Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 09, September-2019

Traffic Analysis and Estimation using Deep Learning Techniques

Shlok Sayani
Department of Computer Engineering
Institute of Technology, Nirma University
Ahmedabad, Gujarat, India

Parth Nanvani
Department of Computer Engineering
Institute of Technology, Nirma University
Ahmedabad, Gujarat, India

***Abstract*—Roads are the most common mode of transportation. People generally choose their private vehicles or public transport for daily commute. The habit of using private vehicles has caused an increase in the number of vehicles. This, along with industrialization has exacerbated the problem of traffic congestion especially in urban areas. Hence, there is a need to manage this problem by making the traffic signals dynamic in nature and improving the road navigation systems by predicting the dynamic traffic flow. In this paper, we have proposed a solution which uses RetinaNet and Long short-term memory for traffic prediction. RetinaNet uses the data from CCTV traffic cameras to detect the vehicles and classify them. This data is then stored along with the time stamp and speed of the vehicles in a file. After the required feature engineering is done, this file is used by the LSTM model to predict the traffic volume by learning the sequences of observations.**

Keywords*— RetinaNet, Long short-term memory, OpenCV, traffic prediction, LSTM

I. INTRODUCTION

The issue of traffic has turned out to be exceptionally grave during the most recent decade particularly in enormous urban areas and on highways connecting different cities. Traffic jams are serious and conspicuous during the times of heavy traffic. As a result of industrialization and urbanization, the living standard of people has increased. This improved way of life inside urban areas has aided people to use their private vehicles as opposed to public transport which results into increase in number of vehicles. In India, since 2000, the road length has increased by 62.2%, but on the counterpart the number of vehicles has increased by a staggering 416.5%. Thus, there is gigantic increment in traffic. This situation has given birth to various other issues. One now as to devote more time in traveling than few years back because of traffic which not only increases the utilization natural resources in terms of fuel but also causes air and noise pollution. Likewise, growing traffic is also responsible for various health problems such as irritation, stress, anxiety, migraine and numerous other medical problems. Even the number of accidents due to traffic has also increased. In India, 78% of accidents are caused by carelessness of the driver and almost two-thirds of accidents take place between 9 A.M. to 9 P.M. These evil impacts are unwanted and consequently consideration is required towards this issue. Simply changing to public transport isn't sufficient.

There is a better and an intelligent solution to solve this problem. The traffic signals that control the system are static in nature, they are tuned by past traffic information they are not dynamic enough to change as per the current flow. Because of this the city still requires traffic police officers to manage smooth traffic flow during peak hours. Traffic information that we gather through various identifiers such as cameras are dynamic furthermore, non-stationary and hence expressing the tally of vehicles with higher certainty is a challenging task. Consequently, to solve this issue we have proposed a model that uses RetinaNet for vehicle detection and classification and Long short-term memory (LSTM) for time series prediction, which uses the dataset of Omaha and Oklahoma traffic cameras to detect vehicles with high accuracy and anticipate the traffic flow in future. With the help of this model we can analyze the traffic flow during various times of a day, on weekends or on any particular time of the year. Because of this traffic signals can be synchronized according to the traffic to guarantee the smooth progression of by and large all traffic of the city.

II. RELATED WORK

Kim et al. [1] proposed a model that composed of LSTM and a CNN, which were used for extracting temporal features and image features. The proposed hybrid model outperformed other single models. For vision-based traffic management, Ernst et al. [2] proposed an airborne wide area traffic-monitoring system called LUMOS. An infrared camera was installed on an aircraft to capture and analyse videos from a height of 600 m. For street view traffic surveillance, camera can be installed at either lateral or overhead viewpoint. The lateral- view cameras preserve car shapes better but are more subject to visual occlusions. Shafie et al. [3] developed a smart video surveillance system for vehicle detection and traffic flow control. Their algorithm was tailored to lateral view and the accuracy decreased with an increase of the camera view angle. Their reported false detection rate was 5.81% (or 12.31%) when the camera view angle was 10 (or 20) degree. On the other hand, the overhead-view cameras offer information over multiple lanes but provide less detail on the shape of individual vehicle. Tai and Song [4] counted vehicles by examining a detection window based on a modified histogram. Kiratiratanapruk and Siddhichai [5] computed the optical flow and tracked vehicles between two horizontal detection lines to estimate the traffic flow. In the case of traffic congestion, vehicles moving close to each other tend to combine into a single object in camera perspectives because of visual occlusions. More et al. [6] proposes the use of Artificial Neural Network for controlling road traffic. It also emphasis on use of Jordan sequential network for predicting the future values, depending upon past and current data. Rahman et al. [7] used machine learning techniques to analyze pedestrian and

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 09, September-2019

bicycle crash by developing macro-level crash prediction models. They developed decision tree regression (DTR) models for pedestrian and bicycle crash.

