Multiple Object detection using Autonomous UAVs based on YOLOv3 algorithm

Murugendrappa N¹, KE Gayathri¹, Harshitha GM¹ and Sony Gupta¹

¹ Department of Electronics and Communication, GM Institute of Technology, Davangere, Karnataka, India

Abstract. Currently, there has been an enormous increase in interest towards autonomous UAVs and applications with respect to autonomous UAVs such as scientific data collection and security monitoring, infrastructure inspection, search and rescue and much more. Recognizing objects visually is a key step in such applications and is very pivotal to build a complete autonomous system. Nevertheless, the chore of detecting an object is very much challenging and is even more difficult when we have low quality images aboard low cost consumer UAVs. More frequently due to the motion of UAVs, images get blurred, noisy and even onboard cameras have low resolution hence object detection becomes more challenging since identifying objects are quite small and these tasks get even more difficult as there is a necessity of real time detection of objects. This paper focuses on all these aspects and brings an outcome for the object detection and identification.

Key words: Convolution Neural Networks (CNNs), Graphics Processing Units (GPUs), Regions with Convolutional Neural Network (RCNN), Unmanned Aerial Vehicles (UAVs), You Only Look Once (YOLO).

1. Introduction

A number of UAV researches have been approached for the detection and identification of certain sorts of objects like people, landmarks, vehicles and pedestrians for autonomous navigation along with landing in real-time. Out of numerous methodologies, there are only a few that consider recognizing multiple objects at a time, in spite of the very fact that identifying several target objects is clearly a key action for many applications of UAVs.

In our view, the main reasons for the space between technical capabilities and application needs are the below mentioned practical but critical limitations:

1. Visual perception algorithms often got to be hand-tuned to particular object and context types.
2. It difficult to create and store a spread of target object models, especially when the objects are diverse in appearance.
(3) Real-time object detection demands high computing power even while detecting single objects, much less when many target objects are involved. However, in several fields, there is a necessity for object detection and tracking effectively while handling occlusions and other complexities. Numerous researches tried for various approaches in detecting an object which largely depends on the application domain. Few research works that made evolution for the proposed work are depicted as follows.

**Figure 1.** Detection of varied objects using a UAV (drone)

Object detection and recognition using CNN is deliberately demanding, it requires high-end Graphics Processing Units (GPUs) which results in utilizing a lot of power and weight precisely for a low-cost and lightweight drone. In this paper, they propose moving the computation to an off-board computing cloud. They apply Regions with CNNs (R-CNNs), a state-of-the-art algorithm, to detect not one or two but hundreds of object types in near real-time (1). And the UAVs which are popping up as a promising technology for environmental as well as infrastructure monitoring it requires the use of computer vision algorithms to examine the information captured from an on-board camera with the evolution of a single-shot object detector grounded with images which can operate between 5-18 frames-per-second for different platforms with an overall accuracy is taken by UAVs(2), which only recognizes an object to a particular class; object detection also must extract accurate locations of objects. Within the state-of-the-art object detection algorithms, several methods to enhance its performance are proposed. Some unplanned experiments are conducted to prove the effectiveness of the network. Also, applying the network as an auxiliary module to the faster R-CNN algorithm and testing them (3), you only look once (YOLO) at the given image to predict the objects that are present and their location. It predicts multiple bounding boxes and class probabilities for the predicted boxes. YOLO achieves more than twice the mean average precision of other real-time systems (4). Remediating the network structure of YOLO v2 and obtain the YOLO-R network to improve the accuracy of detection; one can intend to integrate more pedestrian context features (5). And this method is more than 6 times faster than the Faster R-CNN at the same detection rate and quantitatively analyzes the relationship between sample scale and detection performance under small scale datasets(6). Computer Vision consists of diversified aspects like image recognition; every object present in the particular image is highlighted with rectangular boxes and tags are assigned to the identified object also with the accuracy of each method used for identifying objects (7).

Two technologies have empowered major tasks like object detection and tracking for traffic vigilance systems due to the increased demand in the features in image for efficient algorithms and detection and tracking of urban vehicle dataset.

