

# Building BROOK: A multi-modal and facial video database for Human-Vehicle Interaction research

Xiangjun Peng, Zhentao Huang, Xu Sun



**University of  
Nottingham**  
UK | CHINA | MALAYSIA

University of Nottingham Ningbo China, 199 Taikang East Road, Ningbo, 315100, Zhejiang, China.

First published 2020

This work is made available under the terms of the Creative Commons Attribution 4.0 International License:

<http://creativecommons.org/licenses/by/4.0>

The work is licenced to the University of Nottingham Ningbo China under the Global University Publication Licence:

<https://www.nottingham.edu.cn/en/library/documents/research-support/global-university-publications-licence.pdf>



**University of  
Nottingham**

UK | CHINA | MALAYSIA

---

# Building BROOK: A Multi-modal and Facial Video Database for Human-Vehicle Interaction Research

## **Xiangjun Peng**

University of Nottingham  
Ningbo, Zhejiang, China  
zy22056@nottingham.edu.cn

## **Zhentaoh Huang**

University of Nottingham  
Ningbo, Zhejiang, China  
scyzh3@nottingham.edu.cn

## **Xu Sun**

University of Nottingham  
Ningbo, Zhejiang, China  
xu.sun@nottingham.edu.cn

---

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Copyright held by the owner/author(s).  
*CHI'20*, April 25–30, 2020, Honolulu, HI, USA  
ACM 978-1-4503-6819-3/20/04.  
<https://doi.org/10.1145/3334480.XXXXXXX>

## **Abstract**

With the growing popularity of Autonomous Vehicles, more opportunities have bloomed in the context of Human-Vehicle Interactions. However, the lack of comprehensive and concrete database support for such specific use case limits relevant studies in the whole design spaces.

In this paper, we present our work-in-progress BROOK, a public multi-modal database with facial video records, which could be used to characterise drivers' affective states and driving styles. We first explain how we over-engineer such database in details, and what we have gained through a ten-month study. Then we showcase a Neural Network-based predictor, leveraging BROOK, which supports multi-modal prediction (including physiological data of heart rate and skin conductance and driving status data of speed) through facial videos. Finally we discuss related issues when building such a database and our future directions in the context of BROOK.

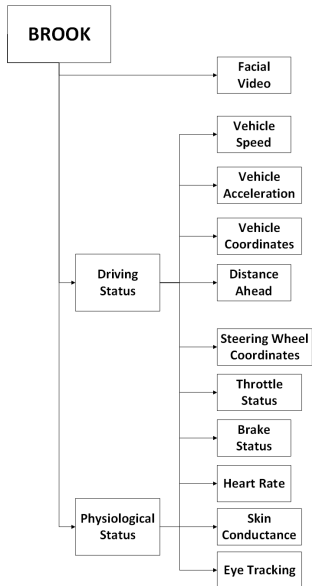
We believe BROOK is an essential building block for future Human-Vehicle Interaction Research. More details and updates about the project BROOK is online at <https://unnc-idl-ucc.github.io/BROOK/>.

## **Author Keywords**

Human-Vehicle Interactions; Computer Vision; Ergonomics.

## Data Release

The current version of BROOK would be released upon approached by re-search groups with ethical approval, please contact at scyzh3@nottingham.edu.cn.



**Figure 1:** The Composition of the Current BROOK Database.

## Introduction

While autonomous vehicles are coming into our daily life and the work on Human-Vehicle Interaction (HVI) steadily progresses, **the question of missing specific data to support design the interaction between driver and the automated vehicle remains largely unsolved**. This would heavily limit future research prototypes and validations. Such limitation would even be amplified, when emerging Machine Learning techniques to improve HVI.

To this end, BROOK<sup>1</sup> was established, as the first attempt, to address and tackle such issue. The goal of BROOK is to over-engineer a multi-modal and facial video database, which aims to support designs and research in HVI. The key idea, behind BROOK, is to **map as many kinds of data flows as possible in the same timeline**, when involved users are under the driving scenarios. Such guiding philosophy allows BROOK to support different research purposes, by selecting different types of data as the sources.

