

Detecting Distracted Driving with Deep Learning

Ofonime Dominic Okon and Li Meng^(✉)

School of Engineering and Technology, University of Hertfordshire,
Hatfield, UK

l.l.meng@herts.ac.uk

Abstract. Driver distraction is the leading factor in most car crashes and near-crashes. This paper discusses the types, causes and impacts of distracted driving. A deep learning approach is then presented for the detection of such driving behaviors using images of the driver, where an enhancement has been made to a standard convolutional neural network (CNN). Experimental results on Kaggle challenge dataset have confirmed the capability of a convolutional neural network (CNN) in this complicated computer vision task and illustrated the contribution of the CNN enhancement to a better pattern recognition accuracy.

Keywords: Distraction detection · Accident prevention · Convolutional neural networks · Kaggle challenge · Triplet loss

1 Introduction

Driving is a complex task and requires a number of skills such as cognitive skills, physical fitness, coordination and, most importantly, attention and concentration of the driver on the driving [1, 2]. Despite of the complex nature of driving, it is common of drivers to get involved in activities that divert their full attention from driving, degrade their driving performance and even lead to fatal accidents. Typical examples of such activities include using a mobile phone, eating or drinking, using a navigation device, grooming, tuning the audio system, and/or talking to passengers, etc. In a report by the National Highway Traffic Safety Administration (NHTSA), it has been estimated that approximately 25 percent of car accidents were due to inattention of drivers [3] and around 50 percent of these accidents were caused by distraction of drivers [4, 5].

With the goal of reducing car accidents and improving road safety, various computer vision based approaches have been proposed. State Farm has initiated a competition called Kaggle competition, which aims to distinguish distracted driving behaviours from safe driving using images captured by a single dashboard camera. This paper presents a solution to the Kaggle challenge by using the latest development in machine learning and computer vision, i.e. deep learning and a convolutional neural network (CNN).

The paper is organized as follows. Section II provides a more in-depth description of the subject of distracted driving. Section III presents the existing computer vision based approaches to the detection of distracted driving. Section IV provides a brief subject review of deep learning and CNNs as well as a detailed description of the CNN

we have adopted for the Kaggle challenge. Furthermore, section IV presents the details about the Triplet loss for the improvement in the accuracy of deep learning classification. Section V explains the Kaggle challenge, describes our experimental setup and compare the results of our two CNN models on the Kaggle images. Finally, Section VI concludes the paper and highlights some remaining challenges.

2 Distracted Driving

Distracted driving is a type of inattention. It has been defined by the American Automobile Association Foundation for Traffic Safety (AAAFTS) as the slow response of a driver in recognizing the information required to complete driving task safely due to some event within or outside the vehicle, which causes the shift of driver attention from the driving task [1, 4, 6]. Distracted driving can be categorized into four main types; visual distraction, auditory distraction, cognitive distraction and biomechanical distraction [7]. Visual distraction is the diversion of driver's visual field while looking within or outside the vehicle to observe any event, object or person [8]. Cognitive distraction is defined as diversion of thoughts from driving due to thinking about other events [9]. Auditory distraction is defined as diversion from driving due to the use of a mobile phone, communicating with other passengers or any other audio device [9]. Biomechanical distraction is diversion due to physical manipulation of objects instead of driving [10]. It is important to note that although distraction is categorized into four different types they do not occur individually but are usually linked with each other. For example, in the activity of answering an incoming call all four types of distractions can be observed: visual distraction when looking at the phone screen to interpret the phone alert and to locate the right button(s) to press; auditory distraction when hearing the alert and when being in the conversation; physical distraction when taking a hand off the wheel to press a button to receive the call; and cognitive distraction when diverting thoughts to the topic of conversation.

A research by the National Highway Traffic Safety Administration NHTSA stated thirteen different sources of distraction, which can be further categorized into technology based, no-technology based and miscellaneous sources [4]. Table 1 presents the common sources of distracted driving as identified by the NHTSA. As shown in Table 1, some technical enhancements in modern vehicles, such as the navigation system and the entrainment system, on one hand are assisting drivers in many ways but on the other hand have become sources of distraction to drivers. Furthermore, it has been predicted by Stutts et al. [11] that number of distraction-related accidents will increase with the enhancements of vehicle technologies.

