A Hybrid Deep Learning Approach for Driver Distraction Detection

Jimiama Mafeni Mase[1], Peter Chapman[2], Grazziela P. Figueredo[1], Mercedes Torres Torres[1]
[1]School of Computer Science, The University of Nottingham
[2]School of Psychology, The University of Nottingham

***Abstract*—The World Health Organisation reports distracted driving actions as the main cause of road traffic accidents. Current studies to detect distraction postures focus on analysing spatial features of images using Convolutional Neural Networks (CNN). However, approaches addressing both spectral and spatial features of images for driving distraction are scarce. Our hypothesis is that deep learning approaches can further be exploited to consider spatial and spectral features, so that the spatial features capture the spatial information within the image and the spectral features capture the spectral correlations among the image channels. This paper introduces a novel driver distraction posture detection method using CNNs and stacked Bidirectional Long Short Term Memory (BiLSTM) Networks to capture the spectral-spatio features of the images. The proposed methodology consists of two stages: first, we automatically learn the spatial posture features using pre-trained CNNs. Subsequently, we utilise BiLSTMs architecture to extract the spectral features amongst the stacked feature maps from the pre-trained CNNs. Our proposed approach is evaluated on the American University in Cairo (AUC) Distracted Driver Dataset, the most comprehensive and detailed dataset on driver distraction postures to date. Results show that our approach beats state-of-the-art CNN models with an average classification accuracy of 92.7%.**

***Index Terms*—Deep learning, image classification, driving distraction postures, Neural Networks, Spatial features, Road traffic incidents, Spectral features**

I. Introduction

The World Health Organisation (WHO) reported 1.35 million deaths in 2018 due to road traffic accidents worldwide. The WHO report attributes the main causes to violations and distractions, such as over-speeding, harsh cornering, daydreaming, cell phone usage and looking at something outside the car. In an attempt to mitigate this problem, researchers have explored the use of artificial intelligence to understand risky driving behaviours and to develop driver assistance and alert systems [1], [2]. However, the number of road traffic deaths has been continuously increasing over the last few years [3]

With the dramatic increase in computational power, deep neural networks have demonstrated impressive performance in automatically extracting image features for computer vision tasks, such as image classification [4], [5] and object detection [6]–[8]. This has caused a shift in image analysis from hand-crafted feature learning (where features are manually derived using expert knowledge) to deep learning. Deep learning models, specifically Convolutional Neural Networks (CNNs), automatically learn spatial features from images by generating feature maps using sliding windows (i.e., kernels) and filters. Current studies in detecting distraction postures have explored different variations of CNNs to extract the spatial information from images for the classification of driving postures, with promising results. However, some distractions are still very difficult to classify due to their spatial similarities with other postures. Such postures may only be accurately detected by analysing their spectral features, which provide additional information about the images.

In this paper, we propose a deep learning architecture that outperforms current state-of-the-art CNN models when classifying distracted driving postures using static images. Our model consists of concatenated CNN and BiLSTM networks. The CNN networks automatically learn the spatial features of the images and the LSTMs extract the spectral correlations among the feature maps produced by the CNNs. With the spectral and spatial features extracted, our model accurately identifies postures in the AUC Distracted Driver Dataset [9], [10], which is the most comprehensive and detailed publicly available driver distraction dataset.

This paper is organised as follows, in Section II we review the literature on driver distraction detection using deep learning techniques and provide an overview of CNNs and LSTMs architectures. Subsequently, we describe our methodology in Section III. In Section IV, we introduce the publicly available driver distracted dataset, describe hyperparameter optimisation of our model and the evaluation protocol. In Section V, the results are presented along with discussion, and Section VI concludes the paper and establishes the opportunity for future work.

II. Background

A. Related Work

Recent studies on driver distraction detection use deep learning methods, which have proven to outperform traditional machine learning techniques. Kim *et al.* [11] proposed a method of detecting driver distraction using RestNet and MobileNet CNN models. However, their study only focused on two types of distraction: looking in-front and not looking in-front postures. Their results on training the models using fine-tuned pre-trained models significantly outperformed training the models from scratch.

