Proceedings of the International Conference on Augmented Intelligence and Sustainable Systems (ICAISS-2022)
IEEE Xplore Part Number: CFP22CB2-ART; ISBN: 978-1-6654-8962-1

DeepSORT: Real Time & Multi-Object Detection and Tracking with YOLO and TensorFlow

1st Abhijeet Pujara
School of Computer Engineering & Technology
MIT World Peace University
Pune, India
abhijeetpujara@gmail.com

2nd Prof. Mamta Bhamare
School of Computer Engineering & Technology
MIT World Peace University
Pune, India
mamta.bhamare@mitwpu.edu.in

2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS) | 978-1-6654-8962-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAISS55157.2022.10011018

***Abstract*—One of the most significant and challenging areas of computer vision is object recognition and tracking, which is extensively utilised in many industries including health care monitoring, autonomous driving, anomaly detection, etc. The tracking of moving objects in videos is actively researched over the past two decades due to its practical applications in many fields such as event analysis, human-computer interaction, crowd analysis, video surveillance, behaviour analysis, etc. The effectiveness of object trackers and detectors has significantly increased with the rapid advancement of deep learning (DL) networks and GPU processing capability. New methods have been presented for object recognition and tracking in video as a result of extensive study in this field. This article addressed the several processes of object tracking in video sequences: object detection, object classification, and object tracking, in order to comprehensively comprehend the key advancements in the object detection and tracking pipeline. Additionally, we thoroughly examine the various approaches available for object recognition, categorization, and tracking.**

Keywords—Background subtraction, Human detection, Object detection, Object tracking, Video surveillance.

I. INTRODUCTION

Object tracking is a critical component in Computer Vision (CV) applications such as human-computer interface, visual surveillance, geolocation, gesture recognition, augmented reality, movie production, and person following [3]. Numerous studies on object identification and counting have been conducted, including one by [6] which looked at object detection with YOLO, object tracking with Deepsort, and counting tracked objects. Even when the tracked items are no longer present in picture frames, Deepsort is still able to identify them.

This survey article builds on earlier research in the field of multi-object identification and tracking and aims to determine the advantages of using research on tracking data [17], [7]. On COCO data, the majority of research is conducted using conventional machine learning techniques as well as certain deep learning techniques. The new methods seek to increase accuracy [18]. But Object detection categorization poses certain difficulties. The performance of an object tracking algorithm is influenced by a number of variables, including different lighting conditions and views, rapid motion, and occlusions [3]. Recent developments in computer vision and unmanned aerial vehicle (UAV) control technology have opened up new possibilities for both military and civilian uses [18].

Object tracking involves a number of phases, including object detection, creating IDs for objects that are discovered, classifying objects, and detection [17]. The YOLO Technique is used for object identification and affirmation. For instance, bouncing boxes are initially created and then placed under the appropriate classes after some time. We employed Kalman channels because they increased the precision of our model and produced better results. Kalman channels are also used to display these states so we can track them till they increase in the edge [16]. Object tracking may be used with 3D disciplines as well as 2D sequence data to estimate the position of the target entity that is visible in the scene [9].

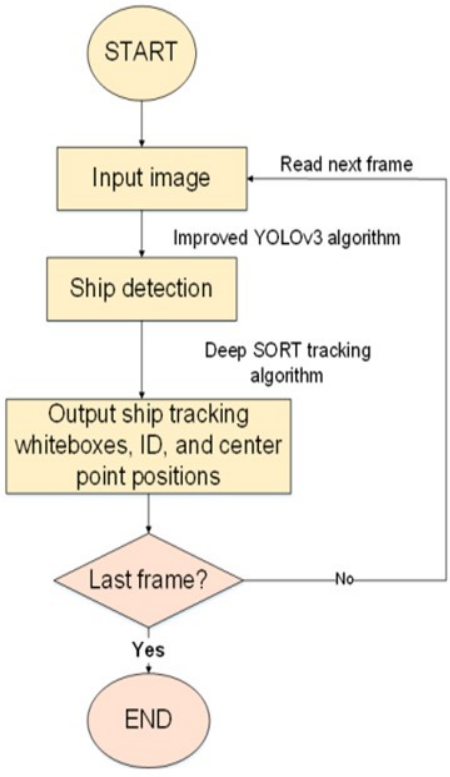

Fig. 1: Flow Diagram of Basic steps in Object Tracking [22].

