Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, June : 2023

Detecting Objects and Estimating Distance using YOLO

A. Sai Pranathi, Ch. Sampath Yadav, C. Rakshitha, Mr. M. Sandeep Kumar (Asst. Professor)
Department of Electronics and Communications Engineering,
Vignana Bharathi Institute of Technology, Hyderabad, Telangana, India
Email Id: saipranathiarroju@gmail.com, cheemalasampath349@gmail.com, chinthojurakshitha@gmail.com

Abstract: ***Our method for measuring the distance between objects using a camera with a monocular lens is a substantial improvement in the field of computer vision. By increasing the prediction vectors and combining the weights of the underlying structure and bounding box regressor, we have created a fully integrated architecture that is both efficient and effective. Our updated loss function, which includes a component for distance estimation, further improves the accuracy of our model. We have also explored two different methods for handling distance estimation: class-agnostic and class-aware. According to our tests, class-agnostic generates fewer prediction vectors and provides superior outcomes against class-aware. This approach has potential applications in various fields, including autonomous driving, robotics, and surveillance systems. Overall, our method represents a significant step forward in detection of objects and estimating distance using monocular cameras. The level of accuracy of the bounding box capability has been increased as a result of the merging of the object identification and distance measurement subtasks.. Our proposed scheme, tested on the KITTI dataset, has yielded a mean relative error of only 11% spanning all eight divisions and an assortment of distances of [0, 150] m. This result demonstrates that our approach is highly competitive with existing methods. Moreover, we have found that our solution does not compromise on inference speed, maintaining an identical rate of 45 frames per second compared to the unmodified YOLO. By combining these two subtasks in a synergistic manner, we have achieved an impressive level of accuracy without sacrificing performance. This approach holds great promise for applications that require both object detection and distance measurement capabilities.***
Keywords : YOLO, distance estimate, object detection, monocular camera

I. INTRODUCTION

Locating and recognising things in images and videos is the challenge of identifying objects in computer vision. It is an essential part of many applications, including as robots, autonomous vehicles, and surveillance. The ability of various nautonomous tools for moving about in a natural atmosphere depends on distance estimate, which is a crucial component of 3D scene identification and orientation. In addition to active capturing devices like lidar or sonar, these devices can also include passive capturing devices like RGB or IR cameras. In contrast to passive devices, which require complex mathematical approaches for computer vision on top to estimate the approximate distance, active devices are more costly but immediately give distance estimation as a cloud of points. The fastest onestage object identification algorithm, YOLOv3, has been chosen for this instance.

II. PROPOSED ALGORITHM

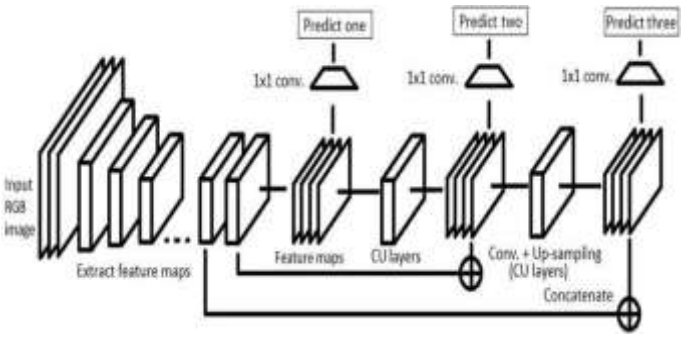
Yolo algorithm: The Yolo algorithm proposes using an end-to-end neural network to predict bounding boxes and class probabilities simultaneously. We think YOLO can be educated to aid with the regression problem of absolute distance measurements since it conducts regression on BB coordinates. The network is designed to use the same internal features in both BB coordinates and line - of - sight prediction, resulting in synergy. It encourages us to train the model from scratch and add the distance measurement component within the YOLO design. The assumption is validated when a distance estimate methodology yields a greater BB accuracy than the baseline YOLO.

The bounding box of an object is determined using YOLO, which then assigns the object to one of the established classes. It is unable to calculate the object's distance by default. Our main objective is to build Dist-YOLO by giving YOLO this capability while maintaining its original qualities. Three fundamental steps must be taken in order to obtain this capability: (A) Add distance-related data to the training dataset labels. (b) Extend each cell's prediction to determine an object's distance. (c) To

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, June : 2023

take object distance into account, alter the YOLOv3 loss function that was utilised during training.

