# Deep CNN Model for Condition Monitoring of Road Traffic: An Application Of Computer Vision

**Dr. P. Hasitha Reddy[1], M. Manjunath[2], M. Rohith[2], N. Manohar Reddy[2], A. Satyanarayana[2]**

[1,2]Department of Computer Science and Engineering

[1,2] Sree Dattha Group of Institutions, Sheriguda, Telangana.

**Abstract:** The traffic surveillance system accumulates an enormous amount of data regarding road traffic each second. Monitoring these data with the human eye is a tedious process and it also requires manpower for monitoring. Deep learning Convolutional Neural Network (DLCNN) can be utilized for traffic monitoring and control. The traffic surveillance data are pre-processed to construct the training dataset. The Traffic net is constructed by transferring the network to traffic applications and retraining it with a self-established data set. This Traffic-net can be used for regional detection in large scale applications. Further, it can be implemented across-the-board. Further, DLCNN is used for prediction of traffic status i.e., dense traffic, low traffic, accident, and fire occurred from test sample. Finally, the simulations revealed that the proposed DLCNN resulted in superior performance as compared to existing model.

**Keywords:** Traffic Control Monitoring, DLCNN, Traffic Surveillance.

## 1. INTRODUCTION

### 1.1 Overview

As urbanization has accelerated, traffic in urban areas has increased significantly, and the similar phenomenon has been appeared in freeways connected to the urban areas as well. The real-time monitoring of traffic on freeways could provide sophisticated traffic information to drivers, so the drivers could choose alternative routes to avoid heavy traffic [1]. Furthermore, long-term records of traffic monitoring will be helpful for developing efficient transportation policies and strategies across urban and suburban areas. Currently, the typical means of monitoring traffic information use closed circuit television (CCTV) or detection equipment. The detection equipment includes loop detectors [2], image detectors [3], dedicated short range communication (DSRC) [4], and radar detectors [5]. In general, CCTVs are installed at fixed locations, and they can monitor the area on the freeway 24 hours a day. CCTV can monitor only limited areas; therefore, multiple CCTV circuits are necessary to monitor a wide range of freeways. However, the installation and maintenance of the multiple CCTV circuits is costly. In addition, it is difficult to detect vehicles in CCTV videos automatically due to the overlapping between vehicles because CCTV usually captures freeways in an oblique direction.

Recently, to overcome the limitations of collecting traffic information through CCTV, video collection methods employing unmanned aerial vehicles (UAVs) are being used [6]. Unlike CCTV, a UAV can monitor a wide range of freeways by elevating its altitude or moving its location, and it can travel to a specific location to observe unexpected situations, such as traffic accidents. Furthermore, a UAV views the freeways in a perpendicular direction, so the vehicles in the recorded videos do not overlap. Currently, however, videos from installed CCTV or operated UAVs are monitored by humans. Therefore, as the number of CCTV circuits and UAVs increases, more human resources are required. Moreover, we can not avoid human error; it is highly demanding to analyze real-time videos to effectively monitor traffic information.

In contrast to CCTV videos taken at a fixed height, the altitude of UAV varies at every time the video is recorded, and sometimes the altitude of UAV changes during recording. If the image scales are not fixed, we are not able to estimate the vehicle's traveling distance on the actual road by simply measuring moving distance of the vehicle in sequential images. Therefore, to determine the exact speed of a vehicle by tracking the vehicle in sequential images, the image scale of each image should be estimated and the changes in the image scale should be taken into account. For example, the scale of the image was obtained by comparing a pre-defined structure on an actual road with its corresponding object in the first frame of a video. This approach requires a pre-definition of a structure for each location; therefore, images without known structures cannot be utilized. Later, the image scale is calculated by comparing the average sizes of vehicles in the images and pre-measured and averaged actual vehicle size. Although these methods have somewhat resolved the restrictions associated with a UAV's flight area, the calculated image scale is not accurate because the size of vehicle varies depend on the types of the vehicles. For instance, a detected vehicle can include sedans, vans, buses, of trucks.

### 1.2 Problem Definition

The Existing method requires a huge amount of Hardware equipments deployed to the Road. Moreover, they are very sensitive to external noise and environmental conditions. It is more accurate when processing a limited number of vehicles, but it does not work well on large scale dataset.

