

The role of Convolutional Neural Network in Vehicle Detection on Spatial - Temporal Road Traffic data

S. Senthilarasi, Research Scholar,
Department of Computer Science,
School of Computing Sciences
Vels Institute of Science, Technology
& Advanced Studies (VISTAS)
Pallavaram, Chennai, India
senarasi80@gmail.com

Dr. S. Kamalakkannan, Associate Professor,
Department of Information Technology,
School of Computing Sciences
Vels Institute of Science, Technology
& Advanced Studies (VISTAS)
Pallavaram, Chennai, India
kannan.scs@velsuniv.ac.in

Abstract. In places of urbanization, people who reside in several places of a city utilizes different types of transportation according to their needs. On regular working days, People travelling from one place to another at peak hours is a difficult task. This induces on-road traffic which turns traffic congestion—low, moderate and heavy. Traffic congestion is still a challenging issue to be solved in different methodologies. This paper is discussed with a few viewpoints applicable at present to measure congestion, Intelligent Transportation System (ITS) in Chennai. Since our study is focused on road traffic congestion, a convolutional neural network is used to detect the vehicles in road traffic by means of a trained neural network to recognize vehicles such as cars, two-wheelers, buses, trucks, vans, auto, etc. To display the road traffic images in different forms, computer vision technology is used. In addition, the Chennai road traffic data set is applied in a grouped bar graph to display the number of persons killed and injured due to traffic congestion from 2014 to 2021.

Keywords: Traffic congestion measures, ITS, CNN, Haar Cascade Classifier, Used Technology

1. Introduction

In recent decades, the count of vehicles moving on the road has been numerously increased for which the requirement of a safe and well-directed transportation system has been created. At this period, Intelligent Transportation System (ITS) has appeared as a systematic way to refine traffic flow, its control and safeness of commoners and vehicle drivers. Intelligent Transportation in the model of autonomous vehicles is trading possible for the welfare of the people, is capable of accomplishing individualistic tasks such as object and detection[1].

The collection of data computed by vehicles in ITS is used to train models in deep learning. For example, A model named CNN is used to plan, control and to enforce the implementation of laws in ITS. An establishment of deep learning in ITS is capable of transforming the road transportation application by saving time of travel, end-users are informed about traffic jams, managing plan of a journey, information of traffic signs time to time in order to avoid accidents and to decrease traffic congestion and pollution[1].

1.1 Present viewpoints to measure congestion

On the basis of different performance criteria, the congestion measures can be classified into the following groups:[2]

1. Speed
2. Travel Time
3. Delay
4. Level of Services
5. Congestion indices

Table 1: Categorization of Congestion measures [2]

Category	Estimation approach	Congestion range	Pros	Cons
lane	Speed Reduction Index (SRI)	>4	Supplies information easier about	It does not contemplate nonrecurring

Speed	Speed Performance Index(SPI)	Levels of different ranges	the related speed of vehicles both in a normal and congested environment	conditions	S.NO.
Time of Travel	Travel amount	NIL	Accounted for both time and space	Not added in capacity	
Delay	Rate of delay	NIL	Flexible to calculate system performance and selects well-planned travel method	Restricted for a particular road type	
	Delay Ratio	NIL	Congestion levels are composed according to different types of roads	Congestion range is not applicable	
Level of Services	Volume to capacity	Levels of different ranges	Understandable by non-technical users	Cannot able to issue continual congestion values	
Congestion indices	Relative Congestion Index (RCI)	>2	Representation of traffic – Spatial-mean performance	Road type, in particular, is limited	
	Congestion Index-Road segment	NIL	Suitable to constitute segment condition	It is applicable only to evaluate the particular segment conditions	

1.2 Chennai – ITS (Intelligent Transportation System)

Table 2: ITS components and its objectives [3]

