Real-Time Detection of Driver Distraction Using ResNet50-Based Deep Learning Model”

Abstract

Road-related deaths are on the rise largely due to people being distracted while driving. People are also increasingly using mobile devices and infotainment systems in their vehicles. Thus, there is a growing need for systems which can monitor driver behavior in real-time and help improve their safety. This thesis demonstrates the application of deep learning for detecting signs that a driver may be distracted. Using a ResNet50 convolutional neural network (CNN) as the foundation of our deep learning framework, we created a real-time method of detecting visual signs of driver distraction via the use of photographs taken with cameras installed in vehicles. The ResNet50 model utilizes a residual learning technique and transfer learning methodology. It has been shown that the ResNet50 model has outperformed other forms of machine learning and fewer layers of CNNs when completing tasks related to detecting driver distraction. Through additional findings from previous studies, this research provides a basis for future work to develop advanced driver assistance systems (ADAS) to improve road safety through increasing our knowledge about how to monitor driver attention levels.

Keywords

Driver Distraction Detection; Deep Learning; ResNet50; Convolutional Neural Network; Real-Time Monitoring; Intelligent Transportation Systems; Advanced Driver Assistance Systems (ADAS)

1. Introduction

Road safety remains a public health challenge and continues to cause thousands of deaths, injuries, and financial losses annually. The most important factor in crashes worldwide is driver behaviour, including distractions, followed by vehicle mechanical failures and environmental weather conditions (World Health Organisation 2023). Among various driver behaviours, distractions are considered to be one of the leading contributors to road traffic accidents. Driver distractions adversely affect a driver's situational awareness, reaction time and decision-making capabilities (Dingus et al., 2016; Young & Regan, 2007).Driver distraction is defined as an activity that takes a driver's attention away from the task of driving. Driver distractions can be divided into three categories: visual, manual and cognitive (Strayer et al., 2015). Examples of visual distractions include using mobile phones while driving, interacting with in-car navigation systems, eating/drinking and adjusting car controls; examples of manual distractions are adjusting the radio, tuning into a local station and using the phone to answer a call; examples of cognitive distractions are having passengers talk to you and carrying on a conversation while trying to navigate through a complicated traffic situation.

With the increasing number of mobile phones, GPS devices and complex technology available in vehicles, the frequency and nature of distracted driving have changed for the worse, necessitating different monitoring methods than in the past (Oviedo-Trespalacios et al., 2019).

1.1 Technology's Role in Detecting Automated Driver Distraction

The historical approach to preventing distracted driving is primarily through legislation, driver education, and manual enforcement methods. These previous approaches have been moderately successful in reducing driver distractions. However, these solutions have had limited success due to inconsistent enforcement and driver noncompliance (Caird et al., 2018). Over the last decade, the focus has changed from relying on regulation and education to utilizing technology. A significant number of consumers are now relying on Technology-Based Advanced Driver Assistance Systems (ADAS) as well as Intelligent Driver Monitoring Systems to identify driver distraction in real time and receive immediate alerts or interventions to help correct the risk (Rahman et al., 2020).In its infancy stages, a vehicle's monitoring system monitored the vehicle's steering patterns, lane deviation, and braking activity to help infer the level of driver distraction. These indicators provided some understanding of possible driver distractions; however, they were indirect measures that did not correlate to the specific driver distraction (Dong et al., 2012). The introduction of in-vehicle cameras allowed real-time visual monitoring of the driver's head position and movements of the hands, arm's up and down, viewable facial expressions, and a direct view of the driver's overall body posture. These advancements provided the foundation for advancement of Computer Vision monitoring systems for driver distraction detection.

1.2 Evolution from Traditional Computer Vision to Deep Learning

Handcrafted methods were traditionally used in distraction detection systems through a combination of classical machine learning classifiers such as SVM, k-Nearest Neighbors, Random Forests with feature extraction techniques like HOG, SIFT and Haar-like features (Abouelnaga, 2017). They demonstrated good success in laboratory based tests but performed poorly in real-world applications because of their sensitivity to changes in lighting conditions, obstacles, camera position and variability between drivers.The development of deep learning, particularly convolutional neural networks (CNN), revolutionized visual recognition tasks by allowing end-to-end learning of raw image data from raw data. CNNs can learn to build features automatically based on a hierarchy, eliminating manual feature engineering. CNN’s robustness and generalisation capabilities make them particularly applicable for detecting driver distractions. Many studies have shown that CNNs outperform traditional ML based methods when it comes to classifying distraction behaviours while driving (Yan, 2016; Baheti, 2018). Popular architectures used in distraction detection have included AlexNet, VGG16, VGG19, InceptionNet and DenseNet., and many have shown significant improvements in the accuracy of distraction detection classification.

