A New Hybrid Architecture for Real-Time Detection of Emergency Vehicles

Eshwar Prithvi Jonnadula and Pabitra Mohan Khilar

Department of Computer Science and Engineering
National Institute of Technology, Rourkela
Email: {714cs1038,pmkhilar}@nitrkl.ac.in

Abstract. VANET is a vital part of wireless networking. Vehicular movement is expanding indefinitely everywhere and is causing terrible problems to daily life. Almost all of the traffic lights now feature a fixed green light sequence and so green light sequence is determined without taking the existence of emergency vehicles into consideration. Consequently emergency vehicles such as ambulances, fire engines, police vehicles etc. are struck in traffic which might cause loss of valuable life and property. In this paper we present a new hybrid architecture for detection of emergency vehicles in real time. This hybrid architecture is based on the mixed features of image processing and machine learning. We also show the percentage decrease in the search space for the processing which results in faster detection of emergency vehicles.

Keywords: VANET, ITS, Emergency Vehicles, Image Processing, Machine Learning

1 Introduction

VANET is the abbreviation of Vehicular Ad-Hoc Network is an integral part of wireless communication. It is a subset of MANET abbreviated as Mobile Ad-Hoc Network. VANET plays a crucial role in smart cities and smart computation. VANET consists of three major components namely Vehicle to Vehicle communication(V2V), Vehicle to Roadside units communication(V2R) and Roadside units to Roadside units communication(R2R) [3]. The primary challenge of VANET is the rapid movement of vehicles. The vehicles interact with others and with the roadside infrastructure units to send and receive relevant information regarding the traffic scenario, emergency information, etc.

In today's world travel and transport has become a standard and non-removable part of life. According to surveys, 40% of people spend 60 minutes on a trip on average each day. Humans and society are becoming more dependent on transportation every day. This problem of increased vehicles leads to congestion of traffic and increased pollution. In Beijing, China alone 400 thousand vehicles are present at the start of the year 2000 and nearly double the number, i.e., 800 thousand vehicles are added to the society at the end of that particular year [18].

The critical feature of traffic management is traffic lights. The traffic light assumes a primary job of smart city and adaptive traffic management. The time interval of green light and continuance of the green light are crucial things which are responsible for adaptive traffic control. In most of the parts of world traffic signals are settled and of a fixed time interval. These fixed traffic signals are however is only suitable for balanced traffic and not suitable for changing traffic conditions. These traffic signals are also programmed in such a way that these do not take the possibility of the presence of emergency vehicles which may come at any instance of time into consideration. In this case, the emergency vehicles get stopped in traffic and might result in loss of life and assets.

The NHTSA - National Highway Traffic Safety Organization collected the vehicle accident information for the USA in the span of the year 1992 to 2011. The data shows that there are about 4500 annual vehicle mishaps which include a lot of emergency vehicles. This report also gives an insight which tells that nearly 300 casualties take place every year despite a good traffic control system [10]. On observing the figures mentioned about the horrific accident, we need to have an intelligent transportation system which needs to take into consideration external factors like weather, traffic blockage and most importantly emergency vehicles as these can save lives [7]. The main difficulty in implementing an above system is the detection of an emergency vehicle which has been detailed addressed in the following section of this paper. This paper presents a new hybrid based method for real-time detection of emergency vehicles and also shows the decrease of search space for the algorithm compared to the previous strategies which are currently being used.

This paper is organized into six sections for a clear understanding of the reader. This paper mainly focuses on ambulances as it is an essential emergency vehicle. So in the future emergency vehicle refers to an ambulance. In the first section, we have explained the basics of VANET and the necessity of identification of emergency vehicles. In the second section, we will be discussing the various works done by the other researchers and authors. In the third section, we will give a basic idea of different types of methods available and useful for the detection of an emergency vehicle. In the fourth section, we see the difficulties involved in the existing practice, and we give a clear explanation of the proposed method for the detection of emergency vehicles. In the fifth section, we show the results of the proposed method discussed in the previous section. In the last and final part, we conclude the paper and give the details of future scope so that it can help other researchers in the right direction.

