Journal of Network Communications and Emerging Technologies (JNCET) www.jncet.org
Volume 8, Issue 4, April (2018)

Object Detection on Roads using Deep Learning and Neural Networks

Nakul K. Warrier
Undergraduate, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India-600089

K. Sathish
Assistant Professor, Department of Computer Science, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India-600089

Abstract – Accidents are very common in this era of countless distractions, so it is imminent that we find a way to reduce and if possible completely overcome this problem. On a moving vehicle, if a camera of any sort is placed, it can capture and transmit the videos simultaneously to a machine that can work on the raw data. Using supervised learning mechanisms we can detect objects frame by frame of a video data and so provide the machine information which can reduce or completely reduce damage to the moving vehicle. The goal of this project is to recognise any object that come within the frame of a camera mount on a moving vehicle to help a system realize if any object can cause harm to the vehicle or to the object.

Index Terms – Machine learning, Tensorflow, OpenCV, Supervised learning, Deep learning, Convolutional neural networks, R-CNN.

1. INTRODUCTION

Machine Learning gives machines the power to understand or give a sense of understanding as to how they can accomplish a task that a human might find hard or even if not hard, time consuming to finish. It is said that a computer can do upto 33,860 trillion calculations per second, which is nothing compared to what a human can achieve in the same amount of time. Supercomputers were a requirement to do some machine learning in the past but a few years earlier, it has come to be possible in the most affordable CPUs with high accuracy and time efficiency.This change allowed Machine Learning to end up open more broadly.

Deep learning is a subset of machine learning. In machine learning algorithms are used to tokenize or break down raw data, understand them, learn them and make some decision based on them. The algorithms in deep learning are layered in such a way that they create an "Artificial neural network". These structured alogorithms are capable of learning and making come unsupervised decisions on its own. These decisions are also called intelligent decisions.

TensorFlow has APIs available in several languages both for constructing and executing a TensorFlow graph. The Python API is at present the most complete and the easiest to use, but other language APIs may be easier to integrate into projects and may offer some performance advantages in graph execution. In general, the tensorflow API's that use deep learning and neural networking will be adapted into the project using python scripts. The python scripts will also use open library OpenCV which will help in rendering the video and displaying in the frame speed specified by the user.

2. RELATED WORK

In order to detect objects using deep learning and neural networks, we need to understand the fundamentals of each component that are going to be used. To give the reader a wider scope of this domain, we have a referred several papers which address the same problem.

There are three ways for detecting motion in image sequences:

- Temporal difference
- Optical flow
- Background subtraction.

2.1. Temporal difference

The Frame difference is arguably the simplest form of background subtraction. The present casing is basically subtracted from the past casing, and if the distinction in pixel esteems for a given pixel is more noteworthy than a threshold (Th), the pixel is considered piece of the frontal area . For an assortment of dynamic conditions, it has solid versatility, however it is for the most part hard to get a total blueprint of moving item, at risk to show up the unfilled marvel, therefore the recognition of moving object is not exact.

2.2. Optical flow

Optical flow method is to calculate the image optical flow field, and do grouping preparing as indicated by the optical stream distribution qualities of picture . This strategy can get the complete development data and recognize the moving article from the background better, be that as it may, an expansive amount of computation, sensitivity to commotion, poor hostile to clamor execution, make it not reasonable for ongoing requesting events.

2.3. Background subtraction

Background subtraction is to utilize the distinction technique for the present picture and background picture to identify moving items. After foundation picture B(X, Y) is acquired, subtract the foundation picture B(X, Y) from the present casing FK(X, Y). In the event that the pixel distinction is more noteworthy than the set limit T, at that point confirms that the pixels show up in the moving item, generally, as the background pixels.

3. PROPOSED MODELLING

The main objective of the proposed system is to make the system fast and accurate. This system is based on python language and uses tensorflow framework as a key to object detection. The object detection API that is provided by google's tensorflow is very accurate comparing to self made machine learning models. The object detection is backed up by RCNN (Regional convolutional neural network) which help in framing the object that are detected by the API. GoPro's are very small action camera which has the ability to capture videos of upto 4k resolution and is very handy due to its very small size and weight. These action cameras help capture videos for implementation. Moreover GoPro's have a 172 degree wide capture ability.