S.NO.	COMPONENTS OF ITS	DESCRIPTION
1	Traffic Control Room	It is a control room in miniature for management of traffic, developed and handled by Chennai traffic police – monitors the condition of traffic by CCTV.
2	Variable Message Sign(VMS) Board	Static messages are provided to raise traffic awareness like wearing a helmet. In a city 31 VMS boards are in action.
3	Traffic Signals	The signal cycles in a fixed state are set manually. In a city 383, signalized intersections are available.
4	CCTV cameras	The core areas intersections are provided with CCTV, handled by Chennai traffic Police to monitor the site condition.
5	E- Challan System	It is an implementation system for contravened vehicles.
6	Highway Traffic Management System (HTMS)	Systems such as Emergency Call Box, Weather monitoring are initiated by concessioner on Outer Ring Road (ORR).
7	Toll Management System	An introduction of RFID established TMS on the BOT division of NH45 and Chennai Bypass.
8	Accident database	It has been built with the support of the World Bank fund. It is utilized by the Transport department.
9	Road Management System	An establishment and usage of National and State highways by road administrators.
10	Vehicle database	Management of vehicle information is evolved in database underneath the plan of e-governance-India.
11	R&D activities on ITS	IIT Madras carried out various R&D activities like count of traffic, pre-trip information, etc.

2. Specialization in the classification of images – Convolutional Neural Network

Convolutional Neural Network is one of the important categories in Deep Neural Networks, which can identify and categorize specific features from images and are broadly used to examine visual images. Their applications vary from image, video recognition, categorization of images, analysis of medicinal images, Computer Vision (CV) and NLP (Natural Language Processing). The word ‘convolution’ in CNN designates the mathematical based functions of convolution, further, it is a special kind of linear operation. The third function(outcome of images) is obtained by the two fully connected layers before the process of softmax classifier to compute final output probabilities for each class. Since the CNN plays a major role in identifying and classifying images it has been chosen in our road traffic video to detect the vehicles in metropolitan areas like Chennai with the application of Haar-Cascade Classifier.

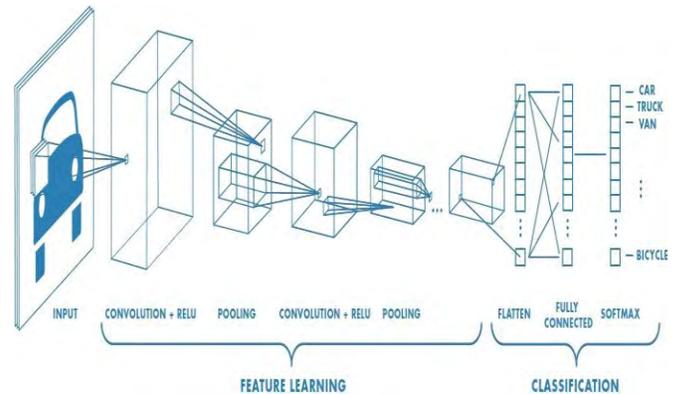


Figure 1: Basic function of a Convolutional Neural Network

2.1 Primary Architecture

The two main parts of CNN architecture are listed below:

1. A convolution-based tool divides and finds the different characteristics of the image for

study in a process with the use of deep learning algorithm to extract the images and classify them into its category or label. This process is known as 'Feature extraction'.

2. A layer termed 'Fully Connected' uses the outcome from the process of convolution and tends to predict the image's class based on the characteristics retrieved in previous stages.

2.2. Layers of Convolution

The layers that build the CNN are listed below:

- Convolutional layers
 - Pooling layers
 - Fully connected layers
- Parameters:
- Dropout layer
 - Activation function

2.2.1 Convolutional layer

The foremost part of a CNN is the convolutive part. It acts as a features separator of images. In this segment, a picture is travelled via a succession of filters otherwise convolution kernels generate current pictures called convolution maps. Some intermediary filters decrease the interaction of the pictures through a local maximum operation. Since the CNN is best in detecting the images from moving videos of any domain, it has been chosen for identifying the vehicles movement on road transport. The concept of CNN with its added supportive features are displayed sequentially. In a CNN, every layer consists two types of parameters such as weights and biases to the input images accordingly.

2.2.1.1 Input volume

A convlayer accepts a volume of size as:

- W1 – Width of neurons
- H1 - Height of neurons
- D1 – Depth of neurons

The outcomes of neurons are obtained in these CNN layers by using the width, height and depth of neurons derived from images, which are further computed by applying the product with their weights, local region in connection to input volume.

The number of features are computed initially with the full size of an input image to estimate the size of the every convolutional layer.

