A Spontaneous Driver Emotion Facial Expression (DEFE) Dataset for Intelligent Vehicles: Emotions Triggered by Video-Audio Clips in Driving Scenarios

Wenbo Li, Yaodong Cui, Yintao Ma, Xingxin Chen, Guofa Li, Member, IEEE, Guanzhong Zeng, Gang Guo, and Dongpu Cao, Member, IEEE

Abstract—In this article, a new dataset, the driver emotion facial expression (DEFE) dataset for drivers’ spontaneous emotions analysis is introduced. The dataset includes facial expression recordings from 60 participants during driving. After watching a selected video-audio clip to elicit a specific emotion, each participant completed the driving tasks in the same driving scenario and rated his/her emotional responses during the driving processes from the aspects of dimensional emotion method and discrete emotion method. The study also conducted classification experiments to recognize the scales of arousal, valence, dominance, as well as the emotion category and intensity to establish baseline results for the proposed dataset. Furthermore, this paper compared emotion recognition results difference through facial expressions between dynamic driving and static life scenarios. The results showed that dynamic driving and static life datasets were different in emotion recognition results. To further explore the reasons for the difference in emotion recognition results, the analysis from the AU (action unit) presence perspective was studied. The results showed significant differences in the AUs presence of facial expressions between dynamic driving and static life scenarios, indicating that drivers’ facial expressions may be affected by the driving task to influence the recognition of drivers’ emotions through facial expressions. Therefore, to accurately recognize the drivers’ emotions to establish a reliable emotion-aware human-machine interaction system, thereby improving driving safety and comfort, publishing a human emotion dataset specifically for the driver is necessary. The proposed dataset will be publicly available so that researchers worldwide can use it to develop and examine their driver emotion analysis methods. To the best of our knowledge, this is currently the only public driver facial expression dataset.

Index Terms—Driving safety, driver emotion, facial expression dataset, spontaneous expression, affective computing, intelligent vehicles

1 BACKGROUND AND RELATED WORK

Driver emotion plays a vital role in driving because it affects driving safety and comfort. Among the 20-50 million non-fatal injuries and 1.24 million fatal road traffic accidents occurring every year worldwide [1], drivers’ inability to control their emotions has been regarded as one of the critical factors degrading driving safety [2], [3]. The rapid development in intelligent vehicles also calls for an emerging demand in the integration of driver-automation interaction and collaboration to enhance driving comfort [4], where driver emotion is one of the critical states.

Driver emotion detection and intervention are emerging topics for the automotive human-machine interaction (HMI) system [5]. Previous studies have shown that drivers believe that emotion detection and intervention systems will help understand and improve their emotional state to improve driving safety and comfort [5], [6]. The novel HMI system in intelligent vehicles provides new opportunities for solving drivers’ emotional disorders [7]. The emotion-aware HMI of the intelligent vehicle can recognize and regulate drivers’ emotions in various ways to improve driving safety and comfort, where precise driver emotion recognition is the prerequisite for developing an emotion-aware HMI system. Therefore, accurately recognizing driver emotions is essential to enhance the driving safety and comfort of intelligent vehicles.

To describe human emotions, discrete emotion theory and dimensional emotion theory were proposed by psychological researchers for classifying emotions [27]. Due to the discrete language words used by humans to describe emotions,
discrete models are well-established and widely-accepted, such as the basic emotions of Ekman et al. [28] and the emotion tree structure of Parrott [29]. Specifically, Ekman et al. categorized discrete emotion models into six basic emotions (happiness, sadness, anger, fear, surprise, and disgust) [28], which are supported by cross-cultural researches showing that humans perceived these basic emotions in a similar form regardless of culture differences [30]. The dimensional emotion models propose that the emotional state can be accurately expressed as a combination of several psychological dimensions, such as the 2D “circumplex model” proposed by Russell [31] and the 3D dimensional model of Mehrabian et al. [32]. In the widely adopted model proposed by Russell [31], the valence dimension measures whether humans feel negative or positive, and the arousal dimension measures whether humans are bored or excited. Mehrabian et al. [32] extended the emotional model from 2D to 3D by adding a dominance dimension, which measures submissive or empowered feelings.

The discrete emotion method is intuitive and widely used in people’s daily lives. However, it fails to cover the whole range of emotions exhibited by humans. The dimensional emotion method is less intuitive and often requires training the participants to use the dimensional emotion method labelling system. Nevertheless, the dimensional emotion method is a more pragmatic and context-dependent approach to describe emotions [27]. In this study, considering the primary emotions of drivers during driving, we combine both the discrete emotion method and dimensional emotion method to describe drivers’ negative emotions (e.g., anger) and positive emotions (e.g., happiness) quantitatively by employing the well-known differential emotion scale (DES) [33] and Self-assessment manikin (SAM) [34].

Driver emotion recognition is often conducted by analyzing driver emotion expressions. The expressions of human emotions consists of facial expressions, speech, body posture and physiological changes. So far, different behavioural measurements (e.g., facial expression analysis, speech analysis, driving behaviour) [35], [36], physiological signal measurements (e.g., skin electrical activity, respiration) [37], [38], or self-reported scales (e.g., SAM) [39] have been applied in driver emotion recognition. Comparatively, physiological measurements are more objective and can be measured continuously. However, this measurement is highly invasive and may affect drivers’ driving performance. Self-reported measurements measure the subjective experience of the drivers when applied correctly, but such measurements cannot take place during the study without interruption. For the study on the driver emotion in the driving environment, it is crucial to use non-invasive and non-contact measurement methods. High intrusiveness has a significant impact on both the driver emotion expression and actual emotional experience, therefore should be avoided [40]. To this end, this study employed facial expression to recognize driver emotions and ensure the continuity of data collection.

Facial expression is a powerful channel for drivers to express emotions [41]. Recent advances in facial expression-based emotion recognition have motivated the creation of multiple facial expression datasets. Publicly available datasets are fundamental for accelerating facial expression research. As shown in Table 1, we summarized the up-to-date representative public available datasets containing facial expressions. These datasets have been used for emotion recognition and to achieve different levels of success. As shown in Table 1, one of the common aspects of these datasets is the collection of participants’ facial expression data in static life scenarios and wild settings. Although facial expression data collected in static life scenarios and wild settings can be employed to recognize emotions using various algorithms, it restricts the application of these algorithms into static life scenarios.

However, driving a car is a complex cognitive process [42], which requires the driver to dynamically respond to driving tasks, such as visual cues, hazard assessment, decision-making, strategic planning [43], [44]. Consequently, driving occupies a lot of the driver’s cognitive resources [45], and cognitive processing is needed to elicit emotional responses [46]. Therefore, driving may affect drivers’ emotional expressions. Due to the influence of driving tasks, drivers’ facial expressions may be suppressed or more subtle when they experience emotional states. As a result, if the above-mentioned algorithms are applied to dynamic driving scenarios, reliable recognition results may not be obtained. Thus, it is necessary to collect drivers’ facial expression data specifically for drivers’ emotional recognition in dynamic driving scenarios, analyze the difference in emotion recognition results between dynamic driving scenarios and static life scenarios, and explore the reasons for this difference.