In this paper, we report our past ten-month study as the first step of BROOK, which we collected real-world users' data from 34 drivers in 11 dimensions. Later, we developed 3 Neural Network-driven predictors for three specific categories, to showcase the power of BROOK database. Finally, we discuss several reflections from this study and identify further directions of this work.

This paper made the following three major contributions:

- We present our work-in-progress BROOK, a multi-modal and facial video database, supported by customized driving scenarios. The current version of BROOK involves 34 drivers and their statistics (as structured in Figure 1).

<sup>1</sup>The name "BROOK" refers to multiple water flows, which stands for the key idea of our database

- We showcase one potential use case of BROOK. Leveraging BROOK, we develop three Neural Network-driven predictors for heart rate, skin conductance and driving speed.
- We quantitatively address the key limitation while developing BROOK, and discuss promising research directions of Human-Vehicle Interaction, with support from BROOK.

## Background and Motivation

**Human-Vehicle Interaction** (HVI) is a growing sub-area of Human-Computer Interactions. Especially, with Autonomous Vehicles coming into daily life, the demand for intelligent and personalized HVI systems becomes intense. Since in-vehicles drivers start to become available from performing necessary driving behaviors. For example, classifications of Driving Styles, in long term, have started to be studied for intelligent and personalized HVI [27, 31, 28, 18, 30].

**Data Support for HVI Research** appears to be a missing part, especially when emerging techniques from other disciplines. As an essential part, domain-specific data support have bred a large number of advances in Data Science and Computer Vision, e.g. ImageNet [3], CIFAR [17, 16], Kinetics [14] and DEAP [15]. All those databases have supported the validations of different designs/prototypes and served as the source information for system developments.

**Our Vision** is that, with the growing importance of Human-Vehicle Interaction research, a publicly available and over-engineered database is essential to support designs of interactive HVI. In response, we are building BROOK, a multi-modal and facial video database, which involves real-world drivers during customized driving scenarios. BROOK consists of both manual and autonomous driving behaviors, with 6 sensors and facial video records, from 34 drivers for around 20 minutes.

Data Types	Size
Facial Video	273 GB
Heart Rates	13.1 MB
Skin Conductance	48.3 MB
Eye-tracking	27.2 GB
Driving Status	60.9 MB

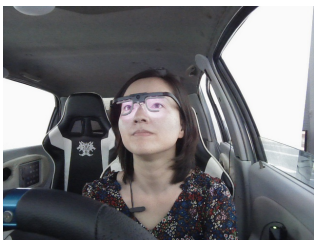
**Table 1:** A Summary of the Current BROOK Database.

### System Configuration

We host our driving simulator on a machine with two Intel Xeon E3-1225 processors and 32GB of RAM, with a NIVIDA RTX2070 GPU to ensure fluently visible scenarios.

	# of Lanes	# of Cars
1st	4	32
2nd	4	60
3rd	8	60
4th	8	92

**Table 2:** Configurations of Used Driving Scenarios and Scenes.



**Figure 2:** An Example of Data Collection Procedure.

### Overview

In this section, we provide an overview of BROOK database. The current version of BROOK consists of 34 drivers, who has driven for around 20 minutes in automated and manual modes separately. BROOK have covered their facial videos, multi-modal data and driving status. Table 1 provides key characteristics of the current BROOK version.

Now BROOK consists of 11 dimensions of data, as shown in Figure 1, including: **Facial Video, Vehicle Speed, Vehicle Acceleration, Vehicle Coordinate, Distance of Vehicle Ahead, Steering Wheel Coordinates, Throttle Status, Brake Status, Heart Rate, Skin Conductance and Eye Tracking.** All data streams are consolidated with each other in the same time flow.

### Hardware and Software Support

In this section, we provide details around hardware and software support, which has been utilized to build BROOK.

**Hardware Support.** We deploy an ultra-clear motion tracking camera (for facial video), a heart rate sensor (for heart rates), a leather electrical sensor (for skin conductance) and a Tobii eye tracker (for Eye-tracking) for participants.

Also, in order to collect detailed driving status information, we have integrated pressure sensors for Throttle and Brake, and we have ported Steering Wheel to obtain spatial changes from drivers during the whole period.

**Software Support.** We use OpenDS [4, 25], a open-source cognitive driving simulator, for driving scenario generations. Also, we utilise ErgoLAB [26] to coordinate sensor data output for storage. Finally, we build an in-framework tool to stream out speed and coordinates as driving status.

