IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134116

MULTIPLE-OBJECTS ANNOTATION AND LOCALIZATION USING YOLO

Janardhan K[1], Bharath Kumar Reddy B[2], Sushmitha R[3], Nageswari G[4], Dharma Teja B[5]

Assistant Professor, Department of Computer Science and Engineering, Rajeev Gandhi Memorial College of Engineering and Technology, Nandyal, Andhra Pradesh.[1]

Students, Department of Computer Science and Engineering, Rajeev Gandhi Memorial College of Engineering and Technology, Nandyal, Andhra Pradesh.[2-5]

Abstract: There have been significant strides in computer vision that result in momentous improvements in object detection and tracking, which form the basis of a number of applications such as surveillance, driverless vehicles and human-computer interaction. This paper proposes an original but complicated method for reliable and precise tracking based on DeepSORT (Deep Simple Online and Realtime Tracking) with YOLOv5 (You Only Look Once version 5). YOLOv5 is an effective detector that performs object detection by looking once on an image or video frame to identify objects as well as their locations. These detection results are then incorporated into the DeepSORT tracking framework, which employs deep learning techniques to consistently track objects across frames. The combination of YOLOv5 and DeepSORT addresses issues of accuracy in detecting as well as reliability in following objects thereby providing a holistic approach to dynamic scenes involving multiple objects. The proposed system detects many different yolov5s and DeepSORT at one time.

Keywords: YOLOv5, DeepSORT.

I. INTRODUCTION

Tracking is a crucial task in deep learning that involves predicting object locations in a video using visual cues from each frame. It requires detecting objects initially, assigning unique identities to them, and continuously monitoring their movement across frames while maintaining their IDs. Based on the tracking method and the number of targets tracked simultaneously, trackers can be classified into single object trackers and multiple object trackers. Single object trackers focus on monitoring a specific object in a sequence of images, ignoring other objects in the scene. They start by identifying the object's position in the first frame and then track it continuously throughout the video. These trackers are known for their efficiency, relying mainly on classic computer vision algorithms like CSRT and KCF.

In contrast to single-object trackers, multiple object trackers (MOTs) excel in following multiple objects in a single frame. Trained on vast datasets, MOTs offer high accuracy, allowing them to track multiple objects from different classes efficiently. Prominent MOT algorithms like DeepSORT, JDE, and CenterTrack effectively handle real-world challenges, making them highly versatile solutions for object tracking.

There are two main ways to track objects: Tracking by detection: This method uses object detectors to find objects in each frame of a video. Then, it links the objects between frames to create tracks. This approach is good for tracking multiple objects and objects that come and go, even if the object detector doesn't work perfectly. Tracking without detection: This method initializes the positions of objects and tracks them over time. It's mainly used in older computer vision algorithms and requires manual input to initialize the tracks. In our plan, choosed DeepSORT as the tracker. It's an updated version of the SORT tracker. SORT is great for tracking precisely and accurately. In the existing system, it can struggle with changing identities too often and not being able to see objects that are blocked from view. DeepSORT fixes these problems by using a better way to match objects together that combines both how they move and what they look like. This new algorithm follows objects not just by watching how they move but also by comparing how they look, which makes the tracking more precise.

In the proposed system use a combination of YOLOv5, OpenCV, and PyTorch to find and follow objects in videos. YOLOv5, which is a deep learning classifier made with PyTorch, is used to find objects. After that, OpenCV is used to give the algorithm videos in real time or as files. The algorithm then follows the objects that were found in the videos. Our method shows how well these cutting-edge algorithms work together to track objects in videos.

DOI: 10.17148/IJARCCE.2024.134116

II. LITERATURE SURVEY

U. P. Nagane and Dr. A. O. Mulani are described about the process of identifying moving objects in a series of video frames is called object tracking. In the domains of computer vision, motion-based recognition, automated surveillance, traffic monitoring, augmented reality, and object-based video compression, among others, real-time object tracking is a difficult challenge. Higher level performance is significantly depends on object detection working accurately. The object detection method is being designed and implemented on a variety of platforms. It covers things like openCV, MATLAB, Simulink, and C programming. Because of its many features, MATLAB programming is the most widely used of these. Easy programming, a collection of toolboxes and Simulink blocks covering many technological domains, and matrix data processing are some of these capabilities. This project shows how to use MATLAB for object tracking and detection.

