Science, Engineering and Technology *Volume 2, Issue 1, Pages 38-43*
www.setjournal.com *https://doi.org/10.54327/set2021/v2.i1.26*

A pratical implementation of deep neural network for facial emotion recognition

Ferroudja DJELLALI, Emir DELJANIN

SUP de Vinci, 6-12 Av. Léonard de Vinci, 92400 Courbevoie, Paris, France

Abstract

People's emotions are rarely put into words, far more often they are expressed through other cues. The key to intuiting another's feelings is in the ability to read nonverbal channels, tone of voice, gesture, facial expression and the like. Facial expressions are used by humans to convey various types of meaning in a variety of contexts. The range of meanings extends from basic, probably innate, social-emotional concepts such as "surprise" to complex, culture-specific concepts such as "neglect". The range of contexts in which humans use facial expressions extends from responses to events in the environment to specific linguistic constructs in sign languages. In this paper, we will use an artificial neural network to classify each image into seven facial emotion classes. The model is trained on a database of FER+ images that we assume is large and diverse enough to indicate which model parameters are generally preferable. The overall results show that, the CNN model is efficient to be able to classify the images according to the state of emotions even in real time.

Keywords: *facial expressions, artificial neural network, image classification, real time predictions*

1 Introduction

Human faces are arguably the most important things we see. We are quick to detect them in any scene, and they command our attention. Faces express a wealth of important social information, such as whether another person is angry or scared, which in turn allows us to prepare for fight or flight. Does this mean facial expressions are universal? It's a question scientists have debated for half a century, and it remains without a definitive answer. Emotions are essential to our lives. They allow us to improve communication between individuals, to ensure a better understanding of the message conveyed and to adapt to a given situation.

Emotion recognition is an important aspect of affective computing, one of whose objectives is the study and development of behavioral and emotional interactions between humans and machines. This can be useful to verify that the person standing in front of the camera is not just a 2-dimension representation [1].

It is also important because it allows the observer to infer the emotional states and intentions of others and to anticipate their actions, but also to regulate its own behaviors accordingly. Thus, the ability to recognize emotions influences one's ability to adapt to the environment and is therefore an essential skill for interpersonal functioning.

Facial recognition of emotions is linked to several domains, for example:

- Marketing: applications to measure customer satisfaction, to predict the products that interest them.
- Security: stress detection.
- Medicine: help to detect certain psychological diseases
- Human-Computer Interaction: support robot.
- Education: distance learning.

2 History introducing emotions

The history of emotions is based on the research of several scientists, starting with Darwin who in 1872 wrote one of the first hypothesis that will influence the research on emotions [2]. He will be followed since the 1960s by several scientists such as Paul Ekman emotions [3], Carroll Izard, Alan Fridlund and Sylvan Tompkins who will try to demonstrate the universality of certain fundamental emotions for human beings. In the 20th century, the history of emotions took a decisive turn and

Published: 30 April 2022
Corresponding author: *Ferroudja DJELLALI (ferroudje.djellali96@gmail.com)*

only grew thanks to the studies of Lucien Febvre, who was one of its precursors in France.

Emotions have been classified according to two categories: simple or complex. According to Paul Ekman simple emotions are happiness, sadness, anger, surprise and disgust. Complex emotions are a combination of simple emotions [4]. As introduction to the complexity of emotions, Figure 1 is represented.

