

Object Detection and Classification for Autonomous Drones



Harit Ahuja, Vedant Kuhar, R. I. Minu

Abstract – Detecting and classifying objects in a single frame which consists of several objects in a cumbersome task. With the advancement of deep learning techniques, the rate of accuracy has increased significantly. This paper aims to implement the state of the art custom algorithm for detection and classification of objects in a single frame with the goal of attaining high accuracy with a real time performance. The proposed system utilizes SSD architecture coupled with MobileNet to achieve maximum accuracy. The system will be fast enough to detect and recognize multiple objects even at 30 FPS.

Index Terms – SSD; FPS; MobileNet; Tensorflow

I. INTRODUCTION

Computer Vision is an interdisciplinary field that has been gaining huge amount of traction in the recent years. Object detection is a basic part of computer vision used for detection of objects like person, face, surveillance etc. Object detection is the combination of classification and localization. In image classification, the input image is processed and labelled accordingly. Localization is the method to determine coordinates of the detected instance and their corresponding boundaries.

In this paper, we deduce a model which consists of object detection and classification on a platform where the processing power is not powerful. We focus on training the MS COCO dataset which is considered to be one of the finest dataset for performing object detection.

The model presented in this paper requires just the CPU to deploy in realtime. The model uses Tensorflow Deep learning API as its backend. Using the power of convolutional neural network the detection accuracy is increased in parallel with the classification accuracy and not by reducing the speed. For obtaining the results, a video recorded from a drone will be used as input to perform the following detections.

This paper focuses the use of SSD architecture and MobileNet model to perform the detection and classification.

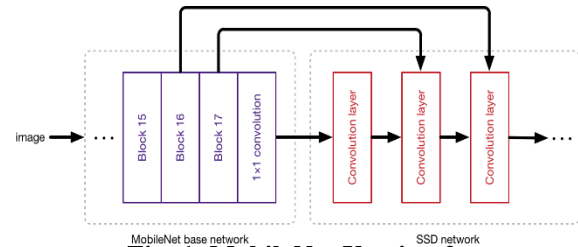


Fig. 1. MobileNet Version 2

II. RELATED WORK

The paper by Bernado Augusto [1] that used 200 field-captured images of plates as the corresponding dataset and trained them using CNN. The formula used for Classification accuracy used is:

$$\text{Classification Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Positives} + \text{Negatives}}$$

The result was determined by identifying the determining the individual region of the corresponding image in the frame.

Shaoqing Ren in [2] written by focuses on the implementation of RPN and Faster R-CNN for detections. It made us of dataset that contained numerous images that were required to build the model and perform the necessary processing.

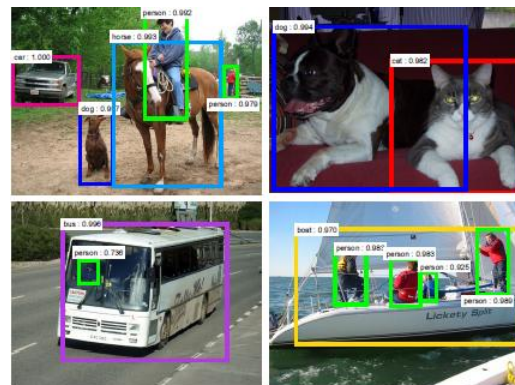


Fig. 2. Objects successfully detected in [2]

The corresponding method is able to specify different objects that are of various sizes and strength. It minimizes the object function using a mathematical formula. The corresponding formula for the image is:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*)$$

Manuscript received on February 10, 2020.
Revised Manuscript received on February 20, 2020.
Manuscript published on March 30, 2020.

* Correspondence Author

Harit Ahuja, Computer Science and Engineering Department, SRM Institute of Science and Technology, Kattankulathur, Chennai, India

Vedant Kuhar, Computer Science and Engineering Department, SRM Institute of Science and Technology, Kattankulathur, Chennai, India

DR. R.I. Minu Computer Science and Engineering Department, SRM Institute of Science and Technology, Kattankulathur, Chennai, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Where ‘i’ is mentioned as the index and p_i is termed as the predicted probability.

This method implements an object based detection system which is able to achieve high accuracy and run at 5-17 frames per second.

Kevin Okiah in [3] discusses the three meta architectures that are SSD, Regional Proposal Network. The dataset used for this paper was COCO (Common Objects in Context) which contains a large number of various kinds of objects.