**Region-based Customization.** We customise our scenarios [9] and scenes [10] based on the common patterns of roads in China, as shown in Table 2. This is because all participants during this study have held issued license from it, where they accumulated their driving experiences.

### Procedure

In this section, we illustrate the whole procedure (about 8 months) while collecting BROOK database. We first prepare necessary resources and tests to ensure the bulk of study run smoothly. Then we performed the bulk of the study to collect necessary data streams, in various different settings and modes separately.

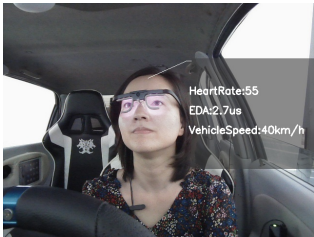
**Preparations.** Participants were first briefed about the study and were screened by a questionnaire to ensure that they were at low risk of motion sickness while experiencing the driving simulation. After the introduction, participants were given five minutes to sit in the simulator by themselves to adapt to the environment. The study only began when we ensured that participants understood the brief and were comfortable with the simulator.

**Data Collection.** The experiment included three driving conditions, as explained in previous sections. Each participant began with the baseline manual mode, with drivers in control of the vehicle operations, following which they undertook the two autopilot experimental drives (standard and personalised versions<sup>2</sup>) which were completed in a counterbalanced order. During the study, driving data was logged. After each driving session, the participants were asked to fill out a questionnaire about the cognitive aspects of their experience, including an assessment of perceived trust, comfort and situational awareness. The study lasted approximately one hour for each participant.

<sup>2</sup>Determined by participants' behaviors in manual mode.

Parameters	Value
Depth/Layers	100
Growth Rate	12
Dense Blocks	4
Compression Factor	0.5
Batch Size	128
Initial Learning Rate	0.1
Training Epochs	50

**Table 3:** Pivotal Parameters about DenseNet for BROOK in details.



**Figure 3:** A Demonstrated Example of Our Prototype.



**Figure 4:** An Overview of Test Accuracy between skin conductance (Orange), heart rates (Green) and Vehicle Speed (Yellow).

## Case Study

In this section, we present an in-vehicle real-time design to estimate driver's multi-modal states (skin conductance and heart rates) and driving status (speed) through their facial expressions only. Powered by BROOK, we present three representatives of our trained models, which are specialised for these three kinds of data flows separately.

**Data Pre-processing.** Before model training, we manually went through BROOK, recorded all "abnormal"<sup>3</sup> data and exclude noisy data, which are caused by the unconscious operations during the setup and abort stages of our study. In total, the ready-to-train BROOK contains 320,000 frames/images, and we split the training, test and validation sets in a ratio of 8:1:1.

**Methodology.** To map data with frames, we simply label every frame of facial video streams with relevant data streams, which are Heart Rates (per minutes), Skin Conductance (uS) and Vehicle Speed (km/h) respectively. For each kind of data flows, we apply DenseNet as the model architecture, since it saves many efforts of hyper-parameter adjustments through feature-map concatenation [8].

In addition, there are two points to highlight. For those "abnormal" data, we re-assigned their labels with the lower-bound values, which are 60 (for Heart Rates), 20.0 (for Skin Conductance) and 120 (for Vehicle Speed).

To support the whole pipeline, we utilise a face detector, extended from OpenCV [22] to crop only the face from camera captures, as the input for our trained models.

<sup>3</sup>"abnormal" refers to those out-of-the-common-range data, which shall be within 60-100 (per minutes) for Heart Rates, 0.0-20.0 (uS) for Skin Conductance and 0-120 (km/h) for Vehicle Speed. And all of "abnormal" issues are confirmed through our video records.

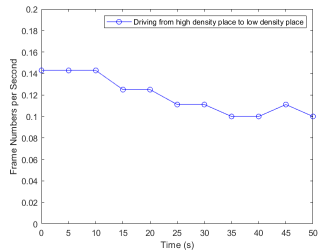
**Data Training.** We trained BROOK on a NVIDIA GeForce GTX 1660 Ti GPU, and details about DenseNet architecture are provided in Table 3. The majority of our settings are based on the recommended settings from DenseNet. Such design choice is because BROOK varies from all other commonly used data sets for existing neural network architectures, which are initially purposed for image classifications (e.g. CIFAR [17, 16]), and we hope to remain its effectiveness as much as possible.

