Object Detection for Autonomous Vehicle

Md.Al-Samiul Amin Rishat (id:17-35746-3)[a], Oishi Chowdhury (id:18-36807-1)[a], Sonaly Akter (id:17-33586-1)[a], (Raffe,Al Shazid(id:17-33356-1)[a]

[a]*Department of Computer Sciences, American International University-Bangladesh*

Abstract

The zone of computer vision is developing ceaselessly with the increasing interaction and improvement to supply a comfortable interaction between human and machines. One of the key perspectives within the handle of computer vision is protest discovery. Either objects can be recognized mostly or close to the first objects. The precision in recognizing the objects can be made strides by using state-of-the-art profound learning models like faster-Regional Convoluted Neural Arrange (faster-RCNN), You Merely See Once show (YOLO), Shot Detector (SSD) etc. Conventional calculations can't recognize objects as efficiently due to its confinements. Though the profound learning models require large amount of information for preparing the dataset, which has more asset and labour intensive in nature. The determination of calculation decides its exactness in object detection as well as its unwavering quality. The acknowledgment and classification of object begins with planning dataset taken after by part the dataset into preparing

Keywords: Deep learning, Computer Vision, Object detection, COCO data set, YOLO, Faster-RCNN, ResNet

1. Introduction

Protest discovery is the sprouting investigate region within the field of computer vision. The ability to distinguish and recognize objects either in single or more than one picture frame can pick up extraordinary significance in different ways as whereas driving the vehicle, the driver cannot distinguish objects appropriately due to the shortage of consideration, reflection of light, anonymous objects etc. which may lead to deadly mischances. In arrange to overcome such perceptible issues, independent vehicles and Progressed Driver

Help System (ADAS) took the liberal assignment of protest discovery and classification. The errand of computer vision is performed within the taking after steps:
1. Classification of protest in image
2. Localization of question in image
3. Question detection
4. Division of image.
The application of protest location can be found in progressed mechanical autonomy, defense systems, reconnaissance frameworks, space investigate, confront acknowledgment and numerous more. The thought of self-driving vehicles has been progressing with the movement in techniques related to the assignment of distinguishing and extricating highlights from the objects. Object detection for self-driven vehicles could be a non-trivial assignment in arrange to explore on the road.[1] The advancement of profound neural systems have changed the viewpoint of computer vision over the conventional strategies. Routine machine learning and computer vision models plays a predominant part within the handle of protest classification, be that as it may the industry presently intensely depends on the profound learning based classifiers. The development of graphical preparing units (GPU) has driven to more proficient and comfort in achieving the task of question classification through profound neural arrange models. These models attempt to memorize imperative highlights comparing to each course that are propelled from the organic structure of neurons in people. Google's Tensorflow is one such machine learning system which works on dataflow programming among a rang of the task.Hubs in TensorFlow speak to numerical operations and the chart edges represent multidimensional clusters called as Tensors.[3] Tensorflow question discovery API is competent in recognizing objects in an picture with great exactness it is additionally able to detect objects in live spilling video with a great degree of exactness in which speed of frames is around 20–30 outlines per second. We propose the utilize of Tensorflow protest discovery API for our dataset to prepare and test the dataset in arrange to distinguish objects effectively for an independent vehicle.

1.1. Objective

The main objective of this work is to detect vehicles in still images which employ the use of the wheels of the vehicle.

The goal of this work is to improve the detection technique from the side view of vehiclesin still images.

The main contributions of this work are:

To modify the existing vehicle detection methods for achieving better quality vehicle detection output. To apply edge detection algorithm and Hough circle detection algorithm in the side view of still vehicle images.[6]

