Traffic Incident Recognition Using Empirical Deep Convolutional Neural Networks Model

Nam Vu(✉) and Cuong Pham

Posts and Telecommunications Institute of Technology, Hanoi, Vietnam
{namvh,cuongpv}@ptit.edu.vn

Abstract. Traffic incident detection plays an important role for a broad range of intelligent transport systems and applications such as driver-assistant, accident warning, and traffic data analysis. The primary goal of traffic incident detection systems in real-world is to identify traffic violations happening on the road in real-time. Although research community has made a significant attempt for detecting on-road violations, there are still challenges such as poor performance under real-world circumstances and real-time detection. In this paper, we propose a novel method which utilizes the powerful deep convolutional neural networks for vehicle recognition task to detect traffic events on the separate lane. Experimental results on real-world dataset videos as well as live stream in real-time from digital cameras demonstrate the feasibility and effectiveness of the proposed method for identifying incidents under various conditions of urban roads and highways.

Keywords: Convolutional neural network · Traffic incident
Vehicle detection

1 Introduction

Automatic incident detection (AID) is a proper solution for traffic incident in intelligent transportation systems (ITS). Recently, AID has attracted research community as it is an indispensable component in modern frameworks of ITS [1]. The existing approaches of AID in literature can be categorized into direct detection and indirect detection [2]. The former determines whether vehicles crash or obstacles occur based on the information acquired by sensors pre-installed on the road. Although these algorithms are simple and relatively effective, their false alarm rate (FAR) is often dependent on the density of the traffic means. The performance is especially not high at rush hours (dense vehicles). The later, however, indirectly detects traffic incidents by analyzing traffic data collected from monitoring stations. This approach might achieve higher detection rate and lower FAR than the former. The detection methods of traffic incident can also be divided into the classical method and the modern intelligent methods.

P. Cong Vinh et al. (Eds.): ICCASA 2017/ICTCC 2017, LNICST 217, pp. 90–99, 2018.
https://doi.org/10.1007/978-3-319-77818-1_9

Classical detection mainly focuses on traffic data captured from digital cameras under resource-constraints such as the limitations of time computation and memory. Due to resource limitation, amount of data might not highly available, classical approaches might struggle with building accurate models for traffic incident detection. Typically AID relies on pattern recognition techniques such as neural networks or support vector machines (SVM). For example, [3] proposed an approach using fuzzy inference for solving AID problem. While artificial neural networks model (ANN) was implemented in [4] by Dipti Work by Xiao et al. [5] built an AID system using multiple kernel Support vector machine (SVM). Somehow similar to [3] on the use of fuzzy logic as inference schema, work by Ren et al. [6] proposed a fuzzy-identification method that is combined with SVM to detect and position traffic incident, and later on for analyzing traffic states. Few works such as [7] exploits hybrid technologies to enhance detection performance.

The recent explosion of big data and sensor technology brings an entirely new approach to this problem. Big data can be collected through various ways such as user-wearable devices, social network data, etc. In [8] GPS data of travelers is utilized to classify anomalous traffic behavior into a different type of traffic incidents, while [9] implemented time and location of traffic congestion detection system by using on-board GPS mounted on probe vehicle. [10] proposed a method applying text mining from Twitter to extract vehicles incident information both highways and arterials as an efficient and low-cost alternative to traditional data sources. With the increasing number of mobile phone users, while participating in traffic, mobile phone usage data has become another source of useful information to detect traffic incidents as reported in [11]. Although big data is becoming a trend in computer science with a huge achievement. In big data, a lot of complicated noise may be included in the real data and need to be handled very carefully.

Although the existing AIDs have made significant progression, the detection accuracies are still not fulfilling for many practical applications. Two main reasons for this challenge are various conditions of lightning, illumination, weather, environment etc. under real-world settings and suitable machine learning techniques. With the rapid increase of computing resources such as GPU and Memory, deep learning approach has become a promising tool effectively used in many computer vision domains including AID. In this work, we propose Convolutional Neural Networks (CNN) model for automatically detecting and positioning of traffic incidents in road lane by using a camera system installed at the traffic lamp sited at the road. Each lane is calibrated using an array of predefined zones to create inputs for the CNN model as well as the position where traffic incidents occur. The sizes of the predefined zones are empirically chosen according to the camera setting position and the real size of the vehicles on the road where the system is deployed. Then, each zone will be considered a single image to feed the CNN model to identify whether there is a vehicle appearing on the predefined zone or not at any time.

The remaining sections of this paper are organized as follows. Section 2 presents our proposed CNN models for traffic incident detection Experiments are present in Sect. 3. We end up with conclusions and discussions in Sect. 4.