To address the above-mentioned limitation, this study introduces a driver emotion facial expression dataset (DEFE) for driver emotion studies in intelligent vehicles. Table 2 presents the details of the experimental design for stimulus material selection, data collection, dataset content, and emotional labels. The performance of different emotion recognition algorithms on DEFE dataset was analyzed in this study. Also, this paper analyzed the difference in emotion recognition results between dynamic driving scenarios and static life scenarios, and explored the reasons for this difference in Action Units (AUs) presence perspective.

The main contributions of this paper can be described as:

- The study provide a new, publicly available dataset DEFE for spontaneous driver emotions analysis. The dataset contains frontal facial videos from 60 drivers, including their biographic information (gender, age, driving age), and subjective ratings on driver emotions (arousal, valence, dominance scales, as well as emotion category and intensity). To the best of our knowledge, this dataset is currently the only public available dataset of driver facial expressions.
- The study compared the classification results of driver emotions on our DEFE dataset using the mainstream classification algorithms. The DEFE dataset supports driver emotion classification from two aspects, dimensional emotion method (arousal, valence and dominance) and discrete emotion method (emotional category and intensity). The comparisons established the baseline results of the introduced dataset with classification accuracy and F1 score.
The study compared emotion recognition results difference through facial expressions between dynamic driving and static life scenarios. The results showed that dynamic driving and static life datasets were different in emotion recognition results. To further explore the reasons for the difference in emotion...
recognition results, this paper conducted the study from the AU presence perspective. The results showed significant differences in the AUs presence of facial expressions between dynamic driving and static life scenarios, indicating that drivers' facial expressions may be affected by the driving task to influence the recognition of drivers' emotions through facial expressions.

The structure of this paper is as follows: Section 2 presents the selection of stimulus materials to induce drivers' emotions. Section 3 introduces the DEFE data collection in detail. Section 4 reported the classification performance of the DEFE dataset and compared the DEFE dataset and the CK+ dataset from emotion recognition to verify whether the driving tasks will affect emotion recognition results. After discovering that there is a difference between the driving scenario and life scenarios in the recognition results, Section 5 explored and discussed the reasons for this difference. The conclusion and future work are shown in Section 6.

2.1 Initial Video-Audio Clips Selection

All initial video-audio clips were selected from the Bilibili website (https://www.bilibili.com/). Bilibili is a Chinese video-sharing website that allows users to upload life videos and tag or add comments to videos through a scrolling commenting system nicknamed "bullet-screen comments." To select the most effective video-audio clips, two research assistants (1 male and 1 female) reviewed more than 500 video-audio clips and conducted the preliminary screening. They were asked to select video-audio clips that lasted 30-120 seconds and contained content to elicit a single target emotion, including a negative emotion (anger), a positive emotion (happy), and a neutral state. Another two research experts (1 male and 1 female) with rich experience in drivers' emotions analysis evaluated each selected video-audio clip. A consensus of the two experts decided the selections of the video-audio clips.

The selected video-audio clips are mainly based on Chinese real-life scenarios and events, such as aggressive driving and chatting. Other video-audio clips selection criteria include: 1) the clip should have a bright and clear background, 2) the clip should contain complete speech segments, and 3) there is only one wanted expressing emotion in the clip. Accordingly, the study selected 18 video-audio clips and checked them further in subjective annotation session, and all video-audio clips are 800 x 480 pixels.

2.2 Subjective Annotation

2.2.1 Participants

Forty-two participants (33 males and 9 females) with ages from 21 to 32 years old (mean age = 24.95 years, SD = 2.62, 42 identified themselves as Chinese) were recruited to participate in this study from Chongqing University. Each participant had a valid driving license with at least one year of driving experience (mean year = 3.55 years, SD = 2.48, range = 1-10 years). All participants signed the consent form to participate in the study and received 25 CNY as financial reimbursement.

2.2.2 Materials

SAM [34] and DES [33] were used to subjective annotation. SAM uses non-verbal graphical representations to assess the arousal, valence, and dominance level. Previous study in [34] has concluded the effectiveness of SAM. This study adopted a 9-point scale (1 = “not at all”, 9 = “extremely”) SAM [34] for evaluation. The DES was used to annotate the
DES is a multi-dimensional self-report scale for assessing an individual’s emotions [51], including ten fundamental emotions: interest, joy, surprise, sadness, anger, disgust, contempt, fear, shame, and guilt. In this study, a 9-point scale DES (1 = “not at all”, 9 = “extreme”) [33] was used to assess the intensity of each self-reported emotional dimension.

### 2.2.3 Procedure

The web-based subjective emotion annotation experiment was conducted to evaluate the video-audio clips. First, each participant was provided with a set of instructions to inform them of the SAM and DES scale definition. Next, the 18 video-audio clips were displayed in a random order, and there was a relatively long break time (3 minutes) between every two clips to avoid interference from the previous one. After watching each video-audio clip, participants finished two questionnaires based on their true feelings. Each participant watched as many videos as he/she wanted and was able to end the rating at any time if he/she felt uncomfortable. In the end, at least 35 assessments were collected for each video-audio clip.

### 2.3 Selection Results

To select the most effective three video-audio clips, this study considered both the SAM and DES results. For each video-audio clip in the SAM data analysis, this study first calculated the mean rating of 35 participants. Then, to maximize the strength of elicited emotions, the study selected video-audio clips that had the strongest participant ratings and, at the same time, a small variation. To this end, for each video-audio clip, the study calculated a normalized valence, arousal and dominance score by maximum-minimum normalization method and then conducted a cluster analysis using the K-means method to identify the clusters of emotions based on the SAM data.

As shown in Fig. 1, the clustering results showed that a total of three emotion categories were generated, which corresponded to the positive emotion (happiness), negative emotion (anger), and neutral, respectively. Fig. 1 also showed each video-audio clip’s rating scores, and each video-audio clip’s distribution results in the valence-arousal-dominance space. The video-audio clip whose rating was closest to the center of each cluster was selected and marked as the representative video-audio clip (highlighted in red circle) of the each cluster [15].

Moreover, to select the video-audio clips that can effectively induce the driver’s emotions, we defined the hit rate, intensity value, and success index according to the DES result’s emotion category and intensity. The hit rate represented the proportion of the 35 participants who chose the target emotion category. The intensity value is defined as the average score of target emotion intensity. The success index equals the sum of the normalized scores of the hit rate and the intensity value. The normalized hit rate and the intensity value are between 0-1 by using the maximum-minimum normalization method.