2.2.1.2 Output Volume

It is also called convolution maps, where:

- W2 – Width of neurons
- H2 – Height of neurons
- D2 – Depth of neurons

$$W_2 = \frac{W_1 - F + 2 \times P}{S} + 1 \quad (1)$$

$$H_2 = \frac{H_1 - F + 2 \times P}{S} + 1 \quad (2)$$

$$D_2 = K \quad (3)$$

Where as,

- F → Filter's spatial extension
- K → Total count of filters
- P → Zero padding
- S → Stride[4]

The output layer is computed by using the computation of input images in convolutional layers.

2.2.2 Pooling layer:

In most of the instances, a convolution layer is accompanied by a pooling layer. The primary objective of this layer is to minimize the measurements of the convolved attributes map to reduce the cost of computation. This is dealt with by reducing the connections among layers and operating independently on each feature map. The average numbers of elements are computed in a preplanned sized Image segment known as average pooling. The pooling layer actually serves as a link in the middle of the convolution layer and the FC layer[5].

The insertion in between successive convlayers by pool layer, appealing a downsampling operation with the structural dimensions such as width and height. Pool layer constructs a volume [W2 X H2 X D2] whereas W2, H2 and D2 are applied in the following equations:

$$W_2 = \frac{W_1 - F}{S} + 1 \quad (4)$$

$$H_2 = \frac{H_1 - F}{S} + 1 \quad (5)$$

$$D_2 = D_1 \quad (6)$$

In the end, a feature separator vector (or) CNN combines the information of output as a unique vector[4].

2.2.3 Fully Connected Layer

The FC layer contains the weights and biases through the neurons and is used in associating the neurons in the middle of two different layers. These layers are generally placed prior to the output layers that build the final CNN architecture with less number of layers. In this layered concept, the previous layer's input images are flattened and placed in the FC layer. Later, the flattened vector undertakes a few more Fully Connected layers, whereas the mathematical functions applies the multiplication to the input of weight matrix ,then it adds to a bias vector for execution[5].

Such fully connected layers are also known as 'dense layers' due to each input is linked to output with the use of learnable weight. In addition, each fully connected layer is accompanied by an activation function[4].

2.3 Detecting Vehicles

In general, the AlexNet contains the inputs images of 227*227*3 colours which are trained along with five layers—CNN architecture added with three layers of Fully Connected Layers. After the addition of all the inputs along with weights fed into the convolution's layers, the role of RELU's activation function is initially applied to decrease the computation size of images such as:

$$F(y) = \max(0,y)$$

Where y is an image.

The pooling layer is also utilised in AlexNet to rescale the smaller data[6].

The structure of networks is modified for experiments and the separated vehicle area is shown in outputs[7].

The Convolutional Neural Network has been trained with 1000 images by using its defined layers to detect the vehicles movement on land transport. The images of vehicles has been predicted using TensorFlow applied in Keras. The classification outputs shows the maximum accuracy obtained in 10 number of Epochs compared to other iterations count due to flexibility.

```

Model: "sequential"
-----
Layer (type)                Output Shape              Param #
-----
conv2d (Conv2D)              (None, 30, 30, 32)       896
max_pooling2d (MaxPooling2D) (None, 15, 15, 32)       0
conv2d_1 (Conv2D)            (None, 13, 13, 64)       18496
max_pooling2d_1 (MaxPooling2 (None, 6, 6, 64)         0
conv2d_2 (Conv2D)            (None, 4, 4, 64)         36928
Flatten (Flatten)           (None, 1024)             0
dense (Dense)                (None, 64)               65600
dense_1 (Dense)              (None, 10)               650
-----
Total params: 122,570
Trainable params: 122,570
Non-trainable params: 0
    
```

Figure 2: Application of Convolution neural network

Table 3: Classification Outputs

Epochs	Iterations count	Training time(s)	Loss in %	Accuracy in %
5 Epochs	05	65	4.2231	0.6714
10 Epochs	10	66	0.8940	0.7038
15 Epochs	15	67	4.2975	0.6783
20 Epochs	20	68	1.8115	0.6907

The table3 is obtained from trained neural network of images, detects the on road vehicles with the use of trained parameters applicable in Convolutional Neural Network. Classification outcomes were achieved for the various number of epochs. The maximum accuracy is obtained at the count of 10 Epochs compared with the other number of Epochs[4]. Also, the loss function compares the target and predicted outcomes to estimate the CNN models the training data of images , where minimum loss is obtained in 10 Epochs compared to other Epochs.

