INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 5, Issue 4)
Available online at: www.ijariit.com

A review on object detection from Unmanned Aerial Vehicle using CNN

Amanuel Ayalew
aamanuel69@gmail.com
Sharda University, Greater Noida, Uttar Pradesh

Dr. Pooja
Pooja.1@sharda.ac.in
Sharda University, Greater Noida, Uttar Pradesh

ABSTRACT

UAV stands for unmanned Aerial Vehicle, which can be as small as birds, regular drones or as big as private aircraft with no pilot on board. Since there is no one onboard UAV is remotely controlled. UAV is currently being used for a different purpose, examples are spy footages, sky view footage, reconnaissance, attacking roles, aerial surveillance, motion picture filmmaking, disaster rescue, parcel delivery, warehouse management, and other uses. Due to UAV multi-functionality and portability especially drones demand is growing faster, therefore, people need systems that work with the UAV (drones) to detect objects in real-time for military, safety reconnaissance, and surveillance. This paper review approaches to detect objects from camera view and from UAV scenes using machine learning algorithms. The growth of computer vision systems initiated better algorithms using huge training and testing datasets, faster GPU and CPU so that systems can achieve state of the art object detection system by training and classifying the data using machine learning approach. Since there are different orientation, background, occlusion in an image, object detection is not an easy task. The goal of object detection is to categorize images and video feeds from UAV into common categories.

Keywords*— Object detection, Unmanned Aerial Vehicle*

1. INTRODUCTION

An Unmanned Aerial Vehicle (UAV), ordinarily known as a drone, is an airplane with a ground base controller that means it does not have a pilot on board. UAVs are a part of an Unmanned Air Ship Framework (UAS); that consists of the controlling system, UAV and their information exchange mechanism. The control mechanism for UAV can be either by using remote control or person control the flight independently by a computer.

Contrasted with a manned flying machine, at first UAVs were used for a purpose which is excessive "not smart, risky or hazardous" for people. Their first application, for the most part, was military purpose, but after that their application area started to get broader started to be applied in into different applications of business, agriculture, site mapping, photography, landscape studies, reconnaissance and others, such as, asset management, target detection, security footage, parcel delivery, policing, and UAV racing competition. Nowadays Regular people have more UAVs than military offices because of the exponential growth of the UAV market. In 2015 more than a million UAVs were sold showing the demand for UAV is getting higher. Currently, UAVs are used for a variety of applications among that these are few of them.

- Management of civil infrastructure assets
- Routine bridge inspection
- Power line surveillance
- Traffic surveying
- Reconnaissance
- Surveillance
- Search and rescue
- Infrastructure inspection

The cameras in the UAVs will take photo and video from the environment and used for processing by the method of object detection and classification methodology and it can be used for the wanted application.

For an object detection application to work efficiently we need to first train our model, classification of objects which means the model has to identify if the image or video of the dataset is a vehicle, people, tree or any other thing. After the model is trained with classification, we need to train the model object detection, which involves in which pixels of the whole image is the target object located and finding out the boundary frame for the object.

In today’s modern world images play a vital role in the technology industry, billions of images are available on social media, most people would rather use images than text or audio file for communication. Approximately 300 million photos are uploaded to Facebook every single day, Infotrends estimates in 2020 cameras and phones will capture 1.4 trillion images. We can say image utilization is growing exponentially, therefore, we need system and applications that will comprise image processing and analysis. Computer vision mainly focuses on analysis from image or video data.

To adequately deal with this image data, we need some thought regarding its substance. Automated processing of image content is valuable for a wide assortment of image related assignments. For a computer system, this implies crossing the supposed semantic gap between the pixel level information in the image documents and the human comprehension of a similar image. Computer vision endeavors to connect to this gap.

2. OBJECT DETECTION TECHNIQUES

The real motivation behind why new object detection algorithms are needed and can't continue with the previous algorithms is the number of output layers in the CNN are not constant they are variable because we don’t know how many times an object of interest can appear in an image and use a CNN to classify the presence of the object within that region so this makes it difficult to build a convolutional network which is directly followed by fully connected layers. The issue in this methodology is that the proposal objects might have any spatial location within the image and the object may also have a different aspect ratio. Consequently, we might need to choose a huge number of regions and that will be very difficult to process computationally. In this way, algorithms like R-CNN, YOLO and so on have been created to find these occurrences and find them quickly.

2.1 R-CNN

To solve the issue choosing the huge number of regions a new algorithm is proposed by Ross Girshick et al. the algorithm uses the principle of selective search and extracts two thousand regions only from the image and these regions are called region proposals. In this manner, presently, rather than classifying an enormous number of regions, the algorithm made it possible to work with the 2000 regions. Selective search algorithm which is stated below:

(a) Produce starting sub-division, to create numerous candidate regions.
(b) Merge same regions recursively into one big region using the greedy algorithm.
(c) Produce the final candidate regions proposal using the generated regions.

