Experimental Design and Implementation of Real Time Priority Management of Ambulances at Traffic Intersections using Visual Detection and Audio Tagging

Sayak Banerjee
School of Electronics Engineering
Vellore Institute of Technology,
Vellore, India

Ritayan Patra
School of Electronics Engineering,
Vellore Institute of Technology,
Vellore, India

Arin Paul
School of Electronics Engineering,
Vellore Institute of Technology,
Vellore, India

Debmalya Chatterjee
School of Electronics Engineering,
Vellore Institute of Technology,
Vellore, India

*Sumit Kumar Jindal
School of Electronics Engineering,
Vellore Institute of Technology,
Vellore, India
sumitjindal08@gmail.com

2023 IEEE 2nd International Conference on Data, Decision and Systems (ICDDS) | 979-8-3503-1416-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICDDS59137.2023.10434545

Abstract – **Transportation of patients to the hospitals in the least possible time is of utmost importance. In this work, we have come up with a unique solution to reduce the transport time of patients to hospitals by reducing the average waiting time at traffic intersections. The Wi-Fi enabled CCTV cameras at the traffic intersections will detect an incoming ambulance using the You Only Look Once (YOLO) v3 object detection algorithm, as well as tagging and classifying ambulance sirens from regular noise for detection of ambulances with high precision. As a prototype, the authors have built a hardware model which can be easily installed in traffic intersections to make the other drivers aware of the incoming ambulance through an LCD display and a buzzer. The successful detection of an ambulance will initiate a prompt response by the system which will reach the corresponding authorities as well as the other car drivers and it will enable creation of a smooth transport corridor for the ambulance.**

Keywords - Ambulance detection, audio tagging and classification, deep learning, object detection, traffic intersection, YOLO-v3

I. INTRODUCTION

Emergency vehicles like ambulances are the first pillar of the health infrastructure of any country. Ambulances play an essential role when it comes to life threatening situations. They quickly transport patients to hospitals requiring serious medical attention. This noble work carried out by the ambulance drivers and concerned authorities are greatly hampered by traffic congestions. Traffic congestion accounts for almost 20% deaths caused due to unavailability of proper medical help. Traffic congestion acts as the last nail in the coffin for a person fighting for his/her life. These traffic congestions are prevalent in each and every country and they impair the health infrastructure [1].

The proposed work tries to address this problem in an innovative way by using sophisticated deep learning and machine learning tools [2]. In most cases, the traffic at intersections is managed by policemen. To avoid human errors, the traffic management systems must be equipped with state-of-the-art technologies which can work in unison with policemen to efficiently identify the essential medical vehicles and provide a clear and fast passage for these vehicles by avoiding the traffic congestion [3][4]. In this work, the images from the CCTV installed along the roads are taken along with audio. The images are taken continuously each second and are analyzed to detect different classes of vehicles. The audio further simplifies the process of detecting ambulances and increases the overall accuracy. Whenever an ambulance is detected, it will immediately notify the policemen as well as other drivers near the crossroads.

The proposed method uses the YOLO-v3 algorithm for detecting the presence of ambulances. The YOLO-v3 algorithm is very fast and gives highly accurate values both in terms of IoU (Intersection over Union) and mAP (mean Average Precision) when compared with other object detection algorithms. The audio tagging and classification is done by generating the frequency Power Spectrum images of the incoming sound signals and feeding it to the Convolutional Neural Network (CNN) for classification. Finally, for alerting policemen and other people, the authors have provided a low-cost hardware implementation of the prototype using Raspberry Pi3 Model B+, Buzzer and LCD as a proof of concept.

979-8-3503-1416-8/23/$31.00 ©2023 IEEE

II. LITERATURE SURVEY

A few systems and notable research have been done to detect an emergency vehicle. The problem is approached by developing an automatic traffic control system which detects and clears the path of the emergency vehicles. For object detection, Yolo -V3 architecture is used. It checks the probability of the occurrence of the essential vehicle being in a particular image segment. In the next part, a deep learning technique called transfer-learning is employed, which uses the model previously trained with a huge dataset. For example, VGG_16 is used as CNN which is trained with the ImageNet dataset. It increases the accuracy of object detection and hence, the overall accuracy of the model [5].

