2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019

Emergency Vehicle Detection on Heavy Traffic Road from CCTV Footage Using Deep Convolutional Neural Network

Shuvendu Roy
dept. of Computer Science and Engineering
Khulna University of Engineering & Technology
Khulna, Bangladesh
bikash.shuvendu@gmail.com

Md. Sakif Rahman
dept. of Computer Science and Engineering
Khulna University of Engineering & Technology
Khulna, Bangladesh
sakif1407023@gmail.com

***Abstract*—A highly populated country like Bangladesh faces too much traffic jam. Sometimes emergency vehicles like ambulance, fire-fighter get stuck in the traffic causing threat to life in many cases. It is important to give priority to this car and help to clear its path. But it is difficult or sometimes impossible for traffic police to handle this. For this reason, we need an automated system that will be able to detect an emergency car in heavy traffic road, let the controller know or automatically navigate other cars to clear its path. In this work, we have proposed an automated system to detect emergency cars from CCTV footage using the deep convolutional neural network. Our method has shown good result in detecting and classifying emergency cars.**

***Index Terms*—Emerhecy vehicles detection, Object detection, CNN**

I. Introduction

Emergency vehicles play an important role in every life-threatening situation. Traffic jam takes more than 20% patient lives in an ambulance but when the patients condition is very serious the percentage of patient death is increased [1]. These are situations when an emergency patient needs to go to the hospital immediately and the ambulance got stuck in the traffic jam. This scenario is dangerous in case of heart patients who needed to be rushed to hospital in time. In traffic jams, many people do not bother to give pass-way for the emergency vehicle and also traffic police cant see which lane they should clear for the ambulance. Therefore many patients lose their lives before reaching hospitals. Traffic jam situation is also a very big challenge for the firefighter teams. In the USA at least 90% cases where a firefighter team is needed, people expect that fighter team would respond within four minutes [2]. In 2013 firefighter team got calls more than 1.2 million times in the USA. Though firefighter responded that places very quickly, 3240 civilian deaths and 15,925 injuries and more than $11.5 billion in property is lost during the year [3]. Every second is very valuable for the firefighter team. Many peoples lives and many properties are lost for delaying the emergency firefighter services on the emergency situation.

We can reduce these problems by introducing an intelligent automated system integrated with a traffic control system that will detect and give priority to emergency cars. We need to a build a system to detect cars and classify it as an emergency or regular car. Deep learning architecture like deep convolutional neural networks have been applied in computer vision and they have produced results not only comparable to human but also superior to a human expert. In this paper, the problem is solved by taking the CCTV footage of the road and detect the emergency vehicle. Images are taken every second with the help of CCTV camera. In every image, every vehicle is detected on the road. After detecting every vehicle, they have classified it into an emergency vehicle and regular vehicles. If an emergency vehicle is found, the computer can notify the traffic police or an automated system to clear its way.

A human brain can easily detect any vehicle easily and visually process it in a fraction of time. But the brain has some limitation, it can monitor or sustain attention not more than half an hour. So the future world needs many artificial brains to do the hard work for helping mankind. This paper contains 2 major part. One is object detection and classification of an image as an emergency vehicle or not. For object detection we use Yolo-V3 [4]. Yolo-V3 architecture is very fast and can process 45 images per second in a computer with a good processor. Yolo-V3 reframes the image and divides it in fixed grids. Then it predicts multiple bounding boxes and finds the probability of an object those boxes. So it does not need complex pipelining and directly optimizes detection performance. In R-CNN [5] is a region proposal based method that generates boxes and runs

978-1-5386-9111-3/19/$31.00 ©2019 IEEE

classification then post-processing for bounding boxes. R-CNN pipeline is slower than YOLO detection. YOLO makes less background error. YOLO method is highly generalizable of object detection. More discussion about this is in session III.

