LEVERAGING COMPUTER VISION FOR EMERGENCY VEHICLE DETECTION- IMPLEMENTATION AND ANALYSIS

Kaushik S
Student, Dept. of Industrial Engineering and Management,
RV College of Engineering,
Bangalore, INDIA
kaushiks.im18@rvce.edu.in

Abhishek Raman
Student, Dept. of Industrial Engineering and Management,
RV College of Engineering,
Bangalore, INDIA
abhishekraman.im18@rvce.edu.in

Dr. Rajeswara Rao K.V.S
Associate Professor
Industrial Engineering and Management
RV College of Engineering
Bangalore, INDIA
rajeswararao@rvce.edu.in

,

***Abstract*- Recent advances in Computer Vision technology have revolutionized the field of Intelligent Transportation Systems. The applications are far reaching- right from traffic monitoring systems to self-driving cars. Most applications entail at least simple, if not advanced image or video analytics at a fundamental level. This paper is an attempt to examine the use of object detection and instance segmentation for emergency vehicle detection, which is indispensable to any Intelligent Transportation System. More particularly, emergency vehicle detection can be programmed into autonomous vehicles as well as traffic signal controllers for preferential signal switching upon encountering emergency vehicles. The architectures implemented are Faster RCNN for object detection and Mask RCNN for instance segmentation. The computational results of these implementations, their accuracies and most importantly, their suitability for emergency vehicle detection in disordered traffic conditions are deliberated. Additionally, the object detection model is contrasted with instance segmentation and the merits and demerits of each are identified, again in the context of emergency vehicle detection.**

***Keywords*- Emergency Vehicle Detection, Computer Vision, Object Detection, Instance Segmentation, Convolutional Neural Network, Faster RCNN, Mask RCNN**

I. INTRODUCTION

To realize an unimpeded movement of emergency vehicles, two synchronous mechanisms must be constructed viz. a mechanism to detect the presence of an emergency vehicle in a traffic congestion and a post detection mechanism that clears the traffic for the emergency vehicle. An array of diverse frameworks has been proposed envisaging these mechanisms, involving various levels of automation i.e. manual, partially automated and completely automated. The primary drawbacks of manual and partially automated systems are lapses in detection and poor responsiveness (higher response time post detection of emergency vehicle). However, an autonomous or completely automated system is more robust with higher detection accuracies as well as swift response times, the proviso being that it has to be programmed in a comprehensive manner to accommodate the possible scenarios of detection (under high traffic, interference, noise etc.). This paper explores state of the art computer vision algorithms as tools to automate the detection of emergency vehicles i.e. the first mechanism of the two necessary mechanisms explained previously. The efficacy of these algorithms is assessed vis a vis two use cases- intelligent traffic signal and autonomous vehicles.

II. RELATED WORK AND MOTIVATION

Rajeshwari Sundar et al [1] have implemented a Radio Frequency Identification (RFID) based detection system in which emergency vehicles equipped with RFID tags are recognized by an RFID reader installed at a signal or a junction. Another RFID based emergency signal detection system was devised by Shajnush Amir et al. [2] using Programmable Logic Controllers (PLCs). An alternate technology is to deploy sophisticated microphones to detect audio signal or emergency siren. Bruno Fazenda et al [3] investigated a cross microphone array system for acoustic signal detection.

With the advent of computer vision, image processing has found applications in emergency vehicle detection as well. Shuvendu Roy and Md. Sakif Rahman [4] employed YOLO object detection for isolating an emergency vehicle in an image. [5] demonstrates the superiority of computer vision over RFID. However, computer vision in the domain of emergency vehicle detection is still in its fetal stages, especially in a developing economy like India. India, in fact presents a unique setting for computer vision, as eccentricities like heavy traffic, lane indiscipline, haphazard and non-homogeneous movement must be factored into any algorithm. It is safe to say that an algorithm modelled on foreign conditions will be rendered ineffective in India. There is very limited literature on computer vision techniques for emergency vehicle detection, especially factoring in the eccentricities of the Indian context and hence the goal of this research is to execute

and contrast a couple of image processing algorithms i.e. object detection and instance segmentation in an exclusive Indian context and further the cause of applying computer vision for emergency vehicle detection.

III. METHODOLOGY

The two selected computer vision techniques for emergency vehicle recognition are object detection and instance segmentation. Specifically designed Convolutional Neural Networks (CNNs) called Faster RCNN and Mask RCNN are employed for object detection and instance segmentation respectively. These CNNs were initially trained, iteration wise, to recognize features of emergency vehicles from their images. This was followed by comprehensive testing of the trained models for detection accuracy on an alternate unseen dataset. The results were categorized as true positives, false positives, true negatives and false negatives.