The following graph is obtained by applying the maximum possibility of vehicles images into CNN to identify the images and get it trained to find the accuracy.

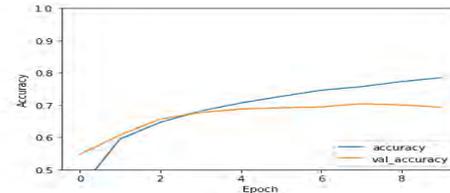


Figure 3: Accuracy rate of trained neural network

3 Vehicles Classification

The vehicles area that exists in the foreground is relinquished by the vehicle detection algorithm is initiated into an alternative Convolutional Neural Network for vehicle recognition. CNN utilizes definite numerals of convolutional.

With the usage of deep learning methodologies, the attributes are avoided by the traditional methods can be extricated, the accuracy of vehicle recognition can be expanded clearly and various convolution kernels are employed to acquire convolution feature pictures as shown in output[7].

3.1. Object detection using Haar-Cascade Classifier

It is the most familiar method used to in-favour-of detecting objects, otherwise known as the Viola-Jones method. In order to get the objects detection, the method has the following points:

- Haar-like feature

- Cascade Classifier
- Integral Images
- AdaBoost learning

3.1.1. Methodology: Block Diagram

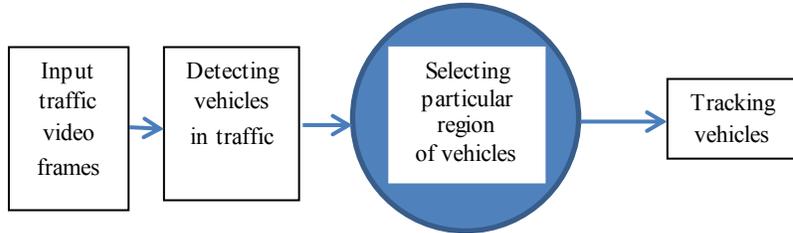


Figure 4: Block Diagram of Vehicle Detection [10]

