Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

# Performance Evaluation of Intelligent Driving Emotion Recognition Model based on Synthetic Dataset in Real Scenes

Hancheng Li[1, *], Yurui Li[2], Yaowen Qian[3]

[1]Chongqing University of Posts and Telecommunications, Chongqing, China

[2]Guangxi Normal University for Nationalities, Chongzuo, China

[3] Shaanxi University of Science & Technology, Xian, China

*Corresponding Author: 2021214941@stu.cqupt.edu.cn

**Abstract.** The paper aims to explore the feasibility of using artificially generated facial expressions with different emotions to enhance the features and increase the data volume for emotion recognition in the context of intelligent cockpit. The paper first introduces the background and significance of the research, which is motivated by the increasing number of private cars in China, the development of intelligent cockpit technology, and the importance of emotion recognition for driving safety. The paper then reviews the existing literature on emotion recognition based on facial recognition, and points out the challenges and limitations of using real datasets, such as ImageNet, which may have high cost, low quality, privacy issues, and inaccurate annotations. The study suggests utilizing synthetic facial expressions that convey a range of emotions, created through advanced deep learning algorithms, as a solution to enhance the precision and reliability of the emotion detection system. The paper further examines the prospective applications and effects of the recommended technique pertaining to the realm of intelligent automotive cockpits and the associated vehicular journey. The paper concludes by summarizing the main contributions and limitations of the research, and suggesting some directions for future work.

**Keywords:** Intelligent Cockpit; Synthetic Datasets; Emotion Recognition.

## 1. Introduction

Over the past few years, there has been a persistent rise in the ownership of private vehicles in China. As of the end of 2020, China, as the world's largest automobile consumer, had a production and sales volume of over 25.2 million vehicles, maintaining its position as the global leader. The total number of domestic civilian vehicles has climbed to 280.87 million, compared with a net increase of 19.37 million in the previous year. In this statistic, the number of private cars is particularly significant, reaching 249.3 million, with an annual growth rate of 17.58 million. In short, the growth of private cars accounts for the vast majority of the growth of civilian vehicles, which shows the increasing importance of private cars in people's travel mode [1].

With the continuous improvement of science and technology and intelligence, the car has made great changes in driving mode, power source and driving experience. The cockpit has also changed from the traditional mechanical and electronic era to an intelligent cockpit [2]. This development has increased the depth of human-vehicle interaction. With the increasing number of privately-owned cars, driving has inevitably become an important part of daily life, and with it comes the issue of driving safety. Driving is a complex behavior that requires drivers to be highly focused and use a significant amount of cognitive resources to judge road conditions and operate in-vehicle systems. Even the slightest distraction during this process can lead to serious consequences [3]. Indeed, existing research has found that emotions play a crucial role in influencing driver cognition, and they may have potential impacts on their cognitive resources. Emotional states such as stress, anger, or fatigue can significantly affect a driver's attention, decision-making, and reaction time. These emotional factors can potentially increase the cognitive load on drivers, leading to decreased performance and compromised driving safety. Recognizing and managing emotions while driving is important for maintaining optimal cognitive functioning and promoting safe driving behaviors [4].

According to research, there are 1.24 million fatal traffic accidents worldwide each year [5], and one of the main causes of accidents that has been proven is the driver's emotional condition [6]. Consequently, by integrating an algorithmic model with a comprehensive dataset, the intelligent cockpit can achieve emotional recognition. This capability enables the detection of the driver's emotional condition, thereby enhancing overall driving safety [7].

However, emotional detection technology based on facial recognition relies heavily on the accuracy and robustness of the deep learning model. Deep learning models require a lot of dataset training. Dataset refers to a structured dataset that can represent a certain range and form a complete description after human beings collect and process it in various social activities [8]. Selecting an appropriate data set enhances the precision of the information, and procuring the data set effectively can expedite the research timeline. Make researchers more comprehensive research technology [9]. ImageNet database is a common real dataset with a rich variety of images, which reflects great potential in deep learning and provides a large amount of training data [10]. The ImageNet dataset comprises an extensive collection of over 14 million URLs pointing to images. Each image has a manually annotated category label. In order to detect the target, more than 1 million images in the dataset are marked with the object enclosure. The disadvantages of ImageNet are also obvious. Most image annotations in the ImageNet dataset are not annotated by specialized technicians, so some wrong data annotations will be generated [11, 12]. Fueled by the swift advancement of state-of-the-art technologies like big data and AI, there is a growing need for data sets within the realm of scientific inquiry.The cost, time consumption, privacy and labeling of real datasets are far worse than synthetic datasets [13]. Therefore, this paper will use the research method of synthetic dataset to analyze the emotions of drivers under the intelligent cockpit.