These 2000 candidate region proposals are warped into a square and inputted to the convolutional neural network which will give us 4096-dimensional feature vector as output. Feature extraction will be done by the CNN and the feature extracted from the images will be comprised by output dense layer and they will be an input for the SVM. The SVM algorithm then identifies if there is an object in the image region proposal. In order to increase the efficiency and accuracy of the bounding box, the algorithm will predict four offset values apart from predicting an object is present in the region. For instance, if we want to detect a person in a specific region proposal the algorithm may detect the presence of a person but the face of the person in the region proposal could be sliced down the middle in this kind of scenario the counterbalance values will change the bounding box of the region proposals.

2.2 Issues of R-CNN

R-CNN systems classify 2000 regions for each image in training and testing, this process consumes time and resource. In order to detect an object using R-CNN, it will take around 47 seconds per a testing image so it is impossible to implement this system in real-time. There is no learning process in the selective search algorithm so this may lead to a selection of bad candidate region proposals.

2.2.1 Fast R-CNN: Fast R-CNN was implemented in order to overcome the problems of R –CNN by the same author. The main problem of R-CNN was its speed and in Fast R-CNN a quicker object detection algorithm was introduced. Both R-CNN and Fast R-CNN use the same methodology except input image is fed to the CNN which gives a convolutional feature map but in the case of R-CNN, it was the region of proposal fed to the CNN. In Fast R-CNN region of a proposal will be distinguished and warped into squares from the convolutional feature map and shape them to a fixed size in order to feed them to a fully connected layer by using RoI pooling layer. To predict the class of the region proposal and the offset values of the bounding box we will be using a softmax layer from the RoI feature vector. In Fast R-CNN convolution is done only once for an image then it will give us the feature map while R-CNN feeds 2000 regions for the CNN. This is the reason why Fast R-CNN is quicker than R-CNN algorithm

2.2.2 Faster R-CNN: In order to find the region of proposals both R-CNN and Fast R-CNN algorithms utilize selective search but it is a slow and tedious procedure influencing the execution of the network. In this way, in order to solve this problem, Shaoqing Ren et al. thought of an object detection algorithm that takes out the selective search algorithm and gives the network a chance to learn the region proposals.

The CNN will receive input image just like Fast R-CNN in order to give a convolutional feature map. To recognize region proposals a new network will be used. In the case of Fast R-CNN, selective search algorithm was used on the feature map to recognize the region proposals but for Faster R-CNN a different network will be used. Then it will follow the same process as Fast R-CNN.

2.2.3 YOLO—You Only Look Once: Unlike the previous object detection algorithms YOLO does not use the region to detect an object from an image. Region-based algorithms do not look at the whole image they only look at some region of the image having the highest probability of an object. YOLO works in a completely different way from region-based algorithms that are bounding boxes and their class probability is predicted in a single convolutional neural network. YOLO first take the input image and changes the image into NxN grid for each grid m number of bounding boxes will be taken then the CNN predicts the class probability and offset values for each grid. Threshold value will be set then if the class probability is below the threshold object will not be detected but if the class probability is greater than the threshold the object will be detected at that specific class probability value.

3. LITERATURE REVIEW

I reviewed different papers that focus on object detection fro UAVs and object detection as a general. Most of the papers mentioned utilization of CNN under YOLO algorithm this is because of YOLO algorithms are so fast that they can be used in real-time object detection. Real-time object detection is the main goal for processing video feeds from Unmanned Aerial Vehicles (UAV). These are the papers I reviewed:

CNN is being used in different applications M.Radovic et al.[1] Used it in civil engineering applications in order to detect target objects from aerial images using autonomous UAV. They tested CNN image recognition and implemented CNN architecture and selection of parameter and its’ tuning for detection and classification of objects in aerial images then demonstrate successful applications of YOLO algorithm from real-time video feed during UAV operation on real-time object detection and classification. The experiment used CNN, YOU ONLY LOOK ONCE YOLO, with a best-performing image size of 448x448x3 and the result was 97.5% accuracy and 97.4% sensitivity.

J. Lee et al.[2] Used Region with convolutional neural network (R-CNN) algorithm to detect an object from drones in real-time by computing the object detection remotely on the cloud because trying to compute it on the drone is computationally challenging and the additional hardware will make the drone unable to fly. It was able to detect hundreds of object types in near real-time with 88% accuracy.

A. Chung et al.[3] Used a micro-Unmanned Aerial Vehicle (UAV) capable of real-time litter detection from video surveillance footage through an ensemble-based machine learning model. They used five different algorithms which are two classifiers and three detectors to determine the strongest models to utilize in the ensemble method. The Classifiers are SVM and CNN, while the detectors are Single Short multi-box Detectors (SSD), region-based fully convolutional network and You Only Look Once (YOLO).

