View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by NOPR

Journal of Scientific & Industrial Research
Vol. 79, June 2020, pp. 509-512

A Deep Learning Approach to Helmet Detection for Road Safety

Tanupriya Choudhury[1]*, Archit Aggarwal[2] and Ravi Tomar[1]

[1]Dept of Informatics, School of Computer Science, University of Petroleum and Energy Studies (UPES), Dehradun, Uttarakhand, India
[2]Amity School of Engineering and Technology, Amity University, Uttar Pradesh, Noida, India

Received 23 July 2019; revised 22 February 2020; accepted 10 May 2020

The rapid growth in the commute and vehicles has made exponential growth in the progress of mankind. This growth besides its positive aspects comes with a concern of saving life on road due to accidents. And, hence the technological advancements in the field of machine learning are required to cope up with the challenges such as road safety and traffic rule violations. According to the survey the majority of the life lost in road accidents is due to the negligence of wearing a helmet on a two wheeler vehicle. The enforcement of the traffic rules regarding this violation proves to be a challenge due to dense population and low rate of detection which is primarily caused by the lack of an automated system to detect the violation and take the necessary action. The growing population and the growing number of vehicles cause the manual systems in place to fail in curbing the issue. The recent advancements in Deep Learning and Image Processing provide an opportunity to solve this problem. This manuscript presents the implementation of a system which detects three objects namely the vehicle, non-usage of a helmet and the number plate of the vehicle under consideration using Tensorflow. Deep learning using the SSD MobileNet V2 is the primary technique used to implement the system. The system has been tested under different use cases with successful results.

Keywords: Object detection, Traffic violations, Motor Vehicles, MobileNet, Tensorflow

Introduction

The road safety is the real time and life saving problem, the use of computer vision and machine learning can lead to find optimal ways to solve this challenging and life threatening problem. In Computer vision the Object Detection is defined as the process of detecting an object in the multimedia such as image or video, using computational methods such as deep learning. There are many algorithms available such as RCNN[1], HOG[2], YoLO[3], SSD MobileNet[4,5], etc. The proposed work uses the SSD MobileNet[4,5] V2 algorithm due to its characteristics like support of transfer learning to detect generic objects. This paper proposes an implementation of SSD MobileNet[4,5] version 2 algorithm with transfer learning via Tensor flow[6] to train a system using a custom dataset made of images acquired from various sources in order to perform detection of three objects namely two-wheeler motor vehicles, the number plate on the two-wheeler and whether the driver of the vehicle is wearing a helmet or not.[7–9]

Materials and Methods

The proposed work utilized a two-stage approach provided by google, i.e., SSD MobilNet Version 2. The image dataset is divided into two parts for training and testing respectively. Fig. 1 and Fig. 2 provide an overview of the entire system, where Fig. 1 illustrates the architecture of the SSD MobileNet V2, and Fig. 2 presents the entire process from training to evaluation. The Model Selection block in Fig. 2 refers to the SSD MobileNet V2 model as presented in Fig. 1.

Proposed Work

The present work is achieved through a classified four phases, i.e. data collection, annotation, training, and evaluation. Table 1 below presents the flow of each phase with its sub-tasks.

The data collection phase is the most crucial and highly dependent phase of any machine learning objective. The data needs to be a heterogeneous mixture to get proper learning of the model, for the mix of both worlds, the approach uses a dataset collected by placing the camera on road to capture traffic whereas, the bulk of data is collected from the open-source video available on social media platform such as youtube.[10] The next phase is to annotate the images to tell the learning algorithm about the area of interest to identify in the image. The importance of annotation can be understood as if the annotations are not performed well; the results will show large