Shraddha Mane and others are described Object tracking and recognition are essential parts of a computer vision algorithm. Robust object detection is a challenge because the scenes are different. Keeping track of the object when it is occluded is the most challenging challenge. As a result, this method uses the TensorFlow object detection API to recognize moving objects. The position of the found object is subsequently fed into the object tracking algorithm. A novel CNN-based object tracking technique is used for accurate object detection. The recommended approach can recognize the object in a variety of occlusion and lighting scenarios. Using the proposed approach on self-generated image sequences, the accuracy was 90.88%. Over the past few years, deep learning has had a significant impact on how society is adjusting to artificial intelligence. Chandan G, Ayush Jain, Harsh Jain, Mohana and others are described that several widely used object detection algorithms include You Only Look Once (YOLO), Single Shot Detector (SSD), Faster-RCNN, and Region-based Convolutional Neural Networks (RCNN). Of these, SSD and Faster-RCNN perform better in terms of accuracy, although YOLO works better when speed is prioritized over accuracy. Deep learning integrates SSD and Mobile Nets to carry out tracking and detecting tasks effectively. This technique maintains performance without sacrificing efficiency in object detection.

Durriya Bandukwala, Muskan Momin, Akmal Khan, Aasim Khan, Dr. Lutful Islam and others are describes a method for identifying and monitoring four categories of vehicles through the analysis of road crossing video footage: cars, buses, trucks, and motorcycles. Vehicle route accounting's initial results show that the method is very promising, with good results in several circumstances, but much more research is required to strengthen these systems' resistance to occlusions and other unanticipated incidents.

III. PROBLEM STATEMENT

From the above studies of the literature survey In the existing system had found a problem statement that is some of them are used to detect only one class and some others are used to detect only vehicles and some are used to detect only four types of vehicles. Our problem statement is multiple classes detection at a time.

IV. OBJECTIVE

In the proposed system use the YOLOv5 algorithm for the object detection and DeepSORT algorithm for the tracking of the detected object.

V. METHODOLOGY

YOLOv5

YOLOv5 belongs to the famous YOLO (You Only Look Once) computer vision model family and is a real-time object detection model. When it comes to images or video frames, YOLO models are meant for effective detection and identification of objects in them. Ultralytics came up with YOLOv5 and released it in 2020. It was able to make notable improvements over its predecessor, YOLOv4, such as speed, accuracy and being user-friendly. It can be mentioned that there are number of platforms where it's easy to use YOLOv5 for real time object detection because of its simplicity, high performance and portability.

YOLOv5 uses a new method for generating the anchor boxes, called "dynamic anchor boxes". The system begins by taking input as video or providing the already recorded video for the detection. After that the input data is processed into the different number of frames for detection. For the task of the detection YOLOv5 uses the dataset or the model to detect the object in the video that is provided as the input for the system. By using the model after detecting the objects the detected objects are represented by using the bounding boxes around the object that are detected. By using the OpenCV In the proposed system can access the input video. The camera or video feeds input, which is processed and divided into individual images (frames).

DOI: 10.17148/IJARCCE.2024.134116

Each frame is then passed through the YOLO (You Only Look Once) detection algorithm, which uses a pre-trained model to classify objects within the frame. The model can be predefined model or can create custom dataset models for detection. The proposed system uses the predefined model. Once the detection is done, it is bounded with the bounding boxes where the object is found.

DeepSORT

DeepSORT, short for Deep literacy for Object Tracking in videotape Sequences, is an advanced tool in the world of computer vision, particularly for tracking objects in videos. It's like having asuper-smart operative who can keep an eye on effects in a videotape. Developed by experimenters, DeepSORT is part of the family of deep literacy models designed to track objects in real- time. Just like YOLOv5, DeepSORT is each about spotting objects in images or videotape frames and keeping track of them as they move around. When you feed a videotape into DeepSORT, it breaks it down into individual frames and starts to dissect each bone.

It's like taking each shot of the videotape and looking for objects in them. DeepSORT uses a combination of ways, including deep literacy, to identify and track these objects. It's really good at this because it learns from lots of exemplifications, just like a operative literacy from once cases. Once it spots an object, it draws a box around it to show where it's in the frame. This helps us understand what is passing in the videotape, like spotting buses on a road or people in a crowd. Overall, DeepSORT is like having a keen- eyed operative helping us make sense of what is passing in vids by keeping tabs on all the moving objects.

VI. DATASET

The proposed system uses the dataset called COCO (Common Objects in Context) dataset, which is a widely used dataset for the object detection and segmentation tasks in the computer vision.

