Acadlore Transactions on AI and Machine Learning
https://www.acadlore.com/journals/ATAIML

Real-Time Prediction of Car Driver's Emotions Using Facial Expression with a Convolutional Neural Network-Based Intelligent System

Pawan Wawage*, Yogesh Deshpande

Department of Computer Engineering, Vishwakarma University, 411048 Pune, India

* Correspondence: Pawan Wawage (pawan.wawage-026@vupune.ac.in)

Received: 07-23-2022 **Revised:** 08-21-2022 **Accepted:** 09-20-2022

Citation: P. Wawage, Y. Deshpande, "Real-time prediction of car driver's emotions using facial expression with a convolutional neural network-based intelligent system," *Acadlore Trans. Mach. Learn.*, vol. 1, no. 1, pp. 22-29, 2022. https://doi.org/10.56578/ataiml010104.

Abstract: When driving, the most crucial factor to consider is your own safety. Driver's must be kept under observation for any potential harmful act, whether intentional or inadvertent, in order to ensure a safe navigation for a driver. As a result, a real-time emotion detection system for a driver has been developed to detect, exploit, and evaluate the driver's emotional state. This paper discusses how to recognize emotions using facial expressions for application in active security systems for drivers. We discuss our research and development of a Convolutional Neural Network-based intelligent system for face image-based expression classification in this paper.

Keywords: Convolution neural network; Deep neural network; Driver behavior; Facial expressions; Human-vehicle interaction

1. Introduction

Driving an automobile has become a difficult undertaking in recent years. It necessitates the driver doing many and simultaneous tasks, including steering, braking, and acceleration, as well as maintaining focus on the road. Drivers also engage in a variety of non-driving tasks, including hands-free calling, navigation, and altering the radio and entertainment system. Due to an increased mental or physical workload, the number of tasks completed while driving demonstrates that it can be tough for the driver.

As a result, most automobile companies are exploring and attempting to lessen the driver's burden and distraction while also making driving safe, fun, and fuel efficient. An in-car system that knows the driver's emotional state and can categorize the driver's conduct is one proposed solution based on this. All seven emotions are present in the driver's emotional state (Happy, Sad, Surprise, Disgust, Fear, Anger, and Neutral). We believe that in the future era of autonomous car driving, understanding the emotional condition of the drivers would be even more vital than it is now.

The issue of interest is recognizing the driver's emotional state and making recommendations for safe driving: Negative as well as positive emotional states have been identified as important factors influencing driving performance and can lead to risky driving behavior [1]. In comparison to other areas of emotion-related sciences like human-computer interaction (HCI), the car industry has previously ignored this topic. The majority of study focuses on detecting driver behavior using data other than the driver's emotions [2].

We conducted this study of recording emotional experiences while drivers completed driving events in live traffic from source to destination because there is a need to understand how drivers' emotions affect driving performance and to establish meaningful driver-vehicle interactions. This paper's contributions can be summarized as follows: We share insights on (1) factors affecting human emotions in the car and (2) the effect of human emotions on driving performance based on our online survey performed in Pune, India. (3) We condense our findings to show the impact of human emotions on driving behavior. We ran an online survey and collected the information about participants such as: Demographics (Age, Gender, Region), Duration as a driving license holder (in years), Estimated average mileage, Positive experiences while driving (joy, love, happiness...), Negative

experiences while driving (sad, anger, frustration, etc.), Effect of Positive/Negative emotions on driving behavior, Reason for the Positive/Negative emotions (traffic, road condition, others, …).

According to the World Health Organization, the global death rate has risen to 1.25 fatalities per year. The driver's emotion is one of the most important elements in a car accident. Accidents are frequently caused by driver distraction brought on by additional activities. The National Highway Traffic Safety Administration (NHTSA) defines distracted driving as any behavior that takes a driver's focus away from the task of driving. There are three types of distractions: manual, visual, and cognitive. We hope to solve the difficulty mentioned above by identifying seven moods in this research (Happy, Sad, Surprise, Disgust, Fear, Anger, and Neutral). We'll begin by taking driver input photographs for our VGG-16 model. The driver's current mood is then returned using this programme. We also make an effort to maintain high accuracy, which is important in real-world applications.

