International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-1, November 2018

An Improvement of Traffic Incident Recognition by Deep Convolutional Neural Network

Hoai Nam Vu, Ngoc Hung Dang

Abstract: Traffic incident is the source of many problems that cause economic and human life damages, especially in developing countries. There has been a great amount of research which focuses on early warning traffic incident on the road. Although some researches are able to achieve promising results, the problem of traffic incident detection is still far from completely solved due to the difficult situations such as weather condition, a group of vehicles traveling at the same time. In this paper, we propose a method, which takes advantage of recent deep learning model in vehicle detection and recognition to detect traffic event on separate lanes. Experimental results on real-world dataset prove that the proposed method is effective in locating incidents happing while ensuring real-time scenario of the system.

Index Terms: Deep Convolutional Neural Network, Traffic Incident, Pattern Recognition

I. INTRODUCTION

Traffic problems have become more and more terrible over the world with the highly increasing number of vehicle. Traffic congestion of developing countries is even worth considering due to the low quality of transportation infrastructure and the consciousness of the people. In addition, traffic incident such as illegal parked vehicle, wrong way vehicle, and the irregular pedestrian appearance commonly happens on the road. These events usually cause unexpected occlusion in current direction of the road, and are the main the reason of the other traffic incident. These situations, if detected early, are able to help reduce the traffic jam by firing an alarm about the incident to the other travelers on the road. This kind of solution is commonly mentioned as Automatic incident detection (AID) in intelligent transportation system (ITS).

Recently, AID has gained huge number of research according to its emergency and necessity in modern framework of ITS [1]. The existing AID algorithms in this area are typically divided into direct detection and indirect detection [2]. The first approach determines whether vehicles crash or obstacle occurs based on the information acquired by sensor installed on the road. Although these algorithm are simple, their false alarm rate (FAR) often relatively high, especially in dense traffic at rush hours. The second approach, however, indirectly detects traffic incidents by analyzing collected data from surveillance stations. This approach mainly focuses on traffic data captured from the digital cameras under resource-constraints such as computation time, memory and computation ability. Due to resource limitation, data for training may not be available, the indirect methods are difficult with building accurate models for traffic incident detection. This research area relies on pattern recognition techniques such as neural network, decision tree, and support vector machine for detecting and recognizing type of vehicles on the road. For example, [3] proposed a method using fuzzy inference for solving AID problem. Artificial neural network was implemented in [4] by Dipti, S.. In [5] Xiao et al. built AID system using multiple kernel Support vector machine (SVM). In [6]Renet. al. proposed a method utilizing fuzzy-identification in combination with SVM to detect and position traffic incident and to analysis traffic state. And even a series of hybrid technologies proposed by author of [7] has achieved more and more attention

The recent explosion of big data and high-performance computing brings a completely new solution to this problem of AID. Big data can be collected through various ways such as using wearable devices, using social media, etc. The author of [8] proposed a method that utilizing GPS data to classify anomalous traffic behavior into the different type of traffic incidents, while [9] implemented time and location of traffic congestion detection system by using on-board GPS mounted on probe vehicles. In [10], text data mined from Twitter has been used to extract vehicles incident information both highways and arterials as an efficient and low-cost alternative to traditional data sources. Smart-phone data has become a useful big data in order to detect traffic incident according to the increasing number of phone users while participating in traffic [11]. Although big data has been a trend in computer science with a promising achievement. The big data problem is that a lot of complicated noise may be mixed with the real data. In such a case, the statistical model of noise usually is similar to data, making classifier more difficulty identifying the correct pattern. In [12], a method of utilizing deep CNN model is proposed in order to recognize the traffic incident occurring in various conditions. The result achieved is promising but the system is still failed to detect and classify vehicles when a group of vehicle travels closely. This situation has commonly happened in dense traffic with a high number of vehicles. In this paper, we propose a method that takes advantage of the result from [12] and improves the performance by applying Faster RCNN [13] framework to detect and separate group of vehicles into individual object in dense traffic.