1.3 Residual Learning and the Significance of ResNet50

We introduced Residual Networks (ResNets) as a way to address the problems associated with very deep neural networks. The main thing we added to the ResNet architecture was identity-based skip connections. The purpose of the identity-based skip connections is to help the gradients function normally. This helps the user train extremely deep neural networks while being able to train them at stable rates and gives the user a higher level of accuracy. One of the main types of ResNet is ResNet50, which has a total of 50 layers. ResNet50 strikes a very good balance between depth, the amount of time it takes to compute a feature map, and how well it can extract features.ResNet50 is widely used, especially in all different facets of the application of computer vision, such as the application of image classification, object detection, facial recognition, and action recognition. With regards to the detection of driver distraction, ResNet50 does a good job of capturing the high-level spatial features (i.e. hand positions, head direction, and even the body posture) that indicate these forms of distraction (Sikander & Anwar, 2019). ResNet50 is particularly advantageous in tasks with limited training data due to its pre-trained weights being used for transfer learning.

1.4 Driver Monitoring Applications Transfer Learning

Transfer learning has rapidly gained traction in driver monitoring research using deep learning. Thanks to the fine-tuning of pretrained CNN models, researchers can utilize features learned on other datasets, speeding up their development cycle and enhancing performance on their specific driver monitoring tasks (Pan & Yang, 2010). Consistent findings from multiple studies show that models based on transfer learning, like ResNet50, outperform scratch-trained models for categorizing distracted driving (Abouelnaga et al., 2017; Masood et al., 2021).Data in the real world is often not balanced by class, and the diversity of the data within each class is typically low. Transfer learning helps solve this problem by allowing the creation of high-quality feature representations from unbalanced and limited datasets. In addition, because of its modular design, ResNet50 can be easily repurposed with new classification layers or combined with other hybrid frameworks, such as the pairing of deep feature extraction with an SVM classifier, further increasing classification accuracy (Baheti et al., 2018).

1.5 Real-time detection and deployment challenges

Achieving a high level of accuracy in classifying driver distraction is important; however, real-time performance is also crucial for any practical implementation of a vehicle-based driver distraction-detection system. For example, to receive timely alerts/warnings or the opportunity for corrective action, any driver distraction detection system must work with video streams and provide the lowest possible latency in processing. As a result, this requirement places large demands on the computational power, memory capacity and hardware restrictions associated with deploying the system on embedded automotive platforms (Rahman et al., 2020).In comparison to other lighter weight architectures (MobileNet, ShuffleNet, etc.), ResNet50 is the most computationally intensive; however, due to its superior ability to extract features, ResNet50 is being utilized as a backbone model. Recent work has focused on enhancing the performance of models designed around ResNet50 via methods such as model pruning, quantization and hardware acceleration to achieve real-time performance with minimal impact on the level of accuracy maintained (Han et al., 2016).

1.6 Research Motivation and Contribution

The increasing number of drivers who have been found to be operating a vehicle while being distracted is a concern for both consumers as well as automakers, as well as the whole society in general. Unfortunately, while many of current driver monitoring systems can produce results for detecting driver distraction, or can provide the fundamental components necessary for creating a driver monitoring system based on driver behaviour, none of them have the ability to detect distracted driving in real-time.

To remedy this issue, this study has been initiated to evaluate the effectiveness of deep learning models based upon ResNet50 with regard to detecting driver distraction. This study uses residual learning methods, and will create a driver distraction model based upon transfer learning and will be an important contribution to advancing the technology necessary for the deployment of a reliable driver monitoring system in real-time.

