FACIAL EMOTION RECOGNITION

Aditi Shrivastava

Abstract

Applications for Facial Emotional Recognition (FER) range from marketing and healthcare to human-computer interaction. It is an important task. This paper suggests a computerized facial emotion recognition model using convolutional neural networks (CNNs). In order to acquire distinguishing characteristics from face photos and categorize them into several emotion categories, the model makes use of CNN's hierarchical feature learning capabilities. Using supervised learning techniques, the CNN model is developed on an extensive set of facial photos tagged with appropriate emotion labels.

Dataset

The dataset utilized for facial emotion recognition is obtained from Kaggle, encompassing distinct sets for training and testing. These sets encompass seven emotion categories: surprise, anger, fear, disgust, sadness, happiness, and neutrality. Each emotional category is underpinned by a comprehensive dataset, facilitating robust training and validation.

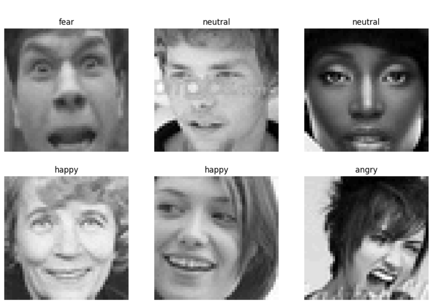

Fig1. Dataset for FER.

The training dataset comprises a total of 28,000 images, each with a pixel size of 48x48 which are almost evenly distributed. The testing dataset is about 7000 images.

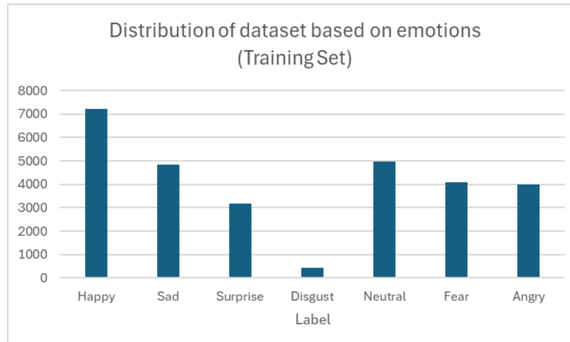

Fig2. Distribution for Training Set

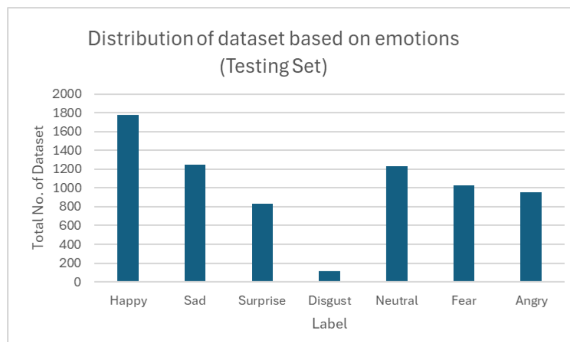

Fig3. Distribution for Testing Set

Proposed Model:

Modules like pandas, tensorflow, numpy, seaborn,and keras were used to perform this study.

Convolutional neural networks, or CNNs for short, are a class of deep learning algorithms that are frequently applied to image processing and recognition applications. It takes inspiration from the receptive field, a small area of the visual field in which individual neurons in the animal visual cortex respond to stimuli. Convolutional, pooling, and fully connected layers are among the layers that CNNs have in order to enable their algorithms to effortlessly and adaptively identify spatially organized hierarchies of attributes from input images. They are extensively employed in many different applications, including facial recognition, object identification, and image categorization.

There are four convolution layers used in this project, and the number of filters in each 2D Convolution layer keeps increasing. The first 2D Convolutional Layers has 128 filters, the second one has 256 filters, the third and fourth layer have 512 filters. The activation used is **"gelu".** MaxPooling2D is performed for each convolutional layer with a pool of 2x2. Max-pooling helps to minimize computation and control overfitting by decreasing the dimension of the input volume.