As shown in Table 3, the selected three video-audio clips all achieved a high hit rate (M = 0.91, range = 0.86-0.97). Besides, in the anger intensity 9-level scale, 35 participants’ target emotion intensity value were all above 7 levels (M = 7.92, range = 7.69-8.33). Table 3 also described the success index of the three stimuli. In addition, the analysis showed that the selection results of video-audio clips through SAM data and DES data were consistent. Eventually, the study selected the three most effective video-audio clips for the DEFE data collection experiment. The content of the stimulus materials was shown in Table 3.

### 3 DEFE DATA COLLECTION

#### 3.1 Ethics Statement

The experimental procedure and the video content shown to the participants were approved by Chongqing University Cancer Hospital Ethics Committee, China. Participants and data from participants were treated according to the Declaration of Helsinki. The participants were also informed that they had the right to quit the experiment at any time. The video recordings of the participants were included in the dataset only after they gave written consent for the use of

#### TABLE 3

| Target emotion | Content                                                                 | Duration (sec) | Hit rate | Intensity value | Success index |
|----------------|-------------------------------------------------------------------------|----------------|----------|-----------------|----------------|
| Anger          | Many people were used in cruel human experiments during the war         | 45             | 0.86     | 8.33            | 1.71           |
| Happiness      | Parents mentor their children to do homework                             | 62             | 0.91     | 7.69            | 1.47           |
| Neutral        | Man drives on city road with nothing happened                           | 48             | 0.97     | 7.74            | 1.64           |

Note: The hit rate represented the proportion of the 35 participants who chose the target category. The intensity value is defined as the mean score of the participants’ target emotion intensity. The success index is equal to the sum of the normalized scores of the hit rate and the intensity value.
their videos for research purpose. A few participants also agreed to use their face images in research articles.

3.2 Participants
Sixty Chinese participants (47 males and 13 females) with aging from 19 to 56 years old (mean \( M = 27.3 \) years, standard deviation \( SD = 7.7 \) years) were recruited to participate in this experiment from Shapingba District, Chongqing, China. Each participant had a valid driving license with at least one year of driving experience (average \( M = 5.5 \) years, standard deviation \( SD = 5.8 \), range = 1-30 years). All participants had normal or corrected to normal vision (36 participants wear glasses) and normal hearing ability. The presence of occlusion such as glasses is a significant research challenge of facial expression recognition; hence participants wearing glasses were included to evaluate the robustness of emotion recognition. All participants received 100 CNY as financial reimbursement for their participation.

3.3 Experiment Setup
The experiments were carried out in a fix-based driving simulator (Fig. 2b) with illumination-controlled (RDS2000, Real-time technology SimCreator, Ann Arbor, Michigan, USA). Fig. 2d shows the front view, which was presented using three projectors, and the rear view was displayed using three LCD screens (one for the rear view in the vehicle and two for the left and right rear views). Another two LCD screens were used to display the dashboard and central stack. The ambient noise and sound of the engine were presented through two speakers. The vibration of the vehicle was simulated through a woofer under drivers' seat. For the presentation of stimuli without changing the internal environment of the driving simulator, as shown in Fig. 2e, the experiment used a 20-inch central stack screen (1,280 × 1,024, 60 Hz) to display the video-audio stimulus materials. A stereo Bluetooth speaker (Xiaomi) was used to play the audio, and the audio volume was set to a relatively loud level. However, each participant was asked before the experiment whether the volume was comfortable and adjusted when necessary for clear hearing. During the experiment, as shown in Fig. 2a, the participants' faces were continuously imaged with a visual camera. The visual face camera was an Pro Webcam C920 (Logitech, Newark, CA.) with a resolution of 1,920 × 1,080 pixels, collecting data at a frame rate of 30 fps. Also, an iPad (Apple) was used for participant self-reported emotion. Fig. 2c shows the overall data collection experiment setup.

3.4 Experiment Protocol
Two driving scenarios on highways were realized in the simulator. The reason for setting these two scenarios is to minimize the impact of complex driving scenarios on driver performance. The first was a practice driving scenario to help participants familiarize themselves with the simulator before the experiment. As shown in Fig. 3a, the practice driving scenario was an 8 km straight section of a four-lane highway with two for each driving direction. The participants were asked to drive on the right lane with speed changes in the range of 80 km/h – 50 km/h – 100 km/h. The second scenario is an emotional driving scenario. As shown in Fig. 3b, the emotional driving scenario was a 3km straight section of the same highway with a posted speed limit of 80 km/h. The participants were asked to drive on the right lane with speed around 80 km/h.

To obtain drivers' emotional drivings data, the study designed an experimental protocol about 45 minutes driving. The protocol was composed of one practice driving, followed by three emotional drivings. emotional drivings included angry driving (AD), happy driving (HD) and neutral driving (ND). Fig. 4 presented details of the protocol. Before the experiment, each participant filled out a basic information questionnaire (gender, age, driving age). Next, they were provided with a set of instructions to inform them of the experimental protocol and the definition of different scales used for self-reported emotions. Then, participants were required to drive a 10-minute practice driving to help them get familiar with the operation and motion performance of the driving simulator. After a short break following practice driving, participants started the three emotional drivings. The corresponding emotion was induced by watching the selected video-audio clip at the beginning of each emotional driving, following by driving with emotion. At the end of each emotional driving session, the participant was required to report his/her...
self-evaluated emotion level using SAM and DES. There was a 3 minutes break between each two emotional drivings. During the entire experiment, if the participants felt any discomfort, they could withdraw from the experiment at any time.

3.5 Self-Reported Emotion

To identify the emotion experienced by participants, the study employed self-reported scales for subjective assessment of emotions. After each driving task, the participants were asked to assess their emotional experience while driving using SAM and DES. The SAM and DES scales were presented to participants by an iPad. In SAM, the valence scale ranged from unhappy to happy, the arousal scale ranged from calm to stimulation, and the dominance scale ranged from submissive (or “without control”) to dominant (or “under control, empowered”). In DES, there were ten emotion dimensions, and each dimension evaluated the intensity of emotions from “not at all” to “extremely”. Each dimension of the SAM scale and the DES scale is represented from one to nine by a Likert scale. If the self-assessments from participants were not consistent with the induced target emotions, the study would use participants’ self-reported data as the ground truth to label the facial video data.

4 DEFE DATA PROCESSING, EVALUATION, AND COMPARISON

In this section, the data processing, classification performance of the DEFE dataset was reported. Furthermore, to verify whether the driving tasks will affect emotion recognition results, the study compared the DEFE dataset and the CK+ dataset from emotion recognition results. In the following subsections, the study introduces data processing, classification protocol, as well as dataset evaluation and comparison, respectively.

4.1 Data Processing

First, the study labelled the facial expression data of 60 drivers according to their self-reported emotion and removed the emotional drivings data that was not successfully induced. Second, the study reported how to split data for drivers’ emotion recognition, including splitting effective video clips from the original data and extracting driver facial expression.

