JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 08, ISSUE 05, 2021

Condition Monitoring of Road Traffic using Deep CNN Model

Ramavath Thukaram, Akavaram Swapna

Department of Computer Science and Engineering

Sree Dattha Group of Institutions, Hyderabad, Telangana, India.

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: Road traffic, deep learning, traffic surveillance.

1. INTRODUCTION

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.

Several methods have been developed to automatically analyze traffic conditions on freeways using videos from CCTV or dash cam. In [7], the vehicle was detected using Mask RCNN [3] from the surveillance video taken with a fixed camera, and the vehicle speed was calculated. In [8], a vehicle was detected in the image using Ada-Boost, which uses multiple weak classifiers to construct a strong classifier. Recently, as artificial intelligence technology has rapidly advanced, road image analysis methods using deep learning are also being proposed. A deep learning-based object detection method, Faster R-CNN [9], was used to detect vehicles in images. There have also been various attempts to automatically measure traffic from road videos taken by UAVs. Most of these methods consist of object detection techniques for capturing vehicles in images and object tracking techniques for identifying the movements of detected vehicles, and the speed of vehicles are calculated at the end. Various types of vehicles were detected from UAV video using Yolo. The vehicle speed was calculated from the results of tracking the vehicle using the moving average of the previous frame and the Kalman filter. A Haar-like feature-based cascade structure is used to detect the location and size of a vehicle in the image with a bounding box, and the convolutional neural network (CNN) method was applied to the detection results to improve the final classification performance. Traffic volume was also calculated by tracking the movement of the vehicle using the KLT-optical flow.

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 the 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 considered. 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 depending on the types of the vehicles. For instance, a detected vehicle can include sedans, vans, buses, or trucks.

2. LITERATURE SURVEY

Abdul et al. studied the deep learning being a subcategory of the machine learning follows the human instincts of learning by example to produce accurate results. Deep learning performs training to the computer framework to directly classify the tasks from the documents available either in the form of the text, image, or the sound. This enables the deep learning neural networks to have a state of art accuracy that mostly expels even 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.

Zoe et al. studied the 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 using 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, using different temporal data segments to understand how contextual temporal data can improve prediction. 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.

Wang et al. prevented the jam escalation of detection of traffic congestion is important for route guidance using in intelligent transport system (ITS. 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 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.

Chen et al. studied the 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 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.

Peixia et al. achieved a deep learning in 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 (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 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 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 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. 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.

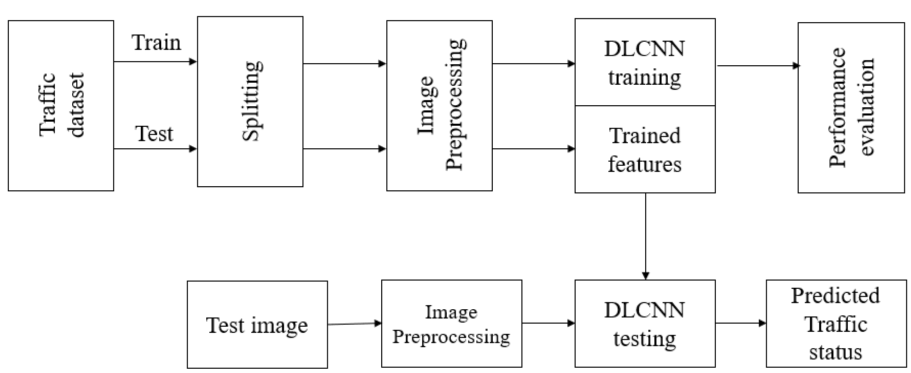


Fig. 1. Proposed method.

ISSN- 2394-5125 VOL 08, ISSUE 05, 2021

Fig. 1 shows the block diagram of the 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 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.

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.

3.4 Proposed DLCNN

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.

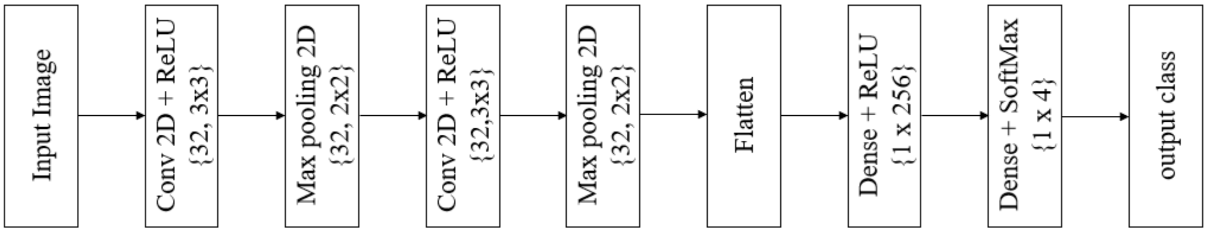


Fig. 2: Proposed ResNet-CNN.