2 Convolutional Neural Networks (CNN) Based Approach to Traffic Incident Detection

CNN represents feed-forward neural network which is the combination of various layers such as the convolutional layers, max-pooling layers, and fully connected layers. Recent studies have shown that deep learning has achieved good results in many areas including the image classification field. In this paper, we utilized the deep learning framework for the vehicle classification problem.

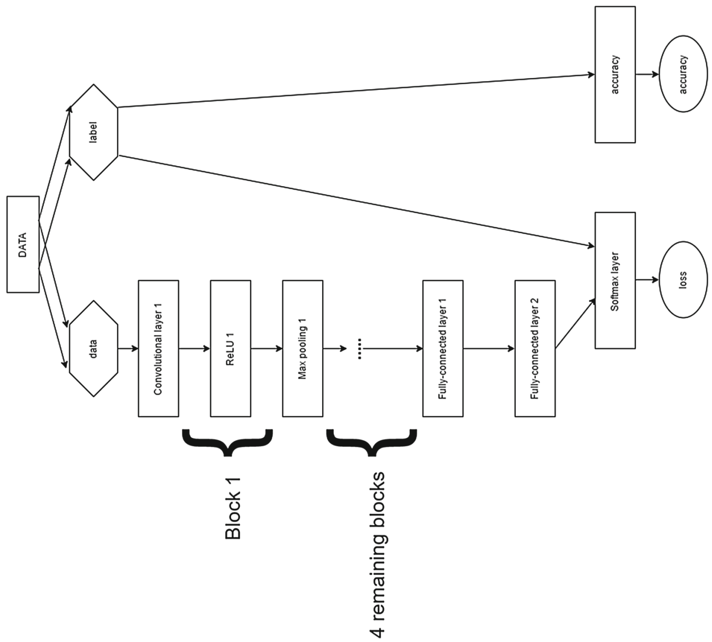

Fig. 1. 5-layers CNN.

2.1 Proposed CNN Model

In this work, we address the problem of traffic incident detection. Firstly, images of the predefined zone are classified into three different classes which are car, human (i.e. motorbike) and unknown objects to verify where vehicles (i.e. car) appear on predefined zone or not. Figure 1 illustrates a 5-layer CNN architecture that takes care of classification task. The final output of CNN model is the probability of pre-defined zone that belongs to one of three classes. These probabilities are so-called confidence scores which are then utilized to verify whether a vehicle passes through the predefined zones. In the field of deep learning, experimentation demonstrates that the deeper the model with as many parameters as possible, the more accurate the results will be. However, the real-time traffic surveillance system always requires the processing time to be fast enough, therefore the 5-layer CNN architecture has been chosen which achieves fast processing time but remaining acceptable accuracy. The proposed CNN model is formed by two convolutional, rectified linear unit (ReLU), max-pooling, fully-connected, dropout and softmax layers.

The convolutional layer is the main part of the CNN model, consisting of a number of trainable filters. Each filter is independently convolved across the input image, calculating the dot product between the entries of the filter and the input image, then producing a feature map of that filter. A convolutional layer of the CNN model helps to detect some specific type of features at some spatial position in the image hierarchically. Therefore CNN model is able to learn feature from simple to complex manner.

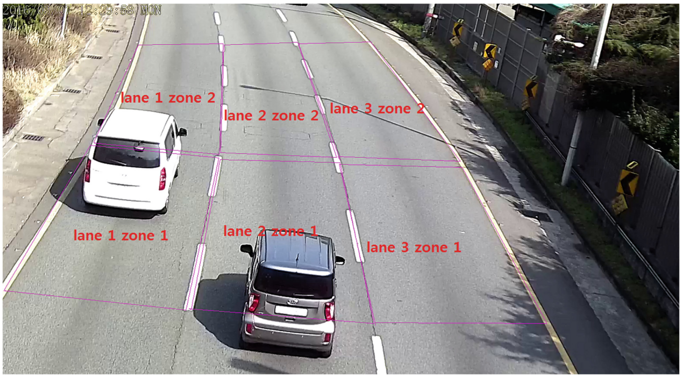

Fig. 2. Lane setting example.

Next to Convolutional layer is a ReLU activation layer, Pooling layer. While ReLU layer will apply an elementwise activation function, where all the negative pixel will be replaced by zero, the Pooling layer will reduce the spatial size of feature maps but remain the most important information.

