

Emergency Vehicle Detection by Autonomous Vehicle

Aparajit Garg, Anchal Kumar Gupta, Divyansh Shrivastava, Yash Didwania, Prayash Jyoti Bora
Department of Information Technology
SRM Institute of Science and
Technology Kattankulathur, India

Abstract—Loss of lives due to the ambulance stuck in traffic can be avoided by incorporating a special system in self-driving vehicles which will move aside the vehicles and let the ambulance pass by. A camera is fixed on the back of car taking live feed of the surroundings. From the feed frames are feed to the deep learning module detecting ambulances from the frames. The technology used in the detection of ambulance from frames of the live feed is the Convolutional Neural Network. The ambulance can just be commuting and need not be in an emergency state always, so no need to provide way to ambulance in these situations. To implement the recognition of emergency state or not, an audio detection module is used. The audio from the surrounding is taken through a microphone and an audio duration of 10 seconds is fed to the sound detection module. The sound detection module uses Support Vector Machine, normally known as SVM and gives out the output of whether there is an ambulance in the surrounding or not. Based on the results of both the modules i.e., Image detection and Audio detection, an ambulance is detected in the surroundings and proper action is taken.

Index Terms—Image Detection, Audio Detection, Convolutional Neural Network(ConvNet), Support Vector Machine(SVM)

I. INTRODUCTION

In this paper, with the introduction of self-driving [2,3] cars with ambulance detection system, we would try to save as many lives as possible. With the implementation of ambulance detection module through image as well as audio module, we would try to reduce the number of false positives and build a redundant model detecting the ambulance in the traffic in the surroundings of the car and then the autonomous car can take the mentioned decision in the program to provide proper way to the ambulance to pass by as its first priority. We will also include a module for obstacle avoidance which will prevent any accidents of the self-driving car preventing loss of property and human or any lives. For the image detection, the dataset has been collected from google images consisting of ambulances and rest other vehicles to make the model learn how to distinguish between ambulances and other than ambulance vehicles and for the audio detection, the google dataset has been used consisting of the sound of emergency vehicles.

II. RELATED WORK

Several attempts has been made in order to detect the ambulance on the road and provide a faster means of commute system to it without any delay with an attempt to save as many human lives as possible. The detection algorithm has been used to alter the lights of the traffic signals[6,8] to help the ambulance go first. But there is no system which prevents the ambulance from getting stuck into the traffic. The sudden change in the traffic light might be the reason of accidents, caused due to fast moving vehicles. The accuracy of the system to detect ambulance in the traffic is not very good and includes lot of false positives.

The existing system detects ambulance using a ripple algo- rithm[4] which does not give that good accuracy and hence there is need of a better system. The ambulance can be detected using images as well as audio and the work done till now is able to detect on the basis of a single mode rather than detecting the presence of ambulance in the surrounding using both i.e., image detection as well as siren detection. With both the modules used together for ambulance detection, false positive rate would decrease and the self-driving cars will move aside on their own. This will prevent the ambulance from becoming the victim of traffic jams and thus saving as many lives as possible.

III. DATA ACQUISITION

A. Dataset for Image Detection

For learning the features of the ambulance and differentiating it from other vehicles the dataset is divided into 3 sections, namely: 1.) Training Dataset : It contains 439 images of ambulance and 372 images of non ambulance from different angles and different views so that the model can learn the features of the ambulance and differentiate it from all other vehicles. 2.) Validation Dataset : It contains 20 percent of the training dataset which is used to validate the accuracy of the model in predicting ambulance. 3.) Test Dataset : It contains 53 images which are used to calculate the accuracy and the effectiveness of the model.

B. Dataset for Audio Detection

The dataset for audio detection or siren detection of ambulance is taken from Google known as Google Audio Dataset. It contains the audio files of different categories of YouTube

videos of duration of 10 seconds and are human labelled to categorize the dataset into proper categories. The dataset is divided into 3 categories: 1.) Balanced train : 22,176 segments are taken from distinct videos having same criteria. It provides minimum 59 examples for each class. 2.) Unbalanced train : 2,042,985 segments are taken from distinct videos, consisting the rest of dataset. 3.) Evaluation : 20,383 segments are taken from distinct videos. It provides minimum of 59 examples for every class being used. The dataset used for siren detection has 5730 videos with an overall duration of 15.8 hours.

The following is the distribution of the dataset: For the purpose of training the classifier we have taken 42 records which contain emergency vehicles siren audio and another 50 records which do not contain any siren audio. 50 records were chosen out of thousands of records which did not contain any type of siren audio so as to balance the dataset with the number of samples containing the siren audio and hence to avoid skewing of data towards one category.