Manuscript published on 30 November 2018.
*****Correspondence Author(s)
Hoai Nam Vu Computer Science Department, Posts and Telecommunications Institute of Techonology, Ha Noi, Viet Nam,
Ngoc Hung Dang, Computer Science Department, Posts and Telecommunications Institute of Techonology, Ha Noi, Viet Nam,

The system has three mains part which are CNN classification, Faster RCNN, and the final result refinement.

The remaining sections of this paper are organized as follows. Section 2 presents our proposed model for traffic incident detection. Experiments are presented in Section 3. We end up with conclusion discussions in section 4.

II. PROPOSED METHOD

In the scenario of traffic surveillance, we calibrate each lane using an array of predefined cell [12] to create inputs for the CNN model and to locate where traffic incident occur. The sizes of the predefined cell are empirically determined according to the camera position and the real size of the vehicles on the road where the system is deployed. Our proposed method has included three main parts, which are CNN (DL Stage 1) for vehicle recognition; faster RCNN (DL Stage 2) for vehicle separation in case of a group of vehicle travels through the predefined cell and cell examination for the final decision. The detail of the system has been shown in Fig. 1.

A. CNN

With the development of computation technology of super network computer, many problems of computer vision and image processing are able to be solved by deep models. There has been several deep learning models proposed recently, such as Recurrent Neural Network (RNN), Deep Boltzmann Machines, Deep Belief Networks and so on. However, one of the most favored deep learning models in image recognition and pattern recognition is CNN. CNN stands for Convolutional Neural Network, represents feed-forward neural network that is the combination of various layers such as the convolutional layers, max pooling layers, and fully connected layers. But what is the main difference between traditional neural network and deep convolutional neural network? Deep convolutional neural network architecture makes explicit assumption that the inputs are images, which allows them to encode certain properties of the image into architecture. Moreover, each neuron in a layer will only be connected to small region of the layer before it, instead of all neurons in traditional fully-connected neural network. That is to say, in traditional fully-connected neural networks, each output of neurons is computed by multiplying entire input V by weights W in that layer. However, in CNN, each output of neuron is computed by multiplying a small local input by the weights W. The weights W are then shared across the entire input. This leads to the fact that neurons that belong to the same layer share the same weights. Weight sharing mechanism is the keypoint of CNNs to make them trainable and efficient model by vastly reducing the amount of parameters used in the network.

Our CNN model setup follows the work that had been done by [12]. Our CNN model consists of 5 blocks of the convolutional layer, ReLU, and pooling layer. They are stacked next to each other to extract useful insights from the input image, introduce non-linearity in our network and make the feature of input image scalable and translation invariant. The final of the network is two fully-connected layers and classification layer. In fully-connected layer, each neuron have full connection to all activations in the previouslayer in the network, as used in the traditional neural network. The last layer of the proposed CNN is SVM classifier. Its function is to predict the accurate class index of the vehicle, human, unknown, and vehicle group.

B. Faster RCNN

After having vehicle group from the CNN classifier of the previous step, we need to separate them into the individual vehicles. The separation process can be done by using an object detection method. Due to the development of computing technology and deep learning, there has been a

