Tracking Socer Player Based on Deepsort Algorithm with YOLOV8 FrameWork

Zein Bilal Khabibullah
Department of Electrical Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
602222201014@student.its.ac.id

Eko Mulyanto Yuniarno
Department of Electrical Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
ekomulyanto@ee.its.ac.id

Reza Fuad Rachmadi
Department of Electrical Engineering
Department of Computer Engineering
Institut Teknologi Sepuluh Nopembe
Surabaya, Indonesia
fuad@its.ac.id

***Abstract*—Tracking is a set procedure that entails assigning an identification to a certain object and subsequently consistently recognizing that object without altering the assigned identification over a sequence of frame images and associating it accordingly.When performing research on object tracking, especially in sports where the object of interest is a human, a resilient technology is necessary to facilitate the tracking process. When the state-of-the-art object detection approach, YOLOV8, is combined with the DeepSORT algorithm, it is anticipated to produce highly accurate and exact outcomes in the tracking and detection of objects. Challenges in multi-object tracking include robustness, oculusion, and identity shifts. In our research, we take advantage of a fusion of YOLOV8 and DeepSORT algorithms to achieve a highly reliable and precise tracking solution. The implementation of the Kalman filter-based motion prediction in DeepSORT allows for the achievement of smooth trajectories, whereas the YOLOV8 deep neural network used assists in precisely recognizing the appearance of objects on the field. The result of our experiment shown the tracking we get is 38% HOTA, 47% DetA, 31% AssA, 68% DetPre, 35% AssRE, 61% AssPr amd 79% LOcA.**

***Index Terms*—Tracking, DeepSORT, YOLO, MOT, Soccer**

I. Introduction

One of the various sports that are widely popular worldwide is soccer [1]. Soccer is a sport that is played between two teams. There are eleven players on each team that can play simultaneously on the soccer field. The objective of this sport is to skillfully maneuver the ball with one's feet across the field and earn points by successfully kicking the ball into the opposing team's goal. This sport is played at a rectangle field within a standard playtime 2 x45 minutes. This sport is organied by FIFA(*Fédération Internationale de Football Association*) the head office of FIFA located in zurich, switchzerland. This organization founded at 1904. The needs of tracking soccer player arise to make a more beter strategy and studying the player performance and behiavour of player during match. Tracking of soccer player in match could help coach to formulate beter strategy and training for thier team.

One of the system that used by FIFA to support the game is VAR(video assistant referee). VAR, or Video Assistant Referee, is a technological equipment that provides additional information to the referee when a judgment needs to be made. This approach employed a group of VAR reviewers in the VAR room. They are observing the match through cameras positioned at multiple locations all over the field, and analyzing the recorded video frame by frame. Through the adopting the VAR principle, we might establish an equivalent system which aids coaches with improving training as well as choosing which strategies are suitable that should be deployed during gameplay according on the player's behavior. The ongoing advancement of technology, particularly in the area of data processing, has a major positive impact on the growth of the sport analytic sector. Soccer is one of the sports that gains from this. There are numerous applications for this technology, one of which is as an automatic tracking system in the field.

Soccer coaches may formulate strategies that correspond with the game. This might be accomplished by utilizing a computer vision field study for machine learning. MOT [2] (Multi Object Tracking) is one area of computer vision that can meet the requirements of this task. As implied by its name, MOT is capable of tracking and simultaneously identifying several objects. The frame ID, object ID, bounding box, confidence score, and any custom data that can be added based on user needs make up the data created by MOT. This unique data is kept in a CSV format. In order to guarantee that MOT has the capability of performing its duties in multiobject tracking, it is necessary to implement a combination of object detection and tracking algorithms. In this particular scenario, we have opted to utilize YOLOv8n [3] [4] for the sole purpose of object detection, while employing DeepSORT for object tracking. A recently published research project, YOLOV8, has demonstrated excellent object detection capabilities with low latency. Additionally, DeepSORT [5] [6] has the advantage of precise tracking for objects that have been identified by YOLO[7]. Prior to initiating the tracking process, it is necessary to complete a series of steps. The initial phase entails identifying the specific object of interest that will be tracked. Once the object of interest has been recognized, it will be assigned an identifier.