2. Literature Survey

For a long time, research on car user interfaces has centered on how to improve the user experience while minimizing distracting impacts on the driver [3]. Approaches to this essential necessity for road safety have evolved in tandem with technological advancements in automobiles [4]. Today's automobile human-machine interaction researchers strive for a natural experience via multi-modal input channels and persuasive skills. Such systems can improve traffic participant safety by observing driving behavior and, if necessary, influencing driving style or regulating speeding. In order to keep track of the driver's mental state, other systems detect and react to it [5].

The concept is based on research that shows feelings can have a significant impact on behavior. When people are in an active mood, they are more difficult to divert, and they are less likely to engage in mentally demanding tasks when they are content. Nass et al. looked into this in the car by providing a voice-based assistant that can modify its voice to the user's emotion. When the system's voice matched the driver's emotion, they reported fewer accidents, improved concentration, and a greater willingness to interact [6].

Data from emotion recognition systems can be coupled with location-based services to detect abnormalities or used to describe the quality of interpersonal engagement. The requirement to comprehend the context of emotions derives from the fact that neither the environment nor user behavior alone can provide a system with a clear picture of what is going on [7], despite the fact that emotions have a significant impact on driving safety. Our poll provides more information about the causes of emotional situations on the road, allowing us to better understand and respond to them [8].

Emotions play a crucial role in safe driving, according to psychological studies [9]. Because the driver's emotional state influences the car's safety and comfort, it's critical to design an in-car emotion identification system that can detect the driver's emotions and warn or raise an alarm if necessary. Negative emotions including grief, wrath, disgust, and fear, according to a psychology study, contribute to rash, inattentive, and rapid driving. Anger and aggression, according to are two emotions that have a strong influence on driving behavior and enhance the likelihood of accidents. Similarly, weariness and stress are two other factors that might lead to unsafe driving [10].

Anxiety, grief, and other intense emotions can all have an impact on driving. It is critical that the driver's emotional state is appropriate for driving. Recognize and make the driver aware of the driver's emotions is the final step in regulating the driver's emotions. Determining the user's mood can be utilized to deliver appropriate responses from the gadget in various situations. Facial expressions are one of the most researched techniques of detecting mood, and it is still one of the most complex subjects in pattern recognition and machine learning science. Deep Neural Networks (DNN) have been widely employed to solve challenges in face expression classification [11].

3. Proposed Approach

We conducted a series of driving tests under real-world settings in order to test the proposed strategy. To eliminate possible variations from actual driving behavior, we chose a genuine scenario rather than a simulation-based scenario. The studies were carried out with the help of eight drivers. They used the identical vehicle and drove in real-world situations.

We collect data in the first stage, such as a video of the driver driving along a specified route. The driver will be accompanied by a passenger in order to record and observe the driver and the divers' actions throughout the journey. The observer is the passenger, who will note both inside-the-vehicle (inner) factors like infotainment system use, AC activation and deactivation, etc., and outside-the-vehicle (outer) parameters like weather and road conditions. These behavioral features can also be used to look at how a driver reacts to things like speed bumps, potholes, and zebra crossings on a particular route. We could capture the driver's footage with the use of a mobile camera. While these qualitative parameters might not be used to develop models, they will undoubtedly aid in our understanding of the different elements that affect driver behavior and driving performance.