4. IMPLEMENTATION

The implementation basically consists of these following modules:

4.1 Video Segmentation

Fig 1.0 A frame from test video

In this module, the input video that we get from a camera that is mount on a moving vehicle has to be converted to a form that has the ability to detect object from the tensorflow API. OpenCV is an open library that helps in rendering the video. This is only applicable when the video is available beforehand. Otherwise they is a way to capture the video of a continuous live feed from a webcam or any other camera that is supported in OpenCV library.

The video is inputted to the videocapture function of the cv2 module. If a webcam is used the integer number 0 is given as the input. This will convert the input video into an object of cv2 module that is split frame by frame. This object can be used in any method to get frame by frame images of the video.

Fig 2.0 A frame from second test video

4.2 Getting Tensorflow Ready

This is the module in which tensorflow models in brought into memory and API's are labelled. The tensorflow that is installed on the system has to be imported and allowed to download a model that gives accurate answers to the object detection of the frames.

- The model that is to be downloaded must be predefined along with it's version. It is usually a compressed file and found at the official website of Tensorflow.
- Provide a path to the frozen detection graph. This is actually the model that is used for object detection. Then the list of strings that is used to add correct label for each box in the objects that are detected are brought in to the program.
- Download the model with it's verified version number. This is achieved using urllib package which allows download from various websites. The file is then extracted and stored in a separate variable.
- Load a frozen Tensorflow model into memory using graphs. Label maps have the ability to map indices to category names, so that when our convolution networks number, we know that it corresponds to some object. Anything that is capable of returning a dictionary mapping integers to appropriate string labels can be used. In this project an internal utility function is used.

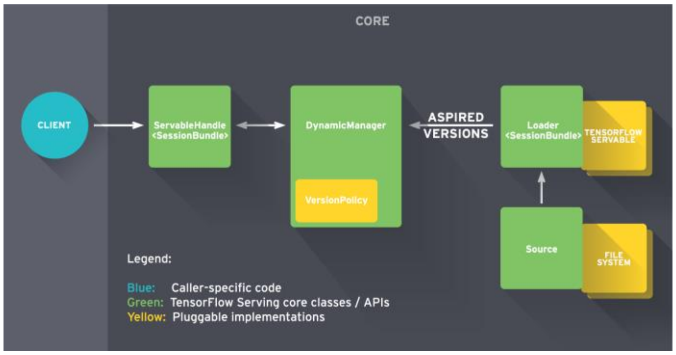

Fig 3.0 Basic architecture of Tensorflow API

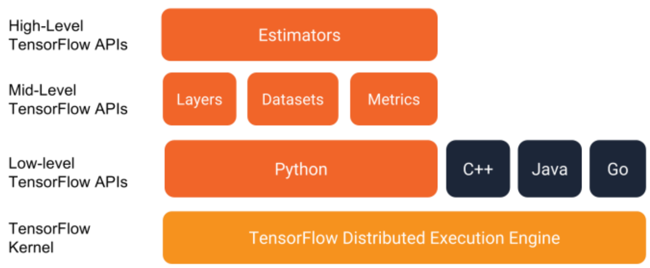

Fig 4.0 Programming stack consisting of multiple API layers

4.3 Object detection

The first and second modules are used here to detect objects frame by frame and display the result in a new window open by cv2 module. It combines the use of the cv2 object that contains the video separated frame by frame and the tensorflow models that have been loaded to the memory to detect and label object.

- A loop is started until the object loses scope. This object is now read to a variable so that it contains the first frame of the input video.
- For each frame in the object:
  - The image is expanded since the models expects images to have shape.
  - Each box represents a part of the image where a particular object was detected.
  - Each score represent how level of confidence for each of the objects.
  - Score is shown on the result image, together with the class label.
  - The session is run with all the above as parameters. This uses the models in the memory to do actual object detection.
  - The visual of the image with the labels and boxes of the detected image is display in a window open by cv2 package.