Studies have been carried out to investigate the impact of distracted driving to car crashes. Stutts et al examined the Crashworthiness Data System gathered from 1995 to 1999 to identify the contribution of different distractions to accidents [11]. Glaze and Ellis focused their study on the distraction sources from within the vehicle and investigated their contributions to car accidents based on the troopers' crash record [12]. Table 2 presents a comparison of the outcomes of these two studies.

Table 1. Different sources of distraction in drivers categorized by NHSTA [4]

Type of distraction	Source of distraction
Technology based distraction	Operating radio of music devices
	Talking or listening on mobile phone
	Dialing mobile phone
	Adjusting climate controls
	Using device/object brought into vehicle
	Using device/controls integral to vehicle
Non-technology based distraction	Eating or drinking
	Outside person, object or event
	Other occupants in vehicle
	Moving object in vehicle
	Smoking related
Miscellaneous	Other distraction
	Unknown distraction

Table 2. Contribution of different distraction sources to vehicle crashes

Distraction type	Stutts et al. study [11]		Glaze and Ellis's study [12]	
	Distraction sources	% of crashes	Distraction source	% of crashes
Technology based	Adjusting radio, cassette, CD*	11.4	Adjusting radio, cassette, CD*	6.5
	Using/dialing mobile phone*	1.5	Using/dialing mobile phone*	3.9
	Adjusting vehicle/climate controls*	2.8	Adjusting vehicle/climate controls*	3.6
	–	–	Technology device*	0.3
	–	–	Pager*	0.1
	Total	15.7		14.4
Non-technology based	Smoking related*	0.9	Smoking related*	2.1
	Other occupant in vehicle*	10.9	Passenger/children distraction*	8.7
	Eating or drinking*	1.7	Eating or drinking*	4.2
	Moving object ahead**	4.3	–	–
	Person, object or event**	29.4	–	–
	–	–	Grooming*	0.4
	–	–	Other personal items*	2.9

(continued)

Table 2. (continued)

Distraction type	Stutts et al. study [11]		Glaze and Ellis’s study [12]	
	Distraction sources	% of crashes	Distraction source	% of crashes
	–	–	Unrestrained pet*	0.6
	–	–	Document*	1.8
	Total	47.2		20.7
Miscellaneous	Other distraction	25.6	Other distraction inside vehicle*	26.3
	Unknown distraction	8.6	–	–
	Object brought in *	2.9	–	–
	Total	37.1		26.3

* Inside Vehicle Distraction Source

** Outside Vehicle Distraction Source

3 Previous Work

This section presents a review of the computer vision based approaches to distraction detection of drivers proposed by researchers in the literature.

Study of driver’s visual behaviour has been widely carried out by researchers since 1960 [13]. Eye glance is considered a valid measure among researchers for the detection of distraction in drivers [14, 15]. In the eye glance approach, the frequency and the duration of a driver’s eye glances for a secondary task are taken to produce a total measure of eyes off the road [13]. Eye glance of the driver can be measured by observing the driver’s eye and head movements using a video sensor. Modern computer vision systems, for example FaceLAB [16], are able to provide real-time measurement of eye glance using head tracking and eye tracking techniques. In a study by Victor et al. [17], the validity of FaceLAB data as the measure for distraction detection has been studied and confirmed. Park and Trivedi [18] also applied SVR for the classification facial features to detect the distracted eye glance in drivers. Relevant facial features were extracted using the global motion approach and colour statistical analysis. Pohl et al. [19] developed a system based on the gaze direction and head position to monitor the distraction in drivers. Instantaneous distraction level was determined and a decision maker was used to classify the distraction level in drivers. Kircher et al. [20] also used the gaze direction as the measure for distraction detection and proposed two different algorithms. Murphy-Chutorian et al. [21] proposed a distraction detection system based on the head position of driver. Localized gradient histogram approach was used to extract the relevant features and were classified using Support Vector Regressor (SVR) to detect the distraction in drivers.