2. **Object detection model yolov3**
YOLOV3 is one of the widely used methods for object detection. Basically, it is the algorithm that explains the working of the code with respect to detecting the objects in the image. Previously, object detection frameworks looked at varied parts of the respective image multiple numbers of times at different scales and processed image classification techniques in order to detect objects. This approach is time consuming and inefficient. It has its own architecture based on CNN and anchor boxes. YOLO in general is an improvement over a region based object detection algorithm. Region based object detection algorithms are much slower and have errors whereas YOLO mainly focuses on speed.

The above mentioned Figure 2 depicts the block diagram and the steps involved in this process are explained below:

The very first step is camera initialization in the drone.

This in turn, searches for an image in order to input the image to the object detection algorithm.

Feature extraction is a framework that decreases the dimensionality of the given data by removing the redundant data. This in turn increases the accuracy in extracting the features.

YOLOv3 provides an input image to which the output is a particular vector that consists of the class probabilities and the coordinates of the bounding boxes. This is done using a sliding window approach with the stride of zero that gives no overlaps. The whole image gets divided into a grid of particular dimensions. The output vector will have \( P_c \) which gives the probability of the object being present in that particular image, \( C_1, C_2...C_n \) gives the probability of the class that a particular object may belong to, \( x \) and \( y \) coordinates of the particular image and the height \( (h) \) and width \( (w) \) of the bounding box.

For multiple objects being present in a single grid, the algorithm uses anchor boxes. Ideally, any number of anchor boxes can be used by the algorithm. With the anchor boxes the output vector will also get modified for a particular grid cell that is the parameters in each output vector that gets multiplied into the number of anchor boxes it has. The YOLOv3 algorithm uses non-maximal suppression to have only one bounding box for each object. The bounding boxes with the confidence less than the threshold will be removed with the non-maximal suppression technique.

3. Result analysis
This is a sample image that is fed to the algorithm and the algorithm is expected to detect and identify the objects present in the image and label them with respect to the particular class assigned. After identification of the objects, each object is assigned with a tag which has the details of the object.
Figure 3. Image before object detection

Figure 4. Image after object detection

Figure 5. Aerial image before detection
Figure 6. Aerial image after object detection

Figure 7. Real time object detection with class probability and bounding box from a video clip

Figure 8. Image before object detection
4. Conclusion
This paper explains about various models used in object detection and recognition techniques. Along with, feature descriptors, bounding boxes and segmentation methods based on an image, a video or live stream in python. We have analyzed high accuracy methods to identify the object of same and different classes along with its probability using CNN for feature extraction and pre-defined weights to transform the input data. The result obtained is marked using a bounding box and a tag is assigned to the respective object with its class and probability.

References
[1] Lipeng S, Zhao S, Li G and Liu B 2019 High accuracy object detection via bounding box regression network Frontiers of Optoelectronics 12 324-31
[2] Lan W, Dang J, Wang Y and Wang S 2018 Pedestrian Detection Based on YOLO Network Model IEEE International Conference on Mechatronics and Automation Changchun, China, 1
[3] Li G, Song Z and Fu Q 2018 “A New Method of Image Detection for Small Datasets under the Framework of YOLO Network 3rd Advanced Information Technology, Electronic and Automation Control Conference (IEAEC) 1031-1035
[4] Sri Sai MR, Rella S, Veeravalli V 2019 Object Detection and Identification A Project Report Thesis 14
[5] Mohana and Ravish Aradhya HV 2019 Object Detection and Tracking using Deep Learning and Artificial Intelligence for Video Surveillance Applications Int. Journal of Advanced Computer Science and Applications (IJACSA) 10 4-8
[6] Felicita JJ and Abinaya D 2019 An Efficient Object Detection Model Using Convolution Neural Network Proc. of the 3rd Int. Conf. on Trends in Electronics and Informatics (ICOEI) 142-147
[7] Geethapriya S, Duraimurugan N and Chokkalingam SP 2019 Real-Time Object Detection with Yolo International Journal of Engineering and Advanced Technology (IJEAT) 8 578-581
[8] Huang L, Yang Y, Deng Y, Yu YD 2015 Unifying Landmark Localization with End to End Object Detection arXiv preprint arXiv:1509.04874
[9] Aradhya HVR 2016 Elegant and efficient algorithms for real time object detection, counting and classification for video surveillance applications from single fixed camera International Conference on Circuits, Controls, Communications and Computing (I4C), Bangalore 1-7
[10] Van Eten A 2018 You Only Look Twice: Rapid multi-scale object detection in satellite imagery in ArXiv preprint, ArXiv:1805.09512