In deep learning based on facial expression recognition, utilizing diverse datasets for model training is a fundamental approach. However, challenges often arise due to the insufficient quantity of data caused by significant variations in facial appearances [14]. This inadequacy in data volume can further result in a lack of precision in model accuracy [14]. In current deep learning approaches, the use of artificially synthesized images is employed to augment training datasets. For instance, in the research conducted by Zhang et al., they proposed a deep end-to-end facial expression recognition model that leverages a face editing method based on Generative Adversarial Networks (GAN). This approach autonomously generates facial images, significantly mitigating the issue of overfitting in facial expression recognition tasks, which often leads to reduced model accuracy [14]. In addition, Niinuma and his team created new images with specific facial action units by aligning the three-dimensional grid with the standardized view and using a model based on the generation of adversal network (GAN). In their analysis, they discovered that when comparing models trained with original facial expression data, those utilizing synthetic expressions demonstrated enhanced results over models that were trained with actual facial expression data [15]. Simultaneously, various deep learning-based facial editing methods exist for synthesizing diverse datasets of facial expressions. For example, Han et al. utilized the StyleGAN facial editing method to synthesize the Chinese facial dataset SZU-EmoDage, demonstrating a feature richness far exceeding that of real facial datasets [16].

The rapid development of artificial synthesis dataset technology is notable. However, there remains a limited amount of research on the impact of models trained on synthetically generated facial expression datasets on driver emotion analysis within the context of intelligent driving. Therefore, in this article, we will use synthetic datasets to study their impact.

This paper primarily focuses on evaluating the precision of facial recognition algorithms that are trained using synthetic data within the realm of smart vehicle technology.

This article consists of six chapters: This “Introduction” section introduces the limitations of real facial datasets in facial recognition technology for intelligent driving, presents successful cases of using synthetic datasets, and outlines the research objectives. This “Methodology” section provides a detailed explanation of the experimental process, including dataset preprocessing, model training, and the utilization of the trained model for facial expression recognition on the fer 2013 dataset. This

“Results” section summarizes all the data obtained from the experiments. This “Discussion” section conducts a rational analysis of the experimental results, discussing limitations. And elucidates the significance of the experiments. This “Conclusion” section summarizes the research objectives and main findings, and proposes future prospects for the study.

## 2. Method

In this paper, FER-stable-diffusion-dataset was used as the basic dataset and supplemented with FER-stable-diffusion-dataset in intelligent driving scenarios. Using the pytorch framework [17], the ResNet-18-based neural network model is used for training, 80% of the dataset is used to train the model, and 20% of the dataset is used for model accuracy self-test, which is hereinafter referred to as "self-test accuracy". After the self-test accuracy reached 85%, the training was completed, and the model was used to test the FER 2013 dataset [17], and the data obtained is hereinafter referred to as "test accuracy".

### 2.1. Dataset Processing

FER-stable-diffusion-dataset comprises both training and validation sets, with digits ranging from 0 to 6 representing seven basic emotions that may occur during driving processes: angry, disgust, fear, happy, neutral, sad, and surprise. After filtering out low-quality data with excessively high fitting and unclear features, it remained a total of 12680 facial data samples. In addition, due to the research carried out in the context of intelligent driving, the FER-stable-diffusion-dataset is supplemented in this paper. 20% of the driver's facial emotion images were added to each emotion category. Using stable- diffusion with the RealVisXl V3.0 Turbo checkpoint merge model.

The size of images in original dataset from 128 x 128 are reduced to 64 x 64 pixels without destroying the sharpness of the face to speed up the model training. At the same time, the value of each pixel in the test dataset is copied to three channels and displayed in the form of color images, which is consistent with the FER-stable-diffusion-dataset to improve the accuracy and feasibility of model recognition. The face image in .PNG format is converted to a grayscale image, and then converted to a .CSV file.

### 2.2. Facial Expression Recognition

#### 2.2.1. Training Basic Model

Using an off-the-shelf facial expression recognition model (CNN), the model is trained on a processed image dataset to learn feature representations for different expressions. At the same time, transfer learning is used for fine-tuning to adapt to the needs of smart cockpit driving.