In an effort to provide efficient solution for accident prevention due to distraction, different researchers have proposed distraction warning/alerts systems in the literature. A forward warning system for distraction system was proposed by Hattori et al. [22], which used the idea of checking if the driver is looking at road based on the visual information captured by an in-vehicle camera. PERLOOK is the parameter proposed by

Jo et al. [23] as a measure to detect the distraction level in drivers in a similar way as the PERCLOS for drowsiness detection. PERLOOK is the percentage of time in which a driver’s head is rotated or the driver is not looking at the road ahead. Higher values of PERLOOK means higher duration of distraction in driver. Nabo [24] used the Smart-Eye [25] software tool for the measurement of PERLOOK to detect the distraction in drivers.

Visual occlusion detection is another approach to detecting distracted driving. It assumes that safe driving does not require the driver to look at the road all the time and short intervals are allowed for performing other tasks, such as tuning the radio or adjusting climate controls. With this assumption, secondary tasks that can be performed within 2 s are classified as ‘chunkable’ and considered acceptable during driving [26, 27]. During the occluded time interval, driver can work with different control devices without getting distracted [28]. Validity of visual occlusion technique for the distraction detection is widely measured by researchers and considered promising approach for measurement of visual distraction in drivers [29–31].

4 Our Deep Learning Solution

4.1 Model A: The Baseline Convolutional Neural Network

AlexNet deep network [32], which was the winner of 2012 ImageNet challenge has been used as the baseline model (Model A) in this work. In ImageNet competition, AlexNet was trained on about 1.3 million real life images of 1000 different classes of objects and has achieved the test error rate of 15.3% [32]. Figure 1 shows the architecture of the AlexNet network that we have modified and used for the Kaggle challenge.

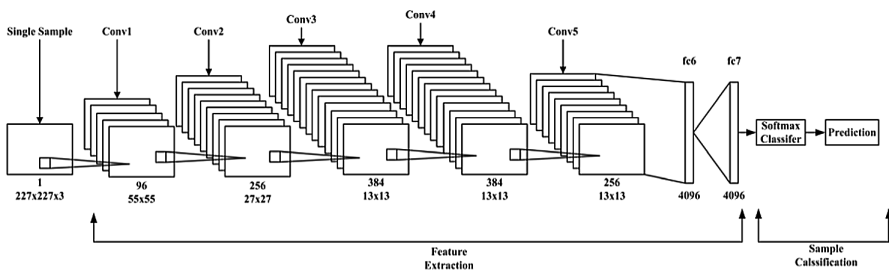


Fig. 1. Modified AlexNet deep learning architecture for Kaggle challenge

The reason behind adopting AlexNet in this work is that AlexNet (or more precisely, the architecture of AlexNet) has demonstrated its ability to learn what to ‘see’ in an image for the purpose of object classification. This ability means, with appropriate training, a CNN with the same architecture as AlexNet will have the ability to recognize objects such as coke cups, phones, pets, driver’s hand etc., all of which are valuable measures in classification of distracted driving.

Each input image to our AlexNet (model A) is $227 \times 227 \times 3$ as defined by the Kaggle challenge. As adopted in the ImageNet competition, the first five layers of network are convolutional layers and provide representation for local features in the images while the last layers are fully connected layers responsible for learning the key features for the given classification task. Our AlexNet extracts 4096 features at fc7 layer and creates a matrix X of the features extracted from all the training images. The dimension of feature matrix X is $m \times 4096$, where m is the number of training images in each batch. In our work, m equals 50. This extracted feature matrix is then fed into the Softmax classifier, which predicts the probabilities of the images in the input batch to the output classes. In Kaggle challenge, there are 10 classes of distracted driving. The output probability values from the Softmax classifier will be compared to the ground truth labels to calculate the following classification loss.

$$\text{logloss} = \frac{1}{N} \sum_i^N \sum_j^M y_{ij} \log(p_{ij}), \quad (1)$$

where N is the total number of images, M is the total number of classes, y_{ij} is the actual class of image and p_{ij} is the predicted class of image.

