IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115193

OBJECT DETECTION USING ARTIFICIAL INTELLIGENCE

Arji Bhandhavi, S Rishika

Student, Computer Science Engineering, Jain School of engineering and Technology, Bangalore, India.

Student, Computer Science Engineering, Jain School of engineering and Technology, Bangalore, India.

Abstract: Autonomous vehicles using Artificial Intelligence (AI) technologies requires various sensors such as radars, lidar, ultrasonic, etc. to have the human visual perception in monitoring the road. Wide angle camera is often used for better coverage and experience for view. Those sensors generate massive amount of data that could be processed with the cloud computing through the wireless communication. The cloud computing may not be a feasible solution, as for real- time detection systems. In this work, we examine the implementation of the deep-learning and real-time object detection on the edge devices that is connected to the wide-angle camera. This system can achieve real-time object detection with a latency of less than 0.2 ms. This model also helps to mitigate the distortion that is introduced by the wide-angle camera. Detection system will be able to warn user of his or her surrounding road conditions.

Keywords: Artificial Intelligence, Deep-Learning, Edge Computing, Real-time Object Detection.

I. INTRODUCTION

Machines have gained AI over the last two decades with speed. Amongst the various applications of Artificial Intelligence, Computer Vision has gained popularity in road safety, driving a fundamental change in transportation safety using the concept of object detection. Objection detection algorithms are used across various industries with standard cameras. On roads, in self-driving vehicles these cameras act as sensors that help an autonomous vehicle sense its surroundings. Most cameras, optical disturbance caused by a mismatch between the camera's photosensor and the optical lenses that the camera uses are common. The presence of such disturbances can inevitably affect the performance of object detection models. In this there will be examine and evaluate some of the present algorithms, also utilizing data to train and improve models. The model is then implemented in the edge devices as part of the mobile Internet of Things (IoT) setup.

II. OBJECT DETECTION ALGORITHMS

Object detection is a part of a broader field of computer vision that enables machines to recognize objects within an image frame.

A. Two-stage Object Detection Algorithms

Two-stage object detection methods are mainly region- based Convolutional Neural Networks (CNN). Two-stage object detection models take a classifier and apply it on various places in which there may be potential objects. These locations are proposed by the regional proposal network. Some of these algorithms include:

1) Deformable Part Models (DPM): Unlike the CNN, one of the first Deformable Part Model (DPM) [1] for object detection published about a decade ago describes the object detection system based on mixture of multiscale part models to represent the high variability of objects, interclass or intraclass alike as in [1]. Interclass variability is due to the physical difference between objects and hence are categorized into different classes and categories. Intraclass variability can be due to difference in illuminations, angle of view, or even color.

2) Region-based Convolutional Neural Networks: Region- based CNN (R-CNN) forms the bases of most two-stage object detection models. R-CNN as in [2] generates potential regions within an image. This selective search technique is a common algorithm that proposes regions which is predicted to contains objects. These proposed regions are considered candidates, or Regions of Interest (RoI). Followed by running a classifier through these proposed boxes. CNN features are extracted independently from the regions to undergo classification, where post-processing is used to refine the bounding boxes, eliminate duplicate detections, and rescore the boxes based on other objects within image.

3) Fast-RCNN: Fast R-CNN [3] defers from R-CNN in the feature extraction stage. Instead of extracting CNN feature vectors independently for each proposed region, fast R-CNN aggregates them into one CNN forward the entire image.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115193

B. One-stage (Fast) Object Detection Algorithms

The primary reason they are considered 'fast' is because these models skip the region proposal stage, running detection directly over a compact statistical sampling of possible locations in which objects might exist, usually a fixed number of predictions over an image grid pre-determined by the algorithm.

The key focus of these detectors is on the inference time rather than on accuracy. State-of-the-art single shot object detectors give much better inference time without sacrificing much on the accuracy. Some of these algorithms include:

1) Single Shot Detector: The Single Shot Multibox Detector (SSD) [4] uses a pyramidal hierarchy that adds several convolutional feature layers of decreasing size.

2) Retina Net: Retina Net [5] focuses on the concept of Focal Loss, where more weights are given to hard, easily misclassified examples (such as a partial object), while easier samples (such as a clear sky) are given less weights.