*Author for Correspondence
E-mail: tanupriya1986@gmail.com

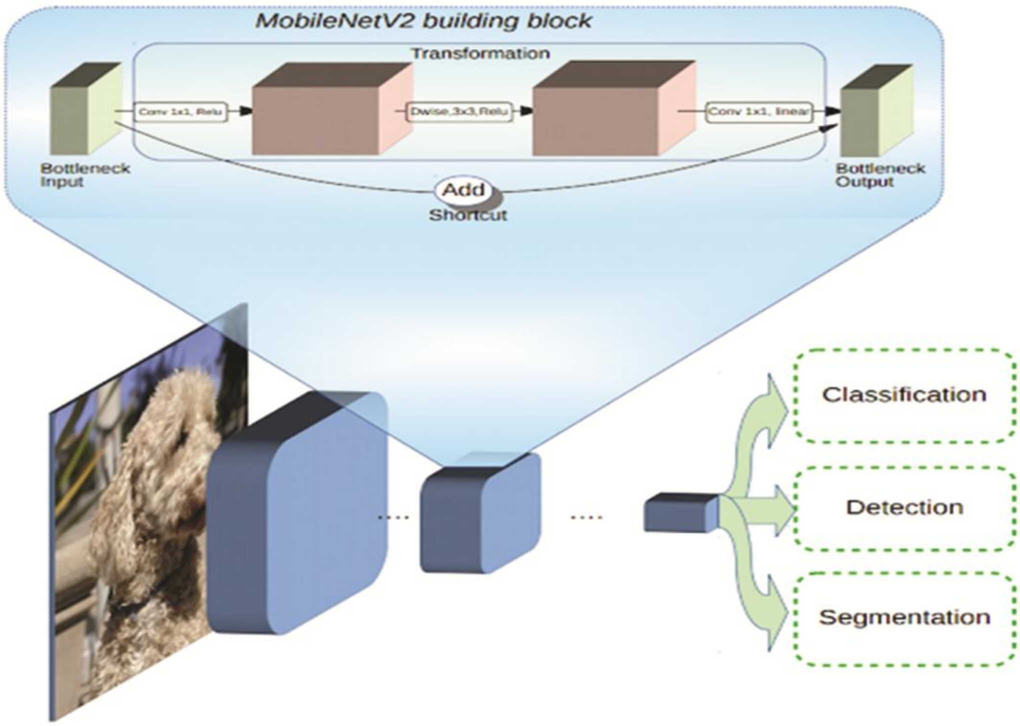

Fig. 1 — SSD MobileNet V2 Architecture

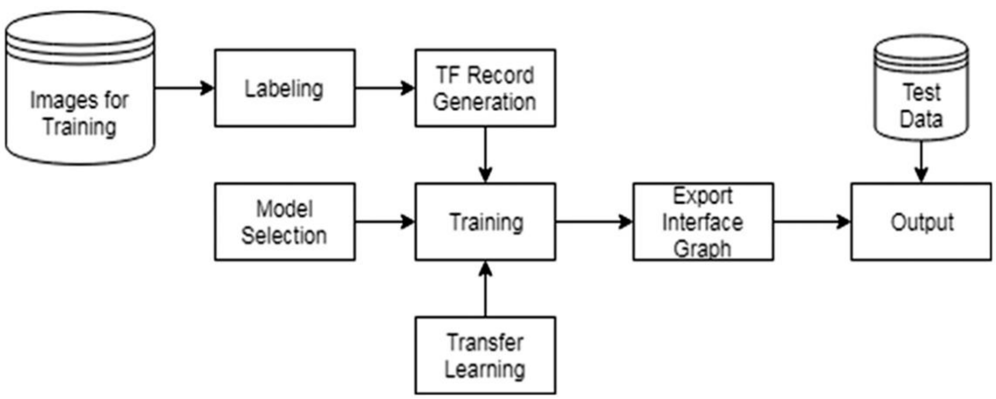

Fig. 2 — The Proposed System Representation

ambiguity that leads to the ultimate failure of the system. The process is achieved using LabelImg an open-source python tool. Once the images are annotated, the data is divided into training and test data. The next phase uses the training data obtained from phase 2. The models such as Faster RCNN were not selected besides providing more accuracy due to their hunger for time and computational requirements. The SSD MobileNet V2 has been chosen as the model for this work due to its characteristics of providing a good trade off between accuracy, time, and speed. And, finally, in phase four the test data is used to validate the model and is evaluated for deployment. The confidence score achieved is satisfactory. To achieve quick on boarding and development, all the phases are implemented using Google Colaboratory.[11]The primary reason for using Colaboratory was the ability to use the Tensor Processing Unit available on the cloud to train the system. Training is a very system intensive process and requires a very powerful GPU or TPU to process the tensors.[12,13]