A popular non-linear activation function in deep learning networks is the Gaussian Error Linear Unit, or Gelu, activation function. Gelu is a proposed activation function that has various benefits for neural network training over alternatives to activation functions as ReLU (Rectified Linear Unit) as well as its variations. Gelu gives the neural network non-linearity, which helps it recognize intricate patterns and correlations in the data. Gelu features a curve that is smooth, which can result in steadier gradients throughout training, in contrast to ReLU, which has jagged transitions at zero. For small values of x, Gelu performs like an estimate to the identity function, which helps to maintain the flow of data through the system and mitigate the vanishing gradient issue also, Gelu provides an easy-to-implement and calculate function that works well with a variety of deep learning architectures.

In order to avoid overfitting, dropout is a regularization approach that randomly changes a portion of input units to zero during training. Here, each convolutional layer is followed by a dropout rate of 0.4.

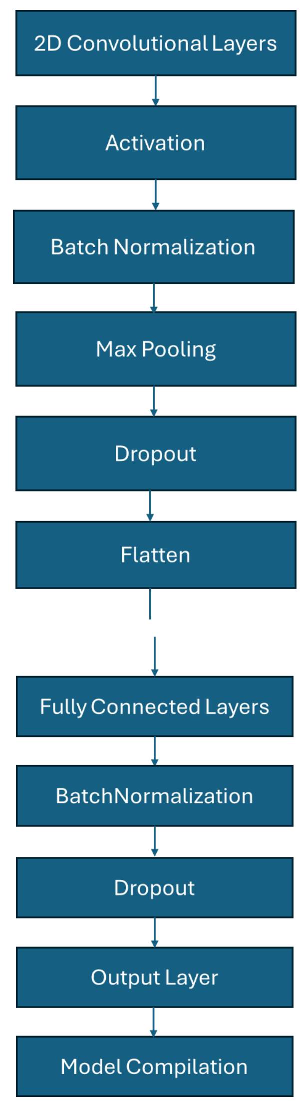

Fig4. Flow Chart

The input volume is flattened into a vector with one dimension by the flatten layer. In order to feed it into the fully connected layers, it transforms the multi-dimensional outcome of the pooling and convolution layers into a vector with one dimension.

Then comes two fully connected layers with 256 and 512 layers, with activation **“gelu” .** The raw scores are transformed into probabilities using the **'softmax'** activation, which represents the likelihood distributions across the resultant classes.

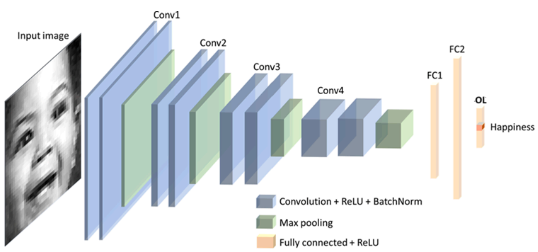

Fig5. Visual Diagram for the Flow

A mathematical function called Softmax can be used to transform a vector of any real value with a probability distribution. When performing multi-class classification problems, it is frequently employed as the final activation function within a neural network. The softmax function produces a vector as its result, each member of which reflects the likelihood of the associated class and which adds up to 1. Softmax is appropriate for classification tasks since it guarantees that the output probabilities add up to 1.

Results

After 100 training epochs, the output shows that the model obtained a training accuracy of roughly 98.2% and a validation accuracy of approximately 50%.

With help of BatchNormalization and activation of GeLU is used for the accuracy of the training set. The accuracy of the training set improved to ~ 98% for 100 epochs.