The COCO dataset contains images that depict our everyday scenes in context. It includes 80 different classes of objects such as person, car , dog, bus among others. The dataset is designed for covering a wide range of object types and scenarios. Each image in the COCO dataset is densely annotated with bounding boxes around the objects. These annotations make the COCO dataset suitable for a variety of computer vision tasks beyond just object detection.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

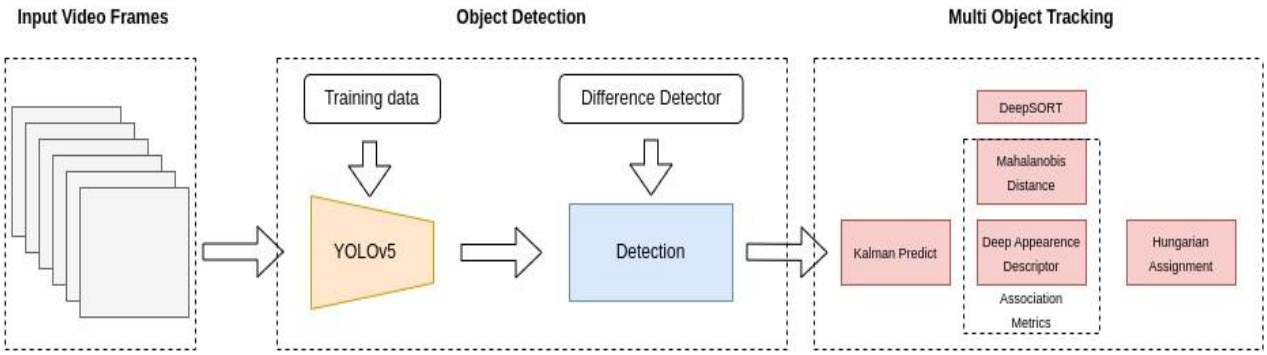
DOI: 10.17148/IJARCCE.2024.134116

Img: Different classes of the COCO dataset

The COCO dataset is known known for its challenges, such as object scale variation, occlusion, and cluttered scenes. These challenges reflect real-world scenarios and help to evaluate the robustness and generalization capabilities of object detection models.

VII. WORKING PRINCIPLE

YOLOv5 is used to detect the objects and DeepSORT is used to track the objects that are detected by the detection algorithm YOLOv5. The detection algorithm operates by dividing the input image into a grid. Every object has its centre. The centre of the image which falls in the grid the grid is responsible for the whole object detection . It first takes an image from the video as the input. The input image/frame passes through a deep neural network, which predicts bounding boxes for the objects present in the video. The YOLOv5 algorithm outputs bounding box coordinates for each of the detected object in the frame.



Img: Flow diagram for the object detection and tacking

Initially the YOLOv5 algorithm detects the objects in each frame of the input (input video file or builtin camera) video stream. DeepSORT extracts deep features from the detected bounding boxes using a pretrained convolutional neural network. These features encode the appearance of the object and are used to match the detections to existing tracks. DeepSORT mainly depends on the Data Assosiation. Which means the DeepSORT associates the detected objects with the existing tracks using the combination of appearance matching and the motion prediction. This association is typically performed using techniques such as Hungarian algorithm, which minimizes the overall cost of assigning detections to tracks. DeepSORT maintains tracks over time, updating the state estimates of the existing tracks based on the appearance and the future velocity or motion of the object and creating the new tracks for the unassociated detections. Tracks may also be terminated if they no longer associated with any detections for a certain number of frames. Finally the DeepSORT outputs the tracks of the objects. The final output consists of tracked objects, including their IDs, bounding boxes and class labels for each of the object in the frame.

VIII. RESULT

The YOLOv5 algorithm detects the objects fastly and accurately which is followed by the DeepSORT that can track the detected objects very fastly and accurately across the frames of the input video stream.

```
Frame 0 Done. YOLO-time:(0.279s) SORT-time:(0.153s)
Frame 2 Done. YOLO-time:(0.184s) SORT-time:(0.155s)
Frame 4 Done. YOLO-time:(0.209s) SORT-time:(0.111s)
Frame 6 Done. YOLO-time:(0.202s) SORT-time:(0.125s)
Frame 8 Done. YOLO-time:(0.209s) SORT-time:(0.113s)
Frame 10 Done. YOLO-time:(0.208s) SORT-time:(0.097s)
Frame 12 Done. YOLO-time:(0.187s) SORT-time:(0.100s)
Frame 14 Done. YOLO-time:(0.192s) SORT-time:(0.076s)
Frame 16 Done. YOLO-time:(0.184s) SORT-time:(0.055s)
Frame 18 Done. YOLO-time:(0.185s) SORT-time:(0.071s)
Frame 20 Done. YOLO-time:(0.191s) SORT-time:(0.089s)
Frame 22 Done. YOLO-time:(0.188s) SORT-time:(0.107s)
Frame 24 Done. YOLO-time:(0.197s) SORT-time:(0.090s)
```