This paper trying an approach of combining YOLOV8n as object identifier, which is a one of a object detection module that works on CNN algorithm, this module esepecialy YOLOv8n is the lightweight version of YOLOv8 that designed on minimum limited computational resources but also provid-

How to cite: Khabibullah, Z.B., Yuniarno, E.M., Rachmadi, R.F. (2025). Tracking Socer Player Based on Deepsort Algorithm with YOLOV8 FrameWork. JAREE (Journal on Advanced Research in Electrical Engineering), 9(1).

ing a balanced result of object detection and speed. Also using DeepSORT as object tracker which is an extended version of SORT(Simple Online and Realtime Tracking) framework. DeepSORT which is an tracking algorithm is needed to fulfil the task needed in this job. We also using the dataset that provided by soccernet which in that datset including various scene of soccer matches that happend in european soccer league.

II. Related Work

1) Object based detection: The object detection method can be classified into two categories: classic image processing methods and machine learning-based approaches. classic methods often rely on handmade features such as Histogram of Oriented Gradient (HOG), while machine learning-based methods utilize machine learning techniques. Meanwhile, the current demands and prerequisites of these daily tasks necessitate an object identification method that is capable of effectively handling the work. A commonly used machine learning model is the Convolutional Neural Network (CNN)[8][9]. The Convolutional Neural Network (CNN) consists of two stages: feature extraction and localization. In the feature extraction stage, the network assesses the likelihood of an object being present in a certain frame. In the localization stage, the network determines the object's location inside the frame. In addition, the categorization of deep-learning object identification depends on the number of stages utilized in its process.

The first one is one-stage detector the second one is two stage detector. The two [10] step process on the two stage detector it is potential region within an image generated and then clasification within the proposed region of intrest. YOLOv8[11][12] is a novel and potent model that has been recently developed for the purposes of object detection, classification, and segmentation. Exceptional performance in terms of both speed and accuracy is achieved by this model, which leverages cutting-edge advancements in computer vision and deep learning. After conducting tests on the COCO dataset, it was determined that YOLOv8 outperforms its predecessors in terms of detection speed and accuracy.

A. Object Tracking

Object tracking is a practical application of object detection. Prior to performing object tracking, it is necessary to accurately identify and detect the desired object. The primary goal of object tracking is to assign a unique identity to the thing being tracked. A commonly used method for object tracking is known as SORT (Simple Online and Realtime Tracking). SORT was employed for the purpose of multi-object tracking. SORT prioritizes simplicity and emphasizes real-time performance. The SORT algorithm operates within the framework of the Tracking-by-detection paradigm. In order for SORT to function, it requires an object detection module that is capable of providing it with object detection across the entire video frame.

Object tracking is a one of the resereach subject that have been conducted

III. Methods

On our resereach we using a combination of YOLOV8 and Deepsort to analysis and visualize the performance of object tracking.

A. YOLOV8

YOLO (You Only Look Once) is a cutting-edge real-time object identification system developed by Joseph Redmon. The method known as YOLO using deep neural network not only accurately detect but also precisely locate the target of interest inside a single frame of a picture. This networks objective is to precisely predicting the probability of class and the bounding box which will apear in the frame of the image or video, while continually improving its effectiveness. The resereach that make use of YOLO showing outstanding solution in case of object detection, which makes it applicable to various research areas associated to object detection. In our research, we utilize YOLOv8n, which can be considered a modified version of YOLOv8, which is renowned for its high speed. This model Developed by the Ultralytics team.