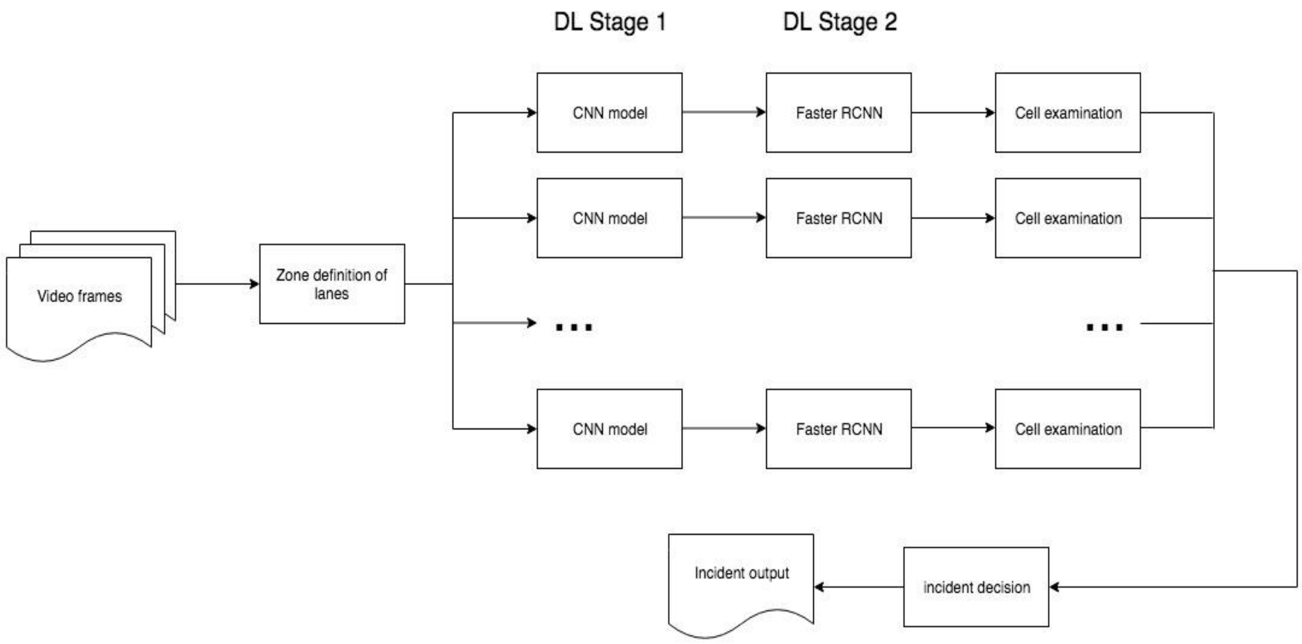

Figure 1. Flowchart of Proposed Method

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

significant improvement in object detection recently. To the best of our knowledge, there are two types of object detector that take advantage of deep convolutional network. These are Fast R-CNN [14] and single shot detector such as YOLO [15]. The former can achieve more accurate results.

Flowchart of proposed method result but not real-time performance, while the latter is super fast but not accurate for detecting the small object. In our proposed system, the detection part is used to separate the group of vehicle into individual one. The detection module is only enabled when a group of vehicles is present in the predefined cell. Then, the detection does not need to be too fast. Therefore, the Faster R-CNN is chosen for use in the proposed system according to its comparable performance. Faster R-CNN is super-accurate and faster than Fast R-CNN to adapt in the real-time system.

Faster R-CNN is an object detection system, which is a combination of two modules. The first one is the region proposal network (RPN) for generating region proposal while the other is a network using these proposals to detect objects. The main difference here with Fast R-CNN is that the later uses selective search to generate the region proposals. The time cost of generating region proposal is much smaller in RPN than selective search, when RPN shares the most computation with object detection network. Briefly, RPN ranks region boxes and propose the ones that most likely contain objects. The output of a region proposal network is a bunch of boxes that will be exanimated by a classifier and regressor to eventually check the occurrence of objects. To be more precise, RPN predicts the possibility of a box being background or foreground, and refine the box. Given the set of object proposal, the Faster R-CNN detector identifies the class label of each region as shown in Fig. 2.