Unawareness of drivers has been a big concern when it comes to traffic control which leads to accidents and delay management of emergency vehicles. The authors addressed this problem by developing a siren detection system which primarily detects the presence of an emergency vehicle's siren sound and thereby alerts the driver. The detection system uses audio tagging and classification, to differentiate sounds of normal traffic with and without the presence of siren sounds from emergency vehicles. The working of the model is based on two network streams for classification. In the first stream, raw data is directly proposed using WaveNet whereas in the second one MLNet is being used which works a 2D representation of audio formed using Mel spectrogram and Mel-Frequency Cepstral Coefficients (MFCC) [6].

A mobile automated model is developed to analyze the one-of-a-kind traits of automobiles in environments irrespective of their inter-connection. This prototype considers the rate outcomes of main automobiles, the impact of traffic signals and any other traffic rules and regulations prevalent in that particular road intersection in order to reproduce the same traffic flow. An assessment of various parameters like automobile velocity, traffic glide, and average tour time are estimated through this model for the above-mentioned scenarios to provide a better traffic flow and reduced waiting time [7].

The vehicles on the roads don't always know about the traffic conditions ahead. This often leads to unending traffic snarls. A new mechanism is proposed where each vehicle is connected in a single system. The system is constantly and automatically updated with real time traffic conditions in that road or of the upcoming roads' intersection. An alternate route can be provided by this centralized system to the vehicles depending upon the traffic condition, thus handling traffic automatically in small areas by providing dynamic routing options to the vehicles. [8].

The authors have developed a smart traffic system using embedded systems tools. Traffic signals at the intersections depend on detecting emergency vehicles along with recognizing the density of traffic in a particular lane. A real-time video stream is used for image acquisition and then the algorithm is processed. The amount of time for which the green signal will be on for a particular direction of the road depends on the density of the vehicles in that particular lane. Object detecting and counting algorithms are used to determine the density of traffic. When an emergency vehicle is detected in a lane the former gets higher priority than all other directions in the traffic intersection and traffic signals are managed accordingly [9].

The WIFI-enabled CCTV cameras installed along the roads are incorporated with some software to calculate the distance of essential vehicles like ambulances from the traffic signal and this information is delivered to concerned traffic personnels. The images obtained from the cameras are processed to their threshold level and the images are enhanced using morphological image processing techniques. Further, the distance between the CCTV camera and the essential vehicle is calculated using the Euclidean formula using MATLAB software. Along with this, other essential parameters like the speed of the vehicle and the traffic density in the road at a particular time are forwarded to the traffic policemen and the concerned personnel to effectively control the traffic flow at the intersections [10].

III. SYSTEM ARCHITECTURE

A. Flowchart

The entire system architecture is described by the flowchart given in **Fig. 1**.