Fig. 3. Detection in different driving conditions in [3]

The above picture shows the object detection under four different driving conditions that is snowy, day, night and Rainy.

The conclusion stated that after comparing the different architectures, the region-based algorithms such as Faster R-CNN and R-CNN tend to have a high accuracy and the SSD based algorithms have faster inference time.

The paper written by Souhail Guennouni in [4] is used to implement the high speed object detection on the embedded platforms which do not have high processing capabilities. The aim of this paper is to use this method to use in surveillance systems with the help of distributed cameras and the detection takes place in a backend server.

The architecture used to implement this method is called the cascade classifier and 4000 positive images were used as the dataset.

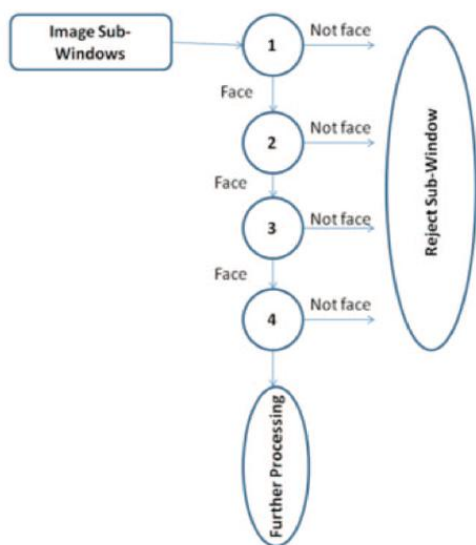


Fig. 4. Cascade classifier used in [4].

This method concludes that object detection can be deployed in different platforms. The next step of this corresponding work will be to improve the other hyper parameters by tuning accordingly.

In [5] Zhong-Qiu Zhao explains in detail about the CNN architecture and then focuses of the generic object detection architectures and then finally it discusses few tips and tricks to improve the object detection. The model developed successfully categorizes the objects in the corresponding frame and saves them as they are detected.

III. IMPLEMENTATION

A. Overview

For achieving high accuracy detection rate in a faster speed, our method uses the combination of Single Shot Detector(SSD) 300 with Mobilenet. Mobilenet is being used as it more suitable for getting implemented where they computational power is less.

B. Obtaining The Dataset

A good dataset will work with the model to give a high precision for detection and classification of the object. For the dataset, we have used the MS COCO (Common Object in Context) dataset. The COCO dataset contains images from everyday scenarios. The corresponding dataset contains images of objects like person, bicycle, car, traffic light etc. The corresponding dataset contains 91 common objects with more than 5000 labeled instances. The advantage of using COCO dataset over ImageNet is because the COCO dataset has more instances per category.

C. Model

The images were resized to size 224 X 224 to lower the cost of the computation. Various filters are applied on the resized images to remove the noise from the images. Then, the images are used in training and testing stage. The model that is used is SSD architecture which has a small 3X3 convolution for feature extraction.

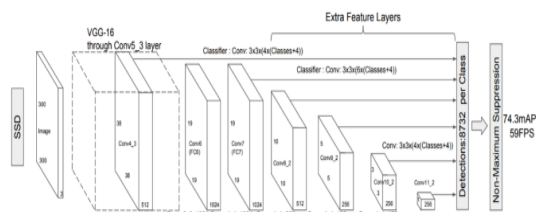


Fig. 5. Single Shot Detector Architecture

After the feature extraction, a feature layer is obtained which is of the size $M \times N$ which contains a channel called ‘p’. For every location, we get bounding box called ‘k’ which has different aspect ratios. The loss function is used to minimize the faulty detections that are due to the displacement of the different detection box.

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c) + \alpha L_{loc}(x, l, g))$$

In the above formula, N is the quantitative instance and alpha is the localization strength.

D. Training

For training of the dataset, 80% of the dataset is kept for training and 20% is kept aside for testing.

The first step in training is the balancing of the class distribution for the corresponding imbalanced data. Next, the features are extracted and selected using the $(\alpha, \beta) - k$ Feature Selection Set selection.

The dataset is trained and classified to create a model. The model will be used to detect the objects during the testing phase to determine the results. The dataset was trained using 16 GB of RAM and NVIDIA GeForce RTX 2080 Ti.

