International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 09 | Sep 2024 www.irjet.net p-ISSN: 2395-0072

Facial Expression Recognition in Advanced Driving Assistance System

Dhruv Nadkar[1], Aditya Avhad[2], Bhushan Gajare[3]

[1]*Professor V.R. Jaiswal, Pune Institute of Computer Technology, Pune, Maharashtra, India*

Abstract - *In the automobile domain, advanced driver-assistance systems (ADASs) are employed to improve safety, however existing ADASs do not take into consideration drivers' conditions, such as whether they are emotionally suited to drive. In the automotive industry, advanced driver-assistance systems (ADASs) are employed to improve safety, however existing ADASs do not consider drivers' circumstances, such as whether they are emotionally fit to drive. Many road accidents and unanticipated situations are caused by driver inattention, which is one of the key characteristics and reasons.Face expression recognition is a relatively new image processing technique that is becoming increasingly important in applications such as driver warning systems. Even when given a noisy input or incorrect data, current algorithms can recognise facial expressions, but they lack accuracy. It's also useless when it comes to dealing with uncontrollable emotions and recognition. Based on Deep Neural Networks and Convolutional Neural Networks, the proposed technique provides a driver warning system that efficiently identifies face expressions (CNN).*

Key Words: **Convolutional Neural Networks (CNNs), Deep Learning (DL), Driver Warning System, Facial Emotion Recognition (FER), Advance Driving Assistance System (ADAS).**

1. INTRODUCTION

Excessive driver distractions, alcohol consumption, and speeding beyond safe limits are widely recognized as major causes of road accidents and mishaps. According to statistics from respective departments, it is observed, another very crucial factor contributing to road accidents is the fatigue condition that a driver experiences while driving. Drivers experiencing mental fatigue also suffer from excessive sleepiness and loss of consciousness after regular intervals. Drivers driving for more than 8 hours a day, undergoing immense physical activities and lack of sleep usually suffer mental fatigue.

In recent years, a rising amount of automation has penetrated the automobile industry. In terms of manual driving, this automation has opened up new options. On-board Advanced Driver-Assistance Systems (ADASs), which are used in automobiles, trucks, and other vehicles, provide exceptional opportunities for increasing the quality of driving, safety, and security for both drivers and passengers. Adaptive Cruise Control (ACC), Anti-lock Braking System (ABS), automotive night vision, drowsiness detection, Electronic Stability Control (ESC), Forward Collision Warnings (FCW), Lane Departure Warning System (LDWS), and Traffic Sign Recognition (TSR) are examples of ADAS technology. The majority of ADASs are electronic systems that adapt and improve vehicle safety and driving quality. By correcting for human mistakes, they have been shown to minimize road deaths. Proposed System uses CNN and deep learning algorithms to detect and predict facial emotion of drivers by keeping track of images and through videos.

2. LITERATURE SURVEY

Face recognition has gotten a lot of attention in recent years as one of the most successful uses of image analysis and comprehension. Automatic FER approaches have been extensively researched for many years, and because the use of the most discriminative options is the most important factor determining a FER method's effectiveness, they'll be divided into two categories: those using hand-sewn options and those using options generated by a deep learning network. One of the most prominent uses of FER is in the Advanced Driving Assistance System.

The driver's gaze was fixed on the specially constructed drowsy driver detection system, which was utilized to assess weariness. Techniques for detecting drowsiness are divided into two groups based on the criteria employed for detection: intrusive and non-invasive methods of detection. The distinction is made depending on whether or not an instrument is attached to and paired with the driver. An instrument is well-connected to the driver in the invasive technique, and the value of that instrument is examined and recorded.