IV. THE OBJECT DETECTION MODEL

A. Principle and Architecture

Object detection is the localization of a particular object in an image by means of generating a bounding box around the object. Conventional convolutional neural networks that perform image classification (Resnet, VGG Net, Inception Net etc.) serve as the backbone for object detection. The principle of transfer learning is employed i.e. these base networks are pre-trained on large pre-existing datasets to minimize the number of computations as well as the number of images required for the custom dataset. They are also made fully convolutional to accommodate the inputs of random dimensions. Additionally, these base networks are supplemented by object detection networks like Faster RCNN, Single Shot Detectors (SSD) and Region based Fully Convolution Networks (R-FCN). Figure 1 depicts a Faster RCNN with a VGG Net base network.

The CNN architecture that has been chosen for implementation is Faster RCNN with a Resnet backbone [6]. The Faster RCNN is itself a combination of two networks, the Fast RCNN and the Region Proposal Network (RPN). The backbone network of the RPN generates an output feature map that is examined by the RPN for possible presence of object under detection. There is extensive parameter sharing between the backbone networks of both the RPN and the Fast RCNN to facilitate computational efficiency.

B. Nature of Dataset and Training

A custom dataset consisting of 400 downloaded images of emergency vehicles was assembled and divided in an 80:20 ratio for training and validation respectively. The nature of the images ensured that traffic congestion, disorderly movement and non-homogeneity in the shapes and sizes of the emergency vehicles were factored into the model. The object detection model was trained in the TensorFlow deep learning platform for 10200 iterations to decrease the values of training loss. The subsequent values of training and validation loss obtained were 0.0086 and 0.0029 (figure 2).

