A Review on Object Detection from Unmanned Aerial Vehicle Using CNN

Amanuel Ayalew and Dr. Pooja
Department of computer science engineering, Sharda University, Greater Noida-201306, India
Associate professor, department of computer science engineering, Sharda University, India
aamanuel69@gmail.com, Pooja.1@sharda.ac.in

Abstract

UAV stands for unmanned Aerial Vehicle, which can be as small as birds, regular drones or as big as private aircrafts with no pilot on board. Since there is no one on board UAV are remotely controlled. UAV are currently being used for 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 at real time for military, safety reconnaissance and surveillance. This paper presents an approach to detect objects from camera view of the UAV using machine learning algorithms. With the availability of large amounts of data, faster GPUs, and better algorithms, we can now easily train computers to detect and classify multiple objects within an image with high accuracy. Detection can be difficult since there are all kinds of variations in orientation, lighting, background and occlusion that can result in completely different images of the very same object. The goal of object detection is to categorize images and video feeds from UAV into common categories.

1 Introduction

An unmanned aerial vehicle (UAV), ordinarily known as a drone, is an airplane without a human pilot on board. UAVs are a segment of an unmanned air ship framework (UAS); which incorporate a UAV, a ground-based controller, and an arrangement of interchanges between the two. The trip of UAVs may work with different degrees of independence: either under remote control by a human administrator or independently by locally available PCs.

Contrasted with manned flying machine, UAVs were initially utilized for missions excessively "dull, filthy or hazardous" for people. While they started for the most part in military applications, their utilization is quickly growing to business, logical, recreational, farming, and different applications, for example, policing, peacekeeping, and observation, item conveyances, elevated photography, horticulture, sneaking, and ramble dashing. Regular citizen UAVs now endlessly dwarf military UAVs, with appraisals of over a million sold by 2015. [Radovic et al., 2017f]

Currently UAVs are used for 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 have to identify if the image or video of the dataset is 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 computer system, this implies crossing the supposed semantic gap between the pixel level information in the image documents and the human comprehension of similar image. Computer vision endeavors to connect to this gap.

2 Object Detection Techniques

In a standard convolutional network pursued by a fully connected layer, the length of the output layer is variable (not constant), this is because the number of occurrences of the objects of interest in an image is not fixed. A naive way to deal with this issue is take different regions of interest from the image, and use a CNN to classify the presence of the object within that region. The issue with this methodology is that the objects of interest may have different spatial locations within the image and different aspect ratios. Consequently, you would need to choose a huge number of regions and this could computationally explode. In this way, algorithms like R-CNN, YOLO and so on have been created to find these occurrences and find them quick.

2.1 R-CNN

To sidestep the issue of choosing a huge number of regions, Ross Girshick ET al.1 proposed a technique where we utilize selective search to extract only 2000 regions from the image and he called them region proposals. In this manner, presently, rather than trying to classify an enormous number of regions, you can simply work with 2000 regions. These 2000 region proposals are created using the selective search algorithm.

These 2000 candidate region proposals are warped into a square and fed into a convolutional neural network that delivers a 4096-dimensional feature vector as output. The CNN acts as a feature extractor and the output dense layer comprises of the feature extracted from the image and the extracted features are fed into a SVM to classify the presence of the object within the image region proposal. In addition to predicting the presence of an object within the region proposal, the algorithm additionally predicts four values which are offset values to increase the accuracy of the bounding box. For instance, given a region proposal, the algorithm would have predicted the presence of an individual but the face of that individual within that region proposal could've been sliced down the middle. In this way, the counterbalance values help in changing the bouncing box of the region proposal.

Regardless it requires a huge amount of time to train the system as you would need to classify 2000 area for each picture.

It can't be executed real time as it takes around 47 seconds for each test picture. The selective search algorithm a fixed algorithm. Hence, no learning is going on at that stage. This could prompt the generation of bad candidate region

2.2 Fast R-CNN

The same author of the past paper(R-CNN) found the solution of some of the disadvantages of R-CNN to introduce a quicker object detection algorithm and it was called Fast R-CNN. The methodology is like the R-CNN algorithm. But, rather than feeding the region of proposal to the CNN, we feed the input image to the CNN to produce a convolutional feature map. From the convolutional feature map, we distinguish the region of proposal and warp them into squares and by using a RoI pooling layer we reshape them into a fixed size so it tends to be fed into a completely associated/connected layer. From the RoI feature vector, we use a softmax layer to predict the class of the proposed region and furthermore the offset values for the bounding box.

The reason "Fast R-CNN" is quicker than R-CNN is because that you do not need to feed 2000 region proposals to the convolutional neural system every time. Rather, the convolution task is done just once per picture and a feature map is produced from it.

2.3 Faster R-CNN

Both R-CNN and Fast R-CNN uses selective search to discover the region proposal. Selective search is a slow and tedious procedure influencing the execution of the network. In this way, Shaoqing Ren et al2. thought of an object detection algorithm that takes out the selective search algorithm and gives the network a chance to learn the region proposals.

