IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

3d Object Detection And Tracking Using Yolov5

[1] Mrs. Uma Mythraye, [2]Shaik Gousiya Begum
[1]Assistant Professor, [2] Assistant Professor
[1]Department of Artificial Intelligence,
[1]Vidya Jyothi Institute of Technology, Hyderabad, India

Abstract—Object detection has developed rapidly in recent years and is being used in modern times. Applications like Biometric Recognition, Surveillance, Industrial Inception, Content-based image retrieval, Robotics, Medical Analysis, Lane Detection mainly uses object detection. Our model mainly includes detecting and tracking vehicles, passengers, cattle on road. Road accidents have been rapidly growing throughout the years in many countries. One of the reasons is cattle on roads. As a solution, we propose a system where in the model takes traffic video as input and detects and tracks objects, detects Humans, Animals, Traffic lights, Vehicles and many other Objects finding the count of the objects per class. For the proposed system, we have used methods like mathematical modeling, parametric and non-parametric methods. Among the non-parametric processes, Machine Learning (ML) is the most famous one. In this project, we will be working on You Only Look Once version-5 for object detection (e.g., bicycle, motorbike, cars) YOLOv5 uses the CSPDarknet53 variant (open-source neural network in C). The model is tested on real-time videos. The environment is greatly influenced by livestock. It is our goal to deal with two issues: 1) managing traffic by managing cattle on the road, and 2) saving cattle's lives. This work will assist the city's traffic management system monitoring and locating livestock. This system has the major advantage of requiring no additional infrastructure and working with the existing one.

IndexTerms - Biometric Recognition, Surveillance, You Only Look Once version-5, CSPDarknet53.

I.INTRODUCTION

In computer vision, 3D object recognition involves recognizing and determining 3D information, such as the pose, volume, or shape, of user-chosen 3D objects in a photograph or range scan. Typically, an example of the object to be recognized is presented to a vision system in a controlled environment, and then for an arbitrary input such as a video stream, the system locates the previously presented object. This can be done either off-line, or in real-time. The algorithms for solving this problem are specialized for locating a single pre-identified object, and can be contrasted with algorithms which operate on general classes of objects, such as face recognition systems or 3D generic object recognition. Due to the low cost and ease of acquiring photographs, a significant amount of research has been devoted to 3D object recognition in photographs.

1.1 Objective

The Main Objective is to detect Humans, Animals, Traffic lights, Vehicles and many other Objects from the Traffic CCTV Camera Feed Using YOLOv5 Algorithm. Vehicular object detection is the heart of any intelligent traffic system. It is essential for urban traffic management. R-CNN, Fast R-CNN, Faster R-CNN and YOLO were some of the earlier state-of-the-art models. Region based CNN methods have the problem of higher inference time which makes it unrealistic to use the model in real time. YOLO on the other hand struggles to detect small objects that appear in groups. To locate and classify vehicular objects from a given densely crowded image using YOLOv5. The model can be implemented in the street for real-time traffic detection which can be used for traffic control and data collection. To implement an intelligent traffic control system reduces congestion; operating costs, increase the capacity of the infrastructure and provide a cattle detection system.

RELATED WORK

2.1.1 Prabhu, Understanding of Convolutional Neural Network (CNN) — Deep Learning, 4th May 2018.[1]

Social networks and microblogging sites such as Twitter are widespread amongst all generations nowadays where people connect and share their feelings, emotions, pursuits etc. Depression, one of the most common mental disorders, is an acute state of sadness where a person loses interest in all activities. If not treated immediately this can result in dire consequences such as death. In this era of the virtual world, people are more comfortable in expressing their emotions on such sites as they have become a part and parcel of everyday lives.

2.1.2 Nilesh R. Mate, Intelligent Transportation System in India- A Review, Journal of Development Management and Communication, Volume II, No.3, ISSN 2348-7739, July-September, 2015.[2]

This paper aims to build a structured literature review of the field of Intelligent Transportation Systems (ITS). In this literature review an effort made to critically evaluate the earlier research work and methodologies related with Intelligent Transportation Systems to study, analyze and evaluate its relevance in today's changing environment. The recently published research articles related to ITS are undertaken for the study.

2.1.3 Shaoqing Ren et al. \Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". In: IEEE Transactions on Pattern Analysis and Machine Intelligence 39.6 (2017).

This paper introduces the Region Proposal Network (RPN), which allows for cost-free region proposals in object detection networks [4]. The RPN is a fully convolutional network that predicts object bounds and scores for each position. It is trained end-to-end to generate high-quality region proposals for detection. By sharing convolutional features, the RPN and Fast R-CNN can be merged into a single network that achieves state-of-the-art object detection accuracy on multiple datasets. This system has a frame rate of 5 fps on a GPU and has been used in winning entries in several competitions. Code for this system is available publicly.