4.2 Model B: CNN Enhanced with Triplet Loss

In this work, triplet loss has been used to fine tune the model A network pre-trained with classification loss to improve the overall accuracy of the model. There are three main components in each triplet, a positive, an anchor and a negative sample as shown in Fig. 2. The aim of applying triplet loss is to minimize the distance between the anchor and the positive during the learning process and simultaneously increases distance between the anchor and the negative during the learning process to improve the classification accuracy of deep networks. Equation 2 represents the mathematical formulation of triplet loss [33].

$$\sum_i^N \max(0, f(x_i^a, x_i^p) - f(x_i^a, x_i^n) + \alpha), \quad (2)$$

where x_i^a represents the anchor feature vector, x_i^p the positive feature vector and x_i^n the negative feature vector; and α is the forced margin between the anchor-to-positive distance and the anchor-to-negative distance. $f(x_i^a, x_i^p)$ is the function which gives the distance between two feature vector. Triplet loss function from this equation tries to set

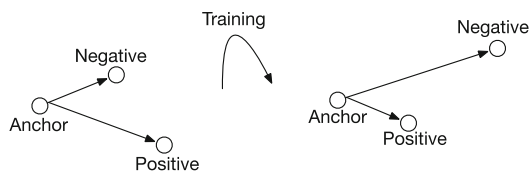


Fig. 2. Working illustration of triplet loss

apart the position samples from the negative samples by a minimum margin of α . The only condition at which the triplet loss will be greater than zero is when $f(x_i^a, x_i^p) + \alpha > f(x_i^a, x_i^n)$.

Random selection of triplets is a slow process and not much efficient for training the network. Triplets that actively contribute to the loss function and hence to improving the accuracy of the network are called hard triplets. Mining hard triplets is an essential step in efficient training of a CNN. Hard triplet selection can be done either offline or online. In offline approach triplets are generated offline for every few steps using the network checkpoint and argmin and argmax of the data are determined. While, in online approach triplets are generated by selecting the positive/negative exemplars from mini-batch [33] during live training. To fasten the convergence of our model B network with triplet loss, offline selection of hard triplets is implemented.

5 Experiments and Results

5.1 Dataset

The Kaggle competition [34] provides a dataset of 80,000 2D images of drivers for data scientists (Kagglers) to classify. Each image in the dataset is captured in vehicle, some with occurrence of distracted activities such as eating, talking on phone, texting, makeup, reaching behind, adjusting radio, or in conversation with other passengers [35]. Table 3 shows the 10 prediction classes defined by the competition.

Overall the dataset has been divided in the ratio of 90%:10% for training and testing the proposed algorithms, respectively. This means from a total of 22424 images in all the Kaggle classes, 20182 are used to train and 2242 to test the two network models.

Table 3. Prediction classes for Kaggle task and number of images in each class [34]

Class0	Class1	Class2	Class3	Class4	Class5	Class6	Class7	Class8	Class9
Safe driving	Texting-right	Talking on phone-right	Texting-left	Talking on phone-left	Operating the radio	Drinking	Reaching behind	Hair and makeup	Talking to passenger
2489	2267	2317	2346	2326	2312	2325	2002	1911	2129

5.2 Experimental Results

This section presents the results of the experiments performed to test the classification accuracy of the two proposed deep learning models as explained in Sect. 4. Overall 5000 maximum iterations were allowed to train the. Figure 3 presents the test accuracy and the test loss of both Models (A: AlexNet+Softmax and B: AlexNet+Triplet Loss) for 5000 iterations with an iteration interval of 500. It has been observed that over the number of iterations classification accuracy improved and both models converged.

Table 4 summarize the results of both algorithms after 5000 iterations. Classification accuracy of 96.8% and 98.7% has been achieved for Model A and Model B, respectively. It is important to mention here that 100% accuracy was achieved for these algorithms when applied to training dataset.

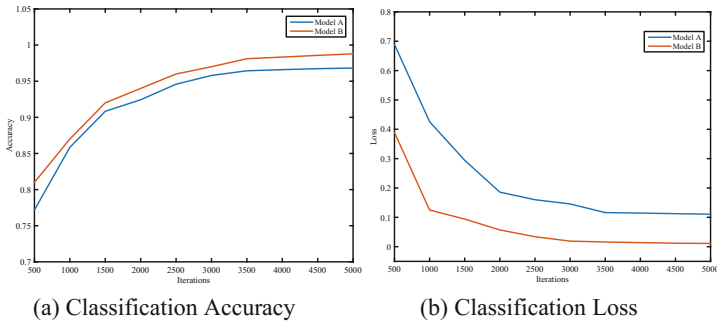


Fig. 3. Classification accuracy and classification loss plots of model A and model B

Table 4. Summary of experimental results for model A and model B

	Test accuracy	Test loss
Model A: AlexNet+Softmax	96.8	0.11
Model B: AlexNet+Triplet Loss	98.7	0.01

5.3 Kaggle Scores

Kaggle provided 22424 images to participants for training their algorithms and asked to submit their classification probabilities for each image in form of excel sheet. Further they tested the submitted algorithms on 79,726 un-labeled images and calculated the loss score for each participant. Kaggle evaluated each submission using a multiclass logloss function as given in Eq. 1.

