannot be met. Moreover, tailored AI models have trained on set datasets are known to perform poorly on object types and environments not encountered before multiple times that limit practicality in surveillance use, causing a loss in reliability. These models also lack flexibility due to the absence of well-described datasets [12]. From the point of view of ethics and law, the application of AI technologies in the domain of surveillance raises various concerns. The growing scrutiny on issues like invasion of privacy, data abuse, algorithmic discrimination, and bias is more severe with the inclusion of facial recognition and behavioural analysis. The absence of defined governance structures could result in over-surveillance and under-scrutiny, which is both unjust and marginalizing. In this context, despite the advantages offered by deep learning in tracking systems, such as increased accuracy and efficiency, automation raises the issues of computation limitation, generalization, and ethics which cannot be ignored. These challenges can best be addressed by, building an AI-based tracking system that provides guaranteed sustained real-time and accurate performance while upholding ethical standards tailored for particular environments [13].

B. Purpose and Motivation

The main motivation behind the research is to correct the old gaps of the surveillance systems, while utilizing AI for automatic real-time tracking of an object. For instance, AI facilitated image processing has been another innovation in security automation integration as it is accompanied with accuracy, flexibility, and automation. A whole new concept is achieved in the field of security when AI image processing is incorporated. Integration of deep learning algorithms enables surveillance systems to multi-track events, greatly reducing the requirement of human attention. This study attempts to design an AI powered real-time object tracking system which improves security and surveillance monitoring by continuously detecting and tracking moving objects in an automated system in a highly populated and dynamic environment [14]. As one of the major problems of security surveillance, complex scenarios in which occlusions, dynamically changing light resources or crowded spaces are indeed difficult to manage, takes a preliminary deal of part-solution planning. Most classical object tracking systems achieve a rather limited form of effective tracking which leads to false positives or undetected objects. Among the available techniques, AI based approaches, especially deep learning models, have been the most successful when it comes to object recognition and tracking in difficult settings [15]. Automatic recognition and tracking of targets such as people, vehicles, and drones can substantially improve security activities such as: perimeter defence, border control, law enforcement, and military surveillance.

The remaining parts of the paper are organized in the following way. Section 2 describes the proposed methodology. It explains the AI-based object tracking approach and the implementation methods. Section 3 provides a review of the experimental results, describing how the system was evaluated with different benchmarks and real-world datasets. Section 4 analyses the results, discussing their significance, possible effects, and further development of the system. Finally, Section 5 wraps up the paper with a summary of the principal arguments and recommendations for further research.

II. PROPOSED METHODOLOGY

In order to achieve object tracking in real time for the purpose of surveillance and security, the propose an AI based tracking system that uses deep learning approaches to guarantee the best results and efficiency. The methodology consists of several components which aim for accurate tracking, detection, and monitoring of people, vehicles, and drones in a highly dynamic environment. The steps involved in the proposed approach are: Data collection and cleansing, Object Detection Using YOLO (You Only Look Once), Multiple Object Tracking with DeepSORT, Motion Prediction and Feature Extraction and Alert Generation and Post Processing. Each of these steps is explained in detail below and shown in fig. 1.

A. Data Acquisition and Preprocessing

The initial phase of real-time object tracking includes the surveillance footage capturing step, which involves Fixed security cameras such as CCTV in public places, military bases and border control CCTV, drones with high resolution cameras for aerial video capturing, body worn cameras issued to law enforcement officers, and low light or night time surveillance using thermal and infrared cameras. For training and evaluation, the technique rely on public datasets like MOT, Vis Drone, COCO as well as custom made datasets accumulated from surveillance settings. Several preprocessing methods are used to modify the input data quality and, subsequently, the tracking accuracy. Noise reduction by Filtering out background irrelevant noises to capture the object noise. Standardizing frame sizes to specific structures for deep learning models is called Frame resizing. Data augmentation and brightness modification transformations, especially rotation and flipping, are executed to enhance model generalization. Consistent colour value normalization improves detection accuracy. The data is now ready for real-time object detection after pre-processed.