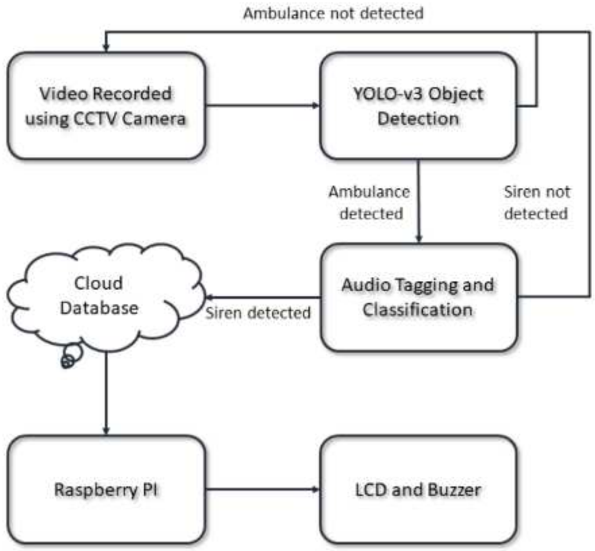


Fig. 1. Flowchart of the proposed system

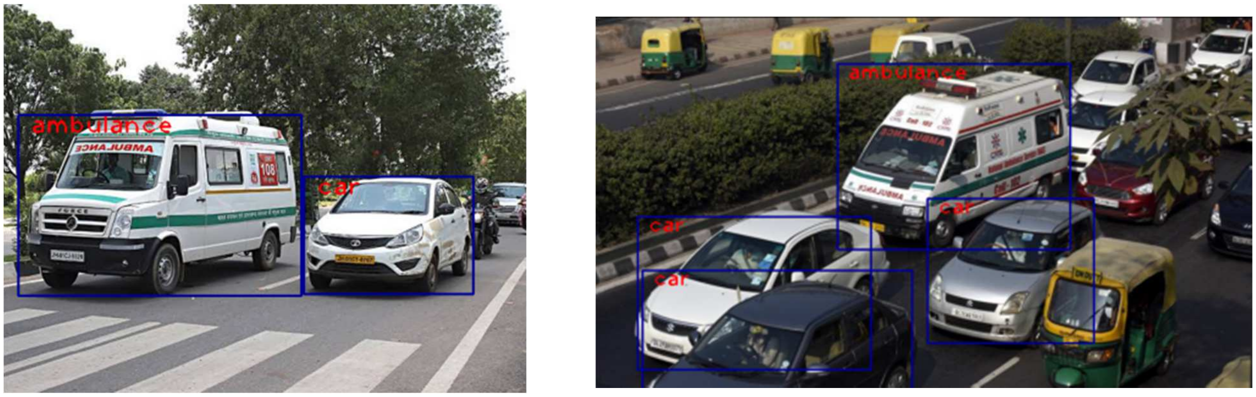


Fig. 2. Ambulance detected in traffic using YOLO-v3 algorithm

B. Audio Tagging Dataset

The dataset – *Emergency Vehicle Siren Sounds* – available as an open-source dataset in the Kaggle platform is used to train the deep learning model for audio tagging and classification. The dataset has 200 samples of ambulance siren and 201 samples of regular traffic noise. Each audio sample ranges between 3-5 seconds.

C. Visual Ambulance Detection

The visual detection of an ambulance from a road full of cars, requires computer vision solutions which use heavy deep learning algorithms for accurate detection and localization. Some of the algorithms used are Convolutional Neural Networks (CNN), Faster Region-based Convolutional Neural Networks (Faster R-CNN) and Single Shot Multi-box Detector (SSD). These algorithms successfully detect any type of vehicles, pedestrians, traffic signs and other objects on the road. However, these algorithms have certain limitations. For example, Faster R-CNN overcomes the high computational time requirement of R-CNN and Fast R-CNN limits itself in interference time. Similarly, SSD outperforms existing algorithms but compromises on accuracy.

YOLO was first introduced in 2015. It revolutionized object detection in all aspects of accuracy and interference time. The authors have used YOLO-v3 which is the third version of the algorithm [11]. This algorithm is very fast and gives highly accurate values both in terms of IoU (Intersection over Union) and mAP (mean Average Precision) when compared to other object detection algorithms. It fits perfectly for real time object detection with minimum lag, which in the proposed case is vehicle detection. It is implemented using a deep convolutional neural network called Darknet-53 which contains a 53-layer network, trained on ImageNet. 53 more layers are being used for detection making it a 106 layers deep architecture [11].

The proposed model is based on a traffic intersection having a camera placed with traffic lights for each direction. The camera records video from which image frames at a rate of 45 frames/sec are being processed by YOLO-v3 algorithm. The algorithm detects the objects present in the road like cars, buses, trucks, ambulances, persons, etc. For each detected object the algorithm draws an anchor box around it and shows the name of the class it belongs to. If the algorithm detects the presence of an ambulance on that particular road, it transmits a message regarding the presence of it. The presence of an ambulance activates the next part of the model which is audio tagging and classification, used to confirm that the ambulance is on emergency service.

The resultant images from the object detection module are visualized in **Fig. 2**.

D. Audio Tagging and Classification

Ambulance detection using only images may not be enough to efficiently detect them. There can be a scenario where an ambulance is not using the siren indicating that it is not transporting any patients to hospital but this can't be understood by the deep learning model unless the audio (siren) is added and processed by deep learning model. If the system allows each and every ambulance which are not essential or not transporting any type of patients, then it will create more traffic congestion and it will nullify the whole effort of adding the artificial brain to the traffic management system.