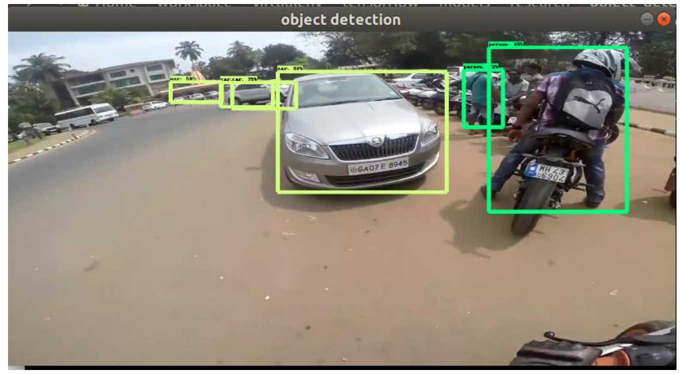

Fig 5.0 Objects detected in frame1

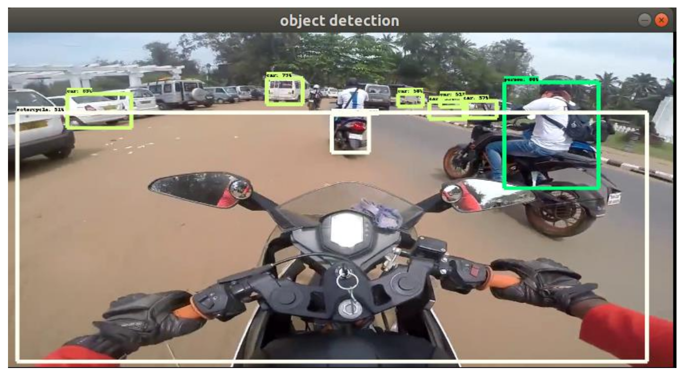

Fig 6.0 Objects detected in frame2

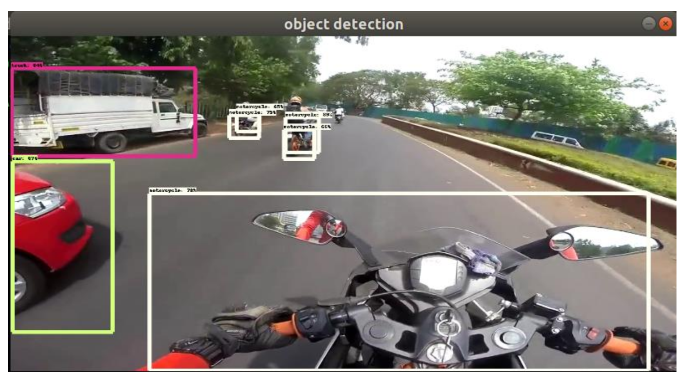

Fig 7.0 Objects detected in frame 3

5. RESULTS AND DISCUSSIONS

After running the learning model on the input video frame by frame, it was observed that all the objects in each frame was not detected but all the important object were detected. All of the detected objects were labeled and boxed accurately. The accuracy of the API is very high in detecting objects even from a very huge distance.

6. CONCLUSION

This study has helped to gain greater insight into the understanding of the major deep learning and neural networks models. It also helped to realize how these concepts tend to work out in the real world. The conclusion is that object detection is key to most of the technologies to follow and that google's Tensorflow API has been trained to accurately all the objects that were detected.

REFERENCES

[1] Zhang Pengyu, Research on road cracks' recognition based on image processing. China: Changsha University of Science & Technology, 2010.

[2] Chuo Eryong, "Development summary of international pavement surface distress automatic survey system", in Detection and Supervision for metrology, Vol.9, pp. 96-99, 2009.

[3] *Li Guoyan, Pavement distress recognition based on image. China: Hebei University of Technology,2008.*

[4] Xue Ying, The detection algorithm research of asphalt pavement rutting. China: Nanjing University of Science and Technology, 2011.

[5] *Zhou Lin, Research and Design of Pavement Crack Detection System Based on Image Processing. China: Taiyuan University of Technology, 2013.*

[6] Cao Maoyong, Digital Image Processing, 1st edn. Beijing: Peking University press, 2007.

[7] Gao Haojun, and Du Yuren, "The application of median filtering on image processing", in Electronic Engineer, Vol.30, pp. 35-36, August 2004.

[8] Chen Xiuxin, Zheng Ya, Yu Zhongzhong, and Jia Kebin, "Image Retrieval Based on Steerable Pyramid Binary Image Projection", in Computer Simulation, Vol.31, pp. 438-442, January 2014.