IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Special Issue-14, May 2019 E-ISSN: 2347-2693

# YOLO Based Object Detection Using Drone

**Shiva Kumar R Naik[1*], Kushal A[2], Lakshmi Narayan S[3], Sreeraam V Chatrapathi[4], Sagar T[5]**

[1,2,3,4,5]School of Computing & Information Technology, REVA University, Bangalore, India

*Corresponding Author: shivakumarnaik@reva.edu.in*

DOI: https://doi.org/10.26438/ijcse/v7si14.181184 | Available online at: www.ijcseonline.org

***Abstract***—The headway of convolutional neural systems (CNNs) and Deep learning (DL) in the previous decade brought about significant upgrades in computer vision. One of the recipients of these advances is the task of object detection, where the goal is to distinguish and locate real-world objects inside pictures or videos. Real-time object tracking on a drone under a dynamic situation has been a difficult issue for a long time, with existing methodologies utilizing off-line calculation or powerful computation units on board. This paper displays lightweight real-time on board object tracking methodology, which varies, from basic image classification in that the AI demonstrate needs to distinguish numerous objects in a single frame, and figure out where these objects are found. The advances in procedures, joined with the improved PC equipment, put real-time object detection well inside the capacities of present day processors. Real-time object recognition is essential for some utilization of Unmanned Aerial Vehicles (UAVs), for example, observation and reconnaissance, search-and-rescue, and foundation assessment. In the previous couple of years, Convolutional Neural Networks (CNNs) have ascended as an unbelievable class of models for recognizing picture content, and are seen as the standard strategy for generally issues.

***Keywords*—**Detection, Drone, Pattern Matching, Privacy Preserving, Security Vulnerabilities, Sensitive Items, Yolo

## I. INTRODUCTION

Lately, with the quick advancement of handling Information technology, astute transportation frameworks have become an imperative method for current traffic management and an inevitable trend. In the meantime with the innovation development and market advancement of UAV (“Unmanned Aerial Vehicle”), which has qualities of being lightweight adaptable and cheap. In rundown, vehicle recognition for aerial imagery assumes an imperative job in engineering applications.

In addition, the innovation depends on artificial intelligence, image processing, machine vision, and other disciplines; In addition, it is also a typical application of interdisciplinary research. Along these lines, it additionally has imperative research criticalness in academics. There are wide scope of uses for unmanned aerial vehicles (UAVs) in the field of civil engineering. A couple of applications include incorporate yet are not constrained to coastline perception, fire fighting, checking vegetation development, glacial perceptions, riverbank degradation surveys, 3D Mapping, woodland observation, Emergency Response & Disaster Management, power line reconnaissance, Infrastructure assessment and monitoring traffic. Drone technology has an incredible effect to simplify our everyday lives and limit issues. This innovation is approaching for reconnaissance, perception, finding and controlling crimes. This is likewise being utilized for civil, military, airborne photography, investigation, agriculture and a lot more purposes. Working guideline depends on PC vision, AI and control framework and so on. This is an airplane without an individual so it is great to use at hazardous and inaccessible spots. No dread of life by using this in basic circumstances and it likewise can also be utilized to investigate certain territories where subsistence of human is beyond imagination.

As UAV, Applications become broad; a more elevated amount of self-sufficiency is required to guarantee its well-being and operational productivity. In a perfect world, an autonomous UAV on a very basic level relies upon microprocessors, sensors, and on-board aircraft intelligence for safe navigation. As of now, the civil and the military automatons have obliged on-board insight to execute Autonomous-flying undertakings. Much of the time, they utilize a Global Positioning System (GPS) for flight task and its sensors for detection of any obstacle and evasion. With the true objective for unmanned ethereal vehicles (UAVs) to be totally self-ruling in decision-making, an on-board intelligence module must be provided with proper data about its prompt environment.

A large portion of the UAVs relies upon coordinated frameworks comprising of elevation, speed, and position control circles to accomplish operational autonomy. In spite of its exhibited unwavering quality, such a framework is by and by restricted in executing exceptionally complex undertakings. Completely self-ruling UAV decision-making can be possible only when the framework can play out the double capacity of object sighting and perception, which is alluded to as object detection and classification in computer vision. While these undertakings work out easily for people, they are dynamic and confusing for machines to perform the task on their own. One of the issues as of now confronting self-sufficient UAV operation is conducting detection and classification activities continuously. To take care of this issue, we have adjusted and tried a CNN ("Convolution Neutral Network") based software called YOLO ("You Only Look Once").

This detection and classification algorithm was balanced and effectively connected to the video feed acquired from the UAV in real-time.