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

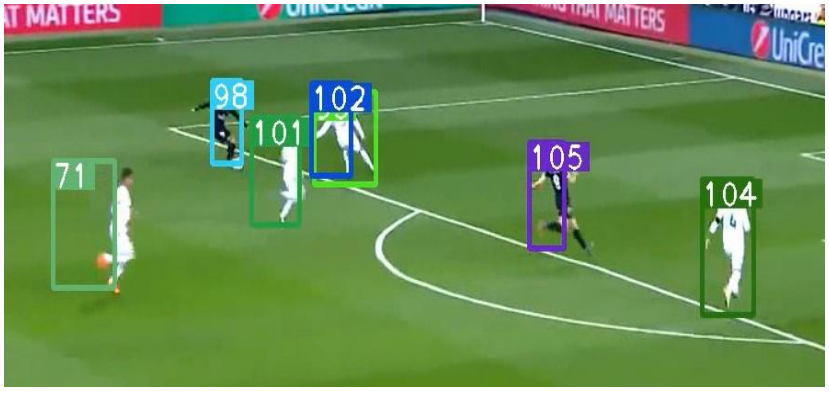
International Journal of Advanced Research in Computer and Communication Engineering

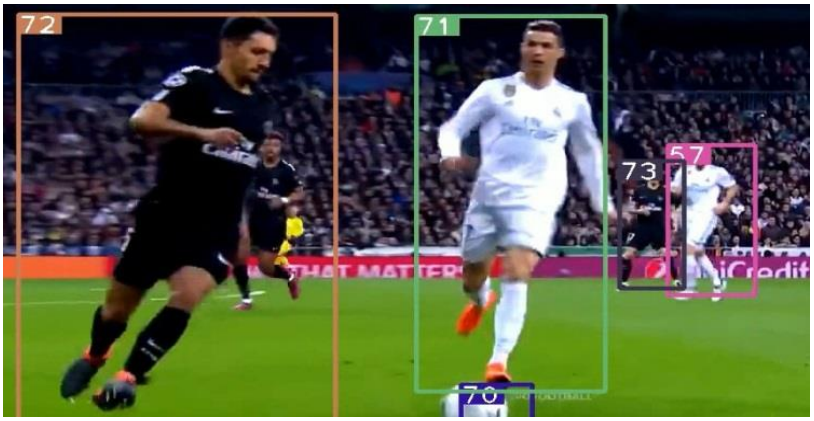
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134116

The above image specifies the time taken for the object detection that is specified as the YOLO-time and the time taken for the tracking that is specified as the SORT-time.

The time taken for the each object is not same because of the some of the objects has some changes in its appearance and velocity.





The above images indicates that an unique id is assigned for each of the object and based on that id the objects are tracked across the frames.

IX. CONCLUSION AND FUTURE SCOPE

The proposed system detects and tracks the 80 different classes across the video using the deep neural algorithms YOLOv5 and the DeepSORT. The Future scope include it can be extended more than the 80 different classes and can be trained with the custom dataset.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134116

REFERENCES

[1]. U. P. Nagane and A. O. Mulani, “Moving Object Detection and Tracking Using Matlab”, Journal of Science and Technology, , Vol. 06, Special Issue 01, August 2021, pp63-66.

[2]. S. Mane and S. Mangale, "Moving Object Detection and Tracking Using Convolutional Neural Networks," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2018, pp. 1809-1813, doi: 10.1109/ICCONS.2018.8662921.

[3]. G. Chandan, A. Jain, H. Jain and Mohana, "Real Time Object Detection and Tracking Using Deep Learning and OpenCV," 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2018, pp. 1305-1308, doi: 10.1109/ICIRCA.2018.8597266.

[4]. Manjunath Jogin, Mohana, “Feature extraction using Convolution Neural Networks (CNN) and Deep Learning” 2018 IEEE International Conference On Recent Trends In Electronics Information Communication Technology,(RTEICT) 2018, India.

[5]. A.O.Mulani and Dr.P.B.Mane, “Watermarking and Cryptography Based Image Authentication on Reconfigurable Platform”, Bulletin of Electrical Engineering and Informatics, Vol.6 No.2, pp 181-187, 2017.

[6]. Wang N., S. Li, A. Gupta, and D. Yeung, “Transferring rich feature hierarchies for robust visual tracking”. Computing Research Repository a. 2015, abs/1501.04587.

