Face2Statistics: User-Friendly, Low-Cost and Effective Alternative to In-Vehicle Sensors/Monitors for Drivers

Zeyu Xiong, Jiahao Wang, Wangkai Jin, Junyu Liu, Yicun Duan, Zilin Song, and Xiangjun Peng

User-Centric Computing Group, University of Nottingham, Ningbo, China
https://unnc-ucc.github.io

Abstract. We present **Face2Statistics**, a comprehensive roadmap to deliver user-friendly, low-cost and effective alternatives for extracting drivers' statistics. **Face2Statistics** is motivated by the growing importance of multi-modal statistics for Human-Vehicle Interaction, but existing approaches are user-unfriendly, impractical and cost-ineffective. To this end, we leverage **Face2Statistics** to build a series of Deep-Neural-Network-driven predictors of multi-modal statistics, by taking facial expressions as input only. We address two outstanding issues of the current design, and then (1) leverage HSV color space; and (2) Conditional Random Field to improve the robustness of **Face2Statistics** in terms of prediction accuracy and degree of customization. Our evaluations show that, **Face2Statistics** can be effective alternatives to sensors/monitors for Heart Rate, Skin Conductivity and Vehicle Speed. We also perform the breakdown analysis to justify the effectiveness of our optimizations. Both source codes and trained models of **Face2Statistics** are online at https://github.com/unnc-ucc/Face2Statistics.

Keywords: Human-Vehicle Interaction · Vision Computing · Image Processing · Deep Neural Networks.

1 Introduction

Modern Human-Vehicle Interaction systems leverage comprehensive statistics of drivers, to perform complex decision-making procedures for personalized Human-Vehicle Interactions. Particularly, the importance of in-vehicle drivers' statistics grows significantly, with the emerging trends of Autonomous Vehicles. To this end, developing mechanisms to provide user-friendly, low-cost and effective methods to extract in-vehicle drivers' statistics, is essential. However, existing methods, for extracting in-vehicle statistics of drivers, directly integrate either respective sensors (e.g. Heart Rate) or vehicle monitors (e.g. Vehicle Speed), which imposes two major challenges in practice.

First, driver experiences would be endangered, if sensors/monitors are directly equipped or deployed for in-vehicle drivers. Existing mechanisms extract drivers' statistics by equipping or deploying sensors/monitors

within vehicles. Though such methods are straightforward and easy-to-deploy, there are huge risks to degrade drivers' feelings regard to comforts, satisfaction and etc. For instance, to monitor drivers' Heart Rate, a Heart Rate sensor needs to be equipped with the driver along the way. However, such arrangements require physical-contact deployments, and this would indeed cause side effects on driving experiences. Furthermore, such side effects can grow exponentially with the number of sensors/monitors, if there are multiple types of statistics to be monitored.

Second, the costs and overheads of directly integrating sensors or monitors for in-vehicle drivers are unexpectedly high. Existing mechanisms require additional data transfer between sensors/monitors and in-vehicle computers, so that Human-Vehicle Interaction systems can take advantage of these statistics. Hence, there are two major issues for the adaptions. One is that extra data transfer, obtained from equipped sensors/monitors, can lead to huge performance degradation in terms of latency, which further affects Quality-of-Services; The other is that re-routing statistics of vehicle statuses, retrieved from in-vehicle monitors, can cause extra costs since internal modifications of vehicle structures might be required.

Therefore, it's apparent that user-friendly, low-cost and effective alternatives, for extracting drivers' statistics, are needed. **Our goal** is to present a comprehensive road-map on how to deliver a series of such alternatives, so that they can **obtain drivers' statistics via an unobtrusive manner**. We make the **key observation** that recent advances of Computer Vision techniques (e.g. by applying Deep Neural Networks for image classifications) can potentially serve as suitable alternatives in such purposes. Such techniques can leverage a large volume of historical data to train models, and then predict future statistics via these models by using only a subset of types from the historical data.

Based on such characteristics, we can choose a unified and unobtrusive source as the input streams, and leverage it to derive different types of drivers' statistics. To this end, we introduce **Face2Statistics**, a series of user-friendly, low-cost and effective alternatives for extracting drivers' statistics. The **key idea** of **Face2Statistics** is leveraging each frame of facial expressions, to predict instant status of in-vehicle drivers. Such prediction is powered by sophisticated, customized and pre-trained Deep Neural Network models, and we achieve this by exploring and examining various types of state-of-the-art Deep Neural Network models. We also design a comprehensive execution pipeline of **Face2Statistics**, so that **Face2Statistics** can be easily visualized/exported/integrated with other sub-systems for Human-Vehicle Interactions, as the sources of drivers' statistics.

Moreover, we address two major inefficiencies of **Face2Statistics** in practice, and provide viable solutions to them: (1) one problem of **Face2Statistics** is the effects of illuminance, which incurs variations of different models in terms of effectiveness; and (2) the other problem of **Face2Statistics** is that different drivers might have distinctive facial features during driving procedures, which substantially degrades the robustness of **Face2Statistics** if we apply the same model for every driver. To strengthen the robustness of **Face2Statistics**, we

introduce two novel optimizations of **Face2Statistics**: (1) to mitigate the effects of illuminance, we leverage HSV encoding instead of commonly-used RGB encoding, for each frame of video captures; and (2) to take advantage of differences across individual drivers, we provide customization supports through *Conditional Random Field*, so that different models can have personalized customization, according to drivers' variations.

We quantitatively examine the performance and robustness of **Face2Statistics** over a state-of-the-art open-sourced data set for Human-Vehicle Interaction. The experimental results demonstrate significant benefits of our approach. Averaged across all drivers, the best of **Face2Statistics** achieves up to 58.44%, 72.68% and 70.25% accuracy in terms of Heart Rate, Skin Conductivity and Vehicle Speed predictions. Compared with two other models, the best of **Face2Statistics** improve the accuracy by 4.52%, 2.68% and 1.41% in terms of Heart Rate, Skin Conductivity and Speed predictions. We also perform the breakdown analysis to illustrate how our proposed optimizations improve the robustness of **Face2Statistics**. We show that both techniques can significantly improve the robustness of **Face2Statistics**. We believe both our proposal and related optimizations are essential in practice.

