Object Detection and Tracking using YOLOv8 and DeepSORT

Atul Yadav, Pratyush Kumar Chaturvedi, Shallu Rani
Chandigarh University, Mohali, Punjab, India
Corresponding author: Atul Yadav, Email: atulyadav9632@gmail.com

Object detection and tracking are critical capabilities for computer vision systems. This research proposes a method for real-time and recorded video-based multiple object detection and tracking in videos, utilizing cutting-edge computer vision algorithms. YOLO, a high-performance convolutional neural network for object detection, and DeepSORT, an algorithm for separating object instances and matching detections across frames based on motion and appearance, are combined to create an object detection and tracking pipeline. The study's findings show how well YOLO's quick object recognition and DeepSORT's dependable object tracking work together to provide accurate and instantaneous object monitoring. The suggested method has a lot of promise for use in areas like object detection, traffic control, and video surveillance, which will improve automation and situational awareness. The findings provided here pave the way for more investigation and practical use of these methods, providing opportunities for future advancements in the domains of artificial intelligence and computer vision. It could have a significant impact on a variety of businesses that rely on precise object perception and tracking.

Keywords: Object Detection, Object Tracking, You Only Look Once (YOLO), Simple Online and Realtime Tracking (SORT), DeepSORT (Simple Online and Realtime Tracking with a deep association metric).

in: *Advancements in Communication and Systems*. Ed. by Ashish Kumar Tripathi and Vivek Shrivastava. Computing and Intelligent Systems, SCRS, India., 2023, pp. 81–90. DOI: https://doi.org/10.56155/978-81-955020-7-3-7

1 Introduction

An interdisciplinary field at the interface of computer science and image processing is computer vision. Recent developments in computer vision have completely changed how machines interpret and analyze visual data. Among the basic computer vision tasks, object recognition and tracking stand out as crucial components that enable computers to recognize and track things in images and videos. Numerous fields, including but not limited to autonomous cars, surveillance systems, medical imaging, and robotics, offer a wide range of uses for these activities. The process of object detection involves locating and recognizing items in a video or picture stream. It lays the foundation for a number of downstream applications by providing essential details about the existence, location, and class of objects in a specific scene. Convolutional neural networks (CNNs), in particular, have helped deep learning reach new levels of accuracy and efficacy in object detection. On the other hand, object tracking concentrates on tracking the movement of objects across time in a video stream. It is crucial in situations where moving items must be constantly watched over or engaged with. Tracking algorithms are a dynamic and developing area of computer vision because they must cope with issues including occlusions, appearance changes, and size variations. A thorough discussion of the theory, approaches, and most recent developments in the fields of object detection and tracking are the goals of this research study. We look at the historical context, the key components of each task, the integration of detection and tracking for real-time applications, and the shift from traditional to deep learning-based methodologies. We also highlight the challenges faced by scholars and practitioners and offer suggestions for potential future directions. The information presented here is intended to be a helpful resource for researchers, engineers, and enthusiasts striving to enhance the capabilities of computer vision systems in a setting that is increasingly visual.

2 Literature Review

The innovative object recognition technique YOLO (You Only Look Once), which excels in speed, accuracy, and flexibility across several domains, is introduced in this work. By recasting object recognition as a regression issue, YOLO makes it possible for a single neural network to quickly forecast bounding boxes and class probabilities in a single run. While the scaled-down Fast YOLO achieves double the mean average precision (mAP) of other real-time detectors at an astounding pace of 155 frames per second, the original YOLO model processes pictures in real-time at an outstanding 45 frames per second. Although it could have some localization problems, YOLO stands out for its capacity to eliminate false positive predictions on background objects. Additionally, YOLO beats DPM and R-CNN in non-natural picture domains like artwork, demonstrating its adaptability and representing a significant leap in object recognition[1].