Similarly, Yan *et al.* [12] examined CNNs to detect driver distraction postures. The authors first pre-trained their model using an unsupervised feature learning method called sparse filtering, and subsequently fine-tuned with CNNs for classification. Their model was evaluated on three datasets: the

Southeast University Driving Posture dataset, and two datasets developed by the authors called *Driving-Posture-atNight* and *Driving-Posture-inReal* datasets. The authors claimed that the Driving-Posture-atNight dataset has 29,410 images and Driving-Posture-inReal dataset has 17,730 images. Results showed high classification accuracy with the three driving posture datasets which outperformed methods using hand-crafted features. However, their datasets have only four distraction postures and are not publicly available for benchmarking.

Furthermore, Majdi *et al.* [13] employ CNNs for detecting driver distraction postures. The authors adopted the U-Net CNN architecture for capturing context around the objects. Their model was trained on the American University in Cairo (AUC) Distracted Driver dataset. Their results show great improvement in accuracy compared to Support Vector Classifiers and other CNN architectures. Likewise, Eraqi *et al.* [9] propose a weighted ensemble of CNNs using four different CNN architectures (i.e AlexNet network [14], InceptionV3 [15] networks, ResNet networks [16], and VGG-16 networks [17]). The CNNs are trained on five different image sources of the AUC distracted driver dataset i.e. raw images, skin-segmented images, face images, hands images, and face and hands images. The results from the individual CNNs show the best accuracy when trained on the raw images. Subsequently, the predictions from the different CNNs are combined using a weighted Genetic Algorithm (GA) and the results from the fusion show improved accuracy compared to the independent CNNs and majority voting fusion. However, training these CNNs is extremely costly with large number of parameters: VGG16 , AlexNet , ResNet50, and InceptionV3 models have 134.3 million, 58.3 million, 25.5 million, and 24.3 million parameters respectively.

The studies reviewed above employ different variations of CNNs to identify driver distraction postures. Their limitations regard: (1) the lack of spectral features that capture the correlations among the feature maps; (2) the lack of well-defined objectives as the studies simply evaluate several state-of-the-art CNN models for image classification and select the best performing model. We address these limitations by proposing a concatenated CNN-BiLSTM architecture, where we train only the last few layers of the pre-trained CNN to capture the more specific spatial features of distracted driving and use stacked BiLSTMs for extracting the spectral correlations within the feature maps. Our methodology is motivated by the state-of-the-art performance of LSTMs in hyper-spectral image classification [18], where LSTMs are used to capture correlations among the spectral channels of static images. The authors' hybrid architecture outperformed state-of-the-art CNNs models including the popular 3D-CNN [19] in classifying hyper-spectral images.

B. Overview of CNNs and LSTMs

1) Convolutional Neural Networks: CNNs [20] are neural networks consisting of filtering (or convolution), pooling and activation layers. The inputs go through the convolution layer, where they are filtered to produce stacked smaller dimensional features (feature maps). The stacked feature maps go through the pooling layer, which downsamples the input representations using a sample-based discretisation process.The activation layer later converts the stacked downsampled data into specific features depending on the activation function that is used (e.g. Rectified Linear Unit (ReLU) converts all negative values to zero and maintains all positive values). These filtering, pooling and activation layers allow CNNs to learn hierarchical discriminative features.

Fig. 1 presents a simple CNN architecture with one convolution, one pooling and one activation layer. Most state-of-the-art CNN models [14]–[17] consist of a concatenation of many of such layers with additional units for Batch normalisation and regularisation.

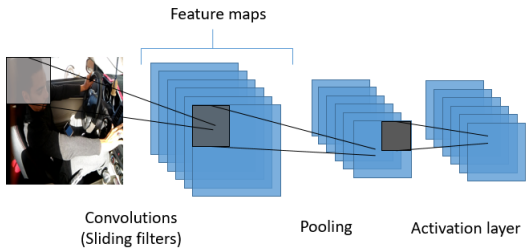

Fig. 1. A simple CNN architecture consisting of convolution, pooling and activation layers

2) Long Short Term Memory Networks: LSTMs [21], [22] are a type of recurrent neural network capable of learning short and long-term dependencies in the data (i.e. connecting previous information to the present task). LSTMs consist of several recurrent neural network layers interacting to produce three core gate layers: forget, input, and output gate layers. The input gate controls which state is updated. The forget gate controls how much information needs to be retained or forgotten, and the output gate decides which part of the cell state is outputted to the next LSTM unit. These gates control information flow into and out of the LSTM cell unit.

Fig. 2 represents a simple LSTM architecture with three inputs: the cell state vector of the previous time step (C_{t-1}), the hidden state vector (h_{t-1}) of the previous time step, and the current input vector (X_t).