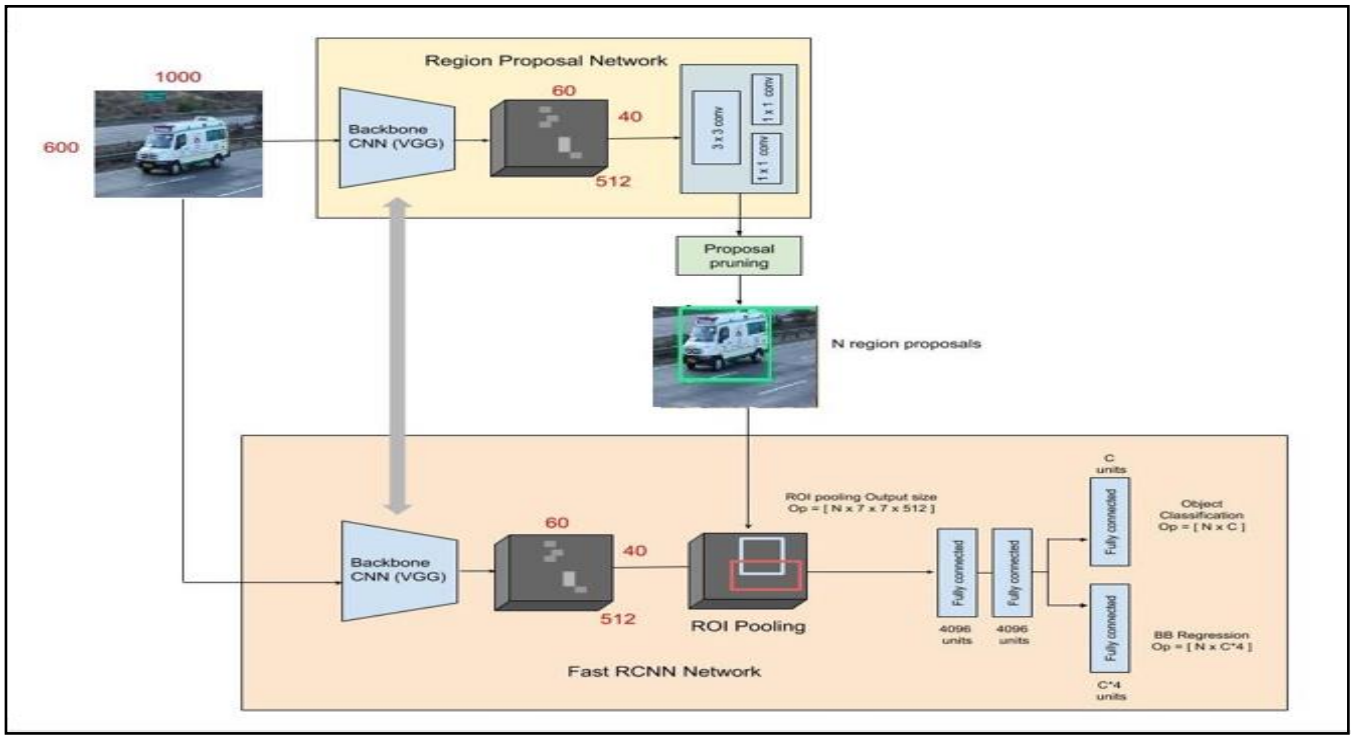

Figure 1- Representation of Faster RCNN

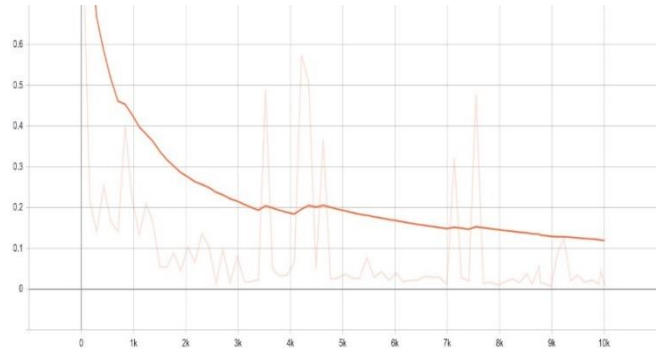

Figure 2 (a)- Training loss vs number of iterations

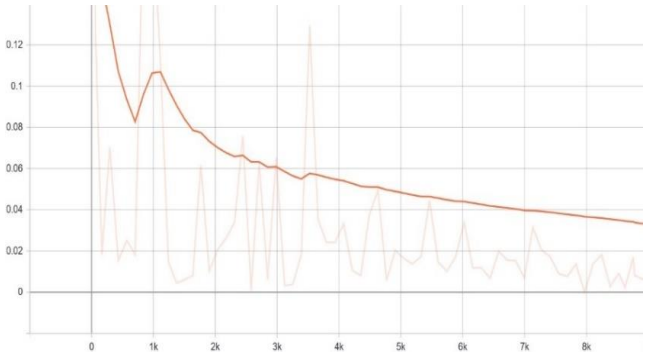

Figure 2(b)-Validation loss vs number of iterations

C. Results

The object detection algorithm recognizes the ambulance even when it is amidst a traffic congestion (figure 3). The output is in the form of a bounding box with detection accuracy in percentage.