In their paper, S. Suchitra, S. Sathya Priya, R. J. Poovaraghan, B. Pavithra, and J. Mercy Faustina [1] propose a Local Octal Pattern-Convolutional Neural Network (LOP-CNN) approach to develop a more efficient Driver Warning System using Deep Learning. The CNN-based feature extraction reduces the semantic gap and enhances overall performance by utilizing Facial Expression Recognition. In their paper, Mira Jeong and Byoung Chul Ko [2], explain that ADAS integrates psychological models, sensors to capture physiological data, human emotion categorization algorithms, and human-car interaction algorithms. Researchers are focusing on issues such as using subtle psychological and physiological indicators (e.g., eye closure) to enhance the detection of dangerous situations like distraction and

drowsiness. They also aim to improve the accurate recognition of risky conditions, such as fatigue, by monitoring driving performance metrics like "drift-and-jerk" steering and detecting vehicle movements in different directions.

"Pixel Selection for Optimizing Facial Expression Recognition Using Eigenfaces," by C. Frank and E. Noth, was a contribution. They devised a technique for facial expression recognition that uses only pixels that are relevant for facial expressions to generate an eigenspace. A training set of face photos with facial emotions selects these relevant pixels automatically.Eigenspace techniques are well-known in the field of face recognition (e.g. [Tur91], [Yam00], [Mog94]). Each person's eigenspace is constructed using numerous images of that person in a traditional face recognition system. To create an eigenspace with training photos, a partial Karhunen-Loeve transformation, also known as principal component analysis (PCA), is utilized.

"Facial Emotion Analysis Using Deep Convolutional Neural Network," by Rajesh Kumar G A, Ravi Kant Kumar, and Goutam Sanyal [6], was published in 2017. They believed that human emotions are mental states of sentiments that develop without conscious effort and are accompanied by physiological changes in facial muscles, resulting in facial expressions. Happy, sad, rage, contempt, fear, surprise, and other key emotions are examples. Facial expressions play a significant part in nonverbal communication since they represent a person's interior sentiments. There has been a great deal of study on computer modeling of human emotion. However, it still lags behind the human visual system by a factor of 13. They are employing deep Convolutional Neural Network (CNN) to provide a better technique to anticipate human emotions (Frames by Frames) and how emotional intensity varies on a face from low to high levels of emotion in this system. The FERC-2013 database has been used to train this algorithm. The results of the suggested experiment are fairly good, and the accuracy attained may stimulate researchers to develop future models of computer-based emotion identification systems.

3. PROPOSED METHODOLOGY

A Convolutional Neural Network, which is also known as CNN, is a distinct type of neural network that primarily focuses on processing information that has a grid-like structure, for example, a picture. A binary illustration of visual information could be a digital image. It possesses a series of constituents arranged in a grid-like fashion that has component values to indicate how bright and which color every pixel has to be.

A CNN generally has 3 layers: a convolutional layer, a pooling layer, and a fully connected layer.

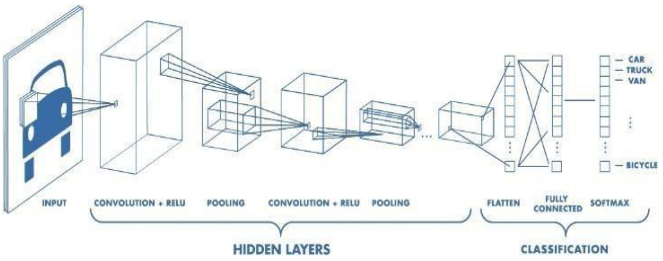

Fig-1: CNN architecture

3.1 Convolutional Layer

The convolution layer is the core component of a CNN, responsible for most of the network's processing power. It performs a dot product between two grids: the learnable parameters (kernel) and a specific portion of the receptive field. Although the kernel is smaller than the image, it holds a significant amount of information. For an image with three (RGB) channels, the kernel's height and width are smaller, but its depth spans all three channels. During the forward pass, the kernel moves across the image's height and width, producing a visual representation of the receptive area. This results in an activation map, a 2D representation showing the kernel's response at each spatial location.