Then we start with the video that was recorded and utilize image segmentation to collect images from the frames. The images would be segmented using segmentation software. The segmented images from the recorded videos

serve as the model's input, therefore using segmentation tools will aid in the dataset preparation. The gathering of the dataset, as well as the subsequent module, might benefit from image processing. Finally, we will consider the datasets that we have gathered from the recorded videos, as well as picture segmentation from those videos. These datasets will aid in the model training. There would be facial cues and expressions that may assist us figure out how the driver is feeling. The 4 to 5 minutes driving videos may have numerous segmented images. Each video may contain a different number of photos depending on the driver. Only photos of the driver's face that are clearly visible will be included in the dataset. We won't be able to measure the driver's true emotions, but we will be able to identify them during that particular driving sequence.

3.1 Participant

Considering the majority of all age groups combined, between the victims who were 18 to 35 years old, there were approximately 35,000 road accident deaths in India in 2020. So, we decided to focus on the drivers of this age category. Eight different drivers, ranging in age from 18 to 35, were photographed while driving for about 5 kilometers. The drivers' video driving time ranged from 4 to 5 minutes each. The driver went through both congested and deserted roads in order to get a variety of expressions. The video was taken with the front camera of the Samsung Galaxy Note 9 at 720p resolution. The driver was immediately in front of the camera. With this we manage to prepare a dataset of 1,500-2,000 frames.

3.2 Extracting Frames from Videos

Python was used to extract about 1,500-2,000 frames from the video after it was taken. Each image was roughly 250 kb in size and had a resolution of 1,280x720. For each of the eight applicants' driving videos, the same method was followed as shown in Figure 1.

(a) Frame 1 (b) Frame 2

Figure 1. (a) Frame 1 and (b) Frame 2

3.3 Cropping Images

In some of the images received, there were two faces in a single image (one of a driver and other of co-passenger or the person sitting on the backseat). The image had to be trimmed for improved neural network training, thus the person other than the driver was removed and only the image of the driver's face was retrieved.

(a) Before cropping (b) After cropping

Figure 2. (a) Before cropping and (b) After cropping

3.4 Data Collection

In the initial stage, we gather data such as a video of the driver travelling a predetermined path. To record and monitor the driver's and the divers' behavior throughout the trip, the driver will travel with a passenger. The driver is noting both inside-the-vehicle (inner) aspects like entertainment system use, AC activation and deactivation, etc., and outside-the-vehicle (outer) parameters like weather and road conditions. The drivers' video driving time ranged from 4 to 5 minutes each. The video was taken with the front camera of the Samsung Galaxy Note 9 at 720p resolution. The driver was immediately in front of the camera. With this we manage to prepare a dataset of 1,500-2,000 frames.

4. Methodology

We assume that the camera is positioned in the vehicle in such a way that it remains stationary and always focuses on the driver's face. As a result, each shot will have at least one face (that of the driver) and some backdrop. We use the Viola and Jones face detection algorithm [8] to detect faces in each frame and crop the discovered face region as the Region of Interest (ROI) image. Figure 2 depicts a couple of the face detection results.

4.1 CNN Training and Testing

It is broken down into two sections: emotion recognition using GGDA and feature extraction using AlexNet [12] and VGG16. Real-valued parameters and binary parameters were both derived from the facial pictures of eight individuals. Each facial image yielded the extraction of a total of 15 parameters, including seven binary parameters and eight real-valued parameters. The parameters' real values were standardized. All fifteen input parameters were used to train generalized neural networks. The seven different facial expressions were represented by seven output nodes (neutral, angry, disgust, fear, happy, sad and surprised). The real valued parameters have a fixed value that depends on the distance measured. This exact figure was calculated using the number of pixels. A value of present (= 1) or absence (= 0) was returned by the binary measurements. All of the real-valued and binary measurements were taught to the neural network on its own. The effectiveness of a number of real-valued and binary characteristics in identifying a certain face expression was retrieved and tested.

The features that didn't provide any useful information about the facial expression seen in the image were removed and didn't appear in the final analysis. Initially, the CNN was given diverse data for different types of moods in order to train it. A clip comprising all of the different types of expressions was created in a unique way. There were happy, sad, furious, panicked, and neutral expressions. The retrieved images for each mood were then annotated accordingly. All of the parameters (listed above) were retrieved using automated approaches from these various videos. The predicted feelings have to be compared to the true emotions in the areas where the annotations were applied [13].