### 1.3 Objective Of Project

The purpose of the proposed work is to make a speedy traffic detection system which reduces the manpower and detected Multiclass problems namely fire detection, accident detection, dense and sparse traffic detection. The main aim is to classify the given input image to dense or sparse based on the trained model from the input dataset.

## 2. LITERATURE SURVEY

### 2.1 Survey on Evolving Deep Learning Neural Network Architectures

**AUTHORS:** Abul Bashar

The deep learning being a subcategory of the machine learning follows the human instincts of learning by example to produce accurate results. The deep learning performs training to the computer frame work to directly classify the tasks from the documents available either in the form of the text, image, or the sound. Most often the deep learning utilizes the neural network to perform the accurate classification and is referred as the deep neural networks; one of the most common deep neural networks used in a broader range of applications is the convolution neural network that provides an automated way of feature extraction by learning the features directly from the images or the text unlike the machine learning that extracts the features manually. This enables the deep learning neural networks to have a state of art accuracy that mostly expels even the human performance. So the paper is to present the survey on the deep learning neural network architectures utilized in various applications for having an accurate classification with an automated feature extraction.

### 2.2 A Novel Online Dynamic Temporal Context Neural Network Framework for the Prediction of Road Traffic Flow

**AUTHORS: Zoe Bartlett et al**

Traffic flow exhibits different magnitudes of temporal patterns, such as short-term (daily and weekly) and long-term (monthly and yearly). Existing research into road traffic flow prediction has focused on short-term patterns; little research has been done to determine the effect of different long-term patterns on road traffic flow prediction. Providing more temporal contextual information through the use of different temporal data segments could improve prediction results. In this paper, we have investigated different magnitudes of temporal patterns, such as short-term and long-term, through the use of different temporal data segments to understand how contextual temporal data can improve prediction. Furthermore, to learn temporal patterns dynamically, we have proposed a novel online dynamic temporal context neural network framework. The framework uses different temporal data segments as input features, and during online learning, the updating scheme dynamically determines how useful a temporal data segment (short and long-term temporal patterns) is for prediction, and weights it accordingly for use in the regression model. Therefore, the framework can include short-term and relevant long-term patterns in the regression model leading to improved prediction results. We have conducted a thorough experimental evaluation with a real dataset containing daily, weekly, monthly and yearly data segments. The experiment results show that both short and long-term temporal patterns improved prediction accuracy. In addition, the proposed online dynamical framework improved predication results by 10.8% when compared with a deep gated recurrent unit model.

### 2.3 Detection of unwanted traffic congestion based on existing surveillance system using in freeway via a CNN architecture trafficNet

AUTHORS: P. Wang, L. Li, Y. Jin, and G. Wang

Detection of traffic congestion is important for route guidance using in intelligent transport system (ITS) to prevent jam escalation. Although the surveillance system has been used in freeway for years, it is hard to automatically identify and report traffic congestion in complicated transportation scene according to various illumination, weather and other disturbances. The detection process based on human eye is time-consuming and tedious as the machine detection accuracy is not high enough to meet the requirements of practical applications. In this paper, a new classifier is proposed using convolutional neural networks (CNN) to generate four TrafficNet based on two championships of ILSVRC including AlexNet and VGGNet. Instead of using fully-connected layers in AlexNet and VGGNet, a support vector machine (SVM) are used after CNN architecture. Congestion and non-congestion images are trained and tested through this new structure. Image database with more than 30000 images are extracted from existing traffic surveillance video and corresponding labels are added manually. With database, those TrafficNet are trained and tested, detection accuracy and training time of those TrafficNet are compared. The experimental results show that the accuracy of proposed method can reach up to 90%, which is much higher than traditional method based on feature extraction without deep learning.

### 2.4 LSTM network: A deep learning approach for short-term traffic forecast

AUTHORS: Weihai Chen,Zheng Zhao,Jingmeng Liu and Peter C. Y. Chen

Short-term traffic forecast is one of the essential issues in intelligent transportation system. Accurate forecast result enables commuters make appropriate travel modes, travel routes, and departure time, which is meaningful in traffic management. To promote the forecast accuracy, a feasible way is to develop a more effective approach for traffic data analysis. The availability of abundant traffic data and computation power emerge in recent years, which motivates us to improve the accuracy of short-term traffic forecast via deep learning approaches. A novel traffic forecast model based on long short-term memory (LSTM) network is proposed. Different from conventional forecast models, the proposed LSTM network considers temporal-spatial correlation in traffic system via a two-dimensional network which is composed of many memory units. A comparison with other representative forecast models validates that the proposed LSTM network can achieve a better performance.

