Artificial Intelligence for Monitoring Vehicle Driver Behavior "Facial Expression Recognition"

Abdelhak Khadraoui, Elmoukhtar Zemmouri
Ecole Nationale Supérieure d'Arts et Métiers Meknès
Moulay Ismail University, Meknes, Morocco

Keywords: Facial Expression Recognition, ADAS, Active Security, Deep Learning, Vehicle Driver.

Abstract: Advanced Driver Assistance Systems (ADAS) are the first step towards autonomous vehicles. They include systems and technologies designed to make more accessible the driver's attitude to prevent accidents and prevent an accident. My thesis project fits into this context. The objective is to design and develop an intelligent, and active safety system capable of detecting the driver's state and alerting in real-time, in case of fatigue, stress, drowsiness (half-sleep), or gestures likely to disturb his attention while driving. We are particularly interested in the recognition of the driver's facial expressions to detect the states of fatigue, stress, or drowsiness. The system will then monitor the driver in real-time and send him personalized alerts and notifications asking him, for example, to stop for a coffee break, to change the music or the temperature inside the car.

1 INTRODUCTION

Abnormal driving behaviour such as drowsiness is defined by the driver's behaviour, which increases the risk of accidents, However, there is still no driver behaviour monitoring system capable of effectively distinguishing different abnormal driver behaviours.

The main contribution of my paper is to propose a CNN-based approach to detect a driver's behaviour based on his face monitoring.

One of the most visible manifestations of human emotions is facial expressions. Recognizing facial expressions allows us to better comprehend how others are feeling. As the demands of human-computer interaction have grown in recent years, a lot of research on facial expression recognition (FER) has arisen.

Convolutional Neural Network is built with five convolutional layers and one fully connected layer. Adam is used as the optimizer, categorical cross-entropy as the loss function and accuracy as the evaluation metric.

1.1 General Background

Emotional factors have a substantial impact on social intelligence, such as communication, understanding, and decision-making; they help in comprehending human behavior. Emotion plays a pivotal role during communication.

Emotion recognition can be done in a variety of ways. It could be either verbal or nonverbal. Face expressions, actions, bodily postures, and gestures are non-verbal modes of communication, while voice (Audible) is an oral method of communication.

1.2 Problem Statement

To predict the correct facial expression out of the seven universal facial expressions: Happiness, Sadness, Surprise, Fear, Disgust, Anger and Neutrality, for the detected face.

1.3 Objectives

The project seeks to satisfy the following objectives:

Analysing the video input or real-time camera input fed to the model and classifying the expression into one of the seven basic facial expressions (Universal expressions): Happiness, Sadness, Surprise, Fear, Disgust, Anger and Neutrality.

It is detecting the face accurately and predicting the emotion using the integrated model.

Khadraoui, A. and Zemmouri, E.
Artificial Intelligence for Monitoring Vehicle Driver Behavior "Facial Expression Recognition".
DOI: 10.5220/0010732500003101
In *Proceedings of the 2nd International Conference on Big Data, Modelling and Machine Learning (BML 2021)*, pages 279-284
ISBN: 978-989-758-559-3

2 RELATED WORK

2.1 Emotion

An emotion is a set of automatic responses to external situations. There are bodily responses, of course: facial movements, heart racing, gestures, sweat running down the face. Each of us has already had these experiences, whether it be during an oral exam or a love encounter.

There are different human emotions, a group of psychologists proposed a list of basic emotions ranging from 2 to 18 categories, Ekman and his group conducted various studies on facial expressions, which led to the conclusion that there are six basic emotions, also called primary emotions: joy, sadness, surprise, fear, anger, and disgust.

Table 1 : different emotions proposed by psychologists

Authors	Emotions
Izard (1977)	Joy, surprise, anger, fear, sadness, contempt, distress, interest, guilt, shame,
Ekman (1992)	Anger, fear, sadness, joy, disgust, surprise

2.2 Comparison of Three Critical Articles Related to Our Contribution:

In one of the most-known works in emotion recognition by Shervin Minaee, Amirali Abdolrashidi have shown a Deep Learning alignment work. This is a robust Convolutional NN-based face calibration algorithm. They propose that the Deep Alignment Network performs face calibrations primarily based on the whole-facial images, as opposed to recent face alignment algorithms, which makes it remarkably accurate with large changes in both initializations and forehead postures.