After the data has been trained, a CNN model is produced, which will be tested against the eight different extracted driver images. For various sorts of parameters, the CNN compares the data from the testing set to the values obtained from the training set. If the value corresponds to a specific mood type, the image is classified into that group. The images could be categorized as happy, angry, sad, panic and sleepy [14].

4.2 Image Processing

To filter and transform our images so that they are prepared to be fed to the model for training, we followed five simple steps [15]. The first step is to convert photos to grayscale. Next, identify the face in the image using OpenCV HAAR Cascade. Step 3 involves cropping the photo to make only the face visible. Then, resize the image to 350*350 pixels in step 4, and step 5 save the picture. Currently, the photographs are perfect for feeding to the model [16]. The color of the image has no effect on how things turn out. In addition, a lot of human photographs are essentially in black and white. We decided to convert the remaining images to grayscale to make all of the images identical so that the model would treat them equally during training, regardless of color.

4.3 Creating Bottleneck Features from VGG-16 Model

Convolution neural networks are challenging to construct from the ground up. Convolution is a computationally expensive technique, so the issue gets significantly worse when we don't have adequate computing power [17]. Convolution and building a CNN layer would take a lot of time, even if we only had a few photos. So, we made the decision to use transfer learning in order to save this unneeded investment.

The idea of transfer learning enables us to use what we have discovered from other previously trained models on our own data. Instead of creating our own custom neural network, we can leverage other well-known pre-trained models and feed them our data to extract the features for our photographs. By default, the convolution layer

develops characteristics for the images. Each pixel in the photos is subjected to a convolution process, producing an array of learned image features that is 'n' dimensional.

The final features of an image that we obtain after a convolution neural network are called bottleneck features. These bottleneck features are the learnt qualities of the pictures, which are transmitted to the MLP, which serves as a top-model [18]. This MLP then minimizes the loss function (function that takes the true values and predicted values as required parameters), updating the CNN kernels/filters as well as the MLP weights. Now, we've selected the VGG-16 pre-trained neural network to produce bottleneck characteristics. 13 of the 16 layers in the VGG-16 network are convolutional layers. Millions of photos from the ImageNet dataset were used to train this neural network.

Thankfully, Keras has this VGG-16 trained network available. We have added ImageNet weights to this pre-trained VGG-16 network as a result. This network was used to process each image individually, producing bottleneck features and storing them in a numpy array. Now all we need to do is use the model, the VGG-16 model [18]. In order to develop bottleneck features for our photos, we employed the predict () algorithm. We then produced bottleneck characteristics as shown in Figure 3 for every one of our images and saved them on our hard drive. We were able to transfer learning from the VGG-16 model to our own challenge using this technique.

