Emotion Recognition Through Facial Expression: A Convolutional Neural Network Approach

Siddharth Mohanty
BITS F312 NNFL
BITS Pilani K K Birla Goa Campus
f20170436@goa.bits-pilani.ac.in

Saurabh Vijay Surana
BITS F312 NNFL
BITS Pilani K K Birla Goa Campus
f20170450@goa.bits-pilani.ac.in

***Abstract*—In this article, we have developed a CNN model to recognises human emotions through facial expression recognition. Our model works in case of static images of human subjects in grey-scale color scheme. The model works well enough with low resolution images of the subjects faces. This is possible as the classification of emotions is based mostly on the inner facial features like the curvature of the eyes and lips, which when simplified is the case of edge detection in an image. Human emotion when not suppressed can accurately be worked out through the permutations and combinations of the eyes and the mouth.In between the layers of the network a simple technique of regularization called "dropout" is used to check the over-fitting of the model. The goal is to classify each facial image into one of the 7 categories(Anger, Disgust, Fear, Happy, Sad, Surprise, Neutral).**

***Index Terms*—CNN, emotion recognition**

I. Introduction

Emotion recognition is a growing field with wide implication in several fields like entertainment, education, e-commerce, health, and security. Majorly the field of robotics which shall be the foundation of humanoids in the near-future, will have better abilities and functionalities with emotion recognition technology, making them more human-like. Additionally, the field of psychology also has to benefit from such technology in case of people with depression who have difficulty processing emotions of others. The data-set given to us consists of 48x48 pixels of gray-scale images with front view of human subject faces. The faces are optimally centered and occupies about the same amount of space in each image. There are seven categories of emotions present in the data-set namely Anger, Disgust, Fear, Happy, Sad, Surprise, Neutral.

II. Data Pre-processing

The first column of the data-set consists of label from 0-6 representing 7 different emotions(0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The second column of the data-set consists of pixels in string format for every image. For every image the string was reshaped to 48x48 matrix and scaled down by a factor of 255. The data-set is split into training set and validation set (75%- 25%). The label data-set is converted to one-hot into 7 classes.

III. Model Architecture

For this purpose of this project, we have built a CNN model which similar to VGG16 [SZ15] architecture. This network has 6 convolutional layers and 2 FC layers along zero padding, max pooling and dropout layers. The first 3 convolutional layers consist of 64 3x3 filters with the activation function ReLU followed by a max pooling layer of size 2x2 and Zero-Padding of size 1x1. The next 3 convolutional layers consist of 128 3x3 filters with the activation function ReLU followed by a max pooling layer of size 2x2. The model is then flattened and connected to FC layer. The 2 FC layers have 1024 neurons with ReLU as the activation function. The output layer has 7 neurons with softmax as the activation function. The model is compiled with categorical cross entropy as the loss function and SGD as the optimizer. L1 and L2 regularization and dropout layers have been added to prevent overfitting.

IV. Results

The model is trained for 100 epochs and batch size 32. After 100 epochs, accuracy on the training set was 98.47% and 57.92% on the validation set. Precision is 0.6150 and Recall is 0.5474 on validation set.

References

[SZ15] Karen Simonyan and Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. arXiv: 1409.1556 [cs.CV].