Mira Jeong and Byoung Chul work on emotion recognition has defined the system "Affective Computing" as the development of recognizable, interpretable systems, devices, and mechanisms that imitate a person's affects through various attributes such as how they look, the depth and modulation of their voice, and any biological signals they may have.

In their Article Kamran Ali, Likin Isler, and Charles Hughes, have shown the facial expression recognition system, which is a real-world application and solves the phases that occurred in the post changes made. The authors have generated several new tests over FER Datasets on these phases and proposed a new "Region Attention Network (RAN)" which itself depicts the importance of the facial landmarks.

2.3 Databases

The proposed databases for processing a facial expression recognition system:

CK+: The extended Cohn-Kanade (known as CK+)

FER2013: The Facial Expression Recognition 2013

JAFFE: This Dataset contains 213 images of the seven facial expressions posed by 10 Japanese female models

FERG: FERG is a database of stylized characters with annotated facial expressions.

The MMI database contains 213 image sequences, of which 205 sequences with frontal view faces of 31 subjects were used in our experiment.

3 METHODS

HA-GAN: The fundamental goal of the HA-GAN method is to learn how to encode expression information from an input image and then transfer that information to a produced image file. The resulting expression images are then subjected to facial expression recognition.

Hierarchical weighted random forest (WRF): For real-time embedded systems, geometric features and the hierarchical WRF is used.

Attentional Convolutional Network (ACN): an attentional convolutional network technique that can focus on parts with a lot of features

VGGNet is invented by the Visual Geometry Group (in Oxford University)

Table 2 : comparison of three critical articles related to our contribution

Articles	Proposed Methods	Methods	Datasets	Accuracy
S Minaee, A. Abdolrashidi, 2019	Attentional Convolutional Network (ACN)	VGG+SVM	FER2013	66.31%
		GoogleNet		65.2%
		ACN		70.02%
		CNN+SVM	JAFFE	95.31%
		ACN		92.8%
		ACN	CK+	98.0%
		ACN	FERG	99.3%
Mira Jeong and Byoung Chul Ko, 2018	WRF	WRF and deep neural network-based	CK+	92.6%
		WRF	MMI	76.7%
Kamran Ali, Ilkin Isler, and Charles Hughes, 2019 []	HA-GAN	HA-GAN	CK+	96.14%
		HA-GAN	MMI	71.87%
		HA-GAN	Oulu-CASIA	88.26%

4 EXPERIMENT

4.1 Face Detection

Face detection is described as the process of identifying and extracting faces (location and size) from an image for use by a face detection algorithm.

4.2 Dataset and Pre-Processing

The Dataset consists of grayscale images of faces at a resolution of 48x48 pixels. The faces have been automatically registered so that each illustration has a central face that occupies roughly the same amount of space. Each face is categorized according to the emotion expressed in the facial expression (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutrality).

The training set consists of 28708 images and the test set consists of 7178 images.

This Dataset was prepared by Pierre-Luc Carrier and Aaron Courville for the Facial Expression Recognition Challenge 2013 (Kaggle).

CNN includes convolutional layers, subsampling layers, and fully connected layers. Convolutional layers are usually described in terms of kernel size. Sub-sampling layers are employed to increase the kernels' position invariance. The two most popular types of sub-sampling layers are maximum-pooling and average-pooling. Layers that are fully connected are similar to those seen in generic neural networks.

Five convolution layers and two fully connected layers are utilized in the project to anticipate seven different face emotions.

4.3 Facial Expression Recognition

A Facial Expression Model class is created to load the model and load the trained weights into the model and predict facial expressions. This model is then used when the application uses real-time feed.

Flask application

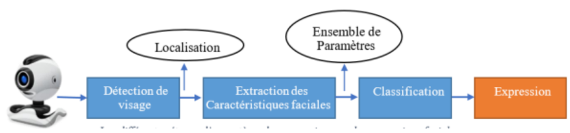

Figure 1: The different steps of a facial expression recognition system

4.4 Algorithm

The FER2013 database is an open-source database that comprises 35887 photos of 48 × 48 faces of various emotions, grouped as follows. It was generated for a project by Pierre-Luc Carrier and

Aaron Courville and later given publicly for a Kaggle contest. 28709 photographs were used as training sets, 3589 images were used as public test sets, and 3589 images were maintained as private test equipment. There are seven labels in total.