We make the following three contributions in this paper:

- We propose **Face2Statistics**, a comprehensive road-map to deliver user-friendly, low-cost and effective alternatives for extracting drivers' statistics. We also provide a comprehensive execution pipeline of **Face2Statistics**, for ease of visualization/exportation/integration with other sub-systems for Human-Vehicle Interactions, as the sources of drivers' statistics.
- We address two major issues of **Face2Statistics** for robustness, and introduce two techniques to mitigate them accordingly. We first leverage HSV encoding, rather than RGB, for better robustness against the effects of illuminance; and then we add *Conditional Random Field* to achieve customization supports for different individuals, to improve the robustness as well.
- We quantitatively examine the performance and robustness of **Face2Statistics**, and the results demonstrate significant benefits of our approach. **Face2Statistics** can achieve highly accurate predictions, in terms of Skin Conductivity, Heart Rate and Vehicle Speed. We also demonstrate that our optimizations can greatly enhance the robustness of **Face2Statistics**.

The rest of this paper is organized as follow. Section 2 introduces the background of **Face2Statistics**. Section 3 gives an overview of **Face2Statistics** and elaborates each component one-by-one. Section 4 elaborates key optimizations of **Face2Statistics** to enhance the robustness. Section 5 presents the experimental methodology to evaluate **Face2Statistics**. Section 6 provides the results and detailed analysis of our experiments. Section 7 gives the discussion in terms of experimental results, model limitation and driver's privacy issues. Section 8 introduces the related works and summary of **Face2Statistics**.

2 Background

In-vehicle drivers' statistics play a significant role in the designs and implementations of modern Human-Vehicle Interaction Systems. In the past two decades, there has been a large volume of research efforts being paid to leverage drivers' statistics, to obtain a better understanding of driving behaviors for personalized Human-Vehicle Interactions [9]. These statistics can be used for various purposes and we only list several examples hereby. In-vehicle drivers' statistics can be used for but not limited to the following: (1) Indicating the safety of road trips [37], (2) Classifying the driving styles and behaviors [18, 36], (3) Classifying the health conditions for old drivers [27], (4) Validating the accuracy of self-reporting properties [30], (5) Detecting driving distraction [28], (6) Detecting road surface and hazard [19]. Therefore, to enhance the driving experience, In-vehicle drivers' statistics are considered essential building blocks in various example usages, and we envision they are becoming more important in the near future of Autonomous Vehicles.

The collection of in-vehicle data is therefore particularly important. For different types of in-vehicle data, researchers have used a variety of measurement methods, which can be classified into two main types: (1) Integrating/deploying specific devices, sensors, and monitors for certain statistics, (2) Re-routing statistics based on a vehicle's internal makeup. For the first type, an example application is heart rate / skin conductivity data collection by applying biosensors. However, wearing biosensors can often be distracting to drivers while driving, which is user-unfriendly and impractical. For the second type, rerouting data (e.g. vehicle speed) from dashboard to human-vehicle interaction system (in digital format) is not cost-effective. We introduce the related works in detail in Section 8. In summary, existing methods for the extraction of In-vehicle Drivers' Statistics are user-unfriendly, impractical and cost-ineffective. Therefore, providing user-friendly, low-cost and effective alternatives to sensors/monitors becomes essential for Human-Vehicle Interaction systems in practice. Our goal is to present a comprehensive road-map on how to deliver a series of such alternatives, so that they can derive drivers' statistics via an unobtrusive manner.

3 Face2Statistics: An Overview

In this section, we give an overview of the designs and implementations of **Face2Statistics**. We elaborate details regarding key components of **Face2Statistics**, including: (1) Image Processing (Section 3.1); (2) Neural Network Models (Section 3.2); and (3) Visualization/Integration Supports (Section 3.3). Particularly, we illustrate multiple attempts to obtain the best prediction models, by exploiting different state-of-the-art Deep Neural Network models (Section 3.2).

3.1 Component 1: Facial Expressions from Video Streams

The first component of **Face2Statistics** is a pre-processing component. The reason why we need to pre-process is that: raw video streams, obtained from an

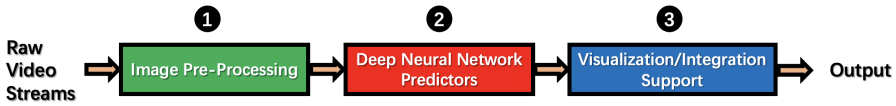

Fig. 1. Execution Pipeline and Components of **Face2Statistics**.

unobtrusive camera, contain many noisy pixels in addition to facial expressions. Therefore, the pre-processing component needs to support two functionalities. First, the pre-processing component needs to retrieve only the facial expressions rather than all pixels within a frame; and second, the pre-processing component may need other techniques to enhance the performance of **Face2Statistics**, such as adjusting encoding schemes of images, varying the size of images, etc.

We implement this component leveraging OpenCV library [5], which is a state-of-the-art image processing library. In total, we achieve three functionalities in our current prototypes: (1) we normalize image color channels to 32×32 pixels of matrix, as recommended by mainstream Computer Vision datasets (e.g. CIFAR-10 [22]) and the-state-of-the-art models (e.g. [21]); (2) we implement a facial detector to crop facial expressions, so that we can only retrieve relevant pixels; and (3) we implement a transformation tool of color matrices, so that we can represent frames in different encoding schemes (e.g. RGB, HSV and etc.)[1].

3.2 Component 2: Deep Neural Network-driven Predictors

The core component of **Face2Statistics** is the Deep Neural Network-driven predictors. The models are the key to achieve accurate predictions of relevant statistics, via only facial expressions. Therefore, it's critical to determine which models fit the best in our scenarios. Hereby, we perform an in-depth comparison across different kinds of models shown in this section, where our comparisons are backed up by our quantitative evaluations (i.e. Section 6.1).

3.2.1 First Attempt: Convolution Neural Networks (CNNs) Convolution Neural Networks (CNNs) achieve great performance in image classification, and representative examples include ResNet [11], DenseNet [14], SENet [13], etc. Though these models are well-suited for image classifications, there is still a gap to determine which one is better in our scenarios. By analyzing relevant data streams, we observe that In-Vehicle Drivers' Statistics (i.e. sequential data) exhibit a high volume of noises and they are potentially misguidance during the inference procedure. Therefore, we believe a model, which can enhance the weights of the particular features, is the best choice among all CNN models.

[1] We use this functionality to investigate how we can enhance the robustness of **Face2Statistics**, as described in Section 4

Hence, our first choice is DenseNet as the representative model derived from CNN models[2]. This is because the architecture of DenseNet provides a simple but effective mechanism to enhance feature expressions: DenseNet takes all inputs from the early layers, merges them and feeds them into the next layer as inputs. More specifically, if the model has N layers in total, a conventional CNN architecture has only N connections, while DenseNet has $\frac{N(N+1)}{2}$ connections. This optimization strengthens the effects of feature extraction from previous layers, mitigates the gradient-vanishing issues, promotes feature reuse, emboldens feature reproduction and significantly reduces the number of parameters.