However, we did some lightweight adjustments of the training epoch, batch size and some other hyper-parameters which would improve the accuracy. For our case, the initial learning rate is set to 0.1, and is divided by 10 at 50% and 75% of the total number of training epochs.

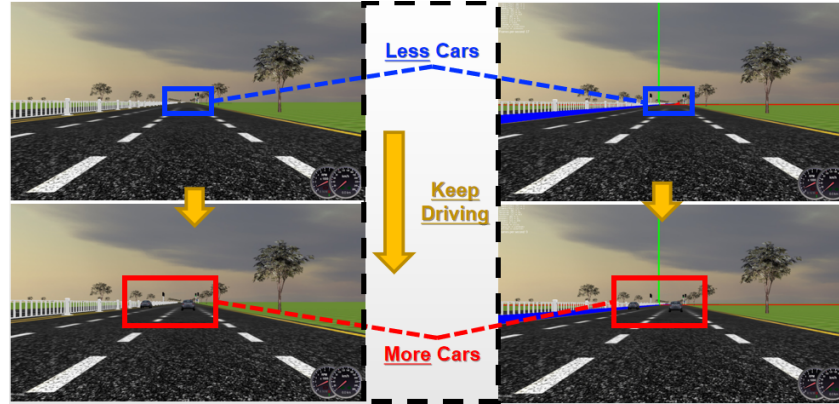
**Showcase.** As Figure 3 shown, our estimation prototype reflects all estimated results directly. The backbone of this prototype is three different models, which are for three different kinds of data flows.

**Accuracy.** We also evaluated our prototype through validation set, in terms of accuracy. The single-crop validation accuracy is reported in Figure 4 as a function of training epochs. The validation accuracy of skin conductance, heart rates and vehicle speed are 82.96%, 57.90% and 58.75% respectively. Our estimation on Skin Conductance is better than other data significantly, it might be caused by the smaller timeslot among data streams, which substantially provides more valuable information during training.

One thing to be highlighted, is that initial hyper-parameter settings are mainly optimized for image classifications. We are confident that more studies through extensive hyper-parameter adjustment could improve the accuracy of our prototypes in the end.



**Figure 6:** Visual Trend of Frequency Changes through the whole procedure, as shown in Figure 5.



**Figure 5:** An Illustration of Our Case Study to Address the Frequency Flip Issue. The right series are equipped with the detailed monitor to record the frequency.

Timespots	Frequency
5	0.142
15	0.124
15	0.124
25	0.113
25	0.113
35	<b>0.096</b>
35	<b>0.094</b>
45	0.119

**Table 4:** Detailed Breakdown of Frequency Changes at Different Time Spots, as the procedure shown in Figure 5.

### Key Limitation

In this section, we address the key limitation, based on our ten-month preliminary tryouts to build and utilize BROOK. We performed a case study to illustrate our concerns in details, which are presented as follow.

**Motivation.** During our studies among different Traffic Scenarios, we found that the frequency of frames, in our driving simulator, might drift while the density of the scenarios changes. Therefore, to quantitatively address this issue, we performed a case study to observe how severe this issue is and whether it would endanger the performance of our Vehicle Speed predictor further.

**Scenario Designs.** We implemented a straight road, with several vehicles coming near. Also, we set our vehicle (in the first-person perspective) at a constant speed to go forward. Figure 5 illustrates the whole procedure in details, and it automatically ends when two vehicle sets have passed one another.

**Timespot Breakdown.** We first visualize the relatively low-precision frequency during the whole process. As shown in Figure 6, the drift has been confirmed, and the frequency would become lower while there are more and more vehicles within the visual field.

To further quantify such issue, we deploy a high-precision frequency monitor for the overall procedure. As shown in Table 4, detailed drifts are quantified with different timespots. As highlighted, we draw the key observation that, when vehicles interleave within the visual range, the frequency would flip into another unit level.