#### 2.2.2. Model Evaluation and Optimization

Use evaluation indicators, such as accuracy, precision, recall, etc., to evaluate the performance of the trained model on the test set. Carry out model optimization based on the evaluation results, including adjusting the model structure, adjusting hyperparameters, adding training data, etc.

#### 2.2.3. Model performance evaluation

The trained model was used to recognize the FER 2013 real facial emotion dataset, and the recognition accuracy was obtained, and the performance of the model in real scenes was evaluated on this basis.

## 3. Results

The validation loss serves as an indicator to assess the effectiveness of machine learning models. The statement describes the model's accuracy on the validation set, indicating the proportion of correctly classified instances relative to the total number of cases evaluated. The validation dataset, a segment of the larger dataset, serves as a tool to evaluate the model's capability to generalize beyond the training data. This evaluation is crucial for preventing the model from becoming too specialized (overfitting) or too general (underfitting) for the data it was trained on. In essence, this metric is an indicator of how well the model is likely to perform on new, unseen data. A reduced validation loss signifies the model's enhanced performance on the validation dataset, reflecting its improved capability to generalize to unseen data.

As shown in the Figure 1, this is a graph that reflects the changes in training losses and verification losses with the increase of the number of training cycles during the experiment. By observing the chart, it can be found that with the increase of the number of training cycles, both training losses and verification losses show a downward trend. However, the decline of verification losses is slower than that of training losses, and there are some fluctuations in the later stage. This may indicate that the model's fitting ability on the training set is stronger than its generalization ability on the validation set, and there may also be some overfitting phenomena.

After 10 rounds of Epoch, the average self-test accuracy of the model is 74.75%, and the test accuracy is about 32% using the trained model to identify the FER 2013 dataset.

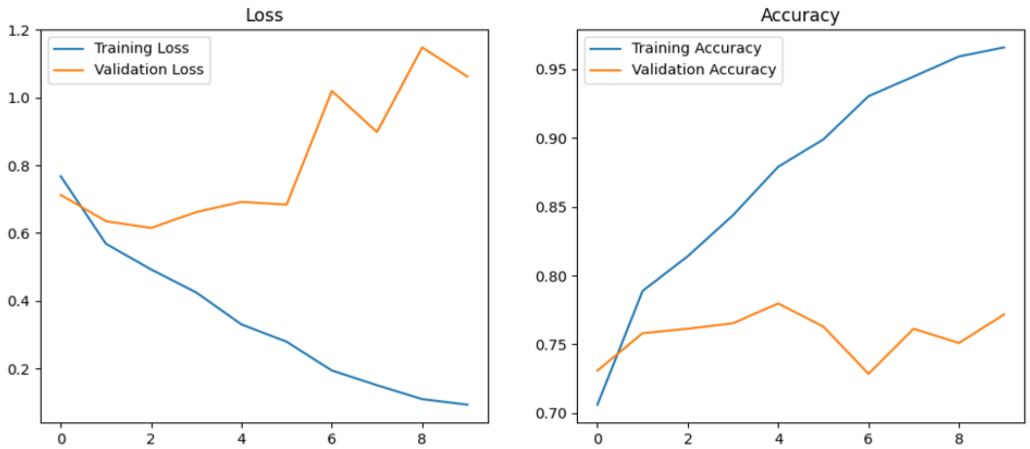


**Figure 1**. Results of training losses and verification losses (Photo credited: Original)

## 4. Discussion

In this study, an intelligent driving emotion recognition model based on synthetic datasets and real-world datasets was used for performance evaluation. Experimental results show that the model trained entirely on synthetic datasets does not perform well in real-world scenarios. After training, the model achieved a self-test accuracy of 74.75%, while the test accuracy on the real-world dataset was only 32%. This difference in performance may stem from the fact that synthetic datasets cannot adequately simulate complex situations in real-world scenarios, resulting in insufficient generalization capabilities of the model.

For the situation where the self-test accuracy is higher than the test accuracy during model training, we speculate that there may be a problem that the model is over-fitting during the training phase, resulting in poor performance in the real world. To enhance the model's effectiveness in practical applications, we recommend that more attention should be paid to the collection and utilization of real-world data in the model training to enhance the model's adaptability to complex situations. Add a certain amount of real-world images to the original dataset and test the most appropriate ratio.

Additionally, it is crucial to explore methods for boosting the model's generalization capabilities and its operational efficiency. Possible solutions include optimizing the model structure, introducing more real-world scene data for training, exploring other emotion recognition algorithms, and more. Through continuous research and improvement, we are expected to improve the effect of the intelligent driving emotion recognition model in practical applications, so as to better serve the development of intelligent driving systems.