3.2.2 Second Attempt: Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) Our second choice is Long-Short-Term Memory Recurrent Neural Network model (LSTM-RNN). This is motivated by the fact that, In-vehicle drivers' statistics are sequential in a continuous timeline. Therefore, Recurrent Neural Networks (RNN) might be a good fit since the design of RNN emphasizes the effects of sequential data streams. To better speculate long-term impacts, Long Short-Term Memory RNN (LSTM-RNN) architecture is proposed to address this challenge [12]. LSTM-RNN mitigates the issues of long-term memory dependency, and the key idea of LSTM-RNN is to selectively determine whether a state shall be kept or forgotten, instead of only considering the most recent states (i.e. as previous RNN models). This provides a more appropriate usage to predict sequential data streams.

3.2.3 Third Attempt: Bidirectional Long-Short-Term Memory (BiLSTM) Recurrent Neural Network (RNN) Our third choice is Bidirectional Long-Short-Term Memory Recurrent Neural Network (BiLSTM-RNN). This is motivated by the fact that, LSTM-RNN is unable to encode information flow from front to back, in the sequence of data streams. To optimize LSTM-RNN, Bidirectional LSTM-RNN (BiLSTM-RNN) is proposed to address such challenges [10]. The key idea of BiLSTM-RNN is to enable both forward and backward across the flow of all layers, so that the learning procedure can be enhanced significantly.

3.3 Component 3: Visualization/Exportation of Predicted Results

The final component of **Face2Statistics** is to provide efficient visualization/exportation of predicted results. Since all previous components of **Face2Statistics** are within the in-vehicle systems, there is no need for extra data transfer or re-rountine. Therefore, we consider three parts of this component: (1) we provide a basic data-processing component for reorganizing/streaming results into flat files, to serve as system logs/records; (2) we provide a high-level API so that other systems can directly deploy **Face2Statistics** via a simple function call; and (3) we also build a visualization example using the API, to justify the feasibility of visualizations.

[2] A prior work showcases that DenseNet is effective in this scenario [15], and an exhaustive study [1] demonstrates the benefits via comparisons between this design and others.

4 Customizing Face2Statistics for Different Individuals

In this section, we introduce two optimization methods for **Face2Statistics** to enhance its robustness in practice. We first elaborate the problem of illumination effects and our solution to this issue (Section 4.1), and then we identify the issues of individual variations and introduce a novel method to customize **Face2Statistics** for different individuals (Section 4.2). Figure 2 shows the pipeline of the detailed optimization.

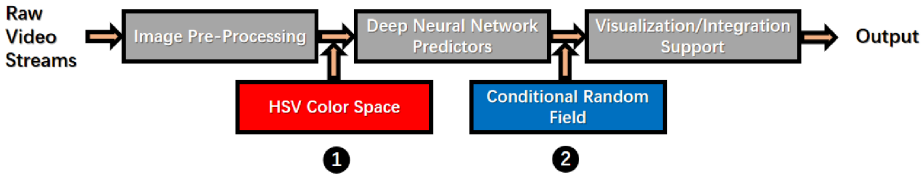

Fig. 2. Pipeline: Detailed Optimization.

4.1 Optimization 1: Mitigating the Effects of Illumination

During our empirical studies, we observe that the effects of illumination can significantly impact the performance of **Face2Statistics**. This is because the environmental factors of driving procedures can have significant variations in different scenarios, and the illumination is an outstanding example. In this section, we elaborate this issue via a concrete example (Section 4.1.1), and provide a viable solution to it (Section 4.1.2). This optimization is backed up by our quantitative evaluations (i.e. Section 6.2)

4.1.1 Issue: The Curse of Illumination Illumination has significant impacts on the quality of facial expressions, as one of the most important environmental factors during the driving procedure. Considering a video stream (i.e. which consists of a sequence of frames/images), the evaluation of facial expression can be highly unstable. This is realistic since there are multiple driving scenarios causing illumination impacts (e.g. weather, daylight, tunnels and etc.). Therefore, the cause of illumination can be highly diverse and such a complicated situation is very hard to be resolved. Therefore, it is critical to provide a robust version of **Face2Statistics** to mitigate the side effects of illumination. To this end, we make the key observation that the impacts of illumination directly reflect on the values from the Color Matrix (i.e. the digital representation of each frame/image). Therefore, we conjecture that it might be possible to mitigate the issues of illumination by using a more stable and robust encoding scheme of frames/images.

4.1.2 Key Idea: Use HSV color space, instead of RGB The conventional representation of frames/images is to leverage "Red, Green and Blue" (RGB) channels, to visualize the images in a digital manner. Though this approach provides chromatic visualizations of frames/images, the impacts of illumination are also significant in this case. This is because that the illumination can significantly vary the values of each channel in RGB, and substantially lead to the degradation of the robustness in terms of **Face2Statistics** since the values are unstable.

Therefore, we propose to leverage "Hue, Saturation, Value" (HSV), instead of RGB, as the color space for **Face2Statistics**. The details of HSV are as follow. HSV color space is represented as a circular cylinder, where a color is defined in cylinder's coordinates [3]. Unlike RGB color space (i.e. using all three color channels to represent colors for each pixel in an image), HSV color space only uses the channel H (Hue) to represent a color, and the other two attributes (i.e. S (Saturation) and V (Value)) indicate the intensity and the brightness of the color. Illumination change can lead to all three channels of RGB change significantly, while for HSV only Saturation and Value will be affected. Therefore, HSV is more stable than RGB in terms of illumination variation, because the number of affected channels, by illumination, is decreased in HSV.

4.1.3 A Comparative Example between RGB and HSV We use a comparative example, by visualizing the scenarios using both RGB and HSV, to give a more straightforward comparison between two approaches. We consider a scenario as follow. A driver is driving to enter a tunnel, where we choose two timespots at "before entering the tunnel" and "after entering the tunnel". In the pair of these timespots, we visualize the captured frames/images using RGB and HSV respectively, and then compare these two pairs to justify the impacts of illumination. Figure 3 presents the visualized frames/images for this example. It's evident that, RGB encoding can be significantly impacted by the illumination since the light is lowered, but HSV has a more robust visualization in this case.

Fig. 3. A comparative example between RGB (left) and HSV (right): the visualized frames/images "before entering the tunnel" and "after entering the tunnel".

Therefore, in this case, HSV color space reduces the variance of illumination among different pixels. The reason behind it can be elaborated into two aspects.

1. If illumination of one color changes, for RGB color space pixel value in all three channels may change in a considerable range, while for HSV color space only saturation and value changes, fluctuation on hue channel has been minimized.
2. HSV color space can better represent the difference between two colors, while color difference in RGB martix may not distinctive.