Audio Classification of various traffic sounds and ambulance sirens are discussed in this work. This audio classification increases the efficiency of the proposed prototype. The authors have used the Mel Scale to classify and separate the ambulance siren from rest of the traffic noises. Each and every sound has a definite frequency and power which creates a unique spectrogram. Using python and librosa library, the authors have created a spectrogram using Mel Scale to easily classify and segregate traffic noise from the ambulance sirens and this spectrogram is known as Mel Spectrogram [12]. Mel Spectrogram is significantly different from normal audio spectrogram. It uses Mel-Frequency in place of frequency on the y-axis and Decibel in place of Amplitude on the x-axis. **Fig. 3** represents a Mel-Frequency Spectrogram.

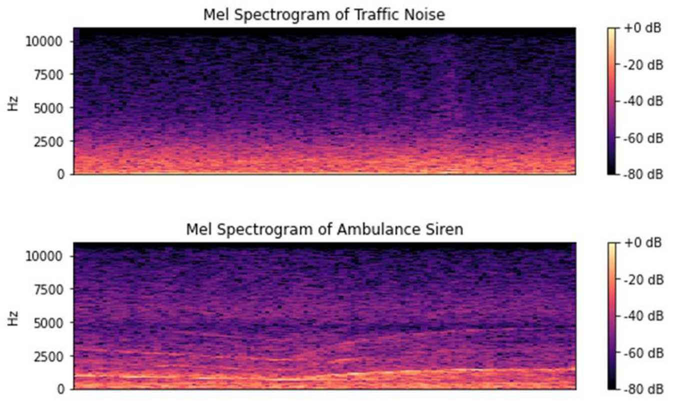


Fig. 3. Mel Spectrogram

MFCC (Mel-Frequency Cepstral Coefficients) represent the Mel Spectrogram. The MFCC can be used in the deep learning model to separate the ambulance siren from the rest of the traffic noise. **Eq. (1)** is the mathematical formulation to convert frequency to Mel-frequency scale.

$$m = 2595\log_{10}\left(1+\frac{f}{700}\right) \qquad (1)$$

where *f* represents the frequency and *m* represents mels.

To calculate MFCC, the first step includes the preprocessing of the incoming audio signal which includes splitting the audio signals in short frames and applying windows. In the second step, each frame is then subjected to NN-point Fast Fourier Transform (FFT) to calculate the frequency spectrum which is known as the Short-Time Fourier Transform (STFT). Then it is passed through Mel scaled Filter Banks. The output energy spectrum is then transformed to get a logarithmic energy spectrum. These energy coefficients are highly correlated. So, to prevent any chances of overfitting and to optimize for much better performance, Discrete Cosine Transform (DCT) is applied to decorrelate the coefficients. The Mel Spectrogram images generated after these operations by the librosa library are fed as input to the CNN model – which classifies between the different traffic noises and ambulance siren [13].

D.1 Deep Learning Architecture

The authors have proposed a deep learning neural network architecture which will classify and tag the corresponding ambulance siren from traffic noises. The deep learning model has 8 layers. The first layer of the architecture is a 2D convolution layer with 64 output kernels. The convolution matrix is used as a dimension of 3x3. The second 2D convolution layer has 128 output kernels. Both the convolution layers are activated using a ReLU activation layer and are followed by the 2D max pooling layers which have a stride size of 2x2. The convolution layers extract features from the input mel-frequency spectrogram images. A dropout layer with a rate of 0.1 is added to control regularization. The extracted features from the CNN layers are flattened into a one-dimensional input vector as input to the fully connected layer. A hidden layer is added with 1028 output neurons. The output layer has 2 neurons for classification into the 2 classes and is activated by the 'Softmax' activation function. The entire deep learning architecture is shown in **Fig. 4**.

