IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

DETECTING AND TRACKING VARIOUS OBJECTS IN A VIDEO STREAM

[1]Dr.P.Babu, [2]SK.Asiff
[1]Faculty, [2]Faculty
[1]Computer Science Department,
[1]Geethanjali Institute of Science and Technology, Nellore.

Abstract: The main objective of object detection is to detect various objects in a video stream with more accuracy and with less computation time. Image classification involves assigning a class label to an image, whereas object localization involves drawing a bounding box around one or more objects in an image. The object detection algorithm YOLO to examine the entire image in a single instance and predicts the bounding box coordinates and class probabilities for these boxes. This project uses Yolo algorithm with Sort tracker to detect and track various objects in a video stream derived from a Webcam and through the pre-recorded video. From the experimental study, the accuracy of object detection varies from 72% to 97%. The existing algorithms take more time to process each frame and accuracy is also less. The Intersection over Union will decide prediction of each object as Good one. The Nonmax suppression technique will get a single prediction per object. The biggest advantage of using YOLOv7 can process 155 FPS. The experiment was conducted in VScode . The video taken from various places were used to examine the effectiveness of the proposed approach. The experimental results show that our proposed approach can effectively improve the accuracy rate in detection of various objects in an image and also track it. with many applications including surveillance, robotics, autonomous vehicles, and augmented reality.

***Index Terms* -Object detection and tracking, computer vision, yolo model, deep learning.**

I. INTRODUCTION

The main goal of this thesis is to detect and track various objects in a video stream. Object detection and tracking are two important tasks in computer vision that involve identifying and locating objects within an image or video stream, and then keeping track of those objects as they move or change position.Object detection refers to the process of locating and identifying objects within an image or video frame, and is typically done using machine learning algorithms such as convolutional neural networks (CNNs) or Haar cascades. These algorithms analyze the visual features of the image, such as edges, textures, and shapes, and use them to recognize specific objects or classes of objects.Object tracking, on the other hand, involves keeping track of an object over time as it moves within a video stream. This can be challenging due to variations in lighting, changes in object appearance, occlusion by other objects, and other factors that can make it difficult to maintain a consistent tracking algorithm. To address these challenges, tracking algorithms often use a combination of visual features, motion models, and machine learning techniques to accurately track objects and predict their future positions.Together, object detection and tracking are used in a wide range of applications, from autonomous vehicles and surveillance systems to robotics and augmented reality. By enabling computers to detect and track objects in real time, these technologies are helping to transform the way we interact with the world around us.

Object detection and tracking are important tasks in computer vision with many real world applications. The main motivation behind object detection and tracking is to enable machines to recognize and understand the visual world around them.One of the key applications of object detection is in autonomous driving. Self-driving cars need to be able to detect and recognize objects in their environment, such as other cars, pedestrians, traffic signs, and traffic lights, in order to make decisions about how to navigate the road safely.Object detection and tracking are also used in surveillance systems for security purposes. By detecting and tracking objects such as people, vehicles, and suspicious packages, security personnel can quickly respond to potential threats.

In addition, object detection and tracking have applications in sports analysis, robotics, and healthcare. For example, in sports analysis, computer vision systems can track the movements of athletes during a game or match to provide coaches with valuable insights into their performance. In healthcare, object detection and tracking can be used to track the movement of medical instruments during surgical procedures.Overall, the motivation behind object detection and tracking is to enable machines to better understand the visual world around them, which has many practical applications in various fields.

II. LITERATURE REVIEW

The literature survey focuses on various approaches and methods related to object detection and tracking, specifically in the context of on-road environments and aerial images. The surveyed papers address the challenges and advancements in detecting and recognizing different objects, such as pedestrians, vehicles, traffic signs, and license plates, using deep learning-based techniques. Here is a summary of the key findings:

Nguyen et al. [1] proposed a framework for robust on-road pedestrian and vehicle detection, recognition, and tracking. The approach combines deep learning with multiple local patterns and depth information, achieving good real-time performance and potential for extension to other object detection tasks.

Deng H. Sun et al. [2] developed a region-based CNN method for detecting vehicles in aerial images. They employed two CNNs to improve accuracy for small object detection, but false detections remained challenging. Future work includes mining hard negative samples and reducing computation time.

Y. Cai et al. [3] presented a scene-adaptive vehicle detection algorithm using the Bagging mechanism. The approach incorporates deep auto-encoder-based feature similarity calculation and achieved a high vehicle detection rate. However, the confidence assignment method lacks a theoretical basis.

Cao et al. [4] proposed a novel framework for deep neural networks that simultaneously performs detection and recognition. Their method achieves significant speed improvements while maintaining object detection performance.