Deep convolutional neural network [6] is the most advanced technology for classification and detection in the image. In this days the object detection and classification capabilities have dramatically improved. The convolutional neural network uses the concept of a kernel to process the image. The values of these kernels are learned during training. A deep convolutional neural network is constructed stacking many of such convolutional layers. But it has two drawbacks; it is hard to train such network and it requires a lot of data to train it. To solve this kind of problems transfer learning [7], [8] can be used. The idea is to use the knowledge, in this case, the weight of a neural network, learned by a network in one application in other application. We have experimented with few different pre-trained model including VGG-16 [9], inception-v3 [10] and xception [11].

II. Related works

The number of vehicles is increasing dramatically and the traffic problems are increasing day by day. So the emergency vehicles such as ambulance, firefighter truck face many difficulties because of a large number of vehicles. A few systems were made for detecting an emergency vehicle. Nellore et al. [12] calculated the distance of an emergency vehicle using the camera and informing the Traffic Management Center (TMC). They used visual sensing technique. A camera was used to record 1920 1080 pixels video with 30 frames per second. First, the images are converted RGB to grayscale then threshold the images. Some morphological operations were applied to each image and after that, they measured the distance between the camera and the emergency car in a different technique like Euclidean distance in MATLAB. Then they sent the distance, speed and count variable data to the traffic management center to control the traffic signal more effectively for an emergency vehicle. On the other hand, Djahel et al. [13] designed an advanced traffic control system that minimized the emergency vehicle congestion level. A traffic management controller architecture was made with the help of fuzzy logic controller for emergency services. Their method had got the control of changing the traffic light, changing the speed limit, lane clearance etc. Fuzzy logic determinates most accurate evaluation of the low, medium, high congestion level.

Another idea was given by Parthasarathi et al. [14] that an intelligent traffic system that implements some embedded system for giving more priority of an emergency vehicle in a traffic control system. They measured the density of the vehicle by the infrared detector but could not work efficiently in real time scenarios. The toy car was used to make the prototype of a traffic model. At first, the images were cropped into the region of interest like only roadside images. The red and the blue color were only selected in the images and measured the distance between each red and the blue color. If the distance is very narrow than it takes a high chance that is an ambulance upper side LED light. The images were taken only the top sight of a road. They also measured the density of the vehicle. The images were turned into the grayscale images and some morphological operation are applied to it. The noise of the images was is removed by the Gaussian filter and counted the set of connected pixels.

