Detection of Distracted Driver using Convolutional Neural Network

Amirhossein Sorour (40270432)
Anahita Jabbari (40204349)
Ghazal Danaee (40295587)
Shima Sargordi (40296231)

Group 4

Abstract. The increasing incidence of road traffic accidents attributable to distracted driving has prompted significant research into automated detection systems. This project explores the use of Convolutional Neural Networks (CNNs), specifically employing the VGG-16 architecture, to detect distracted driving behaviors effectively. We aimed to replicate and expand upon previous studies by utilizing the Distracted Driver Dataset, which captures a variety of driving postures through in-vehicle camera systems. The study engaged in a comprehensive comparative analysis between the original dataset (V1) and an enhanced version (V2), which introduced greater diversity in driver behaviors and environmental conditions. Modifications to the VGG-16 architecture were made to improve computational efficiency by replacing fully connected layers with convolutional layers, allowing for varied input sizes and reducing model complexity. Our experimental results indicate improved accuracy and robustness in distracted driver detection with the modified CNN, achieving up to 94.73% accuracy on the more challenging V2 dataset. The findings underscore the potential of advanced image processing techniques in enhancing road safety by accurately identifying and classifying distracted driving behaviors.

Keywords: Distracted Driving, Convolutional Neural Networks, VGG-16 Architecture, Image Processing, Data Augmentation, Machine Learning, Driver Behavior Analysis, Road Safety

1 Introduction

Based on the global status report on road safety by the World Health Organization (WHO) in 2018, road traffic accidents are the 8th leading cause of death globally[1]. According to their survey, 1.35 million people die in traffic accidents each year and up to 50 million are injured. When a driver's focus and attention stray from the road, the risk of a collision escalates. Distractions impair their performance and reduce awareness, resulting in slower recognition of critical events. Consequently, their ability to respond safely may be compromised, and they might even overlook such events altogether. The National Highway Traffic Safety Administration (NHTSA) in the USA reports that the number of road accidents caused by driver distraction has been increasing in recent years [2]. It mentions that approximately 481,000 drivers use mobile phones while driving during daylight hours. Drivers using mobile phones are approximately four times more likely to be involved in a crash than drivers not using mobile phones. [1] The NHTSA report mentions the death of 3,166 people in 2017 caused by distracted drivers in the USA [3].

According to data from Transport Canada's National Collision Database, distracted driving contributed to an estimated 21% of fatal collisions and 27% of serious injury collisions in 2016. These statistics reflect an upward trend in distracted driving related collisions, increasing from 16% of fatal collisions and 22% of serious injury collisions a decade earlier [4]. NTHSA (National Highway Traffic Safety Administration) describes distracted driving as "any activity that diverts the attention of the driver from the task of driving" which can be classified into Manual, Visual, or Cognitive distraction [3].

In this study, our objective is to replicate the findings outlined in the paper titled "Detection of Distracted Driver using Convolutional Neural Network," authored by Baheti et al. [5]. We aim to conduct a thorough analysis of our work in comparison to the original paper. To accomplish this, we employ VGG-16 architecture proposed by Simonyan et all. [6]. Similar to the original study, we opt for convolutional layers instead of fully connected layers to minimize the number of parameters. Finally, we evaluate our results to confirm their consistency with the findings reported in the paper.