Using the system, YOLO ("You Only Look Once") at an image to anticipate what and where the objects are present. YOLO is refreshingly straightforward. A single convolutional network predicts multiple bounding boxes and class probabilities for those boxes simultaneously. YOLO improves its performance in detection by training itself on full images. This bound together model has a few advantages over customary strategies with respect to object detection. To begin with, YOLO is very quick. We do not need a complex pipeline as we outline detection as a regression problem. To predict detections we run the neural network on a new image. Besides, YOLO accomplishes more than double the mean normal exactness of other real-time systems. Secondly, YOLO reasons out globally about the image while making predictions. Unlike sliding window and region proposal-based techniques, YOLO sees the whole picture amid test time and preparing so it implicitly encodes relevant data about the classes and appearance. Fast R-CNN, which is a top detection method, mistakes the background patches in an image for objects since it cannot see the bigger setting that is present. Whereas, YOLO makes very fewer background errors as compared to that of Fast R-CNN. Third, YOLO learns generalizable portrayals of objects. At the point when trained on natural images and tested on artwork, YOLO outmanoeuvres top detection methods like DPM and R-CNN by a wide margin. Since YOLO is significantly generalizable it is less inclined to break down separate when connected tonew domains or unexpected inputs.

## II. RELATED WORK

### A. *Study on Tensor Flow*

TensorFlow is a MLframework that works everywhere scale and in heterogeneous conditions. TensorFlow uses dataflow graphs to represent computation, shared state, and the operations that mutate that state. It maps the nodes of a dataflow graph across many machines in a cluster, and within a machine across multiple computational devices, including multicore CPUs, general-purpose GPUs, and custom-designed ASICs known as Tensor Processing Units (TPUs). This architecture offers adaptability to the application designer: whereas in previous "parameter server" structures the administration of shared state is incorporated with the framework, TensorFlow empowers designers to explore different avenues regarding novel advancements and training algorithms. TensorFlow supports a variety of applications, with an attention on training and inference on deep neural networks. A few Google administrations use TensorFlow in production, and it has ended up being extensively used for machine learning research. In this paper, we depict the TensorFlow dataflow model and demonstrate the convincing performance that TensorFlow achieves for several real-world applications.

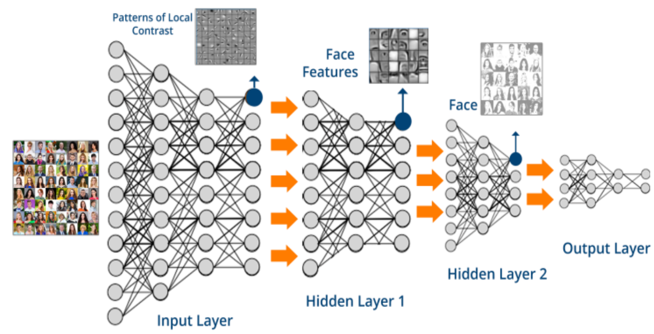


*Fig.1 TensorFlow Working model*

The Client is the fundamental segment in a TensorFlow system. The client with the help of Session interface communicates with the master and one or more worker processes, with each worker process responsible for executing graph nodes on those particular devices as instructed by the master and arbitrating access to at least one computational device such as ("CPU cores or GPU cards"). We have both local implementation and distributed implementations of the TensorFlow interface. The local implementation is utilized when the client, the master, and the worker all run on a single machine in the context of a single OS process ("possibly, with multiple devices, if for example, the machine has many GPU cards installed"). The dispersed usage imparts the vast majority of the code with the local implementation, yet broadens it with help for a domain where the master, client and the workers can all be in various procedures on various machines. In our dispersed condition, these distinctive errands are containers in jobs managed by a cluster scheduling system.

### B. *Unified Detection with YOLO*

The "YOLO" has numerous points of interest over other generally utilized Convolutional Neural Networks (CNNs) software. For instance, numerous CNNs utilize local proposition techniques to propose potential bounding boxes in images. This is trailed by arrangement of bounding box and refinement and the disposal of copies. Finally, based on the number of objects found in the scene, the bounding box is then re-scored. The issue with respect to this technique that they are connected at numerous areas and scales. Areas of an image that has high scoring are considered as detections. This methodology is repeated until a specific detection threshold is met. These calculations are exact and are at present utilized in numerous applications, but they are additionally computationally costly and essentially hard to enhance or parallelize. Therefore, this makes them inadmissible for autonomous Drone applications. Then again, “YOLO” utilizes a solitary neural network to isolate a picture into areas, while anticipating probabilities and bounding box for each region. Each of these bounding boxes are then weighted by the anticipated probabilities. The standard favoured point of view of this methodology is that the entire image is assessed by the neural network, and expectations are made subject to the possibility of the picture, not the proposed territory.