**Why Important?** Such flip could cause significant penalty for driving-status predictor, since all data are streamed out from the simulator and labelled frame-by-frame. In our case, we might have mislabelled much Vehicle Speed data in the first place, which caused accuracy degradation, since we added lots of vehicles for more realistic scenarios.

## **Discussions**

We believe BROOK would be a building block for many future research, which aims to stretch the boundaries of theory and practice, in the context of Human-Vehicle Interaction (HVI). Hereby this moment, we summarize three aspects of potential research agenda, which are imposed with the support of BROOK.

### **Accurate In-vehicle Face-to-Multimodal Predictors.**

Tailed to the need for personalized HVI systems, multimodal data support could bring a lot of benefits to assist decision-making procedures for different drivers. However, adapting multiple sensors for drivers/passengers is not user-friendly and hard to be integrated. We believe BROOK provides a unique series of opportunities to develop and validate more accurate Face-to-Multimodal predictors (e.g. more detailed facial expression [21, 13]) inside the vehicles, which has a great potential to be the data source for future HVI systems in practice.

### **Long Short-Term Characterization of Drivers.**

Current studies have paid a lot of attention on the characterizations and classification of driving styles, but few attempts to adapt the instant reflections from drivers/passengers. We believe BROOK provides offers a great many chances for taking in-depth analysis and modeling of drivers/passengers in both Long and Short Term (e.g. Long Short-Term Memory (LSTM) [7]). Despite driving styles from long-term inference, future HVI systems could be capable of providing real-time reflections on drivers'/passengers' affective states, to further improve its quality of service.

### **Novel Adaptions of Neural Network Models into HVI.**

One of the key constraints, to adapt complicated decision-making procedures into Autonomous Vehicles, is the limitation of computationally power. Even for our demonstration, we faced tremendous performance challenges while train-

ing video-based data set and obtain preliminary results. Compression (e.g. [5, 6]), context-aware combination (e.g. [23, 1, 29]) and reconfiguration (e.g. [2]) would all be useful in the context of HVI. It's clear that, even outside HVI community, BROOK is still quite valuable to various other communities, since the usage and training of video data are still in debts (e.g. [19, 20]) and need more sources to validate novel proposals.

## **Conclusion and Future Work**

We presented BROOK, a multi-modal and facial video database for Human-Vehicle Interaction Research. A ten-month study has merged 34 drivers into the current version of BROOK database, with manual and auto-pilot modes separately. Based on BROOK, we showcased a novel usage of Face-to-Multimodal predictor and evaluate it in details. Also, we have addressed the key limitation of BROOK project, and we briefly discuss the future research agenda, enlighten by BROOK.

We hope BROOK could stimulate more intellectual contributions to the growing Human-Vehicle Interaction field. Since BROOK is a long-term project, our future work would consist of multiple aspects for expansion of it, including collecting valuable data through more realistic simulation support [24], improving simulator-related productivity [12, 11] and providing more variants of BROOK as needed.

More details and updates about the project BROOK is online at <https://unnc-idl-ucc.github.io/BROOK/>.

## **Acknowledgements**

We thanked for all members of User-Centric Computing Group and CHI'20 reviewers for their valuable and insightful feedback. This research is generously supported by LandRover and National Youth Fund No. 123456.