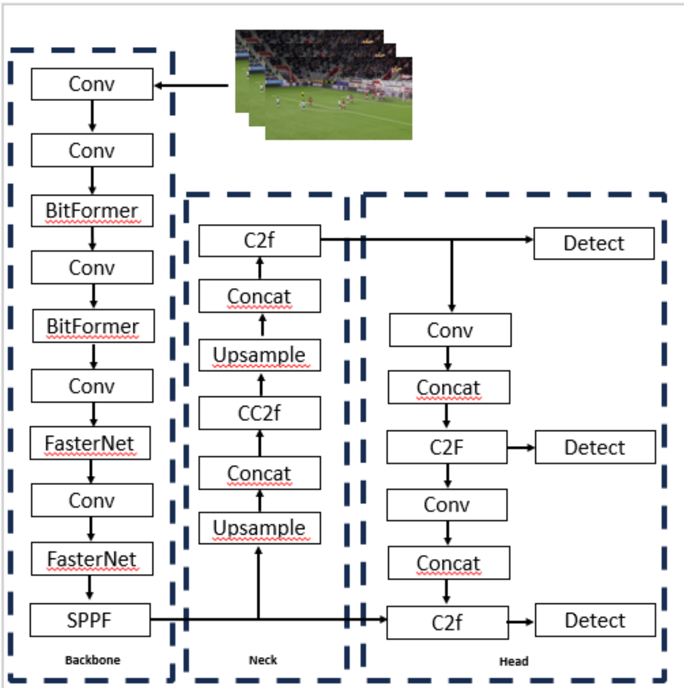

Fig. 1: Architecture YOLOV8n

The figure 1 is showing the architecture of YOLOv8. In the figure 1 the structure of YOLOv8 [13] can be separated into several section that can described as:

- Backbone network: The task of extracting the feature of an image leave at backbone network. The architecture of this network work on the basis of a network called cross-stage partial network(CSPNet) to reducing computing network cost while maintened the accuary.

- Neck: The network that responsible connecting Backbone network and head it called Neck. Neck work as spatial pyramid pooling(SPP) module,which in capturing diffrent scales of feature using diffrent pooling sizes module.
- Detection head: The one who responsible to predicting the components needed for objects detection which is called bounding box and the probability of each class that is showing in frame of picture is named detection head, for each input image or video frame. There are some network that involved in the function of the detection head. The network that responsible for the work of detection head such convulutional layers and anchor boxes.

B. DeepSORT

Simple Online Realtime Tracking, or SORT, was created by A. Bewley et al[14] in for real-time multiple object tracking. The issue of object tracking encompasses various tasks such as detection, estimation of tracks, association of tracks, and creation and removal of track identities. This problem can be effectively addressed by employing the SORT algorithm[15]. The constant velocity Kalman filter[16] is an essential element of the SORT algorithm, which is employed to propagate detections from the current frame to the next frame in order to compute the Intersection over Union (IoU) within the new detection.

One of the most widely used revolutionary object tracking frameworks available today is DeepSORT[17][18], which is an enhanced version of SORT. To generate feature vectors for use as a deep association metric, DeepSORT has incorporated a pre-trained neural network. Considering that DeepSORT[19] was initially developed with a concentration on the Motion Analysis and Re-identification Set (MARS) dataset , a large-scale video-based human reidentification dataset. V. Mandal and Y. Adu-Gyamfi [20]implemented sophisticated object detection and tracking algorithms, which include SORT and DeepSORT, in their study to detect various types of vehicles within their specified area. Nevertheless, they discovered that improvements were required for DeepSORT to effectively track vehicles, as its requirements differ compared to those of human tracking[21]. By taking into account the appropriate modules, the algorithm can be adapted to fulfil their unique requirements. Z. Li, Y. Chen, and Z. Yin [22] have proposed a vehicle tracking method that addresses the issue of vehicle tracking under occlusion through integrating the prior information of the Kalman filter. However, it has been indicated that the suggested approach does not function effectively when the target is absent for an extended duration.

