Review: Multiple Objects Detection and Tracking using Deep Learning Approach

Abhijit Ghodake

Department of Information Technology

VPKBIET, Baramati, India

abhijitghodake2002@gmail.com

Sharim Shaikh
Department of Information Technology

VPKBIET, Baramati, India

shaikhsharim7@gmail.com

Pranav Katkar

Department of Information Technology

VPKBIET, Baramati, India

pranavkatkar18@gmail.com

Ashutosh Deshmukh
Department of Information Technology

VPKBIET, Baramati, India

deshmukhashutosh10@gmail.com

Abstract — Multiple Object Tracking (MOT) is a crucial tool with diverse applications, such as object detection, object counting, and security systems. Precise identification and monitoring of numerous objects are essential in several computer vision uses, such as monitoring, self-driving cars, and computer-human communication. Very little has been done to address occlusion problems in order to enable the best moving object tracking with detection The tracking of visual objects is one of the most important components of computer vision. The process of tracking an object (or a group of objects) across time is called object tracking. Visual object tracking is used to identify or link target items over successive video frames. In this study, we analyze the tracking-by-detection strategy, which includes YOLO-based detection and SORT-based tracking.

This work elucidates a general approach to tracking and recognizing many objects with an emphasis on accuracy improvement. We aim to revolutionize computer vision by applying Non-Maximum Suppression (NMS) and Intersection over Union (IoU) approaches, and by combining the state-of-the-art YOLO NAS algorithm with conventional tracking methods or an alternative version of the YOLO Algorithm for object identification. It is expected that our work will have a major impact on many different applications, enabling more precise and reliable object tracking and detection in difficult real-world scenarios.

***Keywords**—Multiple Object Detection, Kalman Filters Multiple Object Tracking, DeepSORT, YOLO, IoU, NMS.*

I. Introduction

Computer vision has fundamentally modified our approach to the important problem of multiple object recognition and tracking in deep learning. This This method uses many items in pictures or video streams to be watched and identified at the same time using state-of-the-art neural networks. The model uses YOLO for detection and DeepSORT for tracking, together with two crucial methods, Non-Maximum Suppression (NMS) and Intersection over Union (IOU), to improve accuracy. IOU helps reduce overlapping item detections, whereas NMS eliminates redundant detections, especially in cases of occlusion. The model's accuracy and dependability are greatly increased by this combination, making it a vital tool for a variety of applications, including surveillance and driverless vehicles.

The potential for significant advancements in the transportation industry has increased with the advent of autonomous cars in recent years. Numerous benefits are offered by these cars, including as lower emissions, better fuel efficiency, traffic management, and enhanced safety due to fewer human errors.However, accurately identifying and tracking moving objects as possible barriers is a major issue in this discipline. The use of intelligent vehicle recognition and counting is becoming more and more essential in the field of managing transportation infrastructure, particularly highways. Real-time vehicle detection, categorization, and counting are difficult for conventional image-based techniques to accomplish, particularly in challenging real-world environments. In addition to researching YOLO NAS, this project prioritizes effective algorithms, tracks benefits, and develops an adaptive model that combines YOLO and DeepSORT.

The research focuses on improving object tracking using Kalman filters and the YOLO algorithm. It uses motion prediction, feature generation, and DeepSORT for improved accuracy. The model is used in fields like traffic management, crowd assembling, surveillance, and can also be used in animal management. The tracking-by-detection strategy, which includes YOLO-based detection and SORT-based tracking. For traffic studies, identifying a car or a person in an ongoing video is useful.

Population growth necessitates a people counting system. In the field of visual surveillance, accurate people counting and tracking are active study areas. In this work, a novel method for estimating persons and locating them in a series of video frames has been proposed. The image processing technology of object tracking is well-known and has a bright future. Because of deep learning, computer vision, machine learning, etc., the MOT has significantly expanded in recent years. The goal of this work done was to put forth a software solution that manages item list and count by keeping track of the objects. Unlike the basic Yolo object detection tool, which detects all items at once, this MOT system just detects what the user needs to detect, which helps to increase system performance.

