**FACE EMOTION RECOGNITION**

**(Live Class Monitoring)**

**Sampanna Mishra**

**Data Science Trainee**

**AlmaBetter,Bangalore**

**Abstract:**

Emotion is a mental state associated with nervous system associated with feeling , perception,behavioral reactions and a degree of gratification or displeasure.Facial expressions are a form of nonverbal communication.One of the current application of artificial intelligence using neural networks is the recognition of faces in images and video for various applications.Face detection technology can come in handy in various practical life examples.In this paper I have strived to represent a method for identifying seven emotions such as anger, disgust, fear, happy,surprise ,sad and neutral using deep learning techniques.

***Keywords: emotion, facial, artificial intelligence, deep learning,recognition***

**1.Problem Statement**

The Indian education landscape has been undergoing rapid changes for the past 10 years owing to the advancement of web-based learning services, specifically, eLearning platforms.

Global E-learning is estimated to witness an 8X over the next 5 years to reach USD 2B in 2021. India is expected to grow with a CAGR of 44% crossing the 10M users mark in 2021. Although the market is growing on a rapid scale, there are major challenges associated with digital learning when compared with brick and mortar classrooms. One of many challenges is how to ensure quality learning for students.

Digital platforms might overpower physical classrooms in terms of content quality but when it comes to understanding whether students are able to grasp the content in a live class scenario is yet an open-end challenge.

In a physical classroom during a lecture the teacher can see the faces and assess the emotion of the class and tune their lecture accordingly, whether he is going fast or slow. He can identify students who need special attention. Digital classrooms are conducted via video telephony software program (exZoom) where it’s not possible for medium scale class (25-50) to see all students and assess the mood.

Because of this drawback, students are not focusing on content due to lack of surveillance.While digital platforms have limitations in terms of physical surveillance but it comes with the power of data and machines which can work for you. It provides data in the form of video, audio, and texts which can be analyzed using deep learning algorithms. Deep learning backed system not only solve the surveillance issue, but it also removes the human bias from the system, and all information is no longer in the teacher’s brain rather translated in numbers that can be analyzed and tracked.

We will solve the above-mentioned challenge by applying deep learning algorithms to live video data. The solution to this problem is by recognizing facial emotions.

**2.Introduction**

Emotion recognition is a technique that allows reading the emotions on a human face using advanced image processing.Emotion recognition is one of the many facial recognition technologies that have developed and grown through the years. Currently, facial emotion recognition software is used to allow a certain program to examine and process the expressions on a human’s face.

What is the use and why is it important?

Use of technology to help people with emotion recognition is a relatively nascent research area. Facial expressions are a form of nonverbal communication. Various studies have been done for the classification of these facial expressions. There is strong evidence for the universal facial expressions of seven emotions which include: neutral, happy, sadness, anger, disgust, fear, and surprise. So it is very important to detect these emotions on the face as it has wide applications in the field of Computer Vision and Artificial Intelligence. These fields are researching facial emotions to get the sentiments of the humans automatically.

**3.Data Summary**

The model is trained on the FER-2013 dataset downloaded from Kaggle.This dataset consists of 48x48 pixel grayscale images of faces. The training set consists of 28,709 images and the test set consists of 7,178 images.The dataset consists of 7 different facial expressions as follows:

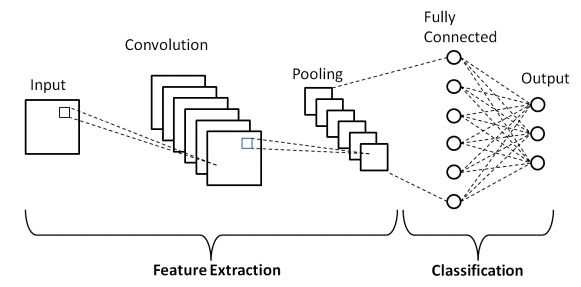


**4.Model Building**

The Convolution Neural Network (CNN) model has been used for processing the training images. Customed CNN model has been built using Conv2D, Max Pooling, Batch Normalization, Dropout and Dense parameters.Activation function used is Relu and softmax.Adam optimizer has been used in the model.The model was trained for 40 epochs.The total parameters generated during the training process is 4,478,727.The output has 7 nodes because our output emotions are seven in number.