The object tracking effect can be augmented in the presence of occlusion by capturing the evident feature of the target for nearest neighbour matching in the real-time target tracking process. At the same time, the issue of target ID loss is mitigated. The main idea of the algorithm is based around a conventional single-hypothesis tracking method, employing a recursive Kalman filter to determine associations amongst data on a frame-by-frame basis.

C. Evaluation Matrics

We using two diffrent of evaluation matrics in our resereach, since there is two stage of in our resereach that is object detection and then object tracking. We use a Avarang precission, precisison call that using a IoU (Intersection over Union) which after that we need to evaluate the precision of tracking that has been done using deepsort, which it we using HOTA to comparing the ground truth againts the coordinate that we get from tracking using deep sort.

IV. EXPERIMENT AND RESULTS

A. Experimental environtment

The software that we use during the experiment for tracking soccer player include following components: CUDA Version: 12.1, Python 3.10.12, and pytorch 2.1.2. The experiment's training process comprises 200 epochs, with a learning rate of 0.001, batch size of 16, Speed: 0.2ms pre-process, 1.5ms inference, 0.0ms loss, 1.8ms post-process per image

B. Dataset

the dataset we use is a dataset provided by SoccerNET[23]. The dataset consist of a sequance image of soccermatch. the dataset is consisted of 2300 datasets from diffrent match of europe lague. the dataset classes consited of player, sport ball, goalkeeeper and refree. this dataset then splited into three part, that is a 10% test image, 20% validation image, and lastly 70% train image.

C. Evaluation indicator

The model performance need to be evaluarted hance we need evaluation metric to contain it precision, mean average prefcision at threshold In order to evaluate the model performance, the evaluation metrics used in this paper contain Precision, mean average precision at IoU Threshold of 0.5 (mAP50), in addition to the number of floating point operations (GFLOPs), and the size of the model parameters (Params). The average accuracy represents the model's average accuracy across various categories. Reduced floating-point operation number suggests that the model uses less processing power for inference, whereas smaller model parameters suggest that the model is more lightweight. The formula is as stated:

$$precision = \frac{TP}{TP + EP} \tag{1}$$

$$mAP50 = \frac{1}{n} = \sum_{n}^{i=1} \int_{1}^{0} P(r)d(r) \tag{2}$$

TP indicates the number of positive samples that were accurately predicted, whereas FP indicates the number of positive samples that were inaccurately predicted.

D. YOLOV8n detection results

In this section, we will be presenting the findings of our research, specifically the outcomes of our trained model for object detection[24]. In our simulation, we train the model using 300 epochs. The utmost change we can achieve is approximately 240 epochs. The model training technique used in this paper is YOLOV8[25]. This model ensures compatibility with edge computing devices and compact sizes.YOLOv8n employed 168 layers and over 3 million parameters, with a computation cost of approximately 8 GFLOPs. The fig 2 its showing the results of our training validation which is showing the example of image that used on validation process, where the object of interest is detected which in that image shown a object of soccer player and refree.

Fig. 2: Validation batch

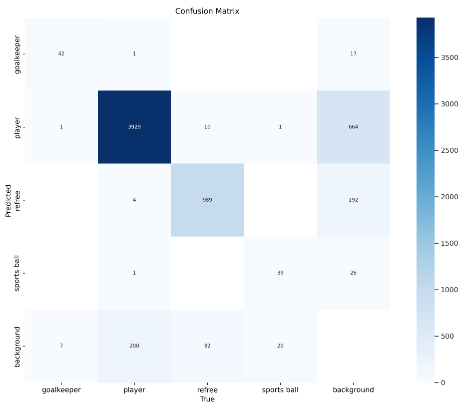

Fig. 3: Confusion matrix

At figure 3 we can see the confusion matrix of our training. This confusion matrix used to visualize the performance of the clacification of object detection. This show the real prediction with real label. At the figure number 3, there are 5 classes which is player, Sports ball, refree, and background. *True Positive* (TP) to0p left quadratic showing a value of 42 that showing how much the right prediction of goal keeper. the bottom left is False negative, which showing how much false object detection from the trained model, the number of failed is 7. in the top right is a value of how much where the goal keeper failed to detected in frame of image. at figure 3 the darker the color showing a higher frequency of TP and TN. THis show the model frequently make the right detection within the object of intrest.