The research on real-time detection and tracking algorithms has grown due to the increasing use of surveillance cameras in security and surveillance. However, challenges persist in the detection and tracking stages. A new algorithm is proposed to detect and track objects from natural scenes captured with real-time cameras with updating till date named as YOLO NAS algorithm. Experimental results show the new algorithm's effectiveness and efficiency, achieving good detection and classification accuracy when compared to other detection and tracking systems.

Artificial intelligence, pattern recognition, and military navigation are crucial in surveillance, security, and transportation. OpenCV is also a popular tool for vehicle detection and tracking, offering low prices, convenient installation, and wide monitoring range. It can also be used on the road. These technologies are essential for improving security, military navigation, and intelligent transportation.

For the monitoring, planning, and management of traffic flow, moving vehicle identification, tracking, and counting is essential. A video-based system doesn't impede traffic and is simple to setup.

There are several applications for object detection and tracking, including autonomous automobiles, security systems, patient monitoring, and others. The difficulties of long-term occlusion, identity switching, and fragmentation in real-time multi-object identification and tracking have been addressed using a variety of techniques. However, it is still unknown how to decrease fragmentation and the number of identity shifts in multi-object tracking and detection. As a result, we suggested a two-stage multi-object identification and tracking method in this study. By integrating sophisticated AI approaches, Computer Vision has established a new standard for picture resolution, object identification, object tracking, and other areas. There are several applications for object detection and tracking, including autonomous automobiles, security systems, patient monitoring, and others.

I. Literature Survey

D. Li et al., [1] put forth similarly video-based Solution applied with adaptive subtracted technology and background technology, Virtual detector and blob tracking technology, virtual detector in OpenCV developments kits. Authors proposes a video-based approach that applies adaptive subtracted background technology in conjunction with virtual detector and blob tracking technologies to analyze the traffic video sequence captured by a video camera. The suggested technique can recognize, track, and moving cars counting reliably, according to experimental findings implemented in Visual C++ code using OpenCV development kits [1].

Yingjie Cai et al., [2] proposed high efficiency and celerity of faster RCNN (Region based Convolutional neural network) and KCF (Kernelized correlation filter) for real time performance and effective validation [2]

Bathija et al., [3] proposed tracking by detection approach using YOLO for detection with SORT for tracking for 6 classes. Accuracy, precision worked by training model by more epochs with fine tuning while detector training. Future Scope for this approach would be more objects detection by input vary with YOLOv1,YOLOv2,YOLOv3

, R-CNN, SSD Deep SORT, Centroid, IoU or tracker, CNN LSTM for more accuracy Visual object tracking is used to identify or link target items over successive video frames.

In this study, we analyze the tracking-by-detection strategy, which includes YOLO-based detection and SORT-based tracking. This study describes a bespoke image dataset that was trained using the YOLO algorithm for six distinct classes, and this model was then employed in movies for tracking by the SORT algorithm. For traffic studies, identifying a car or a person in an ongoing video is useful. This paper aims to analyze the topic and get understanding of it [3].

T. -N. Doan et al., [4] proposed overcoming brightness changing, shadows, partially obscured vehicle problem using Kalman filter algorithm. They studied YOLON4, CNN with DeepSORT with limited hardware Configuration on COCO with CIOU high value frame assigning Lacks on environment Condition. This study's main goal is to create an adaptable model that combines YOLOv4 and DeepSORT. With an emphasis on straightforward, efficient algorithms and the advantages of tracking, the new model can recognize objects with high accuracy and quick computation times. The results of our experiments demonstrate that, for the majority of the field scenarios in our dataset, running at a real-time speed of about 32 FPS, our new approach outperforms the previous one by at least 11% of AP and 12% of AP50 [4].