II. RELATED WORK

Identifying object tracking using machine learning and deep learning models may be done in a variety of ways [16]. as object tracking has captured the interest of several academics, it has significantly aided numerous applications. The discovery about object categorization has created several opportunities [14]. Many machine learning and deep learning techniques, including YOLO, Deep Sort, RNN, Faster RCNN, and CNN, have been used to build object models [8]. In entity disclosure and global positioning systems, Deep SORT is one of the major estimations that is astonishingly quick and accurate [16]. Kalman filters operate on consistent speed and straight discernment and assess the method of the article when that thing is not perceptible due to obstructions like smoke, murkiness, flying birds, etc.

Deep sort based on YOLO, target tracking based on Mosse + KCF, and target tracking based on Siam algorithm are the popular techniques for monitoring pedestrians. Although the target tracking based on the Mosse + KCF algorithm has high speed and minimal hardware needs, it is easy to lose the target since the tracking frame scale is constant and cannot follow the target's changing scale. The YOLO network is utilised for detecting [2] in order to get the best detection result of the previous vehicle.

Let's see how the YOLO and Deep SORT technique's function:[15]

- Using the YOLO method, objects are found and recognised.
- Motion Prediction and feature production include the creation of an estimating model and the modelling of these states using Kalman filters.
- Kalman filters are used to track items in the previous frame, and then the current frame is updated with the newly discovered objects. The next step is the creation of an association for new detection.

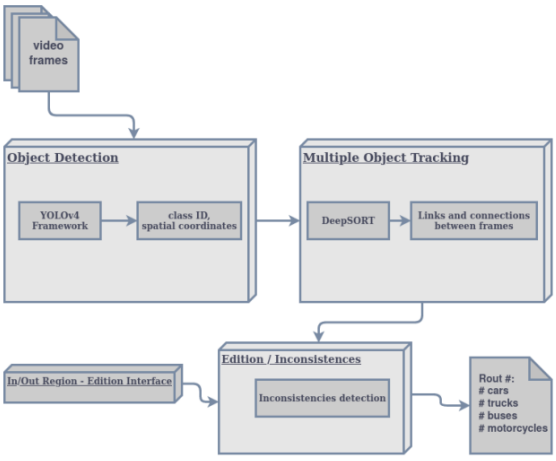

Fig. 2: Block Diagram of the multi vehicle tracking [1].

III. LITERATURE SURVEY

Table 1 lists the results of a survey on current research developments, datasets utilised, benefits, approaches, and research needs.

TABLE I. LITERATURE SURVEY TABLE.

Ref no.	Year	Technique	Dataset	Accuracy
1	2021	YOLOv4 and DeepSORT	Custom	78%
2	2021	YOLOv3-tiny	VOC2012	84%
3	2021	DeepSCT	VisDrone SOT2019	74%
4	2022	YOLOv3 and DeepSORT	Custom	74%
5	2021	YOLOv3 and DeepSORT	Custom	78.29%
6	2022	YOLOv4 and DeepSORT	COCO	80.99%
7	2021	YOLOv3 and DASiamRPN	Custom	76%
8	2021	YOLOv4 and DeepSORT	COCO	84%
9	2021	YOLOv3	COCO	91%
10	2021	Multiple algorithms	Multiple Datasets	-
11	2014	Multiple algorithms	Multiple Datasets	-
12	2012	Multiple algorithms	Multiple Datasets	-
13	2013	Multiple algorithms	Multiple Datasets	-
14	2021	YOLOv-3,4, Faster RCNN, Mask RCNN and SORT, DeepSORT	COCO, MOT	54%, 72%, 92%
15	2021	YOLOv4 and DeepSORT	MOT17	69.8%
16	2021	YOLOv4 and DeepSORT	MOT18	61.03%
17	2022	YOLOLv4 and DeepSORT	CVC-CLINIC, ETILS-Larib, Kvasir-SEG.	73%