2. Literature Review

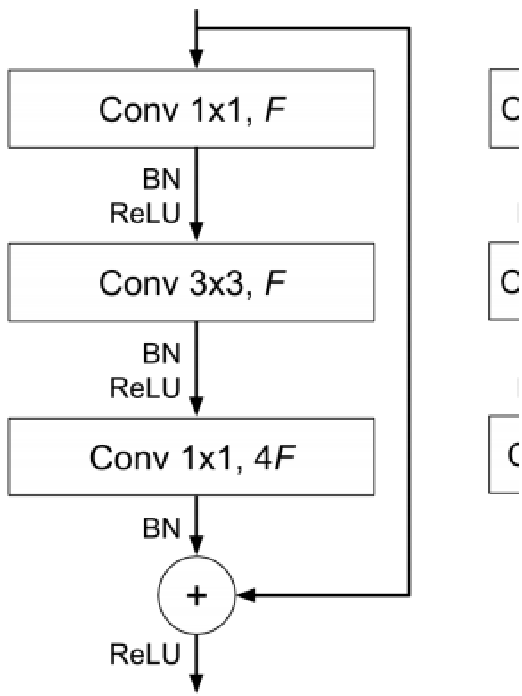
The profound learning models for protest location are able of recognizing the objects but the exactness of location for each demonstrate shifts. In [1] the creators have practically performed protest location utilizing SSD demonstrate, which was exceptionally quick in creating the results but the exactness was exceptionally less. While in case of Faster-RCNN the accuracy of the recognized protest was tall as compared to SSD but the time required to produce the comes about was moreover more as compared to SSD. In [2] the analysts have utilized faster-RCNN for highlight extraction and object detection on the dataset arranged by them for docking space creates at space station. They concluded that in spite of the fact that a few unmistakable highlights that has moo contrast from the environment or any other locale of the station were never recognized. Hence selecting fitting show is of critical significance as per prerequisite and assets available. In [9] the creators have backed the expressivity of Profound Neural Network (DNN) for question locator.But the comes about came at a few computational fetched at training time i.e., one must prepare a organize per protest sort and cover type. Object Location for Independent Vehicle Utilizing TensorFlow. Thus it was basically exceptionally basic to choose a demonstrate which has great exactness as well as quick result generation in arrange to meet the necessity of question location in self driving vehicles. TensorFlow overcomes the issue of both exactness and speed in a live video stream which is comparable to issue of question detection for a driverless vehicle. Moreover, protest location API in TensorFlow at the side the MobileNet neural organize permits us to make dataset with moo computation fetched which can be exported easily for the assignment of question detection.[8]

3. Proposed Method

Picture classification can accomplish different upsets but the major issue is that the model can recognize as it were one lesson per picture. Though in case of question location demonstrate more than one lesson per picture can be classified beside the position of the question in the image with a bounding box around the protest. We started by collecting our dataset of the test seat that we have made for the navigation of the demo test car. The test car was mounted with a camera setup for capturing the pictures of objects on the test seat. The rest handle is taken after in the steps as follows:

3.1. ResNet

Deep neural networks are hard to train because of valishing gradient problem. As we want to update the weights - we need to use backpropagation.



(a) ResNet

Figure 1:

In deep neural networks we face the problem of vanishing gradients. Backpropagation uses chain rule - the repeted multiplication of derivatives - makes the weights extremly small while reaching to the earlier layers. Residual networks are also deep networks. But they provide a mean to counter this vanishing gradient problem. Resnet introduced the concept of skip connection that resolves this problem.