### 2.5 Deep visual tracking: Review and experimental comparison

**AUTHORS: Peixia Li,Dong Wang,Lijun Wang and Huchuan Lu**

Recently, deep learning has achieved great success in visual tracking. The goal of this paper is to review the state-of-the-art tracking methods based on deep learning. First, we introduce the background of deep visual tracking, including the fundamental concepts of visual tracking and related deep learning algorithms. Second, we categorize the existing deep-learning-based trackers into three classes according to network structure, network function and network training. For each categorize, we explain its analysis of the network perspective and analyze papers in different categories. Then, we conduct extensive experiments to compare the representative methods on the popular OTB-100, TC-128 and VOT2015 benchmarks. Based on our observations, we conclude that: (1) The usage of the convolutional neural network (CNN) model could significantly improve the tracking performance. (2) The trackers using the convolutional neural network (CNN) model to distinguish the tracked object from its surrounding background could get more accurate results, while using the CNN model for template matching is usually faster. (3) The trackers with deep features perform much better than those with low-level hand-crafted features. (4) Deep features from different convolutional layers have different characteristics and the effective combination of them usually results in a more robust tracker. (5) The deep visual trackers using end-to-end networks usually perform better than the trackers merely using feature extraction networks. (6) For visual tracking, the most suitable network training method is to per-train networks with video information and online fine-tune them with subsequent observations. Finally, we summarize our manuscript and highlight our insights, and point out the further trends for deep visual tracking.

## 3. PROPOSED SYSTEM

The traffic surveillance system is accumulated with an enormous amount of data regarding road traffic each and every second. Monitoring these data with the human eye is a tedious process and it also requires manpower for monitoring. Deep learning approach (Convolutional Neural Network) can be utilized for traffic monitoring and control. The traffic surveillance data are pre-processed to construct the training dataset. The Traffic net is constructed by transferring the network to traffic applications and retraining it with self-established data set. This Traffic net can be used for regional detection in large scale applications. Further, it can be implemented across-the-board. The efficiency is admirably verified through speedy discovery in the high accuracy in the case study. The tentative assessment could pull out to its successful application to a traffic surveillance system and has potential enrichment for the intelligent transport system in future.

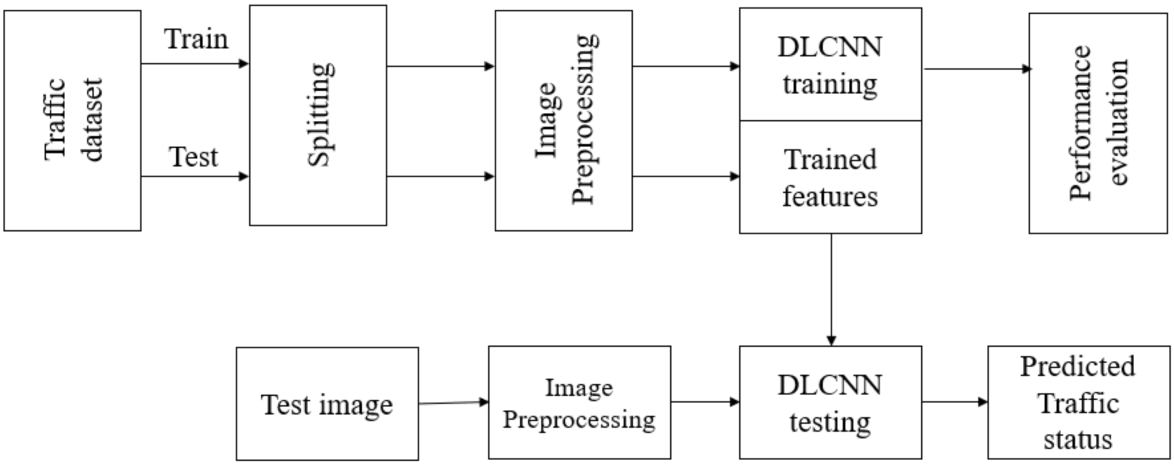


Fig. 1: Proposed methodology.