If we have an input of size W x W x D and a Dout variety of kernels with an abstraction size of F, stride S, and padding amount P, the output volume scale is frequently given by the following formula:

$$W_{out} = \frac{W - F + 2P}{S} + 1$$

Motivation behind Convolution:

Three key principles driving advancements in computer vision research are distribution interaction, parameter sharing, and equivariant representation. Let's explore each in detail. Simple neural network layers rely on matrix operations, where a matrix of parameters represents the interaction between input and output units, meaning every output unit is connected to every input unit. In contrast, convolutional neural networks (CNNs) have more selective interactions by using a kernel smaller than the input. Although images may contain thousands or millions of pixels, the kernel focuses only on relevant information covering tens or hundreds of pixels. This reduces the number of parameters, lowering memory requirements and improving statistical efficiency. If detecting a feature at one location (x1, y1) is useful, it should also be useful at another location (x2, y2). Thus, for creating an activation map, CNNs force neurons to use the same set of weights. Unlike traditional neural networks, where each part of the weight matrix is used only once, CNNs share parameters,

meaning the same weights are applied across different input regions to generate the output.

The layers of convolutional neural networks can have equivariance to translation due to parameter sharing. It states that if we change the input in one method, the output will also be changed in that method.

3.2 Pooling Layer

The pooling layer replaces the network's output at constrained areas using an outline statistic of surrounding outputs. This decreases the abstraction size of the illustration, lowering the needed number of computationsand weights. During the pooling step, each slice of the artwork is handled independently.

The common of the rectangle neighborhood, L2 norm of the rectangular neighborhood, and a weighted average are all pooling functions that support the gap from the center component. The most used option, however, is max pooling, which reports the neighborhood's maximum output.

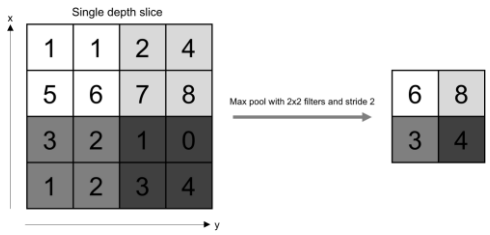

Fig-2: Max Pooling

If we've got an activation map of size W x W x D, a pooling kernel of abstraction size F, and stride S, then the scale of output volume are often determined by the subsequent formula:

$$W_{out} = \frac{W - F}{S} + 1$$

3.3. Fully Connected layer

In a typical CNN, neurons in this layer are fully connected to all neurons in both the previous and next layers. As a result, matrix operations followed by a bias adjustment can be used to compute their values.

The FC layer assists with the illustration's mapping between the input and output.

4. IMPLEMENTATION

4.1 Face Detection

ViolaJones Face Detector is used to recognize and search for faces in photos and videos taken by the camera. Learning is slower in response to this strategy, but detection is faster. This approach doesn't use multipliers, it uses filters from the underlying Haar function. Most searches occur in a single search window. After optimally determining the minimum and maximum frame sizes, the detection frame is moved correctly over that face image and finally a sliding step size is chosen for each recognizable size. It then uses the active shape model to find many available facial landmarks in the face image. In most cases, groups of points are used to represent the shape of an object. Facial landmarks are mapped using an ideal mapping approach based on center of gravity coordinates. Facial landmarks are categorically aligned with their average face shape model using an optimal mapping procedure based on centroid coordinates. The centroid mapping procedure uses these individual face markers to align each face image in the database for an average face shape approach.

4.2 Face Feature Extraction

In the feature extraction step, a 7X5 grid is essentially extracted for each observed facial component, and each grid is represented by a square patch. The aligned face image has a total of 175 meshes represented by the 5 main facial features, including the two corners of the mouth, the tip of the nose, and the two eyes, using the same approach as in the projection. Each grid contains a section of a face image. Uprooted and retrieved, a LOP function handle is created and evaluated as a local function. After that, the CNN is trained and matched using the LOP feature vectors as input. It critically recognizes facial expressions in specific supported categories based on CNN output and combines each LOP feature vector matching with a fully connected network. The LOP function is successfully passed to the CNN classifier for further classification. The technology we propose raises an alarm when the driver's facial expression is detected.