2 Related Works

For the clear understanding of the problem and how the issue has is solved we first need to study the previous works done by different researchers and authors in this particular field of study. We have done a significant amount of research and gathered the best-related works suitable for our paper for a better understanding of the readers.

- V Dangi et al. [4] uses real-time image processing techniques to implement an intelligent traffic controller. Various edge detection algorithms have been used in identifying vehicles.
- S Saravanan et al. [15] develops a system that can count the number of vehicles in a particular field of view which is done using an optical sensor. The traffic signal timing is dynamically changed according to the traffic intensity.
- F. Andronicus et al. [2] use a ripple algorithm to detect an ambulance. This algorithm uses the concept of template matching techniques to compare the input to the existing template.
- V. Parthasarathi et al. [11] use a new idea of the centroid of lights to detect the ambulance. First, the red and blue colors of the ambulance siren are identified using the segmentation method of image processing. Then the ambulance is recognized if and only if the distance between the centroids of the red and blue part is less than a predefined threshold.
- Rajeshwari Sundar et al. [16] uses a basic hardware module called the RFID - Radio Frequency Identification which is installed on every ambulance to identify the presence of the ambulance.
- Bartlomiej Placzek et al. [12] makes use of data gathered by GPS - Geographical Position System and pattern matching algorithms to recognize the presence of emergency vehicles.
- Deepa et al. [5] use the basic concepts of computer vision. The emergency vehicles are recognized using the OCR - Optical Character Recognition method.

3 Ambulance Detection

Ambulance detection is the first and primary step in adaptive traffic lights and smart cities. Only when the ambulance is detected, then the signals can change and give preference to the ambulance. There are many methods for detection of ambulances. Here are a few methods mentioned below.

- **RFID-Radio Frequency Identification:**In this method, every ambulance is fitted with an RFID device which has its frequency. When an ambulance passes through a junction, it is recognized by the RFID reader.
- **OCR-Optical Character Recognition:** In this method, we use the fundamental image processing technique of OCR to detect the key works like "AMBULANCE", "EMERGENCY" etc. to recognize the emergency vehicles.

- **Template Matching:**In this method, the ambulance is detected by comparing the vehicles present on the road to a predefined existing template of the ambulance which is obtained by various edge detection methods.
- **Siren Lights Frequency:**In the method, the ambulance is detected by the frequency of the siren light. In many countries, there are fixed standards for the timing, color and frequency of the siren light of an ambulance. These can be exploited for the identification of the ambulance.
- **Centroid of Siren Lights:**In this method, we detect the ambulance if the distance between the centroid of the blue and red light of siren is less than a predefined threshold.

4 Implementation

4.1 Difficulties of Video Processing

The general method for the implementation of the detection of emergency vehicles is video processing. This task involves various methods as mentioned in the previous section. These may always not give the correct results. However, in our case, we need to be very sure as the detection of emergency vehicles is a crucial factor in the smart city and traffic management. So machine learning techniques are used. This process is called activity recognition.

Action recognition task includes the recognizable proof of various actions from video cuts (a succession of 2D outlines) where the action could conceivably be performed all through the whole span of the video [1]. Despite the great achievement of profound learning classification in the picture, advance in models for video classification has been slower. The following are the reasons for the slow and arduous process of action recognition.

1. **Enormous Computational Cost:** A simple two-dimensional convolution for ordering 100 classes has nearly a million variables to learn whereas similar classification when it comes to the three-dimensional structure has nearly 33 million variables to learn. It takes nearly 100 hours to train three-dimensional convolution network the standard benchmark dataset of UCF101 [17].
2. **Catching Huge Data:** Action recognition includes catching the spatial-temporal data of the frames of video. Furthermore, this data caught must be made up for the movements and motion of the camera. All the details from very small to the main action needs to be precisely captured to recognize the activity accurately.
3. **Lack of Standard Dataset:** The mainstream and benchmark dataset have been UCF101 and Sports1M for quite a while. To get a perfect structure for the convolution neural network for these datasets can be incredibly costly. However, these datasets only contain a few actions like sports which are not useful for most of the problem in the real world. In our case, we need a dataset of ambulance moving in the streets for the action recognition task. This issue of lack of dataset is added but not completely solved with the kinetic dataset [8].