During data collection, each participant completed three emotional driving sessions with average recording data of 405s. Also, the study compiled the self-reported data for each participant. As shown in Fig. 5, the numbers of successfully induced emotional drivers were 52, 56, and 56 for the anger, happy, and neutral driving, respectively. Participants’ self-reported data were used as the ground truth to label driver facial expression data.

As per [52] and [53], the facial expression video sequences 15s after drivers started driving were clipped as the most effective data. Face detection and alignment in
driving environments are challenging due to various poses, illuminations and occlusions (glasses). MTCNN (Multi-task Cascaded Convolutional Networks) is a cascade structure based on deep learning, which is relatively accurate when detecting faces in multiple pose angles and in unconstrained scenes [54]. Hence, the study used MTCNN to track and extract driver face data from each video frame. After extracting driver face expression data, the study obtained a total of 17,310 image frames of driver faces with 64x64 pixel. Therefore, the created dataset contains facial expression videos and images from 60 drivers with the ground truth of dimensional emotion method (valence, arousal and dominance) and discrete emotion method (emotion categories and its intensity). A few examples of the dataset images are provided in Fig. 6, which shows that drivers’ facial expressions varied with the types of emotion, but the variation was weak in some cases during driving, for example, the difference between AD and ND was tiny. Most video clips were challenging to observe peak expressions, and the study also observed that the change of emotion with driving duration was weak, and this phenomenon is probably because the facial expression of emotion was affected by driving tasks.

4.2 Classification Protocol
The study introduced two different types of protocols for driver emotion recognition based on facial expression data. (1) To investigate driver emotion classification results based on the dimensional emotion model, the study proposed three different nine-classification problems: valence, arousal, and dominance. To this end, the SAM scores of participants were used as the ground truth. Each classification (valence, arousal, dominance) on these scales was divided into nine levels (1 = “not at all”, 9 = “extremely”). (2) To study driver emotion classification results based on the discrete emotion model, the study proposed a three-emotion classification protocol, namely anger, happiness, and neutral. Moreover, the study discussed the intensity recognition for anger and happy emotions, respectively. To this end, the DES scores were taken as the ground truth. Each emotion (anger and happiness) intensity was divided into 5 levels (5 = “no emotion”, 9 = “maximum intensity”). It should be noted that the above approach can lead to unbalanced classes for some participants and scales. In light of this, the study included F1 scores in order to report reliable results. The F1 score is a commonly used metric in classification tasks, which considers both precision (P) and recall (R) of the model. It quantifies the correct prediction of the positive samples. When categories are unbalanced, the F1 score will be attenuated [55]. the study additionally used accuracy as another metric. Accuracy quantifies how well the classification correctly identifies or excludes conditions, and it is robust to unbalanced data.

Both the traditional and the deep learning methods for emotion recognition tasks were included in this study. As the most effective traditional method in most classification tasks [13], SVM (Support Vector Machine) was selected to be implemented by the sklearnn toolbox with a linear kernel. As for the deep learning-based classification methods, Xception [56] was applied. The Xception network has been widely adopted in emotion recognition tasks, and many state-of-the-art emotion recognition networks are developed based on the Xception network[57], [58]. For the network, the loss function can be expressed as:

$$L(y, \hat{y}) = - \sum_{j=0}^{M} \sum_{i=0}^{N} (y_{ij} \cdot \log(\hat{y}_{ij})).$$

Equation (1) where $\hat{y}$ is the prediction and $y$ is the ground truth. The above deep learning method used the same training strategy. First, it employed Adam optimizer [59], which has a learning rate of $10^{-3}$ and a weight decay of $10^{-6}$ for training. Second, image augmentations, including random horizontal flips, random crop, and random rotation, were applied on-the-fly to increase the amount of training images effectively. SVM was applied with Intel R CoreTM i5-dual-core CPU. Xception was used with TITAN XP.
4.3 DEFE Dataset Evaluation and Comparison

4.3.1 Dataset Selection for Classification Comparison

Apart from the emotion recognition results for the proposed dataset, the study also selected the CK+ [13] dataset which were collected in static life scenarios as the comparison dataset. CK+ is currently widely used in emotion recognition studies due to it has a large amount of naturally induced facial expression data. The CK+ dataset consists of 123 participants. This dataset was posed and spontaneous by multiple participants whose facial expressions started from neutral to the peak. In the CK+ dataset, 327 sequences have discrete emotion labels including neutral, sadness, surprise, happiness, fear, anger, contempt and disgust. This paper selected the neutral, anger and happiness sequences in this dataset to compare the emotion classification results based on discrete emotion models.

4.3.2 Classification Results

Table 4 shows the average accuracies and F1 scores (average F1 scores for nine classes) for each rating scale (valence, arousal and dominance) when using protocol one on DEFE. The study compared the performances of SVM and Xception on the DEFE dataset. In general, the accuracies when using Xception method were at least 30 percent higher than the accuracy when using SVM. The highest classification accuracy for valence, arousal, and dominance achieved 86.00, 91.54, and 88.17 percent, respectively, when using Xception. In terms of the F1 scores, the highest scores for valence, arousal, and dominance were: 83.73, 91.76, and 79.55 percent respectively, when using Xception.

Similarly, Table 5 shows the average accuracies and F1 scores for the emotion categories (anger, happiness, and neutral) when using protocol two. The study also compared the classification results when using SVM and Xception in Table 5. The results show that both the highest classification accuracy (90.34 percent) and the highest F1 scores (90.21 percent) were obtained when using Xception. Apart from the emotion recognition results on DEFE, Table 5 also presented the comparison results on the CK+ dataset when using the same recognition algorithms. The results show that the recognition results of the CK+ dataset were higher than that of DEFE dataset.

Moreover, Table 5 shows the average accuracies and F1 scores of the intensity classification results on anger and happiness emotions when using protocol two with different algorithms. Five classes of the intensity of anger and happiness were classified based on facial expression data. The results show that the highest classification accuracies for angry and happy driving intensity were 97.60 and 97.88 percent, respectively. The highest F1 scores for angry and happy intensity were 97.88 and 97.59 percent, respectively. It should be noted that in recognition of emotion intensity, the study did not compare the results with other datasets, because there was currently no spontaneous facial expression datasets with emotional intensity labels.

The comparison results in this section showed that dynamic driving and static life datasets were different in emotion recognition results. Due to the influence of driving tasks in the driving scenarios, facial expressions of drivers may be suppressed or more subtle when they experience emotional states. Hence, it is necessary to further explore the reasons for these differences in emotion recognition results.

5 THE AUS PRESENCE DIFFERENCE BETWEEN DYNAMIC DRIVING AND STATIC LIFE(scenarios)

After discovering that there is a difference between the dynamic driving and static life scenarios in the emotion recognition results, the study explore the reasons for this difference from AU presence perspective in this section. In the following subsections, the study first introduces dataset selection for comparison, differential analysis protocol, and results and discussion.