Classification results from the Model A were submitted to Kaggle and were evaluated for the Kaggle score and rank. Table 5 shows the Kaggle submission results for Model A. The rank was determined at the time of submission out of approximately total 2000 submissions.

Table 5. Kaggle submission results for model A

	Kaggle score	Rank
Model A: AlexNet+Softmax	1.54860	500+

6 Conclusion and Future Works

As discussed in Sects. 2 and 3, majority of the existing approaches to the detection of distracted driving rely on information such as eye glance direction and head movement. To estimate such information, methods have been proposed for the extraction of relevant key features from the face/head region of the driver. However, the image data of the Kaggle challenge are provided for classification of different types of behaviors that involve whole body movements of the driver. To complete the Kaggle challenge, one has to first define the discriminative features from the entire body of the driver that

the subsequent classification process can rely on. This is a challenging task as there is hardly any previous work on what are the discriminative features outside the face region. On the other hand, deep learning networks such as CNNs have provided a brand new approach to data mining and knowledge discovery, which is able to learn the discriminative features for a given classification task. The work presented in this paper confirms the above claim by conducting experiments on the Kaggle challenge using two different CNNs with promising results.

References

1. Beirness, D.J., Simpson, H.M., Pak, A.: The road safety monitor: driver distraction (2002)
2. Peters, G.A., Peters, B.J.: The distracted driver. *J. R. Soc. Promot. Health* **121**, 23–28 (2001)
3. Young, K., Regan, M., Hammer, M.: Driver distraction: a review of the literature. *Distracted Driving*, 379–405 (2007)
4. Stutts, J.C., Reinfurt, D.W., Staplin, L., Rodgman, E.A.: The Role of Driver Distraction in Traffic Crashes. Report prepared for AAA Foundation for Traffic Safety, Washington (2001)
5. Wang, J.-S., Knippling, R.R., Goodman, M.J.: The role of driver inattention in crashes: new statistics from the 1995 crashworthiness data system. In: 40th Annual Proceedings of the Association for the Advancement of Automotive Medicine, p. 392 (1996)
6. Treat, J.R.: A study of precrash factors involved in traffic accidents. *HSRI Research Review* (1980)
7. Ranney, T.A., Mazzae, E., Garrott, R., Goodman, M.J.: NHTSA driver distraction research: past, present, and future. In: *Driver distraction internet forum* (2000)
8. Hajime, I., Atsumi, B., Hiroshi, U., Akamatsu, M.: Visual distraction while driving: trends in research and standardization. *IATSS Res.* **25**, 20–28 (2001)
9. Line, D.: The Mobile Phone Report: A Report on the Effects of Using a Hand-Held and a Hands-Free Mobile Phone on Road Safety. Direct Line Insurance, Croydon (2002)
10. Haigney, D.: Mobile Phones and Driving: A Literature Review. RoSPA, Birmingham (1997)
11. Stutts, J., Feaganes, J., Rodgman, E., Hamlett, C., Meadows, T., Reinfurt, D., Gish, K., Mercadante, M., Staplin, L.: Distractions in everyday driving (2003)
12. Glaze, A.L., Ellis, J.M.: Pilot study of distracted drivers. Transportation Safety Training Center for Public Policy (2003)
13. Farber, E., Foley, J., Scott, S.: Visual attention design limits for ITS in-vehicle systems: the society of automotive engineers standard for limiting visual distraction while driving. In: *Transportation Research Board Annual General Meeting, Washington DC USA*, pp. 2–3 (2000)
14. Haigney, D., Westerman, S.: Mobile (cellular) phone use and driving: a critical review of research methodology. *Ergonomics* **44**, 132–143 (2001)
15. Curry, R., Greenberg, J., Blanco, M.: An alternate method to measure driver distraction. In: *Intelligent Transportation Society of America's Twelfth Annual Meeting and Exposition* (2002)
16. Seeingmachines. <http://www.seeingmachines.com/>. Accessed 11 Mar 2017
17. Victor, T., Blomberg, O., Zelinsky, A.: Automating the measurement of driver visual behaviours using passive stereo vision. In: *Proceedings of International Conference Series Vision Vehicles (VIV9)* (2001)
18. Park, S., Trivedi, M.: Driver activity analysis for intelligent vehicles: issues and development framework. In: *Intelligent Vehicles Symposium Proceedings, IEEE*, pp. 644–649 (2005)