18	2021	YOLOv5 And hungarian algorithm	UAV123	86%
19	2022	YOLOv4 and DeepSORT	Vehicle Licence Plate & custom dataset	90.32%
20	2021	YOLOv4 and DeepSORT	Open Images V4 Dataset	82.08%

IV. REVIEW

- According to the survey's results, the most popular algorithms for tracking multiple objects are YOLO, Deep SORT, and RCNN [1], [5], [14].
- Recent studies show that deep learning models, such as Deep Sort, are preferred by researchers over more conventional machine learning models. Common traditional models are Open CV and so on [4].
- For improved prediction and accuracy, several systems employ deep learning models. Better temporal and space complexity is another thing it is capable of [3].
- For multi-object recognition and tracking, some of the most recent YOLO and DeepSort research studies obtained above 90% accuracy [19], [9].
- The most popular models for multi-object tracking, according to an examination of the research, were found to be YOLO and hybrid models [17].
- To achieve object recognition and classification, deep learning may be utilised to extract deep visual characteristics [18].
- Residual Network (ResNet) is utilised for feature extraction following analysis, and Kalman filter is used to extract the object's motion data as well as the expected Tracks [18].
- The deep learning algorithm's quickest target detection network is called YOLO.Fully convolutional neural network describes it. The detection speed has essentially met the real-time requirements thanks to the addition of the anchor mechanism and the ability to input photos of any size [2].
- To achieve high accuracy with minimal time and space complexity, it is advised to employ YOLO, DeepSort, or hybrid models [5].

V. APPLICATIONS

Multi-Object detection and tracking methods are used in many sectors for various purposes.

- With the use of route recognition and accounting in photographs gathered at road crossings, object tracking may be utilised to classify vehicles according to local reality. the installation of traffic signals, the creation of safety lanes, the enlargement of roadways, and other societal needs, to aid in the decision-making of public officials [1], [8], [19].
- Object tracking may be utilised to establish a system for assessing the efficacy of urban sanitation, which will further enhance smart cities' capacity for management and service delivery [4].
- With the help of Tracking of the object, we can detect and track the driver whenever they using their phone while driving [5].
- Tracking can be done efficiently by counting the number of motorcycles for better traffic arrangement [6], [15].
- The use of object detection and tracking can help stop crowds from forming during COVID hours and setting off a warning for all authorities [15].
- It can be used for the navigation of Autonomous Surface Vehicles (ASVs) in the ocean [08].

VI. PROPOSED WORK

Using various machine learning and deep learning techniques, we have articulated and comprehended the most current developments in the field of object tracking. The researchers will find this helpful for understanding and contributing to the work against object categorization. Additionally, using the YOLOv7 and DeepSORT algorithms, we conducted survey analyses against other systems that have recently been suggested in the scientific community.

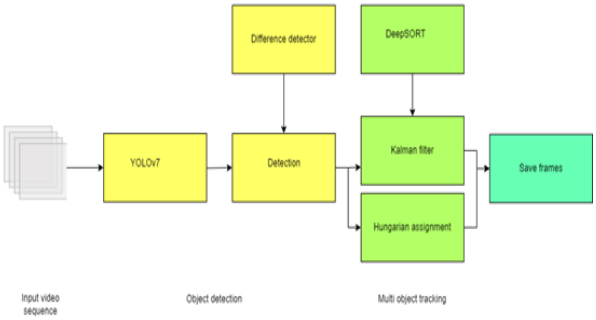

Fig. 3 : Proposed system overview.