In this CNN model, we use 5 blocks of the convolutional layer, ReLU, and pooling layers. They are stacked next to each other to extract useful features from the input image, introduce non-linearity in our network and make the feature of input image scalable and translation invariant. The final is two fully-connected layers and softmax layer. In the fully-connected layer, neurons have full connections to all activations in the previous layer, as seen in regular Neural Networks. Their activations can hence be computed with matrix multiplication followed by a bias offset. The last layer of the proposed CNN model is softmax layer. Its function is to predict the accurate class index of the vehicle, human, unknown classes based on the training dataset.

For Convolutional layers in CNN model, the higher layers usually use larger filters to process more complex part of the input image. Therefore we implement 96 filters for convolutional layer 1, 256 filters for convolutional layer 2, 384 filters for convolutional layer 3 and 4, finally 256 filters for convolutional layer 5. The dimension of two fully-connected layers is 4096 and 1000 respectively. The network is trained using mini batches, where each mini-batch contains 30 images of size 227×227. We also minimize the negative log-likelihood using stochastic gradient descent optimizer provided in Caffe.

Experimental results show that when a vehicle passes through a predefined zone, the confidence score increases gradually to the maximum value, thereafter the confidence score decreases gradually to zero. Basing on this observation, a number of the vehicle passing through specific lane can be counted, several traffic illegal incidents also can be detected such as wrong-way vehicle on one direction road, illegal parking or stopping vehicle, and walking person or motorbike enter highway road.

In order to detect wrong-way vehicle in one direction road, we define the array of predefined zones as shown in Fig. 2. If a vehicle goes through 3 consecutive predefined zones in the wrong direction, then that vehicle is alarmed as wrong way event. In order to detect illegal parking or stopping vehicle on the road, every predefined zone is checked whether there is any zone which has the constant confidence score of a car over a period of time. Walking person and motorbike event can be detected by checking every predefined zones whether their confidence score of human or motorbike is over a pre-defined threshold.

2.2 The System for Traffic Incident Detection

Our system is based on the pre-definition of zones on individual lanes; the definition of zones has been completed as given in Fig. 3. The size of the predefined zones based on the presetting of digital cameras and the actual size of the vehicle traveling on the road to ensure the results outputted from the CNN model to be highly precious. In addition, the right direction of each lane is also known in advance to be an input of the wrong way vehicle event examiner. Figure 2 shows an example of lane setting which is defined by the system operator in our real-time monitoring system.

Our system includes three main steps: first of all, each predefined zone image of the specific lane is extracted to be the inputs of CNN model, then the sequence of confidence scores of each class over previous frames are classified into two separate classes which are an event or no event resulted from the Algorithm 1. After that, in order to examine an event of wrong way car, predefined zones neighbor event has to be audited to ensure whether vehicle traveling on the entire lane is wrong way one. The method of detecting wrong way vehicle event in detail is given in Algorithm 2. Finally, based on result given by the previous step, final incident synthesis stage is utilized to give final system decision of event type.

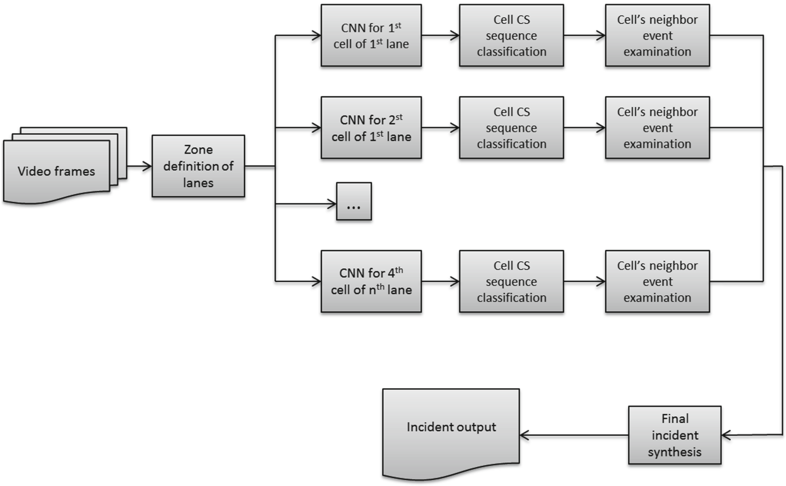

Fig. 3. Flowchart of our proposed method.

Algorithm 1 shows the way the system is implemented to detect whether or not a vehicle passing through each specific lane. This detection process is based on the confidence score of specific class such as car, human and unknown which are the output of CNN model. As given in the algorithm our confidence sequence includes three 10-confidence number arrays of each class. The main task of this algorithm is to examine those arrays separately to detect passing car event happening on each lane. In addition, experimental results show that when a vehicle passes through a predefined zone, the confidence score increases gradually to the maximum value, thereafter the confidence score decreases gradually to zero. Therefore, the examination has been done by using these observations.