2 Previous Works

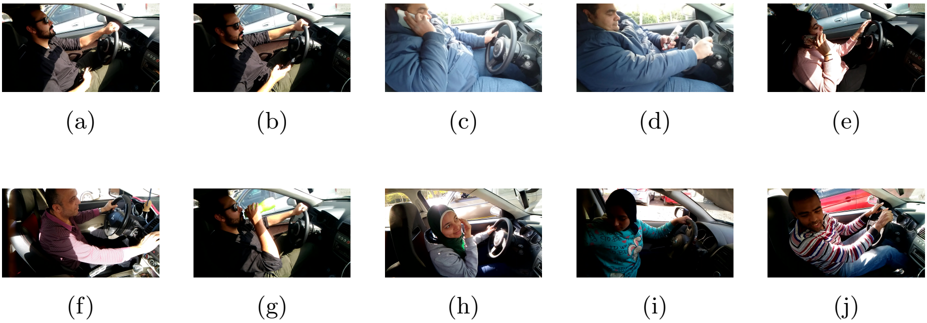
In this section, we will review relevant and significant work from literature about the detection of distracted drivers. Zhang et al. [7] made a database with a camera mounted above the dashboard and utilized the Hidden Conditional Random Fields model to detect cell phone usage. In 2015, Nikhil et al.[8] created a dataset for hand detection in the automotive environment and achieved an average precision of 70.09% using Aggregate Channel Features (ACF) object detector. Seshadri et al.[9] also created their own dataset for cell phone usage detection. The authors used the Supervised Descent Method, Histogram of Gradients (HoG) and an AdaBoost classifier and achieved 93.9% classification accuracy and could process 7.5 Frames Per Second (FPS). Le et al. proposed a faster RCNN based hand and face detector on the above dataset and outperformed earlier methods by achieving an accuracy of 94.2% [10]. The system processed 0.09 FPS for hands on the wheel detection and 0.06 FPS for cell phone usage detection. Automatic Identification of Driver's Smartphone (AIDS) was developed to find the position of the cellphone through analyzing information from sensors available on commodity smartphones [11].

Zhao et al.[12] designed a more inclusive dataset, including side views of the driver engaged in four activities: safe driving, operating the shift lever, eating, and talking on a cellphone. The authors achieved 90.5% accuracy using contourlet transform and random forest. The authors also proposed another approach with PHOG and multilayer perceptron, achieving 94.75% accuracy.[13] In November 2023, Abbass et al.[14] proposed an architecture based on the MobileNet transfer learning model for fast and accurate distracted driver detection, achieving a validation accuracy of 89.63% on the State-Farm dataset. Moreover, in January 2019, Eraqi et al.[15] proposed an Ensemble of Convolutional Neural Networks for distracted driver identification. A reliable deep learning-based solution that achieves an impressive 90% accuracy in identifying distracted drivers and It outperforms individual CNNs and demonstrates robustness across different distraction types.

The University of California San Diego (UCSD), specifically their Laboratory of Intelligent and Safe Automobiles, explored various distracted behaviors in their research. These behaviors included adjusting mirrors, tuning the radio, and operating vehicle controls. To monitor the driver's state, they employed vision-based approaches. In a separate study, Martin et al[16] utilized two Kinect cameras to track the hand positions of drivers from both frontal and rear views. Meanwhile, Ohn-Bar et al. divided the image into distinct regions, such as the steering wheel, instrument panel, and gear area. They then combined classifiers to detect hand presence, which allowed them to infer the driver's actual activity[17] The authors also trained three separate models for the three regions and used a second stage classifier for final inference [18]. They also incorporated the eye information along with head and hand information[19]

3 Dataset Description

The development of the Distracted Driver Dataset has been a progressive effort, resulting in two versions that serve as critical resources for driver posture estimation research. The first version, known as the *AUC Distracted Driver Dataset*, was developed to address the limitations of existing datasets, offering an extensive range of driver behaviors with public access — a first in the field[5].



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Fig. 1: Ten Classes of Driver Postures from the Dataset: **(a)**: Drive Safe, **(b)**: Text Left, **(c)**: Talk Left, **(d)**: Text Right, **(e)**: Talk Right, **(f)**: Adjust Radio, **(g)**: Drink, **(h)**: Hair and Makeup, **(i)**: Reaching Behind, **(j)**: Talk to Passenger.

The second version, *Distracted Driver V2*, expanded upon this foundation with increased diversity and precision.

3.1 Dataset Evolution

The initial *AUC Distracted Driver Dataset (V1)* features 17,308 frames from 31 participants representing seven nations, ensuring a comprehensive portrayal of driving behaviors. Recorded using the rear camera of an ASUS ZenPhone, the dataset includes high-resolution imagery that captures various driver activities within multiple vehicle interiors.

Enhancing the breadth and depth of the initial dataset, *Distracted Driver V2* introduced contributions from 44 drivers, with a more granular approach to labeling and class sampling. The second version focused on collecting new data that better represents the different distracted driving behaviors, with a specific emphasis on the division of training and testing sets based on individual drivers, promoting a more robust validation process [20].

