# Real Time Vehicle Detection and Colour Recognition using tuned Features of Faster-RCNN

Abdullah Tariq
*Intelligent Criminology Research Lab*
*NCAI, Kics, Uet Lahore*
Lahore, Pakistan
abdullah.tariq@kics.edu.pk

Muhammad Zeeshan Khan
*Intelligent Criminology Research Lab*
*NCAI, Kics, Uet Lahore*
Lahore, Pakistan
zeeshan.khan@kics.edu.pk

Muhammad Usman Ghani Khan
*Intelligent Criminology Research Lab*
*NCAI, Kics, Uet Lahore*
Lahore, Pakistan
usman.ghani@kics.edu.pk

2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA) | 978-1-6654-1511-8/20/$31.00 ©2021 IEEE | DOI: 10.1109/CAIDA51941.2021.9425106

***Abstract*—Being the most dominant part of the vehicle, colour anticipate vital role in vehicle identification. Thus, colour also plays significant part in Intelligent Transportation System (ITS) and can be very effective in various applications of ITS. In past, most of the work had done on colour recognition of vehicle are not able to achieve the high accuracy because they rely on hand-crafted feature i.e. Speeded Up Robust Features (SURF), Scale Invariant Feature Transform (SIFT) and Histogram of Oriented Gradient (HOG). In this work, we proposed a solution by utilizing one of the latest deep learning algorithm for the detection of vehicle and the classification of detected vehicles colour. Proposed methodology is based on the tuned features of Faster R-CNN and achieved the good results as compared to current state of the art techniques. In addition to that, this work is also contributes towards the dataset collection of related vehicles being used in Pakistan. Proposed method outperformed the previous works by achieving 95.31% accuracy on testing data. The robust results in terms of accuracy and the generation of dataset depicts the novelty of proposed technique in the literature.**

***Index Terms*—Faster R-CNN, Color Recognition, Vehicle Detection, ITS**

## I. Introduction

Advancement in the field of technology also demands the more efficient and accurate systems for security measures in ITS (Intelligent Transportation System). Aim of Intelligent Transportation System (ITS) is to reduce traffic problems. Instant information about traffic is being provided to user with the help of ITS. Intelligent Transportation System also enhances the safety and increase the security factor of commuters [1]. Being the vital part of ITS, vehicles information extraction is one of the most important area that gains huge attention in recent decade. Colour is one of the most important attribute of vehicles. Colour of the vehicle plays a significance role in video surveillance, ITS, and in smart city. Now a days, in order to detect the criminal activity, it is important to recognize the features of vehicles along with to detect the suspicious objects carried on vehicle. One can face difficulties in recognizing them because of the variations in vehicles colour. Moreover, number plate of the vehicle is also not fully detectable because of partial occlusion due to colour of the vehicle. Approximately 80% of the vehicle area is covered by the paint.

978-1-6654-1511-8/21/$31.00 2021 IEEE

So, automatic detection of vehicle and colour classification plays a important role in suspicious activity detection, law enforcement agencies, video monitoring and for security measures in ITS [1]. Colour of vehicle has been used as a valuable signal for extracting useful information from the vehicle. But, to identify the vehicle colour is difficult task in an uncontrolled environment. The reason behind the difficulty is vehicles colour is prone to be contrived due to different factors such as snow, sun shine and rain. Many techniques have been proposed to overcome these challenges like Normalized RGB Histogram, Feature Context etc. These methods no doubt have excellent performance but can not be able to produce satisfactory results due to some limitations such as at certain angle, frontier pose, illumination. In past, proposed methods were not able to produced desired result as they consider hand crafted features. Our purposed model recognize and classify the vehicle's colour by using deep learning based algorithms. We have fine-tuned the different parameters of Faster RCNN to classify and recognize colour of vehicles. We have also prepared the dataset named as Datasets of Vehicles used in Pakistan. We performed pre-processing and annotation on the collected dataset. There are 5 classes in our dataset such as White, Black, Grey, Red and blue colours of vehicles. Most of the work in the literature till now is failed to distinguished between different colors of the vehicle due to the haze and illumination problems. But our proposed work successfully tackles this problem. Our proposed system is also capable of generating statistical report based on the colour of vehicles enters at the particular premises like highly secured areas i.e. cantonment board, universities, railway stations , airport etc. on the defined time stamp. The proposed system have been shown in Fig.1. We have make the following contributions.