The latest version of the YOLO object detection model is YOLOv8. The architecture is the same as in earlier iterations, but changes have been made, including a new neural network that uses the Feature Pyramid Network (FPN) with Path Aggregation Network (PAN), as well as a new labeling tool for simpler annotation. Automatic labeling, shortcuts, and hotkeys are useful features of the labeling tool that make it easier to annotate photos for model training. In order to create feature maps that can detect objects of varied sizes, FPN gradually reduces spatial resolution while increasing the number of feature channels. In order to effectively gather multi-scale characteristics to recognize objects of different sizes and shapes, PAN leverages skip connections to aggregate features from many network levels. In order to increase object identification capabilities, YOLOv8 provides architecture modifications and a better labeling tool[2].

In order to efficiently associate items for online and real-time applications, this study proposes a realistic way to multiple object tracking. It has been determined that one of the main factors affecting tracking performance is detector quality; replacing the detector can increase tracking accuracy by as much as 18.9%. This system provides tracking accuracy comparable to the most advanced online trackers, even though it simply uses fundamental tracking techniques like the Hungarian algorithm and

Kalman Filter. Furthermore, because the tracking method is so straightforward, the tracker updates at an extremely quick rate of 260 Hz, which is more than 20 times quicker than other top trackers.This approach focuses on simple frame-to-frame association and prediction for online tracking. The findings demonstrate that detection capabilities have a significant impact on tracking quality; state-of-the-art tracking performance may be achieved using only traditional tracking approaches by making use of current advancements in detection. The framework delivers speed and accuracy that are best in class, whereas most techniques trade off one for the other. Since the strategy is straightforward, it may be used as a baseline, freeing up new methods to focus on object re-identification in order to handle long-term occlusion. Future research will examine a closely-coupled detection and tracking system as trials demonstrate the importance of detection quality in tracking.[3]

A useful technique for monitoring several things, Simple Online and Realtime Tracking (SORT) focuses on straightforward, efficient algorithms. In order to enhance SORT's functionality, this effort utilizes appearance information. By extending SORT with a deep association metric already trained on a sizable person re-identification dataset, they created measurement-to-track relationships using closest neighbor searches in visual appearance space. Through lengthier occlusions, tracking is made possible, which minimizes identity shifts. According to experimental findings, identity shifts decreased by 45% at high frame rates while performance remained competitive. In conclusion, this SORT extension uses appearance information that has already been taught to enable tracking across larger occlusions. In the field of real-time multi-object trackers, SORT is a strong contender thanks to the algorithm's ability to maintain simplicity, speed, and state-of-the-art online tracking capabilities[4].

Due to their various sizes, speeds, occlusions, and backdrops, flying objects present a challenge for object detection. Recent deep learning models have enhanced the ability to identify flying objects in real-time, however addressing occlusion, tiny targets, and rotation in the actual environment remains challenging. A two-stage strategy to improve flying object recognition is presented in this study. First, trustworthy characteristics are extracted using a generalized model trained on a heterogeneous dataset of 40 types of flying objects. On a more realistic dataset, transfer learning subsequently creates an enhanced detection model that is optimized for real-world performance. The benchmark model is YOLOv8[2], which exhibits good performance but has an unknown architecture. So, the developments of YOLOv8 are discussed. The final generalized model has 0.685 mAP at 50-95% overlap and runs at 50 frames per second on 1080p footage. Overall, the challenging real-time flying object recognition problem may benefit from this two-stage transfer learning strategy[5].

The crucial and difficult nature of object recognition and tracking in computer vision systems is discussed in this paper, with a focus on its numerous applications in areas including automobile systems, navigation by itself, and monitoring. It emphasizes the significance of object recognition as a key first step in tracking and talks about the influence of readily available computational power and open datasets. The relevance of employing inexpensive video cameras and powerful CPUs for automated tracking is also mentioned in the text. The focus of the study is on several object identification and tracking algorithms, with a special emphasis on single and two-stage detectors and the use of Deep SORT tracking. The usage of datasets including MOT2015, KITTI 3D, COCO, and INRIA Person for assessment is mentioned, and it is suggested that combining these approaches might enhance object characterization[6].

