Using Deep Learning to Detect Driver Distraction in the Australian Naturalistic Driving Study (ANDS) Video Data - Preliminary Results

Mohammed Elhenawy[a], Kristie Young[b], Andry Rakotonirainy[a], Raphael Grzebieta[c,d], Ann Williamson[c]

[a] Centre for Accident Research and Road Safety – Queensland (CARRSQ), Queensland University of Technology, Australia, [b] Monash University Accident Research Centre, Monash University, Australia, [c] Transport and Road Safety (TARS) Research Centre, University of New South Wales (UNSW) Sydney, Australia, [d] Victorian Institute of Forensic Medicine, Monash University, Australia

Abstract

This paper reports preliminary results of investigating the use of machine learning techniques to label distraction related events from video data collected from the Australian Naturalistic Driving Study (ANDS). This offline automatic labeling is designed to replace manual coding and accelerate the data reduction process with the view to save effort and money. We adopted the well-known pre-trained deep learning network Alex to label ANDS video data. The pre-trained network was used as a starting point after modifying the fully connected and classification layers. Then the modified model was retrained using ANDS data. The re-trained network achieved promising results despite low video quality.

Method

Convolutional Neural Networks (CNNs) are deep learning algorithms that are powerful for image classification. During the training, CNN learns the weights of the different layers to yield the best features and the least classification error. Training a CNN from scratch needs a large number of high quality images and long training time to facilitate transfer learning. Transfer learning exploits knowledge gained from solving one problem and applying it to a different but related problemas a shortcut to save time. We used the pre-trained model approach for transfer learning where we choose a pre-trained model (Alex network[1]). Then we modified the fully connected and classification layer to suit the driver distraction task. Finally, we tuned/re-trained the model using ANDS video data. We have chosen transfer learning because it offers the following benefits[2]:

1- Higher classification accuracy before tuning the pre-trained model
2- Higher rate of classification improvement during the training process
3- Higher final classification accuracy at the end of the training process

Results

This paper used six trips of the data collected as part of the ANDS. The videos of six trips were captured from different drivers. The video data were collected using a continuous multi–camera video recording system that captures the driver's face, forward and rear views, and a view of driver interaction with the dashboard and other systems at a rate of 15 Hz [3]. In this early stage of developing the classification model, we focus only on the videos which capture the interaction of the driver with the dashboard and other systems. The videos of five trips were converted into images/frames and were visually inspected to label each image as distracted or non-distracted [4]. The low quality images were removed from the training dataset. We re-trained Alex network using 7449 images extracted from five trips. The video of the sixth trip was processed in the same way as the training data and the resulting 2179 images were used to test the re-trained model. The test result is an indicator of how close the CNN model is to the manual coding.