The YOLO (You Only Look Once) image detector uses an entire convolutional framework and is a one-stage, multi-scale, anchor-based object detector.



Fig(1): convolutional neural network

DIST-YOLO

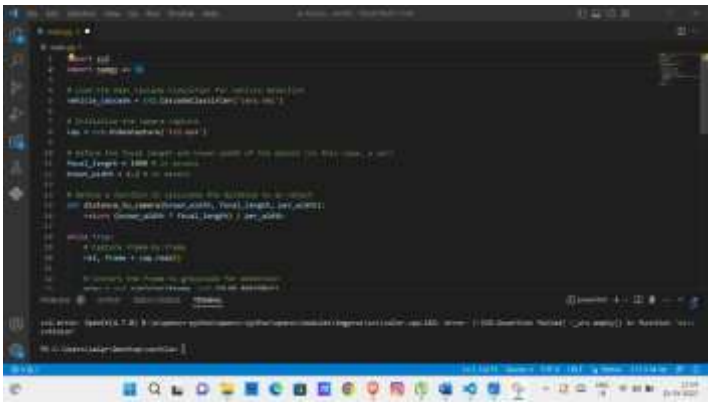
Dist-YOLO is a unique architecture that we define, where the distance information is added to the prediction vectors created by the heads and coupled with the appropriate distance loss function. We demonstrate that YOLO, despite having the same backbone capacity as the original YOLO, recognizes bounding boxes more correctly. We show that a monocular camera equipped with YOLO can accurately calculate an object's distance.

YOLO unifies the entire process into a single design rather than using a region proposal network. It has a decoder and an encoder in it. The encoder serves as a feature extractor. The Darknet-53 backbone serves as its representation, although any backbone can be used. The characteristics are encoded into three output grids using three successive scales. Each grid has cells, and each cell is in charge of identifying items whose centers are within that cell. The regression on the coordinates of the bounding box, alongside class and confidence, is referred to as "detecting". Thus every nucleus in YOLOv3 is capable of identifying three items, and the problem of detecting three similar things is mitigated by the use of anchors.

An anchor is a typical bounding box that was extracted from training data during the preprocessing procedure using the k-means method. A cell is given three anchors during training and inference, and When the intersect over union (IoU) over the box and the anchor are maximised, it recognises a piece of data in its final position. YOLO is much faster than two-stage detectors but results in less accurate detection.

Distances can be calculated and data can be enhanced using lidar data from datasets like KITTI, Waymo Open Dataset, Berkeley DeepDrive Dataset, or nuScenes, among others. When distance is included in the prediction vector, it employs the same features as the bounding box regressor, and those features can be trained to minimize the loss of calculating the distance. This distinguishes it from other YOLO variants that rely on assumptions derived from a training dataset algorithm for determining distance.

III. EXPERIMENT AND RESULT



Fig(2): Programming code for YOLO

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, June : 2023

Fig (3): Included .xml file in code

Fig (4): Input video of vehicles moving on road

Fig (5): Output - object detection and distance estimation

IV.CONCLUSION

We began by using the quick object detector with one stage YOLOv3. We demonstrated During the literature review that while YOLOv3 had been modified for monocular absolute distance estimation, the architecture of YOLO had not yet been completely merged with the distance estimation capability. We did this by boosting the resultant prediction vector, changing the absolute distance estimation factor of the loss function, and explaining how to change the training data. Subsequently, we observed that the streaming analytics for both YOLOv3 and Dist-YOLOv3 was the same.

V.REFERENCE

[1] Rukhovich, D.; Mouritzen, D.; Kaestner, R.; Rufli, M.; Velizhev, A. Estimation of Absolute Scale in Monocular SLAM Using Synthetic Data. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea, 27–28 October 2019, pp. 803–812.

[2] Kumari, S.; Jha, R.R.; Bhavsar, A.; Nigam, A. Autodepth: Single image depth map estimation via residual cnn encoderdecoder and stacked hourglass. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 340–344.

[3]. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.