19. Pohl, J., Birk, W., Westervall, L.: A driver-distraction-based lane-keeping assistance system. *Proc. Inst. Mech. Eng. Part I: J. Syst. Control Eng.* **221**, 541–552 (2007)
20. Kircher, K., Ahlstrom, C., Kircher, A.: Comparison of two eye-gaze based real-time driver distraction detection algorithms in a small-scale field operational test. In: *Proceedings of 5th International Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, pp. 16–23 (2009)
21. Murphy-Chutorian, E., Doshi, A., Trivedi, M.M.: Head pose estimation for driver assistance systems: a robust algorithm and experimental evaluation. In: *Intelligent Transportation Systems Conference*, pp. 709–714. IEEE (2007)
22. Hattori, A., Tokoro, S., Miyashita, M., Tanaka, I., Ohue, K., Uozumi, S.: Development of forward collision warning system using the driver behavioral information. *SAE Technical paper* (2006)
23. Jo, J., Lee, S.J., Jung, H.G., Park, K.R., Kim, J.: Vision-based method for detecting driver drowsiness and distraction in driver monitoring system. *Opt. Eng.* **50**, 127202–127224 (2011)
24. Nabo, A.: *Driver Attention—Dealing with Drowsiness and Distraction*. IVSS, Göteborg (2009)
25. Yunqi, L., Meiling, Y., Xiaobing, S., Xiuxia, L., Jiangfan, O.: Recognition of eye states in real time video. In: *International Conference on Computer Engineering and Technology*, pp. 554–559 (2009)
26. Karlsson, R., Fichtenberg, N.: How different occlusion intervals affect total shutter open time. In: *Presentation at the Exploring the Occlusion Technique: Progress in Recent Research and Applications Workshop*, Torino, Italy (2001)
27. Green, P., Tsimhoni, O.: Visual occlusion to assess the demands of driving and tasks: the literature. In: *Exploring the Occlusion Technique: Progress in Recent Research and Applications Workshop*, Torino, Italy, p. 2004 (2001)
28. Jain, J.J., Busso, C.: Assessment of driver’s distraction using perceptual evaluations, self assessments and multimodal feature analysis. In: *5th Biennial Workshop on DSP for In-Vehicle Systems*, Kiel, Germany (2011)
29. Baumann, M., Rösler, D., Jahn, G., Krems, J., Kluth, K., Rausch, H., Bubb, H.: Assessing driver distraction using occlusion method and peripheral detection task (2003)
30. Baumann, M., Keinath, A., Krems, J.F., Bengler, K.: Evaluation of in-vehicle HMI using occlusion techniques: experimental results and practical implications. *Appl. Ergon.* **35**, 197–205 (2004)
31. Wooldridge, M., Bauer, K., Green, P., Fitzpatrick, K.: Comparison of driver visual demand in test track, simulator, and on-road environments. *Ann. Arbor*, 1001, 48109–42150 (1999)
32. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, pp. 1097–1105 (2012)
33. Tutorial: Triplet Loss Layer Design for CNN. <http://www.cnblogs.com/wangxiaocvpr/p/5452367.html>. Accessed 19 Mar 2017
34. Kaggle Competition: State Farm Distracted Driver Detection. <https://www.kaggle.com/c/state-farm-distracted-driver-detection>. Accessed 12 Apr 2017
35. Liu, D., Sun, P., Xiao, Y., Yin, Y.: Drowsiness detection based on eyelid movement. In: *Second International Workshop on Education Technology and Computer Science (ETCS)*, pp. 49–52 (2010)