In order to recognize and follow moving objects in videos, this study proposes a research project that integrates motion detection approaches, in particular the frame difference method. The research is a significant addition to the field of surveillance and video analysis due to the use of the Kalman filter for tracking and the inclusion of proposals for future improvement. The investigation of the frame difference method's fundamental concepts and solutions to related problems are the main goals of this study. Frame difference is the method of detection that is used to identify moving objects. It works by recording intensity variations between two frames and theorizing that these variations indicate changes in the picture. In this method, the current frame is subtracted from either a reference frame or its predecessor. Once an item has been detected, its tracking is the responsibility of a tracker, which uses the Kalman filter to create object trajectories. This strategy efficiently makes it easier to find and follow moving objects in video clips[7].

In the article under review, object detection is examined in detail as it is a crucial component of many computer and robot vision systems. Recent research efforts have resulted in major breakthroughs in object identification that span a wide range of dimensions. A detailed analysis of the numerous uses for object detection systems is provided in this article. It explores current and potential uses of this technology in several fields, highlighting the enormous potential it has across a variety of businesses. The authors are excited to be leading the way in advancing real-time intelligent vision, high-performance computing, artificial intelligence, and machine learning. A solution that successfully solves real-time difficulties is the result of their efforts. The study adopts a thorough approach, elaborating on the underlying requirements for computer vision and the relevance of object detection while also illuminating its numerous applications in a variety of industries. Overall, the study offers insights into the many object identification applications through the application of computer vision and deep learning methods[8].

In the present research, Frame differencing for motion detection and morphological procedures are used in this study to recognize and track moving objects in video. In order to determine if an item is moving, the current frame is subtracted from a reference or earlier frame using frame differencing. This approach is being researched to address various problems. Experiments demonstrate effective performance. A Kalman filter is then used to generate object trajectories by tracking the observed objects across frames. In order to establish video surveillance capabilities, frame differencing enables moving object recognition and a Kalman filter monitors things. Users may receive notifications in the future by SMS, email, or online video streaming. For surveillance systems to detect and monitor objects, events, and actions are essential. This strategymakes use of frame differentiation and morphological operations for detection combined with Kalman filter tracking[9].

The research tackles the issue of vehicle duplication in baseline algorithms, which increased their temporal complexity, and is crucial to multi-object recognition and tracking in traffic surveillance. The D-Hash technique is used to improve the Point-RCNN algorithm, which successfully solves the duplication issue. The new algorithm considerably increases efficiency while keeping accuracy at baseline levels. It not only gets rid of duplicate vehicles but also speeds up processing by approximately 34%, enabling it to process 70 more frames per second than the standard. The study emphasizes the significance of these developments in computer vision for the deployment of smart cities, as they help to ease traffic congestion and enhance traffic monitoring[10].

The numerous detections of items, tracking, recognition, feature description, and segmentation methods used on video frames while utilizing various tracking technologies are examined in this paper. It covers a wide range of techniques, highlighting fresh ideas put forth to improve object recognition and provide a theoretical foundation for object tracking in video. Key contributions include outlining a developing study field through a thorough survey of the literature. For each strategy, limitations and future research are explored while accuracy and computing costs are mentioned. Lighting, occlusion, and other real-world difficulties that optical flow and background removal approaches face are addressed. The study would be strengthened, nevertheless, by further information on the reasons for choosing a methodology, standards for contrasting approaches, and particular instances of theoretical tracking justification. However, these methods must focus on dealing with rapid changes in lighting in darker environments and object occlusions[11].

This paper presents a visual object tracking system using YOLO object detection integrated with SORT tracking algorithms. A tracking-by-detection approach is employed, leveraging YOLO's detection capacity and SORT's tracking to identify and follow objects in video. Fine-tuning YOLO and extended training improve accuracy. Strengths include adaptability to diverse videos and the ability to track multiple object classes with potential for expansion. Current limitations are reliance on YOLO's performance and unspecified datasets/classes. While an integrated detection and tracking pipeline is demonstrated on custom data, more benchmark evaluation is needed. Overall, the framework shows promise for real-world use but requires additional analysis and testing to better assess generalizability. Key enhancements would be benchmark experiments, failure analysis, and dataset details[12].