## REFERENCES

- [1] Wonmin Byeon, Qin Wang, Rupesh Kumar Srivastava, and Petros Koumoutsakos. 2018. ContextVP: Fully Context-Aware Video Prediction. In *2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2018, Salt Lake City, UT, USA, June 18-22, 2018*. 1122–1126. [http://openaccess.thecvf.com/content\\_cvpr\\_2018\\_workshops/w19/html/Byeon\\_ContextVP\\_Fully\\_Context-Aware\\_CVPR\\_2018\\_paper.html](http://openaccess.thecvf.com/content_cvpr_2018_workshops/w19/html/Byeon_ContextVP_Fully_Context-Aware_CVPR_2018_paper.html)
- [2] Han Cai, Chuang Gan, and Song Han. 2019. Once for All: Train One Network and Specialize it for Efficient Deployment. *CoRR* abs/1908.09791 (2019). <http://arxiv.org/abs/1908.09791>
- [3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Fei-Fei Li. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA*. 248–255. DOI: <http://dx.doi.org/10.1109/CVPRW.2009.5206848>
- [4] Paul A. Green, Heejin Jeong, and Te-Ping Kang. 2014. Using an OpenDS Driving Simulator for Car Following: A First Attempt. In *Adjunct Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Seattle, WA, USA, September 17 - 19, 2014*. 4:1–4:6. DOI: <http://dx.doi.org/10.1145/2667239.2667295>
- [5] Song Han, Huizi Mao, and William J. Dally. 2016. Deep Compression: Compressing Deep Neural Network with Pruning, Trained Quantization and Huffman Coding. In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*. <http://arxiv.org/abs/1510.00149>
- [6] Yihui He, Ji Lin, Zhijian Liu, Hanrui Wang, Li-Jia Li, and Song Han. 2018. AMC: AutoML for Model Compression and Acceleration on Mobile Devices. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VII*. 815–832. DOI: [http://dx.doi.org/10.1007/978-3-030-01234-2\\_48](http://dx.doi.org/10.1007/978-3-030-01234-2_48)
- [7] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (1997), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [8] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. 2017. Densely Connected Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*. 2261–2269. DOI: <http://dx.doi.org/10.1109/CVPR.2017.243>
- [9] User-Centric Computing Group in The Univeristy of Nottingham Ningbo China. 2019a. Customized Driving Scenarios for BROOK. <https://github.com/unnc-idl-ucc/Scenarios-OpenDS>. (2019).
- [10] User-Centric Computing Group in The Univeristy of Nottingham Ningbo China. 2019b. Customized Driving Scenes for BROOK. <https://github.com/unnc-idl-ucc/Scenes-OpenDS>. (2019).
- [11] User-Centric Computing Group in The Univeristy of Nottingham Ningbo China. 2019c. Ommniverse: Configurable Abstractions for Agile Developments of

- Driving Tasks in OpenDS. <https://github.com/unnc-idl-ucc/Omniverse-OpenDS>. (2019).
- [12] User-Centric Computing Group in The Univeristy of Nottingham Ningbo China. 2019d. Onerios: Interactive Graphics User Interface with Automated Layer of Code Generations for Collective Implementations of Driving Scenes. <https://github.com/unnc-idl-ucc/Oneiros-OpenDS>. (2019).
- [13] Natasha Jaques, Jesse H. Engel, David Ha, Fred Bertsch, Rosalind W. Picard, and Douglas Eck. 2018. Learning via social awareness: improving sketch representations with facial feedback. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Workshop Track Proceedings*. <https://openreview.net/forum?id=Bkyn3dJPG>
- [14] Will Kay, João Carreira, Karen Simonyan, Brian Zhang, Chloe Hillier, Sudheendra Vijayanarasimhan, Fabio Viola, Tim Green, Trevor Back, Paul Natsev, Mustafa Suleyman, and Andrew Zisserman. 2017. The Kinetics Human Action Video Dataset. *CoRR* abs/1705.06950 (2017). <http://arxiv.org/abs/1705.06950>
- [15] Sander Koelstra, Christian Mühl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2012. DEAP: A Database for Emotion Analysis ;Using Physiological Signals. *IEEE Trans. Affective Computing* 3, 1 (2012), 18–31. DOI : <http://dx.doi.org/10.1109/T-AFFC.2011.15>
- [16] Alex Krizhevsky. 2009. CIFAR: 80 million Images Datasets. <http://www.cs.toronto.edu/~kriz/cifar.html>. (2009).
- [17] Alex Krizhevsky. 2010. Convolutional Deep Belief Networks on CIFAR-10. In *Technical Report in University of Toronto*. 1–9. <http://www.cs.toronto.edu/~kriz/conv-cifar10-aug2010.pdf>
- [18] Xiaohan Li, Wenshuo Wang, and Matthias Roetting. 2019. Estimating Driver’s Lane-Change Intent Considering Driving Style and Contextual Traffic. *IEEE Trans. Intelligent Transportation Systems* 20, 9 (2019), 3258–3271. DOI : <http://dx.doi.org/10.1109/TITS.2018.2873595>
- [19] Ji Lin, Chuang Gan, and Song Han. 2018. Temporal Shift Module for Efficient Video Understanding. *CoRR* abs/1811.08383 (2018). <http://arxiv.org/abs/1811.08383>
- [20] Ji Lin, Chuang Gan, and Song Han. 2019. Training Kinetics in 15 Minutes: Large-scale Distributed Training on Videos. *CoRR* abs/1910.00932 (2019). <http://arxiv.org/abs/1910.00932>
- [21] Daniel Lopez Martinez, Ognjen Rudovic, and Rosalind W. Picard. 2017. Personalized Automatic Estimation of Self-Reported Pain Intensity from Facial Expressions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2017, Honolulu, HI, USA, July 21-26, 2017*. 2318–2327. DOI : <http://dx.doi.org/10.1109/CVPRW.2017.286>
- [22] Kari Pulli, Anatoly Baksheev, Kirill Korniyakov, and Victor Eruhimov. 2012. Real-time computer vision with OpenCV. *Commun. ACM* 55, 6 (2012), 61–69. DOI : <http://dx.doi.org/10.1145/2184319.2184337>