- Utilized the Faster-RCNN for vehicle detection and colour classification.
- Fine-tuned the different parameters of Faster-RCNN, so algorithm performs good on the locally generated dataset.
- Generation and the annotation of the dataset of vehicles which are mostly used in Pakistan.

## II. Related Work

Vehicle colour recognition and classification is the most actively studied domain from the past few years. Different image

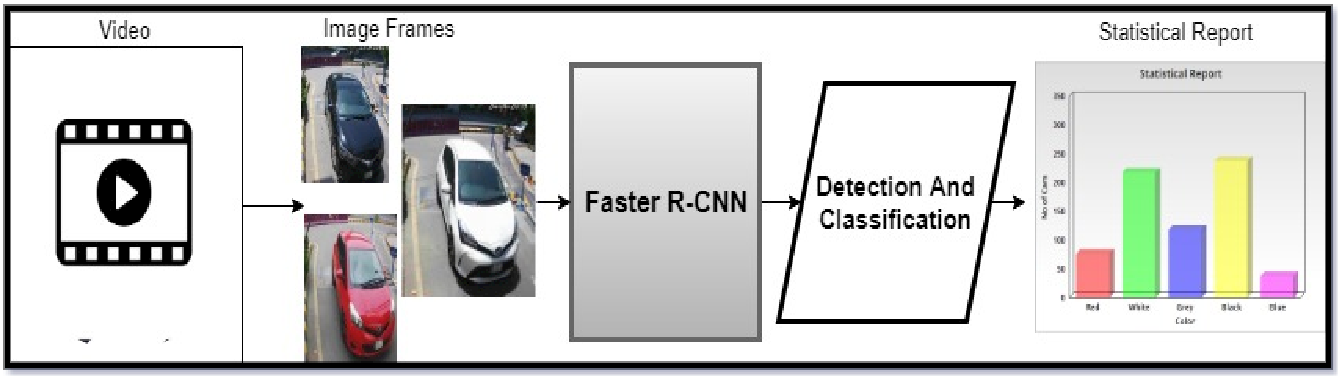


Fig. 1. System Diagram

processing and computer vision algorithms like SURF [2], SIFT [3], HOG [4] have been applied to tackle this problem. N. Baek et al. [5] designed the solution for classification of vehicles colour. In their proposed method, for the elimination of distortions due to intensity changes they adopted the way of converting vehicle image into an HSV (hue-saturation-value). Then, they construct the feature vector, which is a two-dimensional histogram for the hue and saturation pairs. By using SVM, feature vectors classify into different five vehicle colors: yellow, blue, red, black and white.

They have their own dataset of vehicles containing 100 images of each input class, their result shows the success rate of 94.92% on 500 unseen vehicles images. Whereas, Pan Chen et al. [6] proposed a solution for color recognition of Vehicles on urban roads by extracting feature context. This method depends on the Region of Interest (ROI). To overcome the image quality degradation, they performed preprocessing. A classifier is trained to extract color from images just after the Selection of ROI with different weights. They used SVM as a classifier. They collected the datasets of the vehicles on urban roads. They also used videos of urban road traffic as their dataset.

Their experiments have satisfactory results of an average accuracy of 90.6% without preprocessing, 92.2% with preprocessing and 92.4% after made all the improvements.Another solution based on ROI is proposed by E. Dule et al. [7] for vehicle color recognition. They examined two ROI (semi front vehicle and smooth hood peace), they used three method for classification i.e. SVM, K-Nearest Neighbors and Artificial neural network, as a feature sets, they used all of the 16 possible color combination. They classify the vehicle images into different seven colors: yellow, red, black, gray, blue, green, and white. They secured 83.50% accuracy in their experiments. From the recent few years, the field of computer vision is diverted from analytical methods to deep learning neural network methods. The reason behind the diversion of the research community towards the deep learning is the high achievable accuracy on the vision-related task and the availability of the large-scale dataset and computation power. Hu C et al. [8] design a solution for recognizing Vehicle colour in natural scenes.