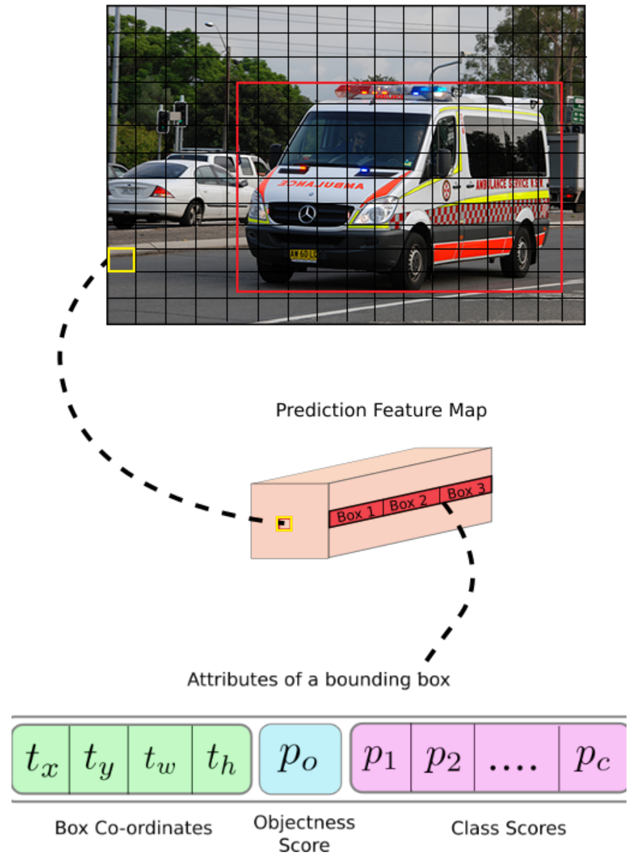


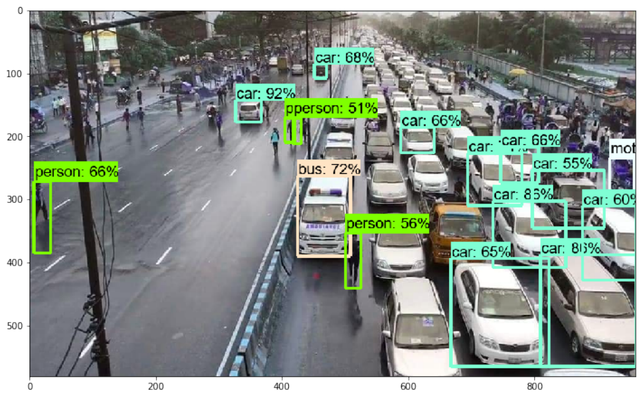
Fig. 1. Prediction of feature maps of a sample image. It predicted three boxes per grid. Each box has box co-ordinate, objectness score, and class prediction

III. Emergency Car Detector

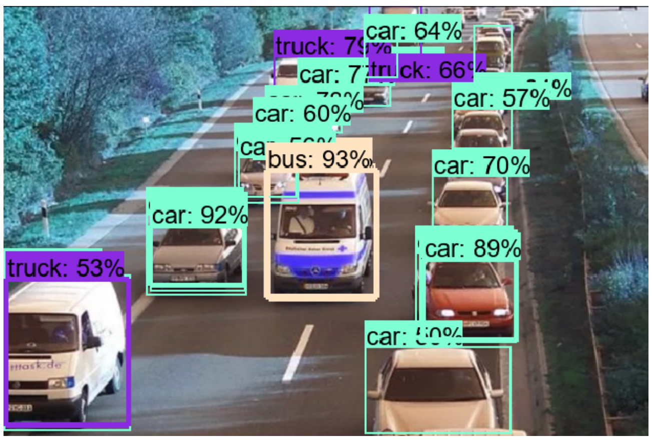
A. Car Detector

The first step towards emergency car detection object detection. For our purpose, the job of the object detector is to train a car detector which will output the images of cars only. Which we will use to classify whether it is an emergency car or a regular car. We have used YOLO-V3 [4] for that. Although it is not the best object detection algorithm it is the fastest model. As we have to run many images to the model which will be generated from the CCTV footage, it is for the detection algorithm be fast.

Yolo-V3 used a deep convolutional neural network called Darknet-53 architecture. It is a 53 layer neural network trained on ImageNet for classification combined with detection layers



(a)



(b)

Fig. 2. Examples of car detection by YOLO object detector

making the total network 106 layers deep. It gives detection at 3 different scales. The image it first divided into a grid. As shown in Fig. 1 each block outputs 3 boxes. Each box represents one object detection at a given center. Each box consists of coordinates of the object, objectness score and class scores. Box co-ordinate consists of the center, hight, and width of the object. Objectness score is the probability of the grid detecting an object. Class probability is the probability of each class the model is trained on. In this case, the model is trained on COCO dataset [15]. It contains 80 classes of object. so, the class score provides the probability of 80 objects.

The relevant classes available in COCO dataset are 'truck', 'bus', and 'car'. All images categorized in these class are passed through the classifier to classify an emergency or not emergency vehicles. Few samples of detection are shown in Fig. 2

B. Emergency classifier

We used a deep neural network for emergency car classification. The variation of neural network that is very popular and accurate for image related task in called Convolutional neural network(CNN) [6]. We had experimented with several different CNN architecture for classification. The is provided in the experiment section. But the problem with shallow CNN network is that it can not capture the patterns in the data for a larger resolution image. Moreover training big neural network requires a lot of data.