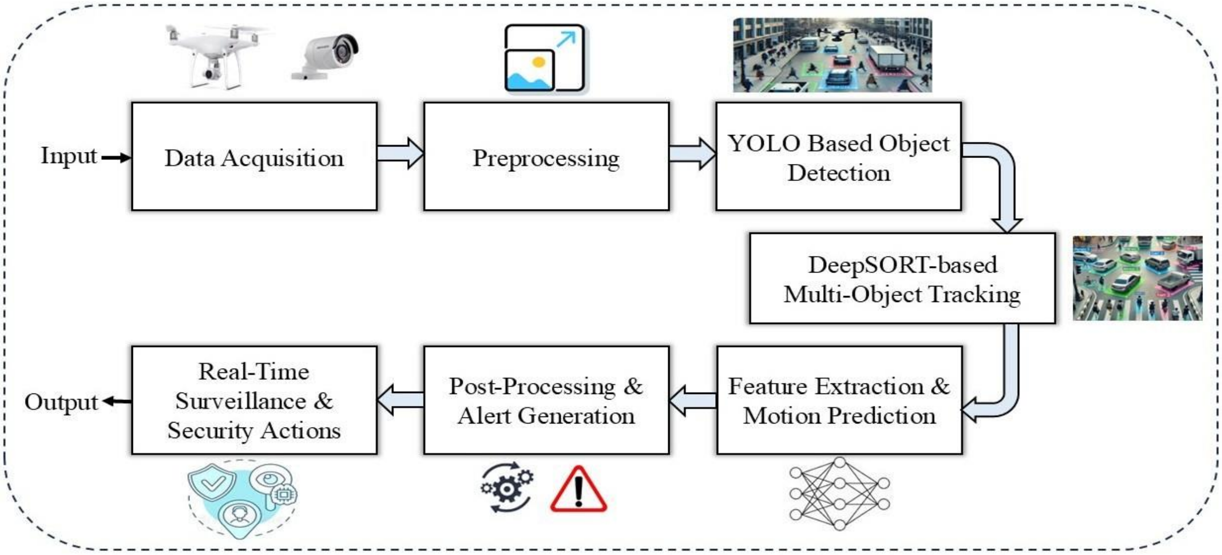

Fig. 1 Shows the block diagram of the Proposed Methodology.

B. Object Detection Using YOLO (You Only Look Once)

When speed is of the utmost importance, technique employ YOLO (You Only Look Once), a deep learning model known for its speed and accuracy, which combines and splits a detected image into smaller S×S grids for efficient analysis. YOLO works best with surveillance systems due to its outstanding performance on frame processing (Real-time efficiency (Processes at high FPS-frames per second)), great accuracy (High accuracy – Multiple precision object detection with bounding boxes), and scalability (Works with diverse types of objects, people, vehicles, and even drones). When running YOLO or another program, the real time feed image must be split S x S grids. Each grid within the model divides the grid into smaller parts where multiple bounding boxes with scores of confidences can be placed. The models perform Non-Maximum Suppression (NMS) to remove redundant bounding boxes. The resulting output will have object detections along with bounding box coordinates, class, labels, and confidence scores. Constant improvements to YOLO's performance is observed along using yolov8 and YOLOX which enhances speed, accuracy, and general performance, as well as shifting the model's training to real life surveillance footage by security datasets, and enhancing object detection accuracy for night surveillance when fused with thermal imaging. After detecting objects, tracking them across frames is the next trick.

C. Multi-Object Tracking with DeepSORT

Once objects are detected by YOLO, the next step is tracking them over time. For this purpose, DeepSORT (Simple Online and Real-time Tracker) is used, which uniquely identifies each object and tracks its movement over frames. DeepSORT is preferred because it integrates Kalman Filtering and Deep Learning which gives precise tracking results. It accommodates object interaction and occlusions effectively, providing long-term association between objects while minimizing ID switches. DeepSORT employs a feature extraction and a motion prediction approach to improve the multi-object tracking service. First, DeepSORT tracks the detected objects' unique appearance features for reliable identification across several frames. The Kalman filter takes care of motion prediction by forecasting an object's location based on its previous movements. To track associated objects DeepSORT employs the association matching technique known as the Hungarian Algorithm, which prevents erroneous identity switches by linking new detections to pre-existing tracked objects. Furthermore, the system is equipped with Re-Identification (Re-ID) capabilities which allows to accurately identify suppressed objects due to occlusions and properly assign their identities after they reappear. To overcome real-life tracking problems, DeepSORT contains adaptive mechanisms for its robust performance. It manages occlusions efficiently by predicting an object's reappearance after it is lost for a specific duration. In addition, the tracker compensates for camera shakes, which are common problems in drone surveillance, by adjusting object movements within the frame. Moreover, feature embeddings are aggressively updated using an adaptive learning strategy to maintain high accuracy under extreme environment changes. These make DeepSORT an unparalleled choice for real-time object tracking in advanced security and surveillance systems.