We bring together different parts of object detection into a solitary neural network system. Our network utilizes features from the whole picture to foresee each bounding box. It likewise predicts entire bounding boxes over all classes for a picture at the same time. This suggests our network reasons comprehensively about the complete picture and all the objects that are present in the image. The configuration of YOLO engages in end-to-end training and speed in real time while keeping up high normal exactness.

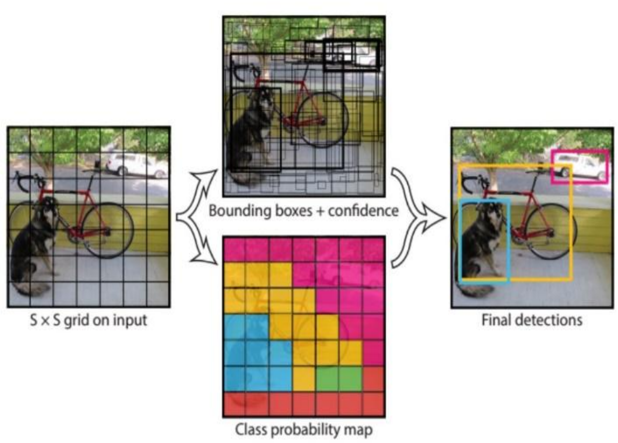


Fig. 2

Our framework partitions the input image into an S × S grid. If the object’s focal point falls into a grid cell, that specific grid cell is in charge of distinguishing that specific object. Each grid cell predicts “B bounding boxes” and “Confidence Scores” for those boxes. The confidence score(s)implies that how beyond any doubt the box contains an object and how exact it supposes the box is that it predicts. Formally we define confidence as $\Pr(\text{Object}) * \text{IOU}^{\text{truth}}_{\text{pred}}$ . On the off chance that an object exists in that cell, the confidence scores ought to be zero. Otherwise, we want the confidence score to be equal to IOU “Intersection over Union” between the predicted box and the ground truth. Each bounding box consists of 5 predictions: confidence, w, h, x and y. The coordinates (x, y) represent the center of the box relative to the bounds of the grid cell. The width and stature are anticipated in respect to the whole image. Finally, the confidence prediction represents the Intersection over Union (IOU) between the predicted box and any ground truth box. Each grid cell also predicts C conditional class probabilities, $\Pr(\text{Class}_i|\text{Object})$. These probabilities are conditioned on the grid cell containing an object. We just foresee one lot of class probabilities per grid cell, regardless of the number of boxes B. At test time, the individual box confidence predictions and the conditional class probabilities is multiplied, which provides us with class-specific confidence scores for each box. These scores encode how well the anticipated box fits the object and thelikelihood of that class showing up in the box.

## III. METHODOLOGY

1) THE ARCHIETECTURE AND WORKING OF DRONE

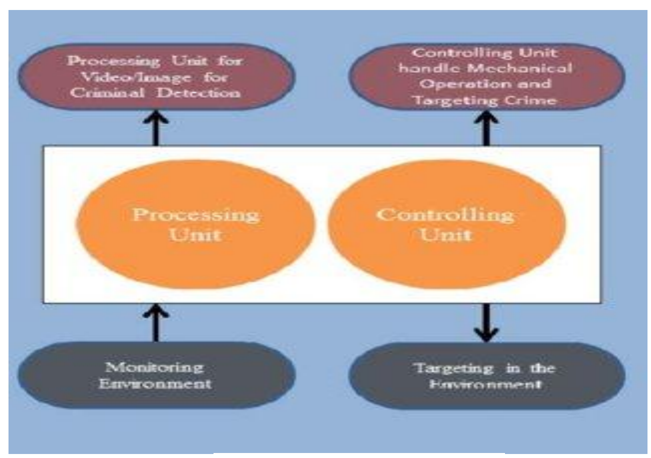


Fig.3 Architecture of drone Model

The structure of proposed UAV is small as it helps in moving effectively in blocked roads amid monitoring and targeting. The UAV depends on High Definition resolution camera for visionary framework. The proposed UAV contains camera to capture images and record videos, weapons for focusing on, control unit reliant on two handling unit for controlling all activities.

UAV depends on two diverse handling units; these two units are connected to one another for sharing of instruction and for controlling all the activities amid checking, for instance, moving between the structures, recording crime scene and focusing on. The principal handling unit mainly deals with image processing operation because the operation of detecting and recognizing the shape of object requires high computational power for detection and classification of object in real time. It is lumbering in field of AI and image processing to discover exact data in image settle on perfect decision for the benefit of extracted information. The second handling unit will control physical advancements of UAV and it is connected to sparing itself from obstacles. The second handling unit likewise controls all the mechanical errands and target objects in the midst of the fly time of drone.