III. PROPOSED SYSTEM

This paper proposes the prediction of traffic volume using LSTM [8] to extract the temporal features of the traffic flow. For achieving high accuracy, features including the speed, category of the vehicle and day of the week are used. The proposed system is divided into four modules:

- Extracting Frames: In this module, the input video that is extracted from the publicly available traffic cameras of Omaha and Oklahoma is read frame by frame by using OpenCV. Each individual frame is then used as an input for the modules that follow.
- Vehicle Detection: The input from the previous module is used to determine the boundary box of the vehicles that are present in the frame. The detection is done using RetinaNet [9] detector. In this module, the vehicle is also classified into one of the following categories: Car, Truck, Bike.
- Vehicle speed detection: In this module, the speed of the vehicle is determined using the difference in the position of the detected vehicles in the individual frames.
- Applying LSTM: The output of the previous modules is augmented with further information about the variables like the day of the week, time stamp and written into a file for further processing. This file is then utilized by the LSTM model for the prediction of the traffic volume over a period of eight minutes.

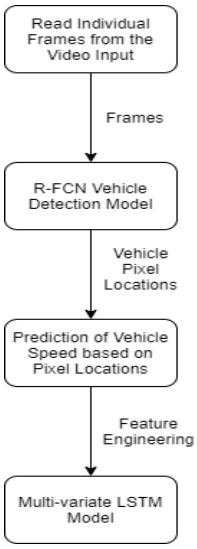

Fig 1. Work flow

IV. DATA INPUT

The data is obtained from the publicly accessible cameras of Omaha and Oklahoma. These cameras capture traffic at several crossroads in the city. They capture the road traffic 24 hours a day, 7 days a week. The cameras have a frame rate of around 20 frames per second and resolution of 352×240. For each camera, the data from 4 cameras was collected for a period of six weeks.

V. WORKING

Detection Model: The dataset consists of congested traffic conditions, low light conditions and a small amount of noise which implies that the detection algorithms need to have a strong classification scope to achieve accurate results. This encouraged us to use RetinaNet which is a one-stage detector and achieves state of the art performance while outperforming two-stage detectors including Faster RCNN [10]. It is a composite network which consists of a backbone network and two sub-networks.

RetinaNet utilizes the Feature Pyramid Network [11] that is built on top of ResNet as its backbone. Feature Pyramid Networks are built in a fully convolutional manner that can take an image of a particular size and output feature maps that are proportionally sized. Doing this reduces the computational cost as different scales of the same image or using multiple feature maps prove to be sub-optimal. The output feature maps are produced at multiple levels where the higher-level feature maps cover a bigger portion of the image and hence are used for detection of larger objects. Whereas, the lower level feature maps are used to to detect smaller objects.

The first subnet is the classification subnet which is a FCN attached to the FPN levels. It is responsible for predicting the probability of the presence of object for A anchors and K object classes. It consists of four 3×3 Conv layers with 256 filters that are followed by ReLU activations. Another 3×3 Conv layer with K×A filters is applied which is followed by sigmoid activation.

The second subnet is a regression subnet which is attached to the feature maps of the FPN in parallel to the classification subnet. It is similar to the classification subnet except the last Conv layer is 3×3 with 4A filters.

The RetinaNet loss function consists of two terms for localization and the other one for classification. It is written as:

$$L = \lambda L_{loc} + L_{cls} \quad (1)$$

where λ Is a hyperparameter for controlling the balance between the losses. The regression loss can be defined as:

$$L_{loc} = \sum_{j\in\{x,y,w,h\}} smooth_{L1}(P_j^i - Tj^i) \quad (2)$$

Where

$$smooth\ L1\ (x) = \begin{cases} 0.5x^2 & for\ |x|<1 \\ |x| - 0.5 & for\ |x|\geq 1 \end{cases} \quad (3)$$

As the smooth L1 loss is less sensitive to outliers, it does not suffer from exploding gradient problem when regression targets are not bounded.