M. Cruz et al., [5] used Three-step approach as people detection, tracking & then people counting. Detection by YOLO v3 pretrained on coco dataset and tracking by DeepSORT. Resulting in 82.76% when restaurant waiting area is not crowded. 66.17 over five days of extensive testing result, which includes extreme Conditions wherein people in the video are densely packed & occluded. The suggested approach involved using video from the facility's current CCTV camera to tally the number of individuals entering and leaving. The dataset has limitations in that it is restricted to a certain geographic and cultural environment. This indicates that most minorities in the area, such men with long hair or blonde hair, have a lower chance of being taken into consideration. Another problem is that the counting system depends on the tracking algorithm. The paper explores the implementation of a vision-based people counting system using surveillance footage from a restaurant establishment. Using Deep SORT and YOLOv3, the system achieves an accuracy of 82.76% in uncrowded conditions and 66.17% in densely packed conditions. Improvements can be made through downsizing frames, retraining models, and exploring other models [5].

M. I. H. Azhar et al., [6] stated that people tracking in crowd surveillance is challenging and needs learned information application. The datasets used is YOLOV3, YOLOV3 tiny, YOLOV3 Custom with different weight size aiming on Occlusion reduction. Powerful GPU deployment, is to be overcome in this aspect of datasets The YOLO approach was used in this study to train a custom image dataset for six separate classes, and the SORT algorithm was then used to track the model in movies. Finding a car or a person in a moving video is helpful for traffic research [6].

K. -H. N. Bui et al., [7] proposed comprehensive framework for multi-class multi-movement vehicle counting across multiple intersections to work with movements and lighting conditions this paper aims to establish a tracking system for NBA and World Cup-related scenes, allowing real-time tracking of each athlete and obtaining relevant track information. The system can also help teachers review competitions and identify student shortcomings. The paper uses cutting-edge deep learning technology, YoloV4, and Deep Sort's advanced version, compared to traditional filtering algorithms, to enhance tracking and detection in the CV field [7].

W. Hou et al., [8] developed Image preprocessing module, Image gray scale, Image binarization and morphological filtering is done at Open CV detection followed by CAMSHIFT algorithm tracking resulting in Gaussian background modelling method tackling Shadow detection with multiple object tracking. Combining motion estimation and Structure information was lacking till models work [8].

Y. Zhang et al., [9] proposed YOLOv4-tiny is a simple neural network architecture for real-time AI platform for detecting people and classification of social distancing based on thermal cameras. This approach is suitable for low-cost embedded devices and is compared to other approaches for real-time detection. The model is applied to videos acquired through thermal cameras for people detection, social distancing classification, and skin temperature measurement. The final prototype algorithm has been deployed in low-cost Nvidia Jetson devices, making it suitable for sustainable smart city surveillance systems. Multi-camera tracking (MCT) has a nascent applicability in real-world applications, but tracking targets across multiple cameras remains a significant computer vision problem. Occlusion, appearance variability, camera motion, and nonrigid object structure are major constraints in MCT. This study provides a comprehensive review of visual object tracking in multi-camera settings, analyzing existing works, adopting problem solving approaches, data association requirements, mutual exclusion constraints, benchmark datasets, and performance metrics. It also examines recent advances and suggests promising future research directions [9].

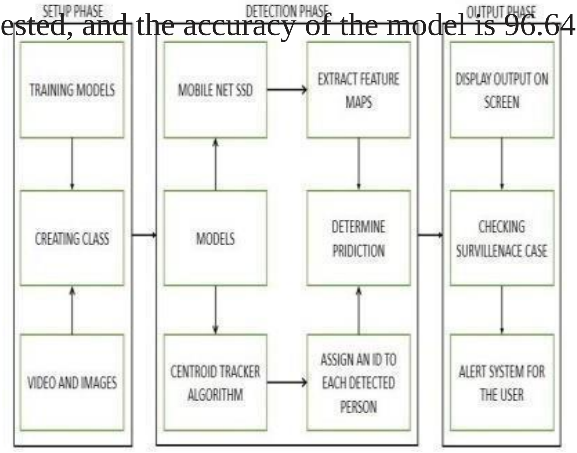
Narinder Singh et al., [10] studied a deep learning-based framework is proposed, using the YOLO v3 object detection model and DeepSORT approach to monitor social distancing. The framework evaluates the mean average accuracy (map), frames per second (FPS), and loss values of the YOLO v3 model against those of other cutting-edge models, such as CNN and SSD. For real-time monitoring, the YOLO v3 with DeepSORT tracking system exhibits the greatest results. [10].