3.2 Proposed System: Detection of vehicles

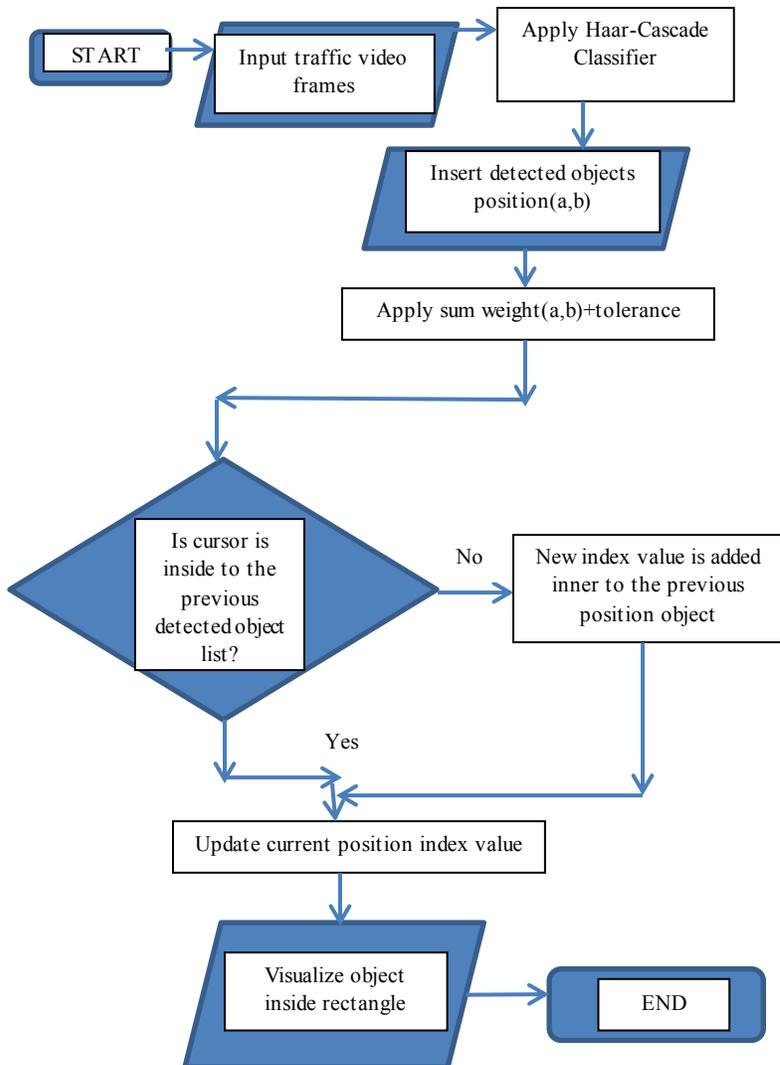


Figure 5: Detection of Vehicles using Haar-Cascade Classifier

3.2.1 Role of Haar-Cascade Classifier in Traffic Congestion:

The first and foremost feature stands ‘Haar-like feature’, which is a rectangular feature that provides a particular implication to an image. The aspect ‘Haar-like’ promotes extensive computational speed based on the pixels count inner the rectangle facet and not based on every picture element of the image. Haar-like aspect value was computed with the use of an integral image while obtaining the values of object detection.

$$S(a,b) = i(a,b) + S(a,b-1) + S(a-1,b-1) + S(a-1, b-1)$$

The above equation is used to calculate the object detection using integral images [11].

The ‘Haar-cascade classifier’ is implemented on road traffic video frames to identify the vehicles moving with the use of bounding box detection method, the proposed algorithm do the process of object detection till the end of video.

3.3 ALGORITHM: Classification of vehicles using Haar-Cascade Classifier

Input: Road traffic video frames

Output: Vehicles detection inside a rectangle

Import the necessary libraries

1. Create the directories to hold the video and its storage extension
2. Capture the road traffic moving video
3. Apply the concept of Haar Cascade Classifier
4. Initialize c to count the number of frames to track the objects
5. While True compute the following steps (a) to (f):
 - a) Read the video frames
 - b) Convert the images of the video to grayscale
 - c) Detect the images in a multi-scale dimension object
 - d) for (x, y, w, h) in objects:
 - i) Draw a rectangle around the tracked objects
 - ii) Display the tracked objects in traffic video frames one by one
 - iii) Resize the images to fit in frames
 - iv) Stop capturing the objects once the video is completed
6. Release the video frames from the classifier
7. Destroy all the windows after execution

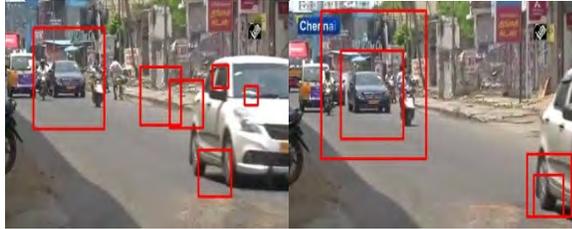


Figure 6: Detection of vehicles by using bounding box method of CNN

4. Used Technology

OpenCV is a large open-source library in favour of computer vision, in the concept of machine learning and image processing. OpenCV subsidizes a broad variety of programming languages such as python, C++, java, etc. It can refine images and videos to find objects, faces also human handwriting. The integration of OpenCV with several libraries, especially Numpy plays an important role to perform numerical operations.

This open-source library is used to execute the following tasks:

- Image resizing
- To set and display entry and exit places, including the presence of vehicles in determined places
- Camera manipulation and avoidance of distortion
- Usage of the context transformation matrix to determine the distance length
- Visualization of data[9]

4.1. Image processing

This technology can process the images in different forms according to the program's complexity. The various forms are listed below:

4.2. Blurring of images

Image blurring indicates the display of images is less clear, which is possible with the support of different minimal pass filter kernels.

Image blurring has the following benefits:

- Noise removal is possible
- Image smoothing is possible
- Removal of low-intensity edges
- Necessary details can be hidden

4.2.1. Significant types of blurring

4.2.1.1 Gaussian Blurring

It is the operation of Gaussian function, which is the most widely used aspect in graphics software to decrease noise on images and decrease its details. It can be processed as preprocessing stage prior to the usage of machine learning (or) deep learning models.