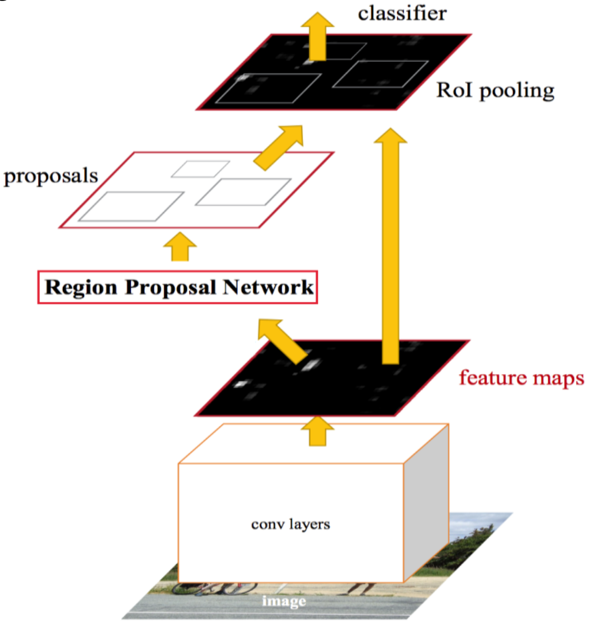

Figure 2. Faster R-CNN

Instead of using the original Faster R-CNN, in our system, only four classes as mentioned are considered. Therefore, our system can meet the real-time processing requirement while still obtain the comparable correct detection rate.

C. Cell Examination

Our system is based on the predefinition of cells on an individual lane; the definition of cells has been completed manually at the initialization phase. The size of the cell based on the pre-setting of digital cameras and the actual size of the vehicle traveling on the road to ensure the results outputted from the CNN and Faster R-CNN model to be highly precious. The system also has to know what is the right way direction of each lane to detect the events related to way direction.

Algorithm 1 Cell Examination

procedure
  begin:
  initialize CNN model
  initialize Faster R-CNN model
  define cell position
  while frame $i^{th} \in$ streaming video **do**
    while lane j^{th} on the road **do**
      if The classification output from CNN is group **then**
        do data association (tracking individual vehicle)
        do checking car pass algorithm on individual vehicle
        do wrong way car detection algorithm on individual vehicle
        return event
      else if the classification output is unknown **then**
        do nothing
        return no event
      else
        do checking car pass algorithm [12]
        do wrong way car detection algorithm [12]
        return event
  end

Figure 3. Cell Examination Algorithm

In the process of analyzing the predefined cell, there will be two cases commonly occurring. The first case is when the output of the CNN model is not a group; the second one is when the output of the CNN model is a group of vehicle. In the first case, our system will perform cell examination as following [12]. In the second case, the output of CNN model is a group of the vehicle traveling through the predefined cell, every cell of the associated lane will be fed into the Faster R-CNN to detect and separate the vehicle in the group. After that, the data association scheme [13] is implemented to track the individual vehicle to get their trajectory. Once the trajectory of the single vehicle is extracted, the analysis steps can be performed in the same scenario as in [12] in order to obtain the traffic incident that happened on the road. Fig. 3 shows the detail of the cell examination algorithm, which implemented in our proposed system.

At the last step, the events discovered through the above algorithms are rechecked and the warning messages are delivered in the form of a message associated with the location of the event. For example, the wrong car incident occurred in lane 1, the illegal parking car occurred in predefined zone 2 in lane 2.

III. EXPERIMENTAL RESULTS

We evaluate the proposed system by using the dataset collected from [12]. The dataset consists of various type of road in Viet Nam, which are highway, urban, and rural. Those videos are recorded under different condition of weather and lightning.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

We chose this dataset in order to compare the performance of our proposed system to the system of [12] and other existing systems. This dataset includes 6 different videos that consist of various traffic incidents to be detected. The resolution of the video is 1920x1080 from the mountable camera on the road.

The performance of the proposed system relies on the object recognition accuracy of CNN and object detection accuracy of Faster R-CNN.