Figure 3- Recognition of ambulance in traffic congestion

V. THE INSTANCE SEGMENTATION MODEL

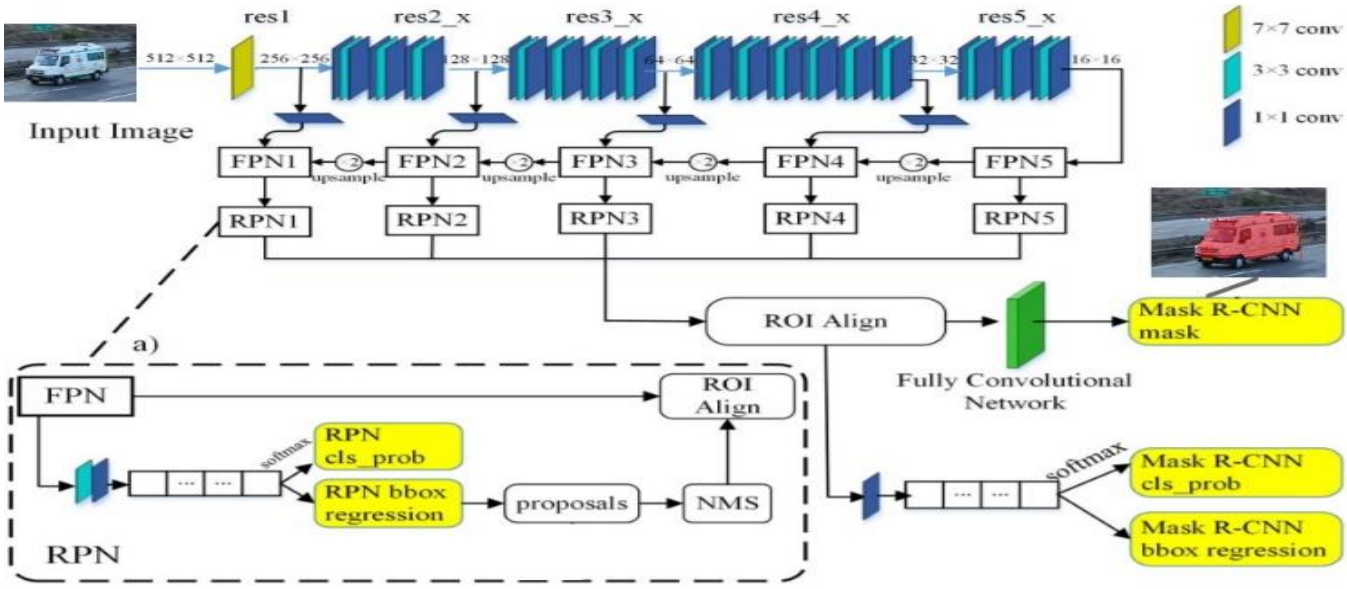

Figure 4- Mask RCNN architecture [6]

A. Principle and Architecture

Instance segmentation is a computer vision technique that detects and characterizes boundaries of a particular object of interest at the pixel level. Instance segmentation techniques primarily employ the use of a flexible framework called Mask Region Based Convolution Neural Network (Mask RCNN) [7]. Mask RCNN is a malleable and state of the art deep neural network which accurately identifies objects in an image, video or a real time feed by enclosing them in a bounding box and concurrently creates a prime segmentation mask for specific instances detected in the feed. It is a multitasking network and surpasses all existing models by blending both bounding box object detection (classification and location) and instance segmentation synchronously.

The Mask R-CNN network (figure 4 [7]) has two principal stages. The first stage anticipates or contemplates the presence of the object in a region of the input image also known as the Region of Interest (RoI). The second stage forecasts the probability, displays Image over Union (IoU) bounding box and the binate mask around the image based on the results of the first stage. Both stages are fused in the backbone. The network has three components namely FPN, RPN and Backbone network architecture. The FPN or Feature Pyramid Network is a top-down or bottom-up architecture and is used as a universal feature extractor. Here the bottom-up architecture is implemented for feature extraction from the feed.

The RPN or Region Proposal Network is a light network that scans the FPN bottom-up and proposes probable regions in the image where the object is likely to be present. It then recognizes various regions by fitting multiple bounding boxes according to certain IoU values. The backbone is a multilayered neural network which obtains feature maps of the input feed. Here ResNet50 is employed as it is not a very deep architecture. Fine tuning helps the model attain higher accuracy with lesser training time.

The ResNet backbone has a rectified linear unit (ReLU) activation, mathematically represented by

$$f(x)=\max(0, x) \tag{1}$$

The Stochastic Gradient Descent algorithm with a mini batch is used to update the weights and momentum for minimizing loss and converging at the most accurate value.

$$\theta = \theta - \eta \cdot \nabla J(\theta; x(i); y(i)) \tag{2}$$

The above equation represents the update rule for stochastic gradient descent where θ represents the parameter matrix of the neural network. η represents the learning rate of the algorithm, ∇ is the gradient calculated for the loss function J over the elements of the feature space x(i) and y(i). As parameter updates are more frequent, the rate of convergence of stochastic gradient descent is quicker than normal batch training.

B. Nature of Dataset and Training-

A custom dataset of 400 images of emergency vehicles was assimilated and annotated using the labelme tool. Masks of respective images were generated post annotation. The model was trained using TensorFlow [9] on 320 images and validated on 80 images resulting in an 80:20 split. This ensured that there was a balance between learning the training data and validation. The

model was first trained for 10 epochs, 1000 steps per epoch for the last layer or heads. The model was then fine-tuned for stage 4 and above (using ResNet50) for 10 more epochs and then it was trained for all stages for 40 more epochs. This resulted in decreasing training loss and validation loss of the model for the specified configurations (figure 6).