**What is CNN?**

A Convolutional Neural Network, also known as CNN or ConvNet, is a class of neural network that specializes in processing data that has a grid-like topology, such as an image.A CNN typically has three layers: a convolutional layer, a pooling layer, and a fully connected layer.The layers are arranged in such a way so that they detect simpler patterns first (lines, curves, etc.) and more complex patterns (faces, objects, etc.) further along.



The **convolution layer** is the core building block of the CNN. It carries the main portion of the network’s computational load.This layer performs a dot product between two matrices, where one matrix is the set of learnable parameters otherwise known as a kernel, and the other matrix is the restricted portion of the receptive field.

The **pooling layer** replaces the output of the network at certain locations by deriving a summary statistic of the nearby outputs. This helps in reducing the spatial size of the representation, which decreases the required amount of computation and weights. The pooling operation is processed on every slice of the representation individually.

**Fully Connected Layers** neurons in this layer have full connectivity with all neurons in the preceding and succeeding layer as seen in regular FCNN. This is why it can be computed as usual by a matrix multiplication followed by a bias effect.The FC layer helps to map the representation between the input and the output.

**Activation Function Used:**

**Relu:**This is most popular activation function which is used in hidden layer of NN.The formula is deceptively simple: 𝑚𝑎𝑥(0,𝑧)max(0,z). Despite its name and appearance, it’s not linear and provides the same benefits as Sigmoid but with better performance.It’s main advantage is that it avoids and rectifies vanishing gradient problems and is less computationally expensive than tanh and sigmoid.

**Softmax:** we use the function at the last layer of the neural network which calculates the probability distribution of the event over ’n’ different events. The main advantage of the function is being able to handle multiple classes.

**Optimizer Used:**

**Adam** is a popular algorithm in the field of deep learning because it achieves good results fast.Adam is a replacement optimization algorithm for stochastic gradient descent for training deep learning models.Adam combines the best properties of the AdaGrad and RMSProp algorithms to provide an optimization algorithm that can handle sparse gradients on noisy problems.Adam is relatively easy to configure where the default configuration parameters do well on most problems.

**LossFunctionUsed:**

Categorical cross entropy is a loss function that is used in multi-class classification tasks. These are tasks where an example can only belong to one out of many possible categories, and the model must decide which one.Used as a loss function for multi-class classification model where there are two or more output labels.

**Metrics Used:**

A metric is a function that is used to judge the performance of your model.**Accuracy** calculates how often predictions equal labels.This metric creates two local variables, total and count that are used to compute the frequency with which y\_pred matches y\_true. This frequency is ultimately returned as accuracy: an idempotent operation that simply divides total by count.

