iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.36144

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI June 2021- Available at www.ijraset.com

Simulation and Detection of Small Drones/Suspicious UAVs in Drone Grid

Arpit Gupta[1], Arnav Gupta[2]
[1, 2]Computer Communication and Engineering, Manipal University Jaipur Jaipur, India

Abstract: Today's technology is evolving towards autonomous systems and the demand in autonomous drones, cars, robots, etc. has increased drastically in the past years. This project presents a solution for autonomous real-time visual detection and tracking of hostile drones by moving cameras equipped on surveillance drones. The algorithm developed in this project, based on state-of-art machine learning and computer vision methods, succeeds at autonomously detecting and tracking a single drone by moving a camera and can run at real-time. The project can be divided into two main parts: the detection and the tracking. The detection is based on the YOLOv3 (You Only Look Once v3) algorithm and a sliding window method. The tracking is based on the GOTURN (Generic Object Tracking Using Regression Networks) algorithm, which allows the tracking of generic objects at high speed. In order to allow autonomous tracking and enhance the accuracy, a combination of GOTURN and tracking by detection using YOLOv3 was developed.

Keywords: Object Detection,Yolo,aerial vehicles,R-CNN.

I. INTRODUCTION

Human beings can easily detect and identify objects in their surroundings, without consideration of their circumstances, no matter what position they are in and whether they are upside down, different in colour or texture, partly occluded, etc. Therefore, humans make object detection look trivial. The same object detection and recognition with a computer require a lot of processing to extract some information on the shapes and objects in a picture.

The use of drones or Unmanned Aerial Vehicles (UAVs), both for military and civilian purposes, has increased in India in the past decade. At the same time, counter-drone systems are also being developed to address the threats posed by UAVs.

In the past several years, India has been seeing more use of drones—or small unmanned aerial vehicles (UAVs)—for various military and civilian purposes. These include reconnaissance, imaging, damage assessment, payload delivery (lethal as well as utilitarian), and as seen recently amidst the COVID-19 pandemic, for contact-less delivery of medicines. The use of drones, however, poses threats to public security and personal privacy. Analysts warn that as Unmanned Aerial Systems (UAS) become less expensive, easier to fly, and more adaptable for crime, terrorism, or military purposes, defense forces will increasingly be challenged by the need to quickly detect and identify such aircraft. The Small Unmanned Aircraft System (sUAS) technologies are continuously evolving: indeed, customized sUAS—i.e.UAVs, micro UAVs, and drones with their controller stations and equipment—can operate without radio frequency (RF) command and control links, and can use automated target tracking, aside from having obstacle avoidance and software-controlled capabilities.

Drones have low Radar Cross Section (RCS), slow speed, and small size—these characteristics make the task of detection difficult, and thereafter, identification and localization even more so. In response, governments and military forces across the world, including in India have developed various approaches to detect these aerial systems. These methods can broadly be classified as radar, video/electro-optical (EO), audio/acoustic, and RF-based.

In particular, UAVs equipped with camera sensors can operate in remote and difficult to access disaster-stricken areas, analyse the image, and alert in the presence of various calamities such as collapsed buildings, flood, or fire in order to faster mitigate their effects on the environment and on the human population. However, the integration of deep learning introduces heavy computational requirements, preventing the deployment of such deep neural networks in many scenarios that impose low-latency constraints on inference, in order to make mission-critical decisions in real-time.

Detection means the technology is able to detect drones. Detection alone usually isn't enough though. A radar that detects drones may also detect birds, for example.

In the field of object detection, tremendous success is achieved, but still it is a very challenging task to detect and identify objects accurately with fast speed. Human beings can detect and recognize multiple objects in images or videos with ease regardless of the object's appearance, but for computers it is challenging to identify and distinguish between things.

That's why classification is useful.

Technology that classifies drones will usually be able to separate drones from other types of objects - like planes, trains, and automobiles, for example.

One step further is identification. Some equipment can identify a particular model of drone, or even identify the drones or controllers digital fingerprint, like a MAC address for example. This level of identification can be handy for prosecution purposes.

Being alerted that a drone is present somewhere in the vicinity is already useful. But your situational awareness, and ability to deploy countermeasures is greatly enhanced if you know the drones (and/or the controller's) exact location. Some equipment will even allow you to track the drone location in real-time.