- [23] Jing Shi, Jia Xu, Boqing Gong, and Chenliang Xu. 2019. Not All Frames Are Equal: Weakly-Supervised Video Grounding With Contextual Similarity and Visual Clustering Losses. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*. 10444–10452. [http://openaccess.thecvf.com/content\\_CVPR\\_2019/html/Shi\\_Not\\_All\\_Frames\\_Are\\_Equal\\_Weakly-Supervised\\_Video\\_Grounding\\_With\\_Contextual\\_CVPR\\_2019\\_paper.html](http://openaccess.thecvf.com/content_CVPR_2019/html/Shi_Not_All_Frames_Are_Equal_Weakly-Supervised_Video_Grounding_With_Contextual_CVPR_2019_paper.html)
- [24] Zilin Song, Shuolei Wang, Weikai Kong, Xiangjun Peng, and Xu Sun. 2019. 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*. 387–391. DOI : <http://dx.doi.org/10.1145/3349263.3351497>
- [25] OpenDS Development Team. 2017. OpenDS - the Flexible Open Source Driving Simulation. <https://opens.dfki.de/>. (2017).
- [26] Beijing Jinfa Tech. 2019. ErgoLab: Human-Machine-Environment Synchronization Platform. [http://www.kingfar.cn/productshow\\_20.html](http://www.kingfar.cn/productshow_20.html). (2019).
- [27] Hanneke Hooft van Huysduynen, Jacques M. B. Terken, Jean-Bernard Martens, and Berry Eggen. 2015. Measuring driving styles: a validation of the multidimensional driving style inventory. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2015, Nottingham, United Kingdom, September 1-3, 2015*. 257–264. DOI : <http://dx.doi.org/10.1145/2799250.2799266>
- [28] Eric Vasey, Sangjin Ko, and Myounghoon Jeon. 2018. In-Vehicle Affect Detection System: Identification of Emotional Arousal by Monitoring the Driver and Driving Style. In *Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2018, Toronto, ON, Canada, September 23-25, 2018*. 243–247. DOI : <http://dx.doi.org/10.1145/3239092.3267417>
- [29] Jing Wang, Yu Cheng, and Rogério Schmidt Feris. 2016. Walk and Learn: Facial Attribute Representation Learning from Egocentric Video and Contextual Data. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*. 2295–2304. DOI : <http://dx.doi.org/10.1109/CVPR.2016.252>
- [30] Wenshuo Wang, Junqiang Xi, and Ding Zhao. 2019. Driving Style Analysis Using Primitive Driving Patterns With Bayesian Nonparametric Approaches. *IEEE Trans. Intelligent Transportation Systems* 20, 8 (2019), 2986–2998. DOI : <http://dx.doi.org/10.1109/TITS.2018.2870525>
- [31] Nidzamuddin Md. Yusof, Juffrizal Karjanto, Jacques M. B. Terken, Frank Delbressine, Muhammad Zahir Hassan, and Matthias Rauterberg. 2016. The Exploration of Autonomous Vehicle Driving Styles: Preferred Longitudinal, Lateral, and Vertical Accelerations. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2016, Ann Arbor, MI, USA, October 24-26, 2016*. 245–252. DOI : <http://dx.doi.org/10.1145/3003715.3005455>