F. Nezhadalinaei et al., [11] worked on KITTI dataset for optimizing motion specify range CRF based spiking Neural networks Synchronous model of tracking for Autonomous driving in two approaches as reinforcement learning. SURF algorithm in semi global stereo visual environment provided better results but exceptions for noisy data. This article introduces an innovative solution to optimize the detection and tracking of motion objects within a specific range of 50 to 80 meters, using the KITTI dataset as a reference. The proposed method, termed CRF-based Deep Spiking Neural Network with Probabilistic Particle Filter (PPF-DSNN), offers a real-time and simultaneous approach to motion object detection and tracking. It leverages a CRF-based deep spiking neural network for feature extraction and employs probabilistic particle filtering techniques for object detection and tracking. The results demonstrate the exceptional efficiency of this approach when compared to existing methods, highlighting its potential to significantly advance the field of Autonomous Vehicles [11].

S. Kumar et al., [12] worked on MOT-A, proposed research work utilizing Kalman Filters for obtaining enhanced results & compared the obtained MOT-A metrics with previous works & results were good. Object detection by Yolo which enable classify objects into 80 classes. object tracking in Kalman filter is used for tracking in the previous frame, & newly detected objects are placed in current frame. All this is done via the DeepSORT, KALMAN Filter improved the accuracy of the proposed model. YOLOv4is used instead of YOLOv3 extension on pretrained COCO model. Zonal Counting by OpenCV library is used for visualizing the zone into the frame & calculate number of objects with unique id.

This research focuses on solving multiple object tracking problems using Kalman filters for enhanced results. The model identifies and tracks objects in a given frame, utilizing the YOLO algorithm to classify objects into 80 classes. Motion Prediction and feature generation are performed using estimation models and Kalman filters. The DeepSORT algorithm is used for tracking and association of newly detected objects. The research aims to find more applications in various domains, such as crowd gathering, animal gathering, and forest rangers [12].

A. KUMAR SINGH et al., [13] proposed detection of Peoples using Single Shot Detector (SSD) replacing VGGI6 with centroid tracker for centroids assigned bounded box. Training & testing having 95.03 TPR and 0.08 FPR making 96.64 % accuracy. Instead of dlib OpenCV was used and SSD to detect in one shot with Mobile Net network Non max suppression and complex dataset augmentation is lacking. This system uses a single shot detector (SSD) mobile net and a centroid tracker to replace the VGG16 base network for better extraction features. The SSD network connects with six convolutional layers for classification. The centroid tracking algorithm uses bounding box coordinates from the SSD to calculate the center of a bounding box, assigning an ID to each person. The highest true positive rate (TPR) of 95.03% and the maximum false positive rate (FPR) of 0.08% are attained by the model using a dataset that contains training and testing data. The provided data set is tested, and the accuracy of the model is 96.64%. [13].



M. Pervaiz et al.,[14] developed model of Gaussian filter and background removal techniques for image approach verification and tracking module Pets Dataset 2009 has been used along S2 dataset. It has drawback as background near objects are omitted out in background removal giving accuracy of 988.14 % for counting 90.14% for tracking. The image was initially preprocessed using a Gaussian filter and background removal methods. Following preprocessing, body point detection and skin verification have been included for human verification. To follow moving objects in video frames for people counting, centroid of silhouettes and jacquard similarity index are created. Experimental findings on the Pets 2009 dataset show that the proposed system outperforms existing state-of-the-art techniques by 8% in terms of tracking accuracy and counting rate. This system should be able to count and monitor persons in crowds with a medium density [14].

Abhinu C G et al.,[15] followed YOLO technology used with the help of Pytorch For object detection, tracking and counting. Extension is as MOT system only detects items that the user needs to detect, which helps to improve system efficiency. In contrast to Yolo's object detection tool, which detects all objects at once. OpenCV is needed to access the camera module, and it also allows us to input video files in various formats. Neural Network (Yolo v5) pretrained Model on COCO dataset for detection. After detection is done, bounded boxes Id are assigned by proposing system using YoLOv5 that can detect objects which were trained and also, they can track and take count of objects in each frame.