Fig. 1 shows the block diagram of proposed method. Initially, TrafficNet dataset is spitted into 80% for training and 20% for testing. Then, dataset preprocessing operation is performed to normalize the entire dataset. The image preprocessing operation converts the all the images into uniform size. Further, DLCNN is used for prediction of traffic status i.e., dense traffic, low traffic, accident, and fire occurred from test sample. The performance evaluation is carried out to show supremacy of proposed method.

### 3.1 Traffic condition dataset

The input dataset of dissimilar classes is gathered from the net. The assessment of output class is set next to the obtained dataset. Four folders namely sparse_traffic, dense_traffic, fire, accident, every folder contains 900 images are generated for training and validation purposes. The folder name represents the class value for classifying output.

### 3.2 Image pre-processing

Digital image processing is the use of computer algorithms to perform image processing on digital images. As a subfield of digital signal processing, digital image processing has many advantages over analogue image processing. It allows a much wider range of algorithms to be applied to the input data — the aim of digital image processing is to improve the image data (features) by suppressing unwanted distortions and/or enhancement of some important image features so that our AI-Computer Vision models can benefit from this improved data to work on. To train a network and make predictions on new data, our images must match the input size of the network. If we need to adjust the size of images to match the network, then we can rescale or crop data to the required size.

we can effectively increase the amount of training data by applying randomized augmentation to data. Augmentation also enables to train networks to be invariant to distortions in image data. For example, we can add randomized rotations to input images so that a network is invariant to the presence of rotation in input images. An augmented Image Datastore provides a convenient way to apply a limited set of augmentations to 2-D images for classification problems.

we can store image data as a numeric array, an ImageDatastore object, or a table. An ImageDatastore enables to import data in batches from image collections that are too large to fit in memory. we can use an augmented image datastore or a resized 4-D array for training, prediction, and classification. We can use a resized 3-D array for prediction and classification only.

There are two ways to resize image data to match the input size of a network. Rescaling multiplies the height and width of the image by a scaling factor. If the scaling factor is not identical in the vertical and horizontal directions, then rescaling changes the spatial extents of the pixels and the aspect ratio.

Cropping extracts a subregion of the image and preserves the spatial extent of each pixel. We can crop images from the center or from random positions in the image. An image is nothing more than a two-dimensional array of numbers (or pixels) ranging between 0 and 255. It is defined by the mathematical function f(x,y) where x and y are the two co-ordinates horizontally and vertically.

**Resize image:** In this step-in order to visualize the change, we are going to create two functions to display the images the first being a one to display one image and the second for two images. After that, we then create a function called processing that just receives the images as a parameter.

Need of resize image during the pre-processing phase, some images captured by a camera and fed to our AI algorithm vary in size, therefore, we should establish a base size for all images fed into our AI algorithms.

### 3.3 Proposed ResNet-CNN

Deep neural network is gradually applied to the identification of crop Traffic conditions and insect pests. Deep neural network is designed by imitating the structure of biological neural network, an artificial neural network to imitate the brain, using learnable parameters to replace the links between neurons. Convolutional neural network is one of the most widely used deep neural network structures, which is a branch of feed forward neural network. The success of AlexNet network model also confirms the importance of convolutional neural network model. Since then, convolutional neural networks have developed vigorously and have been widely used in financial supervision, text and speech recognition, smart home, medical diagnosis, and other fields.

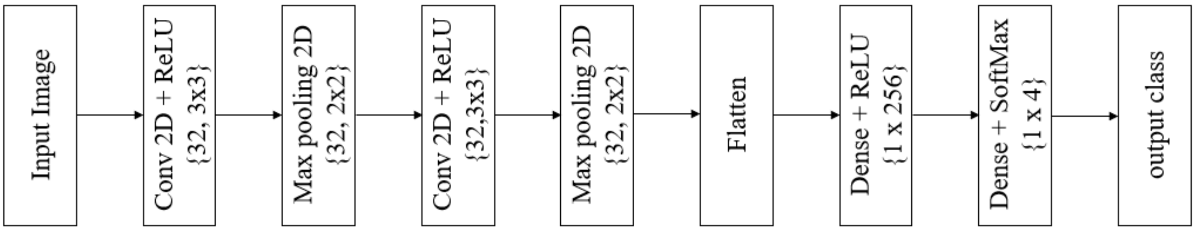


Fig. 2: Proposed DLCNN

Table. 1: Layers description.