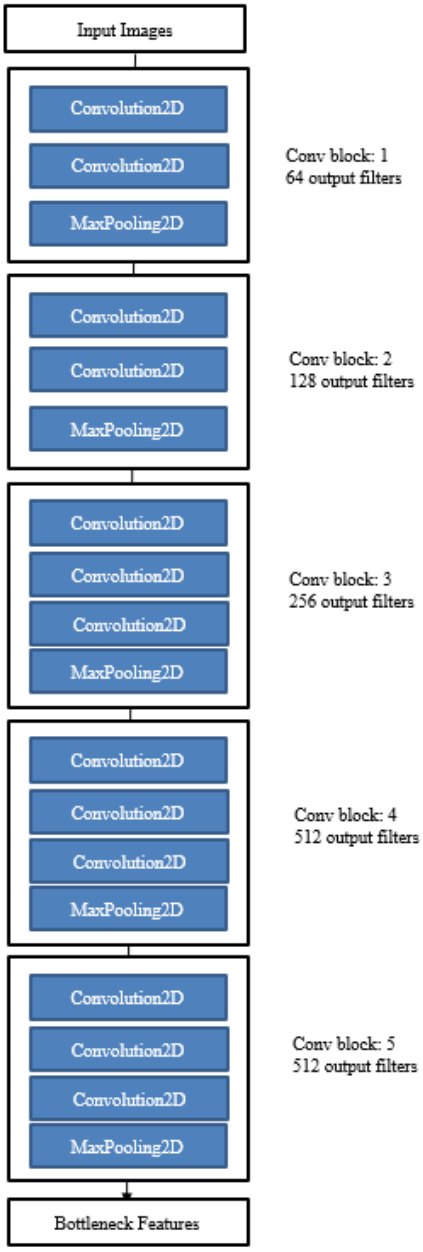

Figure 3. Bottleneck features

5. Results and Discussion

We already have bottleneck features for every single one of our photos. To reduce Multi-Class Log-Loss/Cross-Entropy loss, we must now create a top-model, or MLP model, that will take each image's bottleneck feature one at a time. We developed the neural network depicted in Figure 2 to accomplish this. The diagram above shows that we have a total of five closely connected layers. All of them have RELU activation units. The top layer has 7 softmax units, followed by the fifth layer's 512 activation units, 256 activation units, 128 activation units, and 64 activation units.

Nothing more than a multi-class generalization of logistic regression characterizes these softmax units. In a word, it is multi-class log loss. Seven probability values will be generated, one for each of the seven classifications. The sum of all the probabilities is one. Back-propagation is used to reduce the final cross-entropy loss, which is a result of this. As a result of this training, our MLP model will be able to categorize face expressions in the photos. On human image CV data, we obtained a final accuracy of 87%. The outcomes from epoch 1 to 20 of our model's 20-epoch run are as shown in Table 1.

Table 1. Reduction in losses and increased accuracy

Loss/Accuracy	%
Training Loss	2.45 to 0.04
CV Human loss	2.89 to 0.46
Train Accuracy	17% to 99%
CV Human Accuracy	15% to 87%

Note: After testing on human test data, we got following result: Accuracy on Human Test Data = 82.67%

5.1 Confusion matrix

We can see in the confusion matrix (Figure 4) that some of the classes are more inclined towards the “ANGRY” category. The recall matrix can also be used to verify this. Many images in the “SAD” category are expected to be “ANGRY.” Model is unable to discriminate between the “ANGRY” and “SAD” classes, may be due to the small differences in angry and sad emotions [19, 20] (Table 2).