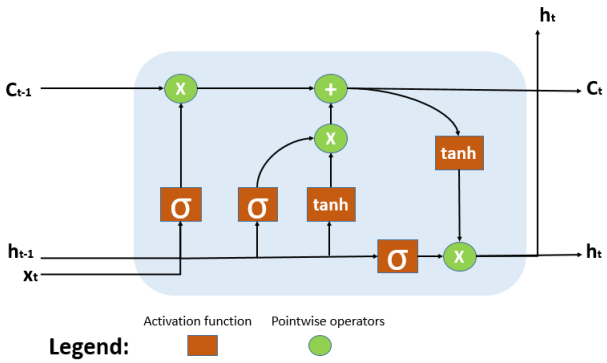

Fig. 2. LSTM cell unit with forget gates

The interactions between the gate layers in the LSTM unit is given by the following equations:

$$
\begin{aligned}
f_t &= \mathrm{sigm}(W_f X_t + U_f h_{t-1} + b_f) \\
i_t &= \mathrm{sigm}(W_i X_t + U_i h_{t-1} + b_i) \\
o_t &= \mathrm{sigm}(W_o X_t + U_o h_{t-1} + b_o) \\
C_t &= f_t \odot C_{t-1} + i_t \odot \tanh(W_c X_t + U_c h_{t-1} + b_c) \\
h_t &= o_t \odot \tanh(C_t)
\end{aligned}
$$

Where f_t is the forget gate's activation vector, i_t is the input or update gate's activation vector, and o_t is the output gate's activation vector. W, U, and b represent the weight matrices and bias vectors which need to be learned during training.

III. Methodology

In this section we present our novel deep learning architecture called C-SLSTM for driver distraction posture detection using CNN and stacked BiLSTMs. C-SLSTM consist of two stages which are trained together end-to-end. The first stage consists of CNNs which extract the spatial features from the posture images and feeds them to the BiLSTMs. The second stage consists of BiLSTMs that learn the spectral correlations among the feature maps to predict the driver's posture. An overview of our proposed solution is shown in Figure 3. We describe each stage in detail below.

A. Pre-trained CNN Inception-V3

Training deep CNNs require large labelled datasets and computational resources, which are not easy to obtain. These issues can be overcome by using deep CNN which have been trained on very large databases for similar tasks (i.e. transfer learning). In this study, we use a pre-trained Inception-V3 CNN model [23] to capture the spatial information in distracted postures. We choose Inception-V3 because of its remarkable performance in image classification and smaller number of parameters (less than 25M parameters) compared to other state-of-the-art pretrained CNN models such as Alexnet and VGGnet.

Inception-V3 is an improved version of Inception [24] with branching within layers that allows abstraction of features at different spatial scales. The model has 16 convolution and mixed layers and a fully connected layer, and is pre-trained on the ImageNet dataset [25] for image classification. In addition, the model has 24.3M trainable parameters. We only train the last 5 layers of the pre-trained network that represent more detailed spatial information of the image. This reduces the number of trainable parameters, thereby, reducing the training cost.

The American University in Cairo (AUC) Distracted Driver dataset consist of 1080 * 1920 images with 3 spectral bands or channels. We preprocessed the images into 299 * 299 with 3 channels for the InceptionV3 CNN model. For each image, the last convolution and mixed layer (known as Mixed_7c) outputs 8 * 8 feature maps with 2048 channels. We remove the fully connected layer and output the feature maps as inputs to the LSTM networks.

B. Stacked Bidirectional Long Short-Term Memory

We use stacked BiLSTMs to learn the spectral features of driving posture images. The output of the CNNs (i.e., 8 * 8 feature maps with 2048 channels) is fed to the BiLSTMs. The BiLSTMs extracts the spectral features in the forward and backward directions by learning the correlations across the channels of the feature maps. This produces two output sequences (one for each direction). We use multiple hidden states to learn the spectral features at deeper spatial scales. The output sequences of the BiLSTMs are concatenated and passed to a fully connected layer to classify the images.

IV. Experiments

In this section we introduce the AUC distracted driver dataset used to evaluate our approach. We also describe the hyperparameters of our model and the evaluation protocol.