Figure 7.1: Gaussian Blurring

4.2.1.2. Median Blurring

It is processed with the use of a median filter, which is a non-linear technique in digital filtering, frequently used to eliminate noise either from an image or its signal. This filtering process is most widely utilized in digitized image processing, under specific conditions, it stores edges at the time of noise removal.



Figure 7.2: Median Blurring

4.2.1.3. Bilateral Blurring

It is operated by means of a bilateral filter, which is a non-linear, preserving edges and reducing noise with a smoothing filter, especially for images. It restores the intensity of every pixel with its weighted average of values derived from nearby pixels. As a result, sharp edges are stored by avoiding weak edges[12].



Figure 7.3: Bilateral Blurring

4.3. Grayscale of images

It is the operation of converting the colour spaces such as RGB, CMYK, HSV to grey shade, which differs absolutely in between the combination of black and white.

Grayscale has its importance in the following aspects:

- Dimensionality reduction
- Decreases model complexity



Figure 8.1: Image before Grayscaleing
 Figure 8.2: Image after Grayscaleing

4.4 Morphological operations

In general, the morphological operations processes images depend on their shapes. The most fundamental morphological operations are: Erosion and Dilation

4.4.1 Erosion Basics

- It weakens the edges of the foreground object
- Used to decline the image features
- Used to remove small white noises
- Separates two connected objects

4.4.2 Dilation Basics

- Object area is increased
- Highlights the features of an image
- After noise removal by erosion, dilation occurs to increase the object area
- It combines the separate object parts



Figure 9.1: Erosion of Image
 Figure 9.2: Dilation of Image

4.5. Image Translation

Translation indicates the object's rectilinear shift (image transferred from one place to another). The transformation matrix can be defined as:

$$\begin{bmatrix} 1 & 0 & tx \\ 0 & 1 & ty \end{bmatrix}$$

Where tx specifies the shift related to the x-axis

ty specifies the shift related to the y-axis

4.5.1 Uses of Image translation

- Hides a portion of the image
- Crops an image
- Shifts an image
- Animates an image using loops



Figure 10.1: Image before translation
 Figure 10.2: Image after translation

5. EXPERIMENTAL STUDY

The study of road traffic data is based on the Chennai region—traffic data set since 2014. More than 40000 data are available in which after preprocessing 3000 data are considered for the research study. Due to spatial and temporal issues in Chennai, commoners' movement by walk as pedestrians, transportation by motor vehicles, cars, trucks, vans and buses occurs at the same place and time yields traffic congestion.

The real road traffic dataset of Chennai region from the year 2014 to 2021 has been derived to show the number of persons killed, injured and victims in Grouped bar graph.

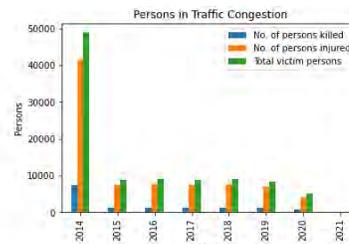


Figure 11: Persons in Congestion

6. DISCUSSION AND FUTURE WORKS:

The spatial-temporal concept is implemented on road traffic congestion and its issues. The traffic images of road transport such as two-wheelers, buses, trucks, vans, pedestrians are captured by using the convolutional neural network. The most commonly used Haar - cascade classifier is implemented to capture the video frames of road traffic to detect objects and draw a rectangle around them for identification. Visualizations of road traffic images have been shown with a technology named OpenCV. Also, Grouped bar graph has been shown

on the study of the Chennai road traffic data set. As future work, the road traffic data has to be further analysed in Recurrent Neural Network for clustering the detected vehicles playing on the road traffic and certain measures to avoid traffic congestion occurring periodically in Chennai regions.