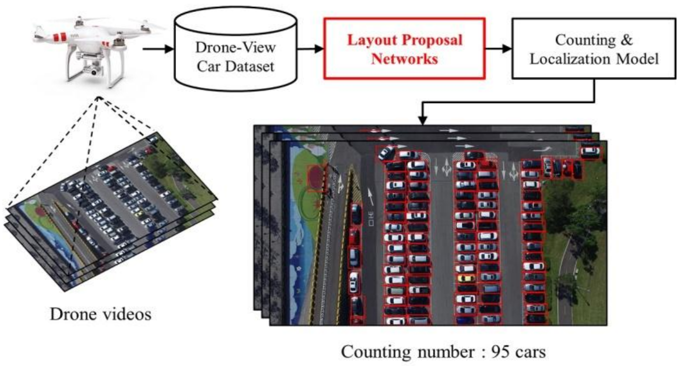


Fig. 4 The Expected result from a drone's footage.

## IV. RESULT

Due to the use of YOLO ("You Only Look Once") over the basic TensorFlow Model, We were able to detect multiple images and due to the use Of High Definition camera that has been used, we could get a better image quality of the objects in focus helping the accuracy of the detection. We conducted experiment to show that our methodology performs effectively in a practical yet controlled condition. In the set of experiments, we center on testing the exactness of object recognition using YOLO, we assess the speed of YOLO for object recognition, comparing running times on a local laptop where communication time is much less predictable but computational resources are much greater. We verify our methodology with the situation of a drone flying around few objects in an indoor domain, as a straight forward simulation of a search-and-rescue or surveillance application.

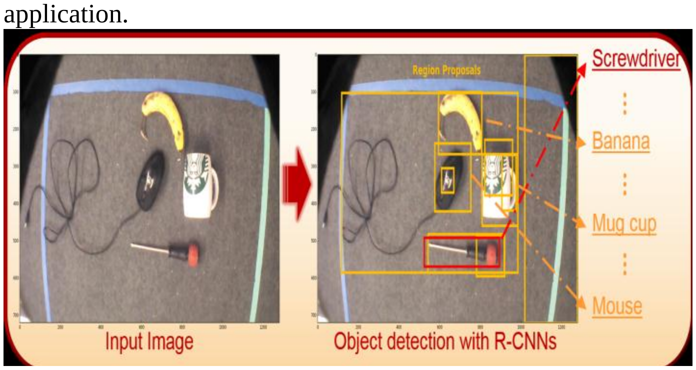


Fig. 5

## V. CONCLUSION

TensorFlow, an adaptable information stream based programming model, just as single machine and circulated usage of this programming model. The frame work is based on the experience in the real world in conducting research and sending more than one hundred ML projects all through a wide scope of Google's products and administrations.

This paper has been of the many use cases of the incredibly powerful Artificial Intelligence – Machine Learning library that is TensorFlow and we are yet to see the full potential of this field and also what third party developers will create using this library. The "YOLO" has been turned out to be increasingly productive. To place that into point of view, "YOLO" is equipped for running systems on video feeds at 150 frames per second. This infers "UAV" can process the video feed from its image acquisition system in less than25ms of latency. This almost prompt reaction time permits UAVs to perform time-sensitive and complex assignments in a productive and precise way.

## REFERENCES

[1]"N.H.Motlagh, T.Taleb, and O.Arouk,"Low-altitude unmanned aerial vehicle-based internet of things services: Comprehensivesurvey and futurePerspectives,"IEEE internet of Things Journal, Vol.3, no.6, Dec.2016"

[2]"Arvind and Rishiyur S. Nikhil. Executing a program on the MIT tagged-token dataflow architecture. IEEE Trans. Comput., 39(3):300–318, 1990."

[3]"Jack Clark. Google turning its lucrative web search over to AI machines, 2015."

[4]"Sherrah, J. Fully Convolutional Networks for Dense Semantic Labelling of High-Resolution Aerial Imagery. Available online: https://arxiv.org/pdf/1606.02585.pdf (accessed on 8 June 2017)."

[5]"Sermanet, P.; Eigen, D.; Zhang, X.; Mathieu, M.; Fergus, R.; LeCun, Y. Overfeat: Integrated Recognition, Localization and Detection Using Convolutional Networks. Available online: https://arxiv.org/pdf/1312.6229.pdf (accessed on 14 June 2017)."

[6] "T. P. Breckon, S. E. Barnes, M. L. Eichner, and K. Wahren, "Autonomous real-time vehicle detection from a medium-level UAV," in Proc. 24th International Conference on Unmanned Air Vehicle Systems, pp. 29.1-29.9, 2009"

[7] "M. Bhaskaranand, and J. D. Gibson, "Low-complexity video encoding for UAV reconnaissance and surveillance," in Proc. IEEE Military Communications Conference (MILCOM), pp. 1633-1638, 2011."

 