Table 3 : Emotion labels of images from the fer2013 database

Index emotion	Emotion	number of images
0	Angry	4593 images
1	Disgust	547 images
2	Fear	5121 images
3	Happy	8989 images
4	Sad	6077 images
5	Surprised	4002 images
6	Neutral	6198 images

This allows us to perform supervised learning.

Examples of the facial expressions of FER2013 shown in the following figure:

Figure 2: Example of images from the fer2013 database

Created a convolutional Neural Network (CNN) with five convolution layers having functions like Conv2D, batch normalization, MaxPooling2D, and the activation function used is Relu. The model is then flattened and is passed through one fully connected layer the activation function used is SoftMax. Next, Adam is used as the optimizer, categorical cross-entropy as the loss function and accuracy as the evaluation metric.

Accuracy and loss function are plotted using PlotLossesKerasTF () function for training and validation sets.

Face detection is completed in the camera.py file using OpenCV and Haarcascade libraries. The real-time inputs from the video streamed, or the camera is taken as individual frames and is then converted into grayscale.

a) The Dataset is split into a training set and a test set. We are now ready to split our model into training, validation, and test sets.

What about the architecture of the model?

[2 x CONV (3,3) — Activation=Relu—Padding = Same] — MAXPooling (2,2)

[2 x CONV (3,3) — Activation=Relu—Padding =Same] — MAXPooling (2,2)

[2 x CONV (3,3) — Activation=Relu—Padding =Same] — MAXPooling (2,2)

[2 x CONV (3,3) — Activation=Relu—Padding =Same] — MAXPooling (2,2)

[2 x CONV (3,3) — Activation=Relu—Padding =Same] — MAXPooling (2,2)

Flatten ()

Dense (1024) — DROPOUT (0.2)

Dense (7, Activation=SoftMax)

MAXPooling (2,2) layers have been added to each convolutional layer's block.
RELU has been picked as an activation function for all convolutional layers.
The total trainable / non-trainable parameters, Model. Summary ()

Total params: 19,935,751

Trainable params: 19,935,751

Non-trainable params: 0

b) Convolutional Neural Network is built with five convolutional layers and one fully connected layer. Adam is used as the optimizer, categorical cross-entropy as the loss function, and accuracy as the evaluation metric.

```
From tensorflow. keras. Optimizers import
Adam.
model.
compile(optimizer=Adam(learning_rate=0.0001)
, loss= 'categorical_crossentropy',
metrics=['accuracy'])
```

Now we can start training our model.

```
history=model.fit (X_train, y_train,
batch_size=64, epochs=100, verbose=1,
validation_data= (X_test, y_test))
```

c) Accuracy and loss function are plotted using PlotLossesKerasTF () function for training and validation sets.

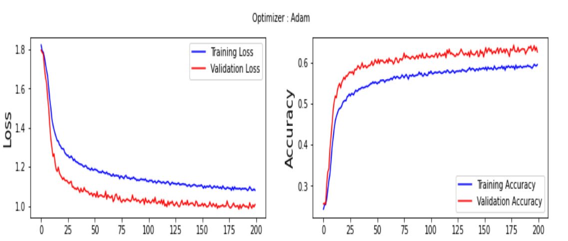

Figure 3: Accuracy and Loss

Thus, the accuracy of learning and testing increases with the number of epochs, reflecting the fact that at each epoch, the model gets more information. If the accuracy is decreased, we will need more information to implement the model, and we also need to increase the number of epochs and vice versa.

Prediction a custom image outside the test set.

Img=image. load_img

'/content/drive/MyDrive/mypicture.jpg'

grayscale=True,target_size= (48, 48))

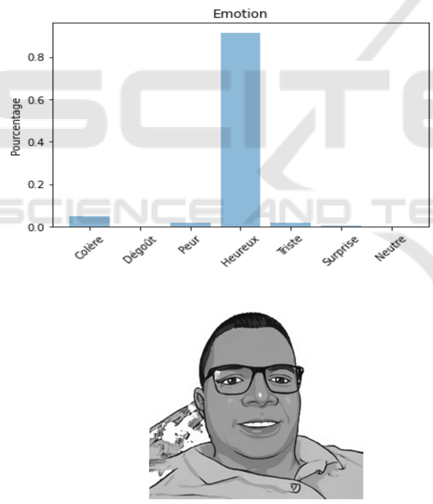

Figure 4: Test our model on photo person happy.

5 RESULTS AND DISCUSSIONS

The project performs the required functions envisioned at the proposal phase. The face is detected accurately, and the emotion is predicted using the integrated model.

