International Journal of Latest Trends in Engineering and Technology
Vol.(14)Issue(3), pp.001-006
DOI: http://dx.doi.org/10.21172/1.143.01
e-ISSN:2278-621X

SURVEILLANCE VIDEO ANALYSIS USING DEEP LEARNING TECHNIQUES FOR TRAFFIC AND CROWD MANAGEMENT

S. Seema[1], Suhas Goutham[2], Smaranita Vasudev[3], Rakshith R Putane[4]

Abstract- In this digital age, the availability of massive amounts of video data have been exploited by deep learning techniques to gain useful insights. Applying deep learning algorithms on surveillance video data can help in the areas of traffic and crowd management. The model proposed for achieving these objectives is Single Shot MultiBox Detector (SSD) with a line of counting approach to count the objects of interest from a surveillance video. The proposed model has been used for analyzing traffic surveillance videos for counting of vehicles on different lanes and make intelligent traffic decisions to prioritize traffic signals based on the traffic densities. As a subcase of traffic management, a Tesseract OCR model is run on surveillance videos to capture the license plate of vehicles violating any traffic regulations. Another use case proposed in this paper involves studying and analyzing the crowd statistics from publicly accessible surveillance video cameras, to handle crowd management in cases of emergencies and huge public gatherings for safety and security. The need for accuracy along with robustness makes deep learning a suitable choice for the use cases enlisted.

Keywords- Deep learning, Neural networks, Single Shot MultiBox Detector, Line of interest counting, Tesseract, Optical Character recognition, Surveillance videos.

1. INTRODUCTION

With the advent of technology in this digital era, surveillance video cameras have been installed in abundance in various publicly accessible places for monitoring purposes. Analyzing the surveillance video data using deep learning models can help in making informed decisions, with an added advantage of accuracy in prediction. This work focuses on analyzing the surveillance videos captured for predicting a count of the objects of interest that is vehicles or people. The scope of the project encompasses to benefit the traffic officials and the police authorities. The societal impact of traffic management revolves around reducing the manual work of monitoring traffic on roads by using the deep learning model’s output for prioritizing signals and hence, managing traffic. Crowd management would play a vital role in efficient handling of emergencies and crowd statistics.

In the past decade, urban traffic congestion has been a concern for commuters leading to delays in commute time and a hurdle for the traffic officials for managing traffic. Intelligent traffic decisions can be taken by monitoring the vehicles and the congestion through surveillance videos. Using the SSD model with the line of interest counting, traffic monitoring can prove beneficial in efficient traffic signal management and help the traffic officials in prioritizing the traffic signals based on the vehicle count predicted by the deep learning model.

Witnessing the current traffic scenarios, cases of signal jumping have been on a rise in the urban areas. As a part of traffic management, to keep a track of the vehicles which have violated the traffic rules and regulations in terms of signal bypassing, Tesseract OCR (Optical Character Recognition) is used to extract the license plate number of vehicles from the surveillance videos. These license plates can be reported under violation of traffic rules and henceforth, resulting in automating the process of monitoring traffic violation and enforcing stricter and rigid traffic regulations in the cities.

The need to look into the crowd counting problem arises to deal with the real-world scenarios such as crowd control for security purposes and emergency management. The right use of crowd counting techniques can help in taking care of unprecedented events such as strikes, stampedes. In case of emergencies and casualties, people counting can provide a rough estimate of the people in the area and this could lead to easier emergency management decisions.

2. LITERATURE SURVEY

The existing research works aligning with our work have been outlined below. Application of the Fast-RCNN framework for detection of vehicle type in a traffic scene has been proposed by Li Suhao and et al. [1]. To deal with challenging road scenarios and external factors, Xun Li and et al. [2] discusses about the vehicle detection from a traffic video using the Darknet framework, by transforming the object detection problem statement to a binary classification problem and solved using the YOLO-vocRV network(more accuracy over the YOLO v2 model).