Sommer et al. [5] applied Fast R-CNN and Faster R-CNN for vehicle detection in aerial images. They found that Faster R-CNN, combined with the Region Proportional Network, provided the best detection results. Future work involves examining the generalizability and adapting the frameworks for rotated bounding boxes.

Nguyen et al. [6] introduced a real-time and robust region proposal network for detecting various vehicles in different driving conditions. The system aimed to improve the performance of Advanced Driver Assistant Systems (ADAS) and could be extended to detecting other obstacles like traffic signs and lights.

Zhu et al. [7] developed a system using the UavCT dataset to estimate real-world city traffic density. Their deep vehicle counting framework (DVCF) achieved vehicle detection, classification, tracking, and counting. The system integrated conventional vision-based algorithms with deep learning-based approaches.

Geng et al. [8] focused on human gesture detection in infrared images using an improved YOLO-V3 network. They enhanced convolutional feature propagation using DenseNet blocks, resulting in better detection performance, especially in low visibility and similar color conditions.

Avramovic et al. [9] proposed a region-of-interest based approach combined with the YOLO architecture for high-performance object detection of various objects.

Tourani et al. [10] developed a real-time License Plate Detection (LPD) and Character Recognition (CR) system for Iranian vehicles. They used two sequential YOLO v.3 deep networks trained on various weather and illumination conditions, achieving high accuracy and real-time performance.

Li et al. [11] introduced ASPP-CenterNet, an approach that incorporates atrous spatial pyramid pooling (ASPP) with CenterNet for road object detection. The method improved detection accuracy for small objects with a similar inference time, making it suitable for autonomous driving applications.

Overall, the literature survey demonstrates the advancements in object detection and tracking using deep learning techniques and highlights the ongoing challenges in achieving high accuracy, real-time performance, and adaptability to different environmental conditions.

III. PROPOSED METHODOLOGY

The problem statement for object detection and tracking is to develop algorithms and models that can accurately identify and locate objects of interest within a given scene, and then track their movements over time. This involves analyzing images or video streams to detect the presence of objects, and then continuously tracking them as they move or interact with their environment. The goal is to enable machines to understand and interact with the world around them, allowing for a wide range of applications such as autonomous driving, surveillance, robotics, and augmented reality. The main challenges in this field include dealing with occlusions, changes in lighting and perspective, and accurately distinguishing between different objects that may have similar appearance.

In this project, we propose a new algorithm to detect and track the various objects in a video stream or else in a live video (through webcam) for object detection we have used a algorithm called YOLO V7 it will detect the objects in a frame with the help of COCO dataset which contains nearly 80 objects over 3,20,000 plus images which are already pretrained .Than the detected images will get labelling's and bounding boxes with accuracy .And SORT algorithm to track the objects in a video stream. The proposed model achieves the highest accuracy of 93.78% and also low accuracy of 72.87% .In this system we can do both object detection and tracking by assigning labels and bounding boxes with accuracy if you want you can just detect the object or else if we want just only track the object or else we can add the labels for object and also remove the labels if you don't want and also you can save the processed video if you want or else leave it without saving and also detect only the specific objects in a video if you want or else change the colour of your labels if you don't like it and also give the unique colour to specific object that are present in your dataset and the main advantage of our proposed system is it will detect and track various objects even in a pre-recorded video.

YOLOv7 provides a greatly improved real-time object detection accuracy without increasing the inference costs. As previously shown in the benchmarks, when compared to other known object detectors, YOLOv7 can effectively reduce about 40% parameters and 50% computation of state- of-the-art real-time object detections, and achieve faster inference speed and higher detection accuracy. In general, YOLOv7 provides a faster and stronger network architecture that provides a more effective feature integration method, more accurate object detection performance, a more robust loss function, and an increased label assignment and model training efficiency. As a result, YOLOv7 requires several times cheaper computing hardware than other deep learning models. It can be trained much faster on small datasets without any pre-trained weights.