2.1.4 Joseph Redmon et al. \You Only Look Once: United, Real-Time Object Detection". In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.

This paper proposed the YOLO approach to object detection frames the problem as a regression task, predicting bounding boxes and class probabilities directly from full images [5]. With a single neural network for the entire detection pipeline, YOLO can be optimized end-to-end for detection performance. YOLO is extremely fast, with a base model processing images in real-time at 45 frames per second and a smaller version, Fast YOLO, achieving 155 frames per second with double the mAP of other real-time detectors. Although YOLO makes more localization errors, it has fewer false positives in the background and learns very general object representations, outperforming other detection methods in cross-domain scenarios.

PROPOSED METHODOLOGY

The methodology for 3D object detection and tracking typically involves several steps, which can vary depending on the specific system and application. Here is a general overview of the methodology:

1. **Data Acquisition**: The first step is to collect data from sensors, such as LIDAR, cameras, or radar. This data can include point clouds, images, or other sensor measurements.
2. **Object Detection:** The next step is to detect objects in the sensor data. This can be done using various techniques, such as deep learning-based object detection models, segmentation-based methods, or clustering algorithms.
3. **Object Tracking**: Once objects are detected, the next step is to track them over time. This can be done using motion-based tracking methods, such as Kalman filters or particle filters, or deep learning-based tracking models.
4. **Object Classification:** After objects are detected and tracked, they can be classified into different categories, such as cars, pedestrians, or bicycles. This can be done using deep learning-based classification models or rule-based methods.

5. **3D Localization**: In order to determine the 3D location of objects, sensor data from multiple sources may be fused together. This can be done using techniques such as triangulation, stereo vision, or sensor fusion algorithms.
6. **Mapping:** The 3D location of objects can then be used to create a map of the environment. This can be useful for applications such as autonomous driving or robotics.
7. **Decision Making**: Finally, the information gathered from the above steps can be used to make decisions, such as controlling a vehicle or alerting a driver to potential hazards.

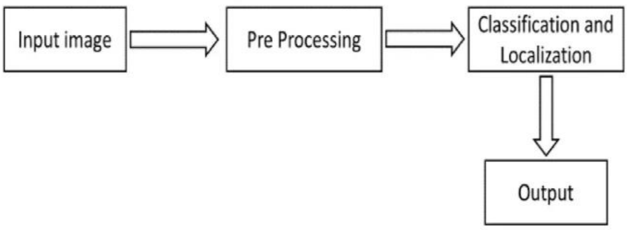
Overall, the methodology for 3D object detection and tracking involves a combination of sensor data processing, deep learning, and machine vision techniques to detect, track, and classify objects in the environment. The output of the system can be used to improve safety, efficiency, and automation in a variety of applications.

3.1 System Architecture

Here's an explanation of the four steps in 3D object detection and tracking:

Input: The first step in the process is to provide a video as input. This video is typically captured by a camera or other imaging device, and it contains a stream of images that will be processed to detect and track objects.

Preprocessing: Once the input video is obtained, it must be preprocessed before object detection and tracking



can begin. Preprocessing involves tasks such as resizing the video frames, normalizing pixel values, and converting the video to grayscale or other color spaces as needed. Preprocessing ensures that the video frames are in a suitable format for analysis and helps to improve the accuracy and speed of object detection and tracking.

Classification and localization: The next step is to classify and localize objects in the video frames. This is typically done using a deep learning model such as YOLOv5. The model is trained to detect and classify objects in the video frames and to output their location information in 3D space. The classification and localization step is the core of the 3D object detection and tracking process, as it identifies and tracks the objects of interest in the video.

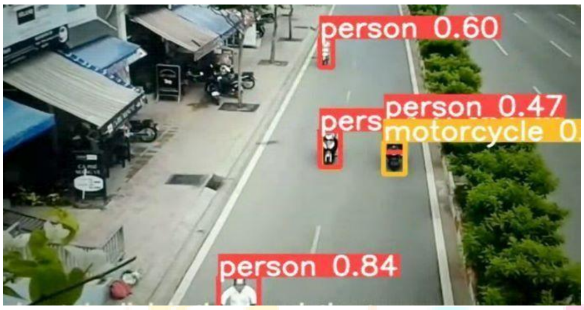
Output: The final step is to output the results of the object detection and tracking process. This may involve displaying the location of objects in the video frames, tracking the movement of objects over time, or providing other information about the objects detected. The output may be in the form of visualizations such as plots or images, or it may be fed into another system for further processing.