Object detection frameworks based on deep learning to handle problems such as clutter, occlusion and low resolution based on different degrees of variations on R-CNN have been outlined in the work of Zhong-Qiu Zhao and et al. [3]. Meng-Ru Hsieh

[1] Professor, Department of Computer Science and Engineering, Ramaiah Institute of Technology, Bangalore, Karnataka, India
[2,3,4] Department of Computer Science and Engineering, Ramaiah Institute of Technology, Bangalore, Karnataka, India

and et al. [4] highlights the drawbacks of regression methods to detect target objects. The dataset used in this work is collected using drones and explains about the usage of deep learning model to detect target objects such as cars by learning from the spatial layout patterns of parking lots.

Xinqing Wang and et al. [5] proposes a new multiple object detection framework in a traffic scenario, namely AP-SSD (Single Shot Multi Box Detector), an improvement to the SSD, by designing a feature extraction convolution kernel library and achieves better results through a dynamic region detection amplification network framework, which improves the recognition accuracy of low-resolution small objects. Peiming Ren and et al. [6] proposes a new model called YOLO-PC (YOLO based People Counting) which is in its capability to ignore irrelevant people in the counting process. Shijie Sun and et al. [7] proposes an efficient method for counting people in real-world cluttered scenes related to public transportations using depth videos. The proposed methodology computes a point cloud from the depth video frame and re-projects it onto a ground plane to normalize the depth information which is later analyzed for identifying potential human heads.

Maksat Kanatov and et al. [8] demonstrates the advantages of using object recognition using deep convolutional neural networks by showing the usage of rectangle filters for object detection to be accurate in person detection from videos and can be evaluated rapidly at any scale. Cong Zhang and et al. [9] proposes a deep convolutional neural network (CNN) for crowd counting, where the network is trained for achieving related learning objectives such as crowd density and crowd count. AMR Badr and et al. [10] introduces an Automatic Number Plate Recognition System (ANPR) using morphological operations, histogram manipulation and edge detection techniques for plate localization and characters segmentation. In order to achieve character classification and recognition, artificial neural networks are also used. The main focus is to experiment deeply and find alternative solutions to the image segmentation and character recognition problems within the License Plate Recognition framework.

Priyanka Prabhakar and et al. [11] presents a License Plate Recognition (LPR) system with the fundamental steps such as detection of number plate, segmentation of characters and recognition of each characters. The system also presents a strong technique for localisation, segmentation and recognition of the characters within the located plate, indicating high accuracy by optimizing numerous parameters that has higher recognition rate than the standard ways.

3. SYSTEM ARCHITECTURE

The system architecture is illustrated in Figure 1. The Android application will be targeted for two kinds of users, namely the traffic and police authorities. This application provides an easy to use interface with options to select the operation to be carried out. The choice of operations include traffic management for traffic signal prioritization, license plate detection (a subcase of traffic management), crowd management and lastly, sending reports such as the vehicle or people count witnessed over a period of time. Based on the user's input of the choice, a request is sent to a Flask server. The Flask server processes this incoming request and runs a test script. Based on the options chosen, surveillance videos are chosen accordingly. For instance, if a traffic management option is chosen, relevant surveillance traffic videos are selected. The video links are sent as input to the deep learning SSD model or to the script running the Tesseract's OCR algorithm based on the operation to be performed.

The SSD model is used for handling two use cases, namely traffic and crowd management. The model is modified to employ the counting approach alongside detecting objects in a video. The model is used for obtaining a vehicle count (for traffic management) or a crowd count (for crowd management). The optical character recognition used by Tesseract is used for license plate detection of vehicles from traffic surveillance videos for monitoring traffic violation, a subcase of traffic management.

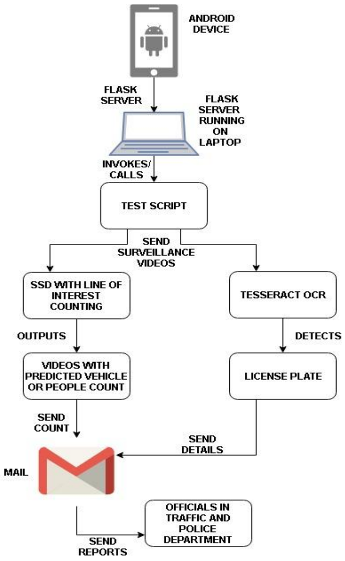

Figure 1. System architecture

4. WORKING

As this application is intended for usage by the traffic and police authorities, for easier monitoring, a mobile application build on an Android platform is developed. The choice of an Android application include easy accessibility, wide usage along with the advantage of an easy-to-use interface. This Android application establishes a connection with a Flask server on a port. The pre-requisite for the choice of port is that no other incoming traffic must be directed to this port on which the Flask server is running.

The Flask server calls a test script which sends surveillance videos as input to the SSD model as per the request sent via the Flask server from an Android client. The model's outputs are sent via email using the Simple Mail Transfer Protocol (SMTP) to the email address registered on the Android client. This summarizes the overall working of sending requests through a server and handling requests by a test script followed by sending the predictions to the user.

5. SINGLE SHOT MULTIBOX DETECTOR MODEL

The deep learning model used in our work is SSD, which leverages deep convolutional neural networks to perform the tasks of classification of objects as well as locating the objects. By localization, it means that SSD provides information regarding where the objects are located. This accuracy is attributed to the mapping of pixels to four floating point numbers. These floating-point numbers represent the bounding boxes. SSD uses VGG16 for the extraction of feature maps, following which object detection takes place using the Conv4_3 layer, where four 3*3 filters are applied. The model makes a total of four predictions at every location (each cell), with each prediction including a bounding box. 21 scores are assigned to 21 classes denoting the proximity of the object to each class. The class with the highest score denotes the class for the object in the bounding box.

In our work, the SSD model is used with Mobile net. The model has been pre-trained on the COCO dataset, which consists of 80 different classes such as people, bicycles, and vehicles and so on. The SSD model is used in conjunction with the line of interest counting approach. Line of Interest (LOI) Counting is used to count the number of objects crossing the line. It differs from the Region-of-interest (ROI) which monitors all locations of a given existing region whereas LOI only needs to monitor the entrances or the exits of a given space.

The objects detected by the SSD model are counted as they pass the line-of-interest set horizontally or vertically at a particular co-ordinate for the video. The line of interest is set horizontally (to a y-co-ordinate) for vehicle counting and is set vertically (to an x-co-ordinate) for crowd counting. This line serves as a reference for counting the objects of interest in a surveillance video.

The SSD model uses OpenCV's libraries to read the surveillance video. The model is tuned to count objects of interest only using the line of interest counting approach. That is, if a given surveillance video consists of various objects in the scene such as dogs, cats and people and vehicles. Given such a video, if one needs to perform vehicle counting, the model ignores all the other objects and only counts the vehicles crossing the given line of interest. The model operates on each frame individually and after processing all frames, a new output video is generated denoting the bounding boxes with the respective class probabilities.

6. OPTICAL CHARACTER RECOGNITION

A deep learning-based text recognition called Optical Character Recognition using Tesseract is used for license plate detection of cars in surveillance videos. In simple terms, Tesseract OCR makes use of a recurrent neural network called Long Short-Term Memory (LSTM) network. The working involves using the OpenCV's libraries to read the input video of vehicles in a scene. The code is written to extract the license plate of vehicles by using the bounding box concept. Each of these detected license plates is passed to the deep learning text recognition algorithm of Tesseract's LSTM. The algorithm's output is the license plate number of the vehicle in the surveillance video.

USE CASES

6.1 SSD model with Mobilenet for traffic management

For cases of vehicle count, the model runs on the videos and generates output videos which shows the count of the number of vehicles detected in a particular surveillance video. Consider a scenario with a traffic signal periodically handling three different lanes, switching signals at every 2-minute intervals, in a round-robin fashion. In this case, three videos are sent as input to the SSD model with the line of interest counting. Three output videos with the respective count of detected vehicles is generated as an output by the model.

The predicted counts of detected vehicles in all the videos is sent as an email using the SMTP protocol to the registered email addresses of the traffic authorities. The mail contains details of the lane number with the associated number of vehicles detected on that lane. Based on this information, traffic authorities can take decisions to prioritize signals based on the lane which shows a periodic higher vehicle count, hence traffic management is dealt.

Based on the higher vehicle count amongst these three lanes, the traffic signals can be prioritized to reduce the traffic congestion on the heavily congested road. Over a period of time, if the traffic official sees a continued congestion on Lane 1 over Lane 2 and Lane 3, traffic prioritization of Lane 1's signal can be considered to reduce the traffic congestion. As shown in Figure 2(a), 2(b) and 2(c), the detected vehicles implies the number of vehicles which have crossed the region of interest line (shown by red in the snapshots).

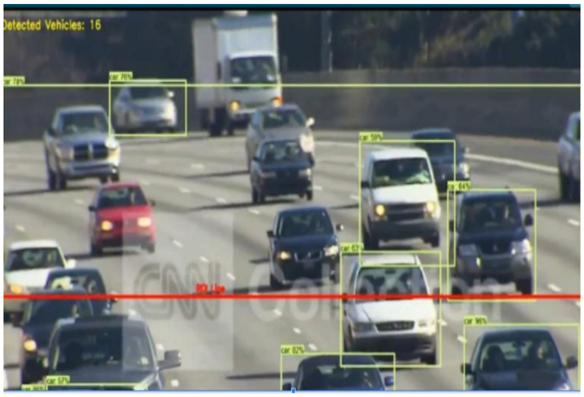

Figure 2(a). Lane 1 with 16 detected vehicles

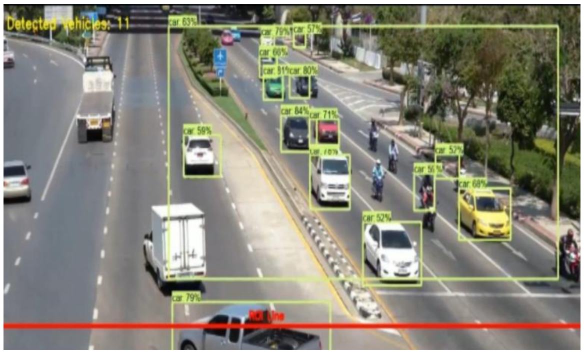

Figure 2(b). Lane 2 with 11 detected vehicles

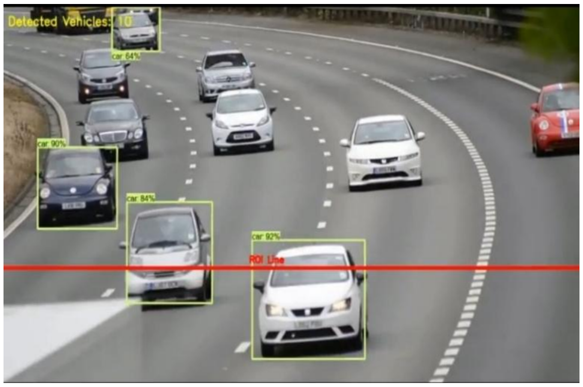

Figure 2(c). Lane 3 with 10 detected vehicles

6.2 Tesseract Optical Character Recognition for traffic violation management

In cases of license plate detection, the optical character recognition techniques used by Tesseract extracts the license plate number of vehicles in a surveillance video. The extracted license plate number can be sent as an email to the traffic officials, thereby automating the process of monitoring road violations, such as signal jumping. Consider a scenario where a vehicle violates the traffic rules and jumps a traffic signal when it is red. To keep a track of these vehicles for violating traffic regulations, the traffic officials can use the model to extract the license plate of the vehicle of interest. This would facilitate an efficient traffic violation management instead of manually noting the number plates of vehicles by seeing surveillance videos. Figure 2(d) shows an image of a car which has violated the traffic signal rules. On applying the model with OCR, the extracted license plate is shown in Figure 2(e).

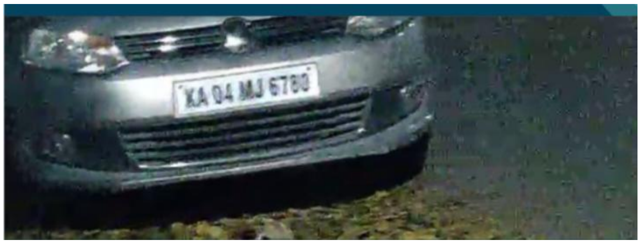

Figure 2(d). A snapshot of one of the frames of the video showing a car.

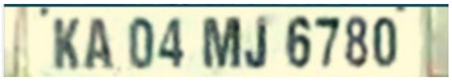

Figure 2(e). The license plate extracted from the video.

6.3 SSD model with Mobilenet for crowd management

For the crowd counting problem, analysis of a surveillance video at the entrance of a building can help in emergency management in case of casualty. Taking an example of any fire in a building, the surveillance video camera installed at the entrance can be considered as an input video and running the YOLO model on this will provide a rough estimate of the people inside the building premises. Based on this count, a mail can be sent to the concerned authorities to operate effectively in case of casualties.

To deal with crowd management for security reasons, consider a public event with multiple gate entries. Consider three different surveillance videos capturing the people entering the event. These videos can be sent as input for the model and the predictions can help in determining the gate which has the most number of people entering, directly implying the need for more security to ensure smooth operation of the event. From Figure 2(f) and Figure 2(g), 11 people exist at time t1. This is how crowd counting can help in emergency management to reduce casualties in a building by an approximate of the people inside a building during any event.

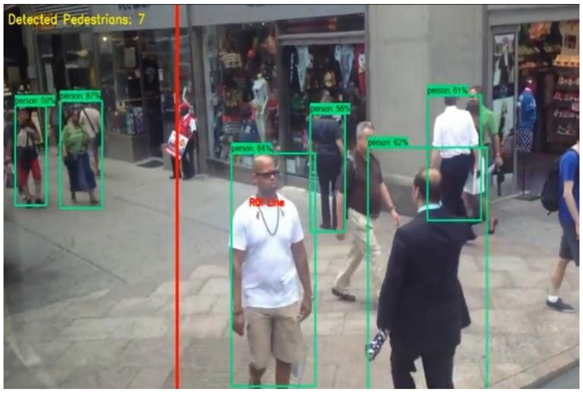

Figure 2(f). Building A’s entrance ‘1’ with a crowd count of 7 people

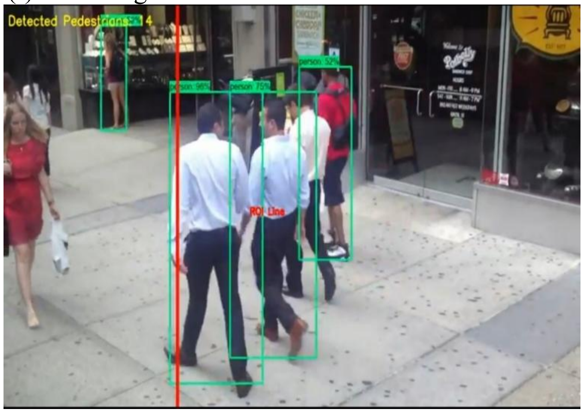

Figure 2(g). Building A’s entrance ‘2’ with a crowd count of 4 people

7. PERFORMANCE EVALUATION

The performance of the model can be evaluated using various metrics. For object counting, the metric of accuracy used is the number of objects detected to the number of objects actually present in the surveillance video. The results for accuracy of vehicle count for traffic management is shown in Table 1 and the results for crowd count are illustrated in Table 2. From Table 1, on an average after running the model on several surveillance videos and noticing similar results, the accuracy of the predicted vehicle count is found to be 78.302%. From Table 2, on an average after running the model on several surveillance videos and noticing similar results, the accuracy of the predicted vehicle count is found to be 85.68%. The accuracy for crowd counting is found to be higher compared to the vehicle count due to heterogeneous factors such as different sizes, shapes of vehicles whereas there is homogeneity in the people feature.

S No.	Actual number of cars in the video	Detected/ Observed cars in the video	Accuracy Percentage (%)
1	9	7	77.77
2	14	11	78.57
3	28	22	78.57
4	46	36	78.26

Table 1. Performance of the SSD model for vehicle counting

S. No.	Actual number of people in the video	Detected number of people in the video	Accuracy Percentage (%)
1	25	22	88
2	38	32	84.21
3	45	39	86.66
4	47	40	85.10

Table 2. Performance of the SSD model for crowd counting

For measuring the performance of license plate detection using the deep learning based optical character recognition, the metric is used is number of mismatched characters in the license plate. Testing on a majority of Indian license plates, observations and results showed that out of the 9 characters visible on the license plate, in 96% of the cases, all the nine characters are predicted correctly. If any noise or disturbance exists in the video, the accuracy drops with a mismatch or error in two of the characters on the license plate.

8. CONCLUSION

From the experimental results and performance evaluation, we can conclude that the deep learning techniques used for traffic management, crowd management and traffic violation management tend to outperform the other existing neural network algorithms. The choice of SSD model over YOLO (You Only Look Once) model is justified as YOLO has a fixed aspect ratio for its grids, which is disadvantageous in YOLO. SSD doesn’t face this problem. Overall, the accuracy and robustness of the SSD model which uses the line of interest counting approach can help in making useful decisions in the areas of traffic and crowd management. Additionally, to handle cases of traffic violation, the Tesseract optical character recognition provides a good accuracy.

Future work would involve improving the accuracy of the model to produce better outputs for vehicle and crowd count under low lighting conditions. Additionally, to improve the license plate detection, the model can be improvised to display the license plates in real-time in a video rather than extracting the license plate from the surveillance video and later using optical character recognition techniques.

9. REFERENCES

[1] Li Suhao, Lin Jinzhao, Li Guoquan, Bai Tong, Wang Huiqian, Pang Yu, “Vehicle type detection based on deep learning in traffic scene,” 8th International Congress of Information and Communication Technology,2018.

[2] Xun Li, Yao Liu, Zhengfan Zhao, Yue Zhang, Li He, “A Deep Learning Approach of Vehicle Multitarget Detection from Traffic Video,” Journal of Advanced Transportation, Volume 2018.

[3] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, Xindong Wu, “Object Detection with Deep Learning: A Review,” Journal Of LATEX Class Files, Vol. 14, No. 8, March 2017

[4] Meng-Ru Hsieh, Yen-Liang Lin, Winston H. Hsu, “Drone-based Object Counting by Spatially Regularized Regional Proposal Network,” August 2017.

[5] Xinqing Wang, Xia Hua, Feng Xiao, Yuyang Li, Xiaodong Hu, Pengyu Sun, “Multi-Object Detection in Traffic Scene Based on Improved SSD”, Published: 6 November 2018

[6] Peiming Ren, Wei Fang, Soufiene Djahel , “A Novel YOLO-based Real-time People Counting Approach”, Third IEEE Annual International Smart Cities Conference 2017.

[7] ShiJie Sun, Naveed Akhtar, HuanSheng Song, ChaoYang Zhang, JianXin Li, Ajmal Mian, “Benchmark data and method for real-time people counting in cluttered scenes using depth sensors” , Journal of LATEX Vol. 14, April 2018.

[8] Maksat Kanatov, Lyazzat Atymtayeva, “Deep Convolutional Neural Network based Person Detection and People Counting System”, 1 Sep, 2018

[9] Cong Zhang, Hongsheng Li, Xiaogang Wang, Xiaokang Yang, “Cross-scene Crowd Counting via Deep Convolutional Neural Networks”, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 12 June 2015.

[10] Amr Badr, Mohamed M. Abdelwahab, Ahmed M. Thabet, and Ahmed M. Abdelsadek, “Automatic Number Plate Recognition System”, Annals of the University of Craiova, March 2011.

[11] Priyanka Prabhakar , P Anupama, S R Resmi, “Automatic vehicle number plate detection and recognition ”, 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), December 2014.