They designed the model based on a deep learning algorithm to detect Vehicle colour automatically. In this, they use the Convolution Neural Network and combine the CNN with the globally use spatial pyramid technique. Their experiments have better accuracy as compare to other conventional approaches. As a classifier, they used a Support Vector Machine (SVM). And then the output from the last three layers is used as features for recognizing the colour. For the experiment purpose, they used the publicly available dataset that released in [6]. They achieved 93.7% accuracy on the validation data. The method proposed by Zhang q et al. [9] for the solution of color recognition of vehicles used the lightweight Convolution Neural Network. For recognition, they first used the lightweight convolutional network and designed it in multi layers Which have five layers, a fully connected layer, three convolutions layers and a global pooling layer. Feature map is divided by using SPM "Spatial Pyramid Matching" and then every SPM region is used to a vector of feature representation. They used the dataset of Pan Chen et al. [6], this dataset consists of 15,601 images of different vehicle colour such as red, yellow, gray, white, green, blue, and black. In their dataset, vehicle images captured from front. They achieved the 94.7% accuracy on the validation data.

Rachmadi R.F et al. [10] proposed another solution for vehicle colour recognition based on deep learning technique using the convolution neural network (CNN). It is considered that; CNN classifies objects by utilizing shape information. But they showed that CNN can classify objects on the basis of colour distribution. In their methodology, input image is passed to run on some CNN architecture after conversion of image into two different colour spaces i.e. CIE lab and HSV. Their dataset consists of 15601 vehicle images, which are of different 8 classes i.e. gray, yellow, cyan, blue, green, black, white, and red. Their model successfully capturing the accuracy of 94.47%. As per studies of the related work, most of the researchers utilized the two different techniques for detection of the vehicle and classification of its different features. But in our proposed system algorithm detects and classify the vehicle within the same network stream and architecture. Moreover,

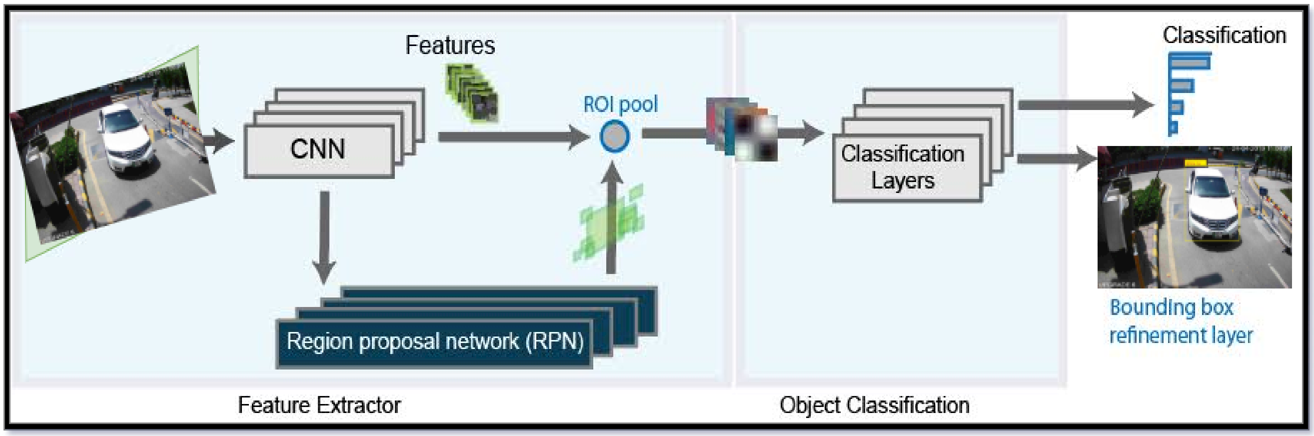


Fig. 2. Network Architecture

we have also fine tuned and add some convolution parameters to achieve some good results locally generated dataset.