Moreover, illegal stopping car and the presence of human or motorbike on the roads which allow only car are able to be detected by algorithm similar to Algorithm 2. In the algorithm, the right direction from predefined zone m to predefined zone n means the real direction which vehicles on the road are allowed to travel, and we assume that the index of predefined zones follows scheme that if a vehicle travels from smaller index zone to larger index zone, that vehicle is considered as wrong way vehicle and an alarm need to be raised. Instead of checking some neighbor predefined zones the algorithm has to check predefined zones individually. When one illegal stopping car or human or motorbike presence events occur in the specific predefined zone in some consecutive frames, then traffic incidents are detected in a specific lane. Illegal stopping car or human and motorbike presence are determined by applying the output of CNN models. Alternatively, the confidence score of human or motorbike as well as car will exceed predefined threshold for each class when human, motorbike and car occurs on the lane. After a while an event of the illegal traffic is detected, a warning message will be sent to the system to inform travelers so that they are able to get appropriate solutions when traveling.

Algorithm 1. Checking car pass

procedure
  begin:
  initialize CCS_Array (Array of car confidence score)
  initialize max _score to 0
  initialize max _threshold to 0.75
  initialize count to 0
  loop through video stream:
  At i^{th} frame:
  Calculate the car confidence score CSSi of the predefined zone.
  if $count < 10$ **then**
    CCS_Array[count] $\leftarrow CCSi$
    $count++$
  else if $count \geq 10$ **then**
    Shift CCS _Array one unit to the left
    CCS _Array[9] $\leftarrow CCSi$
    Checking whether CSS _Array satisfies conditions to be a car pass event
  if CCS _Array increases gradually from left-most to max _score, then decreases to zero **and** $max\_score > max\_threshold$ **then**
    return car pass event
  else
    return no event

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.

3 Experimental Results

In order to evaluate the effectiveness of the proposed system, a set of the testing video is collected on the highway, urban and rural road of Viet Nam. Those videos are recorded under different condition of weather and lightning to accurately evaluate and compare proposed system with other existing systems. This data set includes 6 different videos which consist 66 to 81 traffic event to be detected. The resolution of the video is 1920×1080 with a mountable camera.

It is understandable that the performance of the proposed CNN model relies on object recognition accuracies. Particularly, for the traffic incident detection problem addressed in this study, the accuracy of the object recognizer has a great influence on the accuracy of traffic event identification because each type of traffic events is associated with a particular type of objects. However, this article focuses on detecting traffic events under real-world settings. The accuracy of the CNN identifier is not included in the comparison. Our system is installed using road cameras and video streams are processed on a workstation. Workstations are configured to use Caffe framework [12] with GPU support. Our system is able to work in real-time to monitor several cameras at the same time on the certain monitored track.

Algorithm 2. Wrong way car detection

procedure
  begin:
  Initialize *iEvent_Array* (Array of frame number when event happened)
  Set nFrame to 20
  Assume that the right direction is from zone m^{th} to zone n^{th} $(m > n)$
  while frame i^{th} $\in$ streaming video **do**
    while lane j^{th} on the road **do**
      while $z \in \{1, 2, 3\}$ **do**
        Checking car pass from algorithm 1
        if Checking car pass return event **then**
          iEvent_Array[z] $\leftarrow i$
        if zone z^{th} has passed car zone **and** (i − *iEvent_Array*[z-1]) $<$ $nFrame$ **then**
          return Wrong way event

Table 1. Accuracy comparison

Model	Accuracy	Processing time per frame (millisecond)
SVM + PHOG + GMM	74.5%	15
5-layers CNN	**89.6%**	**12**
GoogleNet CNN	86.4%	86.4
GoogleNet CNN (fine-tune)	89.8%	86.4

We have also compared our proposed method with the state of the art methods as illustrated in the Table 1. Prior to deep learning, the best results in the detection of traffic incidents often rely on manual feature extraction such as PHOG [14–16] and the SVM for object classification. The inputs of the stages based on results derived from the background modeling algorithm. Most of the algorithms used to rely on GMM [17] for background modeling to ensure real-time characteristics. Therefore, these algorithms are heavily influenced by weather conditions because when the weather changes, the conditions of the background also vary. As indicated in the Table 1, the best algorithm is that PHOG + SVM only achieves more than 70% while CNN models show their outperformance at approximately 90%. The CNN model shows that due to the lack of training data in the system, the fine-tune GoogleNet model yielded the highest result with 89.8%. However, our 5-layers model can achieve the approximate result with the faster training and testing time. As shown in Table 1 5-layers CNN model can process one frame within 12 ms compared to more than 80 ms of GoogleNet model. If the big data for training is available our model would yield the same result reported using GoogleNet fine-tune model.