Output Size	ResNet-50	ResNet-101	ResNet-152	ResNeXt-101-32x4d	DarkNet-53	MobileNet	MobileNetV2
112×112	7×7, 64, stride 2				3×3, 32, stride 1 3×3, 64, stride 2	3×3, 32, stride 2 $\begin{bmatrix} 3\times3, dw,s1 \\ 1\times1, 64 \end{bmatrix} \times 1$	$\begin{bmatrix} 1\times1, 32 \\ 3\times3, dw,s1 \\ 1\times1, 16 \end{bmatrix} \times 1$
56×56	3×3 max pool, stride 2 $\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128, C{=}32 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 32 \\ 3\times3, 64 \end{bmatrix} \times 1$ 3×3, 128, stride 2	$\begin{bmatrix} 3\times3, dw,s2 \\ 1\times1, 128 \end{bmatrix} \times 1$ $\begin{bmatrix} 3\times3, dw,s1 \\ 1\times1, 128 \end{bmatrix} \times 1$	$\begin{bmatrix} 1\times1, 96 \\ 3\times3, dw,s2 \\ 1\times1, 24 \end{bmatrix} \times 1$ $\begin{bmatrix} 1\times1, 144 \\ 3\times3, dw,s1 \\ 1\times1, 24 \end{bmatrix} \times 1$
28×28	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 8$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256, C{=}32 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 128 \end{bmatrix} \times 2$ 3×3, 256, stride 2	$\begin{bmatrix} 3\times3, dw,s2 \\ 1\times1, 256 \end{bmatrix} \times 1$ $\begin{bmatrix} 3\times3, dw,s1 \\ 1\times1, 256 \end{bmatrix} \times 1$	$\begin{bmatrix} 1\times1, 144 \\ 3\times3, dw,s2 \\ 1\times1, 32 \end{bmatrix} \times 1$ $\begin{bmatrix} 1\times1, 192 \\ 3\times3, dw,s1 \\ 1\times1, 32 \end{bmatrix} \times 2$
14×14	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 36$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512, C{=}32 \\ 3\times3, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 256 \end{bmatrix} \times 8$ 3×3, 512, stride 2 $\begin{bmatrix} 1\times1, 256 \\ 3\times3, 512 \end{bmatrix} \times 8$	$\begin{bmatrix} 3\times3, dw,s2 \\ 1\times1, 512 \end{bmatrix} \times 1$ $\begin{bmatrix} 3\times3, dw,s1 \\ 1\times1, 512 \end{bmatrix} \times 5$	$\begin{bmatrix} 1\times1, 192 \\ 3\times3, dw,s2 \\ 1\times1, 64 \end{bmatrix} \times 1$ $\begin{bmatrix} 1\times1, 384 \\ 3\times3, dw,s1 \\ 1\times1, 64 \end{bmatrix} \times 3$ $\begin{bmatrix} 1\times1, 384 \\ 3\times3, dw,s1 \\ 1\times1, 96 \end{bmatrix} \times 1$ $\begin{bmatrix} 1\times1, 576 \\ 3\times3, dw,s1 \\ 1\times1, 96 \end{bmatrix} \times 2$
7×7	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 1024 \\ 3\times3, 1024, C{=}32 \\ 1\times1, 2048 \end{bmatrix} \times 3$	3×3, 1024, stride 2 $\begin{bmatrix} 1\times1, 512 \\ 3\times3, 1024 \end{bmatrix} \times 4$	$\begin{bmatrix} 3\times3, dw,s2 \\ 1\times1, 1024 \end{bmatrix} \times 1$ $\begin{bmatrix} 3\times3, dw,s1 \\ 1\times1, 1024 \end{bmatrix} \times 1$	$\begin{bmatrix} 1\times1, 576 \\ 3\times3, dw,s2 \\ 1\times1, 160 \end{bmatrix} \times 1$ $\begin{bmatrix} 1\times1, 960 \\ 3\times3, dw,s1 \\ 1\times1, 160 \end{bmatrix} \times 2$ $\begin{bmatrix} 1\times1, 960 \\ 3\times3, dw,s1 \\ 1\times1, 320 \end{bmatrix} \times 1$
FLOPS (10^9)	3.8	7.85	11.6	8.03	7.14	0.57	0.31
Parameters (10^6)	25.6	44.6	60.2	44.2	40.58	4.3	3.5

Figure 2:

Deeper neural networks are more difficult to train. Residual learning was introduced to ease the training of networks that are substantially deeper than those used previously. ResNet explicitly reformulates the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions. On the ImageNet dataset this method was evaluated with residual nets with a depth of up to 152 layers—8x deeper than VGG nets but still having lower complexity. An ensemble of these residual nets achieves 3.57 persent error on the ImageNet test set. This result won the 1st place on the ILSVRC 2015 classification task. This technique can also be applied to the CIFAR-10 with 100 and 1000 layers.

4. Results

First we analysis the folder in our dataset. there will be 4 folder. we train those folder and show a chart. this training is our training example. also we test the dataset.

Figure 3:

then display 4 of the image class.then train them and found the images belonging to 4 classes and there 88 images.After that display each of items for a batch.then idenitity the blocks, covert the blocks then ResNet50 impliment. we found the leasning rate.
the learning rate is 0.001
then validated the batch and data then create the accuracy graph.