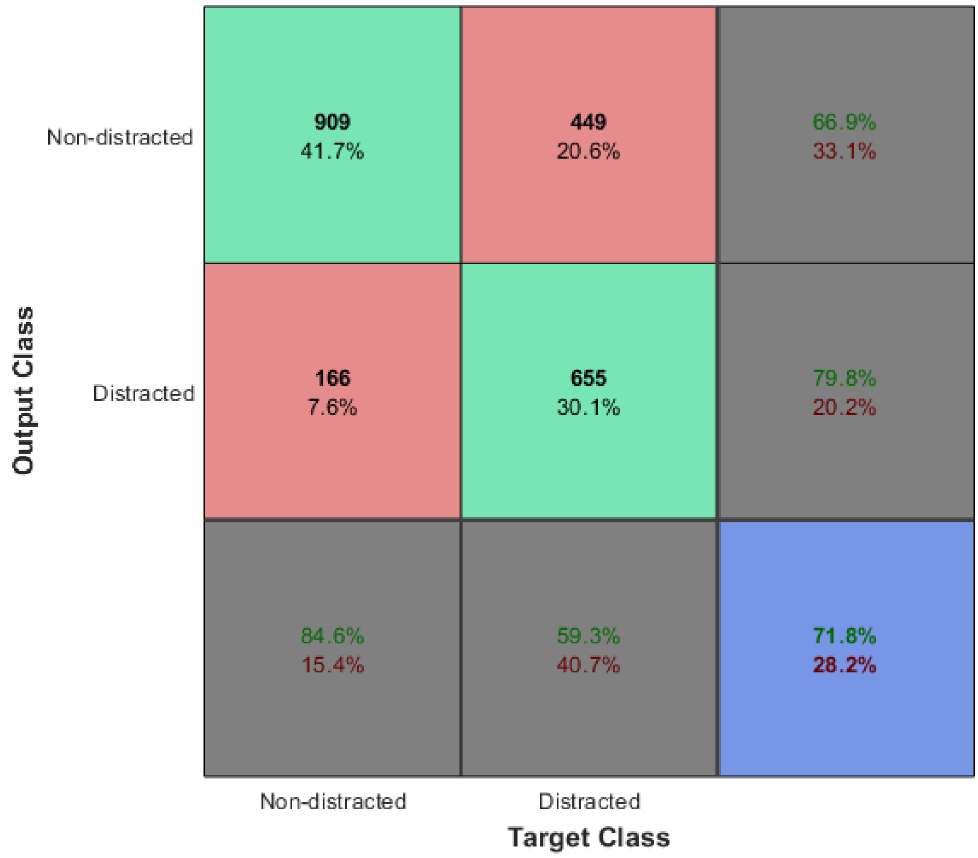

Figure 1. The confusion matrix plot of the re-trained Alex network where target class is the ground truth and output class is the predicted class

The results of testing the re-trained Alex network is shown in Figure 1. As shown in the figure, the accuracy, sensitivity, and specificity are 71.8%, 59.3% and 84.6%, respectively. These measures are promising and suggest that the approach is worth pursuing. The next steps to improve the automatic labeling will include adding much more data for training. Moreover, other models that incorporate the temporal correlation between frames will be used.

Conclusion

Automatic data reduction of video collected from ANDS is very important for downstream analysis due to the sheer size of collected data. Automatic reduction of video data will save effort, money/time and discover new knowledge. However, automatic analysis is challenging because of the data quality and quantity needed for building a good model. The preliminary results using CNN and transfer learning are very promising and our ongoing work to improve the classification model includes; adding more trips to the training dataset and test dataset and adapting recurrent neural network models which take advantage of the temporal correlation between frames to improve classification accuracy.

Acknowledgements

The ANDS study was funded by the Australian Federal Government's Australian Research Council (ARC) under the infrastructure LIEF scheme (grant No. LE130100050) and research Linkage scheme (grant No. LP130100270) that were awarded to the Transport and Road Safety (TARS) Research Centre at the University of New South Wales (UNSW) in partnership with Monash University Accident Research Centre (MUARC), the Centre for Accident Research and Road Safety – Queensland (CARRSQ) at Queensland University of Technology (QUT), the Centre for Automotive Safety Research (CASR) at the University of Adelaide, the Centre for Road Safety, Transport for NSW (TfNSW), NSW's State Insurance Regulatory Authority, Victoria's Transport Accident

Commission, the Victorian State regulator VicRoads, South Australia's Motor Accidents Commission, the National Roads and Motorists Association (NRMA) in NSW, Seeing Machines, the Office of Road Safety, Government of Western Australia, and the Hyundai Motor Company. Supply of equipment and assistance with ANDS data was provided by The Virginia Tech Transportation Institute (VTTI) under the leadership of Prof. Jon Antin.

The Authors acknowledge the contribution of Prof Judith Charlton from MUARC, Prof. Narelle Haworth and Prof Teresa Senserrick from CARRSQ, Dr Jeremy Woolley from CASR to the project. The Authors would like to thank Jan Eusibo and Rainer Zeller from Transport and Road Safety (TARS) Research Centre at the University of New South Wales (UNSW) for their role in managing and operationalising the Australian study and the Sydney site. The contribution of the support staff Jess Hancock from TARS UNSW at the Sydney site and Yik-Xiang Hue, Andrew Lyberopolous and Samantha Bailey from MUARC at the Melbourne site is also gratefully acknowledged. The valuable contribution from John Wall and Ben Barnes from the Centre for Road Safety TfNSW in operationalising the Sydney installation site and the assistance from Hee Loong Wong from Hyundai Motor Company in providing the ANDS service vehicles for both the Sydney and Melbourne sites is also gratefully acknowledged. We thank Dr Karen Stephan for her statistical advice.

The Authors would also like to thank the ANDS participants. Without their contribution to the study, this publication would not have been possible.

Dr Kristie Young's contribution to this paper was funded by an ARC Discovery Early Career Researcher Award (DE160100372).

References

1. Krizhevsky, A., I. Sutskever, and G.E. Hinton. *Imagenet classification with deep convolutional neural networks*. in *Advances in neural information processing systems*. 2012.
2. Emilio Soria, O., et al., eds. *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*. 2010, IGI Global: Hershey, PA, USA. 1-852.
3. Williamson, A., et al. *The australian naturalistic driving study: From beginnings to launch*. in *Proceedings of the 2015 Australasian Road Safety Conference*. 2015.
4. Young, K.L.O., R., Koppel, S., Charlton, J., Grzebieta, R., Williamson, A., Haworth, N., Woolley, J., & Senserrick, T., *What are Australian drivers doing behind the wheel? An overview of secondary task data from the ANDS.* Journal of the Australasian College of Road Safety, 2019.