## III. Methodology

Videos of the moving vehicles is passed to our proposed system as an input at the rate of 25 frames per second. Bounding box specify the Vehicle in input video. Then the proposed model detect the vehicle and classify its colour. In addition to this, statistical report is also generated that infer, which colour of vehicles are being detected in the specific amount of time.

### A. Pre Processing

Images/videos captured by the cameras where traffic is dense are not balanced. As, the images/videos may contain the unwanted objects in it. Moreover, there is possibility that the quality of the images reduces because of some environmental and weather effects. These are the hurdles that may led to difficulties in color recognition. To tackle these influences we eliminated the extra objects from the images and just keep the desired object with the background. Furthermore, methods [11] and [12] are also the parts of preprocessing in our network. After the removal of disturbing objects, the image is more balanced. Additionally, to improve color quality of the image, color contrast technique based on the biased correction and statistical estimation is used [13]. At the end, we have more prominent and accurate image of vehicle.

### B. Proposed Architecture

For the detection and classification of objects, proposed methodology is encouraged by the method of ren et al. [14]. Proposed network consist of two fragments; first one is for detecting the vehicle by using regional proposal and the second one is used for classifying the colour of vehicle. The technique of initial Region Convolution Neural Network had flaws in terms of time-efficiency and computation [15]. Previous methods by ren et al. [14] detect the boundaries of object in image via selective search method. However, the time-efficiency of Region Proposed Network (RPN) is more than the selective search, reason is that it elapse most of the time with detected object classification network. Flow of our proposed architecture is shown in Fig.2.

### C. Vehicle Detection

Region Proposal Network (RPN) detect the vehicle by drawing anchors and give the rectangular box as an output which most probably has vehicle. RPN can handle any size of input image. Anchor is basically a box drawn on the over all image as an output by RPN. Each anchor have objectness score. After scanning the images with the help of anchors, RPN returns the rectangular boxes in which the probability of having the object is maximum. Furthermore for checking of identical boxes in image intersection over union is used.

$$IOU = \frac{OverlapBox}{UnionBox} \tag{1}$$

Loss of RPNs are calculated by the binary labels assigned to anchor drawn on the image. The intersection over union value depicts how much the predicted bounding box is matched with the ground truth values. We have set the 0.7 value for the ROI containing the object with respect to the ground truth value. So, If the value of intersection over union is more than 0.7 then the anchor have positive value showing the object presence. On the other hand, if the value assigned to anchor is less than 0.3 then the anchor have negative value, depicting no object in that particular ROI. Those anchors does not take part in training whose value is not in the range of positive or negative number.

## IV. Backbone Architecture

In our proposed algorithm, vehicle detection and classification architecture based on the three parts. The first part is the backbone architecture which is the basically convolution neural network architecture. After extracting the features from the convolution neural network architecture, features map are further passed to the Region Proposal Network. The region proposal network proposed the regions where the probability