You-Only-Look-Once (YOLO): Compared to other object detection algorithms, You-Only-Look-Once (YOLO) [6] has been considered the state-of-the-art due to its fast inference time and ability to infer objects with only one look at the image. First introduced in late 2015, the YOLO presents a new approach to the object detection tasks, object detection is framed as a regression problem that can enable bounding boxes and class confidence to be predicted over a grid overlaying the image. This enables a single stage detection network that does not need a regional proposal step to run over the image, predicting only a finite number of bounding boxes. Because of that, it is a contending candidate to enable real- time object detection. The author has since developed version 2 [7] and version 3 [8] of the YOLO, with a recent version 4 [9] being conceived by another group of researchers due to its popularity.

We can reasonably conclude that two-stage detectors like Faster R-CNN have better performance in accuracy with some sacrifice in inference speed. On the other hand, one-stage detectors like the YOLO are strong candidates for real-time

III. WIDE-ANGLE CAMERAS AND OPTICAL DISTORTION

The field-of-view of cameras today can range 90 to 170 or even 180 degrees. However, the ineluctable large distortion of the images produced by fisheye lenses cause the images captured with wide-angle lenses to suffer from spatial distortion Fig. 1.

Equation (1) is a mathematical representation of the field-of- view of a camera system, where α is the angular field-of-view in degree, d is the horizontal size of the sensor and f is the focal length of the system, both of which are in millimeters Fig. 2.

Fig. 1. Cameras with a wider filed-of-view (bottom) are able to capture more imagery within the same physical space as opposed to those with a smaller field-of-view (top). Source: Google images.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115193

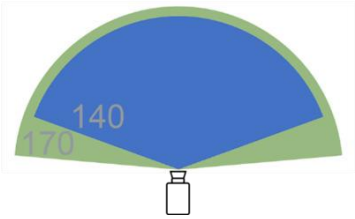

Fig. 2. A representation of a 140 degree and a 170-degree image capturing device

A. EFFECT OF OPTICAL DISTORATION:

Optical disturbance may cause deformation of the image quality and presence that may not necessary be obvious to the human eye. Such distortions can be represented on gridlines that may better present its effect to the observer. Fig. 3 shows an undistorted pattern, while Fig. 4 shows a distorted pattern along with a picture that is optically distorted. The presence of optical distortions can adversely affect the representations of objects that are present within the image.

IV. DATA AUGMENTATION AND PERFORMANCE IMPROVEMENTS

Data augmentation is introduced to create various renditions of comparable scenarios. As a technique, it can better generalize scenarios for the model. Such a method should also not affect the model's inference time, which is a very important aspect in real-time application. The key difference between data augmentation and data pre-processing is that data pre- processing is applied to both the training data and the test.

Fig. 3. Example of an undistorted pattern along with a picture taken with rectilinear lens. Source: Google images.

Fig. 4. Example of a distorted pattern along with a picture taken with wide- angle lens. Though able to fit in more of the physical scenery as compared to Fig. 3, the picture suffers from significant optical (barrel) distortion that can affect the representations of objects present within the image. Source: Google images.

data, whereas data augmentation is only applied to the training datasets. More than just having a more generalized dataset, we wanted to see whether augmenting the composition of the types of images within the training dataset could affect or improve a model's performance. To do this, we created various training datasets, and a test set in which we can standardize our testing.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115193

Fig. 5. A snapshot of Dataset T, which consist of a mixture of normal and optically distorted images.

A. Optically Distorted Dataset

One of the image manipulation technique applicable for our use case is purposefully distorting normal images to give fisheye-like images. However, presents the VOC-360 dataset, which is extrapolated from the existing public dataset PASCAL VOC2012 to give a total of 39,575 images, implemented with a mapping model in MATLAB. To ensure a fair comparison, we created test set Dataset Fig 5 that consists of 30.3 percent normal (undistorted images) and 69.7 percent optically distorted images from various sources. We also created three training Datasets A, B, and C. Each of these datasets are created with different percentages of distorted and normal images as represented in TABLE I.

TABLE I

DATASET COMPOSITIONS

Dataset	Image Composition	
	Distorted (%)	Normal (%)
A	100	0
B	66.6	33.3
C	50	50
T	30.3	69.7

B. Initial Tests and Results

Before we begin training our models, we ran the YOLO- v3 model that is pre-trained on the ImageNet through two test datasets, one consisting of normal, non-distorted images and the other is Dataset T, which consists of both normal and optically distorted images. We summarized the results in TABLE II. Through the results obtained we can prove our initial hypothesis that the performance of an object detector does decrease when subjected to optically distorted images due to the difference in the objects' physical representation that is deformed, whether partially or fully, in optically distorted images.