| Layer Names | No. of filters | Kernel size | Feature size |
|---|---|---|---|
| Conv 2D +ReLU | 32 | 3 x 3 | 62x62x32 |
| Max pooling 2D | - | 3 x 3 | 31x31x32 |
| Conv 2D+ReLU | 32 | 3 x 3 | 29x29x32 |
| Max pooling 2D | - | 3 x 3 | 14x14x32 |
| Flatten | - | 1x6272 | 1x6272 |
| Dense +ReLU |  | 1 x 256 | 1 x 256 |
| Dense + SoftMax |  | 1 x 4 | 1 x 4 |

Convolutional neural networks are generally composed of three parts. Convolution layer for feature extraction. The convergence layer, also known as the pooling layer, is mainly used for feature selection. The number of parameters is reduced by reducing the number of features. The full connection layer carries out the summary and output of the characteristics. A convolution layer is consisting of a convolution process and a nonlinear activation function ReLU. A typical architecture of CNN model for crop Traffic condition recognition is shown in Fig. 2.

The leftmost image is the input layer, which the computer understands as the input of several matrices. Next is the convolution layer, the activation function of which uses ReLU. The pooling layer has no activation function. The combination of convolution and pooling layers can be constructed many times. The combination of convolution layer and convolution layer or convolution layer and pool layer can be very flexibly, which is not limited when constructing the model. But the most common CNN is a combination of several convolution layers and pooling layers. Finally, there is a full connection layer, which acts as a classifier and maps the learned feature representation to the sample label space.

Convolutional neural network mainly solves the following two problems.

1) Problem of too many parameters: It is assumed that the size of the input picture is $50 * 50 * 3$. If placed in a fully connected feedforward network, there are 7500 mutually independent links to the hidden layer. And each link also corresponds to its unique weight parameter. With the increase of the number of layers, the size of the parameters also increases significantly. On the one hand, it will easily lead to the occurrence of over-fitting phenomenon. On the other hand, the neural network is too complex, which will seriously affect the training efficiency. In convolutional neural networks, the parameter sharing mechanism makes the same parameters used in multiple functions of a model, and each element of the convolutional kernel will act on a specific position of each local input. The neural network only needs to learn a set of parameters and does not need to optimize learning for each parameter of each position.

2) Image stability: Image stability is the local invariant feature, which means that the natural image will not be affected by the scaling, translation, and rotation of the image size. Because in deep learning, data enhancement is generally needed to improve performance, and fully connected feedforward neural is difficult to ensure the local invariance of the image. This problem can be solved by convolution operation in convolutional neural network.

**Convolution layer:** According to the facts, training and testing of DLCNN involves in allowing every source image via a succession of convolution layers by a kernel or filter, rectified linear unit (ReLU), max pooling, fully connected layer and utilize SoftMax layer with classification layer to categorize the objects with probabilistic values ranging from [0,1].

Convolution layer as depicted in Figure 4.3 is the primary layer to extract the features from a source image and maintains the relationship between pixels by learning the features of image by employing tiny blocks of source data. It's a mathematical function which considers two inputs like source image $I(x, y, d)$ where $x$ and $y$ denotes the spatial coordinates i.e., number of rows and columns. $d$ is denoted as dimension of an image (here $d = 3$, since the source image is RGB) and a filter or kernel with similar size of input image and can be denoted as $F(k_x, k_y, d)$.

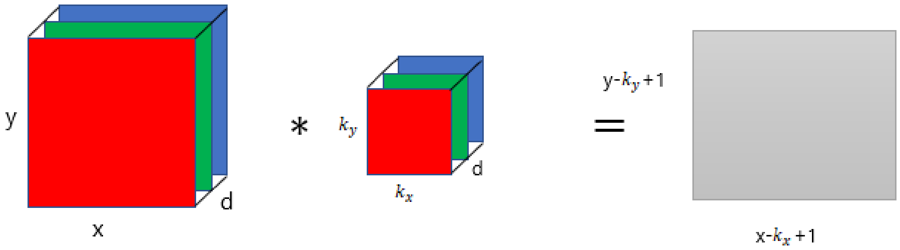
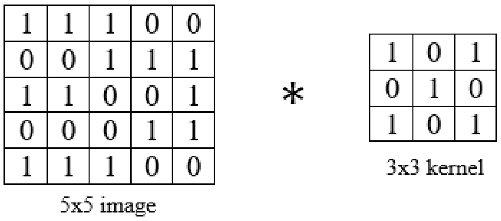


Fig. 3: Representation of convolution layer process.