As seen in Table 2, the accuracy of the algorithm is also affected somewhat by bad weather such as rain and unstable lighting conditions such as in tunnels, while the accuracy of the system has reached nearly 100% with normal weather conditions of sparse time urban road as well as highway. The worst performance video is highway traffic in the nighttime as camera setting has to change to infrared mode. In infrared mode, the input image is in the form of grayscale image instead of RGB image, therefore, making our system difficult distinguishing between vehicle and background road.

Table 2. Accuracy in various conditions

	Events detected	Total number of events	Accuracy
Highway traffic in day time	68	70	97%
Highway traffic in night time	65	80	81.3%
Highway traffic in rainy weather	67	81	82.7%
Highway traffic in sunny weather	65	66	98.5%
Urban traffic in sparse time	80	81	98.7%
Highway traffic of tunnel	59	70	84.3%

4 Conclusion

This paper proposed an approach which utilizes deep convolutional neural network to detect traffic incidents under real-world settings. Experimental results show that the proposed method is able to outperform previous methods by more than 10% while processing time is 83 FPS with my hardware configuration (Processor: Intel Core i7, GPU: GTX 1050 Ti, Ram: 8 GB). With the detection accuracies are as high as more than 81% even in nighttime and rainy weather (worst condition), and 97% in daytime and sunny have demonstrated that our proposed solution is very potential for practical applications such as driver-assistant or accident warning. Our future work should overcome few limitations such as improving accuracies of detecting and counting the group of vehicles which pass through predefined zones. Another challenge such as alleviating weather conditions impacting on the system's detection accuracies, which can possibly be solved by acquiring more data of weather conditions in training process.

References

1. Khattak, A., Wang, X., Zhang, H.: Incident management integration tool. IET Intell. Transp. Syst. **6**(2), 204–214 (2012)
2. Wang, J., Li, X., Stephen, S., et al.: A hybrid approach for automatic incident detection. IEEE Trans. Intell. Transp. Syst. **14**(3), 1176–1185 (2013)

3. Ahmed, F., Hawas, Y.E.: A fuzzy logic model for real-time incident detection in urban road network. In: Proceedings of International Conference on Agents and Artificial Intelligence, Barcelona, Spain, pp. 465–472, February 2013
4. Dipti, S., Xin, J., Ruey, L.C.: Evaluation of adaptive neural network models for freeway incident detection. IEEE Trans. Intell. Transp. Syst. **5**(1), 1–11 (2004)
5. Xiao, J., Liu, Y.: Traffic incident detection by multiple kernel support vector machine ensemble. In: Proceedings of the IEEE International Conference on Intelligent (2012)
6. Ren, J., et al.: Detecting and positioning of traffic incidents via video-based analysis of traffic states in a road segment. IET Int. Trans. Syst. (ITS) **10**(6), 428–437 (2016). Transportation Systems (ITS), Anchorage, USA, September 2012
7. Tian, Q.F., Chen, Y.Z., Zhang, L.G.: Incident detection in urban freeway traffic based on adaptive algorithm. J. Trans. Eng. Inf. **8**(4), 99–103, 125 (2010)
8. Chakraborty, P., et al.: Outlier Mining Based Traffic Incident Detection Using Big Data Analytics. No. 17-05869 (2017)
9. Asakura, Y., et al.: Incident detection methods using probe vehicles with on-board GPS equipment. Transp. Res. Part C: Emerg. Technol. **81**, 330–341 (2016)
10. Gu, Y., Qian, Z.S., Chen, F.: From Twitter to detector: real-time traffic incident detection using social media data. Transp. Res. Part C: Emerg. Technol. **67**, 321–342 (2016)
11. Steenbruggen, J., Tranos, E., Rietveld, P.: Traffic incidents in motorways: an empirical proposal for incident detection using data from mobile phone operators. J. Transp. Geogr. **54**, 81–90 (2016)
12. Jia, Y., et al.: Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia. ACM (2014)
13. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015)
14. Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. In: Proceedings of the 6th ACM International Conference on Image and Video Retrieval. ACM (2007)
15. Pham, C., et al.: FoodBoard: surface contact imaging for food recognition. In: Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM (2013)
16. Vu, H.N., Tran, T.A., Na, I.S., Kim, S.H.: Automatic extraction of text regions from document images by multilevel thresholding and k-means clustering. In: 2015 IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS), pp. 329–334. IEEE, June 2015
17. Zivkovic, Z.: Improved adaptive Gaussian mixture model for background subtraction. In: Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, vol. 2. IEEE (2004)