Results and Discussions

The input images and videos are fed to the system using Open CV. The video are processed by fragmenting into individual frames and the detection

Table 1 — A four phase approach to the proposed work

	OVERALL PROCEDURE:
1.	**Dataset Collection**
2.	*Define all possible camera views and variations in possible data.*
3.	*Gather Images from all available resources.*
4.	*Split the image base into two divisions namely training and testing.*
5.	*Add blank images to the dataset to provide a better knowledge base.*
6.	**Annotation**
7.	*Label all images to generate XML file with marking coordinates for all detection objects namely H_NoHelmet, H_Bike, H_Numplate.*
8.	*Convert XML files to CSV format*
9.	*Generate the TF Record files for training the system*
10.	**Training**
11.	*Select the model SSD_Mobilnet_v2*
12.	*Define Configuration files.*
13.	*Perform transfer learning using the COCO dataset.*
14.	*Define the number of steps.*
15.	*Define the number of evaluation stages.*
16.	*Train the model*
17.	**Evaluation**
18.	*Export the model to an inference graph.*
19.	*Evaluate the test data.*
20.	*Calculate Accuracy and Confidence Scores.*

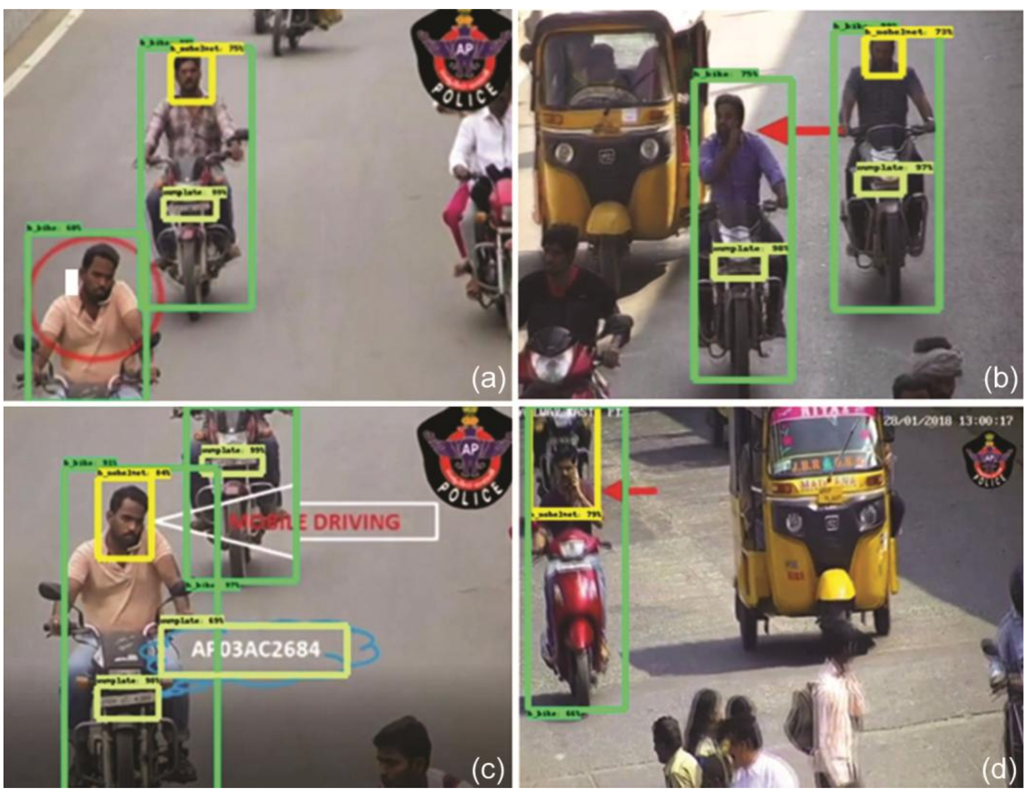

Fig. 3 — Various System Outputs in changing scenarios

is performed on the frames. The large diversity of the target demographic in terms of both animate and inanimate objects remains a challenge as the system is expected to handle this diversity fluently and thus it needs to be trained appropriately to do so. The outputs of the system are shown in Fig. 3, the image shows the tagged objects and it can be observed that the system is able to detect all the desired objects while handling changes in conditions like color, lighting, and the number of objects.