C. Data Augmentation

Data augmentation is a powerful technique which is used in case of less availability of dataset to train upon. The deep learning models require a huge amount of dataset to be trained upon and provide better results with a higher accuracy overall. But in case of less available dataset, a technique called data augmentation is used. In keras, data augmentation is achieved by using the Image Data Generator function which takes in the required parameters which are the basis of processing the available dataset and increasing the size of the dataset. Following are the possible parameters it takes in order to increase the size of the dataset without repeating any image:

Rotation range : It ranges from 0-180 degrees. It is the value within which the images are rotated randomly.

Rescale : It rescales the image or the dataset before any further processing. It takes in the input and process the RGB coefficients of the image which ranges from 0-255 in the range of 0-1.

Width shift: It is a fraction of the total width of the image. It is the value through which the image is translated horizontally.

Height shift : It is a fraction of the total height of the image. It is the value through which the image is translated vertically.

Zoom range : It is the parameter which randomly zooms in the picture. horizontal flip : It is the technique of randomly horizontal flipping of half of the images. It is mostly useful where there is no case of horizontal symmetry i.e., in real world images.

Fill mode : It is the technique of filling in the places of the image which occurs after an operation is applied to it. The pixel values are then need to be filled based upon the nearest values or some other parameter. With these above parameters the dataset is increased multifold and with no repeating images. All the images in the dataset are unique and are used to train the model and prevent overfitting. These increase the chances of good predictions by the deep learning model providing an efficient and reliable model.

TABLE I
AUGMENTATION VALUES APPLIED IN TRAINING DATASET

Type of Augmentation	Values
Horizontal Flip	Yes
Rotation Range	45 degrees
Scaling factor	0.1

IV. IMPLEMENTATION

A. Obstacle Avoidance

The obstacle avoidance algorithm is implemented in raspberry pi. The algorithm operates an ultrasonic sensor HC-SR04[1], which is used to avoid the obstacles in its way and preventing any accidents through the vehicle. The ultrasonic sensor sends out a sound signal to check if there is any obstacle in front of the vehicle or in line of motion of the vehicle. The signal goes on in front until it is reverted back by any obstacle in its way. The signal is received back by the sensor and it has the time taken by the signal to revert back and the speed of the signal through which it travels. Using these 2 measurements with the help of the below formula the distance is calculated between the car and the obstacle.

$$\text{Distance} = \text{Speed} * \text{Time}$$

The travel time for waves is 343 m/s. So, For calculation of distance it will be divided by 2 as signal travels from sensor (HC-SR04) to object and back to sensor module HC-SR-04.

$$\text{TotalDistance} = \frac{343 * \text{TimeOfHigh(Echo)Pulse}}{2} \quad (1)$$

If the distance between the car and the obstacle is greater than the safe distance mentioned, nothing happens. But if the distance is less than or equal to the safe distance, then the motor is triggered and the car is made to avoid the obstacle by either stopping the vehicle or rotating in a direction where there is no obstacle. This provides an accident free commute system for all the vehicles.

B. Image Detection

Ambulance detection using images is used to predict if an ambulance is coming or approaching the car from behind. This is being applied in order to move the vehicle aside and give the ambulance a proper and delay free commute system. The detection of ambulance coming from behind is done using object detection in images in this module. A camera is mounted on the rear end at top of the vehicle to record in the live happenings in the surroundings of the car. The live feed is then used to provide input to the module which then works upon the input to detect if there is any ambulance in the surrounding or not. The frames are taken from the live feed one by one and fed into the deep learning convolutional neural network[5,7] created to process images and detect the ambulance in the images. The frame is resized into 32X32 dimensions which is the size on what the model is trained upon. The pixel values of the image are scaled in the range between 0-1 to provide ease to the model to do

the predictions instead of the range of 0-255 and then the image data is converted into the required array format that the model accepts. The array is then fed into the deep learning model and the model performs its computations in order to learn the features required to differentiate the ambulance from other vehicles. After the module performs its computation on the image fed into the network, the detection of ambulance in the image with the confidence level is given out as the result and if the ambulance is detected, then the car is either stopped or move aside to let the ambulance pass through, giving ambulance the priority of commute.

C. Audio Detection

The audio detection is done in order to detect the presence of any siren which would be accompanied by the emergency vehicle in case of any emergency. When the siren is detected, our car will know that there is an emergency and so the necessary actions are taken. In case there is no siren detected along with the detection of an emergency vehicle through image detection then our car will not perform any action and continue running on the same path. The presence of an emergency vehicle which is detected by camera along with the detection of siren detected through the microphone together will result in an emergency and the required actions will be triggered. In order to detect the presence of a siren we have created a classifier which is used by our car to classify the surrounding audio into two categories that is, either siren is present or there is no siren present. In order to train the classifier we have used Googles Audio Set. Googles Audio set consists of features extracted from 2,084,320 YouTube videos in the form of Tensor flow records. 10 seconds clip is extracted from each video and is labeled manually. There are 527 classes/labels present in total from which three labels are identified as emergency vehicle sirens. The Google Audio set also contains csv files containing YouTube Id for the video segment, start and end time of the video segment and the labels corresponding to the audio present in that video. The features are extracted using VGGish model. For every second of video there is 128 dimensional embedding extracted at 1Hz and so for every 10 second clip of audio from a video there are 128 x 10 dimensional embeddings.