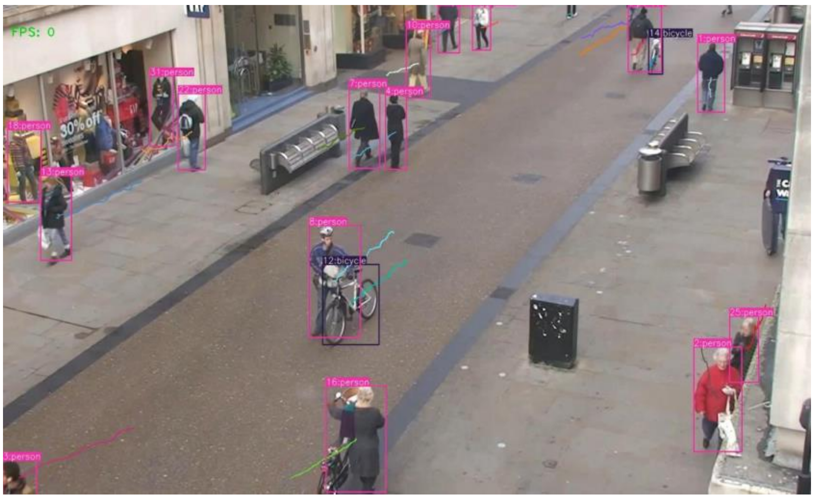

Fig 1. unique tracking color for each object in a video stream

IV. CONCLUSION

In this project, we used one of the most popular implementations of Yolo V7 algorithm with SORT tracker to detect and track various objects in a pre-recorded or through live video by wecam with less computation time and more accuracy. It could process 155 frames per second. YOLO V7 is fast and easy to install and process. The experiment was conducted on a pre-recorded video and also live video by webcam with the objects that are presented in COCO

dataset captured through a webcam and as well as pre-recorded video and it could identify the various objects present in the video, with more detection accuracy in real time and also track those objects. during the in different conditions (clear ,noisy image, dark image,etc.,), during the night, and with different distance from the camera, ranging upto 30m to 300m by assigning class labels and with bounding boxes and also accuracy.

REFERENCES

1) V. D. Nguyen, H. Van Nguyen, D. T. Tran, S. J. Lee and J. W. Jeon, "Learning Framework for Robust Obstacle Detection, Recognition, and Tracking," In IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 6, pp. 1633-1646, June 2017, doi: 10.1109/TITS.2016.2614818.
2) Deng, H. Sun, S. Zhou, J. Zhao and H. Zou, "Toward Fast and Accurate Vehicle Detection in Aerial Images Using Coupled Region-Based Convolutional Neural Networks," In IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 10, no. 8, pp. 3652-3664, Aug. 2017. doi: 10.1109/ JSTARS. 2017.2694890.
3) Y. Cai, H. Wang, Z. Zheng and X. Sun, "Scene-Adaptive Vehicle Detection Algorithm Based on Composite Deep Structure," In IEEE Access, vol. 5, pp. 22804-22811, 2017.doi: 10.1109/ACCESS.2017.2756081.
4) W. Cao, J. Yuan, Z. He, Z. Zhang and Z. He, "Fast Deep Neural Networks with Knowledge Guided Training and Predicted Regions of Interests for Real-Time Video Object Detection," In IEEE Access, vol. 6, pp. 8990-8999, 2018, doi: 10.1109/ACCESS.2018.2795798.
5) L. Sommer, T. Schuchert and J. Beyerer, "Comprehensive Analysis of Deep Learning-Based Vehicle Detection in Aerial Images," In IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 9, pp. 2733-2747, Sept. 2019, doi: 10.1109/TCSVT.2018.2874396.
6) V. D. Nguyen, D. T. Tran, J. Y. Byun and J. W. Jeon, "Real-Time Vehicle Detection Using an Effective Region Proposal-Based Depth and 3-Channel Pattern," In IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 10, pp. 3634-3646, Oct. 2019, doi: 10.1109/TITS.2018.2877200.
7) J. Zhu et al., "Urban Traffic Density Estimation Based on Ultrahigh-Resolution UAV Video and Deep Neural Network," In IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 12, pp. 4968-4981, Dec. 2018, doi: 10.1109/JSTARS.2018.2879368.
8) K. Geng and G. Yin, "Using Deep Learning in Infrared Images to Enable Human Gesture Recognition for Autonomous Vehicles," In IEEE Access, vol. 8, pp. 88227-88240, 2020. doi: 10.1109/ACCESS.2020.2990636.
9) A. Avramović, D. Sluga, D. Tabernik, D. Skočaj, V. Stojnić and N. Ilc, "Neural-Network-Based Traffic Sign Detection and Recognition in High-Definition Images Using Region Focusing and Parallelization," In IEEE Access, vol. 8, pp. 189855-189868, 2020, doi:10.1109/ACCESS.2020.3031191.
10) A. Tourani, A. Shahbahrami, S. Soroori, S. Khazaee and C. Y. Suen, "A Robust Deep Learning Approach for Automatic Iranian Vehicle License Plate Detection and Recognition for Surveillance Systems," In IEEE Access, vol. 8, pp. 201317-201330, 2020, doi: 10.1109/ACCESS.2020.3035992.
11) G. Li, H. Xie, W. Yan, Y. Chang and X. Qu, "Detection of Road Objects With Small Appearance in Images for Autonomous Driving in Various Traffic Situations Using a Deep Learning Based Approach," In IEEE Access, vol. 8, pp. 211164-211172, 2020, doi: 10.1109/ACCESS.2020.3036620.