**CNN based Facial Emotion Recognition System**

**Submitted by**

Nizasujitha B

**Bharat Electronics Limited**

(A Govt. of India Enterprise, Ministry of Defence)

Jalahalli Post, Bengaluru – 560 013, India

**30.08.2022**

**INTRODUCTION:**

Emotion recognition includes recognition of facial expressions from video, spoken expressions from audio, written expressions from text, and physiology as measured by wearables. It is useful and important for security and healthcare purposes. Also, it is crucial for easy and simple detection of human feelings at a specific moment without actually asking them.

**DEEP LEARNING**

Deep learning is a type of supervised machine learning in which a model learns to perform classification tasks directly from images, text, or sound. Deep learning is usually implemented using a neural network. The term “deep” refers to the number of layers in the network—the more layers, the deeper the network

**CNN:**

A convolutional neural network can have hundreds of layers and each layer learn to detect different features of an image. Filters are applied to each training image at different resolutions and size, and the output of each convolved image is used as the input to the next layer. The filters can start as very simple features, such as brightness and edges, and later on it goes deep to extract complex features. Like other neural networks, a CNN is composed of an input layer, an output layer, and many hidden layers in between.

***Image input layer:***

The input layer is the input of the whole CNN. In the neural network of image processing, it generally represents the pixel matrix of the image.

***Convolutional layer:***

Convolutional layers are the major building blocks used in CNN. A convolution layer transforms the input image in order to extract features from it. In this transformation, the image is convolved with a kernel (or filter). A kernel is a small matrix, with its height and width smaller than the image to be convolved. The convolved features of the input image are stacked one after the other to create an output.

***Wout =*** (1)

*Wout x Wout x Dout* is the size of the output image from the Convolutional layer, which depends on the input image size *W1 x W1 x D1*, size of the filter present in the convolutional layer, amount of zero-padding and stride size . For output image size to be equal to the input image size, the following condition is required:

***;*** (2)

***Dout = 1*** (3)

where, *1* is the number of filters in the convolutional layer.

***Batch Normalization Layer:***

Batch normalization (also known as batch norm) is a method used to make artificial neural networks faster and more stable through normalization of the layers' inputs by re-centring and re-scaling

***ReLU layer:***

ReLU is used for filtering information that propagates forward through the network. It takes an elementwise operation on your input and basically if your input is negative, it's going to put it to zero.

***Pooling layer:***

The main objective of this layer is to scale down the magnitude of the convolved feature map in order to minimize the amount of network computation and the number of parameters that must be learned. This is carried out by dropping the contacts among layers and individually works on respective feature map. In a feature map created by a convolution layer, the pooling layer encapsulates the features that are present in a specific area.

*Max pooling:* The pixel with the highest value is chosen to be sent to the output array.

*Average pooling:* For the output array, it determines the average value inside the receptive field.

***Wout =*** (4)

***Dout = D1*** (5)

*Wout x Wout x Dout* is the output image size of the Max Pooling layer, *W1 x W1 x D1* is the size of the image given as input to the Max pooling layer, is the size of the filter and .

Max pooling is a pooling operation that selects the maximum element from the region of the feature map covered by the filter. Thus, the output after max-pooling layer would be a feature map containing the most prominent features of the previous feature map.

***Fully Connected layers:***

Fully Connected layers in a neural network are those layers where all the inputs from one layer are connected to every activation unit of the next layer. In most popular machine learning models, the last few layers are full connected layers which compiles the data extracted by previous layers to form the final output.

***SoftMax layer:***

SoftMax is implemented through a neural network layer just before the output layer. It converts the output of the last layer in your neural network into what is essentially a probability distribution.

***Classification layer:***

A classification layer computes the cross-entropy loss for classification and weighted classification tasks with mutually exclusive classes. The layer infers the number of classes from the output size of the previous layer.