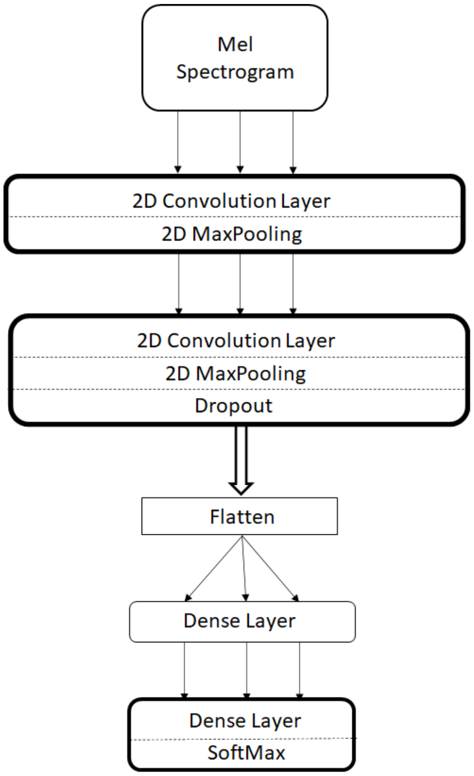


Fig. 4. Deep Learning Architecture

E. Cloud Infrastructure

In this work, the authors have used an opensource cloud infrastructure – Firebase – to store data and act as an intermediary between deep learning model and the Raspberry Pi. It can be integrated with the deep learning model as well as Raspberry Pi through python programming. The data obtained after classification using YOLO-v3 is transferred to Firebase which is then retrieved by the Raspberry Pi for further processing.

F. Hardware Implementation

The hardware equipment employed in this prototype are Raspberry Pi 3 Model B+, LCD and Buzzer. The Raspberry Pi 3 acts as a CPU. The LCD mimics the electric hoarding and the buzzer sends out alerting signals to the people and policemen. The authors have also used 4 pairs of LED lights consisting of red and green light which indicates the traffic light in the cross section of four roads.

IV. RESULTS

The YOLO-v3 object detection algorithm runs at 45fps with real time speed and has a mAP of 63.4%.

The deep learning model for audio tagging and classification is trained on 90% of the dataset. The remaining 10% of the audio samples are used evaluating the test accuracy. The model is trained for 50 epochs. The training accuracy of the model is 92.37% and the test accuracy on the validation set is 88.72%. The F_1 score is 0.89. The result is compiled and sent to the real-time online database.

The data is stored in the Firebase which is retrieved by the Raspberry Pi. In the beginning, there is no ambulance on any of the intersecting roads. So, there is no need to change the traffic signal in any of the paths as it is shown in **Fig. 5.** In **Fig. 5,** since there is no ambulance, Path 1 and Path 3 traffic signals are turned green and similarly, Path 2 and Path 4 traffic signals are turned red indicating normal traffic conditions at the cross section. The phrase "Ambulance Not Detected" is displayed through LCD and the Buzzer is off as shown in **Fig. 6.**

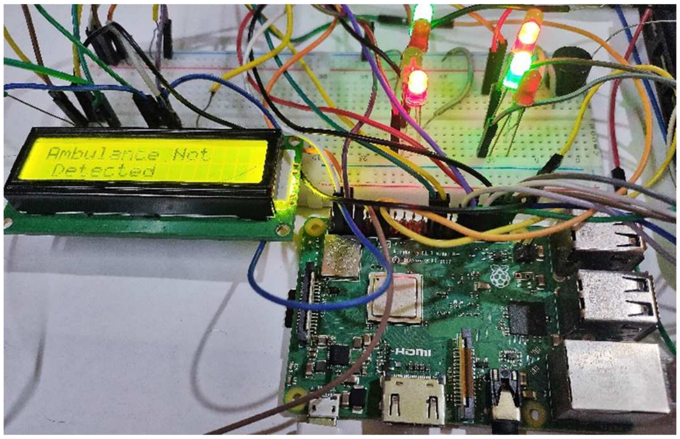
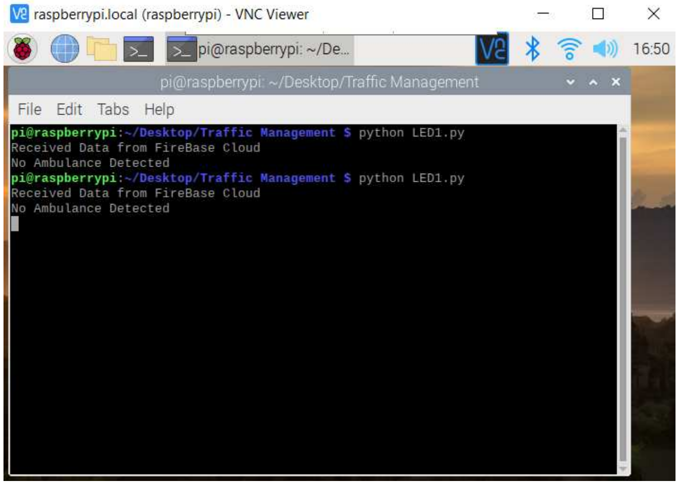