4.2 Proposed Method

The difficulties of performing a machine learning or deep learning techniques have been mentioned in the above subsection. In our case, we cannot afford such a delay as the detection of an emergency vehicle should be done in real time.

So we propose a two-stage hybrid algorithm which combines the power of machine learning and the simplicity of image processing. This method has adopted due to the lack of video dataset of emergency vehicles on the road. As the dataset is not available, the standard machine learning algorithms are not possible to implement.

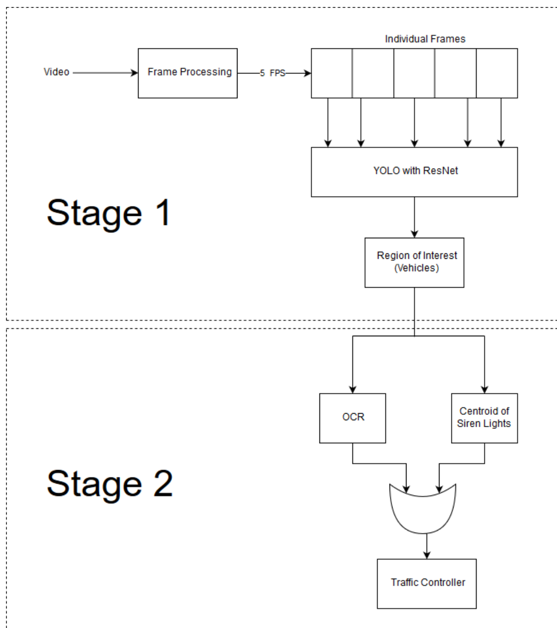

Fig. 1. Architecture of Proposed model

Generally, a video consists of 25 to 30 frames per second. Hence the search space for the existing machine learning and deep learning becomes enormous. This results in multiple methods and large computation time.

So we use only five frames per second, i.e. 200ms for each frame. This process is done to decrease the search space. This process does not result in missing of emergency vehicles as the vehicle cannot escape the camera in a short span of of200 ms.

These five frames are sent to the highly efficient and advanced neural network designed and developed by Google called ResNet which integrated with object detection algorithm called You Only Look Once(YOLO) [14] is used.

This network is used in finding all the vehicles in the five frame. This process is technically called the region of interests.

This region of interests is passed to the modules of image processing in stage 2. Here in our proposed architecture, we use two image processing methods to detect the presence of an emergency vehicle namely Optical Character Recognition and Centroid-based method of the siren bar. The output of these is binary as it only tells if the vehicle is an emergency vehicle or not. In our case, it tells if it an ambulance or a regular vehicle. This output is passed through the binary OR gate. This is done if the any of the image processing modules gives the output as true, i.e. an emergency vehicle is present then the traffic controller will change the traffic lights to give the priority for the emergency vehicle. The OR gate is selected because every time we cannot relay that the image processing modules work correctly in all situations like day or night. So we alert the traffic controller even if any one of the models of image processing says that there is an emergency vehicle present.

4.3 You Only Look Once

YOLO is state of the art object detection algorithm but requires high-end processing hardware for the training of the dataset. In YOLO, object detection is considered as a regression which takes input as pixels and outputs bounding boxes and probabilities of the object being in different classes. YOLO is completely different from the sliding window method. YOLO analyses the whole pixels in the picture at the time of training so that the prediction of class labels in testing can be much faster as compared to other object detection methods. YOLO makes less number of mistakes in detecting background than a Fast R-CNN because YOLO looks at the entire pixels during the training phase [14].