IV. RESULTS

The combination of SSD architecture and the MobileNet model yielded accurate results. The accuracy percentage was approximately 82%. The objects that were detected were enclosed in a bounding box and the name of the classified object was displayed above the bounding box. The proposed model can be run on any low power devices like a drone. It can successfully detect and classify 91 different objects. The results were tested on a pre-recorded video that was recorded from the camera attached to the gimble of a drone. The drone didn't use any external GPU to detect the corresponding images.



Fig. 6. Successful detection and classification of objects.

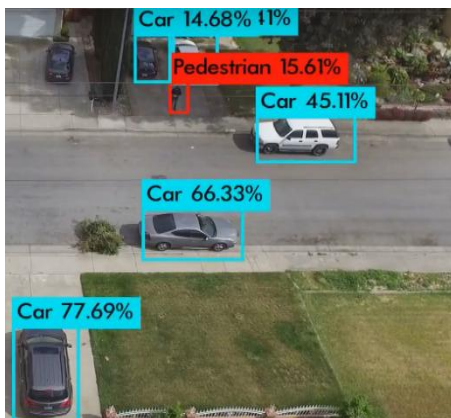


Fig. 7. Detection of car and pedestrian

Different models were compared on their performance that were dependent on mAP, FPS and boxes. mAP (mean Average Precision) is used to determine the accuracy of the corresponding object detection algorithms. FPS is the frame per second by which the following object is detected. Boxes are the bounding boxes that are illustrated on the objects that are detected.

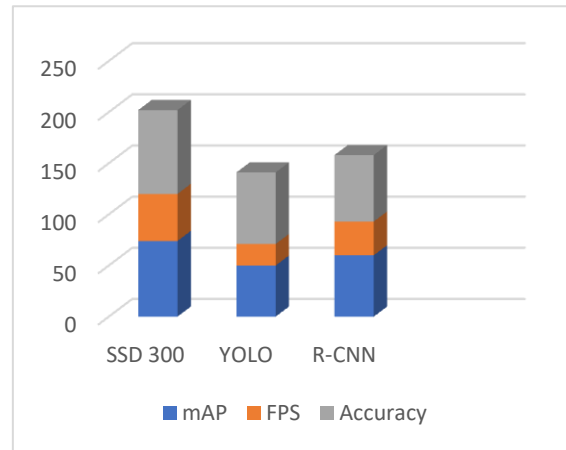


Fig. 8. Comparison of different models

The training time compared were compared among the different model that included SSD 300, YOLO, R-CNN and faster R-CNN.

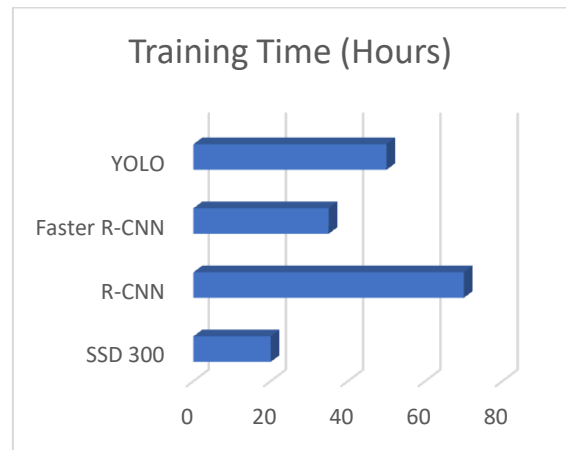


Fig. 9. Training time computed in Hours

So, by interpreting the graphs, it is evident that SSD 300 performs better in most of the aspects. The training time taken is also less compared to the other models. Hence, SSD 300 was the most convenient model to use with the MobileNet that results in high accuracy.

V. CONCLUSION

So, the results conclude that the SSD architecture and the MobileNet Model together is capable of detecting and classifying objects that are present in the images at a very high speed and accuracy. Few results were incorrectly detected as the presented images were not clear or had a lot of noise in them. The system developed proves that object detection can be used in platforms where the computation power is less. This system can be trained for mostly any type of object that has to be detected and classified.

The limitation of this project is that many of the objects are inaccurately detected due to the noise in the frame or bad quality of the video. The second limitation of this paper is the quality of the video, the system cannot process high quality video as the model for detection and classification is built for low computational devices.