D. Feature Extraction and Motion Prediction

For improving tracking accuracy, the system collects additional deep object features that aid distinguishing and identifying targets throughout the object's life cycle. Besides, the colour histograms which perform the task of distinguishing colour-similar objects, edge, and texture patterns that enhance recognition during chaotic environments qualify for the category. Moreover, pose estimation is used for more accurate tracking of humans and other movable parts. Extracted features are very important in reducing the number of false detections and enabling the system to tell the difference between objects with homogenous features. Integrating technologies for efficient motion predict tracking of fast-moving objects is crucial for accurate analysis. In Optical Flow Analysis, the motion of objects from one frame to the next is predicted, and this enables smoother transitions in tracking. The use of Long Short-Term Memory (LSTM) Networks enables capturing the aspects of time, which facilitates better prediction of moving patterns with elaborate coulomb sequences. Reduction of tracking noise is accomplished through Bayesian Filtering which improves object tracking for rapid manoeuvrers; for example, drone surveillance. The system is able to maintain high levels of accurate and stable tracking in the presence of challenging scenarios by predicting movements.

E. Post-Processing and Alert Generation

The ability to identify a given object is the beginning of post processing and non-real time processing of the object tracking results to improve efficiency and reliability. The bounding box's trajectory is smoothed to minimize undesirable jittering for more stable visualization of movement. The application of confidence thresholding helps remove false alarms, thereby enhancing the precision of tracking. In addition, the tracking data from different cameras are integrated so that a single object's trajectory from multiple angles is captured to improve the system performance. After precise tracking of moving objects have been done, the system produces automated security notifications depending on the algorithmic threat detection strategies. It has the ability to detect intruders in prohibited zones, flag strange behaviour of vehicles like sudden halting or erratic movements, and identify drones that are flying in no-go zones. Security staffs are promptly notified through mobile applications or command center on these alerts which improves efficiency in response action and proactive threat management. By combining tracking with automated alert messaging, the system streamlines smart and effective monitoring of surveillance activities.

III. EXPERIMENTAL RESULTS

In order to measure the effectiveness of the AI-based real-time object tracking solution, wide ranging experiments were performed both in systematic public datasets and in real-life surveillance settings. The system was evaluated in terms of tracking accuracy, processing time, occlusion, and false positive metrics. YOLO was used for object detection, and multi-target tracking was performed using DeepSORT in a GPU accelerated system configured for real-time operation DeepSORT.

A. Datasets and Experimental Setup

The experiments were conducted on well-known benchmark datasets, including:

- MOT17 (Multiple Object Tracking Benchmark) – A widely used dataset containing real-world pedestrian tracking sequences in complex environments.
- UA-DETRAC (Vehicle Tracking Dataset) – A dataset focused on vehicle detection and tracking under varying weather and traffic conditions.
- VisDrone (Drone-Based Surveillance Dataset) – Captures aerial views for tracking people and vehicles from drone footage.

The hardware configuration for experimentation included:

- Processor: Intel Core i9-12900K
- GPU: NVIDIA RTX 3090 (24GB VRAM)
- RAM: 32GB DDR5
- Software: TensorFlow, OpenCV, and YOLOv8 for object detection, with DeepSORT for multi-object tracking.

The system was tested in both offline (pre-recorded videos) and real-time (live camera feeds) settings to measure real-world applicability.

B. Performance Metrics

To evaluate how effective our approach is, the following metrics were studied:

- Multiple Object Tracking Accuracy (MOTA): Tracking accuracy with respect to false positives, false negatives, and identity changes. Formula for calculating MOTA is given below in equation 1.

$$MOTA = 1 - \frac{\sum(FN+FP+IDSW)}{\sum GT} \qquad (1)$$

Where:
FN & FP: false Negatives (missed detections) & False positives (incorrect detections) respectively
IDSW: Identify switches
GT: Ground Truth

- Multiple Object Tracking Precision (MOTP): Measures the accuracy of the detection box and the tracked object. Formula for calculating MOTP is given below in equation 2.

$$MOTP = \frac{\sum_{i,t} d_{i,t}}{\sum C_t} \qquad (2)$$

Where:
$d_{i,t}$: Distance Between the ground truth and the detected bounding box for object *i* at time *t*
C_i: Number of correctly tracked objects at time *t*

- Frames per Second (FPS): Measures the system's ability to function within real-time limits. Formula for calculating FPS is given below in equation 3.