Particularly, for the traffic incident detection problem identified in this study, the accuracy of CNN object recognizer has a significant impact on the accuracy of the traffic incident recognition because each type of traffic events is associated with a particular type of objects. However, this paper focuses on detecting the traffic incident in real-world settings. The accuracy of CNN and Faster R-CNN are not included in the comparison. We only evaluate and compare the accuracy of our proposed system with the others by a number of right detection event. Our system is installed on a workstation, which receives the streaming video from the mountable camera on the road. Our workstations are configured to use Caffe framework with GPU support [16]. Our system can achieve real time performance when working with a camera array up to 16 cameras.

We have also compared our proposed method with the method of [12] on their dataset that mentioned. The author of [12] reported that their method had outperformed the existing system and achieved state-of-the-art results in this problem. In this paper, we addressed their problem in case of vehicle group traveling through the predefined cell. We take advantage of Faster R-CNN object detection to overcome their problem. Therefore, our proposed system is able to outperform the result of [12] by over 5%. We compare these algorithms by using F1 score, which calculated from Precision and Recall as follow:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F1\ Score = 2\ x\ \frac{\text{Precision*Recall}}{\text{Precision+Recall}} \quad (3)$$

Where Precision considers how precise your model is out of the predicted positive, how many of those predicted positive are actually positive. Precision is a good measure to determine when the costs of False Positive is high. Recall actually calculates how many of the actual positives our model captures through labeling it as Positive (true positive). Applying the same understanding, Recall shall be the model metric we use to select our best model when there is a high cost associated with False Negative. In this paper, we use the F1 measure to compare our method to the existing method according to the balance between Precision and Recall. The comparison result is given in Table. 1

Table 1. Comparison

Model	F1 Score	Processing Time per Frame (millisecond)
SVM+PHOG+GMM	71.3%	15
5-layers CNN	83.4%	12
5-layers CNN + Faster R-CNN	**88.5%**	**12.8**

It can be clearly seen from the Table. 1 that our method outperforms the method reported in [12] by more than 5% while still achieving the real-time processing with 12.8ms per frame compared to 12ms/ frame of [12]. The F1 score of our proposed system is high; because we can fix the wrong traffic incident recognition when a group of traveling vehicle appears inside the predefined cell.

We have also given the detailed results of various videos recorded in various weather and lightning condition in the dataset as shown in [12]. The F1 score of our proposed system is still affected somewhat by bad weather such as rain and unstable lightning conditions such as in tunnels, while the score of the system has reached nearly 100% with normal weather conditions of sparse time urban road as well as highway. The worst performance occurs in the video of highway traffic at night when the camera setting has to change to infrared mode. In the infrared mode, the input video is converted to grayscale image instead of RGB image, therefore, making our system difficult distinguishing between vehicle and background road. In addition, the rainy weather has also brought a lot of noise to the input of the proposed system. The random distribution of noise appearing decreases the score of the system. The full result of various videos in the dataset is given in Table 2.

Table 2. F1 Score in Various Conditions

	F1 score of [12]	F1 score of our proposed method
Highway traffic in day time	94%	96%
Highway traffic in nighttime	79.2%	80.7%
Highway traffic in rainy weather	80.1%	83.4%
Highway traffic in sunny weather	96.2%	95%
Urban traffic in sparse time	95.7%	97.5%
Highway traffic in tunnel	82.2%	84.4%

IV. CONCLUSION

In this paper, we proposed an approach, which is the combination of CNN object classifier and Faster R-CNN object detector to efficiently detect traffic incidents under real-world settings.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Experimental results show that the proposed method is able to outperform the previous system by more than 5% while the processing time is still 78 FPS with the hardware configuration (Processor: Intel Core i7, GPU: GTX 1050 Ti, Ram: 8Gb). Our proposed system is able to fix the problem of [12] when a group of vehicles passes through the predefined cell. With the detection accuracy by the F1 score are as high as more than 80% even in nighttime and rainy weather (worst condition), and more than 97% in daytime and sunny weather have demonstrated that our proposed solution is very potential for practical applications such as traffic Surveillance or early traffic incident warning. Our future plan should overcome the challenge of alleviating weather conditions that impact on the system performance, which can be solved by collecting more data of various weather conditions in the training process.

