AIRPLANE DETECTION AND RECOGNITION IN OPTICAL IMAGES

Aniket Pathak[1], *Ambuj Kumar*[1], *and Anil Kumar*[2],

[1]National Institute of Technology Delhi, [2]Indian Institute of Remote Sensing, Dehradun
171230012@nitdelhi.ac.in,ambuj01042000@gmail.com,anil@iirs.gov.in

ABSTRACT

Object detection area has been seeing rapid radical changes. Its application in image based sensor's outputs make most challenging task. The objective of object detection technique is to determine objects and classify objects. You Only Look Once or YOLO is one of the popular algorithms in object detection used by the researchers around the globe. Presently in this work, YOLO has been applied for airplane detection. YOLO algorithm presently gave very good results, when generalizing from natural images to optical remote sensing domains.
Although the Faster-R-CNN achieves good detection results, its accuracy is not high enough. To meet high detection accuracy and high-speed performance requirements of real-time operation, Redmon et al. [17] proposed another CNN-based unified target detection method. The proposed method, YOLO, predicts the bounding box and object class probability directly from the complete image in a single estimate. Since the entire detection pipeline is a single network, end-to-end optimization of the detection performance is straightforward.
You Only Look Once or YOLO is one of the popular algorithms in object detection used by the researchers around the globe. According to the researchers at Facebook AI Research, the unified architecture of YOLO is extremely fast in manner. The base YOLO model processes images in real-time at 45 frames per second, while the smaller version of the network, Fast YOLO processes an astounding 155 frames per second while still achieving double the mAP of other real-time detectors. This algorithm outperforms the other detection methods, including DPM and R-CNN, when generalizing from natural images to other domains like artwork.

Fast R-CNN, Faster R-CNN, Histogram of Oriented Gradients (HOG), Region-based Convolutional Neural Networks (R-CNN), Region-based Fully Convolutional Network (R-FCN), Single Shot Detector (SSD), Spatial Pyramid Pooling (SPP-net), YOLO (You Only Look Once).

Index Terms— Object detection, Deep Learning, *You Only Look Once or (YOLO)*

1. INTRODUCTION

Object detection applied in different fields is a computer-based technology belongs to computer vision and image processing area. Object detection deals with detecting occurrences of semantic objects of a specific type objects like; humans, buildings, or cars in natural images and videos [1]. Various research domains related to object detection include face detection, vehicle detection, pedestrian detection, animals and other objects. Object detection has applications in many areas of computer vision, medical science, remote sensing, image retrieval and video surveillance.
Object detection has been extensively used in computer vision tasks such as image marking[2], motion recognition[3], face detection, face recognition, video segmentation. Object detection has been extensively used in tracking moving objects. For example tracking football, tracking cricket bat movement, or moving person tracking in a video.

2. OBJECT DETREDCTION METHODS

Every object class has special features and these helps in detecting the objects – for example all rectangles are bound with straight lines. Object class detection uses special features of an object. For example, when looking for rectangles, objects, which are at a particular distance from camera, are sought. Similarly, when looking for circles, objects that are having center point are sought. A type of similar approach can be used for face detection where eyes, nose, and lips can be identified and features like skin color and distance between eyes can be identified.

Based on latest trend, methods for object detection were generally coming under domain of machine learning based approaches or more latest one called deep learning based approaches. In machine learning approaches, pre-processing step takes place to first define or extract features. Once features are extracted, later applying a technique like; as support vector machine (SVM) for classification. Latest upcoming deep learning techniques are capable to learn from data for point-to-point object detection while extracting features during training, and are typically based on convolutional neural networks (CNN). Machine learning based object detection approaches are like; Viola–Jones object detection framework based on Haar features, Scale-invariant feature transform (SIFT), Histogram of oriented gradients (HOG) features[5]. Secondly upcoming deep learning based object detections approaches are like; Region Proposals (R-CNN[6], Fast R-CNN[7], Faster R-CNN[8], cascade R-CNN[9]), Single Shot MultiBox Detector (SSD) [10]. You Only Look Once (YOLO) [11][12][13][4], Single-Shot Refinement Neural Network for Object Detection (RefineDet) [14], Retina-Net [15][9], Deformable convolutional networks [16][17].

3. DATA DETAILS

Data set used in this study were optical images with 1000 images for training, 1000 images for validation and later 1000 images for testing sample data. Target class identified was airplanes of total 10 types.