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.

JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 08, ISSUE 05, 2021

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.

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]. Fig. 1 discloses the architecture of ResNet-CNN that is utilized in proposed methodology for CBIR system for enhanced feature representation of word image over conventional retrieval systems. Convolution layer as depicted in Fig. 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 $F\left(k_x, k_y, d\right)$.

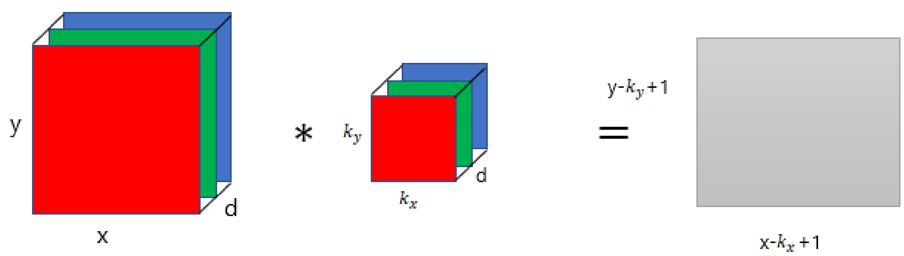


Fig. 3: Representation of convolution layer process.

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

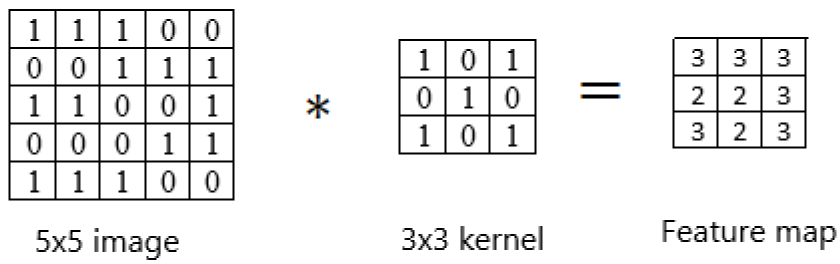
5x5 image

*

1	0	1
0	1	0
1	0	1

3x3 kernel

(a)



(b)

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

3.4.1 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:

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

3.4.2 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.

3.5 Advantages of Proposed System

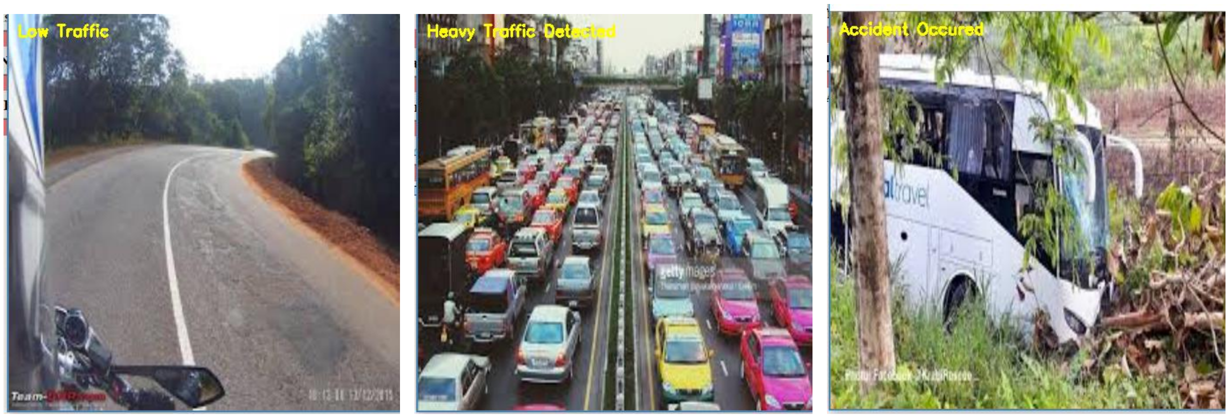
1.The proposed method overcomes the shortcomings of the traditional method. In this digital world deep learning paves, the way for fleet management.
2.Rather than the existing method, the CNN can solve Multiclass problems. This has speedy detection and provides timely information to the public.

4. RESULTS

This section gives the detailed analysis of simulation results implemented using “python environment”. Further, the performance of the proposed method is compared with existing methods using same dataset. Table 2 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. 5 shows the predicted outcomes of the proposed method.

Table.2: Performance comparison.

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



(a) (b) (c)

Fig.5: 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 DLCNN 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.