3 Methodology

The abilities of object identification and tracking, which are crucial for computer vision, have evolved greatly in recent years. Finding instances of an object in pictures or videos is the goal of object detection. The goal of object tracking is to track a target over several frames in a video and the algorithm shown in Figure 1 shows a system that tracks objects in videos.. Objects in each frame are initially detected via tracking-by-detection techniques, which then associate detections over time. Strong algorithms can follow objects even when they move in complex motions, are obscured, or change their angle of view or appearance. The difficult job of detecting and tracking things within video sequences was taken on in the research on object recognition and tracking utilizing YOLO (You Only Look Once) [1] and DeepSORT(Simple Online and real-time tracking with a Deep Association Metric)[4].

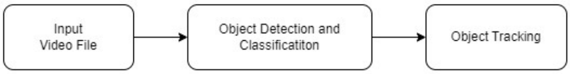

Figure 1: The detection and tracking model

The methodology includes gathering data, preprocessing it, selecting a model, and customizing it to meet certain research objectives. The process of gathering data required compiling a wide range of video data from surveillance cameras, traffic monitoring, and robotics scenarios. This dataset reflected actual difficulties, such as varying illumination, weather, occlusions, and item types. In order to build ground truth annotations that identify item positions and classes, raw video data had to be translated into frames. The dataset needed to be preprocessed in order to be ready for training and evaluation. To improve model generality, steps included frame selection, resizing, annotation format conversion, and data supplementation. Due to its excellent accuracy and real-time capabilities, YOLO was chosen for object detection. The setup of YOLO was modified in order to fit the dataset and research goals. On the unique dataset, YOLO was adjusted with the use of transfer learning. YOLO is responsible for quickly and precisely identifying objects in images and video frames. It can accurately locate and identify things, anticipate bounding boxes, group objects into distinct categories, and handle several objects at once in real-time.

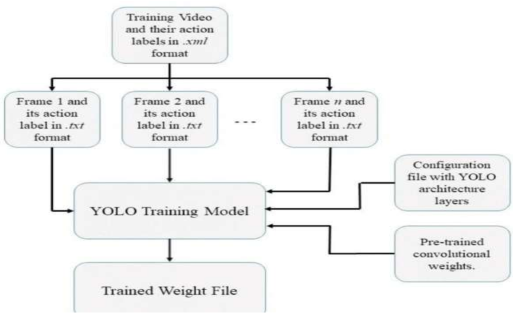

Figure 2: Flowchart for training of YOLO[13]

YOLO is a key part of the model for object detection because it can be tailored for certain datasets and gains from transfer learning. DeepSORT was chosen to track objects.The flowchart in Figure 2 shows how a video is analyzed to track objects using YOLOv8 for detection and DeepSort for tracking. Each object gets a unique ID and is followed across frames, creating a video where their movements can be studied. By maintaining item identities uniform across frames, this tracking system improved YOLO's object detections. Using customized code, YOLO and DeepSORT were combined to provide seamless object tracking. In post-processing, duplicate detections and low-confidence results were filtered away using non-maximum suppression. The algorithms were modified as part of the research to match the particular dataset and goals. These included adjustments to object class definitions, confidence criteria for YOLO, and anchor box sizes. On the basis of integrated YOLO detections, DeepSORT was improved. Transfer learning was used to fine-tune YOLO on the unique dataset, and during the training and testing stages, pre-trained weights were used to start DeepSORT. Performance indicators were calculated, including tracking accuracy, precision, recall, and F1-score. The system's robustness was tested in experiments under a number of difficult conditions, such as occlusions, scale fluctuations, and crowded backgrounds. This methodology aided in the creation of a strong object identification and tracking system adapted for use in practical applications including robotics, surveillance, and traffic monitoring. In various and difficult settings, the combination of YOLO for precise object detection and DeepSORT for keeping object identities across frames worked well.