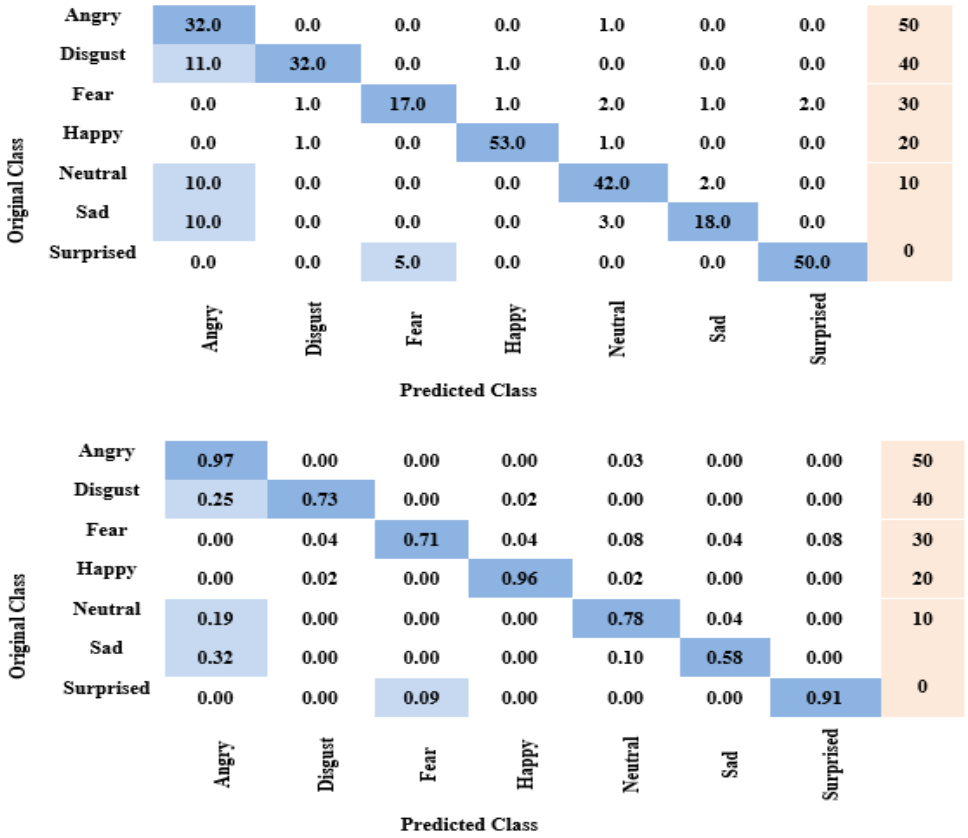

Figure 4. Confusion and recall matrix

Table 2. Emotions and accuracy

Emotions	Accuracy
ANGRY	0.7257885932922
DISGUST	0.009196578525006
FEAR	0.00836753286421299
HAPPY	0.0022135425824671
NEUTRAL	0.130114763975143
SAD	0.12406197190284729
SURPRISE	0.00025709287729

6. Conclusions

According to our findings, a driver's mood has a significant impact on his driving habit. So, if the driver is driving while angry, and the mood remains angry for more than 5 seconds, our system will alert the driver to his angry mood and instruct him to calm down over the car's audible system. It also alerts the driver if the camera records him in a depressed state for more than 5 seconds. This will assist the diver in changing his mood and driving more safely. The emotion data captured with the camera will aid in determining the driver's mood during or before an accident. This will aid study into how accidents are linked to the driver's mood, as well as the development of more sophisticated systems for safer driving for drivers.

We obtained a decent outcome, but there is still a lot of room for improvement. We need a lot more human images with a lot of variety among them to attain higher accuracy. To improve accuracy, we can fine tune the last two or three convolution blocks of the VGG-16 layer. We can also determine the driver's mood by listening to his voice.

Author Contributions

Conceptualization; methodology; software; validation; formal analysis; investigation; data curation, Pawan Wawage. Writing—original draft preparation, Pawan Wawage; writing—review and editing; supervision; project administration, Yogesh Deshpande. All authors have read and agreed to the published version of the manuscript.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] M. T. Harandi, C. Sanderson, S. Shirazi, and B. C. Lovell, "Graph embedding discriminant analysis on Grassmannian manifolds for improved image set matching," *In CVPR 2011*, Colorado Springs, CO, USA, June 20-25, 2011, IEEE, pp. 2705-2712. https://doi.org/10.1109/CVPR.2011.5995564.

[2] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (CVPR 2001), Kauai, Hawaii, USA, December 8-14, 2001, IEEE, pp. 1-1. https://doi.org/10.1109/CVPR.2001.990517.

[3] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," *In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, (CVPRW 2010), San Francisco, CA, USA, June 13-18, 2010, IEEE, pp. 94-101. https://doi.org/10.1109/CVPRW.2010.5543262.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint*, vol. 2014, 2014. https://doi.org/10.48550/arXiv.1409.1556.

[5] M. T. Harandi, C. Sanderson, S. Shirazi, and B. C. Lovell, "Graph embedding discriminant analysis on Grassmannian manifolds for improved image set matching," *In CVPR 2011*, Colorado Springs, CO, USA, June 20-25, 2011, IEEE, pp. 2705-2712. https://doi.org/10.1109/CVPR.2011.5995564.