This section will go into great length on the project's process and the flowchart that follows it. It goes through the method's design for detecting moving objects for traffic control. The technique flowchart for this system's work is shown in Figure 3. In this, we suggested a YOLOv7 and DeepSORT-based Object recognition and tracking on the huge syntenic dataset and custom dataset. Machine learning (ML) and deep learning (DL) tasks are combined in the implementation. The primary goal usage of YOLO is for object detection, whereas the remaining purpose of deepSORT is for object tracking. Figure 4 shows that the YOLOv7 model is more precise, acceptable, and quick than the prior one.

As shown in fig. 4, the neural network-implemented method YOLOv7 delivers quicker object detection with higher accuracy. The image is split into a number of grids using YOLOv7, each with a set size. The detection of items encircled by a grid is the responsibility of every grid. Every grid cell contains the anticipated bounding boxes and confidence scores for each box. These confidence ratings show the degree to which the model is certain that the box includes a certain object as well as how precisely it confirms that the box contains that specific object. The confidence score is provided as 0 [17] if there isn't an item inside the box. This method drastically saves computation because cells from the picture retain both recognition and detection concurrently, but it leads to a large number of redundant predictions because several cells may forecast the same item with different bounding boxes.

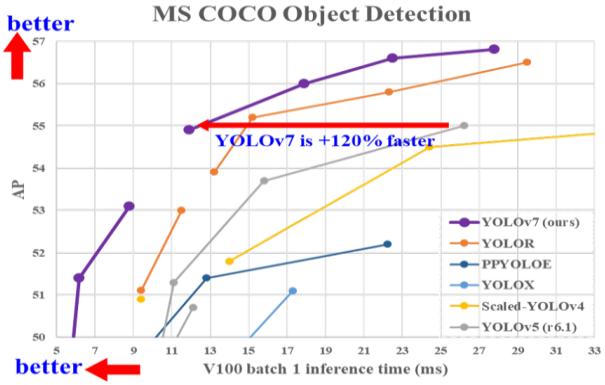

Fig. 4 : Comparison with other real-time object detectors with YOLOv7 [21].

Non-maximal suppression is a strategy that YOLOv7 employs to address this problem. This method suppresses or ignores the bounding boxes with lower probability values. The bounding box with the highest probability score is chosen for this purpose by YOLOv7. Following that, the bounding boxes with the largest Intersection over Union (IoU) with the present high probability bounding box are suppressed. The process is continued until the final bounding box with accurate object detection is obtained.

A modification of Simple Online and Realtime Tracking executes the DeepSORT algorithm (SORT). With a bounding box overlap association metric and frame-by-frame data association using the Hungarian technique, the SORT algorithms conduct filtering (Kalman Filter) in picture space. By including a re-identification on the identified objects across the frames based on a pre-trained CNN to generate bounding box appearance descriptors, the DeepSORT approach enhances resilience against misdetections and occlusions.

Let's see how the YOLO and DeepSORT technique's function:

Through the use of the YOLO Technique, objects are identified and confirmed. For instance, bouncing boxes are first created before being classified under known categories. Motions When an appraisal model is created, prediction and feature ageing take place. Kalman channels are utilised to display these states so we can track them till their augmentation in the edge. Finally, with the support of the kalman filter, the following takes place: recently identified items are placed in their present packing, after which a link is formed for fresh revelation, and this is sustained over time for discovering them.

VII. CONCLUSION

The many approaches to do object recognition, object categorization, and object tracking were covered, as well as the various phases involved in tracking an item from a video sequence. Each of these strategies' benefits and drawbacks were explored. We intend to present an algorithm in the future that addresses the shortcomings of the current object detection techniques and tracks things in a video while being able to manage numerous objects and occlusions. However, there are still a variety of difficult problems that need to be resolved, and further study is needed to significantly enhance the performance, generalizability, and effectiveness of the existing algorithms.

REFERENCES

[1] L. Gatelli et al., "Counting, Classifying and Tracking Vehicles Routes at Road Intersections with YOLOv4 and DeepSORT," 2021 5th International Symposium on Instrumentation Systems.

[2] L. Li and Y. Liang, "Deep Learning Target Vehicle Detection Method Based on YOLOv3-tiny," 2021 IEEE.