A. The American University in Cairo Distracted Driver Dataset

The AUC Distracted Driver dataset [9], [10] is the largest, most comprehensive publicly available dataset for driver distraction identification. The dataset captures most real-world distracted driving postures (up to 10 postures): safe driving (c0), text right (c1), right phone usage (c2), text left (c3), left phone usage (c4), adjusting radio (c5), drinking (c6), reaching behind (c7), hair or makeup (c8), and talking to passenger (c9). The dataset was captured using an ASUS ZenPhone rear camera (Model Z00UD), and consists of 1080 * 1920 pixel images. The dataset contains information for 44 drivers. 38 drivers are used in the training set and 6 drivers in the test set. Table I shows the number of images in the training and test sets for each driving posture.

TABLE I
DESCRIPTION OF AUC DISTRACTED DRIVER DATASET

Types of driving postures	Number of images in training set	Number of images in test set
c0	2,440	266
c1	1,305	133
c2	862	114
c3	744	100
c4	950	90
c5	753	90
c6	733	63
c7	691	63
c8	698	66
c9	1,379	138

B. Metrics

Table II presents the hyper-parameters for the BiLSTMs. The optimisation algorithm (optimiser) trains the neural network by minimising the sum of errors between the predicted values and actual values, i.e. the cost function. The learning rate controls how the weights are updated with respect to the estimated error. Dropout is a regularisation technique to

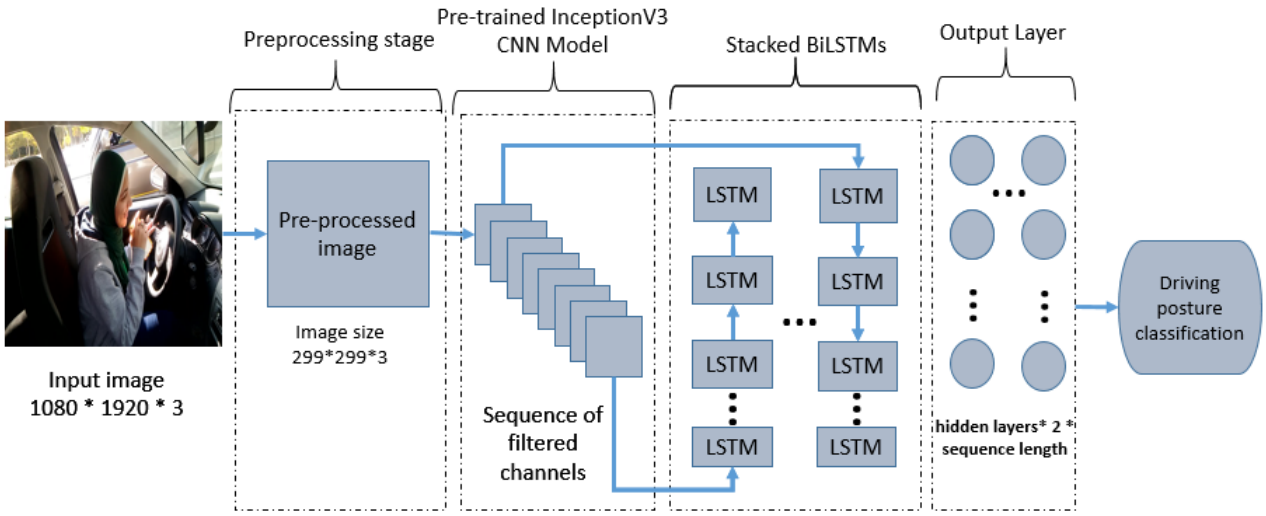

Fig. 3. Our proposed CNN-BiLSTM architecture to detect driving distraction postures

TABLE II
DETAILED CONFIGURATION OF THE BILSTM

Parameters	Tested values	BiLSTM
Input features	64	64
Hidden size	32, 64, 128	128
Number of layers	1, 2, 3	1
Batch size	16, 32, 64	32
Dropout	No, 0.5, 0.6, 0.7	No
Learning rate	0.00001, 0.0001, 0.001	0.0001
Optimizer	Adam, SGD	Adam

reduce overfitting, where network nodes are dropped during training. Batch size defines the number of instances to be propagated (i.e. forward propagation) before updating the model's parameters. Number of layers are the number of LSTM memory cells, and hidden size is the number of hidden states. The levels of abstraction of features increases over time proportionally to the hidden states. The input features represent the size of each feature map (8 * 8).