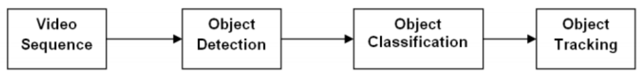

Figure 1-Steps of Object Tracking

Object detection is a fundamental topic in computer vision and an important component of sensor-based situational awareness systems for autonomous driving. With the emergence of convolutional neural networks (CNNs), object detection schemes have evolved significantly in the past few years. The development of various detection frameworks, such as faster region-based CNN (R-CNN) and single-shot multibox detector (SSD) has led to considerable advancement in the newly invented state-of-the-art technologies for object detection and identification.

II. LITRETURE REVIEW

In the field of object detection, tremendous success is achieved, but still it is a very challenging task to detect and identify objects accurately with fast speed. Human beings can detect and recognize multiple objects in images or videos with ease regardless of the object's appearance, but for computers it is challenging to identify and distinguish between things.

The first part of this chapter gives a brief summary of related work done on UAV based object detection whereas the second part explains the evolution of several deep learning models widely used for object detection.

In recent years, autonomous Unmanned Aerial Vehicles (UAVs) especially drones equipped with cameras have become very popular with the wide variety of their applications such as surveillance, aerial mapping, search and rescue, infrastructural inspection, precision agriculture etc.

Understanding visual data collected from these drones autonomously, has been the area of increasing

Interest.

Visual object detection is one of the prominent aspects of applications of drones and is critical to have in fully autonomous systems. However, object detection with drones is very challenging and complex as it is affected by various imaging conditions such as noise in the images,

Blur, low resolution, small target sizes etc. The task is even more difficult because of the limited computational resources available on the drones and the need for almost real time performance in

Many applications such as navigation, traffic management etc. Traditionally hand-tuned features were used for object detection and recognition. With the Breakthrough of deep learning using Convolutional Neural Networks there was a striking performance increase in dealing with these computer vision tasks. The key idea is to learn object features and model from raw pixel data of image. Detecting and identifying multiple objects in an image is hard for machines to recognize and classify. However, a noteworthy effort has been carried out in the past years in the detection of objects using convolutional neural networks (CNNs). In the object detection and recognition field, neural networks were in use for a decade but became prominent due to the improvement of hardware and new techniques for training these networks on large datasets. In, researchers showed that the CNNs inherit the advantages of deep learning, which makes their results in the field of object detection and recognition greatly improved compared with the traditional methods. Researchers had made many efforts to use stochastic gradient descent and backpropagation to train deep networks for object detection. Those networks were able to learn but were too slow in practice to be useful in real-time applications; the technique showed that stochastic gradient descent by backpropagation was effective in training CNNs. CNNs became in use but fell out of fashion due to the support vector machine as in and other simpler methods like linear classifiers.

In recent years, with the updating of computer hardware, especially GPU technology, the deep learning algorithms have been rapidly developed when solving problems in the fields of pattern recognition and image processing, and are more efficient and precise than traditional algorithms. Therefore, this paper uses a deep learning algorithm, YOLO, to achieve vehicle detection.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI June 2021- Available at www.ijraset.com

Considering the high-efficiency, one-stage object detection attracts more attention, proposed SSD method, which spreads out anchors of different scales to multilayers with ConvNet and enforces each layer to predict objects at a certain scale. It is a deconvolution single-shot detector (DSSD), which combines with SSD and augments them with deconvolution layers to introduce additional large-scale context for object detection, improving accuracy. The proposed method also features fusion single-shot multibox detector (FSSD) to enhance SSD with a novel and lightweight feature fusion module.

They concatenate features from multiple layers at different scales, followed by down sampling blocks to generate new feature pyramids, which are fed to multibox detectors to predict final detected results. YOLO uses a single feedforward convolutional network to predict object categories and locations, which can arrive at 5 fps. Then, YOLOv2 is proposed to improve YOLO in several aspects, such as using high-resolution layers, adding batch normalization on each convolution layer, and employing convolution layers with anchor boxes to predict bounding boxes instead of fully connected layers. With the development of the basic network, YOLOv3 —whose accuracy for human detection can reach 76% on a random dataset is proposed by replacing the backbone network with darknet-53 and employing multiscale features to detect the object. However, it still cannot work well on UAV-viewed (small) object detection due to lack of corresponding training data and limited receptive field

III. OBJECT DETECTION

Object detection is an important computer vision task that deals with detecting instances of visual objects of a certain class (such as humans, animals, or cars) in digital images. The objective of object detection is to develop computational models and techniques that provide one of the most basic pieces of information needed by computer vision applications.