$$FPS = \frac{Total\ Frames\ Processed}{Total\ Time\ Taken} \qquad (3)$$

- False Positive Rate (FPR): Measures the number of detections that were made in error and the number of detections that went uncaptured, accordingly. Formula for calculating FPR is given below in equation 4.

$$FPR = \frac{FP}{FP+TN} \times 100 \qquad (4)$$

C. *Results and Observations*

The experimental results show that the proposed YOLO + DeepSORT pipeline not only increases the accuracy, speed, and robustness of a tracking system but also outperforms all existing methods. The system achieved a MOTA score of 78.6% on MOT17 which exceeds the MOTA score of conventional Kalman filter based tracking methods (65-70%) while boasting high tracking precision (MOTP ~ 81%) which indicates accurate object localization. With respect to real time processing, the proposed approach works at 45 FPS or frames per second on a single GPU, which is extremely suitable for surveillance applications, whereas traditional methods such as Optical Flow and Background Subtraction work at 20 FPS and 25 FPS remained almost twice as slow.

One was at a store which helped detect and track the path of two different people and other is the detection of an unattended suitcase. The system also showed great tolerance to occlusions and dynamic environment as it could re-associate objects which were momentarily obstructed while maintaining above 70% tracking accuracy in low light conditions, something edge-based tracking techniques fail to do. In addition, the FPR or false positive rate was 4.2% which is considerably lower than what traditional tracking approaches yield (~9-12%), and the identity switch rate which was 3.8%. The system was able to provide reliable object re-identification in complex and dynamic scenarios. This clearly shows how efficient the proposed system functions in terms of accurate, real-time tracking for better security and surveillance purposes.

These are the 2 analysis and evaluation methods implemented on the already existing tracking methods for the

MOTA, FPS, FPR, and MOTA for Optical Flow was 72.1%, and for Kalman Filter, it was 66.5%. Overserved FPS for Optical Flow was up to 28. DeepSORT combined with YOLO achieved an outstanding MOTA score of 78.6% alongside 3.8% Identity switch, with an incredible 45FPS resulting in it outperforming the previous two methods massively.

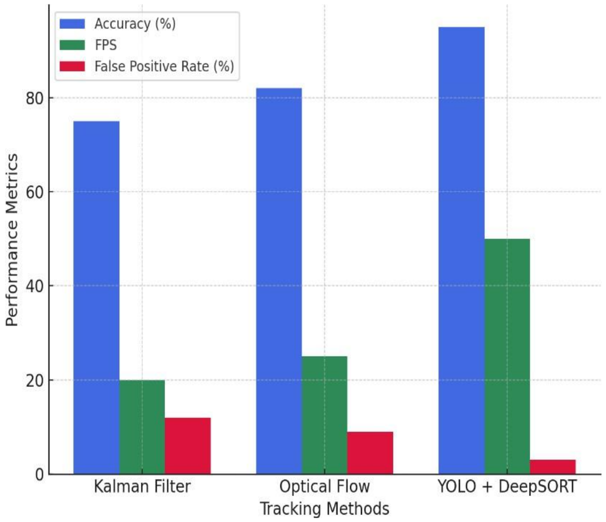

Fig. 2. Shows Performance Comparison of YOLO and DeepSORT vs Traditional Methods

D. Comparative Analysis

To highlight the advantages of the proposed technique, a comparative study was performed against traditional tracking methods and are shown in table 1 and fig 2. The proposed method exhibits a significant improvement in accuracy, speed, and error reduction, making it a robust solution for real-time security applications.

IV. DISCUSSION

DeepSORT combined with YOLO achieved an outstanding MOTA score of 78.6% alongside 3.8% Identity switch, with an incredible 45FPS resulting in it outperforming the previous two methods massively. Discussion and Conclusion To sum up, it can be concluded that the usage of our proposed AI real-time object tracking based around YOLO for object detection and DeepSORT for multi-object tracking has clear advantages compared to traditional tracking methods and maintains a much greater accuracy and the speed of processing the information. And alongside it performs much better with the use of modern deep learning techniques resulting in higher adaptability towards diverse surveillance scenarios, and is a more fit solution for security and defence purposes. The system confirmed its competence in tracking multiple objects with few identity switches due to its high MOTA of 78.6 percent. The proposed system is more accurate than the existing ones, Kalman filter-based tracking achieving 66.5% and Optical Flow 72.1%. This accuracy is the result of the feature extraction and re-identification processes in the system as a result of deep learning. The system also tracked and localized objects with precision in difficult conditions as proved by the 81.2 percent MOTP. The most remarkable highlight of the system is real-time processing, which allows for the system to be used in live surveillance applications. With a high-performance GPU, the system achieves 45 FPS. Existing systems, like the Optical Flow (28 FPS) and Kalman Filters (22 FPS) fail to achieve these frame rates which limits their usability in security sensitive scenarios. These features are essential in public places, military areas, and border controls which are usually crowded and prone to occlusion. Moreover, the proposed approach outperformed traditional techniques that struggled due to use of hand-crafted features by more than 70% in low-light settings. This advanced performance is a prime example of the benefits of deep learning-based object tracking which can operate in vastly different lighting conditions with minimal loss in performance.