Table 1: Comparison of the Distracted Driver V1 and V2 Datasets

	V1	**V2**
Contributions	First dataset for distracted driving. Random split.	More data and drivers. Precise labeling. Driver-based split.
Info	31 drivers	44 drivers
License[1]	License V1	License V2

[1] Dataset is not available for public download; see website for more details.

3.2 Camera Setup

Both datasets employed a consistent camera setup for image capture, prioritizing safety and adaptability to different car interiors. An adjustable arm strap attached the camera to the roof handle above the front passenger seat, and the data collection was conducted while the vehicle was stationary. The datasets' imagery was captured using the rear camera of an ASUS ZenPhone (Model Z00UD), producing videos with a resolution of 1080x1920 pixels. These were later segmented into individual frames. To accommodate various vehicle interiors, the camera was affixed to the roof handle above the front passenger seat using an adjustable arm strap, and data collection was conducted with the vehicle stationary to prioritize safety [20] [5].

3.3 Labeling

Labeling of the dataset was facilitated by a custom-built, multi-platform action annotation tool, which was created using Electron, AngularJS, and JavaScript. This tool's open-source availability significantly enhances the dataset's usability for the research community [21].

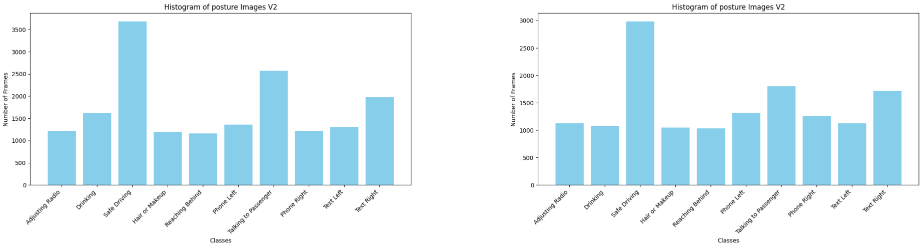
3.4 Statistics

The datasets provide a comprehensive look into distracted driving behaviors, featuring contributions from a diverse pool of participants and a rich distribution of driving postures.

The AUC Distracted Driver Dataset (V1) included 31 participants from 7 different nations, contributing to a diverse representation of driving behaviors. The collected data yielded a total of 17,308 frames,were segmented into the following classes as shown in Figure 1: Adjust Radio, Drink, Drive Safe, Hair & Makeup, Reach Behind, Talk Left, Talk Passenger, Talk Right, Text Left, and Text Right.

Expanding upon V1, the Distracted Driver Dataset (V2) introduced additional data from 44 participants across 7 countries: Egypt, Germany, USA, Canada, Uganda, Palestine, and Morocco. This iteration showcased a gender distribution of 29 males and 15 females and increased environmental diversity by recording in five different car models. A total of 14,478 frames were captured and distributed over the same ten classes. Each frame was labeled manually to ensure accurate representation of the distracted driving behavior [20].

Figure 2 displays the class distributions for the V1 and V2 datasets, providing a visual comparison of the datasets' composition. These histograms are instrumental for understanding the variety of driving postures captured and their respective frequencies, which are crucial for the development of robust distracted driver detection models.