The classification loss is a variant of the focal loss and can be defined as:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 09, September-2019

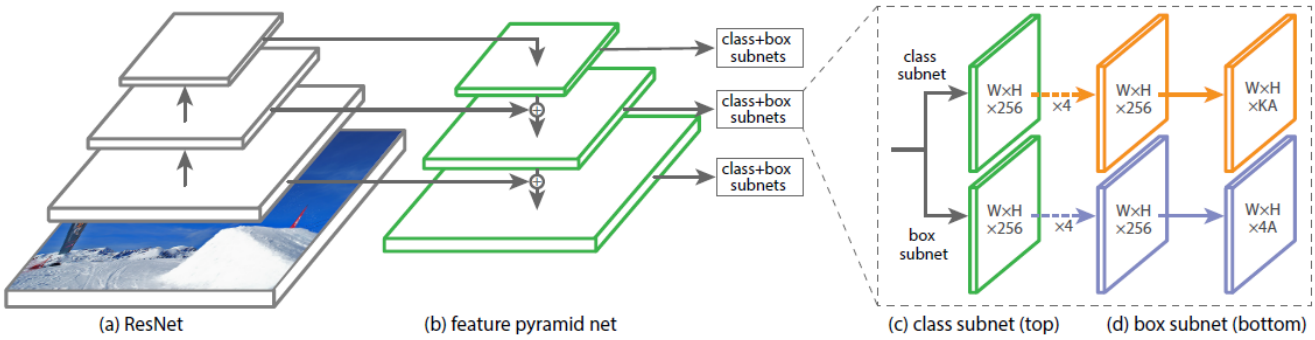

Figure 2: RetinaNet Architecture

$L_{cls}=-\sum_{i=1,k}(y_i \log(p_i)(1-p_i)^{\gamma}\alpha_i + (1-y_i)\log(1-p_i)p^{\gamma}{}_i(1-\alpha_i))$

(4)

Where k = number of classes, y_i is 1 if ground-truth belongs to ith class, p_i indicates the predicted probability of the ith class, γ is a focusing parameter, α is a weighting parameter.

The focusing parameter γ is introduced to reduce the reduce the effects of class imbalance. α is added to address the imbalances in the classes. It can be set as inverse class frequency or the weighting factor α for class 1 and 1-α for class -1. The focusing parameter is used to decrease the losses of easily classified examples. As the value of increases, the network focuses training more on the hard examples.

Feature Engineering: Feature engineering is the process of utilizing domain knowledge to create and manipulate features to increase the predictive power of machine learning models. It is one of the most important steps in solving a machine learning problem as better features will lead to the development of a more accurate model.

In our dataset, there exists a relationship between the time stamp and the number of vehicles. The time stamp format (DD-MM-YYYY HH:MM) makes it difficult for algorithms to establish an ordinal relationship. So, we need to extract the parts of the time stamp into different columns which are more informative for the model. So, we create four features:

- Date [1-31]
- Month [1-12]
- HourOfDay [0-23]
- MinuteOfHour [0-59]

The other features of the dataset are:

- Traffic Flow: Number of vehicles passing the Region Of Interest per hour.
- Speed: The speed of the identified vehicles which is calculated using the location of the vehicle and performing image subtraction.
- Weekday/Weekend [0-1]: Value indicating whether it is the weekend or a weekday.

Traffic Flow Prediction: Recently, Recurrent Neural Networks and Long Short-Term Memory have been applied very successfully to time-series forecasting tasks. The road traffic possesses periodic characteristics as the traffic is likely to be higher during commuting hours or if it is a weekend. Similarly, the traffic flow will be lesser during midnight or early morning. So, in order to predict the flow of traffic in a certain window, the information of earlier windows is vital as

the general patterns of traffic for each day are similar. In this paper, we have developed a LSTM based traffic estimation model that captures the temporal dependencies of the data.

LSTM is chosen over RNNs as they suffer from vanishing/exploding gradients problem. LSTM does not suffer from these problems as it is able to forget previous values and update new memories.