REFERENCES

1. Khattak, A., Wang, X., Zhang, H.: “Incident management integration tool”, IET Intell. Transp. Syst., 2012, 6, (2), pp. 204-214
2. Wang, J., Li, X., Stephen, S., et al.: “A hybrid approach for automatic incident detection”, IEEE Trans. Intell. Transp. Syst., 2013, 14, (3), pp. 1176-1185
3. Ahmed, F., Hawas, Y.E.: “A fuzzy logic model for real-time incident detection in urban road network”. Proc. Int. Conf. on Agents and Artificial Intelligence, Barcelona, Spain, February 2013, pp. 465-472.
4. Dipti, S., Xin, J., Ruey, L.C.: “Evaluation of adaptive neural network models for freeway incident detection”, IEEE Trans. Intell. Transp. Syst., 2004, 5, (1),
5. Xiao, J., Liu, Y.: “Traffic incident detection by multiple kernel support vector machine ensemble”. Proc. IEEE Int. Conf. on Intelligent.
6. Ren, Jianqiang, et al. “Detecting and positioning of traffic incidents via video-based analysis of traffic states in a road segment.” IET Intelligent Transport Systems 10.6 (2016): 428-437.
7. Tian, Q.F., Chen, Y.Z., Zhang, L.G.: “Incident detection in urban freeway traffic based on adaptive algorithm”, J. Trans. Eng. Inf., 2010, 8, (4), pp. 99 - 103, 125.
8. Chakraborty, Pranamesh, et al. Outlier Mining Based Traffic Incident Detection Using Big Data Analytics. No. 17-05869. 2017.
9. Asakura, Yasuo, et al. "Incident detection methods using probe vehicles with on-board GPS equipment." Transportation Research Part C: Emerging Technologies (2016).
10. Gu, Yiming, Zhen Sean Qian, and Feng Chen. "From Twitter to detector: Real-time traffic incident detection using social media data." Transportation research part C: emerging technologies 67 (2016): 321-342.
11. Steenbruggen, John, EmmanouilTranos, and Piet Rietveld. "Traffic incidents in motorways: an empirical proposal for incident detection using data from mobile phone operators." Journal of transport geography 54 (2016): 81-90.
12. Vu N., Pham C. (2018) Traffic Incident Recognition Using Empirical Deep Convolutional Neural Networks Model. In: Cong Vinh P., Ha HuyCuong N., Vassev E. (eds) Context-Aware Systems and Applications, and Nature of Computation and Communication. ICCASA 2017, ICTCC 2017. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 217. Springer, Cham.
13. Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis & Machine Intelligence, (6), 1137-1149.
14. Girshick, Ross. "Fast r-cnn." In Proceedings of the IEEE international conference on computer vision, pp. 1440-1448. 2015.
15. Redmon, J., Divvala, S., Girshick, R., &Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
16. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., ... & Darrell, T. (2014, November). Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM international conference on Multimedia (pp. 675-678). ACM.

Hoai Nam Vu He graduated from Ha Noi University of Science and Technology as electronic engineer (honor) in 2013. He has received Master degree in Electronics and Computer Science department, Chonnam National University, South Korea in 2015. After that, he spent 1 year as research scientist at IVSTech Company, Gangnam, Seoul, South Korea. From 2016, he has been a lecturer in Information Technology department, Posts and Telecommunications Institute of Technology, Ha Noi, Viet Nam. His research area includes image processing, machine learning and document analysis.

Ngoc Hung Dang He received his M.S. degree in Information Technology of Posts and Telecommunications Institute of Technology, Ha Noi, Viet Nam in 2013. From 2013 he has been a lecturer in Information Technology department, Posts and Telecommunications Institute of Technology, Ha Noi, Viet Nam. His research interests include image processing, machine learning and intelligent computing.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)