Figure 4 is the result of training model yolov8n, there show the result of training where it is the value precision is 75.4%, precision 75.4%, recal 77%, and then MAP50 is 82.3

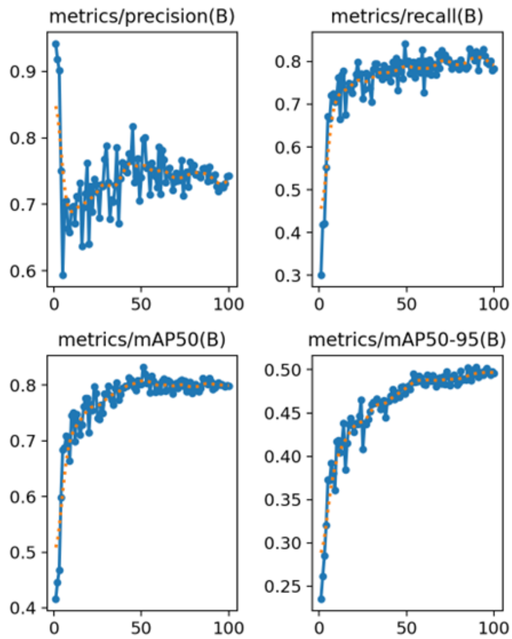

Fig. 4: Training Result

E. Deepsort object tracking results

In this section, we will be presenting the findings of our research, specifically the outcomes of our trained model for object tracking. The performance of object tracking is evaluated using a evaluation modul called HOTA matrics that matric consisting of HOTA, DetA, AssA, DetRe, DetPr, AssRe, AssPR, and LocA. At Figure 5 its the results of our tracking modul at tracking soccer player. The process of evaluating the result we comparing a ground truth 0f already prepared dataset for testing the program that consisted of diffrent number of match, the data we use for testing tracking modul of this resereach is cosisted of 50 sets of datasets which every datasets consisted of 750 continous frame of soccer match and thier repctive ground truth. This figure is showing the results of the tracking we get is 38% HOTA, 47% DetA, 31% AssA, 68% DetPre, 35% AssRE, 61% AssPr amd 79% LoCA. The x-axis is showing the value results of the detection menwhile the y-axis is showing the confidance(alpha) of the tracking module.