4.2 Optimization2: Customizing Face2Statistics for Different Individuals

4.2.1 Issue: One Face2Statistics for Everyone is Impractical The core component of **Face2Statistics** is the Neural Network model, and it's clear that a single model can not achieve the best performance for all individuals. This is because the facial expressions of different individuals vary significantly due to their distinctive inborn facial features. To this end, we make the key observation that personalized parameters are deployed successfully in other use cases, such as estimation of self-reported intensity of pains [26], the prediction of driving states with joint time series [41], the prediction of driving behaviors in lane changes [6] and etc. Therefore, we envision that it might also be important to provide customization/personalization supports for **Face2Statistics** as well.

4.2.2 Key Idea 1: Customization via Conditional Random Field Conditional Random Field (CRF) is a conditional probability distribution model, by assuming the output distribution constituting a Markov random field [43]. More specifically, under the condition of the specific parameter (i.e. quantified by corresponding metrics), the output would be constrained based on the probabilistic distribution. Leveraging Hammersley Clifford theorem [38], the joint probability distribution, represented by linear chain CRF, can be expressed as the function of adjacent nodes. Therefore, CRF provides the constraints to ensure that the final prediction results are valid. These constraints can be automatically learned by the CRF when training the models.

Since existing Neural Network models don't have relevant constraints, based on different individuals while performing the inference. Therefore, we propose to leverage CRF for customizing different models. The key insight here is that, if we can measure the differences of the individuals quantitatively, we can feed these quantified values into CRF. Then the models can be adjusted during the training procedure, and they are customized according to these quantifications. In our case, we consider a standard pairwise CRF model, as shown in Equation 1 [3].

[3] In Equation 1, A_μ is the association (observation matching) potential for modeling dependencies between the class label m_μ and the set of all observations μ. x_μ is the real-valued SVM response on the pixel (or node) μ. N_μ is the neighborhoods of pixels μ (a subset of the full spatial coordinate system S from above). $I_{\mu\nu}$ is

$$p(m|x) = \frac{1}{Z} \exp\left(\sum_{\mu \in S} A_u(m_\mu, x_\mu) + \sum_{\mu \in S} \sum_{\nu \in N_\mu} I_{\mu\nu}(m_\mu, m_\nu, x_\mu)\right), \quad (1)$$

4.2.3 Example: Models w/out CRF or w/ CRF Figure 4 provides a comparative example between model structures with/without CRF supports. In Figure 4, W_1, ..., W_n represent the input values, which can be regarded as different attributes respectively. Figure 4-(A) demonstrates the workflow of existing model prediction (i.e. models without CRF supports). There are three steps. ❶ all input values $W_{1..n}$ are passed into all models; ❷ the models perform the inference to generate a matrix of attributes; and ❸ the generated matrix of attributes are considered as results, and they might be inappropriate since the aggregations from all models may produce an unreasonable sequence.

On the contrary, to validate the output sequence, CRF provides the constraints to adjust the results. Figure 4-(B) demonstrates the workflow of CRF-integrated model prediction (i.e. models with CRF supports). There are four steps. ❶ all input values $W_{1..n}$ are passed into all models; ❷ the models perform the inference to generate a matrixes of attributes; ❸ the generated matrix of attributes are adjusted via CRF, and such adjustments rely on an assigned parameter (i.e. we denoted it as the personalized parameter); and ❹ the adjusted matrix of attributes are considered as results. Therefore, based on the personalized parameter, we can adjust the output tags to ensure the order of the tag results more reasonable, which achieves the customization/personalization of **Face2Statistics**.

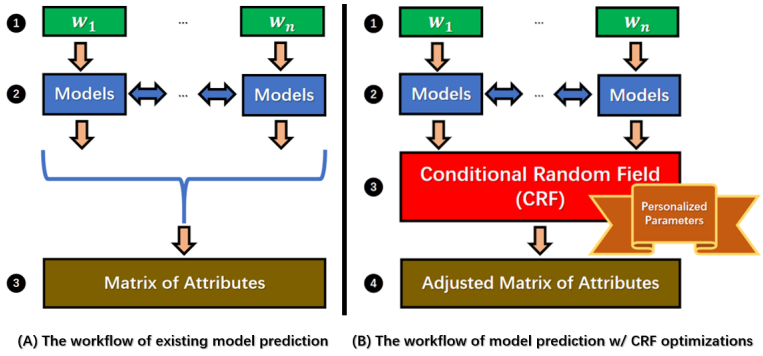

Fig. 4. A comparative example of model workflow, between (A) models without CRF and (B) models with CRF.

the interaction (local-consistency) potential for modeling dependencies between the levels of neighboring elements. Z is the partition function: a normalization coefficient (sums over possible labels)

4.2.4 Key Idea 2: Measuring Personalized Parameters via Pearson Correlation Coefficients Since CRF requires the personalized parameter to guide the adjustments, it's important to decide how to quantitatively measure the differences between individuals' facial expressions. To provide a quantitative examination of personalized parameters, we choose Pearson Correlation Coefficients (PCC) as the metric. PCC examines the correlation between two matrices, and it scales well when the data consists of more dimensions [20].

We elaborate how to use PCC in this case. Assuming a pair of data sets, X and Y, the distance is represented by the difference between targeted matrices, and the attributes are pixels from difference image matrices respectively. Therefore, we construct the examination of the two-dimensional correlation, and the values of the correlation range from -1 to 1. Equation 2 shows the detailed formulation of PCC, where X_i is a sample in set X, Y_i is a sample in set Y, $\overline{X}$ is the average value of set X, $\overline{Y}$ is the average value of set Y.

$$PCC = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}}. \tag{2}$$

5 Experimental Methodology

In this section, we introduce details about our experimental methodology. We first cover the details of our implementations (Section 5.1), and then we elaborate details with regard to datasets and models (Section 5.2).

5.1 Implementation Details

We implement **Face2Statistics** using: (1) OpenCV library for data pre-processing; and (2) Tensorflow for all different Neural Network Models and the CRF model. We use these to ensure the practicality of **Face2Statistics** prototype. We evaluate different variants of **Face2Statistics** using a machine with Intel multi-core i9-9700 CPU and an AMD Radeon PRO 5600 GPU.