7. CONCLUSION

This paper represents the road traffic images, certain points to measure traffic congestion happening in places consisting of extreme levels of population density while compared with rural places. Since, the Convolutional Neural Network plays an important role in capturing images, the traffic flow and the categorization of vehicles has been studied using it. The total count of epochs and their iterations have been studied on traffic images based on a pre-trained neural network to find its accuracy. The proposed model of the Haar cascade classifier is used to detect the vehicles in traffic video frames. The aim of the classifier draws a rectangle around the detected vehicle to recognize it as an object. Traffic images and their different forms have been shown in outputs with the use of OpenCV technology. The study of road traffic that occurs in Chennai has been applied in a Grouped bar graph to show the persons killed and injured due to traffic congestion in numbers since 2014. The paper benefits the readers to know the importance of CNN and Haar-Cascade Classifier in image classification. The limitations of this model can recognize the objects in frames and draws a rectangle around it, where it cannot identify the name, speed, number plate of the vehicles at the time of traffic congestion. The proposed study of this paper helps to extend the classification of images and their detection with its name to form clusters, which can be extended using Recurrent Neural Network and LSTM to prove the increased level of accuracy.

8. REFERENCES:

1. Deepika Sirohi, Neeraj Kumar and Prashanth Singh Rana (2020) 'Convolutional neural networks for 5G-enabled Intelligent Transportation System: A Systematic review', *Elsevier*, pp.459–498.
2. Tanzina Afrin and Nita Yoda (2020), 'A survey of Road Traffic Congestion measures towards a sustainable and Resilient Transportation System', *MDPI*, pp. 1–23.
3. Data Collection Survey for Chennai Metropolitan Region, Intelligent Transport Systems (2017).
4. Nadia Jmour, Sehla Xayen and Afef Abdelkrim (2018), 'Convolutional Neural Networks for Image Classification', *IEEE*.
5. Upgrad.com/blog/basic-cnn-architecture
6. PingWang, Li Li, Yinli Jin and Guiping Wang (2018), 'Detection of Unwanted Traffic Congestion based on existing surveillance system using in freeway via a CNN-architecture Traffic Net', *IEEE*.
7. Ming Lan Sheng, Chunfang Liu, Qi Zhang, Lu Lou and Yu Zheng (2018), 'Vehicle Detection and Classification using Convolutional Neural Networks', *IEEE*.
8. Sampath Kumar. V, Vanjinathan. J and Priyadarshini Pradeep (2015), 'Traffic Management at Sterling Road Intersection in Chennai', *International Journal of Applied Engineering Research*.
9. Kiril Khazukov, Vladimir Shepelev, Tatiana Karpeta, Salavat Shabiev, Ivan Slobodin, Iraki Charbadee and Irina Alferova (2020), 'Real-time Monitoring of Traffic Parameters', *Journal of Big data*, pp. 1–20.
10. Nikhil Chhadikar, Priyanka Bhamare, Krushna Patil and Sangeeta Kumari (2019), 'Image Processing based Tracking and Counting Vehicles', *IEEE*.
11. Moch Ilhaan Ramadhani, Agus Eko Minarno, Eko Budi Cahyono (2018), 'Vehicle classification using Haar Cascade Classifier method in Traffic Surveillance System', *Kinetik*, pp. 57–66.
12. [geeksforgeeks.org/python-image-blurring-using-OpenCV/](https://www.geeksforgeeks.org/python-image-blurring-using-OpenCV/)
13. Ayushi Sharma, Jyotsna Pathak, Muskan Prakash, JN Singh, 'Object Detection using OpenCV and Python', *IEEE 2021*
14. Abdullah Tariq, Muhammad Zeeshan Khan, Muhammad Usman Ghani Khan, 'Real time Vehicle Detection and Colour Recognition using tuned features of Faster-RCNN', *IEEE 2021*
15. Praveen Kumar Maduri, Preeti Dhiman, Shalini Rathour, Eeshaan Dutt, Aman Gupta, 'Smart Eye of Traffic Management & Control System', *IEEE 2021*
16. Mohamed Othmani, 'A vehicle detection and tracking method for traffic video based on traffic R-CNN', *Springer Link*, March 2022
17. Divya Sharma, Shilpa Sharma, Vaibhav Bhatnagar, 'Automated Vehicle Speed Estimation and License Plate Recognition for Smart Cities Development', *IEEE 2022*
18. Mohana, Rahul Kejriwal, Ritika H.J, Arpit Arora, "Artificial Intelligence(AI) enabled Vehicle Detection and counting using Deep Learning", *IEEE 2022*
19. Abirami T, Nivas C, Naveen R, Nithishkumar T.G, "Deep Learning based Traffic Analysis of Motor cycles in Urban City", *IEEE 2022*