To solve around these problems a method called transfer-learning is used. The idea of transfer-learning is to use a deep convolutional neural network which is pre-trained with a big dataset like ImageNet. We can use this model as a feature extractor with a newly defined fully connected layer on top of the pre-trained model. Then we can train a few of the previous layers of the trained model. This process is called fine-tuning the pre-trained model. For example VGG-16 [9] is a deep convolutional neural network with is trained with ImageNet dataset. We added and trained the fully-connected layer and fine-tuned one block of the pre-trained convolutional neural network (Conv Block 5). Fig. 3 shows the model architecture of a VGG-16 model and the part of the model that is trained, fine-tuned and left untouched(Frozen).

IV. Experiments

A. Experimental Datasets

For vehicle detection we use the Common Object in Context (COCO) dataset [15]. COCO is a dataset which is a large scale object detection, captioning and segmentation. COCO has some features like recognition in context, segmentation in object and 330,000 images. This dataset segmentation and detection object found in everyday life in our natural environments. We used a pre-trained model for object detection which has 80 classes. There are many objects that could detect on the road like a bicycle, car, motorcycle, bus, truck etc. All the vehicle on road can be detected in this object detection method.

For emergency vehicle classification, we use 2 classes. The classes are the emergency vehicle and the regular vehicle. For regular vehicle class, we use Stanford Universitys cars dataset [10]. There are 8144 images of the regular car, sports car, van, SUV, truck etc. Form this dataset we select 1500 images of different types of vehicles for training and testing the deep neural network. For emergency vehicle class we collect ambulance and firefighter truck form google search. There are many images of ambulance or firefighter truck in Wikipedia, which image are used to describe an ambulance or fire engine. Around 800 images are collected of the ambulance and 700 images are collected of the fire truck. This 1,500 emergency vehicle images also split randomly 90% for training and 10% for testing purposes. Fig. 5 shows a few samples of the training image.

B. Data augmentation

Data augmentation is an efficient process to use with image data to remove overfitting and train the model better. It is a