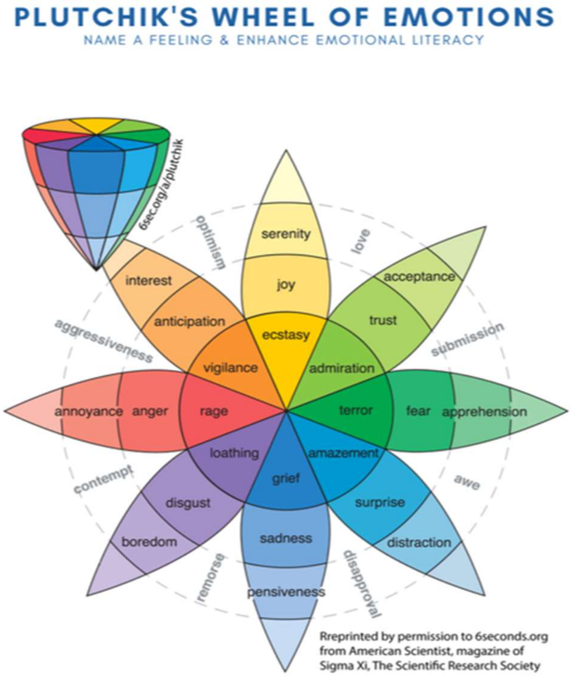


Figure 1. Model of Robert Plutchik [5]

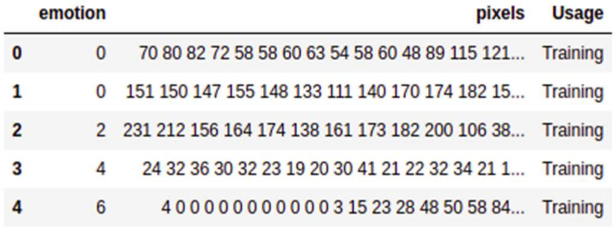
3 Dataset exploration

3.1 Dataset details

The evaluation of facial expression recognition and emotional recognition methods requires the use of one or more databases. The model used in this paper is trained on the FER+ (Facial Expression Recognition PLUS) dataset that was published on the International Conference on Machine Learning (ICML). The FER+ Dataset is an extension of the original FER dataset, where the images have been re-labelled into one of 8 emotion types: neutral, happiness, surprise, sadness, anger, disgust, fear, and contempt shown on Figure 2. The data consists of 48x48 pixel grayscale images of faces. The training set consists of 28,709 examples.

The public test set used for the leaderboard consists of 3,589 examples. The final test set consists of another 3,589 examples. In Table 1 a sample of the datasets are classified.

Table 1. Sample of dataset



	emotion	pixels	Usage
0	0	70 80 82 72 58 58 60 63 54 58 60 48 89 115 121...	Training
1	0	151 150 147 155 148 133 111 140 170 174 182 15...	Training
2	2	231 212 156 164 174 138 161 173 182 200 106 38...	Training
3	4	24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 1...	Training
4	6	4 0 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84...	Training

The images on Figure 2 are represented in way to orthogonally divide each emotion. And the procedure on how to represent these are are:

- Emotion**:** is of numerical type, designates the class of the image (0 to 6)
- Pixels**:** numerical representation of the images
- Usage**:** text that designates the usage of the image (either for training or for evaluation).

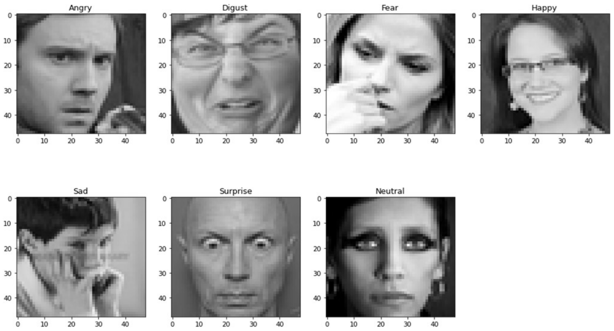


Figure 2. Sample of dataset for seven types of emotion

3.2 Labels distribution and statistics

For a better understanding of the FER+ data, we decided to display the histogram to see the distribution of the emotion classes (Figure 3).