C. Optimisation

With the range of hyperparameters in Table II, we carried out experiments using our hybrid model by training and validating on the AUC distracted driving posture dataset. The AUC dataset contains a training set (38 drivers) and a test set (6 drivers) as described in Section IV-A. We split the training set by driver into new training (80% of the training set) and validation (20% of the training set) sets. Therefore, drivers in the training set are not found in the validation and test sets. The validation data was used to obtain the optimal hyperparameters. The test set was used to benchmark our model with the state-of-the-art CNN models discussed in the literature that used the same datasets and trained variations of InceptionV3. By evaluating the validation loss of each hyperparameter, the following optimal hyperparameter values were obtained: input size = 64, hidden size = 128, number of layers = 1, batch size = 32, dropout = No, learning rate = 0.0001, and optimiser = Adam. The experiments were executed on a graphics processing unit (GPU) using 4 CPU cores and 6GB RAM. Our code was implemented in Pytorch with an epoch size of 50 for each experiment.

Due to space constraints, we only present the validation loss of the optimizers and learning rates as these have great effect on the learning process of neural networks. Fig. 4. shows the validation loss when the model is evaluated using Adam and Stochastic Gradient Descent (SGD) optimizers. Adam optimizer clearly yields better performance with faster convergence compared to SGD. Similarly, Fig. 5. shows the validation loss of the model when evaluated with three learning rates i.e. 0.00001, 0.0001 and 0.001 . The learning rate of 0.0001 performs better than the rest after 30 epochs.

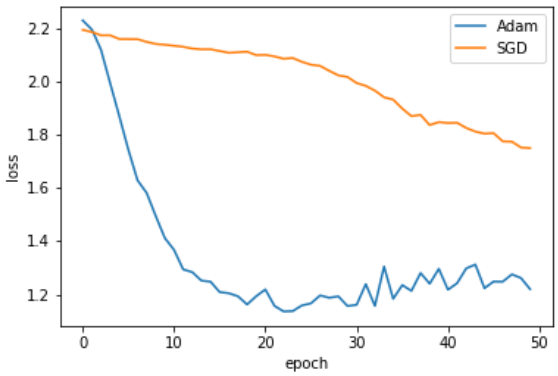

Fig. 4. Selecting optimisation algorithm

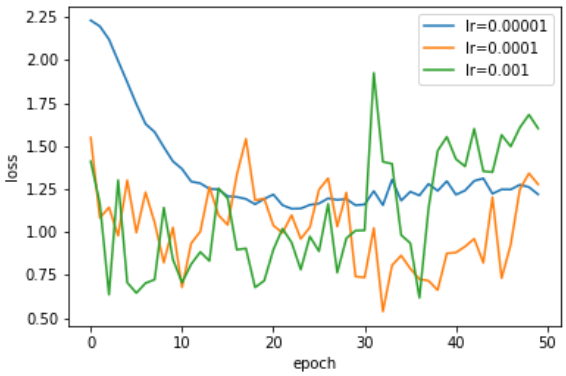

Fig. 5. Selecting learning rate

D. Evaluation

After optimising our model, we evaluated the classification performance using the average confusion matrix, average classification accuracy, average precision, average recall and average F1-score across the different distracted driving postures after 20 runs.

V. Results and Discussion

To provide a comprehensive evaluation of performance, we compare our methodology with the reported results of state-of-the-art CNN models which have been benchmarked on the AUC distracted posture dataset (i.e. CNN VGG-16 [9], CNN Resnet50 [9], and ensemble of InceptionV3 CNNs using Genetic Algorithm (GA) [9]). We also trained and benchmarked our model with a CNN InceptionV3 and CNN InceptionV3 + 1-directional stacked LSTMs. Table III presents the average classification accuracy of C-SLSTM on the test split of the AUC distracted posture dataset after 20 runs with comparison to state-of-the-art CNN models, CNN InceptionV3 CNN and CNN InceptionV3 + 1-directional stacked LSTMs.

Our model, C-SLSTM, beats state-of-the-art CNN models with an average classification accuracy of 92.7% (standard deviation of 0.94%) and average Negative Log-Likelihood (NLL) of 0.279 (standard deviation of 0.023) after 20 runs. An average precision of 92.8% (std = 0.45), recall of 92.7% (std = 0.46) and f1-score of 92.8% (std = 0.45) was achieved. In addition, our model significantly outperforms the InceptionV3 CNN model due to its ability to learn the spatial and spectral features of the images. Also, extracting the correlations between the channels in the forward and backward directions (bi-directional) further improves the classification of one-directional LSTMs from 89.8% to 92.7%.