Like Fast R-CNN, the image is given as an input to a convolutional network which gives a convolutional feature map. Rather than using selective search algorithm on the feature map to recognize the region proposals, a different network is used to predict the region proposals. The predicted region proposal are then reshaped using a RoI pooling layer which is then used to classify the image inside the proposed region and predict the offset values for the bounding boxes.

2.4 YOLO—You Only Look Once

The majority of the past object detection algorithms use regions to localize the object inside the image. The network does not look at the entire image. Rather, portions of the image which have high probabilities of containing the object. YOLO or You Only Look Once is an object detection algorithm entirely different from the region based algorithm seen above. In YOLO a single convolutional network predicts the bounding boxes and the class probabilities for these boxes.

How YOLO functions is that we take a picture and split it into an SxS grid, inside every one of the grid we take m bounding boxes. For every one of the bounding box, the network yields a class probability and offset value for the bounding box. The bounding boxes having the class probability over a threshold value is chosen and used to find the object from the image.

YOLO is orders of magnitude faster 45fps (frames every second) than other object detection algorithms. The problem of YOLO calculation is object detection is hard if the objects are small within the image, for instance it may experience issues in identifying a flock of birds. This is because of the spatial constraints of the algorithm.

3 Literature Review

I reviewed different papers that focus on object detection 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 are the main goal for processing video feeds from unmanned aerial vehicles (UAV). These are the papers I reviewed in a tabular form:

NO	Article Title	Author	Year	Objective	Methodology	Result
1	Object recognition in aerial images using convolutional neural network	Matija Radovic, Offei and Qiaosong Wang	2017	Test CNN image recognition algorithms that can be used for autonomous UAV operations in civil engineering applications, present the CNN architecture and parameter selection for detection and classification of objects in aerial images, demonstrate successful applications of this algorithm on real time object detection and classification from the video feed during UAV operation. [Radovic et al., 2017]	CNN, YOU ONLY LOOK ONCE YOLO, with best performing image size 448x448x3	97.5% accuracy 97.4% sensitivity 0% specificity
2	You Only Look Once: Unified, Real-Time Object Detection	Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi	2016	Present YOLO approach to object detection, frame object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. Since the whole detection pipeline is a single network, it can be optimized end-to-end directly on detection performance. [Redmon et al., 2015]	convolutional neural network YOLO on PASCAL VOC dataset and ImageNet 2012 dataset, fast R-CNN	accuracy of 88% on the ImageNet 2012 validation set, fast R-CNN accuracy 71.6%
3	Real time object detection for unmanned aerial vehicles based on cloud-based convolutional neural network	Jangwon Lee, Jingya Wang, David Crandall, Selma Sabanovic and Geoffrey Fox	2017	Moving the computationally demanding object recognition to a remote compute cloud instead of trying to implement it on the drone itself, detect hundreds of object types in near real time[Lee et al., 2015]	Region with convolutional neural network (R-CNN)	88% accuracy 200 confidence per image
4	Cloud Computed Machine Learning Based Real-Time Litter Detection using Micro-UAV Surveillance	Ashley Chung, Sean Kim, Ethan Kwok, Michael Ryan, Erika Tan, Michael Ryan, Ryan Gamadia	2018	a micro-unmanned aerial vehicle (UAV) capable of real-time litter detection from video surveillance footage through an ensemble-based machine learning model, measure the performance of five different algorithms which are two classifiers and three detectors to determine the strongest models to utilize in the ensemble method.[Chung et al., 2018]	Classifiers SVM and CNN, detectors single short multibox detectors (SSSD), region based fully convolutional network, you only look once (YOLO)	CNN accuracy 51.73 average F1 score 0.330' SVM F1 score0.623, SSD accuracy 50%, F1 score 0.56, R-FCN F1 score 0.53, YOLO F1 score 0.40
5	Deep Drone: Object detection and tracking for smart drones on embedded system	Song Han, William Shen, Zuozhen Liu	2016	Propose Deep Drone embedded system framework, to power drones with vision: letting the drone to do automatic detection and tracking. Implement the vision component which is an integration of advanced detection and tracking algorithms. Implement system onto multiple hardware platforms, including both desktop GPU (NVIDIA GTX980) and embedded GPU (NVIDIA Tegra K1 and NVIDIA Tegra X1) and evaluated frame rate, power consumption and accuracy on several videos captured by the drone.[Han et al., 2016]	detection algorithm running Convolutional Neural Network (CNN) and a tracking algorithm using HOG feature and KCF.	Detection using faster R-CNN accuracy 62.0% with 0.17s runtime. Tracking using KCF algorithms