4.3 Facial Expression Recognition Using CNN

Convolutional neural networks are one of the best known and ideal ways to represent network topologies in the field of deep learning. In the field of identification and classification of facial images, he has undoubtedly gained popularity. ConVnet quickly processes the generated data in the form of packages of multiple array sets. A CNN can take raw image data as input and run it through feature extraction and reconstruction techniques as part of an existing learning algorithm. In terms of the structure of weight distribution networks, they are similar to biological neural networks. As a result, the overall complexity of the network model and the number of weights associated with it are reduced. Directions such as transforms, scales, gradients, and other transformations do not affect the CNN. The deep learning facial emotion recognition method is very reliable. An additional preprocessing approach for fast training on input face

pictures based on facial physics. CNNs are the preferred network model among all existing deep learning models. In the CNN-based approach, the input image is convolved with a filter bank of convolutional layers to obtain the extracted vector feature maps. When each feature vector map is integrated and connected to a fully connected network, facial expressions are identified naturally. CNNs are made up of numerous layers. Convolutional, MaxPooling, and fullyConnected layers are the three types of layers you can use. The first convolutional layer is applied to the main components of the face, such as the left eye, right eye, the two corners of the mouth and the tip of the nose. A rectangular grid of neurons. Each neuron in the convolutional layer receives and processes input from the rectangular region of the previous face component layer, and the weight of this rectangular section is the same for all neurons in the convolutional layer. The result is the convolution of the previous layer image and the weights simply define the convolution filter. Additionally, each convolutional layer can contain multiple meshes that use inputs from previous layers, most with completely different filters. When the convolutional layer is completed, a pooling layer is added. The merge layer subsamples small rectangular pieces of face components in the previous convolutional layer to produce a single composite output from that block. Pooling can be done in a number of ways, such as using a maximum or average number of neurons per block, or using a linear combination of neurons per learned block. The merged layer that counts can be the maximum merged layer. This means that the maximum number of blocks being merged in the face element is taken into account. Finally, multi-layer fully connected layers are used to perform high-level modes of interpretation within neural networks when applied to multiple layers of convolution and max pooling. This stratified layer collects input from all neurons in the previous face component layer and connects with each neuron. A fully connected layer is not followed by a convolutional layer because a fully connected layer is not spatially constrained in the same way as a convolutional layer. The LOPCNN approach is shown in Figure 3. For example, suppose you have aconvolutional layer after your neural layer. To compute a pre-nonlinear input for units in a layer, we need to add a weight filter component from the cells of the previous layer and provide an output.

4.4 Quantization Process and Indexing

After obtaining local feature descriptors, each descriptor is quantized into sparse codewords in the quantization process and indexing step, and the same approach is applied to all components of the same face image to identify different semantic codewords.

4.5 SVM based Recognition

Support Vector Machine (SVM) is a supervised machine learning technique that efficiently generates hyperplanes in high-dimensional space to perform classification or recognition tasks. The goal of SVM training is to choose the optimal hyperplane that maximizes the difference between the two classes. SVMs work based on binary classifiers that divide or classify data into two groups.

5. WORKFLOW

The activity will begin with the definition of realistic UCs, which will subsequently be implemented in an automobile context based on various User Stories (USs) (as well as in a driving simulator). Simultaneously, the system architecture will be designed in order to determine which sensors and communication needs should be included in the development.

The software components will then be trained through an iterative experimental data gathering effort.

Finally, the state detection system and the Decision Support System (DSS) will be incorporated in the driving simulator.

6. RESULTS

By analyzing the previous model and keeping in mind its pros and cons given in the reports. We have divided the process of optimizing the model into 1 test case which is face-aging or Face-age invariance. As shown in the Confusion matrix classification algorithms performs 73%

on anger expression, 74.4% on disgust, 47.1% on fear, 94.8% for happiness, 78.8% on sadness and for surprise it performs 82.0%.