(a) V1 Dataset

(b) V2 Dataset

Fig. 2: Histogram of Posture Images

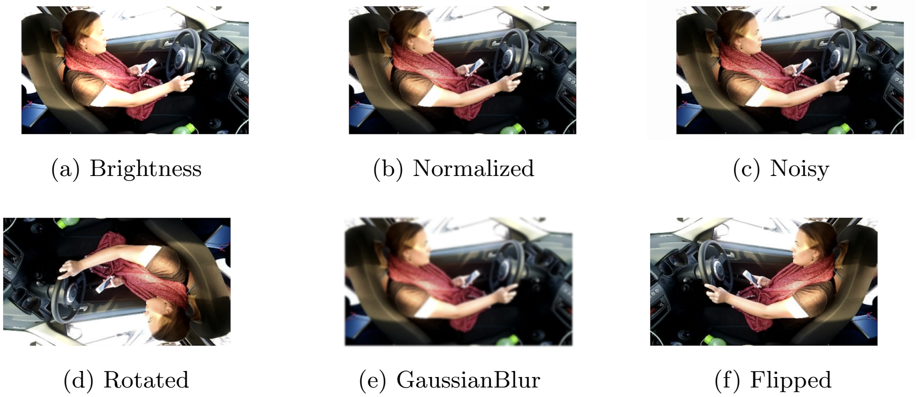
3.5 Benchmarking and Access

Managed by the Machine Intelligence group at the American University in Cairo (MI-AUC), the dataset is intended for use in benchmarking the performance of distracted driver detection systems. Access is granted for academic and research purposes, subject to MI-AUC's approval to maintain the integrity and non-commercial status of the research conducted using this dataset. [22].

3.6 Data Augmentation and MATLAB environment

Initially, our model training utilized the first Distracted Driver dataset and yielded results comparable to those reported in the original paper. For the next stage, we employed the second, more varied dataset to evaluate model performance. Given our primary use of Python, we chose MATLAB for data augmentation, leveraging its extensive capabilities to enrich our dataset artificially. MATLAB's comprehensive support for various augmentation techniques enables us to simulate diverse conditions and perspectives, vital for training resilient machine learning models.

In our approach to enhance the diversity of our image dataset and to improve model generalization, we employed several data augmentation techniques. *Rotation* was used to randomly rotate images by a specified angle, simulating the effect of varying orientations. *Flipping* involved mirroring images horizontally or vertically, which helps the model learn from different perspectives. *Color Jittering* adjusted the brightness, contrast, and saturation of images, preparing the model to handle variations in lighting and color conditions effectively. *Noise Injection* added random noise to images, mimicking real-world imperfections and helping in robust feature learning. *Normalization* scaled pixel values to a uniform range, typically between 0 and 1, standardizing input features for more stable model training. Lastly, *Gaussian Blur* was applied to soften images, reducing high-frequency noise and helping the model focus on more significant structures within the images. Each of these techniques was strategically implemented to simulate real-world variability, thereby enhancing the model's ability to generalize well to new, unseen data. Figure 3 illustrates these augmentation techniques applied to a sample image.



(a) Brightness (b) Normalized (c) Noisy

(d) Rotated (e) GaussianBlur (f) Flipped

Fig. 3: Different types of Data Augmentation

4 Methodology

In this project, our goal was to utilize Convolutional Neural Networks (CNNs), the most widespread type of neural networks for image-related tasks. CNNs have significantly advanced image classification due to their ability to extract detailed features from images by leveraging spatial hierarchies. We trained a CNN model on the *Distracted Driver* dataset, which comprises two versions outlined previously. Initially, we trained our model using *Distracted Driver V1*, and subsequently, for the final phase, we utilized *Distracted Driver V2*, incorporating data augmentation techniques within the MATLAB environment. Our model employs the VGG-16 architecture, a prominent CNN framework designed by Simonyan and Zisserman[6]. We introduced specific modifications to the standard VGG-16 to tailor it for distracted driver detection. VGG-16 is celebrated for its depth and efficacy in deep network construction. The architecture of the original VGG-16 model is depicted in Figure 4.

4.1 VGG16 Architecture

The VGG architecture, formulated by Simonyan and Zisserman from the Visual Geometry Group at the University of Oxford, stands as one of the most prominent models in the computer vision community. Emphasizing simplicity and depth, it excels in both image classification and localization tasks. Among the various iterations of VGG outlined in [6], VGG16 is notable for its performance, consisting of 13 convolutional layers with 3x3 filter sizes. It employs ReLU activation and 2x2 max pooling with a stride of 2, alongside a categorical cross-entropy loss function.