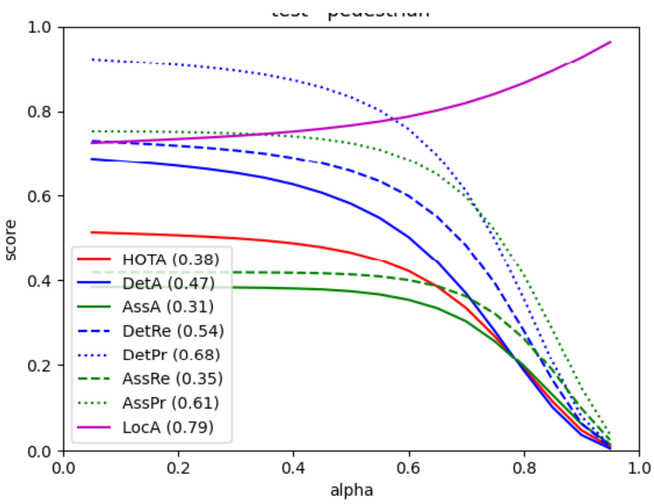

Fig. 5: HOTA result

V. Conclusion

In this paper, we propose a YOLOv8n as object detection and Deepsort as a object tracking. the result we get is the object detection of YOLOV8n getting a result of 75% in precision, meanwhile the results of the object tracking with DeepSORT model getting a low results of HOTA that is a 38% the Detection accuary (DetA) is aorund 47%, 31% AssA, 68% DetPre, 35% AssRE, 61% AssPr amd 79% LoCA. This result might happend because the lack of variation of datasets or also the lack of a modul to determine the color of the uniform of the soccer player, goal keeper and the refree. Beside that there also a problem with ID asignment because the lack of unique Appearance feature because the smilar uniform. we will considering adding a variety number of datasets also adding uniform detection module to ensuring precise id labeling. Compared to the resereach of DeepPlayer-Track: Player and Referee Tracking With Jersey Color Recognition in Soccer[6] that have been done by Banoth Thulasya Naik, and co. The method we use showing a lower results at resereach they do they achieving results of MOTA 94%. This might happend because they adding a foccused clacification of jersey color so in case resulting higher performance in object detection and tracking.

Acknowledgment

Thanks to datsets that have been provided by SoccerNET team. So this resereach can be finished

References

[1] Donald T Kirkendall. "Evolution of soccer as a research topic". In: *Progress in Cardiovascular Diseases* 63.6 (2020), pp. 723–729.

[2] Silvio Giancola et al. "SoccerNet 2022 Challenges Results". In: *Proceedings of the 5th International ACM Workshop on Multimedia Content Analysis in Sports*. ACM, 2022. DOI: 10.1145/3552437.3558545. URL: https://doi.org/10.1145%2F3552437.3558545.

[3] Xiaoxiao Feng, Anhu Ren, and Hua Qi. "Improved Highway Vehicle Detection Algorithm for YOLOv8n". In: *2023 9th International Conference on Mechanical and Electronics Engineering (ICMEE)*. 2023, pp. 444–448. DOI: 10.1109/ICMEE59781.2023.10525289.

[4] Aura Syafa Aprillia Radim et al. "Detecting Vehicles using YOLOv8n in Edge Computing Dashcam". In: *2023 3rd International Conference on Intelligent Cybernetics Technology Applications (ICICyTA)*. 2023, pp. 89–94. DOI: 10.1109/ICICyTA60173.2023.10428952.

[5] Yunlong Wang et al. "Underwater Target Tracking Technology Based on YOLO v4 and Deepsort". In: *2023 IEEE International Conference on Image Processing and Computer Applications (ICIPCA)*. 2023, pp. 238–241. DOI: 10.1109/ICIPCA59209.2023.10257818.

[6] Banoth Thulasya Naik et al. "DeepPlayer-Track: Player and Referee Tracking With Jersey Color Recognition in Soccer". In: *IEEE Access* 10 (2022), pp. 32494–32509. DOI: 10.1109/ACCESS.2022.3161441.

[7] A Anish et al. "Enhancing Surveillance Systems with YOLO Algorithm for Real-Time Object Detection and Tracking". In: *2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS)*. 2023, pp. 1254–1257. DOI: 10.1109/ICACRS58579.2023.10404710.

[8] Behnam Farsi et al. "On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach". In: *IEEE access* 9 (2021), pp. 31191–31212.

[9] G. Priyadharshini and D. Raveena Judie Dolly. "Comparative Investigations on Tomato Leaf Disease Detection and Classification Using CNN, R-CNN, Fast R-CNN and Faster R-CNN". In: *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*. Vol. 1. 2023, pp. 1540–1545. DOI: 10.1109/ICACCS57279.2023.10112860.

[10] Kavitha R and Nivetha S. "Pothole and Object Detection for an Autonomous Vehicle Using YOLO". In: *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*.

2021, pp. 1585–1589. DOI: 10.1109/ICICCS51141.2021.9432186.

[11] Jialiang Yuan et al. "FEB-YOLOv8: A Steel Surface Defect Detection Algorithm Based on Improved YOLOv8s". In: *2023 3rd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*. 2023, pp. 629–633. DOI: 10.1109/CEI60616.2023.10527815.