Limitations

There might be some lag in the display due to GPU memory constraints.

The project is hosted through a local server.

5.1 Future Scope

The project's goals were carefully kept within the parameters of what was deemed possible given the time and resources available. As a result, this fundamental design can be vastly enhanced. This design appears to be a working micro scale model that might be built up to a much larger scale, based on what has been detailed. The following suggestions are offered as possibilities for expanding this project in the future:

a) An external camera input can be taken instead of the laptop camera.
b) The project can be hosted on a cloud platform after beautifying the flask app.
c) The Dataset can be made more balanced.

6 CONCLUSION

In this paper, we proposed a model for a neural network architecture for facial emotion recognition. we compared several models and found that the best model is the cnn+vg16 in the hidden layers,

The results improve with an error that cancels out as we increase the number of iterations.

REFERENCES

P. Roshan, P. Helonde 2020, Python Code for Face Recognition From Camera - Real Time Face Recognition Using Python, https://matlabsproject.blogspot.com/2020/11/python-code-for-face-recognition-from.html

Khan, Sajid Al, 2014 Robust face recognition using computationally efficient features, https://content.iospress.com/articles/journal-of-intelligent-and-fuzzy-systems

P. Ekman," 1992 An argument for basic emotions," CognEmot.,.

Shervin Minaee, Mehdi Minaei, Amirali Abdolrashidi 2021, Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network.

CK+, The most common lab-controlled dataset, http://www.con sortium.ri.cmu.edu/ckagree/

https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge

The Japanese Female Facial Expression (JAFFE) Dataset, 1998, http://www.kasrl.org/jaffe.html

MMI Facial Expression Database, 2002, https://mmifacedb.eu/

Philipp Terhörst, et al. · Fraunhofer 2020, Unsupervised Enhancement of Soft-biometric Privacy with Negative Face Recognition, https://deepai.org/publication/unsupervised-enhancement-of-soft-biometric-privacy-with-negative-face-recognition

Subodh Lonkar, Facial Expressions Recognition with Convolutional Neural Network, https://export.arxiv.org/ftp/arxiv/papers/2107/2107.08640.pdf

Shervin Minaee, Amirali Abdolrashidi, 2019 Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network, Expedia Group, University of California, Riverside

Mira Jeong and Byoung Chul Ko, 2018 Driver's Facial Expression Recognition in Real-Time for Safe Driving, Department of Computer Engineering, Keimyung University, Daegu 42601, Korea,

Kamran Ali; Ilkin Isler; and Charles Hughes, (2019) Facial Expression Recognition Using Human to Animated-Character Expression, Translation, Synthetic Reality Lab, Department of Computer Science University of Central Florida Orlando, Florida, Hacettepe University, Turkey

I.Michael Revina, W.R.Sam Emmanuel, Face Expression Recognition with the Optimization based Multi-SVNN Classifier and the Modified LDP Features, https://www.sciencedirect.com/science/article/abs/pii/S104732031930152X

jokker99 2020, CNN-Face-Emotion-Recognition, https://github.com/jokker99/CNN-Face-Emo

Hrithik Katoch mar 2018, Twitter Emotion Recognition using RNN, http://katoch.medium.com

Gaurav Sharma Sep 18, 2020, Facial Emotion Recognition (FER) using Keras, https://medium.com/analytics-vidhya/facial-emotion-recognition-fer-using-keras-763df7946a64

Bastien L 2021, https://www.lebigdata.fr/machine-learning-et-big-data

P. Kim, 2017 MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence.

https://www.researchgate.net/figure/Comparison-of-different-types-of-machine-learning_fig6_325928183.

Reynolds, C. and Picard, R. W. 2001 Designing for Affective Interactions. In Proceedings from the 9th International Conference on Human-Computer Interaction, New Orleans,

Ekman, P. and Friesen, W. V. 1978. The Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto: Consulting Psychologists Press.

Donato, G. et al. 1999 Classifying Facial Actions. IEEE Transactions on Pattern Analysis and Machine Intelligence,

Turk, M., Pentland, A. Eigenfaces 1991 for recognition. Journal of Cognitive Neuroscience,

Evaluer les émotions, 2015, https://uxmind.eu/2015/04/09/evaluer-les-emotions/

Furkan Kınlı 2018, [Deep Learning Lab] Episode-3: fer2013, https://medium.com/@birdortyedi_23820/deep-learning-lab-episode-3-fer2013-c38f2e052280