5.2 Dataset and Neural Network Models

We use the BROOK dataset [16, 29], an open-sourced and multi-modal dataset with facial videos for adaptive and personalized Human-Vehicle Interaction designs, to evaluate **Face2Statistics**. BROOK contains 11 dimensions of data, covering drivers' physiological statistics and vehicle status, which are collected from 34 drivers in a 20-minute driving process. We choose three dimensions of data streams: Vehicle Speed, Skin Conductivity and Heart Rate, with drivers' facial images as the input. Skin Conductivity and Heart Rate are representative data to monitor drivers' physiological status (e.g. stress, distraction), and Vehicle Speed is an intuitive data type to reflect the instantaneous driving context in different time spots. Note that any type of data streams can be used in **Face2Statistics**, as long as there are historical information to train the models. We cover all models, as described in Section 3.2, during our experiments.

6 Experimental Results

In this section, we present the experimental results and the analysis. Our evaluations aim to answer three questions: ❶ which Neural Network models, for **Face2Statistics**, has the best performance for different types of data streams? (Section 6.1) ❷ how our optimization, using HSV instead of RGB, would benefit **Face2Statistics**? (Section 6.2) ❸ how our optimization, adding CRF supports, would benefit **Face2Statistics**? (Section 6.3)

6.1 Comparisons among Different Neural Network Models

Figure 5 reports the results of training and validation accuracy, in terms of DenseNet, LSTM and BiLSTM. We make three key observations. First, BiLSTM, LSTM and DenseNet achieve the best validation accuracy, in terms of Skin Conductivity, Heart Rate and Vehicle Speed respectively. This is because different types of data streams exhibit different symptoms, and the choices of models may vary as well. Second, the trend of training accuracy mostly correlates with the trends of validation accuracy. The only exception occurs when using BiLSTM for Skin Conductivity. This is because Skin Conductivity has a high frequency of occurrence in the time series, and BiLSTM has an advantage in terms of such structure. Third, all models achieve an extremely high training accuracy, across all types of data streams. The lowest precision of training accuracy occurs, when applying LSTM for Heart Rate prediction (i.e. 96.8%).

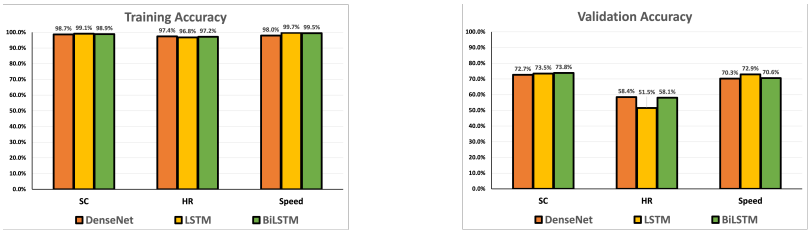

Fig. 5. Training (left) and Validation Accuracy (right) of all models, for Skin Conductivity (SC), Heart Rate (HR) & Vehicle Speed (Speed).

6.2 Comparisons between HSV and RGB

Figure 6 reports the results of training and validation accuracy, in terms of applying HSV and RGB on BiLSTM. We make two key observations. First, BiLSTM, using HSV, achieves better validation accuracy in terms of all types of data streams (i.e. Skin Conductivity, Heart Rate and Vehicle Speed), compared with BiLSTM using RGB. More specifically, BiLSTM, using HSV, improves the validation accuracy by 1.0%, 4.9% and 1.1% for the predictions of Skin Conductivity, Heart Rate and Vehicle Speed. This is because HSV provides more robust representation in terms of illumination, and this provides consistent benefits in prediction. Second, the trends of training accuracy can't reflect the trends of

validation accuracy. More specifically, for the predictions of Heart Rate, the validation accuracy are positively correlated with the training accuracy. However, for the predictions of Skin Conductivity, the validation accuracy is negatively correlated with the training accuracy. We believe this research direction for further characterizations deserves more efforts.

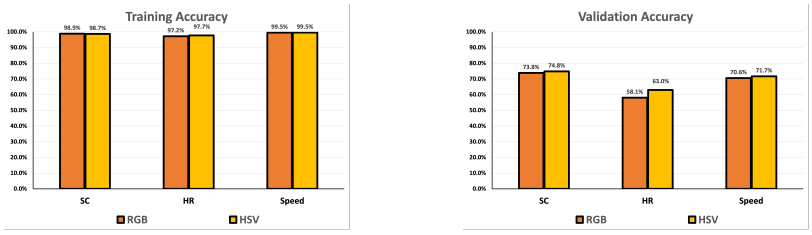

Fig. 6. Training (left) and Validation Accuracy (right) of BiLSTM using RGB or HSV, for Skin Conductivity (SC), Heart Rate (HR) & Vehicle Speed (Speed).

6.3 Comparisons between Models w/ or w/out CRF

Figure 7 shows the comparative validation accuracy of BiLSTM w/ and w/out CRF support for four different drivers. The improvement of validation accuracy in all the cases proves the effectiveness of CRF for customization support. We notice that the prediction accuracy of the model with CRF is higher than that without CRF for all attributes. More specifically, BiLSTM, with CRF, increases the average validation accuracy by 0.95%, 0.44%, and 0.82% of Skin Conductivity, Heart Rate and Speed. This is because CRF provides customization/personalization of **Face2Statistics**, and this enables **Face2Statistics** to adapt different individuals accordingly.

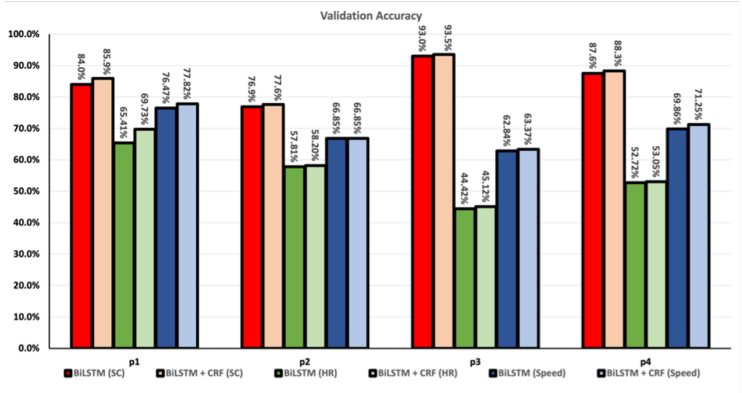

Fig. 7. Validation Accuracy of BiLSTM w/ and w/out CRF for different individuals.

6.4 Takeaways from Experimental Results

We discuss our findings from the experimental studies. First, there is no one-size-fits-all solutions to all types of data streams, in terms of Neural Network model selections. Our evaluations show that both feature-reuse (i.e. DenseNet) and sequence-awareness (i.e. LSTM and BiLSTM) are important in particular cases. Second, image representation can have great impacts in terms of validation accuracy, and it's essential to consider a suitable one in driving scenarios. Our evaluations suggest that HSV is a better choice to mitigate the impacts of illumination, compared to RGB. Third, customization/personalization supports are also essential for practical deployments, and there are more rationales to be formed for guiding future efforts. Our evaluations suggest that CRF has great potentials in practice, to provide customization/personalization supports for different individuals.