In YOLO the image is divided into a number of square pieces using grid. Each grid has the duty of detecting the object at its grid center. Each grid gives the output of bounding boxes and the probability of object belonging to different classes. Some of these bounding boxes which have a low Intersection over Union value with the anchor boxes predefined are eliminated. For each output there are five values corresponding to it. X coordinate of the center of bounding box, Y coordinate of the center of bounding box, width of bounding box, height of bounding box and the highest probability of which class the object belongs to.

4.4 ResNet

A Residual Network commonly referred to as ResNet is a type of neural network that uses skip connections to bypass or skip some layers in the neural network. One of the main reason the skip connections are used is the effect of vanishing gradient in a deep neural network. By theses skip connection the activation function value from one layer can be used in the layer which is deep down in the network. This effectively takes into consideration the initial weights of the network even after the vanish gradient problem occurs [6]. These skip connection are responsible for the simplicity of the neural networks by using a less number of layer in training phases. Skip connections reduce the effect of vanishing gradient problem as the data flows through less number of layers bypassing the intermediate layers using these skip connections. So the features are retained even at the end of the network.

5 Results and Analysis

5.1 Results

The above architecture is implemented in Python 3.5. We have taken the pre-trained weights from Darknet [13] used in YOLOv2 network which can identify vehicles on the road. In the first stage of the architecture, the five frames are sent to YOLOv2 with ResNet to get the regions of interest.

Fig. 2. Input for YOLO V2 with ResNet

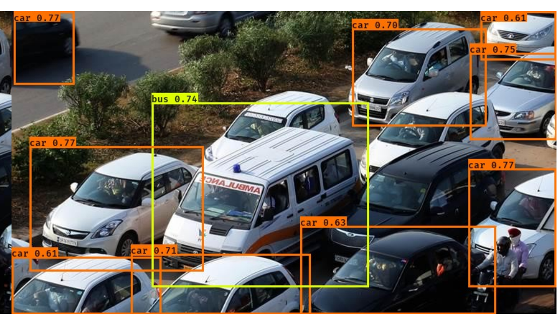

Fig. 3. Output with the regions of interest

After Stage 1 we get all objects present in the frame. But we aim to consider the three out of 80 different types of classes of COCO dataset [9] as mentioned below.

1. Car
2. Bus
3. Truck

We crop out this region of interests and send them to stage 2. Here we show the processing of 'bus' object for the understanding of the readers. We send this cropped portion to the OCR module to detect the text 'Ambulance'. Since in some cases the ambulance will be the mirror image we flip the image accordingly so that the character can be recognized easily.

Fig. 4. Input for OCR module after flipping

Fig. 5. Text recognized by the OCR

The output of the OCR module will be all the text that it recognizes in the image. These can also be garbage texts present in the frame. But we are particularly interested in the word **AMBULANCE**. Once this word is recognized, we send the control information to the traffic controller module.

5.2 Analysis

The primary purpose of this hybrid architecture is to decrease the search space. In this subsection, we do an approximate calculation of search space reduced compared to the simple video processing techniques in the identification of emergency vehicles.

Video Processing
In this method consider we have an HD camera installed at the traffic junction. Generally, the videos have a 25-30 FPS. Let us consider the best case of 25 FPS.
Each frame in single channel=1280*720 pixels
Each frame in all three channels=1280*720*3 pixels
Each Second=1280*720*3*25 pixels
This search space is approximately 70 Million pixels per second.

Proposed Hybrid Architecture
Consider the same camera installed. But in our architecture, we use only 5 FPS.
In Stage 1
Each frame in single channel=1280*720 pixels

Each frame in all three channels=1280*720*3 pixels
Each Second=1280*720*3*5 pixels
This search space is approximately 14 Million pixels per second.
In Stage 2
Consider 50% of pixels objects detected.
Let us assume 50% of those objects are cars, buses and trucks. Let the remaining 50% be pedestrians , motorcycles etc.
One module=14 M * 0.5 * 0.5
Two modules=14 M * 0.5 * 0.5 * 2
This search space is approximately 7 Million pixels per second.
So the total search space for the hybrid architecture is 14+7=21 Million pixels per second.
As we can see the percentage decrease in search space is nearly 70%.
This comparative analysis is done considering the best case conditions for video processing and average case conditions for the hybrid architecture.