5.1 Dataset Selection for AUs Presence Comparison

This study conducted a differential analysis of AUs presence between dynamic driving and static life scenarios by comparing the DEFE and JAFFE datasets. The static life dataset, Japanese Female Facial Expression (JAFFE) dataset [8], was selected as a baseline. Given the East-Asian cultural background with the small difference (the facial expression configuration of people with different cultural backgrounds is different), the JAFFE dataset was the most optimal control group for our DEFE dataset because of the excluded most cultural bias[60], [61]. Ten East-Asian females posed JAFFE dataset with seven emotional expressions (happy, anger, disgust, fear, sad, and neutral). Each female had two to four examples for each emotion. In total, there are 213 grayscale facial expression images in this dataset. Since DEFE only includes two emotions (anger and happiness), the study also selected JAFFE’s anger and happiness expressions for analysis. Meanwhile, gender differences may affect the results, so that the study removed the male drivers from the initial DEFE dataset.
5.2 AUs Presence Differential Analysis Protocol
Each participant’s facial expressions were evaluated by observing subtle changes in facial features. The Facial Action Coding System (FACS) [62] is a systematic approach to describe what a face looks like when facial muscle movements have occurred. There are 44 coded facial muscle movements, namely AUs, in FACS according to the presence and intensity of facial movements. Ekman et al. further proposed that facial emotion expressions could be coded as a combination of several AUs. Figs. 7a and 7b display the common FACS [63] codes for anger and happiness, respectively, and (c) presents the AUs description for anger and happiness [63].

![Facial Action Coding System (FACS) codes](image)

**Table 7.** Facial Action Coding System (FACS) codes can be used to describe the facial configuration in adults. (a) and (b) display the common FACS codes for anger and happiness, respectively, and (c) presents the AUs description for anger and happiness [63].

5.3 Results and Discussion
5.3.1 Statistical Analysis Results
Statistical analysis results of AUs presence are shown in Table 6. For happiness, the results show that AU6 and AU12 movements could be observed in both JAFFE and DEFE. However, compared with JAFFE, the presence frequencies of AU 6 and AU 12 in DEFE were significantly lower (p < 0.01). For anger, the results show that AU4, AU5 and AU23 movements could be observed in both JAFFE and DEFE, and there are significant differences (p < 0.01). Moreover, the study found that AU7 related to anger from DEFE did not appear in the anger expressions from JAFFE.

![Sample images of facial expressions in JAFFE and DEFE](image)

**Table 6.** Statistics Analysis Results of AUs’ Presence in Anger and Happiness Cross DEFE and JAFFE Dataset

|          | AU 4 | AU 5 | AU 6 | AU 7 | AU 23 |
|----------|------|------|------|------|-------|
| Anger    | JAFFE Average: 0.333, Std: 0.045 | DEFE Average: 0.264, Std: 0.374 |
|          | t-test: 5.589***, df: 1,364*** | t-test: 7.375***, df: 1,364*** |
| Happiness| JAFFE Average: 0.361, Std: 0.177 | DEFE Average: 0.384, Std: 0.184 |
|          | t-test: 2.950***, df: 1,364*** | t-test: 4.576***, df: 1,364*** |

Note: p < 0.01 :***, 0.01 < p < 0.05 :**

coded as 1 and anger as 0. If the relationship coefficients of AUs had differences in the two datasets, it could be concluded that some AUs performed differently between dynamic driving and static life scenarios. It should be noted that positive coefficient means the AU is related to happiness and negative coefficient means the AU is related to anger.
driving task, which requires concentration during driving, and the concentration may decrease the presence of AUs near eyes. On the other hand, the presence frequencies of AU7 and AU23 were lower in JAFFE, which maybe because of the difficulties to express negative emotions in Japanese culture [65].

5.3.2 Logit Regression Results

The logit regression results are shown in Table 7. According to our regression results, in JAFFE, for happiness, the coefficients of AU6 and AU12 were consistent with the results from FACS [63], which means AU6 and AU12 were related to happiness. However, only the results of AU12 are significant ($p < 0.01$). For anger, the coefficients of AU4, AU5, and AU23 were consistent with the results from FACS [63], which means AU4, AU5, and AU23 were related to anger. The results of AU4, AU5, and AU23 are significant ($p < 0.01$). Interestingly, AU7 (lid tightener) presence shows that AU7 was related to happiness, which was different from previous researches[63]. In DEFE, only the result of AU4 was significant($0.01 < p < 0.05$), and the coefficient was consistent with the research in FACS, indicating that AU4 had a significant predictive ability for anger. Other AUs were not observed with significant results. Overall, for AUs presence, AU4 (Brow Lowerer), AU5 (Upper Lid Raiser), AU6 (Cheek Raiser), AU7 (Lid Tightener), AU12 (Lip Corner Puller), AU23 (Lip Tightener) are significant differences between dynamic driving and static life scenarios. The presence of AU4, AU5, AU6 and AU12 are higher in static life scenarios, indicating that AU4, AU5, AU6 and AU12 in dynamic driving scenarios may be affected by the main driving tasks, which suppresses the facial expression of drivers’ emotions. Meanwhile, the presence of AU7 and AU23 is higher in dynamic driving scenarios, which may be because Japanese culture suppresses the expression of negative emotions [65]. As for logit regression results, there are also significant differences between dynamic driving and static life scenarios. For anger, the results in dynamic driving scenarios show that only AU4 is significantly related to anger, while in static life scenarios AU4, AU5, and AU23 are all significantly related to anger. For happiness, the logit regression results in dynamic driving scenarios show that there is no significant correlation between AUs and happiness, but the results in static life scenarios show that AU12 is significantly related to happiness. These significant differences were most likely due to the main driving tasks, which reduced the frequency and amplitude of facial muscle movements. Due to the limitation of JAFFE data amount, these results may require further investigations.

6 Conclusion and Future Work

In this work, a dataset for the analysis of spontaneous driver emotions elicited by video-audio stimuli is presented. The dataset includes facial expression recordings from 60 participants during driving. After watching each of the three video-audio clips selected to elicit specific emotions, each participant completed the driving tasks in the same driving scenarios and rated his/her emotional responses in this driving process from the aspects of dimensional emotion method and discrete emotion method. These self-reported emotions include the scales of arousal, valence, and dominance as well as emotion category and intensity. The study selected these three video-audio clips using the SAM and DES scales, which ensured the effectiveness of these stimulus materials aimed at the Chinese cultural background. Furthermore, the study conducted the classification experiment for the scales of arousal, valence, and dominance as well as emotion category and intensity to establish baseline results for the proposed dataset in terms of accuracy and F1 scores, and these results were significantly higher than the results for random classification.