7 Discussion

7.1 Model Limitations

To address the limitations of the current models, we optimize **Face2Statistics** by utilizing HSV color space to mitigate the effects of illumination, and using CRF to customize for different individuals. Beyond the optimizations described in our paper, we believe there are still an extensive amount of opportunities to explore in the future. There are mainly two aspects. First, it's potential to exploit *Adaptation in different driving scenarios*. Currently we only consider general driving scenarios in the context, we can add more information to **Face2Statistics** (e.g. weather, road congestion, etc.) in the future. Furthermore, these information can also make **Face2Statistics** to assist other contexts in Human-Vechile Interaction (e.g. User Trust [34], Driving Styles [42], etc.) Therefore, **Face2Statistics** can be much more personalised for training and prediction. Consequently, this requires a high degree of adaptation of the data set for re-collection and more robust deep neural networks to classify multiple attributes. Second, enabling *Adaptation in different angles of facial expression video streams* is also promising for future works. In practice, **Face2Statistics** can be used in different vehicles and drivers. For different drivers, there are often different seat adjustments. In this case, though we can set a instruction for users to place the camera in the correct position, the camera angle of the driver's facial expression may still change when adapting to different types of vehicles. This can also be an important contributor to affect the processing of facial expressions and the robustness of **Face2Statistics**. Some works to explore detailed facial expressions (e.g. [39]) can help with this issue. Moreover, exploiting advanced simulation infrastructure and toolkits (e.g. [17, 32, 33, 40]) for data collection is also important for developing effective data-driven Human-Vehicle Interactive systems.

7.2 Driver's Privacy and Ethics

Privacy and ethical issues of **Face2Statistics** can be considered as general problems in the context of Human-Vehicle Interaction. For users, before data collec-

tion, **Face2Statistics** demands the approval from drivers (e.g. sign terms-of-use) and guarantees that there are no abuse of all collected data for other purposes. For systems, to achieve the real-time prediction, **Face2Statistics** needs to capture driver's facial expression, and transfers raw data to the back-end. There are two types of back-ends: (1) online servers and (2) local machines. For an online server, the data transmission procedure can incur potential issues of privacy protection. This is because both facial expression images and corresponding attributes may be vulnerable to leakage. A potential solution is to encode the driver's facial data by blurring before the data transformation, and decode when the data has been received by the server. Some works to explore advanced privacy protection to address this issue (e.g. [23–25]). As for a local machine, avoiding data transmission provides a natural guarantee for data isolation, but this requires more computational power on the local machine to facilitate with the needs of local computation. Therefore, enabling data protection in the context of privacy and ethical issues for Human-Vehicle Interaction systems like **Face2Statistics** is an interesting direction to balance the tradeoffs, and there are already some works starting to explore this part (e.g. [8]).

8 Related Works

In this section, we give an overview related works of **Face2Statistics**. We firstly introduce existing methods for extracting drivers' statistics (Section 8.1). Next, we compare existing methods and identify key limitations of them in practice (Section 8.2). Finally, we compare existing methods with **Face2Statistics** to justify the novelty of our work (Section 8.3).

8.1 Existing Methods for Extracting Drivers' Statistics

In order to collect In-vehicle drivers' statistics, existing methods can be grouped into two approaches. One approach is to directly integrate/deploy particular devices for specific types of statistics (e.g. Heart Rate Monitor); and the other approach is to re-routine statistics from internal compositions of a vehicle (e.g. Vehicle Speed). In the following, we discuss these two approaches in details.

1. Integrating/Deploying particular devices/sensors/monitors for specific types of statistics. Previous efforts (as covered in Section 2) show personal statistics of In-Vehicle Drivers are important to explore a large design space. To obtain personal statistics of In-Vehicle Drivers, existing solutions directly integrate/deploy particular devices/sensors/monitors to retrieve specific types of statistics. We use commonly-used types of statistics to elaborate, including Heart Rate, Skin Conductivity and Eye-Tracking. (1) to supply Heart Rate/Skin Conductivity statistics, drivers need to equip these bio-sensors for data collections [2, 7, 35]; and (2) to supply Eye-Tracking statistics, both drivers and vehicles need to equip parts of the Eye-Tracking system for data collection [31].

2. Re-routing statistics from internal compositions of a vehicle. Previous efforts also demonstrate the importance of vehicle statuses, while a particular

driver is driving. To obtain vehicle statistics from a specific driver, existing solutions directly re-routine statistics from internal composition of a vehicle. For instance, to supply Vehicle Speed statistics, existing methods need to re-routine the information (i.e. displayed in dash board) to digital formats for Human-Vehicle Interaction systems.

8.2 Limitations of the Existing Methods

Existing methods are sufficient to leverage In-Vehicle Drivers' Statistics, to explore the design space of their usages. However, there are three major issues to adapt these methods: (1) user-unfriendliness of equipping devices while driving; (2) practical challenges of integrating/deploying devices; and (3) extra costs of vehicle design and manufacture. Note that these issues can further endanger the practicality of more designs, which consider In-Vehicle Drivers' Statistics as the source of data inputs. Hereby, we identify these issues and discuss them in details.

1. User-unfriendliness of equipping devices while driving. Existing approaches can lead to user-unfriendliness when collecting In-Vehicle Drivers' Statistics. As covered in Section 8.1, the collection of personal statistics requires drivers to equip multiple devices/sensors/monitors. Though it's acceptable for in-lab simulations/filed studies, it's not user-friendly for the majority of drivers in practice (e.g. equipping Heart Rate/Skin Conductivity monitors). Moreover, such integration/deployments might lead to more serious issues. For instance, the deployments of Eye-Tracking devices might affect the driver's field of vision, which incurs risks of driving safety in practice [4].

2. Practical challenges of integrating/deploying devices. Existing approaches have significant challenges of integrating/deploying devices in practice. As covered in Section 8.1, the collection of both personal and vehicle statistics demands a massive amount of extra data transfer across individual systems. Such transfer can incur large overheads in supplying statistics for Human-Vehicle Interaction systems. Though it might be applicable for off-line analysis during in-lab/field studies, this design choice is very difficult to deliver a high level of Quality-of-Services in practice.