6	DroNet: Efficient convolutional neural network detector for real-time UAV applications	Christos Kyrkou, George Plastiras, Theocharis Theocharides, Stylianos I. Venieris and Christos-Savvas Bouganis	2018	Explores the trade-offs involved in the development of a single-shot object detector based on deep convolutional neural networks (CNNs) that can enable UAVs to perform vehicle detection under a resource constrained environment such as in a UAV. presents a holistic approach for designing such systems; the data collection and training stages, the CNN architecture, and the optimizations necessary to efficiently map such a CNN on a lightweight embedded processing platform suitable for deployment on UAVs.[Kyrkou et al., 2018]	four different structures (SmallYoloV3, TinyYoloVoc, TinyYoloNet, and DroNet)	95% accuracy and performance only $5-6$ FPS
7	Faster R-CNN: towards real time object detection with region proposal networks	Shaoqing Ren, Kaiming He, Ross Girshick and Jian Sun	2015	Introduce a Region Proposal Network (RPN) that shares full-image convolutional features with the detection network, thus enabling nearly cost-free region proposals.[Ren et al., 2015]	Convolutional neural network (VGG net), PASCAL VOC 2007	Accuracy of PASCAL VOC 2007 with detector fast R-CNN with ZF is 59.9%, Accuracy of PASCAL VOC 2012 with detector Fast R-CNN and VGG is 70.4%
8	Fast and furious: Real time end-to – end 3d detection, tracking and motion forecasting with single convolutional network	Wenjie Luo, Bin Yang and Raquel Urtasun	2017	Build a novel deep neural network that is able to jointly reason about 3D detection, tracking and motion forecasting given data captured by a 3D sensor.[Luo et al., 2017]	region proposal networks (RPN), Mask-RCNN	For detection accuracy is 80.9%, Motion Forecasting have recall value of 92.5%
9	Near real-time object detection in RGBD data	Ronny Hansch, Stefan Kaiser and Olaf Helwich	2017	An object detection pipeline that runs in near real-time and produces reliable results without restricting object type and environment. Object detection with RGBD cameras that can be used for autonomous home robot or other applications[Hansch et al., 2017]	Hough forest ensemble learning framework capable of classification and regression.	Baseline detection performance for RGB feature is an average AUC of 0.576, depth value 0.768 AUC
10	Object Detection Based on Fast/Faster RCNN Employing Fully Convolutional Architectures	Yun Ren , Changren Zhu, and Shunping Xiao	2018	design deep convolutional networks (ConvNets) of various depths for feature classification, especially using the fully convolutional architectures, demonstrates how to employ the fully convolutional architectures in the Fast/Faster RCNN [Ren et al., 2018]	Faster RCNN on the VOC2007	mAP (%) of GoogleNet 64.0% Inception v2 65.2% Inception v3 68.9% ResNet 50 70.8%
11	YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers	Rachel Huang, Jonathan Pedoeem and Cuixian Chen	2018	Develop a real-time object detection model to run on portable devices such as a laptop or cellphone lacking a Graphics Processing Unit (GPU). [Huang et al., 2018]	YOLO-LITE,	A mAP of 33.81% for PASCAL VOC dataset and 12.26% for COCO dataset. YOLO-LITE runs at about 21 FPS on a non-GPU computer and 10 FPS after implemented onto a website with only 7 layers and 482 million FLOPS.

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 state-of-the-art algorithm used for real time object detection with the 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.

References

[Radovic et al., 2017] Matija Radovic, Offei and Qiaosong Wang, *Object Recognition in Aerial Images Using Convolutional Neural Networks*. *J. Imaging*, vol. 3, no. 4, p. 21, 2017.

[Redmon et al., 2015] Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi, *You Only Look Once: Unified, Real-Time Object Detection*, 2015.

[Lee et al., 2015] Jangwon Lee, Jingya Wang, David Crandall, Selma Sabanovic and Geoffrey Fox, *Real-Time Object Detection for Unmanned Aerial Vehicles based on Cloud-based Convolutional Neural Networks*, 2015.

[Chung et al., 2018] Ashley Chung, Sean Kim, Ethan Kwok, Michael Ryan, Erika Tan, Michael Ryan and Ryan Gamadia, *Cloud Computed Machine Learning Based Real-Time Litter Detection using Micro-UAV Surveillance*, pp. 1–10, 2018.

[Han et al., 2016] Song Han, William Shen and Zuozhen Liu, *Deep Drone: Object Detection and Tracking for Smart Drones on Embedded System*, pp. 1–8, 2016.

[Kyrkou et al., 2018] Christos Kyrkou, George Plastiras, Theocharis Theocharides, Stylianos I. Venieris and Christos-Savvas 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.

[Luo et al., 2018] Christos Kyrkou, George Plastiras, Theocharis Theocharides, Stylianos I. Venieris and Christos-Savvas Bouganis, *Fast and Furious: Real Time End-to-End 3D Detection, Tracking and Motion Forecasting with a Single Convolutional Net*, pp. 3569–3577, 2018.

[Hänsch et al., 2017] Ronny Hansch, Stefan Kaiser and Olaf Helwich, *Near Real-time Object Detection in RGBD Data*, no. Visigrapp, pp. 179–186, 2017.

[Ren et al., 2018] Yun Ren , Changren Zhu, and Shunping Xiao, *Object Detection Based on Fast/Faster RCNN Employing Fully Convolutional Architectures*, *Math. Probl. Eng.*, vol. 2018, pp. 1–7, 2018.

[Pedoeem et al., 2018] Rachel Huang, Jonathan Pedoeem and Cuixian Chen, *YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers*, 2018.