Moreover, the study compared emotion recognition results difference through facial expressions between dynamic driving and static life scenarios. The results showed that dynamic driving and static life datasets were different in emotion recognition results. To further explore the reasons for the difference in emotion recognition results, this paper conducted the study from the AU presence perspective. The results showed significant differences in the AUs presence of facial expressions between dynamic driving and static life scenarios, indicating that drivers’ facial expressions may be affected by the driving task to influence the recognition of drivers’ emotions through facial expressions. Therefore, to accurately recognize the drivers’ emotions to establish a reliable emotion-aware human-machine interaction system, thereby improving driving safety and comfort, publishing a human emotion dataset specifically for the driver is necessary.

As with any study, the present research has limitations. One limitation is that although the above analysis has found differences in recognition results and AUs...
presence, the present study cannot completely rule out the possibility that this difference is affected by factors other than the driving task (e.g., the differences in the classification accuracies were due to that the video-audio clips cannot completely trigger participants’ intended emotions in driving scenarios). Therefore, future research should further explore the reasons for these differences. Also, due to the difficulty of recruiting female drivers, in this experiment, we recruited a 3:1 male to female ratio, which is basically the same as the male to female ratio of Chinese drivers [66]. However, public datasets should notice the balance of gender of participants. Therefore, future research should maintain a balanced gender ratio as much as possible.

Furthermore, in this study, the driver’s emotions were induced by predefined video-audio clips. Although existing studies have shown that video-audio clips can induce driver emotions reliably [47], [48], [49], [50], and the driver’s emotions while driving are often triggered by events in life scenarios (e.g., the driver suddenly recalls a sad thing or receives an angry call while driving) [47], [50], the driver’s emotion can best be triggered from driving tasks, because the emotions in real driving might be more intensive and extensive. In the future, we will carry out the study to induce drivers’ emotions through specific driving scenarios. A large number of survey samples will be used to build a driving scenario library that induces drivers’ emotions. Also, the library will be used and shared in future drivers’ emotion research.

The DEFE dataset will be made publicly available after the work is published to allow researchers to evaluate their algorithms on an off-the-shelf driver facial expression dataset and investigate the possibility of applying them to applications. The DEFE data set provides the possibility to study emotion recognition from different emotion models simultaneously. Meantime, DEFE data can also be used to analyze the difference between dynamic driving and static life. Also, there are facial occlusions in DEFE, such as glasses and hands, which increases the complexity of facial expression recognition which is a significant research challenge. The source codes of this paper proposed approaches can be found at https://github.com/ (The link will be available in our final accepted paper).

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China under Grant 2019YFB1706102. The authors would like to thank Zehua Pan, Peizhi Wang, Qianjing Hu, Mingqing Tang, Bingbing Zhang, HuaiFei Xiao, and Jiyong Xue for their assistance.

This work was supported by the National Key Research and Development Program of China under Grant 2019YFB1706102. The authors would like to thank Zehua Pan, Peizhi Wang, Qianjing Hu, Mingqing Tang, Bingbing Zhang, HuaiFei Xiao, and Jiyong Xue for their assistance.

REFERENCES

[1] W. H. Organization, Global Status Report on Road Safety. Geneva, Switzerland: World Health Organization, 2015.
[2] L. James, Road Rage and Aggressive Driving: Steering Clear of Highway Warfare. Buffalo, NY, USA: Prometheus Books, 2009.
[3] G. Li et al., “Influence of traffic congestion on driver behavior in post-congestion driving,” Accident Anal. Prevent., vol. 141, 2020, Art. no. 105508.
[4] G. Li et al., “Risk assessment based collision avoidance decision-making for autonomous vehicles in multi-scenarios,” Transp. Res. Part C-Emerg. Technol., vol. 122, 2020, Art. no. 102920.
[5] M. Braun, R. Chadowitz, and F. Alt, “User experience of driver state visualizations: A look at demographics and personalities,” in Proc. IFIP Conf. Hum. Comput. Interaction, 2019, pp. 158–176.
[6] S. Li, T. Zhang, N. Liu, W. Zhang, D. Tao, and Z. Wang, “Drivers’ attitudes, preference, and acceptance of in-vehicle anger intervention systems and their relationships to demographic and personality characteristics,” Int. J. Ind. Ergonom., vol. 75, 2020, Art. no. 102899.
[7] G. Li, Y. Yang, and X. Qu, “Deep learning approaches on pedestrian detection in hazy weather,” IEEE Trans. Ind. Electron., vol. 67, no. 10, pp. 8889–8899, Oct. 2020.
[8] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, and J. Budynek, “The Japanese female facial expression (JAFFE) database,” in Proc. IEEE Int. Conf. Autom. Face Gesture Recognit., 1998, pp. 14–16.
[9] D. Lundqvist, A. Flykt, and A. Öhman, “The karolinska directed emotional faces (KDEF),” CD ROM Dept. Clinical Neurosci. Psychol. Sect. Karolinska Institutet, vol. 91, no. 630, pp. 2–2, 1998.
[10] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, “Web-based database for facial expression analysis,” in Proc. IEEE Int. Conf. Multimedia Expo., 2005, Art. no. 5.
[11] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, “A 3D facial expression database for facial behavior research,” in Proc. IEEE Int. Conf. Autom. Face Gesture Recognit., 2006, pp. 211–216.
[12] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, “Mapi-ple,” Image Vis. Comput., vol. 28, no. 3, pp. 807–813, 2010.
[13] P. Lucey, J. F. Cohn, T. Kanade, G. McKeown, and J. Hanratty, “The bel-ief,” Image Vis. Comput., vol. 28, no. 1, pp. 191–199, 2010.
[14] O. Langner, R. Dotsch, G. Bijlsma, D. H. Wigboldus, S. T. Hawk, and A. Van Knippenberg, “Presentation and validation of the rad-boud faces database,” Cogn. Emot., vol. 24, no. 8, pp. 1377–1388, 2010.
[15] S. Koelsch et al., “DEAP: A database for emotion analysis; using physiological signals,” IEEE Trans. Affect. Comput., vol. 3, no. 1, pp. 18–31, First Quarter 2012.
[16] I. Sneddon, M. McKorrie, G. McKeown, and J. Hanratty, “The bel-fast induced natural emotion database,” IEEE Trans. Affect. Comput., vol. 3, no. 1, pp. 32–41, First Quarter 2012.
[17] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and I. Matthews, “DISFA: A spontaneous facial action intensity data-base,” IEEE Trans. Affect. Comput., vol. 4, no. 2, pp. 151–160, Second Quarter 2013.
[18] F. Ringeval, A. Sonderregger, J. Sauer, and D. Lalanne, “Introducing the recola multimodal corpus of remote collaborative and affective interactions,” in Proc. IEEE Int. Conf. Autom. Face Gesture Recognit., 2013, pp. 1–8.
[19] S. Du, Y. Tao, and A. M. Martinez, “Compound facial expressions of emotion,” Proc. Nat. Acad. Sci. USA, vol. 111, no. 15, pp. E1454–E1462, 2014.
[20] R. Zhang et al., “BP4D-spontaneous: A high-resolution spontaneous 3D dynamic facial expression database,” Image Vis. Comput., vol. 32, no. 10, pp. 692–706, 2014.
[21] S. Happy, P. Patnaik, A. Routray, and R. Guha, “The indian spontaneous expression database for emotion recognition,” IEEE Trans. Affect. Comput., vol. 8, no. 1, pp. 131–142, Jan. 2017.
[22] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, “Training deep networks for facial expression recognition with crowd-sourced label distribution,” in Proc. ACM Int. Conf. Multimodal Interaction, 2016, pp. 279–283.
[23] C. F. Benitez-Quiroz, R. Srinivasan, and A. M. Martinez, “Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 5562–5570.
[24] S. Zafeiriou, C. Benitez-Quiroz, C. G. Zisserman, and J. S. Bartlett, “Aff-wild: Valence and arousal’in-the-wild database,” in Proc. ACM Int. Conf. Multimodal Interaction, 2018, pp. 147–156.
[25] S. Zafeiriou, D. Laptev, and I. Laptev, “The karolinska directed emotional faces (KDEF),” CD ROM Dept. Clinical Neurosci. Psychol. Sect. Karolinska Institutet, vol. 91, no. 630, pp. 2–2, 1998.
[26] A. Molahosseini, B. Hasan, and M. H. Mahoor, “Affectnet: A database for facial expression, valence, and arousal computing in the wild,” IEEE Trans. Affect. Comput., vol. 10, no. 1, pp. 15–31, First Quarter 2019.