The confusion matrices of our model and InceptionV3 Genetic Algorithm (GA) are shown in figures. 6 and 7 respectively. We observe that the most misclassified postures by the InceptionV3 GA model are "reaching behind" (c7) and "talking to passenger" (c9) with a 19.27% false prediction. The model appears to mistake "reaching behind" for "talking to passenger" postures. This is because the driver's head and body have the similar spatial positions in both postures, as shown in Fig. 8. Our approach, however, distinguishes between these postures with far more accuracy, reaching 1.5% false detection. Furthermore, the "talking to passenger" (c9) posture is the least correctly identified posture in the InceptionV3 GA model, with an accuracy of 76.6%. This posture is correctly classified by our model with an accuracy of 92.5%. And the least identified posture, i.e., "safe driving" (c0) is recognised by our approach with an accuracy of 86.5%. Lastly, InceptionV3 GA model has overall more false predictions above 5% (i.e. 7 in total) compared to our model, which has only 3 false predictions above 5% (indicated by the grey colour fillings on the matrices). Therefore, extracting both temporal and spatial features of images helps to better identify driving postures than the InceptionV3 GA model and other state-of-the-art CNN models for the dataset investigated.

TABLE III
Driver distraction classification results compared to state-of-the-art methods using the AUC 'split-by-driver' Distracted Driver test dataset

Model	Loss (NLL)	Accuracy (%)
VGG-16 [9]	1.2466	76.13
Resnet50 [9]	0.6615	81.69
Ensemble of InceptionV3 with GA-Weighted algorithm [9]	0.6400	90.06
InceptionV3	0.5723	84.41
InceptionV3-LSTM	0.4445	89.82
C-SLSTM	**0.2793**	**92.70**

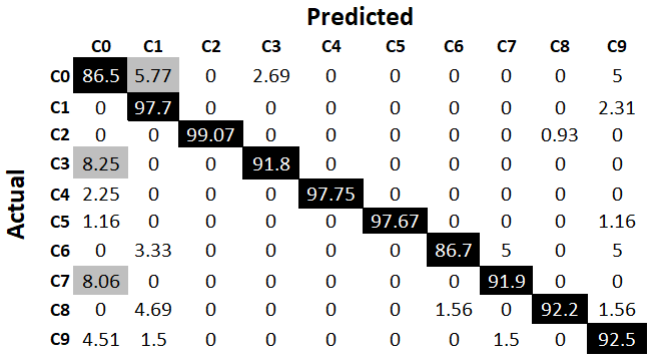

Fig. 6. Confusion matrix of C-SLSTM on AUC distracted driving postures test dataset after 20 runs

Actual \ Predicted	C0	C1	C2	C3	C4	C5	C6	C7	C8	C9
C0	87.57	10.98	0	0.29	0	0	0.29	0.86	0	0
C1	0	96.71	0	1.88	0	0	0	1.41	0	0
C2	0	0	97.42	0	0	0	0	0.52	0.52	1.55
C3	8.33	0	0	91.67	0	0	0	0	0	0
C4	11.76	0	0	0	87.06	0	0	0	1.18	0
C5	0	0	0	0	0	100	0	0	0	0
C6	0.67	0	0	0	0	0	81.82	10.49	6.99	0
C7	0	0	0	0	0	0	0	100	0	0
C8	2.74	0	0.68	0	3.42	0	0	7.53	84.93	0.68
C9	1.38	2.75	0	0	0	0	0	19.27	0	76.6

Fig. 7. Confusion matrix of ensemble InceptionV3 GA network on AUC distracted driving postures test dataset depicted from [9]

Fig. 8. Most confusing driving postures for InceptionV3 GA network

VI. Conclusions and Future Work

Distracted driving is one of the major causes of road traffic accidents worldwide. Therefore, monitoring and detecting driver distraction postures can help in the development of Advanced Driver-Assistance and alert systems to mitigate the problem. In this paper, we presented a hybrid deep learning technique that captures the spatial-spectral features of images for the classification of distraction postures. Our architecture outperforms current state-of-the-art CNN models in detecting distracted driving, with an accuracy of 92.7% when trained and tested on the publicly-available AUC distracted driver dataset.