For the purpose of training the classifier we have taken 42 records which contain emergency vehicles siren audio and another 50 records which do not contain any siren audio. Records containing siren audio are termed as positive samples while the records which do not contain siren audio are termed as negative samples. These records are then saved into a data frame. Each row of the data frame contains all the 1280 embeddings for a 10 second long video along with a label value. The positive records are labeled as 1 and the negative records are labeled as 0. This data frame of dimension 92 x 1281 is then needed to train a classifier. X axis contains all the 1280 embeddings from each row while the Y axis contains labels 0 or 1 depending on the row containing a positive or a negative record. The data is split into training and testing data with 80 percent being the training data and 20 percent being

the testing data. The classifier used is SVM. After training the classifier that is SVM, we got an accuracy of 95 percent. The confusion matrix as well as the classification report are available in the following pages. The model is then saved. This saved model that is the SVM classifier is then ported to our car for the classification of audio present in its surrounding. The audio in our cars surrounding is captured using a microphone which is connected to the Rasberry pi. The microphone detects audio from the surrounding and saves it as a wav file. This wav file is then provided as an input to the VGGish model. The VGGish model converts the wav file into spectrogram and finally extracts 128 dimensional embeddings for each second of audio. These extracted embeddings are then fed as the input to the model saved earlier (SVM classifier). The classifier gives the result as positive or negative so the car can act accordingly.

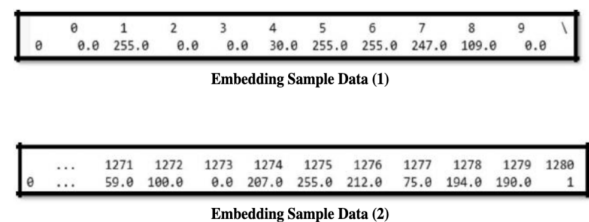


Fig. 1. Embedding sample.

If both the modules detect ambulance through image and audio then the car is made to move aside giving path to the emergency vehicle to pass through.

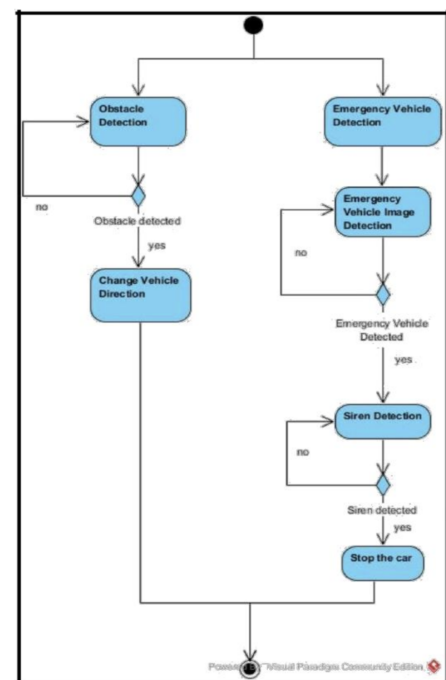


Fig. 2. Flow Diagram.

V. RESULTS

The graphs in Fig.3 and Fig.4 depicts the loss and accuracy of the model. In the accuracy graph both the training and the validation accuracy have less fluctuation which shows the model does not underfit and the variation shows that there is no overfitting either. The highest accuracy in the training set is around 83 percent and in validation set it is around 87 percent.

The loss graph has a decline flow showing no underfitting and there is no substantial overlapping which shows there is no overfitting either. The minimum training loss achieved is around 37 percent and the minimum validation loss achieved is around 34 percent.

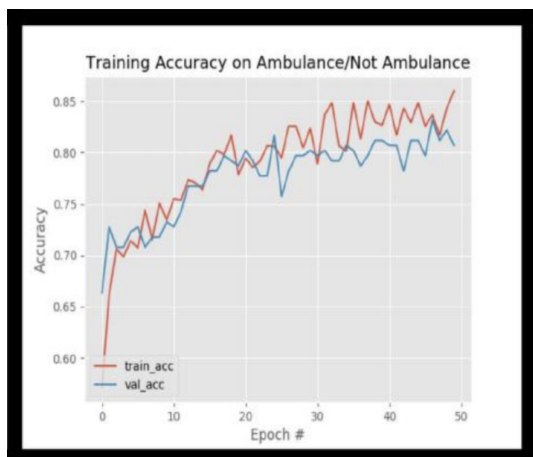


Fig. 3. Training and Validation Accuracy(Image Detection).