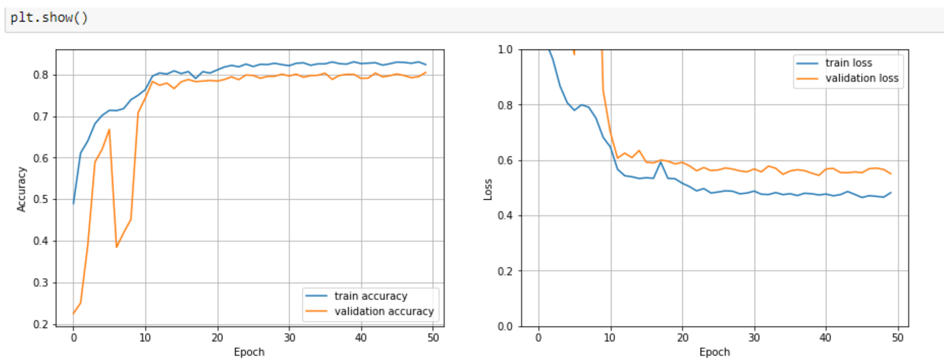


Figure 4:

5. Discussion

The proposed vehicle discovery calculation can be connected as it were in side see of vehicle pictures as wheels are uncovered as it were within the side see of the vehicle pictures. In spite of the fact that there are many other circular objects within the environment, this strategy can effectively evacuate those undesirable objects. The victory rate is very amazing.

6. Conclusion

In this work we utilized the assignment of protest discovery for self-driving vehicle by using TensorFlow API taken after by MobileNet neural arrange. The productivity in discovery for objects is approximately 85.18 persent, which is over normal, but the rate of result generation is very quick. The misfortune per step or age is 2.73 (beneath 3) that oversees the unwavering quality of the show. As for presently we have tried the show on the dataset arranged from the testbench. The demonstrate works fine in recognizing question in an picture but for multiple objects in an picture the bounding box shifts from one protest to another inconsistently. Such irregularity can be overcome by expanding the computation fetched as well as dataset. We are arranging to expand the show on the genuine electric vehicle for performing question acknowledgment and

7. Reference

1. Shetty, J., Jogi, P.S., Pandian, D., et al. (eds.): Study on Different Region Based Object Detection Models Applied to Live Video Stream and Images Using Deep Learning. Proceedings of the International Conference on ISMAC in Computational Vision and BioEngineering 2018 (ISMAC-CVB). Lecture Notes in Computational Vision (2018). https:// doi.org/10.1007/978-3-030-00665-56

2. Fomin, I., Gromoshinskii, D., Bakhshiev, A., Kryzhanovsky, B., et al. (eds.): Advances in Neural Computation, Machine Learning, and Cognitive Research. SCI, vol. 736. https://doi. org/10.1007/978-3-319-66604-412

3. Lin, C., Li, L., Luo, W., Kelvin, C.P., Wang, J.G.: Transfer learning based traffic sign recognition using inception-v3 model

4. Saha, S., Tairin, S., Khaled, M.A.B., Saha, S., et al.: An efficient traffic sign recognition approach using a novel deep neural network selection architecture. In: Proceedings of IEMIS 2018, vol. 3. https://doi.org/10.1007/978-981-13-1501-574

5. Talukdar, J., Gupta, S., Rajpura, P.S., Hegde, R.S.: Transfer learning for object detection using state-of-the-art deep neural networks. In: 5th International Conference on Signal Processing and Integrated Networks (SPIN). 978-1-5386-3045-7/18

6. Abadi, M., et al.: TensorFlow: large-scale machine learning on heterogeneous distributed systems. arXiv:1603.04467, pp. 1–19 (2016)
7. Sapp, A.S., Ng, A.Y.: A fast data collection and augmentation procedure for object recognition. In: Proceedings of the AAAI, Chicago, IL, USA, pp. 1402–1408 (2008)

8. Hoffman, C., Thiagarajan, D.: Continuity report: Revisiting grocery recognition using tensorflow, to be published