The system focuses on object identification, tracking, and counting utilizing Pytorch and YOLO "You Only Look Once" Technology. This MOT system also detects just things that are needed to be detected by the user and helps to improve the efficiency of the system, in contrast to the generic yolo object detection tool which detects all objects at once [15].

Mohan Gowda V et al., [16] stated RCNN used for object detection Similarity and improved Sqrt Cosine MOT benchmark dataset reducing identity switching and fragmentation. The difficulties of long-term occlusion, identity switching, and fragmentation in real-time multi-object identification and tracking have been addressed using a variety of techniques. However, it is still unknown how to decrease fragmentation and the number of identity shifts in multi-object tracking and detection. As a result, uniqueness, and Improved Sqrt Cosine Similarity is used in the second stage to help track the object [16].

Nuha H. Abdulghafoor et al., [17] gave introduction of new algorithm to track and defect objects by integrating essential components by integrating essential components and improve Kalman filter performance implementation on Nvidia platform. Overcoming congested and extremely packed scenes Lighting contrast is a disadvantage. To identify and follow objects in real-time camera-captured natural situations, a novel method is put forward. This approach improves real-time identification and tracking performance by combining deep learning networks with principal component analysis. Comparing the novel algorithm to previous detection and tracking systems, experimental findings demonstrate its efficacy and efficiency in obtaining good detection and classification accuracy. [17].

Gomaa, A. et al., [18] researched on pixel by pixel classification networks and regression networks used for improved detection after K-means clustering refining with YOLOv$_2$ with average time performance of 98 9% to 1.24 frames per second by background subtraction based CNN approach(BS- CNN). The report presents a method for combining detection and MOT algorithms, utilizing Faster RCNN and Kalman Filter for real-time performance. The goal is to detect pedestrians and predict their trajectories, utilizing process MOT (multiple object tracking). Despite various approaches, many issues remain unsolved. The method is demonstrated on four video recordings from a standard dataset, demonstrating the effectiveness of the system. The experimental results on test sequences demonstrate the reliability of the method. The combination of RCNN and Kalman Filter enhances the performance of the system [18].

Kusuma T et al.,[19] stated compressed domain and track moving vectors (MVs) and block coding modes (BCM) from compressed bitstream This paper presents an efficient real-time approach for detecting and counting moving vehicles in videos, based on YOLOv2 and features point motion analysis. The strategy works in two phases: vehicle detection and counting. Different convolutional neural networks are K-means clustering and KLT tracker. The second phase uses temporal information of detection and tracking feature points to assign vehicle labels with

correspondingtrajectoriesandcountthem.

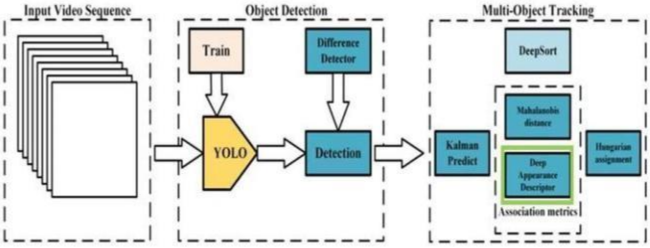
Experimental results show that the proposed scheme generally outperforms state-of-the-art strategies, increasing average time performance by 93.4% and 98.9% [19].

Saponara et al., [20] worked on YOLOv4 tiny along with Bird's-eye view technology provided good performance in terms of accuracy and precision When a video is being sequenced, multiple things tracking is a method of ensuring distinct and consistent ownership of objects. employing the "Spatio Temporal Markov Random Field model (ST-MRF)," a moving tracking sequence of H.264/AVC-compressed footage is created.

The suggested technique monitors moving vectors (MVs) and blocks coding modes (BCMs) from a compressed bitstream and acts on a compressed domain. The findings of this study imply that the volume of the object detection algorithm accurately represents the tracking system's overall effectiveness. Finally, we look into how Deep SORT performance is impacted by the usage of visual definitions during the tracking stage of a tracking system. According to the findings of this study, the volume of the discovery algorithm represents the tracking-by-detection system's overall effectiveness.