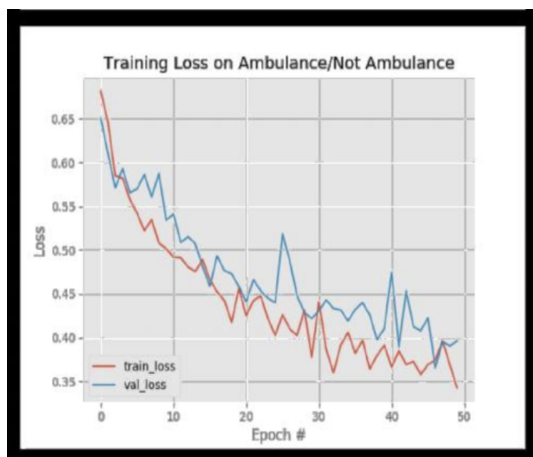


Fig. 4. Training and Validation Loss(Image Detection).

After training the classifier over the features data from Google's audio set, the accuracy of 95 percent was achieved. The confusion matrix gives one false positive and zero false negatives with nine true positives and true negatives.

VI. CONCLUSION

In this paper, an autonomous driving system is proposed which can be incorporated in the vehicles with an emergency

	precision	recall	f1-score	support
0	0.90	1.00	0.95	9
1	1.00	0.90	0.95	10
micro avg	0.95	0.95	0.95	19
macro avg	0.95	0.95	0.95	19
weighted avg	0.95	0.95	0.95	19

Fig. 5. Test Data Classification Report(Audio Detection).

[[9 0]
[1 9]]

Fig. 6. Confusion Matrix(Audio Detection).

detection system and an obstacle avoiding system. For the avoidance of the obstacle in the way of the vehicles, an ultrasonic sensor is used which uses the concept of ECHO, to detect any obstacle in front of the car in a specified range. The ultrasonic sensor sends out a signal to detect if there's any obstacle in front of the car or not. The signal is then received by the sensor after getting reflected back from an object. The distance is then calculated between the sensor and the obstacle and if the distance is less than or equal to the mentioned distance of obstacle avoidance then the vehicle is turned or stopped until the obstacle is removed.

The paper also proposes an ambulance detection system with the help of the images and sound. The ambulance detection through images is achieved with the help of camera mounted on top of the vehicle. The camera records the live feed of the surroundings and the frames from the recording is taken and fed into the proposed Deep Learning Convolutional Neural Network which gives out the percentage of the presence of ambulance in the frame. The use of sound in this paper is an advantage in the recognition of the ambulance in the surrounding. A microphone is placed at the rear end of the car which takes in the surrounding noise and then converts it into a 128-dimensional feature vector which is used to recognize if an ambulance siren is there in the surrounding or not.

The result of both the modules is then combined and the output is given whether there is an ambulance in the surrounding or not.

REFERENCES

- [1] Johann Borenstein Yoram Koren, Obstacle Avoidance with Ultrasonic Sensors, IEEE JOURNAL OF ROBOTICS AND AUTOMATION, VOL. 4, NO. 2, APRIL 1988, pp. 213-218
- [2] Design and Implementation of Autonomous Car using Raspberry Pi International Journal of Computer Applications (0975 8887) Volume 113 No. 9, March 2015
- [3] Stewart Watkiss, Design and build a Raspberry Pi robot [Online], available at: <http://www.penguintutor.com/electronics/robot/rubyrobot-detailedguide.pdf>
- [4] F.Andronicus, Maheswaran, "Intelligent Ambulance Detection System", International Journal of Science, Engineering and Technology Research (IJSETR), Volume 4, Issue 5, May 2015

- [5] Rohit Tiwari, Dushyant Kumar Singh, "Vehicle Control Using Raspberry and Image Processing", Innovative Systems Design and Engineering www.iiste.org ISSN 2222-1727 (Paper) ISSN 2222-2871 (Online) Vol.8, No.2, 2017
- [6] R. Cucchiara, M. Piccardi, P. Mello, (2000) Image analysis and rule-based reasoning for a traffic monitoring system, IEEE Trans.On Intell.Transport.Syst., vol1. No.2.
- [7] T. Schlegl, S. M. Wald stein, W.D.Vogl,(2015)Predicting Semantic descriptions from medical images with convolution neural network in Information Processing inMedical Imaging. Springer, pp. 437448.
- [8] P. Parodi, G. Piccioli, A feature-based recognition scheme for traffic-scenes, Proc. IEEE Intell. Vehicle, pp. 229-234, 1995.
- [9] L. Gao, C. Li, T. Fang, Z. Xiong, Vehicle detection based on color and edge information, Proc. Image Anal. Recog., vol. 5112, pp. 142-150, 2008.