For future work, we plan on exploring optimisation techniques to further reduce model complexity and parameters. This will be essential for the development of real-time detection systems. In addition, our model is limited in detecting new types of distracted postures i.e. distracted postures which are not found in the AUC distracted driver dataset. Therefore for future work, we plan on exploring unsupervised anomaly detection techniques for distinguishing between "safe driving" and "distracted driving". Lastly, we plan on acquiring video or sequential data of driving distraction to improve detection by capturing the temporal dynamics of naturalistic driving.

References

[1] G. P. Figueredo, U. Agrawal, J. M. M. Mase, M. Mesgarpour, C. Wagner, D. Soria, J. M. Garibaldi, P.-O. Siebers, and R. I. John, "Identifying Heavy Goods Vehicle Driving Styles in the United Kingdom," *IEEE Transactions on Intelligent Transportation Systems*, 2018.

[2] U. Agrawal, J. M. Mase, G. P. Figueredo, C. Wagner, M. Mesgarpour, and R. I. John, "Towards real-time heavy goods vehicle driving behaviour classification in the united kingdom," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 2330–2336.

[3] "Global status report on road safety: World Health Organization; 2018. https://www.who.int/violence_injury_prevention/road_safety_status/2018/."

[4] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1891–1898.

[5] T. Xiao, T. Xia, Y. Yang, C. Huang, and X. Wang, "Learning from massive noisy labeled data for image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2691–2699.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[7] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks," *IEEE Geoscience and remote sensing letters*, vol. 11, no. 10, pp. 1797–1801, 2014.

[8] W. Ouyang, X. Wang, X. Zeng, S. Qiu, P. Luo, Y. Tian, H. Li, S. Yang, Z. Wang, C.-C. Loy *et al.*, "Deepid-net: Deformable deep convolutional neural networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2403–2412.

[9] H. M. Eraqi, Y. Abouelnaga, M. H. Saad, and M. N. Moustafa, "Driver distraction identification with an ensemble of convolutional neural networks," *Journal of Advanced Transportation*, vol. 2019, 2019.

[10] Y. Abouelnaga, H. M. Eraqi, and M. N. Moustafa, "Real-time distracted driver posture classification," *arXiv preprint arXiv:1706.09498*, 2017.

[11] W. Kim, H.-K. Choi, B.-T. Jang, and J. Lim, "Driver distraction detection using single convolutional neural network," in *2017 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2017, pp. 1203–1205.

[12] C. Yan, F. Coenen, and B. Zhang, "Driving posture recognition by convolutional neural networks," *IET Computer Vision*, vol. 10, no. 2, pp. 103–114, 2016.

[13] M. S. Majdi, S. Ram, J. T. Gill, and J. J. Rodríguez, "Drive-net: Convolutional network for driver distraction detection," in *2018 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*. IEEE, 2018, pp. 1–4.

[14] Z.-W. Yuan and J. Zhang, "Feature extraction and image retrieval based on alexnet," in *Eighth International Conference on Digital Image Processing (ICDIP 2016)*, vol. 10033. International Society for Optics and Photonics, 2016, p. 100330E.

[15] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[16] S. Targ, D. Almeida, and K. Lyman, "Resnet in resnet: Generalizing residual architectures," *arXiv preprint arXiv:1603.08029*, 2016.

[17] W. Yu, K. Yang, Y. Bai, T. Xiao, H. Yao, and Y. Rui, "Visualizing and comparing alexnet and vgg using deconvolutional layers," in *Proceedings of the 33 rd International Conference on Machine Learning*, 2016.

[18] Q. Liu, F. Zhou, R. Hang, and X. Yuan, "Bidirectional-convolutional lstm based spectral-spatial feature learning for hyperspectral image classification," *Remote Sensing*, vol. 9, no. 12, p. 1330, 2017.

[19] S. Hussein, K. Cao, Q. Song, and U. Bagci, "Risk stratification of lung nodules using 3d cnn-based multi-task learning," in *International conference on information processing in medical imaging*. Springer, 2017, pp. 249–260.

[20] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE transactions on neural networks*, vol. 8, no. 1, pp. 98–113, 1997.

[21] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[22] D. Rengasamy, H. P. Morvan, and G. P. Figueredo, "Deep learning approaches to aircraft maintenance, repair and overhaul: a review," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 150–156.

[23] A. Sajjanhar, Z. Wu, and Q. Wen, "Deep learning models for facial expression recognition," in *2018 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2018, pp. 1–6.

[24] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[25] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.