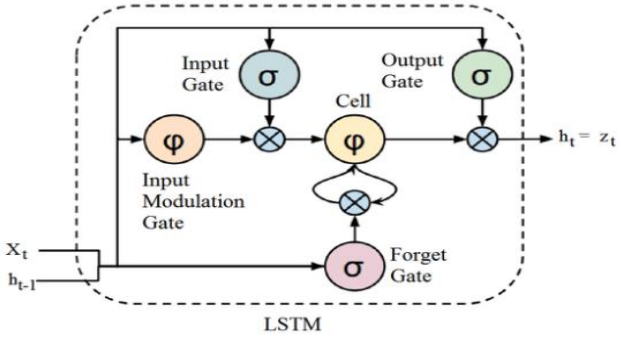

Fig 3. LSTM cell

There are three gates that are used to determine whether the current value has to be forgotten (forget gate ft), the input has to be read (input gate it) or whether the new cell value has to be output (output gate ot). There is a input modulation gate ct. The gates and cell update are defined as follows:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1}) \quad (5)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1}) \quad (6)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1}) \quad (7)$$

$$\sim c_t = \varphi(W_{xc}x_t + W_{hc}h_{t-1}) \quad (8)$$

$$c_t = f_t . c_{t-1} + i_t . \sim c_t \quad (9)$$

$$h_t = o_t . \varphi(c_t) \quad (10)$$

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 09, September-2019

Here, the W matrices are the network parameters. The equations express a multivariate time series with N variables of size T where x = (x_1, x_2, …, x_N) and x_T represents the t-th observation of all variables.

C. Experimental Results:

LSTM Structure		Training Error Values			Test Error Values		
Hidden Layers	No. of neurons	RMSE	MAE	RMSLE	RMSE	MAE	RMSLE
4	16	6.44	4.36	0.64	5.64	4.09	0.56
4	32	6.43	4.35	0.65	5.61	4.08	0.56
6	16	6.65	4.35	0.58	5.81	4.06	0.49
6	32	6.62	4.35	0.6	5.79	4.05	0.48

Table 1

VI. ANALYSIS

In this section, the details of model training, measurements, results and experimental discussion is provided.

A. Training the network:

The LSTM network used in our experiments is composed of hidden layers whose size is 32. The analysis of the number of hidden layers included is in the following section. The activation function used for the hidden layers is tanh function. The parameters are initialized uniformly over the region of [-1.00,1.00] [12]. Adam's algorithm has been used for optimization as the traffic data is noisy and Adam's algorithm is effective on noisy data. The execution is terminated after 15 epochs have been completed. The window size for the LSTM taken as 8 after to incorporate the previous observed values of traffic. The batch size is taken to be 32 and the rest of the hyperparameters are tuned using cross-validation.

B. Evaluation Criteria:

We have used the Root Mean Square Error, Mean Absolute Error, ɔg Error to measure the accuracy of the ı w.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \quad (11)$$

$$\text{MAE} = \frac{\sum_{i-1}^{n}|y_i - x_i|}{} \quad (12)$$

$$\text{RMSLE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\log(x_i) - \log(y_i))^2} \quad (13)$$

Structure: The structure of the LSTM network directly influences the accuracy of the predictions [13] as shown in Table 1. Hence, it is important to determine the best structure for the LSTM network. As shown in the table, 4 hidden layers and 32 number of neurons in hidden layers provide maximum accuracy.

The RMSE of LSTM model is 5.61 on the test dataset which represents a very competitive prediction. The below plot (Fig 4.) shows the predicted values for the training as well as test data and compares it to the truth values.

Moreover, to further illustrate the advantage of using LSTM over other algorithms, an Artificial Neural Network was trained on the same features and the same data. The RMS Error obtained by this method was 12.48. This shows that LSTM networks are a very powerful tool when the temporal dependencies of time-series data need to be acquired.

VII. CONCLUSION

In this paper, we worked on the problem of traffic volume prediction with data from traffic surveillance cameras. We successfully extracted the road traffic information which included the speed and categories of the vehicles using the state-of-the-art RetinaNet detector. After performing feature engineering, we incorporated a LSTM model which was used for forecasting the traffic for short-term as well as long-term periods. This system predicted the traffic with a very competitive root mean square error of 5.61 which indicates that the system can be used effectively in real life scenarios. It has several possible uses including optimization of traffic signal times at crossroads and as a backend route calculator for navigation systems.