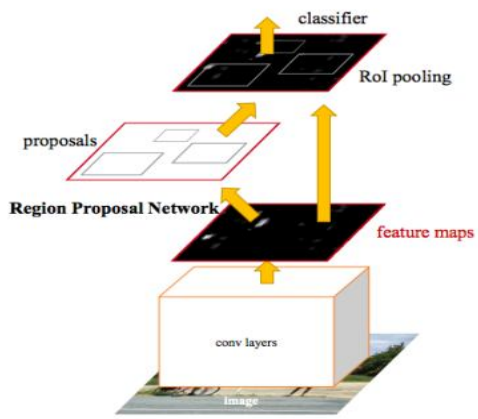

Figure 2-A Regression Sub-network

A. Difficulties and Challenges in Object Detection

Despite people always asking “what are the difficulties and challenges in object detection?”, actually, this question is not easy to answer and may even be over-generalized. As different detection tasks have totally different objectives and constraints, their difficulties may vary from each other. In addition to some common challenges in other computer vision tasks such as objects under different viewpoints, illuminations, and interclass variations, the challenges in object detection include but not limited to the following aspects: object rotation and scale changes (e.g., small objects), accurate object localization, dense and occluded object detection, speed up of detection, etc.

B. RCNN

The idea behind RCNN is simple: It starts with the extraction of a set of object proposals (object candidate boxes) by selective search [17]. Then each proposal is rescaled to a fixed size image and fed into a CNN model trained on ImageNet (say, AlexNet [18]) to extract features. Finally, linear SVM classifiers are used to predict the presence of an object within each region and to recognize object categories. RCNN yields a significant performance boost on VOC07, with a large improvement of mean Average Precision (mAP) from 33.7% (DPM-v5 [20]) to 58.5%. Although RCNN has made great progress, its drawbacks are obvious: the redundant feature computations on a large number of overlapped proposals (over 2000 boxes from one image) leads to an extremely slow detection speed (14s per image with GPU). Later in the same year, SPPNet [19] was proposed and has overcome this problem.

C. Fast RCNN

In 2015, R. Girshick proposed Fast RCNN detector [22], which is a further improvement of R-CNN and SPPNet [21, 19]. Fast RCNN enables us to simultaneously train a detector and a bounding box regressor under the same network configurations. On VOC07 dataset, Fast RCNN increased the mAP from 58.5% (RCNN) to 70.0% while with a detection speed over 200 times faster than R-CNN. Although Fast-RCNN successfully integrates the advantages of R-CNN and SPPNet, its detection speed is still limited by the proposal detection (see Section 2.3.2 for more details). Then, a question naturally arises: "can we generate object proposals with a CNN model?" Later, Faster R-CNN [23] has answered this question.

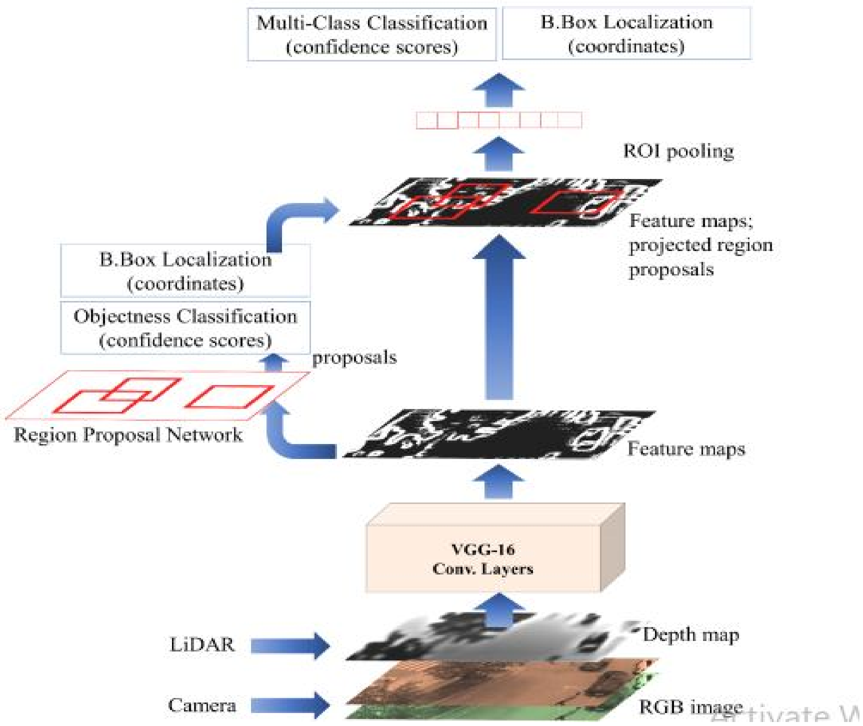

Figure 3-Object detection in cnn

D. Faster RCNN

In 2015, S. Ren et al. proposed Faster RCNN detector [23, 24] shortly after the Fast RCNN. Faster RCNN is the first end-to-end, and the first near-real-time deep learning detector (COCO mAP@.5=42.7%, COCO mAP@[.5,.95]=21.9%, VOC07 mAP=73.2%, VOC12 mAP=70.4%, 17fps with ZFNet [25]). The main contribution of Faster-RCNN is the introduction of Region Proposal Network (RPN) that enables nearly cost-free region proposals. From R-CNN to Faster RCNN, most individual blocks of an object detection system, e.g., proposal detection, feature extraction, bounding box regression, etc., have been gradually integrated into a unified, end-to-end learning framework.