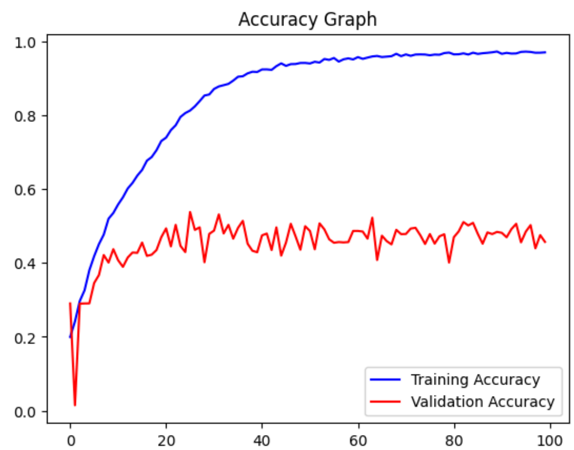

Fig6. Accuracy graph of training and validation

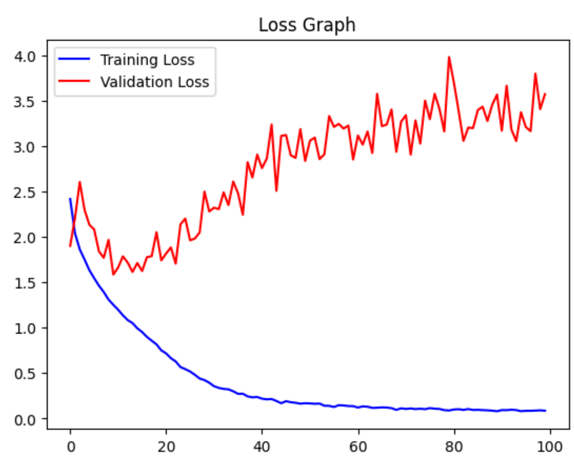

Fig7. Loss graph of training and validation

Conclusion

In this project, facial expression recognition models are developed using the Convolutional Neural Network. The dataset is taken from Kaggle which has seven different expressions, angry, sad, fear, disgust, happy, surprise and neural. After 50 epochs, the training accuracy observed is 96.82% and validating accuracy is 51.29%.

For the future, the work would be on overcoming the overfitting problem which was caused in this project. Larger and more varied datasets may improve the model's capacity to identify emotions across a range of facial expressions, lighting scenarios, and demographics.

Without BatchNormalization, the overfitting issue is resolved but the accuracy of both the training set and validating set is compromised. It could be possible to increase accuracy and robustness by experimenting with cutting-edge deep learning architectures like models built on transformers or attention methods and also. learn to operate and inference times can be greatly accelerated by optimizing the code and making full use of GPU acceleration. This enables rapid exploration and model iteration.

Investigating methods for deciphering and displaying model predictions can boost openness and confidence in the system, aiding in the discovery of biases and enhancing the face emotion detection system's overall functionality and also, study on more complex models and datasets by speeding up model training and testing by utilizing cloud-based systems for distributed computation and training on strong GPU clusters. One can be equipped to make wise judgments, streamline processes, and solve problems in an efficient manner through the project lifecycle by receiving comprehensive education on deep learning technology and best practices.

Applications that need low latency and high throughput can be made possible by looking into ways of improving the inference pipeline for processing in real time, which includes modeling, quantization, pruning, or deploying on edge devices.

Reference:

[1] E. Sariyanidi, H. Gunes, et A. Cavallaro, Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition

[2] C.-N. Anagnostopoulos, T. Iliou, I. Giannoukos, Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011

[3] Wafa Mellouk a, Wahida Handouzi a,Facial emotion recognition using deep learning: review and insights

[4] T. Jabid, M.H. Kabir, O. Chae,Robust Facial Expression Recognition Based on Local Directional Pattern

[5] S. Zhang, L. Li, et Z. Zhao, Facial expression recognition based on Gabor wavelets and sparse representation

[6] G. Zhao, X. Huang, M. Taini, S.Z. Li, M. Pietikäinen, Facial expression recognition from near-infrared videos

[7] M. Pantic, L.J.M. Rothkrantz, Toward an affect-sensitive multimodal human-computer interaction

[8] B. Fasel and J. Luettin. Automatic facial expression analysis: a survey pattern recognition
[9] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern
Recognition
[10] Shuangyu Wei, Paige Tien, Wuxia Zhang, Zhichen Wei, Zu Wang, John Kaiser Calautit,DeepVision based detection for energy-efficiency and indoor air quality enhancement in highly polluted spaces
[11] Facial Emotion Recognition using Keras Tensorflow, Deep Learning, Python by
Hackers Realm Kaggle, Competion