[12] Jin Miao et al. "An Electric Grid Sign Detection Method Based on the Improved YOLOv8n". In: *2024 10th International Symposium on System Security, Safety, and Reliability (ISSSR)*. 2024, pp. 220–223. DOI: 10.1109/ISSSR61934.2024.00033.

[13] Juan Terven, Diana-Margarita Córdova-Esparza, and Julio-Alejandro Romero-González. "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas". In: *Machine Learning and Knowledge Extraction* 5.4 (2023), pp. 1680–1716.

[14] Nicolai Wojke, Alex Bewley, and Dietrich Paulus. "Simple online and realtime tracking with a deep association metric". In: *2017 IEEE International Conference on Image Processing (ICIP)*. 2017, pp. 3645–3649. DOI: 10.1109/ICIP.2017.8296962.

[15] Marcellino Marcellino et al. "Comparative of Advanced Sorting Algorithms (Quick Sort, Heap Sort, Merge Sort, Intro Sort, Radix Sort) Based on Time and Memory Usage". In: *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)*. Vol. 1. 2021, pp. 154–160. DOI: 10.1109/ICCSAI53272.2021.9609715.

[16] Priya Shree Madhukar and L.B. Prasad. "State Estimation using Extended Kalman Filter and Unscented Kalman Filter". In: *2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3)*. 2020, pp. 1–4. DOI: 10.1109/ICONC345789.2020.9117536.

[17] Yuqiao Gai, Weiyang He, and Zilong Zhou. "Pedestrian Target Tracking Based On DeepSORT With YOLOv5". In: *2021 2nd International Conference on Computer Engineering and Intelligent Control (ICCEIC)*. 2021, pp. 1–5. DOI: 10.1109/ICCEIC54227.2021.00008.

[18] Zo.A.T. Rakotoniaina et al. "LIV-DeepSORT: Optimized DeepSORT for Multiple Object Tracking in Autonomous Vehicles Using Camera and LiDAR Data Fusion". In: *2023 IEEE Intelligent Vehicles Symposium (IV)*. 2023, pp. 1–7. DOI: 10.1109/IV55152.2023.10186759.

[19] Zipei Pan et al. "Research on Volleyball Players Tracking Based on Improved DeepSORT". In: *2022 4th International Conference on Communications, Information System and Computer Engineering (CISCE)*. 2022, pp. 591–595. DOI: 10.1109/CISCE55963.2022.9851084.

[20] Vishal Mandal, Lan Uong, and Yaw Adu-Gyamfi. "Automated Road Crack Detection Using Deep Convolutional Neural Networks". In: *2018 IEEE International Conference on Big Data (Big Data)*. 2018, pp. 5212–5215. DOI: 10.1109/BigData.2018.8622327.

[21] Jialue Fan et al. "Human Tracking Using Convolutional Neural Networks". In: *IEEE Transactions on Neural Networks* 21.10 (2010), pp. 1610–1623. DOI: 10.1109/TNN.2010.2066286.

[22] Chuan Hu, Yimin Chen, and Junmin Wang. "Fuzzy observer-based transitional path-tracking control for autonomous vehicles". In: *IEEE Transactions on Intelligent Transportation Systems* 22.5 (2020), pp. 3078–3088.

[23] Anthony Cioppa et al. "SoccerNet-Tracking: Multiple Object Tracking Dataset and Benchmark in Soccer Videos". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 3491–3502.

[24] M. Rizk and I. Bayad. "Human Detection in Thermal Images Using YOLOv8 for Search and Rescue Missions". In: *2023 Seventh International Conference on Advances in Biomedical Engineering (ICABME)*. 2023, pp. 210–215. DOI: 10.1109/ICABME59496.2023.10293139.

[25] Bowen Yan, Ying Wang, and Yongze Liu. "Improved cartoon face object detector based on YOLOv8". In: *2024 4th International Conference on Neural Networks, Information and Communication Engineering (NNICE)*. 2024, pp. 864–867. DOI: 10.1109/NNICE61279.2024.10498701.