E. You Only Look Once (YOLO)

YOLO was proposed by R. Joseph et al. in 2015. It was the first one-stage detector in deep learning era [26]. YOLO is extremely fast: a fast version of YOLO runs at 155fps with VOC07 mAP=52.7%, while its enhanced version runs at 45fps with VOC07 mAP=63.4% and VOC12 mAP=57.9%. YOLO is the abbreviation of "You Only Look Once". It can be seen from its name that the authors have completely abandoned the previous detection paradigm of "proposal detection + verification". Instead, it follows a totally different philosophy: to apply a single neural network to the full image. This network divides the image into regions and predicts bounding boxes and probabilities for each region simultaneously. Later, R. Joseph has made a series of improvements on basis of YOLO and has proposed its v2 and v3 editions [28, 29], which further improve the detection accuracy while keeps a very high detection speed. In spite of its great improvement of detection speed, YOLO suffers from a drop of the localization accuracy compared with two-stage detectors, especially for some small objects. YOLO's subsequent versions [28, 29] and the latter proposed SSD [27] has paid more attention to this problem.

F. Single Shot MultiBox Detector (SSD)

SSD [27] was proposed by W. Liu et al. in 2015. It was the second one-stage detector in deep learning era. The main contribution of SSD is the introduction of the multi-reference and multi-resolution detection techniques (to be introduce in Section 2.3.2), which significantly improves the detection accuracy of a one-stage detector, especially for some small objects. SSD has advantages in terms of both detection speed and accuracy (VOC07 mAP=76.8%, VOC12 mAP=74.9%, COCO mAP@.5=46.5%, mAP@[.5,.95]=26.8%, a fast version runs at 59fps). The main difference between SSD and any previous detectors is that the former one detects objects of 5 different scales on different layers of the network, while the latter ones only run detection on their top layers.

G. Quantitative Evaluation

The overall goal of this project is to measure the detection accuracy of identifying targets in videos. To measure detection accuracy:
Recall = Number of Detected Targets in all Frames Number of Ground-Truth Targets in all Frames.
Precision = Number of Detected Targets in all Frames Number of Detected Objects in all Frames.

Device	R-CNN (s)	YOLOv1 (s)	The modified YOLOv1 (s)
GPU time/image	6.9	0.14	0.11

This indicates that appearance information can complement the miss-detection due to the error in motion estimation. In addition, deep learning method can fully take advantage of manually labeled training dataset with over 95% classification accuracy.

H. Data Availability

The data used to support the findings of this study are
Available from the corresponding author upon request.

Dataset	Year	Description	Cites
TAS	2008	Consists of 30 images of 729x636 pixels from Google Earth and ~1,300 vehicles.	415
OIRDS	2009	Consists for 900 images (0.08-0.3m/pixel) captured by aircraft-mounted camera and 1,800 annotated vehicle targets. url: https://sourceforge.net/projects/oirds/	32
LEVIR	2018	Consists of ~22,000 Google Earth images and ~10,000 independently labeled Targets (airplane, ship, and oil-pot).url:https://pan.baidu.com/s/1geTwAVD	15
DOTA	2018	The first remote sensing detection dataset to incorporate rotated bounding boxes. Consists of ~2,800 Google Earth images and ~200,000 instances of 15 classes. url: https://captain-whu.github.io/DOTA/dataset.html	32

Table 1-Comparison of datasets

IV. CONCLUSION

Real-time object detection/tracking in HD videos is of great importance for video surveillance Detecting small objects in large scenes has long been a challenge. After carefully analyzing the different ways of detection of unidentifiable aerial vehicles in an economically and operatively gaining way. We went ahead to use the most efficient method which was an object detection model using YOLO methodology.

Also, we presented a holistic view of the drones/UAVs domains and provided detailed explanation and classification of their use in various domains and for different purposes, in addition to the different lethal/non-lethal security solutions as part of drones/UAVs countermeasure.