Table 2 — Maximum, Average and Minimum Confidence scores seen in the video processed using the proposed system

Object Name	Maximum Confidence Score	Average Confidence Score	Minimum Confidence Score
H_BIKE	99%	87%	70%
H_NOHELMET	99%	81%	68%
H_NUMPLATE	99%	84%	71%

The model was trained until the checkpoint for the eight thousand iterations was reached. The final loss value at this checkpoint was 1.12. Table 2 represents the minimum, maximum, and average confidence score of the system under all the test conditions.

The variation in confidence is noted due to the change in conditions of testing. This change in the confidence factor under different conditions shows the adaptability of the system. The system performs well for all input cases. The confidence can also been seen in the output images shown in Fig. 3. The images presented here are taken from the video processed using the system.

Conclusion and Future Work

The proposed work concludes with developing the system capable of identifying the objects using the transfer learning from the custom dataset. Thus, the system shows that it is possible to reach the goal of completely automating the process of enforcement of traffic laws while enabling additional functionality like record keeping while completely eliminating manual inputs. The primary focus for future additions is to increase the accuracy of the system by training it to adapt to a wider variety of vehicles and individuals and also to add the processing of individual number plates to create a database of defaulters as gathered by the system. The overall system also presents the need for a larger monitoring and verification system to avoid false positive cases from being added to the database.

References

1 Boonsirisumpun N, Puarungroj W and Wairotchanaphuttha P, Automatic Detector for Bikers with no Helmet using Deep Learning, *Proc 22nd Int Comp Science & Eng Conf* (*IEEE)* 2018, 1–4.

2 Silva R., Aires K. and Veras R., Helmet Detection on Motorcyclists Using Image Descriptors and Classifiers, *Proc 27th SIBGRAPI Conf on Graphics, Patterns and Images* (Springer) 2014, 141–148.

3 https://www.dlology.com/blog/how-to-train-an-object-detection-model-easy-for-free/.

4 https://machinethink.net/blog/mobilenet-v2/.

5 https://ai.googleblog.com/2018/04/mobilenetv2-next-generation-of-on.html.

6 https://www.youtube.com/watch?v=v1dv0RWpGmM&t=19s.

7 https://github.com/tzutalin/labelImg.

8 Zhao Y, Xiang X, Zhang R, Guo L, and Wang Z, Longitudinal Control Strategy of Collision Avoidance Warning System for Intelligent Vehicle Considering Drivers and Environmental Factors, *2018 IEEE Intelligent Vehicles Symposium (IV)* (Changshu) 2018, 31–36.

9 Tomar R, Sastry H & Prateek M, Establishing Parameters for Comparative Analysis of V2V Communication in VANET, *J Sci Ind Res*, **79** (2020) 26–29.

10 Joshi T, Badoni P, Choudhury T & Aggarwal A, Modification of Weiler-Atherton Algorithm to Address Loose Polygons, *J Sci Ind Res*, **78** (2019) 771–774.

11 Dahiya K, Singh D and Mohan C, Automatic detection of bike-riders without helmet using surveillance videos in real-time, *Proc 2016 Int Joint Conf on Neural Networks*, 2016, 3046–3051.

12 https://github.com/EdjeElectronics/TensorFlow-Object-Detection-API-Tutorial-Train-Multiple-Objects-Windows-10.

13 Phadnis R, Mishra J and Bendale S, Objects Talk - Object Detection and Pattern Tracking Using TensorFlow, in *Second Int Conf on Inventive Comm and Comput Technol*, (2018) 1216–1219.