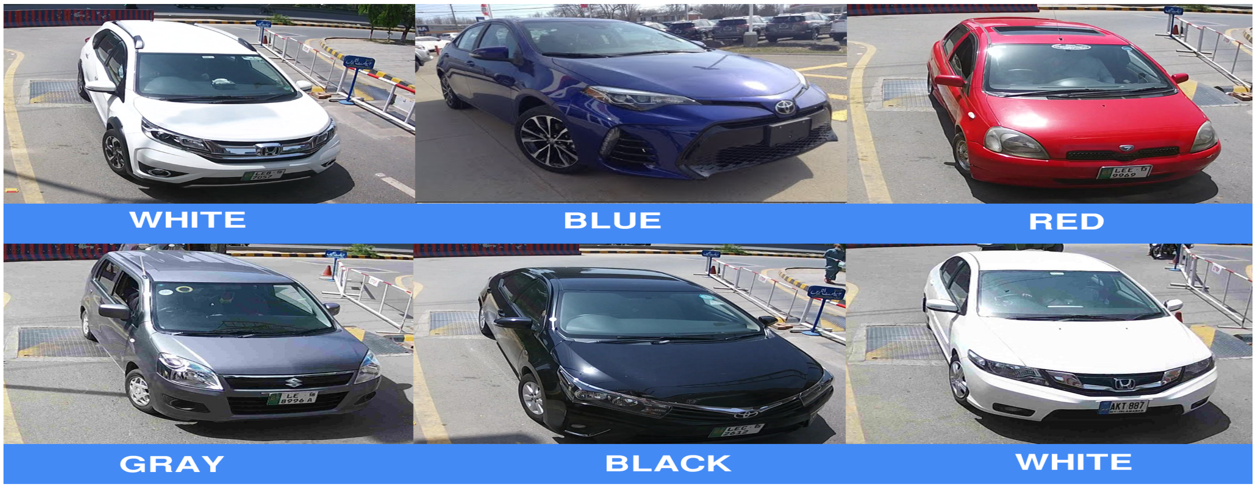


Fig. 3. Sample frames from our generated dataset

of the specific objects founds to be maximum. The regions proposed by the Region Proposal Network are further passed to the classification network. This classification network contains the fully connected layer which classify the detected objects as well as the Region of Interest where the specify object is present in the frame. Our backbone architecture contains the 5 convolution and 2 fully connected layers. The input image of our proposed architecture have the image size of 227x227 with three channels. Every convolution layer and fully connected layer is followed by an activation function Relu. 1st convolution layer has 96 filters having the size of 55x55. After that max pooling is applied the size of filters is now reduced to 27x27. Similarly, in 2nd convolution layer filter size of 27x27 is applied with 256 filters and then after max pooling, filter size is reduced to 13x13 with same numbers of filters. In third convolution layer output of 2nd convolution is passed with the filter size of 13x13 having 384 filters. In 4th convolution layer the size of image remains same with same number of filter before and after the max pooling but in 5th convolution layer, size of the image reduces to 6x6 with 256 numbers of filters. For preventing our model from overfitting, dropout rate is set which change randomly. After that we have two fully connected layer, 1st fully connected layer flatten the features in 4096 that receives from last convolution layer. After this, in 2nd fully connected layer we mapped these features to the number of classes then the Softmax function classify the class with highest probability.

## V. Dataset

The efficiency and accuracy of deep-learning based approaches are primarily depend on the dataset and computation power of resources [16]. These methods require the precise dataset (balanced and without redundancy) in accurate format (images ,video ,sound). The vehicle dataset available for this [6] is not according to our requirements as the proposed network is designed for Pakistan local community. Hence, it is highly needed that we collect the dataset of our own, for the training of the proposed network. For the collection of dataset we use camera stream having the specification of 5 Mega-Pixel at the main entrance in one of the educational institute. We exploited several 7 minutes chunks of 2 days camera stream; comprises of 10 frames per second. After collection of the dataset we perform preprocessing (skip the images which have more than one vehicle in it ) on the images extracted from camera stream and then do annotation conferring to the classes in network. Dataset consists of five classes: red; black; grey; white; blue. We have images from different views i.e. frontal view, both sides view, arial view and back view, There are 500 images in each of the above mentioned class. To summarize, we have a total of 2500 images with 15 camera streams of 10 minutes videos (average length) having quality $1920 \times 1020$. For the productivity of research community the generated dataset will be available freely on world-wide-web. Sample of dataset images shown in Fig.3.