## 5. Conclusion

Overall, this study highlights the challenges of deploying CNN models trained solely on synthetic datasets in real-world scenarios, offering valuable insights for future research. It emphasizes the need to address the complexities of transitioning from simulated to authentic environments. Moving forward, there is significant potential in exploring innovative approaches to seamlessly integrate synthetic and real-world datasets. This integration could enhance the model's robustness and improve its generalization abilities across diverse real-world settings. By bridging the gap between synthetic and authentic data, researchers can mitigate the limitations associated with each dataset type, ultimately fostering the development of more reliable and adaptable models. Such advancements are particularly relevant in fields like emotion recognition for intelligent driving systems, promising safer and more intuitive driving experiences. In essence, this collaborative effort not only drives progress in intelligent driving but also contributes to broader advancements in artificial intelligence, enabling practical applications across various domains. As researchers continue to explore these avenues, they are likely to pave the way for transformative breakthroughs in the integration of synthetic and real-world data, potentially propelling the field of machine learning into new frontiers.

## Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

## References

[1] Nbs. Statistical Communiqué of the People's Republic of China on National Economic and Social Development in 2020 [EB/OL] (2021).

[2] Li H, Wang X, Yu S. Development trend of human-computer interaction in intelligent cockpit[J].Times Automobile, 23 (2022):16-18.

[3] Li W, Cui Y, Ma Y, et al. A Spontaneous Driver Emotion Facial Expression (DEFE) Dataset for Intelligent Vehicles: Emotions Triggered by Video-Audio Clips in Driving Scenarios[J]. IEEE Transactions on Affective Computing, (2021) 14.

[4] Mesken J, Hagenzieker M P, Rothengatter T, et al. Frequency, determinants, and consequences of different drivers' emotions: An on-the-road study using self-reports, (observed) behaviour, and physiology[J]. Transportation research part F: traffic psychology and behaviour, 10(6) (2007): 458–475.

[5] Organization WH. Global status report on road safety 2015[M]. World Health Organization, (2015).

[6] James L. Road rage and aggressive driving: Steering clear of highway warfare[M]. Prometheus Books, (2000).

[7] Zepf, S.; Hernandez, J.; Schmitt, A.; Minker, W.; Picard, R. Driver emotion recognition for intelligent vehicles: A survey. ACM Comput. Surv. (CSUR) 53 (2020), 1–30.

[8] Li L, Yu H, Yi Z, et al. Quantitative Measure of the Value of Scientific Datasets from Perspective of Altmetrics[J].Information Studies: Theory & Application, 9 (2020):47-52.

[9] Zhang T, Wang C, Yu C, et al.Scientific Research Literature Dataset in the Era of AI: Characteristic Laws and Development Directions[J].Library and Information Knowledge, 40(05) (2023):39-49.

[10] Deng J, Dong W, Socher R, Li J, Li K & Li F, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, (2009) 248-255.

[11] J. Buolamwiniand, T. Gebru. Gendershades: Intersectional accuracy disparities in commercial gender classification[C]. In: S. A. Friedler and C. Wilson, (eds.). Proceedings of Proceedings of the 1st Conference on Fairness, Accountability and Transparency, volume 81 of Proceedings of Machine Learning Research, New York, NY, USA: (2018). 77--91.

[12] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He. Aggregated residual transformations for deep neural networks[C]. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2017).

[13] Wu H. Research on deep learning method of image processing based on synthetic dataset[D]. Lanzhou University, (2021).

[14] Zhang X, Zhang F & Xu C, "Joint Expression Synthesis and Representation Learning for Facial Expression Recognition," in IEEE Transactions on Circuits and Systems for Video Technology, 32(3) (2022 )1681-1695.

[15] Niinuma, K., Ertugrul, I. O., Cohn, J. F., & Jeni, L. A. Synthetic Expressions are Better Than Real for Learning to Detect Facial Actions. IEEE Winter Conference on Applications of Computer Vision. IEEE Winter Conference on Applications of Computer Vision, (2021), 1247–1256.

[16] Han S, Guo Y, Zhou X, Huang J, Shen L, Luo Y. A Chinese Face Dataset with Dynamic Expressions and Diverse Ages Synthesized by Deep Learning. Sci Data. 10(1) (2023):878.

[17] FER-stable-diffusion-dataset FER 2013.