IV. RESULTS AND DISCUSSION

4.1 Introduction

3D object detection and tracking in computer vision involve identifying and locating objects in a 3D space using sensors such as cameras, LiDAR, and radar. The goal is to estimate an object's 3D bounding box, which includes its position, orientation, and dimensions. Accuracy can be measured using evaluation metrics such as AP, mAP, and IoU. High accuracy is crucial for applications such as autonomous driving, robotics, and augmented reality as it improves safety, efficiency, and enables advanced features like object recognition and classification.

Detecting the object Output Image/Video



Showing the detected object with score

From the above figure, the object detection model identifies persons with a confidence of 60-84 percent and motorcycles with a confidence of 60-80 percent in a bounding box. which is output from the YOLOv5 model. Identification of type of vehicle is significant during challan generation.

CONCLUSION

We have done 3D object detection which is very useful in the current scenario and also in future. It is used to detect objects, humans, cattle and other things in the given feed. It is used for detection and tracking. It has many applications like biometric recognition, Surveillance, Industrial Inception, Content-based image retrieval, Lane Detection, Smart traffic management system and many more. We have developed it in the context of detecting the objects in traffic feed.

We used YOLOv5 for object detection. YOLO v5 uses the CSPDarknet53 variant (open-source neural network in C), which initially has a 53-layer system trained on ImageNet**.** We applied the transfer learning concept to optimize the model by using pre-trained convolutional weights.

The primary requirement for the traffic system is monitoring the traffic. So, we proposed a model which takes traffic live feed video as input and detects objects, finds the count of the object per class, detection and tracking of objects, cattle, human and other objects. We have used the COCO dataset for the object detection model and customized our dataset for detection of classes. We trained the model on these datasets and tested it on real-time videos. We used a Yolov5 object detection algorithm to get detections from video.

References

[1] Ross Girshick et al. \Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation". In:2014 IEEE Conference on Computer Vision and Pattern Recognition.2014, pp.580{587. doi:10.1109/CVPR.2014.81.

[2] Nilesh R. Mate, Intelligent Transportation System in India- A Review, Journal of Development Management and Communication, Volume II, No.3, ISSN 2348-7739, July-September, 2015 .

[3] Prabhu, Understanding of Convolutional Neural Network (CNN) — Deep Learning, 4th May 2018.

[4] Shaoqing Ren et al. \Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". In: IEEE Transactions on Pattern Analysis and Machine Intelligence 39.6 (2017), pp. 1137{1149. doi:10.1109/TPAMI.2016.2577031.

[5] Joseph Redmon et al. \You Only Look Once: United, Real-Time Object Detection". In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016, pp. 779{788. doi: 10.1109/CVPR.2016.91.

[6] Natthariya Laopracha and Khamron Sunat. \Comparative Study of Computational Time that HOG-Based Features Used for Vehicle Detection".In: Recent Advances in Information and Communication Technology 2017.Ed. by Phayung Meesad, Sunantha Sodsee, and HerwigUnger. Cham:Springer, 2018, pp. 275{284. isbn: 978-3-319-60663-7.

[7] Geethapriya S, N. Duraimurugan, S.P. Chokkalingam, "Real-Time Object Detection with Yolo", International Journal of Engineering and Advanced Technology (IJEAT)

[8] Astha Gautam, Anjana Kumari, Pankaj Singh: "The Concept of Object Recognition", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 3, March 2015.

[9] V. Gajjar, A. Gurnani and Y. Khandhedia, "Human Detection and Tracking for Video Surveillance: A Cognitive Science Approach," in 2017 IEEE International Conference on Computer Vision Workshops, 2017.

[10] Alexe, B., Deselaers, T., and Ferrari, V. (2010). "What is an object?" in ComputerVision and Pattern Recognition (CVPR), 2010 IEEE Conference on (San Francisco,CA: IEEE), 73–80. doi:10.1109/CVPR.2010.5540226

[11] Wei Jia, Shiquan Xu, Zhen Liang, Yang Zhao, Hai Min, Shujie Li, and Ye Yu. Real-time automatic helmet detection of motorcyclists in urban traffic using improved YOLOv5 detector. IET Image Processing, page ipr2.12295, jun 2021

[12] Chenyi Chen, Ming-Yu Liu, Oncel Tuzel, and Jianxiong Xiao. R-CNN for Small Object Detection., 10115 LNCS:214–230, nov 2016.

[13] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "Tph-yolov5: Improved yolov5 based on transformer prediction head for object detection on drone captured scenarios," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, October 2021, pp. 2778–2788