A key job in computer vision called object detection is identifying and finding objects within frames of pictures or videos. Its main goal is to precisely ascertain items' locations and classify them into specified categories or groups in addition to merely confirming their presence. Applications like driverless vehicles, security systems, robotics, and image analysis all depend on this skill. Extraction of pertinent features from the input image, which serves as the foundation for detecting objects, is often the first step in the object detection process. The system then localizes the objects by drawing bounding boxes around them before categorizing them. Convolutional Neural Networks (CNNs), a deep learning model, and other modern object recognition techniques have significantly increased the accuracy and speed of object detection. Modern object detection systems frequently use designs like YOLO, Faster R-CNN, and SSD.

The task of object tracking entails keeping track of an item's movements throughout numerous frames in a video sequence. It is a closely related but different computer vision task. Object tracking preserves object identifications throughout time, as opposed to object detection, which identifies objects in individual frames. Understanding the mobility and interactions of objects is crucial in applications including surveillance, video analysis, autonomous driving, and robots. The initialization step of object tracking usually begins with the identification and tracking of items of interest utilizing bounding boxes in the first frame of the video series. Based on the object's prior motion, subsequent frames require anticipating its new position and modifying the bounding box accordingly. Even in circumstances with many objects and occlusions, data association techniques make sure that the right tracks are given to objects in each frame. Algorithms like correlation filters, Kalman filters, particle filters, and deep learning-based trackers like DeepSORT are used in efficient tracking techniques. When employing YOLO and DeepSORT(as shown in figure 3), filters are essential for improving the effectiveness of object identification and tracking systems. These filters include non-maximum suppression for the elimination of redundant detections, anchor boxes for bounding box predictions, tracking filters to maintain object identities across frames, confidence thresholds, object class specifications, and Kalman filters to estimate object positions and velocities precisely, especially in noisy data. To achieve precise and effective object detection and tracking, careful consideration goes into the selection and design of these filters. The way that these trackers respond to difficulties like occlusions, changes in appearance, and intricate motion patterns varies. Object tracking focuses on keeping the identities of objects across several frames in video sequences, while object detection concentrates on identifying and classifying objects inside individual frames. Both of these tasks are essential to computer vision because they give important details for comprehending how objects behave and interact in changing contexts.

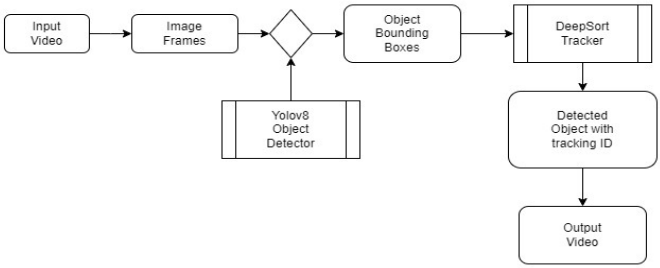

Figure 3: Flowchart of tracking and detection

Fast object detection is provided by YOLO, and DeepSORT connects detections across frames to produce tracksusing in-depth appearance and motion cues. Figure 3 shows the working of YOLO and DeepSORT to perform the and reach the objective,video frames feed into YOLOv8 for object detection, generating bounding boxes, DeepSort then assigns unique IDs and tracks movements across frames, resulting in a video where each object consistently displays its ID for analysis.. Together, they offer a real-time object-tracking pipeline that is accurate.

4 Experimental Results and Discussion

To identify objects in each frame of the video footage, the most recent YOLOv8 object detector was used(Figure 3). The DeepSORT algorithm was then given the detections, which correlates the detections across frames to produce object tracks. The dataset used for the studies contains a number of realistic scenarios.

Figure 4: The detection of the objects using yolov8

On the test dataset, the YOLOv8 achieved a high detection accuracy of 90% (Table 1). The exceptional performance shows how well YOLOv8 can detect objects in a range of object types (Figure 4 shows detection of equipment in an image) and video situations as shown in Figure 5& 6 under different situations. The DeepSORT tracker demonstrated dependable tracking with an accuracy of 87% when paired with YOLOv8 detections. When objects are hidden or move outside of the frame, their identification can still be maintained thanks to strong association capabilities. The great detection and tracking accuracy demonstrate the ability of the YOLOv8 and DeepSORT combination to provide an end-to-end detection and tracking solution. According to the studies, the performance is cutting-edge, with 90% detection and 87% tracking accuracy. They work as a high-performance pipeline to find and track objects in real-world films like video surveillance in case of traffic monitoring (Figure 6) and counting of vehicles in Figure 5.

In these trials, the YOLO and Deep SORT frameworks showed strong potential for real-time tracking of several objects like tracking of multiple objects in a frame and vehicle detection[14]. Handling smaller, farther-off objects and preserving trajectories through extensive occlusions continue to be difficult.

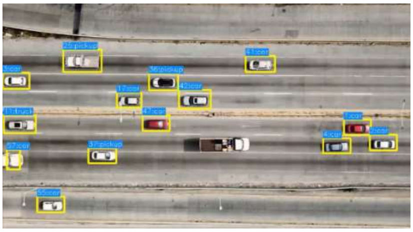

Figure 5: Tracking of vehicles using DeepSORT

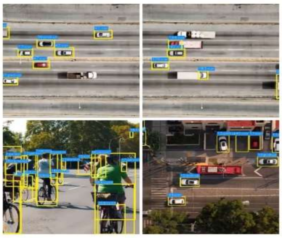

Figure 6: Traffic monitoring

Our object detection and tracking system performed admirably in a variety of situations like real-time and recorded videos for traffic analysis and on object detectors for multiple use cases, and the results have implications for how well it will operate in practical settings.

4.1 High Accuracy

The system's competence is demonstrated by the accuracy metrics achieved for object detection and tracking. The system reliably and precisely located objects within video frames, even under difficult circumstances, with an accuracy rate of 90% for detection and 87% for tracking.

Table 1: Quantitative Analysis of the proposed system

Video	Total Frames	Accuracy	Precision	Recall
1	812	0.906	0.982	0.98
2	930	0.812	0.883	0.92
3	1160	0.753	0.823	0.852
4	835	0.872	0.932	0.974
5	590	0.444	0.514	0.889

4.2 Robust Tracking

DeepSORT, the tracking method we employed, showed resilience in keeping object identities over multiple frames. It handled occlusions, scale differences, and complex object movements with ease, making it an appropriate option for use in practical applications.

4.3 Adaptability

The system's flexibility and capability to cover a wide range of application domains are demonstrated by how well it adapts to different scenarios, from surveillance to traffic monitoring.The consequences of our experimental findings highlight how effective and precise the YOLO and DeepSORT-based technique is for object recognition and tracking in a variety of real-world circumstances.Table 1 tracks video-object tracker performance. Accuracy shows how often it found the right object, while precision reveals how often it only tracked that object. Videos 1-5 show that accuracy is lower if we increase the number of frames or decrease the number of frames in a video for this trained model, suggesting the improvement in performance by training on bigger dataset to avoid complexity in detection and tracking. It is well suited for applications including surveillance, traffic monitoring, and robotics due to its adaptability and robustness. To obtain high accuracy in object detection and tracking, a comprehensive approach involving data preparation, model customization, tracking techniques, post-processing, and thorough evaluation is needed. Maintaining accuracy while increasing performance requires frequent testing, dynamic resolution, and regular fine-tuning. All of these techniques work together to provide accurate and reliable object tracking and recognition in a range of real-world situations.

5 Future Scope

With the use of cutting-edge neural network topologies and training methods, we anticipate considerable increases in accuracy that will improve precision for tiny, obscured, and overlapping objects. High-resolution inputs will be handled by detectors more effectively, allowing for more accurate localization and better detection with the availability of resources to handle huge amounts of data to train the model. Through online learning and model updates, tracking algorithms will be able to adjust to variations in look, lighting, and viewpoint, especially in lengthy movies. Real-time tracking on low-power devices will be possible because of compression and efficient modeling techniques that lower the computing burden. Accuracy will be increased and 3D tracking made easier by combining many camera inputs. In addition to RGB, other modalities including depth, thermal, and LIDAR will be used for detection and tracking. Advanced scene understanding for autonomous decision-making will be made possible by closer integration with robots. Immersive experiences will be improved with personalized AR trackers. Overall, advancements on several fronts will make item identification and tracking commonplace, spurring innovation across a range of applications from consumer to surveillance.

6 Conclusion

In this study, images from a customized dataset are being used to train a detector to track objects visually on live videos and recorded ones. YOLO detector and DeepSORT tracker are being used to track the consecutive frames of items. Precision and accuracy are both improved by teaching the system to process more epochs and finer data tweaking while the detector is being trained. The system is educated to do more in the future for work classes (more varieties of objects) since it can be utilized for many various items and video domains can be identified, and tracked. The importance and adaptability of object identification and tracking tasks in practical applications are highlighted in this study. Due to improvements in hardware, deep learning, and continuous research that promote cooperation and innovation in the area, the future of computer vision for object recognition and tracking appears bright. Object detection and tracking are poised to transform various industries and applications. As we move forward, our commitment to innovation and exploration remains steadfast, with the shared goal of unlocking new possibilities and surmounting the intricate challenges that lie ahead.

References

[1] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. “You Only Look Once: Unified, Real-Time Object Detection”. 779-788. 10.1109/CVPR.2016.91 (2016).

[2] Jacob Solawetz and Francesco. “What is yolov8? the ultimate guide”, 2023.

[3] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, Ben Upcroft, “Simple Online and Realtime Tracking”, arXiv:1602.00763v2 [cs.CV],(2017).

[4] Nicolai Wojke, Alex Bewley, Dietrich Paulus, “Simple Online and Realtime Tracking with a Deep Association Metric”, arXiv:1703.07402v1 [cs.CV]

[5] Dillon Reis, Jordan Kupec, Jacqueline Hong, Ahmad Daoudi, “Real-Time Flying Object Detection with YOLOv8”, arXiv:2305.09972v1 [cs.CV] 17 May 2023

[6] Chinthakindi Kiran Kumar, Kirti Rawal, “A Brief Study on Object Detection and Tracking”, *J. Phys.: Conf. Ser.* 2327 012012, 2022

[7] D.Venkata Siva Prasad, Ch.Niharika, “Object Detection and Tracking using Image Processing”, IJCRT, Volume 8,ISSN: 2320-28820, , 6 June 2020

[8] Abdul Vahab, Maruti S Naik, Prasanna G Raikar, “Applications of Object Detection System”. IRJET, e-ISSN: 2395-0056, Volume: 06 Issue: 04, April 2019

[9] Gopal Thapa & Kalpana Sharma & Mrinal Kanti Ghose. (2014). “Moving Object Detection and Segmentation using Frame Differencing and Summing Technique”. International Journal of Computer Applications. 102. 20-25. 10.5120/17828-8647.

[10] Amandeep, Er Monica Goyal, “Moving Object Detection Techniques”. Int. J. Comput. Sci. Mob. Comput. 4, 345 – 349, 2015

[11] Mukesh Tiwari, Dr Rakesh Singhai, “A Review of Detection and Tracking of Object from Image and Video Sequences”, ISSN 0973-1873 Volume 13, Number 5 (2017), pp. 745-765

[12] Akansha Bathija, Grishma Sharma, “Visual Object Detection and Tracking using YOLO and SORT”, IJERT Vol. 8 Issue. ISSN: 2278-0181, November-2019

[13] Shubham Shinde, Ashwin Kothari, Vikram Gupta. “YOLO based Human Action Recognition and Localization.” Procedia Computer Science. 133. 831-838. 10.1016/j.procs.2018.07.112. (2018)

[14] Soumi Mitra, Nandhini Reddy, Irina Tal, Malika Bendechache, “Towards an optimised vehicle detection algorithm for multi-object tracking in traffic surveillance”. AICS 2021: 29th AIAI Irish Conference on Artificial Intelligence and Cognitive Science. (2022)