4. METHODOLOGY

In this study YOLOv5 object detection algorithms is implemented which is released recently on 9 June 2020. YOLOv5 released only four days back of releasing of YOLOv4 by an unofficial author Glenn Jocher. There are lots of controversies about the selection of the name "YOLOv5" and other stuff. Glenn introduced PyTorch based version of YOLOv5 with exceptional improvements. Hence he has not released any official paper yet.

This version is pretty amazing and outperforms all the previous versions in terms of COCO AP and got near to EfficientDet AP with higher FPS. For the generalization and accuracy purpose transfer learning is used to extract features and last few layer of model is retrained to get detection in our custom images.

The purpose of YOLOv5 is not to achieve the best Map, but instead:

- Ease of use
- Exportability
- Memory requirements
- Speed
- Map
- Market size

Our object detection deep neural network is implemented in pytorch on google colab GPU.

To train our detector we take the following steps:

- Install YOLOv5 dependencies
- Preparation of dataset for YOLOv5 Object Detection. Label is converted into 'text' format form 'xml' format.
- Define YOLOv5 Model Configuration and Architecture
- Train a custom YOLOv5 Detector
- Evaluate YOLOv5 performance
- Visualize YOLOv5 training data
- Run YOLOv5 Inference on test images
- Export Saved YOLOv5 Weights for Future Inference

5. RESULTS AND DISCUSSION

In this work, we implement a novel and fast multi-object detection approach that fully utilizes the complementarity of the optical images to robustly identify multiple objects (i.e. different types of aeroplane). The experimental results on the validation data is shown in below figure 1.

Figure 1 a)

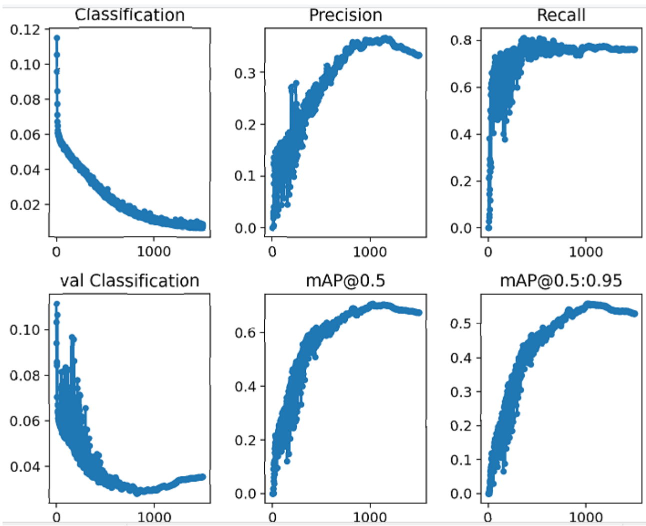

Figure 1 b)

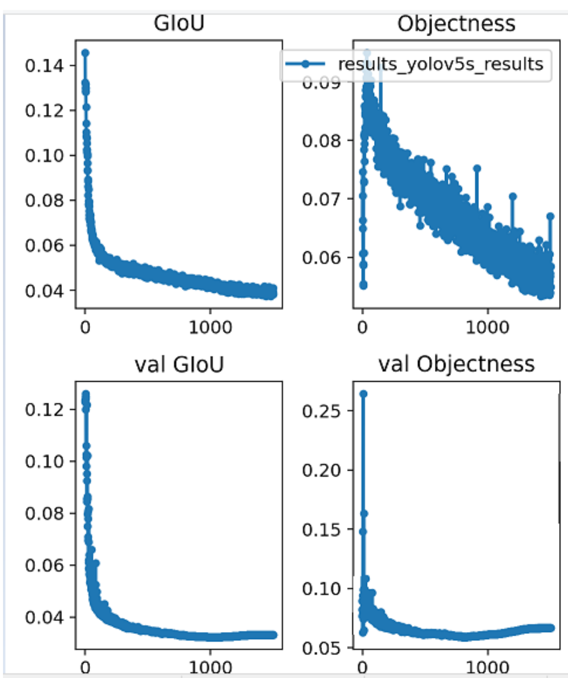

Figure 1 (a & b) Illustration of various performance evaluating parameters

We train our model for 1500 iteration and various observations are plotted in Figure I. It is clearly found that our precision, recall, Map are improving with respect to epochs. It's about 1500 iteration when these things gone in saturation. Hence no need of further trainings is required.

We also tested our model on validation data, and some samples of detection capability of YOLOv5 is shown in Figure 2.