Fig. 5. Ambulance is not detected and the traffic lights are not changed


raspberrypi.local (raspberrypi) - VNC Viewer
pi@raspberrypi: ~/De...
16:50
pi@raspberrypi: ~/Desktop/Traffic Management
File Edit Tabs Help

```
pi@raspberrypi:~/Desktop/Traffic Management $ python LED1.py
Received Data from FireBase Cloud
No Ambulance Detected
pi@raspberrypi:~/Desktop/Traffic Management $ python LED1.py
Received Data from FireBase Cloud
No Ambulance Detected
```

Fig. 6. Data is retrieved and no ambulance is detected

In another scenario, an ambulance is detected and the data along with the path number is sent to Firebase and then to the Raspberry Pi. The traffic signal on Path 4 turns green while on other paths the traffic signal is turned to red as shown in **Fig. 7**. Moreover, the LCD which acts as an electric hoarding board and the Buzzer which acts as a speaker in the prototype are displaying the message "Ambulance on Path 4" and emitting warning sound respectively as shown in **Fig. 8**. This will alert the concerned traffic personnel and reduce the traffic at the intersection before the ambulance reaches there and thus it will have a clear and fast passage to the hospital. This action is performed across all traffic intersections on the path.

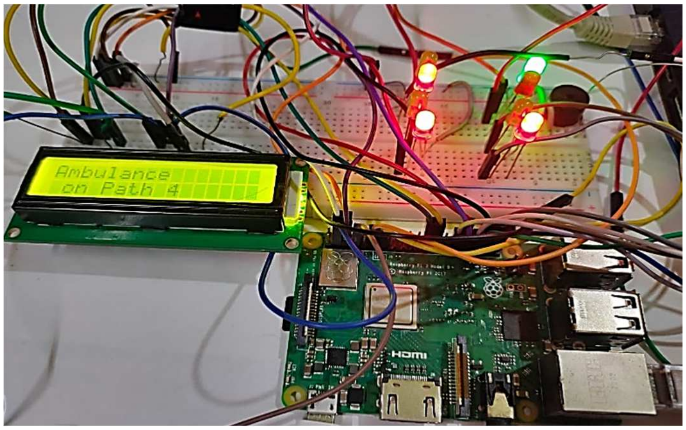
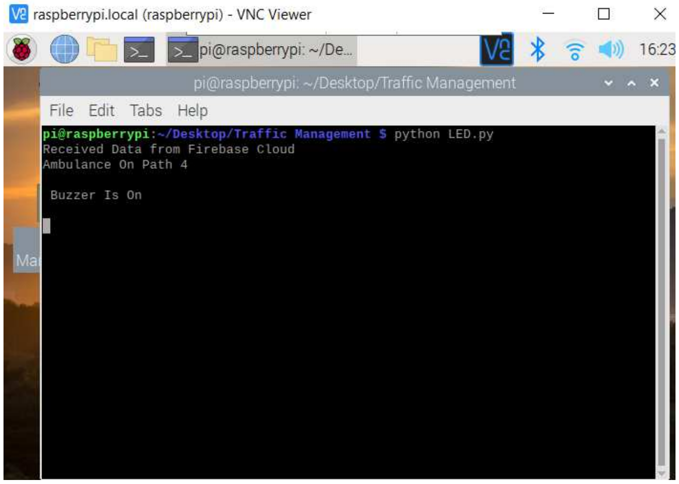


Fig. 7. Ambulance on Path 4 and the traffic lights are changed accordingly


raspberrypi.local (raspberrypi) - VNC Viewer
pi@raspberrypi: ~/De...
16:23
pi@raspberrypi: ~/Desktop/Traffic Management
File Edit Tabs Help

```
pi@raspberrypi:~/Desktop/Traffic Management $ python LED.py
Received Data from Firebase Cloud
Ambulance On Path 4

 Buzzer Is On
```

Fig. 8. Data retrieved from cloud and the Buzzer is on upon detection of ambulance

The traffic signals will not be changed abruptly upon detecting an ambulance as it will lead to some serious accidents. The cameras and the microphones along the roads must be installed much before the intersection to facilitate early detection. This will give enough time to change the signals slowly and steadily alerting the police as well as other drivers. In this way, the prototype can run smoothly with high effectiveness facilitating easy movement of the ambulances on the roads and thereby reducing early fatalities due to non-availability of health personnel.