S. Han et al.[4] In this publication embedded system framework of Deep Drone was proposed to add an automatic detection and tracking feature for drones. The major objective of the author was to implement the vision component of the drone which is a combination of advanced detection and tracking algorithms to several hardware platforms, which contains both desktop GPU (NVIDIA GTX980) and embedded GPU (NVIDIA Tegra K1 and NVIDIA Tegra X1) to evaluate the frame rate. A tracking algorithm (CNN) using HOG feature and KCF were used and an accuracy of 62.0% was achieved also by using R-CNN 0.17sec runtime attained.

C. Kyrkou et al.[5] author proposes using deep Convolutional Neural Networks (CNNs) to explores the trade-offs involved in the development of a single-shot object detector that can enable UAVs to make vehicle detection under a resource-constrained environment such as in a UAV. For such cases, the author presented a general approach for the data collection and training stages, a CNN architecture, and the optimizations necessary to efficiently map. YOLO algorithm was implemented with four different structures (SmallYoloV3, TinyYoloVoc, TinyYoloNet, and DroNet) which gave an overall accuracy of 95% and performs only at $5-6$ FPS.

S. Ren et al. [6] Region Proposal Network (RPN) was proposed by sharing full-image convolutional features with the network detection, that enabled approximately cost-free region proposals. The author used Convolutional neural network (VGG net) on PASCAL VOC 2007 dataset. Accuracy of PASCAL VOC 2007 with detector fast R-CNN with ZF is 59.9% and Accuracy of PASCAL VOC 2012 with detector Fast R-CNN and VGG is 70.4%. A completely original model that explains about the prediction, detection, and tracking of autonomous drivers was proposed.

W.LUO et al.[7] Shows a deep neural network given data captured by a 3D sensor using Region Proposal Networks (RPN) that was able to show 3D detection, tracking and motion forecast, Mask-RCNN. For detection accuracy is 80.9%, Motion Forecasting has recall value of 92.5%.

R. Hansch et al.[8] Propose near real-time object detection network that produces reliable results with no restriction of environment and object type. Object detection with RGBD cameras that can be used for autonomous home robot or other applications. Hough forest ensemble learning framework is applied which is capable of classification and regression. Baseline detection performance for RGB feature is an average AUC of 0.576 and depth value 0.768 AUC.

In another article, The full employment of convolutional architectures in the Fast/Faster RCNN on VOC 2007 dataset was demonstrated by designing deep convolutional networks (ConvNets) of various depths for feature classification by Y. Ren et al.[9].

According to J. Redmon et al. [10] using a single neural network, it was able to detect bounding boxes and their class probability of region proposal of images in one evaluation. YOLO approach was used to detect objects which frame object detection as a regression problem. The network can be directly optimized end-to-end on detection performance because the whole detection pipeline is a single network and was able to detect objects real-time at baseline performance of 45 fps.

4. CONCLUSION

In this paper, I reviewed different papers focused on general object detection and object detection from UAV, in most of the papers YOLO is used and mentioned as an effective model. YOLO Model is a state-of-the-art algorithm used for real-time object detection with a baseline of 45 frames per second (fps). It is one of the most effective models for UAV feed images and videos because these data need to be detected in real-time and trained directly on full images. YOLO likewise sums up well to new spaces making it perfect for applications that depend on fast, robust object detection.

5. REFERENCES

[1] M. Radovic, O. Adarkwa, and Q. Wang, "Object Recognition in Aerial Images Using Convolutional Neural Networks," J. Imaging, vol. 3, no. 4, p. 21, 2017.

[2] J. Lee, J. Wang, D. Crandall, S. Sabanovic, and G. Fox, "Real-Time Object Detection for Unmanned Aerial Vehicles based on Cloud-based Convolutional Neural Networks," 2015.

[3] A. Chung, S. Kim, E. Kwok, M. Ryan, E. Tan, and R. Gamadia, "Cloud Computed Machine Learning-Based Real-Time Litter Detection using Micro-UAV Surveillance," pp. 1–10, 2018.

[4] S. Han, W. Shen, and Z. Liu, "Deep Drone: Object Detection and Tracking for Smart Drones on Embedded System," pp. 1–8, 2016.

[5] C. Kyrkou, G. Plastiras, T. Theocharides, S. I. Venieris, and C. S. Bouganis, "DroNet: Efficient convolutional neural network detector for real-time UAV applications," Proc. 2018 Des. Autom. Test Eur. Conf. Exhib. DATE 2018, vol. 2018–Janua, pp. 967–972, 2018.

[6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, 2017.

[7] W. Luo, B. Yang, and R. Urtasun, "Fast and Furious: Real-Time End-to-End 3D Detection, Tracking, and Motion Forecasting with a Single Convolutional Net," pp. 3569–3577, 2018.

[8] R. Hänsch, S. Kaiser, and O. Helwich, "Near Real-time Object Detection in RGBD Data," no. Visigrapp, pp. 179–186, 2017.

[9] Y. Ren, C. Zhu, and S. Xiao, "Object Detection Based on Fast/Faster RCNN Employing Fully Convolutional Architectures," Math. Probl. Eng., vol. 2018, pp. 1–7, 2018.

[10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2015.