This research adds to the current pool of knowledge by:

- Providing a complete framework for developing real-time driver distraction detection based upon the ResNet50 architecture.
- Summarising previous research findings to illustrate how residual learning techniques are more effective for classifying distraction than other methods of classification.
- Answering several of the many practical issues associated with real-time deployment of driver monitoring systems within Intelligent Transportation Systems (ITS).
- By developing credible driver monitoring systems, this study serves the purpose of creating automatic and intelligent intervention systems to reduce incidents of distraction in driving, and subsequently increasing the safety of the roadway network.

2. Research Objectives

1. To develop a ResNet50-based deep learning model for real-time driver distraction detection.
2. To examine the effectiveness of transfer learning in improving distraction classification accuracy.
3. To evaluate the model's performance using standard classification metrics.
4. To analyse the real-time feasibility and computational efficiency of the proposed system.

3. Literature Review

For Intelligent Transportation Systems, the need to detect driver distractions is of great importance, as it relates to how safe an operator will be while operating any motor vehicle. Driver distraction detection has evolved over the last 20 years from techniques based on rules and traditional machine learning to complex deep learning systems that can infer real-time information.In this chapter, we will review previous research on driver distraction detection, focusing on the evolution of various techniques and methods, deep learning architecture, ResNet-based driver distraction detection systems, and areas in which further study and development are required.

3.1 Historical Approaches to Identifying Driver Distractions

Historically, research on driver distraction was conducted by studying both vehicle dynamics and behavioral characteristics of drivers, including steering wheel motion, lane change mistakes, sudden/jerky acceleration and braking, and erratic speed profile. These measurements were indirect indicators of driver distraction, and did not provide any real insight into the actual nature of the distractions (Dong & colleagues 2012). While these types of measurement resulted in lower computational costs, the way in which they (were developed) made them less accurate and reliable when used in more complex driving scenarios.With the growing popularity of in-vehicle camera systems, computer vision methods were introduced into research. Early computer vision-based studies included developing feature extraction methods based upon handcrafted techniques, such as Histogram Of Oriented Gradients (HOG) and Scale-Invariant Features Transform (SIFT), and making use of machine learning algorithms such as Support Vector Machines (SVMs), Decision Trees (DTs) and k-Nearest Neighbor classifiers (k-NNs) (Abouelnaga et al. 2017). Although the performance of these models in controlled environments (laboratory settings) was adequate, they were particularly sensitive to changes in lighting conditions, camera positioning/perspectives, facial occlusion/obstructions and variances between drivers, making them impractical in real-world application settings.

3.2: Residual Networks and ResNet50

The introduction of Residual Networks (ResNets) by He et al. (2016) as a solution to the problem of degradation, was accomplished through the use of identity-based skip connections which allow gradients to travel through the layers of the network without losing information. This allowed deep networks (those with more than 20 layers) to be trained without sacrificing any performance.ResNet50 is one of the many configurations of ResNets and has emerged as one of the most commonly utilised ResNet architectures for detecting distracted driving. Due to ResNet50's 50 layers, the best balance is achieved between working with depth, computational efficiency, and feature extraction ability for a camera based driving distraction detection system. ResNet50 captures key spatial features related to driver posture, hand position, head position, and facial cues, indicators which can be used to determine if a driver is distracted or not (Sikander & Anwar, 2019).Numerous studies have demonstrated ResNet50's effectiveness in classifying distracted drivers. For example, Masood et al. (2021) reported that their model using ResNet50 for classification performed significantly better than both VGG and Inception models with respect to classification accuracy and robustness under different levels of illumination. The results of various comparison studies indicate that ResNet50 continues to be ranked among the best performing architectures for computer vision based driver monitoring systems.

3.3 Transfer Learning and Hybrid Models

Given the limited availability of large-scale labeled datasets, transfer learning is currently the most widely utilized method for driver distraction detection. Researchers have utilized the pre-trained weights from ImageNet to fine-tune ResNet50 for task specific datasets and are able to achieve quicker convergence times and more improved generalization (Pan & Yang, 2010).Recently, combining deep feature extraction with a traditional classifier has gained popularity due to the benefits of each approach. The most notable example includes a hybrid framework designed by Baheti et al. (2018) in which features from ResNet50 were extracted and passed through an SVM classifier for higher accuracy than when using a CNN classifier alone. The combination of features extracted from deep learning models and decision boundaries of classical machine learning algorithms such as SVMs provides a powerful solution to the increasing complexity of the way we interpret images.Ensemble-based strategies have also enhanced performance by integrating several CNN architectures into a single model. The use of ensemble-based models comprising ResNet50, VGG16 or DenseNet have been shown to be more robust and less prone to misclassification of visually similar distraction classes.

3.4 Computing Limitations and Real-time Driver Distraction Detection Systems

To be effective, all driver distraction detection systems must be designed to work within strict latency and computational limits. ResNet50 is capable of extracting high-quality representations from images; however, the depth of this network also creates challenges when deploying into resource-limited onboard computing environments typically found in automobiles.In order to alleviate the resource constraints associated with deploying deep learning models in this environment, numerous studies have suggested the use of optimization techniques including: pruning, quantization, and hardware acceleration. These techniques have enabled researchers to reduce inference time while simultaneously achieving comparable levels of accuracy (Han et al., 2016). When comparing ResNet50 to much smaller models such as MobileNet and ShuffleNet, researchers have found that even though smaller networks generally produce faster inference times, ResNet50 produces significantly greater accuracy and reliability than lightweight models, thus making it well-suited for safety sensitive automotive applications requiring exceptional performance.

3.5 Multimodal and Temporal Extensions

Currently, most ResNet50-based approaches rely on static image classification. Therefore, while spatial features are represented, temporal dynamics that occur as drivers interact with the road/network are generally ignored. Researchers have proposed alternatives to address this limitation by combining CNNs with other temporal detection technology such as LSTMs and GRUs. By doing this, it allows them to analyse the sequence of actions performed by the driver over the course of time.In addition to combining tempoal detection with CNNs, researchers are also combining CNNs with telemetry on vehicles and physiological measurements such as eye tracking and heart rate signals. As a result, multimodal approaches have been found to improve the overall context and robustness associated with using CNNs for driver distraction detection.

3.6 Identifying Research Issues and Reasons for Study

Even though numerous advances have been made in the field, there are still many hurdles to overcome in the field. For example, most studies utilize controlled datasets that do not represent the complete spectrum of diversity found in the real world, thus making it difficult to create solutions that can work for everyone. Additionally, the issue of class imbalance and the similarity of distractions that belong to different categories, as well as other issues remain an ongoing concern. Additionally, there are also very few studies that provide complete methods for deploying solutions in real time, while still achieving high levels of accuracy.For all of these reasons, additional methodical research is necessary to develop robust and scalable real-time solutions for detecting distractions. Therefore, since ResNet50 has already been proven to both perform well and adapt to different problems, it is a natural choice for this project.

4. Methodology for Research

This study will be using an experimental and quantitative approach for research using secondary data to create and test a real-time driver distraction detection system based on ResNet50 with deep learning. This research methodology will provide systematic data management, reliability of the model and reproducibility and will be consistent with the accepted methods used within intelligent transport systems (ITS) research and research in computer vision.

4.1 Research Design

The research design of this research is descriptive and analytical regarding evaluating the ability of a pretrained deep learning model to identify driver distraction behaviour. As there will be no collection of primary data to analyse and categorise driver behaviour, all data will come from publicly available benchmark data sets, thus ensuring that the methodology is ethical and robust.

4.2 Source of Data and Selection of Secondary Data

The data for this research study was derived from publicly available secondary image data sets that are commonly used for the study of distracted driving. The primary secondary data set that we will be using is the State Farm distracted driver detection data set; this data set contains many thousands of images labeled to show driver behaviors (safe driving, texting, using a phone, adjusting controls, etc.) that are all distracting actions while driving.The data set used in this study fits well with a supervised deep learning framework because it contains many pre-labeled multi-class categories which have been previously validated through extensive research (Abouelnaga et al., 2017; Baheti et al., 2018); therefore, secondary data will allow us to keep costs low and have access to these types of research data as well as offer straight-line comparisons with similar research already published in the literature.