The next step of improving this work can be enhanced by using the concept of parallelism. This can work by running separate tasks in a parallel order. By using this concept, the algorithm will be able to perform accurate detections on high quality video as well. Another step to improve the data can be done by creating a module for filtering the frames to remove the noise to avoid mismatch between the detected objects. We are also intrigued to see how our proposed model performs on different weather conditions and situations.

REFERENCES

1. de Oliveira, B. A. G., Ferreira, F. M. F., & da Silva Martins, C. A. P. (2018). Fast and Lightweight Object Detection Network: Detection and recognition on resource constrained devices. *IEEE Access*, 6, 8714-8724.
2. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91-99).
3. Simhambhatla, R., Okiah, K., Kuchkula, S., & Slater, R. (2019). Self-Driving Cars: Evaluation of Deep Learning Techniques for Object Detection in Different Driving Conditions. *SMU Data Science Review*, 2(1), 23.
4. Guennouni, S., Ahaitouf, A., & Mansouri, A. (2014, October). Multiple object detection using OpenCV on an embedded platform. In *2014 Third IEEE International Colloquium in Information Science and Technology (CIST)* (pp. 374-377). IEEE.
5. Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 30(11), 3212-3232.
6. Viola, P., & Jones, M. (2001). Robust real-time object detection. *International journal of computer vision*, 4(34-47), 4.
7. Lienhart, R., & Maydt, J. (2002, September). An extended set of haar-like features for rapid object detection. In *Proceedings. international conference on image processing (Vol. 1, pp. I-I)*. IEEE.
8. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).
9. Papageorgiou, C., & Poggio, T. (2000). A trainable system for object detection. *International journal of computer vision*, 38(1), 15-33.
10. Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2009). Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, 32(9), 1627-1645.
11. Borji, A., Cheng, M. M., Jiang, H., & Li, J. (2015). Salient object detection: A benchmark. *IEEE transactions on image processing*, 24(12), 5706-5722.
12. Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448)
13. Peng, X., & Schmid, C. (2016, October). Multi-region two-stream R-CNN for action detection. In *European conference on computer vision* (pp. 744-759). Springer, Cham.
14. Poirson, P., Ammirato, P., Fu, C. Y., Liu, W., Kosecka, J., & Berg, A. C. (2016, October). Fast single shot detection and pose estimation. In *2016 Fourth International Conference on 3D Vision (3DV)* (pp. 676-684). IEEE

AUTHOR'S PROFILE



Harit Ahuja, is current pursuing his Bachelor's degree in Computer Science and Engineering from SRM Institute of Science and Technology, Chennai, to be completed in the year 2020. His area of interest includes Artificial Intelligence, Machine Learning and its various applications with business analytics and image processing. He aims to study the various ways in which Machine learning can be used to further improve upon the existing approaches using real world evidence and various learning techniques. His recent focus has been on leading research efforts to develop advanced machine learning, data mining for deriving data-driven and to solve real world problems.



Vedant Kuhar, is currently pursuing his Bachelor's degree in Computer Science and Engineering from SRM Institute of Science and Technology, Chennai, to be completed in the year 2020. His research interest spans topics in Machine learning and its intersection with healthcare applications and/or business analytics. He aims to study the various ways Machine learning can be used to further improve upon the existing

approaches using real world evidence and various learning techniques. His recent focus has been on leading research efforts to develop advanced machine learning, data mining and visual analytics methodologies for deriving data-driven insights from real world healthcare data to facilitate learning health systems. His current work primarily includes developing statistical and Machine learning methods to generate insights in the healthcare.



Dr R I Minu, is presently working as Associate Professor in the Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur. She had received her BE degree in Electronic and Communication Engineering from Bharadhidasan University, 2004 and ME degree in Computer Science Engineering from Anna University, 2007. She obtained her Ph.D. degree in Computer Science Engineering from Anna University, 2015, her thesis titled "Ontology Enhanced Image retrieval for Semantic Web". She has totally 13 years of teaching experience. Her research interest spans from computer vision and Internet of Things. Her profound research is in reducing the dimensionality of an image and to provide Semantic to it. In recent years she tries to implement those algorithm to the IoT devices and extended her research in those areas. She is an active participant of IIT-Bombay e-Yantra project initiative (Robotic), also she received the funding from MHRD for setup robotics lab. Some of her projects are funded by TNSCST student Projects Scheme and CSI students chapter had recognized her project. She has published more than 50 research papers in refereed International Journals and International Conferences as well.