3. Extra costs of vehicle design and manufacture. The only way to ensure a high-level of Quality-of-Services, is to incorporate the methods as parts of vehicle designs and manufactures. However, this is very challenging and can incur extra costs. For currently available devices/sensors/monitors, merging them into one single vehicle demand significant efforts in both hardware and software. As for in-vehicle statistics, the digitization of relevant statistics (e.g. Vehicle Speed, etc.) might also require extra efforts and costs in production.

Therefore, we can summarize the above three limitations as (1) user-unfriendly in practice; (2) impractical in production; and (3) not cost-effective. These limitations can further endanger the practicality of advanced Human-Vehicle Interaction designs, which relies on In-vehicle Drivers' Statistics.

8.3 Novelty of Our Approach

mechanism addresses all limitations mentioned above. Firstly, taking facial expressions only as input is **user-friendly**, as drivers do not need to equip biosensors which distracts them while driving. Next, with regard to data transfer, **Face2Statistics** only requires a video stream of the user's facial expressions and all in-vehicle data is handled via model prediction, which can significantly **reduce the cost** of data transfer and the subsequent risks of data leakage. Second the ease of use of **Face2Statistics** makes it a great **advantage in terms of vehicle designs and manufactures**. **Face2Statistics** can predict drivers' statistics in real time by simply installing a video recorder in the vehicle. Third, **Face2Statistics** enables personalised models for different drivers and eliminates the effects of illumination on facil expressions.

9 Conclusions

We propose, optimize and evaluate **Face2Statistics**. **Face2Statistics** provides a comprehensive road-map to deliver user-friendly, low-cost and effective alternatives for extracting drivers' statistics. Using **Face2Statistics**, in-vehicle drivers are not required to equip sensors/monitors for extracting relevant statistics, which improves the user-friendliness, practicality and efficiency. **Face2Statistics** takes facial expressions as the only input, and provides effective predictions of relevant statistics. We identify two major issues in terms of robustness, and provide viable solutions to optimize **Face2Statistics**. Our evaluations confirm the effectiveness of **Face2Statistics** in representative data types, and justify the benefits of our proposed optimizations. We release both source codes and trained models of **Face2Statistics** online at https://github.com/unnc-ucc/Face2Statistics.

10 Acknowledgements

We thank anonymous reviewers in HCI'22 and AutomotiveUI'21 for their valuable feedback. We thank for all members of User-Centric Computing Group at University of Nottingham Ningbo China for the stimulating environment. An earlier version of this work is at [15].

References

1. Abbas, Q., Alsheddy, A.: A methodological review on prediction of multi-stage hypovigilance detection systems using multimodal features. IEEE Access **9**, 47530–47564 (2021). https://doi.org/10.1109/ACCESS.2021.3068343
2. Asada, H.H., Shaltis, P., Reisner, A., Rhee, S., Hutchinson, R.C.: Mobile monitoring with wearable photoplethysmographic biosensors. IEEE engineering in medicine and biology magazine **22**(3), 28–40 (2003)
3. Berk, T., Brownston, L., Kaufman, A.: A new color-namiing system for graphics languages. IEEE Annals of the History of Computing **2**(03), 37–44 (1982)

4. Blignaut, P.J., Beelders, T.R.: Trackstick: a data quality measuring tool for tobii eye trackers. In: Morimoto, C.H., Istance, H.O., Spencer, S.N., Mulligan, J.B., Qvarfordt, P. (eds.) Proceedings of the 2012 Symposium on Eye-Tracking Research and Applications, ETRA 2012, Santa Barbara, CA, USA, March 28-30, 2012. pp. 293–296. ACM (2012). https://doi.org/10.1145/2168556.2168619, https://doi.org/10.1145/2168556.2168619
5. Bradski, G., Kaehler, A.: Learning OpenCV: Computer vision with the OpenCV library. " O'Reilly Media, Inc." (2008)
6. Butakov, V.A., Ioannou, P.: Personalized driver/vehicle lane change models for adas. IEEE Transactions on Vehicular Technology **64**(10), 4422–4431 (2014)
7. Dao, D., Salehizadeh, S.M., Noh, Y., Chong, J.W., Cho, C.H., McManus, D., Darling, C.E., Mendelson, Y., Chon, K.H.: A robust motion artifact detection algorithm for accurate detection of heart rates from photoplethysmographic signals using time–frequency spectral features. IEEE journal of biomedical and health informatics **21**(5), 1242–1253 (2016)
8. Duan, Y., Liu, J., Jin, W., Peng, X.: Characterizing Differentially-Private Techniques in the Era of Internet-of-Vehicles. Technical Report-Feb-03 at User-Centric Computing Group, University of Nottingham Ningbo China (2022)
9. Erzin, E., Yemez, Y., Tekalp, A.M., Erçil, A., Erdogan, H., Abut, H.: Multimodal person recognition for human-vehicle interaction. IEEE MultiMedia **13**(2), 18–31 (2006)
10. Graves, A., Mohamed, A.r., Hinton, G.: Speech recognition with deep recurrent neural networks. In: 2013 IEEE international conference on acoustics, speech and signal processing. pp. 6645–6649. Ieee (2013)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
12. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997)
13. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
14. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017)
15. Huang, Z., Li, R., Jin, W., Song, Z., Zhang, Y., Peng, X., Sun, X.: Face2multimodal: In-vehicle multi-modal predictors via facial expressions. In: 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. p. 30–33. AutomotiveUI '20, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3409251.3411716, https://doi.org/10.1145/3409251.3411716
16. Jin, W., Duan, Y., Liu, J., Huang, S., Xiong, Z., Peng, X.: BROOK Dataset: A Playground for Exploiting Data-Driven Techniques in Human-Vehicle Interactive Designs. Technical Report-Feb-01 at User-Centric Computing Group, University of Nottingham Ningbo China (2022)
17. Jin, W., Ming, X., Song, Z., Xiong, Z., Peng, X.: Towards emulating internet-of-vehicles on a single machine. In: AutomotiveUI '21: 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Leeds, United Kingdom, September 9-14, 2021 - Adjunct Proceedings. pp. 112–114. ACM (2021). https://doi.org/10.1145/3473682.3480275, https://doi.org/10.1145/3473682.3480275