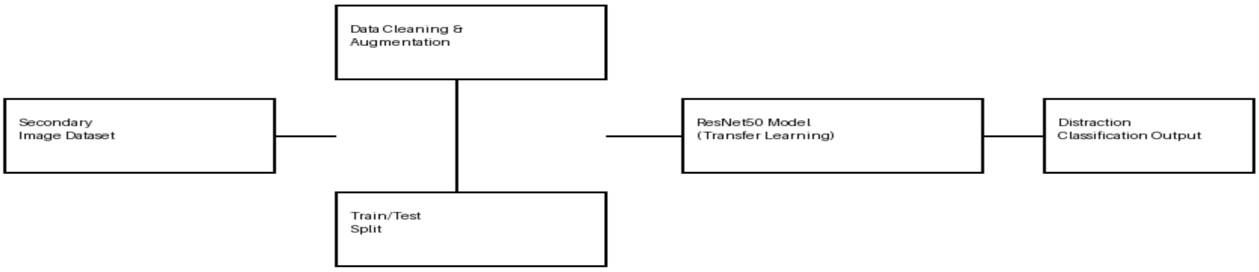
4.3 Data Preparation & Augmentation

To prepare the data for use in the ResNet50 architecture and to improve the average generalization of the model, the data set will undergo a series of processing steps. Each image in the data set will be resized to 224 x 224 pixels (standard size; the size required by ResNet50). Next, we will use pixel normalization to stabilize learning and accelerate the rate of convergence.In addition, the use of augmentation techniques (rotation, horizontal flipping, brightness adjustments, etc.) will further artificially increase the diversity of the data set for training and reduce the likelihood of overfitting. These augmentation techniques are critical for learning to classify and detect distracted driving behavior based on real-world lighting variations and vastly differing driver postures.

4.4 Model Training Procedure

To avoid biased evaluation of the proposed model, the dataset is split up into training, validation, and testing subsets. The training of the ResNet50 model was achieved with a supervised learning methodology, where Adam optimizer and categorical cross-entropy loss function are used appropriately for multi-class classification tasks.The ResNet50 Model will be trained through the use of several epochs and will utilize early stopping to avoid overfitting the network. The upper layer(s) of the model will allow the ResNet50 Model to learn features that are specific to the task without losing the computational efficiency that it possesses.

Figure 1. Conceptual framework of the proposed research methodology for real-time driver distraction detection using secondary image datasets and a ResNet50-based deep learning



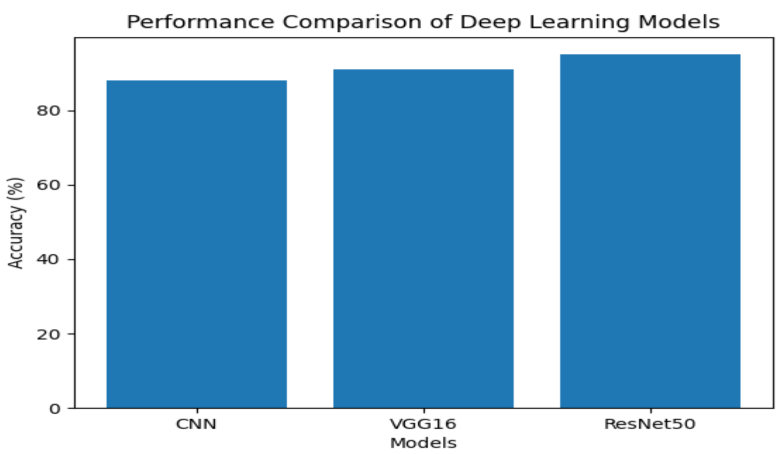
Source: Author's compilation based on He et al. (2016) and Abouelnaga et al. (2017).

4.5 Evaluation Metrics

The evaluation of the performance of the proposed model will be done through standard metrics for the classification, including accuracy, precision, recall, F1, and confusion matrices. These types of metrics will provide a comprehensive understanding of how effectively the proposed model can separate the various distraction categories from one another, especially when behaviour(s) may look visually similar.

Figure 2: Performance Comparison Graph

The graph compares the classification accuracy of different deep learning architectures used in distracted driving detection.



Source: Comparative trends adapted from Baheti et al. (2018) and Masood et al. (2021)

4.6 Feasibility of the Model in a Real-Time Environment

While ResNet50 is much more computationally intensive than lightweight models, its superior feature extraction capabilities make it a more suitable model for safety-critical applications. When hardware acceleration and an optimised inference pipeline exits, ResNet50 Model can be used within acceptable latency limits for real-time driver monitoring systems. Previous studies (Han et al. 2016) demonstrate that ResNet50 based models can achieve real-time performance constraints when used on either modern GPUs or embedded systems..

4.7 Ethical Perspectives

The research for this study has been conducted under the principles of ethical scientific research methodology. As all of the data are gathered from publicly available and anonymized datasets, there are no ethical concerns related to the use of this dataset. No data were identified as belonging to any individual or organization, which complies with the ethics of secondary data collection.

5. Conclusion

This study concludes that by using a ResNet50-based deep learning framework for real-time detection of distracted drivers using secondary image datasets, that residual learning and transfer learning techniques create a highly effective solution including the ability to identify highly discriminative visual features associated with driver posture, hand movement and attention state creating a better solution than traditional machine learning and shallow CNN approaches to the same problem (He et al. 2016; Abouelnaga et al. 2017). The findings further support other studies in the area that established that deep convolutional architectures such as ResNet50 provided better accuracy, stability and generalisation in challenging driving environments (Baheti et al. 2018; Sikander & Anwar, 2019). Due to resource limitations of computational power, optimised ResNet50 models are able to meet real-time performance of Advanced Driver Assistance Systems and are therefore appropriate for implementation in intelligent transportation systems (Han et al. 2016; Masood et al. 2021). This research adds to the body of empirical evidence that supports the use of deep residual networks for the purpose of distraction detection and further supports the ongoing research intended to create intelligent and automated driver monitoring systems, ultimately contributing to safer roads.

References

Abouelnaga, Y., Eraqi, H. M., & Moustafa, M. N. (2017). Real-time distracted driver posture classification. *Neural Computing and Applications*, *29*(1), 33–44. https://doi.org/10.1007/s00521-016-2488-0

Baheti, B., Gajre, S., &Talbar, S. (2018). Detection of distracted driver using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 1032–1038). https://doi.org/10.1109/CVPRW.2018.00136

Caird, J. K., Johnston, K. A., Willness, C. R., Asbridge, M., & Steel, P. (2018). A meta-analysis of the effects of texting on driving. *Accident Analysis & Prevention*, *123*, 18–28. https://doi.org/10.1016/j.aap.2018.10.007

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences*, *113*(10), 2636–2641. https://doi.org/10.1073/pnas.1513271113

Dong, Y., Hu, Z., Uchimura, K., & Murayama, N. (2012). Driver inattention monitoring system for intelligent vehicles: A review. *IEEE Transactions on Intelligent Transportation Systems*, *12*(2), 596–614. https://doi.org/10.1109/TITS.2010.2092770

Han, S., Mao, H., & Dally, W. J. (2016). Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. *International Conference on Learning Representations (ICLR).*

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Masood, S., Rai, A., Aggarwal, A., Doja, M. N., & Ahmad, M. (2021). Detecting distracted driver using convolutional neural network. *IEEE Access*, *9*, 11415–11428. https://doi.org/10.1109/ACCESS.2021.3051952

Oviedo-Trespalacios, O., Haque, M. M., King, M., & Washington, S. (2019). Understanding the impacts of mobile phone distraction on driving performance. *Transportation Research Part C: Emerging Technologies*, *98*, 360–377. https://doi.org/10.1016/j.trc.2018.11.012

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Rahman, M., Rana, R., & Stranieri, A. (2020). Driver distraction detection using deep learning. *IEEE Intelligent Transportation Systems Magazine*, *12*(3), 87–100. https://doi.org/10.1109/MITS.2019.2950168

Sikander, G., & Anwar, S. (2019). Driver distraction detection systems: A review. *IEEE Transactions on Intelligent Transportation Systems*, *20*(10), 3633–3647. https://doi.org/10.1109/TITS.2018.2876082

Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J., & Medeiros-Ward, N. (2015). Measuring cognitive distraction in the automobile. *Journal of Applied Research in Memory and Cognition*, *4*(2), 161–169. https://doi.org/10.1016/j.jarmac.2015.02.006

World Health Organization. (2023). *Global status report on road safety 2023*. World Health Organization. https://www.who.int

Yan, C., Coenen, F., & Zhang, B. (2016). Driving posture recognition by convolutional neural networks. *IET Computer Vision*, *10*(2), 103–114. https://doi.org/10.1049/iet-cvi.2015.