A. Abbreviations and Acronyms

UAV- Unmanned Aerial Vehicles
sUAS-Small Unmanned Aircraft System
RCNN- Convolutional Neural Networks
CNN- Convolutional Neural Networks
YOLO- You Only Look Once
EO- Electro-Optical
RCS- Radar Cross Section
SSD- Single-Shot Detector
SPPNet-Spatial Pyramid Pooling Network
GPU- Graphical Processing Unit
DPM- Deformable Parts Model
MAP- Mean Average Precision
SVM-Support Vector Machine
COCO- Common Objects in Context
D-SSD- Deconvolution Single-Shot Detector
F-SSD- Fusion Single-Shot Detector
RPN -Region Proposal Network

B. Equations

Loss function for object localization-

$$\mathcal{L}(\hat{y}, y) = \begin{cases} (\hat{y_1} - y_1)^2 + (\hat{y_8} - y_8)^2 + \ldots + (\hat{y_9} - y_9)^2 & , y_1 = 1 \\ (\hat{y_1} - y_1)^2 & , y_1 = 0 \end{cases}$$

Image classification and localization algorithm

$$y_{i,j} = [p_c \quad b_x \quad b_y \quad b_h \quad b_w \quad c_1 \quad c_2 \quad c_3 \quad c_4]^T$$

Target variable for a multi-class image classification

$$\hat{y} = c_i$$

V. ACKNOWLEDGMENT

The work was also presented as a minor project under the guidance of Dr.Vijay Pal Singh Dhaka, Head of department Computer Science and Communication.

REFERENCES

[1] U https://towardsdatascience.com/yolo-object-detection-with-opencv-and-python-21e50ac599e9
[2] https://deepnote.com/@deepnote/A-social-distancing-detector-using-a-Tensorflow-object-detection-model-Python-and-OpenCV-KBcEvWejRjGyjy2Ynx
[3] https://heartbeat.fritz.ai/detecting-objects-in-videos-and-camera-feeds-using-keras-opencv-and-imageai-c869fe1ebcdb
[4] https://www.kaggle.com/bulentsiyah/semantic-drone-dataset
[5] https://medium.com/nanonets/how-we-flew-a-drone-to-monitor-construction-projects-in-africa-using-deep-learning-b792f5c9c471
[6] https://blog.christianperone.com/2015/01/real-time-drone-object-tracking-using-python-and-opencv/
[7] https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=8599&context=etd
[8] https://circuitcellar.com/research-design-hub/write-an-object-tracking-drone-application/
[9] https://nanonets.com/blog/real-time-object-detection-for-drones/
[10] https://www.pyimagesearch.com/2017/09/18/real-time-object-detection-with-deep-learning-and-opencv/
[11] https://towardsdatascience.com/how-to-detect-objects-in-real-time-using-opencv-and-python-c1ba0c2c69c0
[12] https://towardsdatascience.com/how-to-train-your-own-object-detector-with-tensorflows-object-detector-api-bec72ecfe1d9
[13] https://www.hindawi.com/journals/mpe/2020/4013647/#abstract
[14] https://ieeexplore.ieee.org/document/8682048?denied=
[15] https://ieeexplore.ieee.org/document/9050881

[16] https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XXXVIII-1-C22/107/2011/isprsarchives-XXXVIII-1-C22-107-2011.pdf

[17] K. E. Van de Sande, J. R. Uijlings, T. Gevers, and A. W.Smeulders, "Segmentation as selective search for object recognition," in Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011, pp. 1879–1886.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks,"in Advances in neural information processing systems, 2012,pp. 1097–1105.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in European conference on computer vision.Springer, 2014, pp. 346–361

[20] R. B. Girshick, P. F. Felzenszwalb, and D. McAllester,"Discriminatively trained deformable part models, release 5," http://people.cs.uchicago.edu/ rbg/latentrelease5/

[21] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580–587.

[22] R. Girshick, "Fast r-cnn," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440–1448.

[23] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn:Towards real-time object detection with region proposal networks," in Advances in neural information processing systems, 2015, pp. 91–99.

[24] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn:towards real-time object detection with region proposal networks," IEEE Transactions on Pattern Analysis & Machine Intelligence, no. 6, pp. 1137–1149, 2017.

[25] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in European conference on computer vision. Springer, 2014, pp. 818–833.

[26] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.

[27] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y.Fu, and A. C. Berg, "Ssd: Single shot multibox detector,"in European conference on computer vision. Springer, 2016,pp. 21–37.

[28] J. Redmon and A. Farhadi, "Yolo9000: better, faster,stronger," arXiv preprint, 2017.

[29] "Yolov3: An incremental improvement," arXivpreprint arXiv:1804.02767, 2018.