18. Khodairy, M.A., Abosamra, G.: Driving behavior classification based on oversampled signals of smartphone embedded sensors using an optimized stacked-lstm neural networks. IEEE Access **9**, 4957–4972 (2021). https://doi.org/10.1109/ACCESS.2020.3048915, https://doi.org/10.1109/ACCESS.2020.3048915
19. Kortmann, F., Hsu, Y., Warnecke, A., Meier, N., Heger, J., Funk, B., Drews, P.: Creating value from in-vehicle data: Detecting road surfaces and road hazards. In: 23rd IEEE International Conference on Intelligent Transportation Systems, ITSC 2020, Rhodes, Greece, September 20-23, 2020. pp. 1–6. IEEE (2020). https://doi.org/10.1109/ITSC45102.2020.9294684, https://doi.org/10.1109/ITSC45102.2020.9294684
20. Kosov, S., Shirahama, K., Grzegorzek, M.: Labeling of partially occluded regions via the multi-layer crf. Multimedia Tools and Applications **78**(2), 2551–2569 (2019)
21. Krizhevsky, A., Hinton, G.: Convolutional deep belief networks on cifar-10. Unpublished manuscript **40**(7), 1–9 (2010)
22. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images (2009)
23. Liu, J., Jin, W., He, Z., Ming, X., Duan, Y., Xiong, Z., Peng, X.: HUT: Enabling High-UTility, Batched Queries under Differential Privacy Protection for Internet-of-Vehicles. Technical Report-Feb-02 at User-Centric Computing Group, University of Nottingham Ningbo China (2022)
24. Martin, S., Tawari, A., Trivedi, M.M.: Balancing privacy and safety: Protecting driver identity in naturalistic driving video data. In: Boyle, L.N., Burnett, G.E., Fröhlich, P., Iqbal, S.T., Miller, E., Wu, Y. (eds.) Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Seattle, WA, USA, September 17 - 19, 2014. pp. 17:1–17:7. ACM (2014). https://doi.org/10.1145/2667317.2667325, https://doi.org/10.1145/2667317.2667325
25. Martin, S., Tawari, A., Trivedi, M.M.: Toward privacy-protecting safety systems for naturalistic driving videos. IEEE Transactions on Intelligent Transportation Systems **15**(4), 1811–1822 (2014)
26. Martinez, D.L., Rudovic, O., Picard, R.: Personalized automatic estimation of self-reported pain intensity from facial expressions. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 2318–2327. IEEE (2017)
27. Nishiuchi, H., Park, K., Hamada, S.: The relationship between driving behavior and the health condition of elderly drivers. Int. J. Intell. Transp. Syst. Res. **19**(1), 264–272 (2021). https://doi.org/10.1007/s13177-020-00240-3, https://doi.org/10.1007/s13177-020-00240-3
28. Omerustaoglu, F., Sakar, C.O., Kar, G.: Distracted driver detection by combining in-vehicle and image data using deep learning. Appl. Soft Comput. **96**, 106657 (2020). https://doi.org/10.1016/j.asoc.2020.106657, https://doi.org/10.1016/j.asoc.2020.106657
29. Peng, X., Huang, Z., Sun, X.: Building BROOK: A Multi-modal and Facial Video Database for Human-Vehicle Interaction Research pp. 1–9 (2020), https://arxiv.org/abs/2005.08637
30. Porter, M.M., Smith, G.A., Cull, A.W., Myers, A.M., Bédard, M., Gélinas, I., Mazer, B.L., Marshall, S.C., Naglie, G., Rapoport, M.J., et al.: Older driver estimates of driving exposure compared to in-vehicle data in the candrive ii study. Traffic injury prevention **16**(1), 24–27 (2015)

31. Silva, N., Blascheck, T., Jianu, R., Rodrigues, N., Weiskopf, D., Raubal, M., Schreck, T.: Eye tracking support for visual analytics systems: foundations, current applications, and research challenges. In: Krejtz, K., Sharif, B. (eds.) Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications, ETRA 2019, Denver , CO, USA, June 25-28, 2019. pp. 11:1–11:10. ACM (2019). https://doi.org/10.1145/3314111.3319919, https://doi.org/10.1145/3314111.3319919
32. Song, Z., Wang, S., Kong, W., Peng, X., Sun, X.: First attempt to build realistic driving scenes using video-to-video synthesis in OpenDS framework. In: Adjunct Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2019, Utrecht, The Netherlands, September 21-25, 2019. pp. 387–391. ACM (2019). https://doi.org/10.1145/3349263.3351497, https://doi.org/10.1145/3349263.3351497
33. Song, Z., Duan, Y., Jin, W., Huang, S., Wang, S., Peng, X.: Omniverse-OpenDS: Enabling Agile Developments for Complex Driving Scenarios via Reconfigurable Abstractions. In: International Conference on Human-Computer Interaction (2022)
34. Sun, X., Li, J., Tang, P., Zhou, S., Peng, X., Li, H.N., Wang, Q.: Exploring Personalised Autonomous Vehicles to Influence User Trust. Cogn. Comput. **12**(6), 1170–1186 (2020). https://doi.org/10.1007/s12559-020-09757-x, https://doi.org/10.1007/s12559-020-09757-x
35. Tamura, T., Maeda, Y., Sekine, M., Yoshida, M.: Wearable photoplethysmographic sensors—past and present. Electronics **3**(2), 282–302 (2014)
36. Toledo, T., Lotan, T.: In-vehicle data recorder for evaluation of driving behavior and safety. Transportation Research Record **1953**(1), 112–119 (2006)
37. Toledo, T., Musicant, O., Lotan, T.: In-vehicle data recorders for monitoring and feedback on drivers' behavior. Transportation Research Part C: Emerging Technologies **16**(3), 320–331 (2008)
38. Wallach, H.M.: Conditional random fields: An introduction. Technical Reports (CIS) p. 22 (2004)
39. Wang, J., Xiong, Z., Duan, Y., Liu, J., Song, Z., Peng, X.: The Importance Distribution of Drivers' Facial Expressions Varies over Time!, p. 148–151. Association for Computing Machinery, New York, NY, USA (2021), https://doi.org/10.1145/3473682.3480283
40. Wang, S., Liu, J., Sun, H., Ming, X., Jin, W., Song, Z., Peng, X.: Oneiros-OpenDS: An Interactive and Extensible Toolkit for Agile and Automated Developments of Complicated Driving Scenes. In: International Conference on Human-Computer Interaction (2022)
41. Xing, Y., Lv, C., Cao, D., Lu, C.: Energy oriented driving behavior analysis and personalized prediction of vehicle states with joint time series modeling. Applied Energy **261**, 114471 (2020)
42. Zhang, Y., Jin, W., Xiong, Z., Li, Z., Liu, Y., Peng, X.: Demystifying Interactions Between Driving Behaviors and Styles Through Self-clustering Algorithms. In: Krömker, H. (ed.) International Conference on Human-Computer Interaction (2021). https://doi.org/10.1007/978-3-030-78358-7_23, https://doi.org/10.1007/978-3-030-78358-7_23
43. Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.H.: Conditional random fields as recurrent neural networks. In: Proceedings of the IEEE international conference on computer vision. pp. 1529–1537 (2015)