The output obtained from convolution process of input image and filter has a size of $C\left((x - k_x + 1), (y - k_y + 1), 1\right)$, which is referred as feature map. An example of convolution procedure is demonstrated in Figure 4(a). Let us assume an input image with a size of $5 \times 5$ and the filter having the size of $3 \times 3$. The feature map of input image is obtained by multiplying the input image values with the filter values as given in Figure 4 (b).



| | | | | |
|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 0 |
| 0 | 0 | 1 | 1 | 1 |
| 1 | 1 | 0 | 0 | 1 |
| 0 | 0 | 0 | 1 | 1 |
| 1 | 1 | 1 | 0 | 0 |

5x5 image

*

| | | |
|---|---|---|
| 1 | 0 | 1 |
| 0 | 1 | 0 |
| 1 | 0 | 1 |

3x3 kernel

(a)

$$\underbrace{\begin{array}{|c|c|c|c|c|}\hline 1&1&1&0&0\\\hline 0&0&1&1&1\\\hline 1&1&0&0&1\\\hline 0&0&0&1&1\\\hline 1&1&1&0&0\\\hline\end{array}}_{\text{5x5 image}} * \underbrace{\begin{array}{|c|c|c|}\hline 1&0&1\\\hline 0&1&0\\\hline 1&0&1\\\hline\end{array}}_{\text{3x3 kernel}} = \underbrace{\begin{array}{|c|c|c|}\hline 3&3&3\\\hline 2&2&3\\\hline 3&2&3\\\hline\end{array}}_{\text{Feature map}}$$

(b)

Fig. 4: Example of convolution layer process (a) an image with size $\mathbf{5} \times \mathbf{5}$ is convolving with $\mathbf{3} \times \mathbf{3}$ kernel (b) Convolved feature map

**ReLU layer**: Networks those utilizes the rectifier operation for the hidden layers are cited as rectified linear unit (ReLU). This ReLU function $\mathcal{G}(\cdot)$ is a simple computation that returns the value given as input directly if the value of input is greater than zero else returns zero. This can be represented as mathematically using the function $max(\cdot)$ over the set of 0 and the input $x$ as follows:

$$\mathcal{G}(x) = \max\{0, x\}$$

**Max pooing layer**: This layer mitigates the number of parameters when there are larger size images. This can be called as subsampling or down sampling that mitigates the dimensionality of every feature map by preserving the important information. Max pooling considers the maximum element form the rectified feature map.

**SoftMax classifier:** Generally, softmax function is added at the end of the output as shown in Figure 5, since it is the place where the nodes are meet finally and thus, they can be classified. Here, X is the input of all the models and the layers between X and Y are the hidden layers and the data is passed from X to all the layers and Received by Y. Suppose, we have 10 classes, and we predict for which class the given input belongs to. So, for this what we do is allot each class with a particular predicted output. Which means that we have 10 outputs corresponding to 10 different class and predict the class by the highest probability it has.

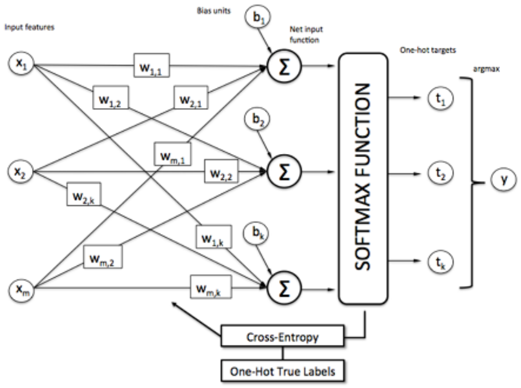


Fig. 5: Traffic condition prediction using SoftMax classifier.

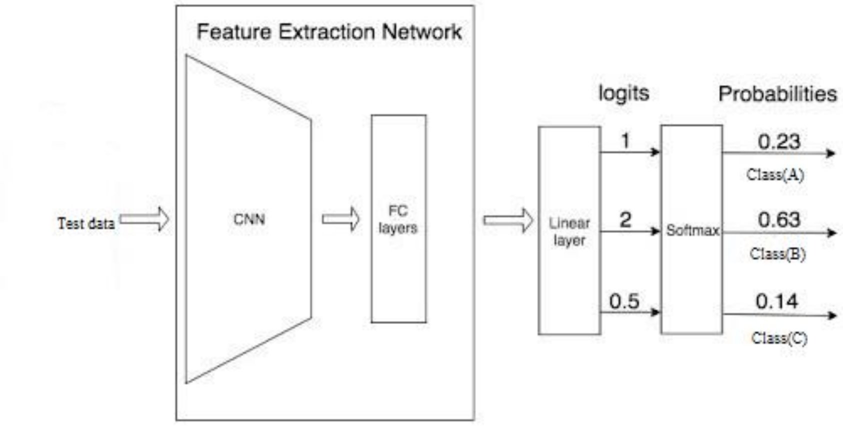


Fig. 6: Example of SoftMax classifier.