## VI. Training Parameters and Results

Our proposed architecture is implemented by utilizing the deep learning framework i.e. TensorFlow. NIVIDIA 1080 Ti of 11 GB memory is used for the training of our model. System took about 18 hours in completion of our proposed Faster R-CNN network for 2,00000 epochs. 80% of our data is used in training and 20% utilized in evaluation of learned model as shown in table I. Furthermore, for the calculation of the loss

| Class | Training Images | Test Images |
|---|---|---|
| Grey | 400 | 100 |
| Red | 400 | 100 |
| Black | 400 | 100 |
| Blue | 400 | 100 |
| White | 400 | 100 |

TABLE I: No of training and testing images

of our proposed model we used the most frequently function i.e. Mean Square Error (MSE). loss comparison on different epochs shown in the Fig.4. MSE compute the difference of square of each data point (between predicted variables and actual variables) and provide the sum after dividing by total number of data points. Once the model have been trained, it gives 15 frame per second on the inference. Learning rate of our model is dynamic. As, for first thousand epochs learning rate is $(10)^{-}1$ and it decreases to 10% i.e. $(10)^{-}2$ after each 20 thousands epochs. Stochastic Gradient Descendant (SGD) is being used for the optimization of weights. Moreover, to prevent our model for being over fit, early stop function is applied. This function works by stopping the network process at the stage where model head towards over fitting. Over fitting of the model (neural network) happens because of the weak control over the learning process of model. If the model gets over fit, it is difficult for the model to perform well on unseen data.

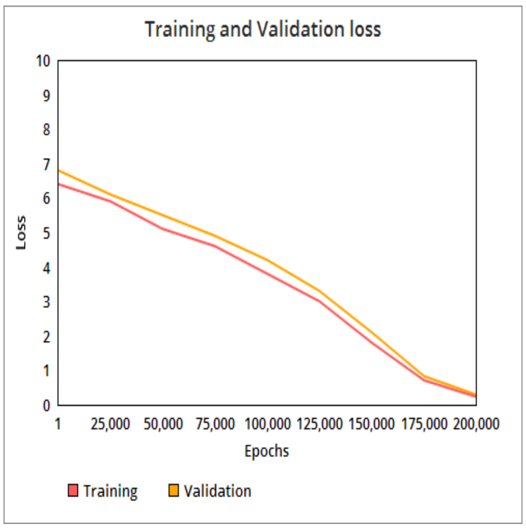


Fig. 4. Loss copmarison on epochs

$$MSE=\frac{\sum_{i=1}^{k}(P_i - P_i^l)^2}{k} \quad (2)$$

Where k is the total number of available data points, $P_i$ is

| Paper | Methodology | Accuracy | Dataset |
|---|---|---|---|
| Own | Faster R-CNN | 95.31% | Self-Generated |
| [1]et al. | Lightweight CNN | 94.73% | Pan Chen[2] |
| [3]et al. | Spatial Pyramid | 93.78% | Pan Chen[2] |
| [2]et al. | Extracting Feature Context | 92.49% | Self-Generated |

TABLE II: Comparison with other methods

the actual label and $P_i^l$ is the predicted label. Our proposed Faster R-CNN model achieved the total accuracy of 95.31% on validation data. Accuracy on validation and on training data is shown in Fig.4. Our architecture outperform the Qiang Zhang et.al [9] model, Chuanping Hu et al. [8] model and Pan Chen et al. [6] model in terms of accuracy. Comparison by above mentioned models with our architecture is given below in Table II. It is clearly seen that our model achieve high accuracy as compare to the light-weight CNN network by Qiang Zhang et.al [9], Spatial Pyramid technique which combine with CNN by Chuanping Hu et al. [8] and ROI based extracting feature context by Pan Chen et al. [6].

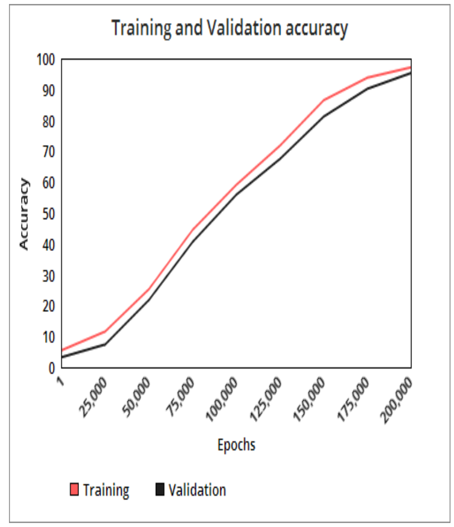


Fig. 5. Training and Validation accuracy

## VII. CONCLUSION

In this paper, we proposed an effective method for vehicle color classification in videos and images. This proposed architecture is based on Faster R-CNN. In comparison with previous methods for vehicle color recognition. From Fig.5 experimental results of our proposed methodology clearly shows the effectiveness of our model that it is highly efficient and perform well in the cloudy weather. Our future works are to make our model efficient in the low light environment i.e. at night. As, in the low light the prediction of blue and black vehicle is not as much accurate and make it able to classify other attributes of vehicle.