C. Results

The algorithm was tested on a dataset of images of emergency vehicles stuck in traffic on Indian roads . The outputs are masks generated on the object which is enclosed in a bounding box with the class name and accuracy (figure 7).

```
Configurations:
BACKBONE                       resnet50
BACKBONE_STRIDES               [4, 8, 16, 32, 64]
BATCH_SIZE                     1
BBOX_STD_DEV                   [0.1 0.1 0.2 0.2]
COMPUTE_BACKBONE_SHAPE         None
DETECTION_MAX_INSTANCES        50
DETECTION_MIN_CONFIDENCE       0.998
DETECTION_NMS_THRESHOLD        0.3
FPN_CLASSIF_FC_LAYERS_SIZE     1024
GPU_COUNT                      1
GRADIENT_CLIP_NORM             5.0
IMAGES_PER_GPU                 1
IMAGE_CHANNEL_COUNT            3
IMAGE_MAX_DIM                  512
IMAGE_META_SIZE                14
IMAGE_MIN_DIM                  448
IMAGE_MIN_SCALE                0
IMAGE_RESIZE_MODE              square
IMAGE_SHAPE                    [512 512   3]
LEARNING_MOMENTUM              0.9
LEARNING_RATE                  0.001
```

Figure 6- Specified configurations

Figure 7- Identification of ambulance in traffic

VI.

VI. ANALYSIS OF USE CASES

The embedded systems in a traffic signal can be programmed to accept an input from the detection unit whenever an emergency vehicle is detected and subsequently switch the signal to green from red. A reliable and robust system that can accurately detect an emergency vehicle and fast track its flow through heavy city traffic is an asset to any Intelligent Transportation System or Smart City venture. Autonomous vehicles can also have built-in emergency vehicle detection capabilities to allow priority movement of ambulances, fire engines etc. In both use cases, it is essential to ensure that there are sufficient computational resources for the execution of the computer vision models.

Both object detection as well as image segmentation differ from conventional image classification in the sense that they identify the location/coordinates of the object under detection. On the other hand, an image classifier would simply assign a particular label to image when the object it is trained to detect is found in the image.

For the intelligent traffic signal application, a conventional image classifier would be ineffective as it is necessary to identify the lane in which the emergency vehicle is present, so as to switch the signal for that particular lane. An object detection model would be an ideal fit for this application.

In the case of an autonomous vehicle, greater precision is required as the vehicle will have to maneuver itself based on the spatial extent of the emergency vehicle. In this case, an instance segmentation model, which traces the emergency vehicle by performing pixel wise classification, would be the best fit. Although the object detection model will generate a bounding box, it will be unable to provide the exact coordinates of the emergency vehicle.

VII. TESTING THE ALGORITHMS

The following table (figure 8) encapsulates the accuracy of the models, tested on a dataset of 100 images.

	True positive	True negative	False positive	False negative
Object Detection	74	7	8	11
Instance Segmentation	83	9	6	2

Figure 8- Tabular representation of accuracy

The cumulative accuracy will be the sum of True positives and True negatives i.e. 81 % for object detection and 92 % for instance segmentation.

VIII. CONCLUSION

Two different algorithms in computer vision, object detection and instance segmentation, have been described for emergency vehicle detection and localization. Both algorithms are effective in identifying an emergency vehicle from a cluster of vehicles, and hence can be leveraged for applications in intelligent transportation systems like deployment of smart traffic signal and autonomous vehicle.

REFERENCES

[1] R. Sundar, S. Hebbar and V. Golla, "Implementing Intelligent Traffic Control System for Congestion Control, Ambulance Clearance, and Stolen Vehicle Detection," in *IEEE Sensors Journal*, vol. 15, no. 2, pp. 1109-1113, Feb. 2015

[2] S. Amir, M. S. Kamal, S. S. Khan and K. M. A. Salam, "PLC based traffic control system with emergency vehicle detection and management," 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kannur, 2017, pp. 1467-1472.

[3] B. Fazenda, Hidajat Atmoko, Fengshou Gu, Luyang Guan and A. Ball, "Acoustic based safety emergency vehicle detection for intelligent transport systems," 2009 ICCAS-SICE, Fukuoka, 2009, pp. 4250-4255.

[4] S. Roy and M. S. Rahman, "Emergency Vehicle Detection on Heavy Traffic Road from CCTV Footage Using Deep Convolutional Neural Network," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox's Bazar, Bangladesh, 2019, pp. 1-6.

[5] A. Raman, S. Kaushik, K. V. S. R. Rao and M. Moharir, "A Hybrid Framework for Expediting Emergency Vehicle Movement on Indian Roads," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 2020, pp. 459-464, doi: 10.1109/ICIMIA48430.2020.9074933.

[6] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031

[7] L. Cai, T. Long, Y. Dai and Y. Huang, "Mask R-CNN-Based Detection and Segmentation for Pulmonary Nodule 3D Visualization Diagnosis," in IEEE Access, vol. 8, pp. 44400-44409, 2020.

[8] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778.

[9] Waleed Abdulla, "Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow." https://github.com/matterport/Mask_RCNN. (2017).