In Fig. 6, and we must predict what is the object that is present in the picture. In the normal case, we predict whether the crop is A. But in this case, we must predict what is the object that is present in the picture. This is the place where softmax comes in handy. As the model is already trained on some data. So, as soon as the picture is given, the model processes the pictures, send it to the hidden layers and then finally send to softmax for classifying the picture. The softmax uses a One-Hot encoding Technique to calculate the cross-entropy loss and get the max. One-Hot Encoding is the technique that is used to categorize the data. In the previous example, if softmax predicts that the object is class A then the One-Hot Encoding for:

Class A will be [1 0 0]

Class B will be [0 1 0]

Class C will be [0 0 1]

From the diagram, we see that the predictions are occurred. But generally, we don't know the predictions. But the machine must choose the correct predicted object. So, for machine to identify an object correctly, it uses a function called cross-entropy function.

So, we choose more similar value by using the below cross-entropy formula.

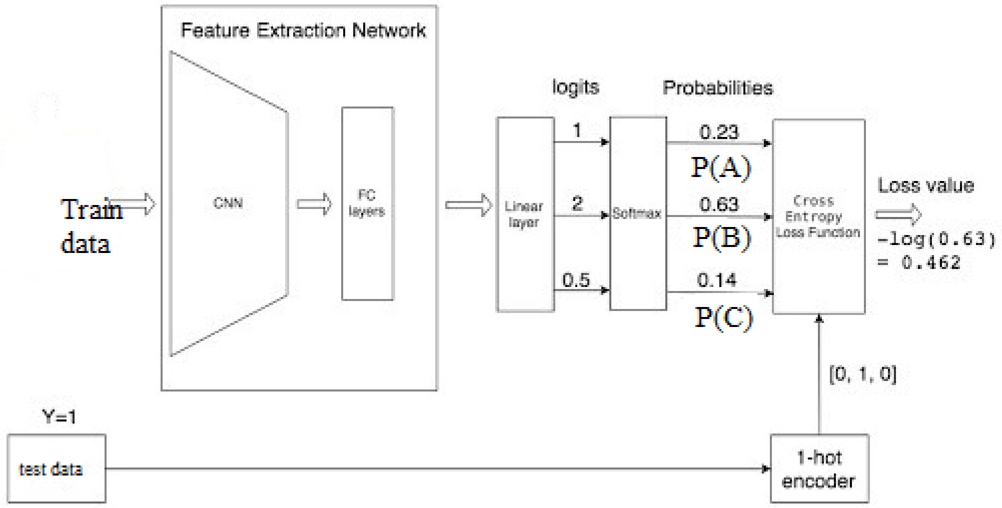


Fig. 7: Example of SoftMax classifier with test data.

In the above example we see that 0.462 is the loss of the function for class specific classifier. In the same way, we find loss for remaining classifiers. The lowest the loss function, the better the prediction is. The mathematical representation for loss function can be represented as: -