## ACKNOWLEDGMENT

We would like to express our earnest gratitude to Intelligent Criminiolgy Lab, National Center For Artificial Intelligence, AlKhawarzimi Institute of CS, UET Lahore for their support, dedication, technical sessions and knowledge sharing effort.

## REFERENCES


[1] Khan MZ, Khan MU, Irshad O, Iqbal R. Deep Learning and Blockchain Fusion for Detecting Driver's Behavior in Smart Vehicles. Internet Technology Letters.:e119.

[2] Hsieh JW, Chen LC, Chen DY. Symmetrical SURF and its applications to vehicle detection and vehicle make and model recognition. IEEE Transactions on intelligent transportation systems. 2014 Feb;15(1):6-20.

[3] Psyllos AP, Anagnostopoulos CN, Kayafas E. Vehicle logo recognition using a sift-based enhanced matching scheme. IEEE transactions on intelligent transportation systems. 2010 Jun;11(2):322-8.

[4] Wei Y, Tian Q, Guo J, Huang W, Cao J. Multi-vehicle detection algorithm through combining Harr and HOG features. Mathematics and Computers in Simulation. 2019 Jan 1;155:130-45.

[5] N. Baek , S.M. Park , K.J. Kim , et al. , Vehicle color classification based on the sup- port vector machine method, in: Advanced Intelligent Computing Theories and Applications. With Aspects of Contemporary Intelligent Computing Techniques, 2007, pp. 11331139 .

[6] P. Chen, X. Bai, and W. Liu, Vehicle color recognition on urban road by feature context, IEEE Trans. Intell. Transp. Syst., vol. 15, no. 5,pp. 23402346, Oct. 2014.

[7] Dule E, Gokmen M, Beratoglu MS. A convenient feature vector construction for vehicle color recognition. InProc. Int. Conf. Neural Netw. WSEAS 2010 Jun 13 (pp. 250-255).

[8] Hu C, Bai X, Qi L, Chen P, Xue G, Mei L. Vehicle color recognition with spatial pyramid deep learning. IEEE Transactions on Intelligent Transportation Systems. 2015 Oct;16(5):2925-34.

[9] Zhang Q, Zhuo L, Li J, Zhang J, Zhang H, Li X. Vehicle color recognition using Multiple-Layer Feature Representations of lightweight convolutional neural network. Signal Processing. 2018 Jun 1;147:146-53.

[10] 7. Rachmadi R.F., Purnama I. Vehicle color recognition using convolutional neural network[ J ]. arXiv: 1510.07391 , 2015.

[11] K. He, J. Sun, and X. Tang, Single image haze removal using dark channel prior, in Proc. IEEE Conf. CVPR, 2009, pp. 19561963.

[12] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, Digital Image Processing using MATLAB. Knoxville, TN, USA: Gatesmark Publishing, 2009.

[13] Finlayson GD. Colour and illumination in computer vision. Interface focus. 2018 Aug 6;8(4):20180008.

[14] S.Ren,K.He,R.Girshick,andJ.Sun,FasterR-CNN:Towardsreal-time object detection with region proposal networks, in Proc. Adv. Neural Inf. Process. Syst., 2015, pp. 9199

[15] R. Girshick, Fast R-CNN, in Proc. IEEE Int. Conf. Comput. Vis., 2015, pp. 14401448.

[16] Khan MZ, Harous S, Khan MU, Iqbal R, Mumtaz S. Deep Unified Model For Face Recognition based on Convolution Neural Network and Edge Computing. IEEE Access. 2019 May 23.