Despite VGG16's efficacy in diverse computer vision applications, its practicality is hindered by the sheer number of parameters, totaling nearly 140 million. The original VGG16 architecture incorporates two dense layers, each with 4096

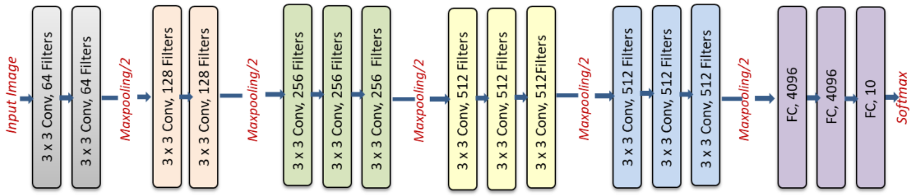


Fig. 4: Original VGG-16 architecture that uses 3x3 convolutions throughout and fully connected layers of dimension 4096 [5]

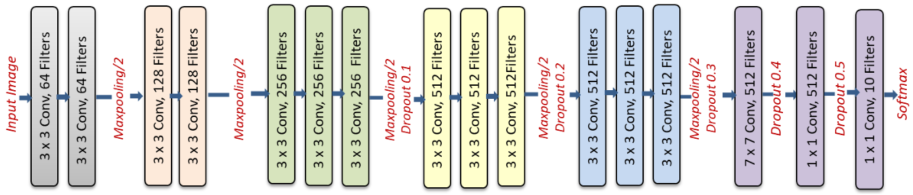


Fig. 5: Fully convolutional VGG-16 Architecture where FC layers are replaced by convolutional layers [5]

units, which are computationally intensive and a significant source of the model's parameters, leading to costly operations. Additionally, this model's reliance on fully connected layers restricts input to fixed sizes and can increase the susceptibility to overfitting, especially as the parameter count escalates.

Given these limitations and the impracticality for deployment on devices with limited resources, we propose adapting the architecture to a fully convolutional neural network, and elaborate on this proposed architecture in the next section.

4.2 Proposed Modified VGG16 Architecure

The original VGG-16 architecture suffers from a significant drawback: its large number of parameters, totaling nearly 140 million. This is primarily due to the presence of fully connected layers, which are both computationally intensive and consume a substantial portion of these parameters. Moreover, networks with fully connected layers are restricted to fixed input sizes. To address these limitations, we propose a solution: replacing the fully connected layers with 1×1 convolutional layers. This modification transforms VGG-16 into a fully convolutional neural network, significantly reducing the number of parameters and enabling it to accommodate varying input sizes. This approach allows for more efficient usage of the network, making it well-suited for our project. The architecture of this modified VGG-16 model is depicted in Figure 5.

4.3 Implementation and Experimental Setup

After loading the images into our dataset, we resize them to a standard size of 224×224 pixels. Additionally, we apply a preprocessing step by subtracting the per-channel mean of the RGB planes from each pixel of the image. This operation effectively centers the cloud of data around the origin along each dimension, aiding in the normalization of the input data.

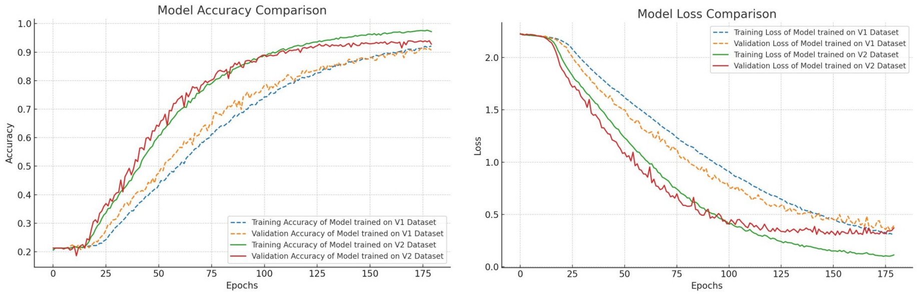
Moving on to the implementation of the modified VGG-16 model, the key distinction from the original architecture lies in the last fully-connected layers. For the initial thirteen convolutional layers, we utilize 3×3 filters, ReLU activation functions, and 2×2 max pooling with a stride of 2. In contrast, the final three layers employ convolutional layers with filter sizes of 7×7 and 1×1, respectively. After each of these layers, dropout regularization is applied to mitigate overfitting, as illustrated in Figure 5. Following the last layer, we incorporate a softmax activation function, which classifies the images into predefined categories. Notably, the initial layers of the CNN serve as feature extractors, while the final layer functions as the classifier. To ensure consistency in the dimensions of the final layer, we incorporate a Global Average Pooling operation at the end.

For the loss function, we opt for categorical cross-entropy due to the classification nature of our task. This loss function quantifies the discrepancy between the predicted class probabilities and the true labels, facilitating effective training of the model.

5 Results and Discussion

The comparative analysis of the two models trained on distinct dataset versions, V1 and V2, reveals significant insights into their performance over epochs, as depicted in Figure 6. The first model, trained on the V1 dataset, shows a consistent decrease in training loss, indicated by the blue dashed line, which is mirrored, though less smoothly, in the validation loss shown in orange. This pattern suggests a reasonable generalization from training to validation data, but with signs of potential overfitting as indicated by the slight divergences between training and validation loss lines.

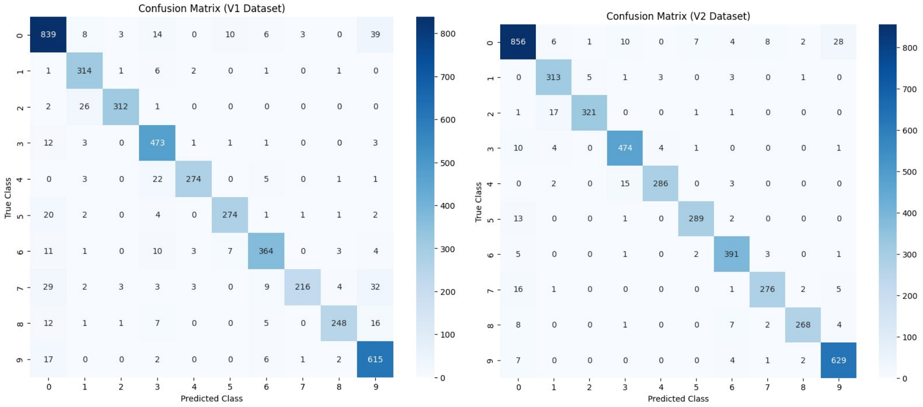
In contrast, the second model, trained on the V2 dataset, demonstrates a sharper decline in both training (green) and validation (red) losses, which converge more closely than those of the first model. This implies not only a faster rate of learning but also better generalization, as the gaps between the training and validation losses are narrower. This demonstrates the enhanced quality of the V2 dataset, which is more diverse and robust, with a larger volume of data and more accurate labels. The accuracy plots further substantiate these findings, with the second model achieving higher validation accuracy more rapidly compared to the first model.



(a) Training and Validation Accuracy

(b) Training and Validation Loss

Fig. 6: Comparison of Model Performance on V1 and V2 Datasets



(a) Model Trained on V1 Dataset

(b) Model Trained on V2 Dataset

Fig. 7: Confusion Matrix

Upon evaluating performance on the test set, the model trained on the V1 dataset achieved an accuracy of 90.72%, whereas the model trained on the V2 dataset surpassed this with a higher accuracy of 94.73%. These results further illustrate the effectiveness of the V2 dataset in enhancing the model's predictive accuracy and overall robustness. Additionally, Figure 7 displays the confusion matrices for models trained on V1 and V2 datasets, offering a visual representation of the models' classification abilities. In both matrices, the concentration of higher values along the diagonals indicates a majority of correct predictions, which is an encouraging sign of model accuracy.

In conclusion, this project demonstrates the enhanced capabilities of convolutional neural networks in detecting distracted driving, significantly improving road safety by accurately identifying driver distractions.