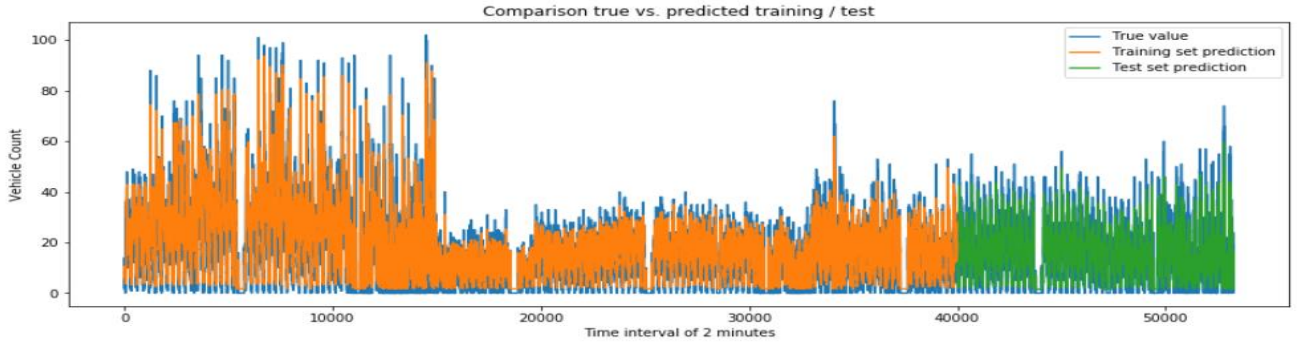

Figure 4

REFERENCES

[1] Kim T, Kim HY: 'Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data'. PLoS ONE 14(2): e0212320, 2019

[2] Ernst, I., Sujew, S., Thiessenhusen, K., Hetscher, M., Rabmann, S., Ruhe, M.: 'LUMOS – Airborne traffic monitoring system'. Int. Conf. Intelligent Transportation System, 2003

[3] Shafie, A., Ali, M., Hafiz, F., Ali, R.: 'Smart video surveillance system for vehicle detection and traffic flow control', J. Eng. Sci. Technol., 2011, 6, (4), pp. 469–480

[4] Tai, J., Song, K.: 'Background segmentation and its application to traffic monitoring using modified histogram'. IEEE Int. Conf. Networking, Sensing and Control, 2004, pp. 13–18K. Elissa, "Title of paper if known," unpublished.

[5] Kiratiratanapruk, K., Siddhichai, S.: 'Vehicle detection and tracking for traffic monitoring system'. IEEE Region 10 Conf. TENCON, 2006, pp. 1–4

[6] Rohan M., Abhishek M., Sheetal R., Rahul B., V K. Pachghare:' Road Traffic Prediction and Congestion Control using Artificial Neural Networks'. 2016 International Conference on Computing, Analytics and Security Trends (CAST) College of Engineering Pune, India. Dec 19-21, 2016

[7] Rahman, M.s., Abdel-Aty M., Hasan S., Cai Q,:'Applying machine learning approaches to analyze the vulnerable roadusers' crashes at statewide traffic analysis zones'. Journal of Safety Research, 2019.

[8] Sepp Hochreiter , Jürgen Schmidhuber, Long Short-Term Memory, Neural Computation, v.9 n.8, p.1735-1780, November 15, 1997

[9] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, ''Focal loss for dense object detection,'' in Proc. ICCV, 2017.

[10] Shaoqing Ren , Kaiming He , Ross Girshick , Jian Sun, Faster R-CNN: towards real-time object detection with region proposal networks, Proceedings of the 28th International Conference on Neural Information Processing Systems, p.91-99, December 07-12, 2015, Montreal, Canada

[11] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, "Feature pyramid networks for object detection", *Proc. 2017 IEEE Conf. Comput. Vision Pattern Recog.*, pp. 936-944, 2017.

[12] Sebastian Raschka, "Python Machine Learning", ISBN 13 9781783555130, September 2015.

[13] Akram A. Moustafa "Performance Evaluation of Artificial Neural Networks for Spatial Data Analysis", Contemporary Engineering Sciences, Vol. 4, No. 4, 149-163, 2011