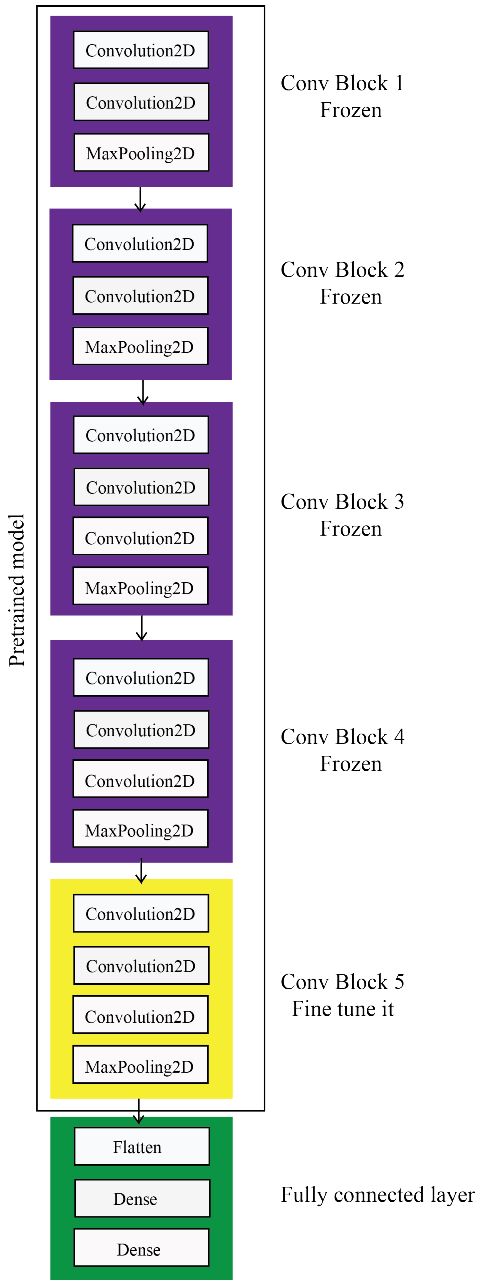


Fig. 3. Architecture of the VGG-16 model. Here pre-trained model section is trained on ImageNet dataset. The fully connected layer is the portion we have added on top of the pre-trained model. "Frozen" represents the portion of the model is not changed(fine-tuned) at all during the training process. The yellow layers are fine-tuned with the emotion dataset.

process of generating more data from the available data by rotating, scaling, translating, cropping and other transformation. An embedded data loader does the job at train time. When it loads the data, it applies the transformation and generates a slightly different image each time. As the images of our dataset are nicely collected we carefully applied only a little amount of augmentation. The augmentation types and amount is summarized in the table below:

TABLE I
VALUES OF AUGMENTATION APPLIED TO THE TRAINING IMAGE

Type of Augmentation	Values
Rotation	-20° to 20°
Horizontal Flip	Yes
Scaling factor	1.1

C. *Experimental setup*

Instead of using regular stochastic gradient descent to train the model we used Adam [16] optimization algorithm. Rather than using a fixed learning rate, Adam uses an adaptive learning rate. It has two parameters, beta1 and beta2 to control the learning rate. These are called momentum parameters. The parameters of the optimizer are given in the table below:

TABLE II
PARAMETERS OF ADAM OPTIMIZER

Parameter	Values
Learning rate	0.001
Beta 1	0.9
Beta 2	0.999

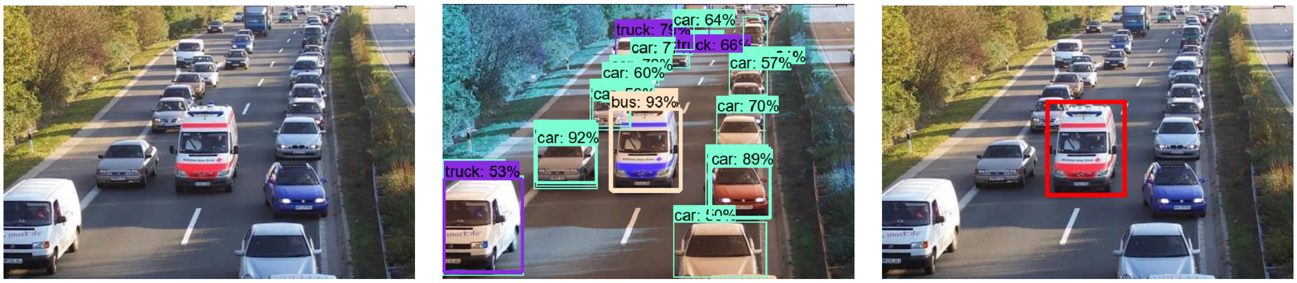
V. RESULTS

Before we experiment on our model we built a basic 2-layer convolutional neural network to find a baseline accuracy. We conducted several experiments with some hyper-parameters and regularization techniques. The accuracies of different experiments are given in the table below.