References

1. World Health Organization: Global status report on road safety 2018 (2018) Accessed: April 15, 2024.
2. Federal Communications Commission: Dangers of texting while driving (Accessed 2024) Accessed: April 15, 2024.
3. National Highway Traffic Safety Administration: Distracted driving (Accessed 2024) Accessed: April 15, 2024.
4. Transport Canada: Distracted driving (Accessed 2024) Accessed: April 15, 2024.
5. Baheti, B., Gajre, S., Talbar, S.: Detection of distracted driver using convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. (2018) 1032–1038
6. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
7. Zhang, X., Zheng, N., Wang, F., He, Y.: Visual recognition of driver hand-held cell phone use based on hidden CRF. In: Proceedings of 2011 IEEE International Conference on Vehicular Electronics and Safety. (July 2011) 248–251
8. Das, N., Ohn-Bar, E., Trivedi, M.M.: On performance evaluation of driver hand detection algorithms: Challenges, dataset, and metrics. In: 2015 IEEE 18th International Conference on Intelligent Transportation Systems. (September 2015) 2953–2958
9. Seshadri, K., Juefei-Xu, F., Pal, D.K., Savvides, M., Thor, C.P.: Driver cell phone usage detection on strategic highway research program (shrp2) face view videos. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). (June 2015) 35–43
10. Le, T.H.N., Zheng, Y., Zhu, C., Luu, K., Savvides, M.: Multiple scale faster-rcnn approach to driver cell-phone usage and hands on steering wheel detection. In: Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops. (Jun. 2016) 46–53
11. Park, H., Ahn, D.H., Park, T., Shin, K.G.: Automatic identification of driver's smartphone exploiting common vehicle-riding actions. IEEE Trans. Mobile Comput. **17**(2) (Feb. 2018) 265–278
12. Zhao, C.H., Zhang, B.L., He, J., Lian, J.: Recognition of driving postures by contourlet transform and random forests. IET Intelligent Transport Systems **6**(2) (June 2012) 161–168
13. Zhao, C., Zhang, B., Zhang, X., Zhao, S., Li, H.: Recognition of driving postures by combined features and random subspace ensemble of multilayer perceptron classifiers. Neural Comput. Appl. **22** (2012) 175–184
14. Abbass, M., Ban, Y.: Mobilenet-based architecture for distracted human driver detection of autonomous cars. Electronics **13**(2) (2024) 365
15. Eraqi, H.M., Abouelnaga, Y., Saad, M.H., Moustafa, M.N.: Driver distraction identification with an ensemble of convolutional neural networks. Journal of Advanced Transportation **2019** (2019) Article ID 4125865
16. Martin, S., Ohn-Bar, E., Tawari, A., Trivedi, M.M.: Understanding head and hand activities and coordination in naturalistic driving videos. In: Proc. IEEE Intell. Veh. Symp. Proc. (Jun. 2014) 884–889
17. Ohn-Bar, E., Martin, S., Trivedi, M.M.: Driver hand activity analysis in naturalistic driving studies: Challenges, algorithms, and experimental studies. J. Electron. Imag. **22**(4) (2013) Art. no. 041119

18. Ohn-Bar, E., Trivedi, M.: In-vehicle hand activity recognition using integration of regions. In: Proc. IEEE Intell. Veh. Symp. (IV). (Jun. 2013) 1034–1039
19. Ohn-Bar, E., Martin, S., Tawari, A., Trivedi, M.M.: Head, eye, and hand patterns for driver activity recognition. In: Proc. 22nd Int. Conf. Pattern Recognit. (Aug. 2014) 660–665
20. Abouelnaga, Y., Eraqi, H.M., Moustafa, M.N.: Real-time distracted driver posture classification (2017) Accessed: 2024-03-17.
21. Abouelnaga, Y.: Action annotation tool (2017) Accessed: 2024-03-17.
22. The Machine Intelligence Group at the American University in Cairo: License agreement for the auc distracted driver dataset (2017) For access and usage terms, contact the MI-AUC group.

Table 2: Member Contributions

Name	**Amirhossein Sorour**	**Anahita Jabbari**	**Ghazal Danaee**	**Shima Sargordi**
Role	Coding and Implementation, Research, Writing Report (Methodology section in first and final report, Result and Discussion section in final report), Presentation for Methodology Section, Editing Final Draft of Report	Coding and Implementation, Research, Writing Report (Dataset Description in first and final report), Preparing Presentation Slides, Presentation for Dataset section	Coding and Implementation, Research, Writing Report (Introduction and Previous Works in first and final report), Presentation for Introduction and Previous Work sections	Coding and Implementation, Research, Writing Reports (Discussion and Future work in the first report and Abstract in the final report), Presentation for Results, Discussion, and Future Works sections.