This shows that our model is capable of detecting various objects with much confidence. It also yields an acceptable accuracy and can be implemented in real time as it takes about 21ms/per image in detection of multiple objects associated with that image., which means it can be executed rapidly and implemented online and that its performance is competitive with current popular methods. Because of its prominent learning capability, it avoids the problems associated with the effects of environmental illumination changes.

Figure 2 a)

Figure 2 b)

Figure 2 (a & b) Pictures of test data showing the detection results

It should also be mentioned, that the trained YOLO v.5's CNN was able to detect an airplane in the image, even if its contours were obscured by another object, ground, or in pretty different conditions, for example, photos of airplanes in the air taken from the bottom, but size in pixels of the airplane must be relatively big. If an airplane is not fully shown in the image, the network recognizes it only when most of it is present, that has been expected.

6. CONCLUSION

We implemented YOLOv5, a unified model for object detection. Our model is simple to construct and can be trained. directly on full images. Unlike classifier-based approaches, YOLO is trained on a loss function that directly corresponds to detection performance and the entire model is trained jointly. Fast YOLO is the fastest general-purpose object detector in the literature and YOLO pushes the state-of-the-art in real-time object detection. YOLO also generalizes well to new domains making it ideal for applications that rely on fast, robust object detection.

7. REFERENCES

1. Dasiopoulou, Stamatia, et al. "Knowledge-assisted semantic video object detection." IEEE Transactions on Circuits and Systems for Video Technology 15.10 (2005): 1210–1224.
2. *Ling Guan; Yifeng He; Sun-Yuan Kung (1 March 2012). Multimedia Image and Video Processing. CRC Press. pp. 331–. ISBN 978-1-4398-3087-1.*
3. Wu, Jianxin, et al. "A scalable approach to activity recognition based on object use." 2007 IEEE 11th international conference on computer vision. IEEE, 2007.
4. *Dalal, Navneet (2005). "Histograms of oriented gradients for human detection"(PDF). Computer Vision and Pattern Recognition. 1.*
5. *Ross, Girshick (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation" (PDF). Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE: 580–587. arXiv:1311.2524. doi:10.1109/CVPR.2014.81. ISBN 978-1-4799-5118-5.*
6. *Girschick, Ross (2015). "Fast R-CNN" (PDF). Proceedings of the IEEE International Conference on Computer Vision: 1440–1448. arXiv:1504.08083. Bibcode:2015arXiv150408083G.*
7. *Shaoqing, Ren (2015). "Faster R-CNN". Advances in Neural Information Processing Systems. arXiv:1506.01497.*
8. Jump up to:[a] [b] *Pang, Jiangmiao; Chen, Kai; Shi, Jianping; Feng, Huajun; Ouyang, Wanli; Lin, Dahua (2019-04-04). "Libra R-CNN: Towards Balanced Learning for Object Detection". arXiv.org. Retrieved 2020-05-23.*
9. *Liu, Wei (October 2016). SSD: Single shot multibox detector. European Conference on Computer Vision. Lecture Notes in Computer Science. 9905. pp. 21–37. arXiv:1512.02325. doi:10.1007/978-3-319-46448-0_2. ISBN 978-3-319-46447-3.*
10. *Redmon, Joseph (2016). "You only look once: Unified, real-time object detection". Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. arXiv:1506.02640. Bibcode:2015arXiv150602640R.*
11. *Redmon, Joseph (2017). "YOLO9000: better, faster, stronger". arXiv:1612.08242 [cs.CV].*
12. *Redmon, Joseph (2018). "Yolov3: An incremental improvement". arXiv:1804.02767 [cs.CV].*
13. *Zhang, Shifeng (2018). "Single-Shot Refinement Neural Network for Object Detection". Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition: 4203–4212. arXiv:1711.06897. Bibcode:2017arXiv171106897Z.*
14. *Lin, Tsung-Yi (2020). "Focal Loss for Dense Object Detection". IEEE Transactions on Pattern Analysis and Machine Intelligence. 42 (2): 318–327. arXiv:1708.02002. Bibcode:2017arXiv170802002L. doi:10.1109/TPAMI.2018.2858826. PMID 30040631.*
15. *Zhu, Xizhou (2018). "Deformable ConvNets v2: More Deformable, Better Results". arXiv:1811.11168 [cs.CV].*
16. *Dai, Jifeng (2017). "Deformable Convolutional Networks". arXiv:1703.06211[cs.CV].*
17. *Redmon, J.;Farhadi, A. (2017) 'YOLO9000:Better, Faster, Stronger, In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525*