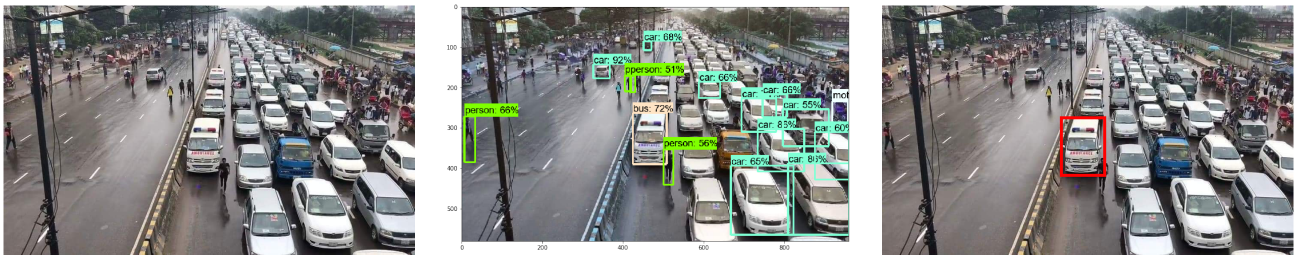
TABLE III
ACCURACIES FOR DIFFERENT INPUT SIZES

Input image size	Accuracy
(48×48)	97.97%
(64×64)	98.39%
(128×128)	97.51%
(224×224)	97.97%
(360×360)	97.97%

As we can see from the table above, the simple model produced the best result with dimension 64×64. Although larger images have more information to process and give a better result, a shallow convolutional neural network cannot take advantage of this as it does not have enough parameters and complex structure to capture such patterns in a large image.



(a) Original image (b) Detected Car (c) Detected emergency car



(d) Original image (e) Detected Car (f) Detected emergency car

Fig. 4. Examples of emergency car detection with our method. The top image shows a good example of detection. The below image shows a bad example of car detection by the object detector, which missed a few of the car but got the close ambulance. The emergency classifier did good in both cases.

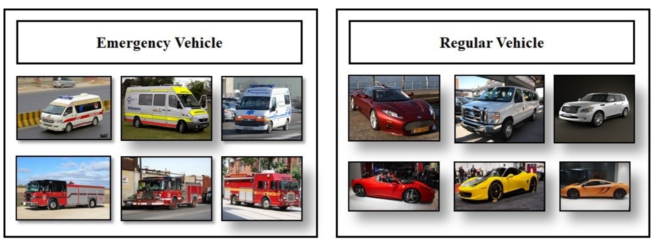


Fig. 5. Samples from training data

We conducted experiments with few different models. we used image size of 224×224 Here are the results of all models:

TABLE IV
COMPARISON OF THE ACCURACIES WITH DIFFERENT PRE-TRAINED MODEL IN KDEF DATASET

Pre-train model	Accuracy
2-layer CNN	98.39%
VGG-16 [9]	99.73%
Inception-v3 [17]	97.57%
Xception [11]	98.84%

Fig. 4 shows two examples of detecting an emergency car in the road.

VI. Conclusion

In this work, we have proposed a model that can detect emergency cars on a heavy traffic road. A populated country like Bangladesh faces too much traffic on the road and because of that emergency car like ambulance and fire-service fall into trouble middle of the road. Our model will solve this problem. It can be embedded with CCTV to track emergency can and give priority in that road to pass the emergency can. With this automated process, no human effort will be required to manually help such scenario. Our model has achieved impressing results in detecting and identifying emergency cars of all kinds.

References

[1] H. K. Kanzaria, M. A. Probst, and R. Y. Hsia, "Emergency department death rates dropped by nearly 50 percent, 1997–2011," *Health Affairs*, vol. 35, no. 7, pp. 1303–1308, 2016.

[2] L. C. Smeby Jr *et al.*, *Fire and emergency services administration: Management and leadership practices*. Jones & Bartlett Publishers, 2013.

[3] M. J. Karter, *Fire loss in the United States during 2010*. National Fire Protection Association Quincy, MA, 2011.

[4] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[5] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[6] Y. LeCun, Y. Bengio *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[7] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[8] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Advances in neural information processing systems*, 2014, pp. 3320–3328.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[11] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *arXiv preprint*, pp. 1610–02 357, 2017.

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

[13] S. Djahel, N. Smith, S. Wang, and J. Murphy, "Reducing emergency services response time in smart cities: An advanced adaptive and fuzzy approach," in *Smart Cities Conference (ISC2), 2015 IEEE First International*. IEEE, 2015, pp. 1–8.
[14] 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.
[15] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.
[16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.