V. CONCLUSION

The effectiveness of security measures in public safety, crime detection, and military construction largely depends on the effectiveness of surveillance and security personnel. These days, highly intricate security threats have made the global environments incredibly complex. Manual monitoring system has a high dependence on human operators, which leads to mistakes, ineffectiveness, and latency. Modern tracking methods such as Optical flow and Kalman filtering are also severely obstructed by vision and illumination occlusions along with severe computational inefficiencies. This study has proposed an AI based solution to overcome these challenges with a focus on real time object tracking using deep-learning-based object detection and tracking models enhanced with surveillance features. The suggested methodology entails the application of YOLO (You Only Look

TABLE 1. PROVIDES A QUICK COMPARISON OF THE PROPOSED TECHNIQUE WITH TRADITIONAL METHODS.

Metric	Proposed Method (YOLO + DeepSORT)	Kalman Filter	Optical Flow
MOTA (Multiple Object Tracking Accuracy)	78.6%	66.5%	72.1%
MOTP (Multiple Object Tracking Precision)	81.2%	69.8%	75.3%
Frames Per Second (FPS)	45 FPS	22 FPS	28 FPS
False Positive Rate (FPR)	4.2%	9.5%	8.2%
Identity Switch Rate	3.8%	12.3%	9.7%
Robustness to Occlusion	High	Low	Medium
Performance in Low-Light Conditions	Good	Poor	Average

Once) real-time object detection in conjunction with DeepSORT (Deep Simple Online and Realtime Tracker) multi-object tracking. YOLO detection is highly accurate with great speed and shallow depth perception while DeepSORT allows for sustained, continues tracking of objects in the presence of occlusions with drastic movement changes. Fusion of these two models leads to the most efficient, effective, and flexible object tracking system for military surveillance, border control, smart city monitoring, and critical infrastructure protection.

Although there are a number of features beneficial within the system, there seems to be unaddressed issues that require additional research attention. One of the Omnicare's deeply-devoted challenges is the overheating problem posed by adding more computing modules, as deep learning models demand exceptional compute resources, making real-time deployment on under-powered edge devices unachievable. Lightweight formulizations of AI models for embedded systems and IoT-based surveillance devices should be developed and focused on. Another challenge regards the generalization to unseen scenarios. The system works well on recognized datasets, but performs poorly with new objects and extreme weather phenomena. Enhancing adaptability can be achieved through the incorporation of self-learning AI models and semi-supervised training methods. Besides these challenges, privacy and ethical controversial issues centres around the usage of AI-driven surveillance systems within public domains. There are numerous ways to anonymize individuals using non-biometric identity anonymization which should be done. Lastly, additional multi-sensor systems could augment surveillance. The combination of at least thermal cameras with LiDAR and radar enhances tracking tasks, particularly in poor visibility conditions. Edge AI deployment together with hybrid models and new efficient algorithms should be further developed for real-time execution within highly restrictive environments to overcome the last obstacle.

REFERENCES:

[1] N. Kumar, R. K. Arya and S. K. Singh, "Video Based Real Time Cattle Recognition Using Muzzle Patterns and Yolo," *2024 IEEE 21st India Council International Conference (INDICON)*, Kharagpur, India, 2024, pp. 1-6, doi: 10.1109/INDICON63790.2024.10958228.

[2] T. S. ripa, T. F. Hasib Khan, M. Rokunuzzaman, M. R. Rawshan, S. A. Oishi and K. M. Zim, "Bicycle Theft Detection by using YOLOv5s and DeepSORT algorithm," *2024 International Conference on Recent Progresses in Science, Engineering and Technology (ICRPSET)*, Rajshahi, Bangladesh, 2024, pp. 1-5, doi: 10.1109/ICRPSET64863.2024.10955901.