579

Li et al.: Spontaneous Driver Emotion Facial Expression (DEFE) Dataset for Intelligent Vehicles: Emotions...

[27] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, “A survey of affect recognition methods: Audio, visual, and spontaneous expressions,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 1, pp. 39–58, Jan. 2009.

[28] P. Ekman et al., “Universals and cultural differences in the judgments of facial expressions of emotion,” J. Pers. Soc. Psychol., vol. 53, no. 4, 1987, Art. no. 712.

[29] W. G. Parrott, Emotions in Social Psychology: Essential Readings. London, U.K.: Psychology Press, 2001.

[30] P. Ekman and W. V. Friesen, “Constants across cultures in the face and emotion.,” J. Pers. Soc. Psychol., vol. 17, no. 2, 1971, Art. no. 124.

[31] J. A. Russell, “A circumplex model of affect,” J. Pers. Soc. Psychol., vol. 64, no. 6, 1980, Art. no. 1161.

[32] A. Mehrabian, “Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament,” Current Psychol., vol. 14, no. 4, pp. 261–292, 1996.

[33] J. J. Gross and R. W. Levenson, “Emotion elicitation using films,” Cogn. Emot., vol. 9, no. 1, pp. 87–108, 1995.

[34] M. M. Bradley and P. J. Lang, “Measuring emotion: The self-assessment manikin and the semantic differential,” J. Behav. Ther. Exp. Psychiatry, vol. 25, no. 1, pp. 49–59, 1994.

[35] X. Wang, Y. Liu, F. Wang, J. Wang, L. Liu, and J. Wang, “Feature extraction and dynamic identification of drivers’ emotions,” Transp. Res. Part. F-Traffic Psychol. Behav., vol. 62, pp. 175–191, 2019.

[36] G. Li, S. E. Li, B. Cheng, and P. Green, “Estimation of driving style in naturalistic highway traffic using maneuver transition probabilities,” Transp. Res. Part. C-Emerg. Technol., vol. 74, pp. 113–125, 2017.

[37] P. Wan, C. Wu, Y. Lin, and X. Ma, “On-road experimental study on driving anger identification model based on physiological features by roc curve analysis,” IET Intel. Transp. Syst., vol. 11, no. 5, pp. 290–298, 2017.

[38] D. T. Cordaro, R. W. Levenson, H. J. Park, Y. N. Kim, and J. H. Hwang, “Wearable mobile-based emotional response-monitoring system for drivers,” IEEE T. Hum-Mach. Syst., vol. 47, no. 5, pp. 636–649, Oct. 2017.

[39] L. Malta, C. Miyajima, N. Kitaoka, and K. Takeda, “Analysis of real-world driver’s frustration,” IEEE Trans. Intell. Transp. Syst., vol. 12, no. 1, pp. 109–118, Mar. 2011.

[40] C. Bussos et al., “Analysis of emotion recognition using facial expressions, speech and multimodal information,” in Proc. Int. Conf. Multimodal Interfaces, 2004, pp. 205–212.

[41] L. Yang, I. O. Ertugrul, J. F. Cohn, Z. Hammal, D. Jiang, and L. Yang, “Facial expression recognition methods: Audio, visual, and spontaneous expressions,” Proc. IEEE Conf. Multimodal Interfaces, 2019, pp. 538–544.

[42] J. A. Groeger, Understanding Driving: Applying Cognitive Psychology to Complex Everyday Tasks. London, U.K.: Psychol. Press, 2000.

[43] G. Li et al., “Drivers’ visual scanning behavior at signalized and unsignalized intersections: A naturalistic driving study in china,” J. Safety Res., vol. 71, pp. 219–229, 2019.

[44] G. Li, Y. Yang, X. Qu, D. Cao, and K. Li, “A deep learning based image enhancement approach for autonomous driving at night,” Knowl.-Based Syst., vol. 213, 2020, Art. no. 106617.

[45] T. Lajunen, D. Parker, and H. Summala, “The manchester driver behaviour questionnaire: A cross-cultural study,” Accident Anal. Prevent., vol. 36, no. 2, pp. 231–238, 2004.

[46] T. Brosch, K. R. Scherer, D. M. Grandjean, and D. Sander, “The impact of emotion on perception, attention, memory, and decision-making,” Swiss Med. Wkly., vol. 143, 2013, Art. no. w13786.

[47] X. Wang, Y. Liu, F. Wang, J. Wang, L. Liu, and J. Wang, “Feature extraction and dynamic identification of drivers’ emotions,” Transp. Res. Part. F-Traffic Psychol. Behav., vol. 62, pp. 175–191, 2019.

[48] H. Hu, X. Zhou, Z. Zhu, Q. Wang, and H. Xiao, “A driving simulator study of young driver’s behavior under angry emotion,” SAE International, Warrendale, PA, SAE Tech. Paper 2019-01-0398, 2019.

[49] W.-L. Li et al., “Visual-attribute-based emotion regulation of angry drivers’ behaviors,” IEEE Trans. Intel. Syst. Mag., to be published, doi: 10.1109/TIIM.2021.3050890.