[3] K. Agrawal, R. Lal, H. Patil, S. Kannaiyan and D. Gupta, "DeepSCT: Deep Learning Based Self Correcting Object Tracking Mechanism," 2021.

[4] Z. Wang, J. Feng, X. Feng, X. Liu and X. Zhao, "Evaluation Algorithm of Urban Road Sanitation Quality Based on Machine Vision," 2022.

[5] S. Carrell and A. Atapour-Abarghouei, "Identification of Driver Phone Usage Violations via State-of-the-Art Object Detection with Tracking," 2021

[6] B. Tongmonwit and K. Tamee, "Increasing Efficiency in Counting Number of Motorcycle with Object Relation Matching," 2022.

[7] C. He, X. Zhang, Z. Miao and T. Sun, "Intelligent vehicle pedestrian tracking based on YOLOv3 and DASiamRPN," 2021.

[8] D. F. Duarte, M. I. Pereira and A. M. Pinto, "Multiple Vessel Detection and Tracking in Harsh Maritime Environments," OCEANS 2021.

[9] N. M. Krishna, R. Y. Reddy, M. S. C. Reddy, K. P. Madhav and G. Sudham, "Object Detection and Tracking Using Yolo," 2021.

[10] Pal, S.K., Pramanik, A., Maiti, J. *et al.* Deep learning in multi-object detection and tracking: state of the art. *Appl Intell* **51,** 6400–6429 (2021).

[11] Kirubaraj Ragland, P. Tharcis, "A Survey on Object Detection, Classification and Tracking Methods", 2014 International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181 IJERTV3IS110458 Vol. 3 Issue 11.

[12] Kinjal A Joshi, Darshak G. Thakore, "A Survey on Moving Object Detection and Tracking in Video Surveillance System", 2012, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3.

[13] Shilpee A. Dave, Dr. M.S. Nagmode, Aditi Jahagirdar, "Statistical Survey on Object Detection and tracking Methodologies", 2013 International Journal of Scientific & Engineering Research, Volume 4, Issue 3, ISSN 2229-5518.

[14] V. -T. Pham, T. -H. Tran and H. Vu, "Detection and tracking hand from FPV: benchmarks and challenges on rehabilitation exercises dataset," *2021 RIVF International Conference on Computing and Communication Technologies (RIVF)*, 2021 S. Kumar, Vishal, P. Sharma and N. Pal, "Object tracking and counting in a zone using YOLOv4, DeepSORT and TensorFlow," *2021*

[15] S. Kumar, Vishal, P. Sharma and N. Pal, "Object tracking and counting in a zone using YOLOv4, DeepSORT and TensorFlow," *2021*

[16] J. S. C. Reddy, C. Venkatesh, S. Sinha and S. Mazumdar, "Real time Automatic Polyp Detection in White light Endoscopy videos using a combination of YOLO and DeepSORT," 2022 1st International Conference on the Paradigm Shifts in Communication, Embedded Systems, Machine Learning and Signal Processing *(PCEMS)*, 2022.

[17] X. Luo, R. Zhao and X. Gao, "Research on UAV Multi-Object Tracking Based on Deep Learning," *2021 IEEE*

[18] R. S. Charran and R. K. Dubey, "Two-Wheeler Vehicle Traffic Violations Detection and Automated Ticketing for Indian Road Scenario," 2022

[19] M. A. Bin Zuraimi and F. H. Kamaru Zaman, "Vehicle Detection and Tracking using YOLO and DeepSORT," *2021*

[20] Yang Jie, Lilian Asimwe Leonidas, Farhan Mumtaz,and Munsif Ali "Ship Detection and Tracking in Inland Waterways Using Improved YOLOv3 and Deep SORT"

[21] Wang, C. Y., A. Bochkovskiy, and H. Y. M. Liao. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv 2022."

[22] Jie, Y., Leonidas, L., Mumtaz, F., & Ali, M. (2021). Ship detection and tracking in inland waterways using improved YOLOv3 and Deep SORT. *Symmetry*, *13*(2), 308.