6 Conclusion and Future Scope

6.1 Conclusion

An expanded number of vehicles does not only boosts the response time of emergency vehicles and also hikes the chances of them ending up in accidents.

This paper presents a way to detect the presence of emergency vehicles using a new hybrid algorithm which combines the power of machine learning and the simplicity of image processing. We also saw that the search space of the architecture is decreased.

6.2 Future Scope

The output of both video processing and audio processing can be used together. They can be integrated into a system for a module for intelligent traffic controller. This traffic controller can give priority to the emergency vehicles if detected by the audio and video processing sub modules. This detection of emergency vehicles plays a significant role in smart city development.

References

1. Deep Learning for Videos: A 2018 Guide to Action Recognition, http://blog.qure.ai/notes/deep-learning-for-videos-action-recognition-review
2. Andronicus, F., Maheswaran: Intelligent ambulance detection system. International Journal of Science, Engineering and Technology Research (IJSETR) **5**(5) (2015)
3. Bhoi, S.K., Khilar, P.M.: Vehicular communication: a survey. IET Networks **3**(3), 204–217 (2013)
4. Dangi, V., Parab, A., Pawar, K., Rathod, S.: Image processing based intelligent traffic controller. Undergraduate Academic Research Journal (UARJ) **1**(1) (2012)

5. Deepa, Navya, K., Manisha, K., Manu, M., Kshama, S.: Smart detection of emergency vehicle in traffic. International Journal of Current Engineering and Scientific Research (IJCESR) **5**(4) (2018)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
7. Jonnadula, E.P., Khilar, P.M.: Comparison of various techniques for emergency vehicle detection using audio processing. Machine Learning, Image Processing, Network Security and Data Sciences (March 2019)
8. Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., Viola, F., Green, T., Back, T., Natsev, P., et al.: The kinetics human action video dataset. arXiv preprint arXiv:1705.06950 (2017)
9. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision. pp. 740–755. Springer (2014)
10. Nellore, K., Hancke, G.P.: Traffic management for emergency vehicle priority based on visual sensing. Sensors **16**(11), 1892 (2016)
11. Parthasarathi, V., Surya, M., Akshay, B., Siva, K.M., Vasudevan, S.K.: Smart control of traffic signal system using image processing. Indian Journal of Science and Technology **8**(16) (2015)
12. Placzek, B., Golosz, J.: The in-town monitoring system for ambulance dispatch centre. arXiv preprint arXiv:1706.03699 (2017)
13. Redmon, J.: Darknet: Open source neural networks in c. http://pjreddie.com/darknet/ (2013–2016)
14. Redmon, J., Farhadi, A.: Yolo9000: Better, faster, stronger. arXiv preprint arXiv:1612.08242 (2016)
15. Saravanan, S.: Implementation of efficient automatic traffic surveillance using digital image processing. In: Computational Intelligence and Computing Research (ICCIC), 2014 IEEE International Conference on. pp. 1–4. IEEE (2014)
16. Sundar, R., Hebbar, S., Golla, V.: Implementing intelligent traffic control system for congestion control, ambulance clearance, and stolen vehicle detection. IEEE Sensors Journal **15**(2), 1109–1113 (2015)
17. Tran, D., Ray, J., Shou, Z., Chang, S.F., Paluri, M.: Convnet architecture search for spatiotemporal feature learning. arXiv preprint arXiv:1708.05038 (2017)
18. Zhang, J., Wang, F.Y., Wang, K., Lin, W.H., Xu, X., Chen, C.: Data-driven intelligent transportation systems: A survey. IEEE Transactions on Intelligent Transportation Systems **12**(4), 1624–1639 (2011)