The tracking algorithm for object detection in video object tracking is known as "TBD". This method detects moving objects in a frame by associating them with previous frames using a tracking algorithm. Convolutional neural networks have improved object detection accuracy, enabling the use of TBD for multiple object tracking. Motion detection plans can be divided into video pixel domain and compression domain approaches.

The paper focuses on building a novel moving target monitoring model within a compressed domain using a Spatio-Temporal Markov Random fiend (ST-MRF) model. The tracking-by-detection system uses YOLO detection algorithm and Deep SORT tracking algorithm, which does not use appearance information in the tracking stage [20].



II. Methodology Work

Author	Paper Name	Year	Dataset	Algorithm	Limitations/Suggestions
Da Li, Bodong Liang, Weigang Zhang	Real-time Moving Vehicle Detection, Tracking, and Counting System Implemented with OpenCV	2014	KITTI	OpenCV	Computational resources, environmental conditions complex backgrounds and real time constraints.
Fan Bu, Yingjie Cai, Yi Yang	Multiple Object Tracking Based on Faster-RCNN Detector and KCF Tracker	2016	KITTI	Faster R-CNN, Kalman Filter and KCF	Promote accuracy of data assignment and apply Kalman filter prediction algorithm to 2D Slam map
Akansha Bathija, Prof. Grishma Sharma	Visual Object Detection and Tracking using YOLO and SORT	2019	Custom Image dataset	YOLO, SORT	System is limited to detect pedestrian and vehicles
Thanh-Nghi Doan, Minh-Tuyen Trouong	Real-time vehicle detection and counting based on YOLO and Deep SORT	2020	COCO, OpenImage	YOLOv4, Deep SORT	Environmental Conditions such as nighttime and heavy rain
Meygen Cruz, Jefferson James Keh, Ramiel Deticio, Carl Vincent Tan, Elmer Dadios	A People Counting System for Use in CCTV Cameras in Retails	2020	COCO	YOLOv3, DeepSORT	Dependency of the counting system on the tracking algorithm
Muhmand Izham Hadi Azhar, Fadhlan Hafizhelmi Kamaru Zaman, Nooritawati Md. Tahir, Habibah Hashim	People Tracking System Using DeepSORT	2020	YOLOv3, YOLOv3 tiny, YOLOv3 custom	YOLOv3, DeepSORT	Tracking Could be improve by providing a reliable and accurate dataset
Khac-Hoai Nam Bui, Hongsuk Yi, Jiho Cho	A Vehicle Counts by Class Frameworks using Distinguished Regions Tracking at Multiple Intersections	2020	CVPR AI City Challenge 2020 dataset	YOLO, DeepSORT	Determining the locations of distinguished regions to improve the tracking process
Wei Hou, Dongsheng Xia, Hokeyung	Video Road vehicle detection and tracking based on OpenCV	2020	-	OpenCV, CAMSHIFT Algorithm	Difficult in detection of overlapping objects

Jung					
Yao Zhang, Zhiyong Chen, Bohan Wei	A Sport Athlete Object Tracking Based on Deep Sort and Yolo V4 in Case of Camera Movement	2020	INRIA Person Dataset	YOLOv4, DeepSORT	Trying to optimise results in future by using higher algorithm versions.
Narinder Singh Punn,	Monitoring COVID-19 social distancing	2020	COCO	YOLOv3, DeepSORT	Continuous evaluation and improvement, integrating

Sanjay Kumar Sonbhadra, Sonali Agarwal and Gaurav Rai	with person detection and tracking via fine-tuned YOLO v3 and DeepSORT techniques.				with warning systems, redundancy and backup systems.
Fahimeh Nezhadalinaei, Lei Zhang, Mohammad Mahdizadeh, Faezeh Jamshidi	Motion Object Detection and Tracking Optimization in Autonomous Vehicles in Specific Range with Optimized Deep Neural Network	2021	KITTI	CRF-Based Deep Spiking Neural Network with Probabilistic Particle Filter	Videos from urban area, climates issues such as rain, fog or snow
Dr. Shailender Kumar, Vishal, Pranav Sharma, Nitin Pal	Object tracking and counting in a zone using YOLOv4, DeepSORT and TensorFlow	2021	COCO	YOLOv4, DeepSORT	Bad weather, unsuitable lighting conditions during night
Aman Kumar Singh, Dheeraj Singh, Mohit Goyal	People Counting System Using Python	2021	PASCAL	SSD, VGG16, Centroid tracking algorithm	small objects detection
Mahwish Pervaiz, Ahmad Jalal, Kibum Kim	Hybrid Algorithm for Multi People Counting and Tracking for Smart Surveillance	2021	Pets 2009 dataset	Gaussian Filter and background removal techniques, Skin verification and body point detection, Centroid of silhouettes and jacquard similarity index	Resolve occlusion
Abhinu C G, Aswin P, Kiran Krishnan, Bonymol Baby	Multiple Object Tracking using Deep Learning with YOLO V5	2021	COCO	YOLOv5	Limited accuracy in crowded scenes, handling occlusions, training data quality.
Mohan Gowda V, Megha P Arakeri	Real Time Multi-Object Tracking based on Faster RCNN and Improved Deep Appearance Metric	2021	Custom dataset	Faster RCNN, Sqrt cosine similarity	Dark environment conditions
Nuha H. Abdulghafoor,	A novel real-time multiple objects	2022	MOT16	SSD, Kalman Filter	Emergence of noise, change in speed

Hadeel N. Abdullah	detection and tracking framework for different challenges				
Ahmed Gomaa, Tsubasa Minematsu,	Faster CNN-based vehicle detection and counting strategy for fixed camera scenes.	2022	KITTI	YOLOv2, KLT tracker	Limited generalizability, data imbalance, integration with traffic flow management systems.
Motaz M. Abdelwahab, Mohammed Abo-Zahhad, Rin-ichiro Taniguchi					
Kusuma T, Dr. Ashwini K	Multiple Object Tracking using STMRF and YOLOv4 Deep SORT in Surveillance Video	2022	MOT16	YOLOv4, DeepSORT	Accuracy, re-identification accuracy, evaluation metrics.
Sergio Saponara, Abdussalam Elhanashi, Quinghe Zheng	Developing a real-time social distancing detection system based on YOLOv4-tiny and bird-eye view for COVID-19	2022	CrowdHuman	YOLOv4-tiny,	Occlusion handling and crowded area compatible development needed.

III.Future Scope and Challenges

While the YOLO and DeepSORT combination has shown remarkable progress in object detection and tracking, there are ongoing challenges, including handling crowded scenes, object re-identification, and further enhancing tracking performance in complex environments. Future research should focus on addressing these challenges and improving the algorithms' adaptability.

1) Improved Real-time Performance: Enhancing the real-time performance of YOLO and DeepSORT, particularly for applications like autonomous vehicles, where low latency is critical.

This involves optimizing the algorithms and utilizing more efficient hardware acceleration.

2) Robustness in Complex Scenes: focus on making YOLO and DeepSORT more robust in highly dynamic and crowded scenes including addressing challenges like object occlusion, scale variations, object interactions, and diverse environmental conditions.

3) Object Re-Identification: Developing more advanced techniques for object re-identification in scenarios where objects might temporarily leave the frame or change appearance significantly is essential to maintain tracking consistency.

4) Cross-Model Object Tracking: Integrating other sensory data sources, such as LiDAR or radar, with YOLO and DeepSORT can improve tracking performance in scenarios with limited visibility or challenging lighting conditions

Collecting and annotating large and diverse datasets for training and evaluation of YOLO and DeepSORT remains a challenge. The availability of high-quality annotated data for various domains is crucial. Adapting these algorithms to complex real-world scenarios with multiple moving objects, dynamic backgrounds, and changing lighting conditions is an ongoing challenge. Detecting and tracking object interactions, such as pedestrian-vehicle interactions, and analyzing object behaviors for anomaly detection require more advanced models and techniques. Integrating

YOLO and DeepSORT with broader systems, such as traffic management or surveillance, presents integration and interoperability challenges that need to be addressed. Evaluating the algorithms in real-world scenarios to prove

their reliability and effectiveness in various applications is a challenge due to the need for extensive field testing. Future research and development in the field of multiple object detection and tracking using YOLO and DeepSORT should strive to overcome these challenges and leverage the opportunities for innovation and applications in a wide range of industries, ultimately contributing to safer and more efficient technological solutions.