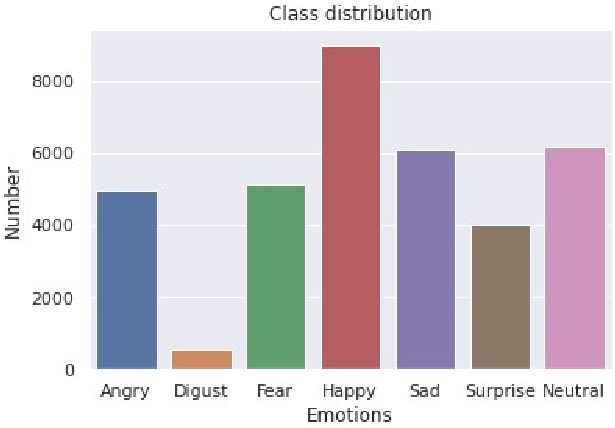
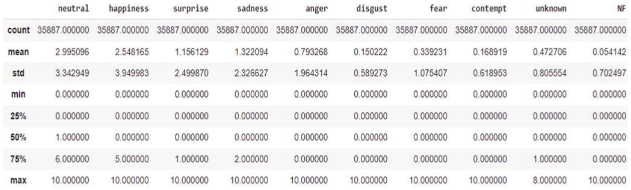


Figure 3. Distribution of the emotion classes on the whole dataset

For statistics detail, we used the command describe of pandas libraries, the results are presented in the Table 2.

Table 2. Statistical information on dataset



	neutral	happiness	surprise	sadness	anger	disgust	fear	contempt	unknown	NF
count	35887.000000	35887.000000	35887.000000	35887.000000	35887.000000	35887.000000	35887.000000	35887.000000	35887.000000	35887.000000
mean	2.995096	2.548165	1.156129	1.322094	0.793268	0.150222	0.339231	0.168919	0.472706	0.054142
std	3.342949	3.949983	2.499870	2.326627	1.964314	0.589273	1.075407	0.618953	0.805554	0.702497
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75%	6.000000	5.000000	1.000000	2.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000
max	10.000000	10.000000	10.000000	10.000000	10.000000	10.000000	10.000000	10.000000	8.000000	10.000000

4 Dataset pre-processing

4.1 Image normalization

The images in the database vary in many parameters which can affect directly on recognition accuracy and performance. These are some difficulties such as rotation, brightness and illumination changes even for the same person's images. To address this problem, a normalization of the face image such as detecting, de-noising and some other preprocessing such as correcting the rotation is performed. The image brightness and contrast variations increase the complexity of the problem.

In image processing, normalization is a process that changes the range of pixel intensity values. Applications include photographs with poor contrast due to glare, for example. Normalization is sometimes called contrast stretching or histogram stretching. In more general fields of data processing, such as digital signal processing, it is referred to as dynamic range expansion [6].

4.2 Image Cropping

The original face images have background information that is not important and could make the output to be less accurate. The cropping region also tries to remove facial parts that do not contribute to the expression.

Cropping is the removal of unwanted outer areas from a photographic or illustrated image [7]. The process usually consists of the removal of some of the peripheral areas of an image to remove extraneous unwanted parts from the picture, to improve its framing, to change the aspect ratio, or to accentuate or isolate the subject matter from its background.

4.3 Training and validation datasets

As mentioned in paragraph 3.1, Table 3 represents the dataset their dispersion, size and distribution to be able to validate our model and have a better accuracy.

Table 3. Datasets description

Dataset	Size
Training	28709
Validation	3589
Testing	3589

5 Organization and structure

The structure of the paper is as follows:

- First, we have training and validation images that we have processed, normalized.
- Train the model and make sure we have a good accuracy, then made predictions with images taken by the webcam in real time.
- Finally analyze the errors and display the performances, restart the training if they are not very efficient.

A better overview of our model in represent on Figure 4.

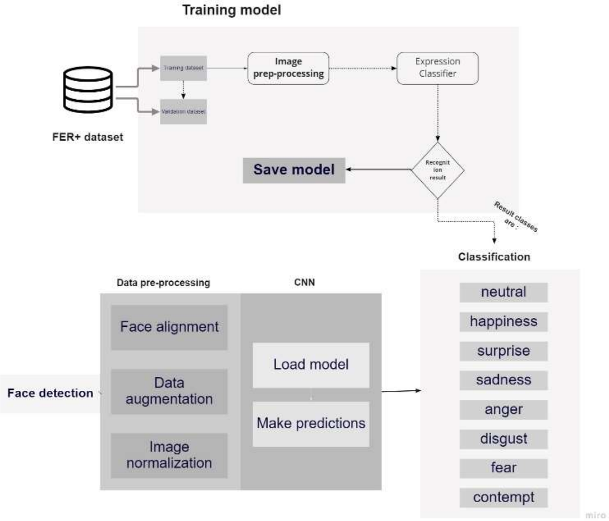


Figure 4. Process description schema

6 Details on structure

A Convolutional Neural Network (ConvNet/CNN) [8] is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. A collection of such fields overlaps to cover the entire visual area. The Figure 5 represents a part of the CNN architecture, and will be explained in detail in the paragraph 6.2.

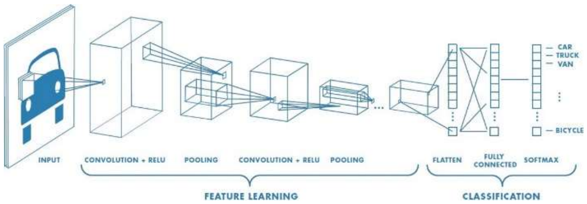


Figure 5. The CNN architecture [9]

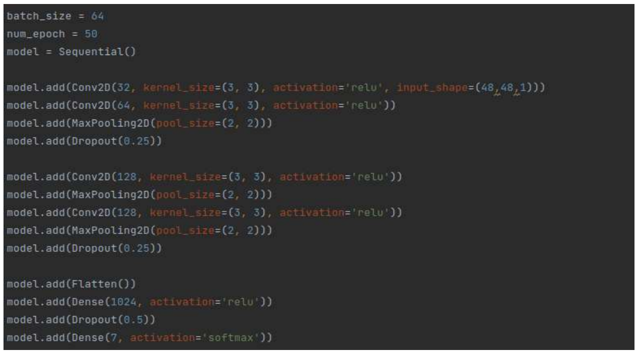
6.1 Technical information of the model

The parameters of the model are described on Table 4, as all the parameters needed to be calibrated and adjusted.

Table 4. Parameters and values description

Parameters	**Values**
Learning rate	0.0001
Epochs	50
Batch size	64
Activation of hidden layer	Relu
Activation of output	Softmax
Optimiser	Adam
Loss function	Categorical_crossentropy
Evaluation metrics	Accuracy

Here are some lines of code to illustrate the model as well as the assessment of its functionality presented in Figure 6. (as well as for replication purposes).



```
batch_size = 64
num_epoch = 50
model = Sequential()

model.add(Conv2D(32, kernel_size=(3, 3), activation='relu', input_shape=(48,48,1)))
model.add(Conv2D(64, kernel_size=(3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))

model.add(Conv2D(128, kernel_size=(3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Conv2D(128, kernel_size=(3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))

model.add(Flatten())
model.add(Dense(1024, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(7, activation='softmax'))
```

Figure 6. Model implementation

6.2 Model explanation

This work consists of using a simple Convolutional Neural Network with linear one-vs-all SVM at the top. The network receives an 48x48 image as an input and then returns the confidence of each expression as an output.

The first layer of the CNN is a convolution layer that applies a convolution kernel of 3x3 and outputs an image. This layer is followed by a subsampling layer that uses Max-pooling with kernel size 2x2. The activation function used is 'Relu' for each convolution as referred to Figure 7.

Once the filters were applied, we passed the resulting vector on two layers:

- A hidden layer with 1024 neurons
- An output layer, with 7 neurons corresponding to each class and softmax function.

The optimization method used is ADAM [10]. Adam combines the concept of momentum with AdaGrad [11].

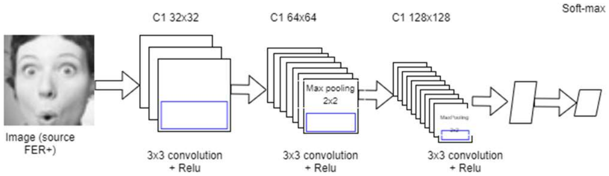


Figure 7. The architecture of the model

6.3 Softmax

In probability theory, the output of the softmax function can be used to represent a categorical distribution – that is, a probability distribution over J different possible outcomes. The softmax function is also known to be used in various multi-class classification methods, for the example in this case of a CNN model which is used to classify the probability of each emotion which refers to Eq. (1).

$$s(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \tag{1}$$

For classification problems using deep learning techniques, it is standard to use the softmax Eq. (1) or 1-of-K encoding at the top. Here we have 7 possible classes, the softmax layer has 7 nodes denoted by $s(x_i)$, where i = 1, . . ., 7. $s(x_i)$ specifies a discrete probability [12].

Dropout is a technique to reduce overfitting during model training. The term "Dropout" refers to the removal of neurons in the layers of a Deep Learning mode [13].

7 Result Analyses

7.1 Result predictions

Here are some results of model prediction on real time images using the camera

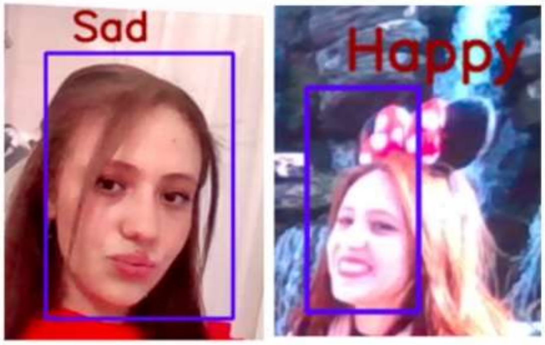


Figure 8. Sample of real time predictions (Author's Original Picture)

The prediction displayed on the Figure 8. is an image taken in real time with a function that uses openCV library to launch webcam, capture the face, apply the prediction and finally take a picture.

7.2 Loss and accuracy analyses

The following Figure 9. explains the evolution of the loss function and the accuracy as a function of the number of epochs.

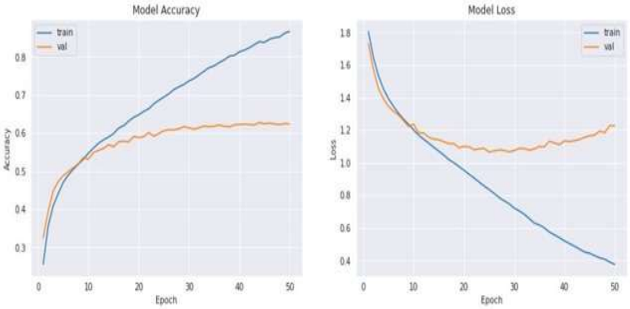


Figure 9. Plot accuracy and loss

The model accuracy on training data is 62% and 85% on testing one.

7.3 Confusion matrix

We want to measure the quality of an automatic classification system of the images represented by the FER+ database [14] as well as the accuracy on a scalable model. Images are classified into seven classes as explained previously. Suppose our classifier is tested. For this, we want to know:

-how many images will be falsely estimated as another emotion (false alarms) by the classifier and

-how many images will be considered as the right emotion by the classifier.

To analyze the performance and predictions of the model, a confusion matrix is displayed on the Figure 10. as a technique for summarizing the performance of a classification algorithm used in this case study.

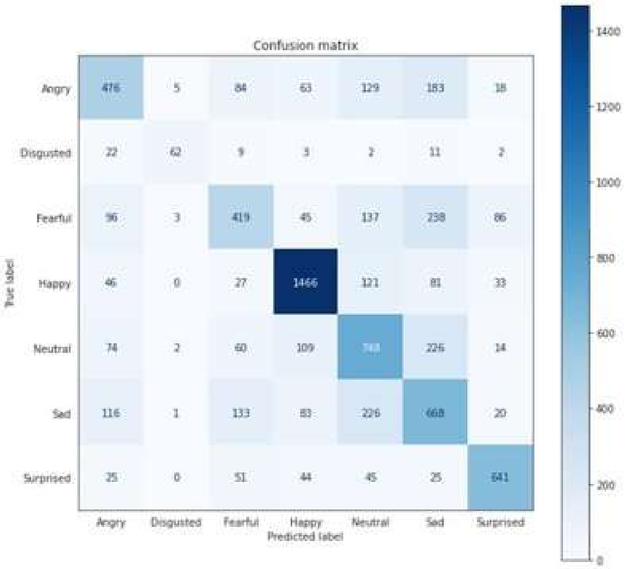


Figure 10. Confusion matrix on testing dataset

8 Conclusion

The proposed paper is designed to develop a real time system to detect, recognize and classify human face emotions. Training a neural network can take a long time, ranging from a few hours to a week, depending on the size of the data source and the complexity of the model.

The limitations of our datasets are most related to the variable settings of the FER+ database. In particular, the assumption that this dataset is large and diverse enough to indicate which model settings are generally preferable. The model used in this paper allowed us to have a good classification of the images according to the emotions on the faces.

The technology of facial expression recognition has enormous market potential and, in the near future, it will enhance most human computer interfaces.

References

[1] W. Swinkels, L. Claesen, F. Xiao and H. Shen, "SVM point based real-time emotion detection," in *2017 IEEE Conference on Dependable and Secure Computing*, Taipei, 2017.

[2] C. Darwin, “The expression of the emotions in man and animals, 3rd edn (ed. Ekman P.),”. London: Harper Collins; New York: Oxford University Press, 1998.

[3] P. Ekman, “An argument for basic emotions” *Cognition and Emotion*, vol. 6, no. 3-4, pp. 169-200, 1992.

[4] P. Ekman, “Are there basic emotions?,” *Psychological Review*, vol. 99, no. 3, pp. 550–553, 1992.

[5] R. Plutchik. “The nature of emotions”, *American Scientist*, vol. 89, no. 4, pp. 344-350, 2001.

[6] R.C. González, R.E Woods, “Digital Image Processing”, Prentice Hall, pp. 85, 2007.

[7] O. Corcoll, “Sementic Image Cropping”, *arXiv:2107.07153*, pp.1-48, 2021.

[8] A. Krizhevsky, I. Sutskever, and G.E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, pp. 1097-1105, 2012.

[9] MD. Zeiler and R. Fergus.,“Visualizing and understanding convolutional networks”, in *European Conference on Computer Vision*, 2014, pp. 818-833.

[10] DP. Kingma, J. Ba. “Adam: A method for stochastic optimization”, *arXiv:1412.6980*, 2014, pp. 1-15.

[11] J. Murphy “An overview of convolutional neural network architectures for deep learning”, in *Microway Inc*, 2016, p. 11

[12] M. Wang, S. Lu, D. Zhu, J. Lin and Z. Wang, "A High-Speed and Low-Complexity Architecture for Softmax Function in Deep Learning," in *2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, 2018, pp. 223-226.

[13] S. Hahn, H. Choi. “Understanding dropout as an optimization trick”, *Neurocomputing*, vol. 398, pp. 64-70, 2020.

[14] E. Barsoum; C. Zhang; C. C. Ferrer; Z. Zhang, “Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution,” *arXiv:1608.01041*, 2016.