V. CONCLUSION

A real time traffic management system has been implemented in this work which will detect the presence of ambulances on the road using CCTV. Audio tagging and classification has been used to confirm that the ambulance is on emergency service. Therefore, the authors have successfully built a prototype using YOLO-v3 for visual detection of ambulances. For audio tagging, Mel spectrograms of the surrounding sounds have been fed to a deep learning model to classify whether the ambulance is in emergency service or not. The positive output from the visual detection and audio tagging, generates an interrupt which is communicated to the hardware prototype in order to provide priority to the ambulance by stopping the ongoing traffic on other roads. Thus, the whole system allows faster arrival in hospital which means faster access to medical treatment. This faster mode of travel facilitates an average decrease in waiting time at the intersection thereby increasing the chances of survival of the patient.

REFERENCES

[1] K. V. Arya, S. Tiwari and S. Behwalc, "Real time vehicle detection and tracking," 2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2016, pp. 1-6, doi: 10.1109/ECTICon.2016.7561327.

[2] R. A. Bedruz, E. Sybingco, A. Bandala, A. R. Quiros, A. C. Uy and E. Dadios, "Real-time vehicle detection and tracking using a mean-shift based blob analysis and tracking approach," 2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), 2017, pp. 1-5, doi: 10.1109/HNICEM.2017.8269528.

[3] Jia L., Wu D., Mei L., Zhao R., Wang W., Yu C. (2012) Real-Time Vehicle Detection and Tracking System in Street Scenarios. In: Zhao M., Sha J. (eds) Communications and Information Processing. Communications in Computer and Information Science, vol 289. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-31968-6_70

[4] Xin Li, Xiao Cao Yao, Y. L. Murphey, R. Karlsen and G. Gerhart, "A real-time vehicle detection and tracking system in outdoor traffic scenes," Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., 2004, pp. 761-764 Vol.2, doi: 10.1109/ICPR.2004.1334370.

[5] "Emergency Vehicle Detection on Heavy Traffic Road from CCTV Footage Using Deep Convolutional Neural Network" " Shuvendu Roy", " Md. Sakif Rahman", 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019

[6] "Acoustic-based Emergency Vehicle Detection Using Convolutional Neural Networks", Van-Thuan Tran, and Wei-Ho Tsai, Member, IEEE

[7] Y. Zhao and P. Ioannou, ''A traffic light signal control system with truck priority,'' IFAC-PapersOnLine, vol. 49, no. 3, pp. 377–382, 2016.

[8] H. Chai, H. M. Zhang, D. Ghosal, and C. N. Chuah, ''Dynamic traffic routing in a network with adaptive signal control,'' Transp. Res. C, Emerg. Technol., vol. 85, pp. 64–85, Dec. 2017.

[9] V. Parthasarathi, M. Surya, B. Akshay, K. M. Siva, and S. K. Vasudevan, "Smart control of traffic signal system using image processing," Indian Journal of Science and Technology, vol. 8, no. 16, 2015.

[10] K. Nellore and G. P. Hancke, "Traffic management for emergency vehicle priority based on visual sensing," Sensors, vol. 16, no. 11, p. 1892, 2016.

[11] Zhao, Liquan & Li, Shuaiyang. (2020). Object Detection Algorithm Based on Improved YOLOv3. Electronics. 9. 537. 10.3390/electronics9030537.

[12] A. Meghanani, A. C. S. and A. G. Ramakrishnan, "An Exploration of Log-Mel Spectrogram and MFCC Features for Alzheimer's Dementia Recognition from Spontaneous Speech," 2021 IEEE Spoken Language Technology Workshop (SLT), 2021, pp. 670-677, doi: 10.1109/SLT48900.2021.9383491.

[13] M. H. Tanveer, H. Zhu, W. Ahmed, A. Thomas, B. M. Imran and M. Salman, "Mel-spectrogram and Deep CNN Based Representation Learning from Bio-Sonar Implementation on UAVs," 2021 International Conference on Computer, Control and Robotics (ICCCR), 2021, pp. 220-224, doi: 10.1109/ICCCR49711.2021.9349416.