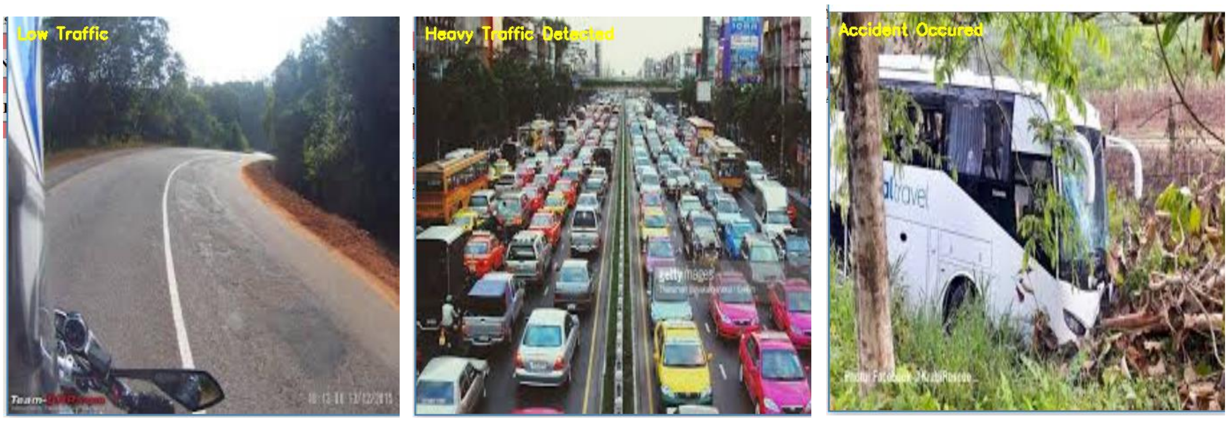
$$LOSS = np.sum(-Y \ * \ np.log(Y\_pred))$$

## 4. RESULTS AND DISCUSSION

This section gives the detailed analysis of simulation results implemented using "python environment". Further, the performance of proposed method is compared with existing methods using same dataset. Table 1 compares the performance of the proposed method with existing methods. Here, the Proposed CNN resulted in superior accuracy as compared to the existing NB and SVM. Fig. 8 shows the predicted outcomes of proposed method.

Table.2: Performance comparison.

| Method | NB [7] | SVM [9] | Proposed CNN |
|---|---|---|---|
| Accuracy values | 89.5 | 90 | 99.3 |



(a) (b) (c)

Fig. 8: Predicted outcomes, (a) Classifier predicted as low traffic, (b) Classifier predicted as heavy traffic, (c) Classifier predicted as accident occurred.

## 5. CONCLUSION

The Convolution neural network is approached to identify the condition of road congestion without human intervention. This is anticipated to deploy deep learning in various realistic functions. The proposed CNN for training and validation is considered as a multi class problem. As a future enhancement, the traffic conditions are detected on the traffic videos on real-time. This can be done by video splitting technique are found and the traffic condition on every frame. Real time traffic detection on video is quite important research for developing countries like India.

## REFERENCES

[1] Bashar, A. (2019). “Survey onEvolving Deep Learning Neural Network Architectures”. Journal of Artificial Intelligence, 1(02), 73-82.

[2] Zoe Bartlett et al, “A Novel Online Dynamic Temporal Context Neural Network Framework for the Prediction of Road Traffic Flow” IEEE Access, vol.7, 2019.

[3] H. Lei et al., ``A deeply supervised residual network for HEp-2 cell classification via cross-modal transfer learning," Pattern Recognit., vol. 79, pp. 290302, Jul. 2018.

[4] P. Wang, L. Li, Y. Jin, and G. Wang, ``Detection of unwanted traffic congestion based on existing surveillance system using in freeway via a CNNarchitecture trafficNet," in Proc. 13th IEEE Conf. Ind. Electron. Appl., May/Jun. 2018, pp. 11341139.

[5] X. Zhu, Y.Wang, J. Dai, L. Yuan, and Y.Wei, ``Flowguided feature aggregation for video object detection," in Proc. ICCV, Mar. 2017, pp. 408417.

[6] Z. Zhao. Chen, X.Wu, P. C. Chen, and J. Liu, ``LSTM network: A deep learning approach for short-term traffic forecast," IET Intell. Transp. Syst., vol. 11, no. 2, pp. 6875, Mar. 2017.

[7] P. Li, D. Wang, L. Wang, and H. Lu, ``Deep visual tracking: Review and experimental comparison," Pattern Recognit., vol. 76, pp. 323338, Apr. 2018.

[8] P. Wang and J. Di, ``Deep learning-based object classification through multimode fiber via a CNNarchitecture SpeckleNet," Appl. Opt., vol. 57, no. 28, pp. 82588263, 2018.

[9] J. Zhao, Z. Zhang, W. Yu, and T.-K. Truong, ``A cascade coupled convolutional neural network guided visual attention method for ship detection from SAR images," IEEE Access, vol. 6, pp. 5069350708, 2018.

[10] T. Pamula, ``Road traffic conditions classification based on multilevel filtering of image content using convolutional neural networks," IEEE Intel. Transp. Syst. Mag., vol. 10, no. 3, pp. 1121, Jun. 2018.