[50] E. Roidl, F. W. Siebert, M. Oehl, and R. Höger, “Introducing a multivariate model for predicting driving performance: The role of driving anger and personal characteristics,” J. Safety Res., vol. 47, pp. 47–56, 2013.

[51] C. E. Izard, Patterns of Emotions: A New Analysis of Anxiety and Depression. Cambridge, MA, USA: Academic Press, 2013.

[52] D. O. Bos et al., “Eeg-based emotion recognition,” Influence Vis. Auditory Stimuli, vol. 56, no. 3, pp. 1–17, 2006.

[53] R. W. Levenson, L. L. Carstensen, W. V. Friesen, and P. Ekman, “Emotion, physiology, and expression in old age,” Psychol. Aging, vol. 6, no. 1, 1991, Art. no. 712.

[54] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint face detection and alignment using multitask cascaded convolutional networks,” IEEE Signal Process. Lett., vol. 23, no. 10, pp. 1499–1503, Oct. 2016.

[55] L. A. Jeni, J. F. Cohn, and F. De La Torre, “Facing imbalanced data-recommendations for the use of performance metrics,” in Proc. Hhumaine Assoc. Conf. Affect. Comput. Intell. Interaction, 2013, pp. 2245–2251.

[56] O. Arriaga, M. Valdenegro-Toro, and P. Plöger, “Real-time convolutional neural networks for emotion and gender classification,” 2017, arXiv:1710.07557.

[57] C. Pramerderofer and M. Kamilp, “Facial expression recognition using convolutional neural networks: State of the art,” 2016, arXiv:1612.02903.

[58] S. Li and W. Deng, “Deep facial expression recognition: A survey,” IEEE Trans. Affect. Comput., to be published, doi: 10.1109/TAFFC.2020.2981446.

[59] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014. [Online]. Available: https://arxiv.org/abs/1412.6980.

[60] R. E. Jack, O. G. B. Garrod, H. Yu, R. Caldara, and P. G. Schyns, “Facial expressions of emotion are not culturally universal,” Proc. Nat. Acad. Sci. USA, vol. 109, no. 19, pp. 7241–7244, 2012.

[61] D. T. Cordaro, R. Sun, D. Keltner, S. Kamble, N. Hudder, and G. McNeil, “Universals and cultural variations in 22 emotional expressions across five cultures,” Emotion, vol. 18, no. 1, 2018, Art. no. 75.

[62] P. Ekman, W. V. Friesen, and J. C. Hager, Facial Action Coding System: The Manual. Salt Lake City, UT, USA: Research Nexus, 2002.

[63] L. F. Barnett, R. Adolphs, S. Marisella, A. M. Martinez, and S. D. Pollak, “Emotional expressions reconsidered: challenges to inferring emotion from human facial movements,” Psychol. Sci. Public Interest, vol. 20, no. 1, pp. 1–68, 2019.

[64] T. Baltrusaitis, P. Robinson, and L.-P. Morency, “Openface: An open source facial behavior analysis toolkit,” in Proc. IEEE Winter Conf. Appl. Comput. Vis., 2016, pp. 1–10.

[65] D. Mlhaloum and P. Ekman, “American-Japanese cultural differences in intensity ratings of facial expressions of emotion,” Motivat. Emot., vol. 13, no. 2, pp. 143–157, 1989.

[66] C. P. G. of the People’s Republic of China, “The national car ownership exceeded 200 million for the first time,” 2018. [Online]. Available: http://www.gov.cn/xinwen/2019-01/13/content_5357441.htm.
Yintao Ma received the BSc degree in engineering mechanics from the University of Illinois in Urbana Champaign, in 2018. She is currently working toward the MSc degree with the Cognitive Autonomous Driving Laboratory, Department of Mechanical and Mechatronics Engineering, University of Waterloo, ON, Canada. Her research interests include machine learning, image processing, and facial expression recognition.

Xingxin Chen received the BSc degree from Nanjing University, Nanjing, China, in 2018. He is currently working toward the master of applied science (MASc) degree with Waterloo Cognitive Autonomous Driving (CogDrive) Laboratory, Department of Mechanical and Mechatronics Engineering, University of Waterloo, Canada. His research interests include domain adaptation, transfer learning, computer vision.

Guofa Li (Member, IEEE) received the PhD degree in mechanical engineering from Tsinghua University, Beijing, China, in 2016. He is currently an assistant professor in mechanical engineering and automation with the College of Mechatronics and Control Engineering, Shenzhen University, Guangdong, China. His research interests include driving safety in autonomous vehicles, driver behavior and decision making, computer vision, machine learning, and human factors in automotive and transportation engineering. He is the recipient of the Young Elite Scientists Sponsorship Program by SAE-China (2018), the Excellent Young Engineer Innovation Award from SAE-China (2017), and the NSK Sino-Japan Outstanding Paper Prize from NSK Ltd. (2014).

Guanzhong Zeng received the BSc degree in mechatronics engineering from Chongqing University in Chongqing, China, in 2018. He is currently working toward the MSc degree with the Advanced Manufacturing and Information Technology Laboratory, Department of Automotive Engineering, Chongqing University, Chongqing, China. His research interests include facial expression recognition, fine-grained image recognition and gaze estimation.

Gang Guo received the BS, MS, and PhD degrees in automotive engineering from Chongqing University, Chongqing, China, in 1982, 1984, and 1994, respectively. He is currently the chair and professor with the Department of Automotive Engineering, Chongqing University. He also serves as the associate director for the Chongqing Automotive Collaborative Innovation Center. He has authored and coauthored more than 100 refereed journal and conference publications. His research interests include intelligent vehicle, multi-sense perception, human-machine interaction, brain computer interface, intelligent manufacturing, and user experience. He is a senior member of the China Mechanical Engineering Society and the Director of the China Automotive Engineering Society. He is also a member of the China User Experience Alliance Committee.

Dongpu Cao (Member, IEEE) received the PhD degree from Concordia University, Canada, in 2008. He is the Canada Research Chair in Driver Cognition and Automated Driving, and currently an associate professor and director of Waterloo Cognitive Autonomous Driving (CogDrive) Lab, University of Waterloo, Canada. His current research focuses on driver cognition, automated driving and cognitive autonomous driving. He has contributed more than 200 papers and three books. He received the SAE Arch T. Colwell Merit Award in 2012, and three best paper awards from the ASME and IEEE conferences. He serves as an associate editor for the IEEE Transactions on Vehicular Technology, IEEE Transactions on Intelligent Transportation Systems, IEEE/ASME Transactions on Mechatronics, the IEEE Transactions on Industrial Electronics, IEEE/CAA Journal of Automatica Sinica and ASME Journal of Dynamic Systems, Measurement and Control. He was a guest editor for Vehicle System Dynamics and the IEEE Transactions on SMC: Systems. He serves on the SAE Vehicle Dynamics Standards Committee and acts as the co-chair of IEEE ITSS Technical Committee on Cooperative Driving.

For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdli.