IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⋇ Impact Factor 7.39 ⋇ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11788

Vehicle Accident Detection using alternating Convolutional and Max-pooling layers in a CNN

Nishanth Rao[1], **Vinodakumar**[2]

Student, Information Science and Engineering, R.V. College of Engineering, Bengaluru, India[1]

Student, Information Science and Engineering, R.V. College of Engineering, Bengaluru, India[2]

Abstract: According to the WHO, approximately 1.3 million people die every year due to traffic related accidents. One of the factors determining survival probability is the response time of the paramedics. This variable depends on the actions of onsite pedestrians or the person responsible for traffic proceedings at the site of accident. We propose the use of a CNN with alternating max-pooling layers to classify images as accident and non-accident to automatically detect an accident over CCTV camera footage. One problem plaguing this field is the lack of available accident datasets, so we prepare the dataset from various sources. The results (training accuracy of 93.78% and validation accuracy of 82.65%) show that the use of alternating max-pooling layers help in accident detection even when the vehicle is not in the centre of the frame.

Keywords: accident detection, convolutional neural network, image classification, deep learning

I. INTRODUCTION

Road traffic accidents are more prevalent in developing countries than developed ones. They lead to considerable loss, not only in terms of life but also economically. According to the WHO statistics [10], the risk factors include – speeding, driving under the influence of psychoactive substances, unsafe road infrastructure, inadequate law enforcement, etc. People from poorer socioeconomic origins are more likely to be involved in traffic accidents, even in high-income countries. The leading cause of death for children and young adults aged 5-29, again according to [10], is road traffic injuries.
One of the variables effecting the survivability of the accident victim is the time taken by the paramedics to reach the scene of the accident. This variable depends on several factors, such as traffic density, time taken by the pedestrians on site to call for help, alertness of the traffic, if any.
Various methods have been proposed by several authors for automatic detection of accidents. Some involve localizing the area of accident before classifying. Others classify the entire image as ‘accident’ or ‘non-accident’. The problem faced by the implementors of the second method is that the CNN model trained often times struggles to detect accidents when not in the middle of the frame, owing to the lack of quality data in this domain. In this paper, we train a CNN for image classification, classifying the images as ‘accident’ or ‘non-accident’. The use of alternating max-pooling layers is predicted to mitigate the drop in accuracy when dealing with accidents that are not in centre of the frame.

II. LITERATURE SURVEY

Ankit Shat et al [1] worked on resolving the lack of public data for research on automatic detection and prediction of accidents by providing a spatiotemporally annotated dataset. The authors propose the use of a Faster-RCNN with LSTM along with Context Mining for accident prediction. The use of CM and ACM compliments the accuracy for small pedestrian detection. They achieve an average precision of 47.25% but more importantly create a dataset for future projects.

E. P. Ijjina et al [2] propose a framework that capitalizes on Mask R-CNN for accurate object detection followed by an efficient centroid based object tracking algorithm for surveillance. The probability is determined based on speed and trajectory anomalies in a vehicle after an overlap with another vehicle. The framework was evaluated on diverse conditions such as broad daylight, low visibility, rain, hail, and snow using the proposed dataset. The framework achieves a detection rate of 71% and a false alarm rate of 0.53%. The model suffers from ineffectiveness in high density traffic due to inaccuracies in vehicle detection and tracking.

Aparajith Srinivasan et al [3] present a method for detecting accidents using DETR (Detection Transformers) and Random Forest Classifier. Objects are first detected using the DETR and the features are fed to a Random Forest Classifier for frame wise classification. Each frame is classified as either accident or non-accident. A total count of predicted

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⋇ Impact Factor 7.39 ⋇ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11788

accident frames from any 60 continuous frames of the video are considered using a sliding window technique before the final decision is made. Simulation results show that the proposed system achieves 78.2% detection rate in CCTV videos.

E.P Ijinna at al [4] propose a system to overcome the challenge proposed by limited regions of surveillance. They propose an approach for detecting accidents from dashboard camera video using computer vision-based techniques. The variation in visual information during accidents is analysed to design a computationally less expensive accident detection model for practical use. As a result, accidents are identified from the change in visual information during an accident. It is evaluated on the new Dashboard Video Accident Detection (DVAD) dataset proposed by the authors.

Sergio Robles-Serrano et al [5] propose an automated DL-based method capable of detecting traffic accidents on video. The proposed method assumes that traffic accident events are described by visual features occurring through a temporal way. Therefore, a visual features extraction phase, followed by a temporary pattern identification, compose the model architecture. The visual and temporal features are learned in the training phase through convolution and recurrent layers using built-from-scratch and public datasets. An accuracy of 98% is achieved in the detection of accidents in public traffic accident datasets, showing a high capacity in detection independent of the road structure.

K. B. Lee et al [6] propose the use of Object Detection and Tracking System (ODTS) in combination with FRCNN for monitoring of unexpected events using a CCTV camera in tunnels. A deep learning model in ODTS was trained with a dataset of event images in tunnels to achieve Average Precision (AP) values of 0.8479, 0.7161 and 0.9085 for target objects: Car, Person, and Fire, respectively.

Dinesh Signh et al [7] propose a framework which automatically learns feature representation from the spatiotemporal volumes of raw pixel intensity instead of traditional hand-crafted features. The proposed framework extracts deep representation using denoising autoencoders trained over the normal traffic videos. The possibility of an accident is determined based on the reconstruction error and the likelihood of the deep representation. The intersection points of the vehicle's trajectories are used to reduce the false alarm rate and increase the reliability of the overall system. The proposed method is able to detect on average 77.5% accidents correctly with 22.5% false alarms on real accidents videos captured under various lighting conditions.

E. Batanina et al [8] implement a deep learning model for car accident detection using synthetic videos while adapting the model, using domain adaptation (DA), to real videos from CCTV traffic cameras. The synthetic data are rendered using a video game. This allows the authors to overcome the lack of data available.

S. Veni et al [9] propose a method where the accident is detected by the dispersion in the motion field of the vehicles during collision. Motion field of the road is obtained from the optical flow of the video frames. The moving objects in the frames are segmented and tracked. The dispersion in the angle vector of the optical flow is derived for each of the moving objects. The dispersion of angle vector for each object is monitored, and deviation of the same from the threshold is determined as an accident. The harshness of the accident can be found by the range of dispersion of the motion field.

III. DATA PREPARATION

Images of varying resolutions and quality were hand-picked from different sources including other datasets such as CADP for accident images, DETRAC for non-accident images. The CADP is a novel accident dataset by [1], which is spatio-temporally annotated. It proves to be one of the few quality datasets available for the task. The annotations in this dataset are of no real concern to the CNN model are therefore not used. DETRAC is a vehicle detection dataset which consist of a large number of images taken on highways, lanes etc. Some images were also picked from general sources like google images. These images were labelled as accident or non-accident.
All 3 channels of colour are used and therefore there is no conversion of images to grayscale. The images are regularized to 250 x 250 dimension and the colour values are normalized. Figure 1 is an example image from the dataset once it has been pre-processed.

DOI: 10.17148/IJARCCE.2022.11788

Figure 1: Sample Image corrected to 250x250

IV. ARCHITECTURE

A CNN is constructed with alternating convolutional layers (with increasing feature count or 32,64,128) and max pooling layers, followed by flattening and a fully connected dense layer (256). 'Adam' optimizer is used to calculate the accuracy metric and 'binary crossentropy' is used as the loss function for binary classification of the images. The CNN is constructed using Keras (Tensorflow). The CNN is trained of the aforementioned dataset. The probability of the selected class by the CNS is displayed over the image. Live video or stored video clips may be fed to the model as frames and the model runs on a per frame basis to detect accidents.

V. RESULTS

The results for training and validation are acquired during the training of the model. As seen in Figure 2, the CNN acquired a final training accuracy of 93.78% and a validation accuracy of 82.65%. Figure 3 shows the loss. The model was trained only for 10 epochs as the available data is very scarce. Further training would lead to overfitting.

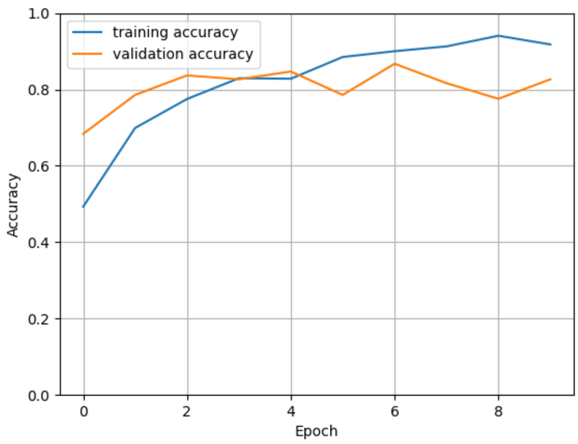

Figure 2: Training and Validation Accuracy

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁑ Impact Factor 7.39 ⁑ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11788

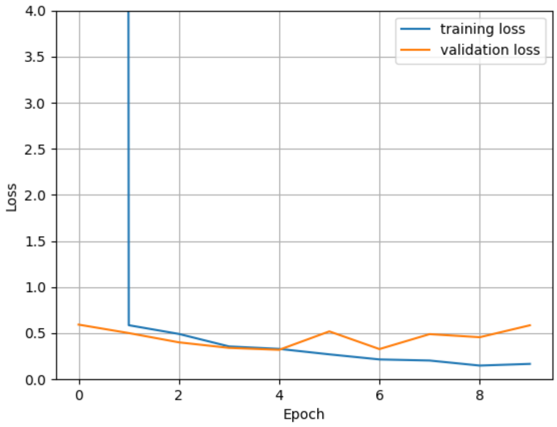

Figure 3: Training and Validation Loss

A randomized batch of testing data was formed for checking the testing accuracy. Since the batch size is 64, 5 randomized batches were made. Figure 4 shows the results of testing.

Figure 4: Results of testing on a batch

The following table shows the results over 5 such batches –

Table 1: Results from 5 batches

Batch Number	Accuracy %
1	92.18
2	89.06
3	90.625
4	90.625
5	89.06

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⋇ Impact Factor 7.39 ⋇ Vol. 11, Issue 7, July 2022
DOI: 10.17148/IJARCCE.2022.11788

VI. CONCLUSION

Computer Vision is a rapidly growing field and there exists several methods for solving tasks of classification. As it stands, the model on its own cannot be used in real life situations, without a human operator to mitigate errors, because the problem demands an accuracy of upwards of 99% with next to zero percent false calls. The max-pooling layers seem to have mitigated the need for object localization to some extent, as evident from the fact that the same accuracy is maintained, even if the accident is in different part of the frame. With a more rigorous and diverse dataset, the CNN's accuracy and the environment under which it gives satisfactory results may improve. Using spatiotemporally annotated data with an RNN may lead to a more natural solution to the problem. One way to improve the current methodology would be to use a probability heat map with the CNN to display the area of interest, i.e., the accident.

REFERENCES

[1] A. P. Shah, J. Lamare, T. Nguyen-Anh and A. Hauptmann, "CADP: A Novel Dataset for CCTV Traffic Camera based Accident Analysis," 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2018, pp. 1-9, doi: 10.1109/AVSS.2018.8639160.

[2] E. P. Ijjina, D. Chand, S. Gupta and K. Goutham, "Computer Vision-based Accident Detection in Traffic Surveillance," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019, pp. 1-6, doi: 10.1109/ICCCNT45670.2019.8944469.

[3] A. Srinivasan, A. Srikanth, H. Indrajit and V. Narasimhan, "A Novel Approach for Road Accident Detection using DETR Algorithm," 2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA), 2020, pp. 75-80, doi: 10.1109/IDSTA50958.2020.9263703.

[4] E. P. Ijjina and S. K. Sharma, "Accident detection from dashboard camera video," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019, pp. 1-4, doi: 10.1109/ICCCNT45670.2019.8944520.

[5] Robles-Serrano S, Sanchez-Torres G, Branch-Bedoya J. Automatic Detection of Traffic Accidents from Video Using Deep Learning Techniques. *Computers*. 2021; 10(11):148. https://doi.org/10.3390/computers10110148

[6] K. B. Lee and H. S. Shin, "An Application of a Deep Learning Algorithm for Automatic Detection of Unexpected Accidents Under Bad CCTV Monitoring Conditions in Tunnels," 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), 2019, pp. 7-11, doi: 10.1109/Deep-ML.2019.00010.

[7] D. Singh and C. K. Mohan, "Deep Spatio-Temporal Representation for Detection of Road Accidents Using Stacked Autoencoder," in IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 3, pp. 879-887, March 2019, doi: 10.1109/TITS.2018.2835308.

[8] E. Batanina, I. E. I. Bekkouch, Y. Youssry, A. Khan, A. M. Khattak and M. Bortnikov, "Domain Adaptation for Car Accident Detection in Videos," 2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA), 2019, pp. 1-6, doi: 10.1109/IPTA.2019.8936124.

[9] Veni, S., Anand, R., Santosh, B. (2021). Road Accident Detection and Severity Determination from CCTV Surveillance. In: Tripathy, A., Sarkar, M., Sahoo, J., Li, KC., Chinara, S. (eds) Advances in Distributed Computing and Machine Learning. Lecture Notes in Networks and Systems, vol 127. Springer, Singapore, doi: 10.1007/978-981-15-4218-3_25

[10] World Health Organization, "*Road Traffic Injuries*". https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries (Accessed: 20 June 2022)