**EMOTION RECOGNITION:**

Detection of facial expressions from photos and videos, vocal expressions from sounds and audios, communication context from texts, and wearable-measured physiology are all included in emotion recognition. Facial emotion recognition is a technology that uses advanced image processing to interpret the emotions on a person's face. The dataset FER2013 consists of 48 x 48 pixels grayscale images of faces. The faces have been placed automatically so that they are roughly centered and take up a similar amount of space in each image. Each face is assigned to one of seven categories, namely:

* Anger
* Disgust
* Fear
* Happiness
* Sadness
* Surprise
* Neutrality

Fig. 1 Sample images for the seven emotions

The facial emotion is recognized using the CNN architecture here. Batch normalization layer and Rectified Linear Unit (ReLU) are used after the convolutional layer because the batch normalization layer accelerates and stabilizes the network by re-centring and re-scaling the inputs of the layers and the ReLU filters the information that flows through the network. If the input is negative, ReLU simply makes it zero by performing an element-wise operation on it.

Adam is an optimization algorithm that can be used instead of the classical stochastic gradient descent procedure to update network weights iterative based in training data. This algorithm is used to accelerate the gradient descent algorithm by taking into consideration the 'exponentially weighted average' of the gradients. Using averages makes the algorithm converge towards the minima in a faster pace.

**RESULTS:**

The validation accuracy of the emotion recognition system is obtained as 84 %.

* Solver name: Adam
* Initial Learning Rate: 0.0001
* Learning rate schedule: Constant
* Maximum Epochs: 30
* Dataset: FER2013

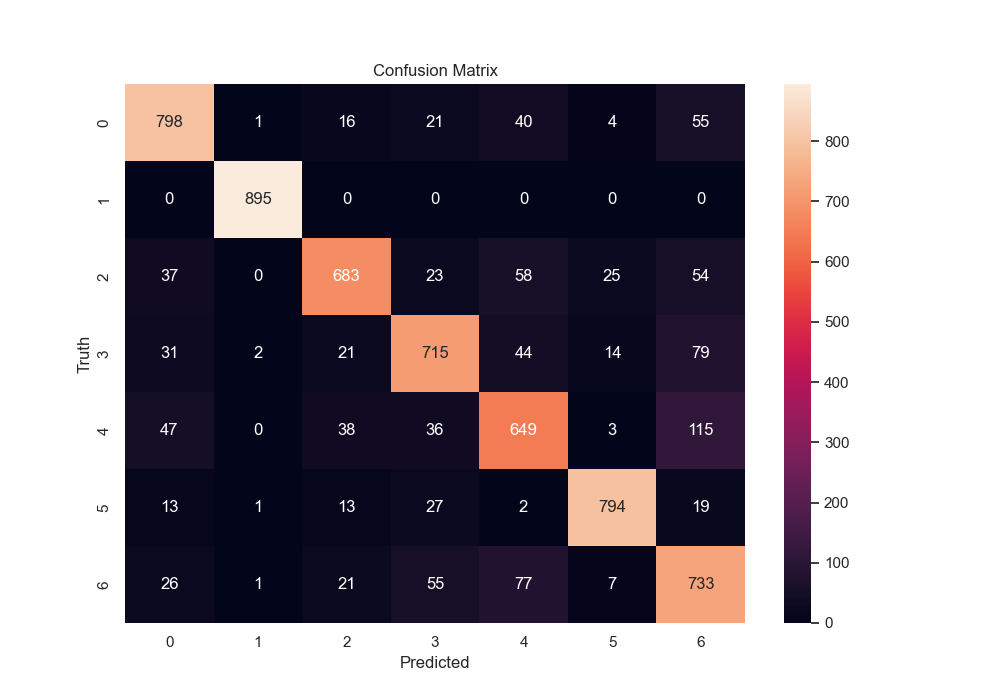
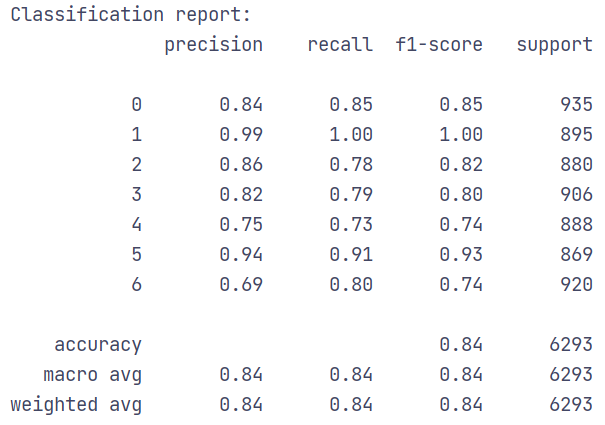
Fig. 2 Confusion Matrix

Fig. 3 Classification Report