TABLE II

INITIAL RESULTS

Model	Test Set	mAP
YOLOv3	Partial VOC2012	57.69%
YOLOv3	Dataset T	31.61%

C. More Training and Results

As opposed to training on more object classes, our models are trained to detect up to 20 different object classes, most of which are objects commonly seen on the road, such as pedestrians, cars, and bicycles. The objective of this is to streamline our model for road application and reduce com- mutational power while increasing the detection performance of the system. Each of the models was trained up to 40,000 iterations, and with their parameters set similarly to ensure no biasness when running through the test dataset.

The model trained with a mixture of the types of images performed better. Having a model trained purely on distorted images such as that of Model A may not necessarily improve the model's performance.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115193

TABLE III

RESULTS WITH MODELS FROM DATASETS

Model	Test Set	mAP
Dataset A	Dataset T	28.80%
Dataset B	Dataset T	36.72%
Dataset C	Dataset T	30.13%

Through the experiments,
we can see that having the goal that allows the model to be able to better generalize will give us better performance, in this case in the form of having a more equal ratio between normal and distorted images.

Fig. 6. A real-time video being fed into and inferenced offline on the Atlas 200 DK

V. IMPLEMENTATION WITH EDGE COMPUTING

An edge device refers to any hardware deployed at the circumstance where input data, in our case a video feed, is required to be analyzed. Traditionally, cloud computing has been used to develop and deploy edge devices by enabling computational capabilities on edge device. However, edge devices are generally distributed, and linking edge devices to traditional cloud computing technologies introduces
challenges on network capacities that are limited by current technologies and capabilities. Implementing algorithms on an edge device can directly address data transmission challenges, in turn improving the overall speed of the system. We tried in to use a standard System-on-Chip (SoC) to try out real-time edge computing. Significant external hardware accelerator is needed in, which can be achieved with a more efficient chip design.
The Huawei Atlas 200 DK is used in this work. As opposed to a standard SoC such as the latest Raspberry Pi 4B, the Atlas 200 DK can run a full-size CNN with minimal power consumption. It is also possible to use the Atlas 200 DK for other applications, such as detecting whether a person is wearing mask on the streets. Fig. 6 shows the snapshot of demonstration of the detection.

VI. CONCLUSION

Our work focuses on the edge computing capabilities of the Huawei Atlas 200 DK. The hardware can run a full suite of solution that enables IoT devices to not only connect to the cloud, but also real-time on the edge. We utilized data augmentation to improve performance on application- specific object detection models. An optimal mix of images that can give us better accuracy. Future works in this area may include introducing distortion parameters into the model inputs and training of the algorithm such that the model will determine This can be in the form of a user input and can give the model a more representative data in which its performance can be calibrated based on the types of cameras being used.

ACKNOWLEDGMENT

We want to acknowledge the Electronics Design Lab for their support during the Covid-19 circuit breaker period and thereafter. Special thanks to **Huawei Singapore** (**Ms. Cather- ine Liu, and group**) for the technical support, training, and sponsorship.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022
DOI: 10.17148/IJARCCE.2022.115193

REFERENCES

[1] P. F. Felzen szwalb, R. B. Girshick, D. Mcallester, and D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1627–1645, 2010.
[2] R. Girshick, J. Donahue, T. Darrell. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014.
[3] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), 2015.
[4] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and
A. C. Berg, "SSD: Single Shot MultiBox Detector," Computer Vision – ECCV 2016 Lecture Notes in Computer Science, pp. 21–37, 2016.
[5] T. Lin, P. Goyal, R. Girshick, K. He and P. Dollar, "Focal Loss for Dense Object Detection," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 2999-3007, doi: 10.1109/ICCV.2017.324.
[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[7] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
[8] J. Redmon, and A. Farhadi, "YOLOv3: An Incremental Improvement", arXiv e-prints, 2018
[9] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection", arXiv e-prints, 2020.
[10] J. Fu, I. V. Bajic, and R. G. Vaughan, "Datasets for Face and Object Detection in Fisheye Images", arXiv e-prints, 2019.
[11] T. H. Teo, Y. S. Tan, and W. M. Tan, "Deep-Learning learning by design," 10th Annual International Conference on Computer Science Education: Innovation and Technology (CSEIT 2019), pp.87-90, Thai- land, 2019, doi:10.5176/2251-2195 CSEIT19.163.
[12] T. H. Teo, W. M. Tan and Y. S. Tan, "Tumour Detection using Con- volutional Neural Network on a Lightweight Multi-Core Device," 2019 IEEE 13th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC), Singapore, Singapore, 2019, pp. 87-92, doi: