Advanced International Journal of Multidisciplinary Research

E-ISSN: XXXX-XXXX editor@aijmr.com

Volume 1, Issue 1, July - August 2023

Facial Expression Recognition with Convolutional Neural Networks

Prof. Prakash Sangle

Assistant Professor, Department of Computer Engineering,
Veermata Jijabai Technological Institute, Mumbai

Abstract

Facial expressions are one form of approach by which people convey their feelings, as they are a potent platform for interaction tool. Recognizing face expressions one of the exciting and effective jobs in public interaction since facial expressions are important in nonverbal interaction. Facial Expression Recognition (FER) is current study topic when it comes to AI, with numerous recent experiments utilizing CNN's. In following study, it shows how to classify FER utilising CNNs and static pictures without performing any feature extraction or pre-processing work. The research also provides examples of pre-processing methods, such as face detection and lighting adjustment, to increase future accuracy in this field. The chin, mouth, eyes, nose, and eyebrows are among the most recognisable face characteristics that are extracted utilising feature extraction. We also talk about the literature review, our CNN design, the difficulties with max-pooling, and how dropout helped us get higher performance. In a classification job with seven classes, we achieved the efficiency of 61.7%in the FER2013 as opposed to the state-of-the-art classification accuracy of 75.2%.

Keywords: Facial Action Coding System (FACS), Convolutional Neural Networks (CNN's), Facial Expression Recognition (FER), Pre-processing, and Features of Extraction.

INTRODUCTION

Since, social interaction involves verbal as well as nonverbal communication, facial emotions are crucial. Example of nonverbal interaction is facial expressions, which communicate important indications of communication including eye contact. Gestures and body language are additional forms of nonverbal communication. Humans are adept at recognising and comprehending faces and facial expressions. Nonetheless, creating an automated system that achieves the same comprehension continues to be challenging response changes in head positions, occlusions or illumination, the retrieval of facial emotion data, the recognition of face characteristics, or classification of emotion are just a few of the issues that surround this subject.

Recognition of facial expressions in AI (FER) is a significant way of study that has uses in a variety of sectors, including advertising, entertainment, e-learning, healthcare, security, enforcement agencies, and social humanoid automations.

Automatic facial expression recognition has applications in a variety of sectors, including business intelligence, psychological research, multiplayer networking, and others that imply human computer connections.

There are some facial expressions that are universally understood. When Ekman and Friesen completed and created the Facial Action Coding System (FACS) in 1978, they discovered that there are six facial

emotions that seem to be shared by all cultures, including pleasure, sorrow, surprise, anxiety, rage, and disgust. These same emotions are presented for categorization in research tasks like Kaggle's Face Expression Recognition challenge, combined with an additional seventh emotion—the neutral emotion. Yet, due of variations in illumination and different head postures, it might be difficult for computers to distinguish these typical human expressions in the real world. According to a latest report released, CNNs successfully used the FER2013 datasets. The history and research will be covered in greater detail in the section on related projects. In order to accurately categorize photos of face images into distinct expressions with CNN models, the goal of this article is to create a unique architecture from scratch. It also demonstrates extraction of features and pre-processing techniques.

Literature Review

In the late nineteenth century, A French neurosurgeon Guillaume-Benjamin-Amand Duchenne de Boulogne, had an interest in physiology and sought to comprehend how the face muscles produced facial emotions since he felt that they were inextricably tied to a person's soul. In order to do this, he employed an electrical probe to cause muscular spasms. He then used recently invented photographic technologies to capture images of his patients' faces, capturing the twisted emotions he had managed to produce. He presented his findings in journal "The Process of Human Physiognomy", together with images of the induced facial expressions, in 1862. Photos of his subjects with a distinct look on both the side of their faces are shown in Figure. 1, which serves as an illustration from his publication. Charles Darwin later utilised this paper as a significant source for the publication "The Appearance of Emotion in Man and Creatures," published in 1872 and focused on the biology of behavioral. Photographers have recently rediscovered Duchenne de Boulogne's book as a significant piece of photography fine arts. Nonetheless, as was mentioned in the introduction, Ekman is without a doubt one of this century's most important scholars in the area of emotional expression.

Figure.1. The Method of Human Physiognomy.

Convolutional Neural Networks (CNNs) were used in 2017 by Kampeland Pramerdorfer to achieve state-of-the-art, which is 75.25% accuracy in FEB2014. Using parameters of 1.8 meters, 1.6 meters, and 5.3 meters, the authors constructed on a collaborative of CNNs utilizing VGG, and ResNet with depths of 10, 16, and 33. The authors utilised the face photos as provided in the dataset using histogram equalization for lighting adjustment. For the purpose of augmenting the training data, they applied horizontal mirroring and randomly compressed the photos to a 48 × 48 pixel size. Moreover, they employed stochastic descent incline to minimize the cross-entropy defeat with an energy cost of 0.9 while training the design for up to 300 epochs. Other settings such as activation functions of 0.1, batch

size of 128, and gradient descent of 0.0001 were fixed.

Using the difficult Kaggle face expression dataset, Zhang et al. employed a Siamese Network to provide a technique for deciphering social connection behaviours from photos and attained a test accuracy of 75.1%. The researchers designed an extraction of features technique, patch-based registration, and focused on feature integration via earlier fusion in addition to using numerous datasets with different labels to expand the training data.

Kim et al. suggested an assembly of Cnn models and showed that using both registered as well as unregistered versions of provided face photos during training and testing is beneficial. Also on FER2013 data - set, the authors attained an accuracy rate of 73.73%. Also, they carried out illumination standardization and Interface for traditional 2-D association, both of which are openly accessible for landmark detectors. Based on consequences of facial significant identification, they executed registration strategically in order to prevent the registered mistake.

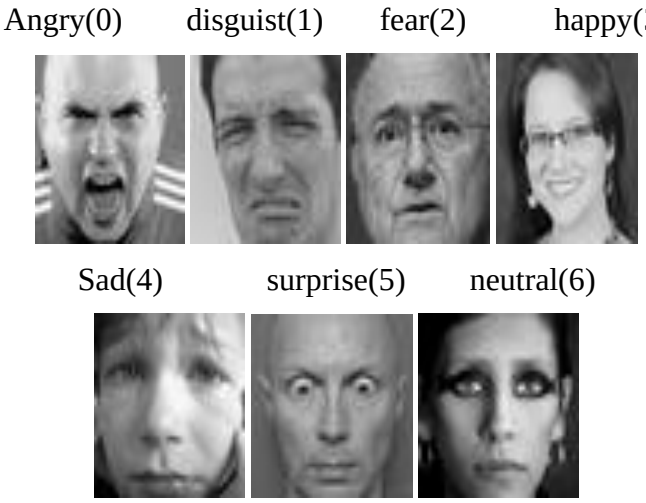


Fig.2.Example images from the FER2013 dataset with labels.

Using Kaggle's Face Emotion Recognition Contest datasets, Raghuvanshi with Choksi experimented with several topologies including Convolutional Networks and techniques like fractional max-pooling and fine-tuning, finally attaining an efficiency of 48%.

DATASET

The International Conference on Machine Learning (ICML) 2014 Workshop on Problems in Classification Tasks featured a presentation of FER2014 datasets. On Kaggle's FER Initiative, a sizable dataset known as FER2014 is available to the general audience. The FER2013 dataset includes 28,709 training, 3,589 validations, and 3,589 testing photos, totaling 35,887 facial crops. Figure. 2 shows pictures of seven distinct expressions along with the labels that go with them (0 represents anger, 1 represents disgust, 2 represents fear, 3 represents happiness, 4 represents sorrow, 5 represents surprise, and 6 represents neutral). Each image has a grayscale resolution of 48 by 48 pixels. The human accuracy of such datasets, according to Ian Goodfellow, is approximately 65.5%.

TECHNICALWORK

We will go through our Convolutional architecture and methods in this part to further enhance its precision on FER2013. The project is divided into three parts, as described in the following:

A. *Pre-processing*

Prior to the featured extraction procedure, pre-processing can be utilised to improve FER performance of the system. Pre-processing an image entails a number of steps, including the recognition and aligning of images, lighting, posture, opacity, and data preprocessing correction.

The images in the FER2013 datasets are automatically recorded such as having comparable required specifications and are roughly centred in the photos. Using the Haar Cascade classifier, facial recognition is shown in Figure. 3.

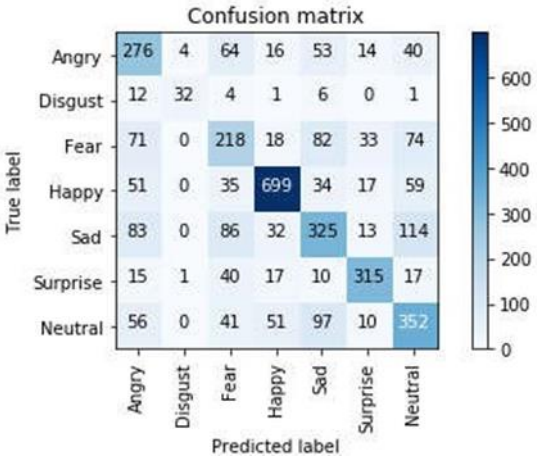
In an attempt to improve efficiency as in future, this research tests with facial recognition to record the majority of the facial and apply various algorithms to the face picture. To create a more reliable approach which can solve the transparency, lighting, and head posture difficulties, we are further improving the facial recognition technique.

The emotion identification rate may be poor and extraction of features may be more challenging when photographs are taken in different types of lighting and emotion characteristics are occasionally incorrectly recognised. The centre picture in Figure. 3 serves as an example of lighting adjustment achieved using contrast enhancement.

B. Feature Extraction

The source data must be converted into a collection of characteristics in order to extract face characteristics. Researchers can use extraction of features to condense enormous amounts of data into a manageable collection, enabling speedier computing. The 8 most noticeable facial features were retrieved using the dlib facial feature classifier pre-trained on the iBUG 300-W dataset [12], [13], and [14]. These features also included eyebrows, both eyes, the nose, the internal as well as outside contours of the mouths, and the jawline. The final image in Figure. 3 serves as an example of extraction of features. Yellow-hued features were retrieved from this image, including the right and left eyes, nose, both the inner as well as outer contours of the mouth.

C. CNN Architecture



FER is just one of the many applications for visual analysis that frequently use CNNs. Early in twenty-first century, multiple reviews of FER research found that Convolutional networks are effective for both changes in size and changes in face location. Additionally, they were discovered to score higher than multilayer perceptron (MLP) while analysing novel variations in face pose. Many facial expression recognition issues, including translations, rotations, subject independent, and scale partially invariant, were resolved by researchers using CNN.

Developing an optimum architectural and topology is a highly difficult undertaking, as seen in Figure. 4. The following traits were used to train our model:

- Using "RELU" as the activation function, there are 6 convolutional layers;
- 3 max-pooling, each of which is followed by convolution layers, with the first two using diameter (3,3) and step size (2,2) and the third using pool dimensions (2,2) and step size (2,2).

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 46, 46, 128)	1280
conv2d_2 (Conv2D)	(None, 44, 44, 128)	147584
max_pooling2d_1 (MaxPooling2	(None, 21, 21, 128)	0
conv2d_3 (Conv2D)	(None, 19, 19, 128)	147584
conv2d_4 (Conv2D)	(None, 17, 17, 128)	147584
max_pooling2d_2 (MaxPooling2	(None, 8, 8, 128)	0
dropout_1 (Dropout)	(None, 8, 8, 128)	0
conv2d_5 (Conv2D)	(None, 6, 6, 128)	147584
conv2d_6 (Conv2D)	(None, 4, 4, 128)	147584
max_pooling2d_3 (MaxPooling2	(None, 2, 2, 128)	0
flatten_1 (Flatten)	(None, 512)	0
dense_1 (Dense)	(None, 1024)	525312
dropout_2 (Dropout)	(None, 1024)	0
dense_2 (Dense)	(None, 7)	7175

Total params: 1,271,687
Trainable params: 1,271,687
Non-trainable params: 0

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 46, 46, 128)	1280
conv2d_2 (Conv2D)	(None, 44, 44, 128)	147584
max_pooling2d_1 (MaxPooling2	(None, 21, 21, 128)	0
conv2d_3 (Conv2D)	(None, 19, 19, 128)	147584
conv2d_4 (Conv2D)	(None, 17, 17, 128)	147584
max_pooling2d_2 (MaxPooling2	(None, 8, 8, 128)	0
dropout_1 (Dropout)	(None, 8, 8, 128)	0
conv2d_5 (Conv2D)	(None, 6, 6, 128)	147584
conv2d_6 (Conv2D)	(None, 4, 4, 128)	147584
max_pooling2d_3 (MaxPooling2	(None, 2, 2, 128)	0
flatten_1 (Flatten)	(None, 512)	0
dense_1 (Dense)	(None, 1024)	525312
dropout_2 (Dropout)	(None, 1024)	0
dense_2 (Dense)	(None, 7)	7175

Total params: 1,271,687
Trainable params: 1,271,687
Non-trainable params: 0

Fig.4.Illustration of CNN Architecture

- 2 drop over with a 0.2 grade.;
- two dense layers, one of which is dense and has the activation function "RELU," and the other of which is dense and has the activation function "Softmax.";
- There are 1.2 million total variables and generative model characteristics.

CONCLUSION

Without using any pre-processing and feature extraction techniques, the 6 Convolutional layer architecture used in this work functioned admirably and attained a FER2014 test precision of 61.69%.

For the 7 distinct emotional expressions, the ensembles of convolutional neural networks' state-of-the-art experimental results is 75.2%. We ran a number of trials with various batch sizes as well as epochs, but the greatest test efficiency was achieved with 512 and 10 as the batch size and epoch values.

The result obtained and weight training have been saved through HDF5 layout. All classification techniques were installed in Keras which use Tensorflow as a server side, as well as instructed on a personal computer config of CPU I7 x64, 32 Gb Of ram, as well as dedicated Eight GigaByte NVIDIA Graphics GTX 1080.

The clustering algorithm for the testing data is shown in Figure. 5, in which rows are true labels as well as columns are predicted labels. Mostly in testing dataset, there seem to be 467 "angry" circumstances, and we accurately predicted 276 of them. One more illustration is where we accurately predicted 699 out of the 895 "happy" incidences.

In accordance with the training data mentioned in Figure 2, Figure 6 displays appropriately anticipated emotions. The y-axis on the expression detection bar graph shows percentages, as well as the x-axis shows emotions. The percentage of each emotion is displayed on the classification bar.
We examined the classification models, and found that, with accuracy rates of 43.95% and 49.77%, accordingly, the responses of fear and sadness were the most often misclassified pictures. Figure. 7 displays the incorrectly labelled photos together with their actual and anticipated labels. For an example, one may use the third image, whose true label is ‘sad,’ however the algorithm predicted fear, sad, & furious by 43%, 40%, & 15% percent efficiency, respectively, as well as disgust, surprised, and neutrality at 2% reliability overall.

DISCUSSION AND FUTURE WORK

Without using any pre-processing or extraction of features approaches, its individual 6 Convolutional layer design functioned well and achieved a FER2013 accuracy rate of 61.7% in a seven-class classified job.

We describe the difficulties and potential future research in this part in order to further enhance accuracy rate mostly on FER2013 datasets. The optimum network design might be difficult to find in supervised learning. As shown in Figure. 8, we found Cnn model using a heuristic algorithm and will continue to search for a more powerful network in the upcoming years. To increase accuracy, we will also use pre-processing and features of extraction methods that were covered in the scientific work part. Also, because the training data had a 99.64% high accuracy, we encountered an overfitting problem. Data augmentation is therefore an essential stage in deep FER. Generally, a deep learning toolkit includes data augmentation to address the overfitting problem.

REFERENCES

1. Hossain, Sanoar, et al. "Fine-grained image analysis for facial expression recognition using deep convolutional neural networks with bilinear pooling." Applied Soft Computing 134 (2023): 109997.
2. Sarvakar, Ketan, et al. "Facial emotion recognition using convolutional neural networks." Materials Today: Proceedings 80 (2023): 3560-3564.
3. Shahid, Ali Raza, and Hong Yan. "SqueezExpNet: Dual-stage convolutional neural network for accurate facial expression recognition with attention mechanism." Knowledge-Based Systems 269 (2023): 110451.
4. Srivignesh, R., and A. Anand. "Facial Expression Recognition using Convolutional Neural Network and Haar Classifier." 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF). IEEE, 2023.
5. Gautam, Chahak, and K. R. Seeja. "Facial emotion recognition using Handcrafted features and CNN." Procedia Computer Science 218 (2023): 1295-1303.

6. Pan, Jiahui, et al. "Multimodal emotion recognition based on facial expressions, speech, and EEG." IEEE Open Journal of Engineering in Medicine and Biology (2023).
7. Kim, Jieun, and Deokwoo Lee. "Facial expression recognition robust to occlusion and to intra-similarity problem using relevant subsampling." Sensors 23.5 (2023): 2619.
8. R. G. Harper, A. N. Wiens, and J. D. Matarazzo, Nonverbal communication: the state of the art. New York: Wiley, 1978.
9. A.Savva,V. Stylianou, K. Kyriacou, and F. Domenach, “Recognizing student facial expressions: A web application,” in 2018 IEEE Global Engineering Education Conference (EDUCON), Tenerife, 2018, p. 1459-1462.
10. https://en.wikipedia.org/wiki/Paul_Ekman,Jan2020.
11. J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, “The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions,” IEEE Transactions on Affective Computing, vol. 5, no 1, p. 86-98, janv. 2014.
12. U. Ayvaz, H. Gürüler, and M. O. Devrim, “USE OF FACIAL EMOTION RECOGNITION IN E-LEARNING SYSTEMS,” Information Technologies and Learning Tools, vol. 60, no 4, p. 95, sept. 2017.
13. Febrian, Rio, et al. "Facial expression recognition using bidirectional LSTM-CNN." Procedia Computer Science 216 (2023): 39-47.