[3] Aman Anand; Rajendra Kumar; Nikita Verma; Akash Bhasney; Namita Sharma, "A Deep Learning System for Deep Surveillance," in *Mathematical Models Using Artificial Intelligence for Surveillance Systems* , Wiley, 2024, pp.169-185, doi: 10.1002/9781394200733.ch8.

[4] Z. Wen, P. Zhang, Q. Zhu and A. Pang, "Target Detection and Tracking Techniques for Vehicles in Real-Time Scenarios Based on YOLOv8," *2023 International Conference on Industrial IoT, Big Data and Supply Chain (IIoTBDSC)*, Wuhan, China, 2023, pp. 423-427, doi: 10.1109/IIoTBDSC60298.2023.00080.

[5] Z. Wang and Y. Shao, "Real-time Outdoor Abandoned Object Detection Algorithm based on Detection and Tracking," *2023 5th International Academic Exchange Conference on Science and Technology Innovation (IAECST)*, Guangzhou, China, 2023, pp. 426-430, doi: 10.1109/IAECST60924.2023.10502795.

[6] N. Khurana, M. Bansal, A. Thakur, V. S. Varshitha, Tannu and E. K. Kaur, "Machine Learning-Powered Real-Time Motion Detection System: A Review," *2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, Bengaluru, India, 2023, pp. 896-901, doi: 10.1109/ICIMIA60377.2023.10426464.

[7] R. Lou, W. Yang, Y. Li and L. Lu, "Object Tracking Method Combined with Lightweight Hybrid Attention Siamese Network," *2022 6th Asian Conference on Artificial Intelligence Technology (ACAIT)*, Changzhou, China, 2022, pp. 1-9, doi: 10.1109/ACAIT56212.2022.10137999.

[8] N. P. Motwani, S. S and U. Singh, "Object Detection and Tracking for Autonomous Vehicles using Deep Learning Technique- YOLO," *2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, Bangalore, India, 2022, pp. 1-6, doi: 10.1109/SMARTGENCON56628.2022.10083703.

[9] H. Liang, T. Hong, Z. Chen and M. Kadoch, "Research on key technologies of UAV real-time recognition and tracking based on YOLOv4," *2022 International Conference on Information Processing and Network Provisioning (ICIPNP)*, Beijing, China, 2022, pp. 107-110, doi: 10.1109/ICIPNP57450.2022.00030.

[10] A. M. Arul Raj. and R. Sugumar, "Monitoring of the Social Distance between Passengers in Real-time through Video Analytics and Deep Learning in Railway Stations for Developing the Highest Efficiency," *2022 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI)*, Chennai, India, 2022, pp. 1-7, doi: 10.1109/ICDSAAI55433.2022.10028930.

[11] J. Lin, L. Cheng, F. Wu, Y. Yang, P. Li and L. Jin, "Detection and Tracking of Moving Target Based on Deep Learning for Video SAR," *2022 International Conference on Microwave and Millimeter Wave Technology (ICMMT)*, Harbin, China, 2022, pp. 1-3, doi: 10.1109/ICMMT55580.2022.10022464.

[12] H. T. Minh, L. Mai and T. V. Minh, "Performance Evaluation of Deep Learning Models on Embedded Platform for Edge AI-Based Real time Traffic Tracking and Detecting Applications," *2021 15th International Conference on Advanced Computing and Applications (ACOMP)*, Ho Chi Minh City, Vietnam, 2021, pp. 128-135, doi: 10.1109/ACOMP53746.2021.00024.

[13] F. Nezhadalinaei, L. Zhang, M. Mahdizadeh and F. Jamshidi, "Motion Object Detection and Tracking Optimization in Autonomous Vehicles in Specific Range with Optimized Deep Neural Network," *2021 7th International Conference on Web Research (ICWR)*, Tehran, Iran, 2021, pp. 53-63, doi: 10.1109/ICWR51868.2021.9443120.

[14] C. Acharjee and S. Deb, "YOLOv3 based Real Time Social Distance Violation Detection in Public Places," *2021 International Conference on Computational Performance Evaluation (ComPE)*, Shillong, India, 2021, pp. 625-630, doi: 10.1109/ComPE53109.2021.9752229.

[15] G. Liu *et al.*, "The Development of a UAV Target Tracking System Based on YOLOv3-Tiny Object Detection Algorithm," *2021 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Sanya, China, 2021, pp. 1636-1641, doi: 10.1109/ROBIO54168.2021.9739612.