IV. Conclusion

The combination of YOLO (You Only Look Once) and DeepSORT (Deep Learning for Single Object Tracking)

algorithms has emerged as a powerful solution for multiple object detection and tracking in computer vision applications.

This review paper has highlighted the significant contributions and advancements in this field, emphasizing the following key points. YOLO's real-time object

detection capabilities have greatly improved the accuracy and efficiency of identifying and localizing multiple objects within a single frame or image.
Its ability to process images rapidly and simultaneously detect multiple objects in different categories has made it a preferred choice for many applications. DeepSORT extends YOLO's capabilities by incorporating deep learning for object tracking. This allows for the tracking of multiple objects across frames, even in challenging scenarios with occlusions and object interactions.

Deep Sort’s use of Kalman filtering and data association techniques ensures robust and reliable tracking. The YOLO and DeepSORT combination excel in real-time applications, making it suitable for a wide range of applications, including surveillance, autonomous vehicles, and robotics.
The ability to maintain high tracking accuracy while processing video feeds at high frame rates is a significant advantage.

REFERENCES

[1] D. Li, B. Liang and W. Zhang, "Real-time moving vehicle detection, tracking, and counting system implemented with OpenCV," 2014 4th IEEE International Conference on Information Science and Technology, Shenzhen, China, 2014, pp. 631-634, doi: 10.1109/ICIST.2014.6920557.

[2] Bu, Fan, Yingjie Cai and Yi Yang. “Multiple Object Tracking Based on Faster-RCNN Detector and KCF Tracker.” (2016).

[3] Bathija, A. and Sharma, G. (2019), “ Visual Object Detection and Tracking Using Yolo and Sort”. International Journal of Engineering Research Technology, 8, 705-708.

[4] T. -N. Doan and M. -T. Truong, "Real-time vehicle detection and counting based on YOLO and DeepSORT," 2020 12th International Conference on Knowledge and Systems Engineering (KSE), Can Tho, Vietnam, 2020, pp. 67-72, doi: and E. Dadios, "A People Counting System for Use in CCTV Cameras in Retail," 2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), Manila, Philippines, 2020, pp. 1-6, doi: 10.1109/HNICEM51456.2020.9400048.

[6] M. I. H. Azhar, F. H. K. Zaman, N. M. Tahir and H. Hashim, "People Tracking System Using DeepSORT," 2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, 2020, pp. 137-141, doi: 10.1109/ICCSCE50387.2020.9204956.

[7] K. -H. N. Bui, H. Yi and J. Cho, "A Vehicle Counts by Class Framework using Distinguished Regions Tracking at Multiple Intersections," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 2020, pp. 2466-2474, doi: 10.1109/CVPRW50498.2020.00297.

[8] W. Hou, D. Xia and H. Jung, "Video Road vehicle detection and tracking based on OpenCV," 2020 International Conference on Information Science and Education (ICISE-IE), Sanya, China, 2020, pp. 315-318, doi: 10.1109/ICISE51755.2020.00076.

[9] Y. Zhang, Z. Chen and B. Wei, "A Sport Athlete Object Tracking Based on Deep Sort and Yolo V4 in Case of Camera Movement," 2020 IEEE 6th International Conference on Computer and Communications (ICCC), Chengdu, China, 2020, pp. 1312-1316, doi: 10.1109/ICCC51575.2020.9345010.

[10] Punn, Narinder Singh, Sanjay Kumar Sonbhadra, Sonali Agarwal, and Gaurav Rai. "Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques." arXiv preprint arXiv:2005.01385 (2020).

[11] 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