[6] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, (CVPR 2001), Kauai, Hawaii, USA, December 8-14, 2001, IEEE, pp. 1-1. https://doi.org/10.1109/CVPR.2001.990517.

[7] S. Vora, A. Rangesh, and M. M. Trivedi, “On generalizing driver gaze zone estimation using convolutional neural networks,” *In 2017 IEEE Intelligent Vehicles Symposium*, (IV 2017), Los Angeles, CA, USA, June 11-14, 2017, IEEE, pp. 849-854. https://doi.org/10.1109/IVS.2017.7995822.
[8] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, “Web-based database for facial expression analysis,” *In 2005 IEEE International Conference on Multimedia and Expo*, (ICME 2005), Amsterdam, Netherlands, July 6, 2005, IEEE, pp. 5-5. https://doi.org/10.1109/ICME.2005.1521424.
[9] C. E. Izard, “Emotion theory and research: Highlights, unanswered questions, and emerging issues,” A*nnu. Rev. Psychol*., vol. 60, pp. 1-25, 2009. https://doi.org/10.1146/annurev.psych.60.110707.163539.
[10] F. Eyben, M. Wöllmer, T. Poitschke, B. Schuller, C. Blaschke, B. Färber, and N. Nguyen-Thien, “Emotion on the road-necessity, acceptance, and feasibility of affective computing in the car,” *Adv. Hum. Com. Int*., vol. 2010, Article ID: 263593, 2010. https://doi.org/10.1155/2010/263593.
[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84-90, 2017. https://doi.org/10.1145/3065386.
[12] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, “Emotion recognition using facial expressions,” *Procedia Comput. Sci*., vol. 108, pp. 1175-1184, 2017. https://doi.org/10.1016/j.procs.2017.05.025.
[13] P. Suja, S. Tripathi, and J. Deepthy, “Emotion recognition from facial expressions using frequency domain techniques,” *Adv. Intell. Syst*., vol. 264, pp. 299-301, 2014. https://doi.org/10.1007/978-3-319-04960-1_27.
[14] Q. R. Mao, X. Y. Pan, Y. Z. Zhan, and X. J. Shen, “Using Kinect for real-time emotion recognition via facial expressions,” *Front Inform Tech. El*., vol. 16, no. 4, pp. 272-282, 2015. https://doi.org/10.1631/FITEE.1400209.
[15] Y. I. Tian, T. Kanade, and J. F. Cohn, “Recognizing action units for facial expression analysis,” *IEEE T. Pattern Anal*., vol. 23, no. 2, pp. 97-115, 2001. https://doi.org/10.1109/34.908962.
[16] D. K. Jain, P. Shamsolmoali, and P. Sehdev, “Extended deep neural network for facial emotion recognition,” *Pattern Recogn. Lett*., vol. 1, no. 120, pp. 69-74, 2019. https://doi.org/10.1016/j.patrec.2019.01.008.
[17] Y. Chen, Z. Zhang, L. Zhong, T. Chen, J. Chen, and Y. Yu, “Three-stream convolutional neural network with squeeze-and-excitation block for near-infrared facial expression recognition,” *Electronics-Switz*, vol. 8, no. 4, Article ID: 385, 2019. https://doi.org/10.3390/electronics8040385.
[18] S. B. Wu, “Expression recognition method using improved VGG16 network model in robot interaction,” *J. ROBOT*, vol. 2021, Article ID: 9326695, 2021. https://doi.org/10.1155/2021/9326695.
[19] Y. Said and M. Barr, “Human emotion recognition based on facial expressions via deep learning on high-resolution images,” *Multimed Tools Appl*., vol. 80, pp. 25241-25253, 2021. https://doi.org/10.1007/s11042-021-10918-9.
[20] B. C. Ko, “A Brief review of facial emotion recognition based on visual information,” *Sensors*, vol. 18, no. 2, pp. 401-401, 2018. https://doi.org/10.3390/s18020401.