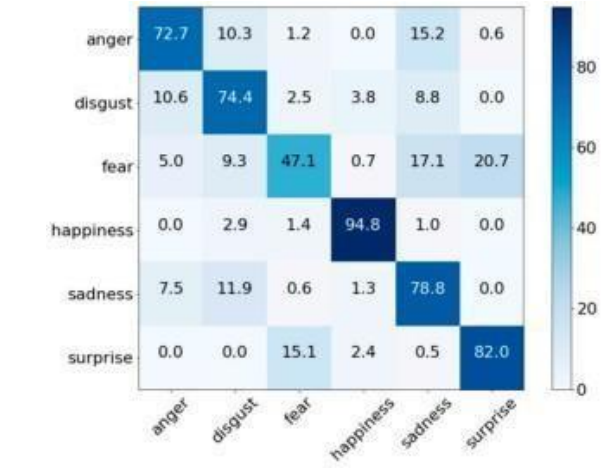

Fig-3: Confusion Matrix

Fig-4: Angry Expressions

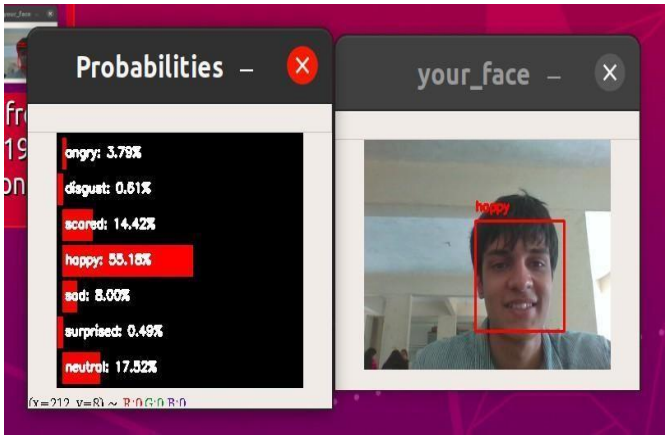

Fig-5: Happy Expressions

6. CONCLUSION

This paper provides a thorough review of the latest technologies in ADAS, highlighting both the opportunities for enhancing driving quality and safety and the limitations, such as some systems only using basic mechanisms to consider the driver's state or ignoring it entirely. The study centers on emotional and cognitive analysis in ADAS development, with a focus on machine learning techniques like Convolutional Neural Networks. The review stresses the need to improve classifiers that assess both cognitive and emotional states.

REFERENCES

[1] S.Suchitra, S.Sathya Priya, R.J. Poovaraghan, B.Pavithra, J.Mercy Faustina, "Intelligent Driver Warning System using Deep Learning-based Facial Expression Recognition"International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-3, September 2019

[2] Mira Jeong and Byoung Chul Ko, "Driver's Facial Expression Recognition in Real-Time for Safe Driving" Keimyung University, Daegu 42601, December 2018.

[3] Lee B G, Jung S J and Chung W Y "Real-time physiological and vision monitoring of vehicle driver for non-intrusive drowsiness detection", IET Commun.,2011

[4] Reza Azmi and Sam iraYegane,"Facial Expression Recognition in the Presence of Occlusion Using Local Gabor Binary Patterns", 20th Intl Conf. on Electrical Engg., May 15-17, Tehran, Iran

[5] Qintao Xu,Najing Zhao, "A Facial Expression Recognition Algorithm based on CNN and LBP Feature," 2020 IEEE 4th Information Technology,Networking,Electronic and Automation Control Conference (INTEC 2020).

[6] "Facial Emotion Analysis using Deep Convolution Neural Network", Rajesh Kumar G A, Ravi Kant Kumar, Goutam Sanyal, IEEE, 2017.

[7] Jeong, M.; Ko, B.C.; Kwak, S.; Nam, J.Y. Driver Facial Landmark Detection in Real Driving Situations. IEEE Trans. Circuits Syst. Video Technol. 2018, 28, 2753–2767.

[8] Hasani, B.; Mahoor, M.H. Facial Expression Recognition Using Enhanced Deep 3D Convolutional Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 2278–2288.

[9] Eduard Zadobrischi, Lucian-Mihai Cosovanu, Mihai Negru and Mihai Dimian, "Detection of Emotional States Through the Facial Expressions of Drivers Embedded in a Portable System Dedicated to Vehicles" Serbia, Belgrade, November 24-25, 2020.

[10] Mollahosseini, A.; Chan, D.; Mahoor, M.H. Going deeper in facial expression recognition using deep neural networks. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016; pp. 1–10.