**Model Callbacks:**

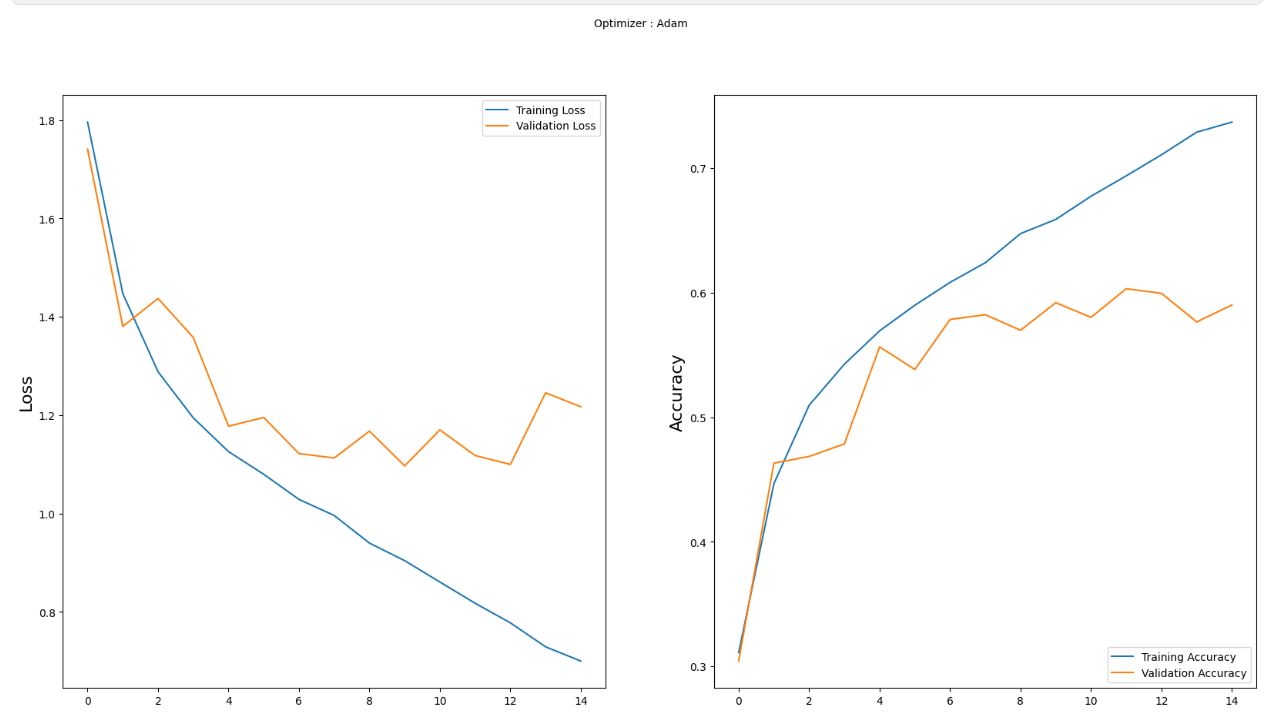
**ModelCheckpoint** this function of keras callbacks is used to save the model after every epoch. We just need to define a few of the parameters like where we want to store, what we want to monitor, define the path and then assigned val\_accuracy to be monitored

**EarlyStopping** is used to stop the model training in between. This function is very helpful when your models get overfitted. It is used to stop the model as soon as it gets overfitted. We defined what to monitor while saving the model checkpoints. We also need to define the factor we want to monitor while using the early stopping function. We will monitor validation loss for stopping the model training.

**ReduceLearningRate** when a metric has stopped improving.Models often benefit from reducing the learning rate by a factor of 2-10 once learning stagnates. This callback monitors a quantity and if no improvement is seen for a 'patience' number of epochs, the learning rate is reduced.

​**5.Model Evaluation**

A convolutional neural network can be evaluated using the ‘evaluate’ method. This method takes the test data as its parameters. Before this, the data is plotted on the console using ‘matplotlib’ library and ‘imshow’ methods.The accuracy versus epoch data is visualized.This is done using matplotlib library.The model is evaluated, and the loss and accuracy are determined.



**6.WebApp Creation**

Streamlit which is an open source web framework has been used to build face emotion recognition web app.OpenCV, an open source computer vision and machine learning software library has also been used for real time face reading.The model weights were saved in json and h5 file format which were later used.Created a function FaceEmotion to detect multiple faces in video camera which further provides a bounding box around faces and predicts face emotion.

Streamlit library does not provide the live capture feature itself ,instead uses a third party API.

Therefore,I used streamlit-webrtc which helped to deal with real-time video streams.

**7.Cloud Deployment**

Then this model was deployed on heroku platform.The web app link is:

<https://xcv-app.herokuapp.com/>

**8.Conclusion**

Finally built a Face Emotion Recognition webapp using streamlit and deployed on heroku cloud, which predicts the face emotions on live webcam.The total epochs considered initially for the training the images was 45.Too many epochs can lead to overfitting of the training dataset, whereas too few may result in an underfit model. However by usage of early stopping which allowed to specify an arbitrary number of training epochs and stop training once the model performance stops improving on a hold out validation dataset.The model created with CNN layers gave training accuracy 73.69% and validation accuracy 59.01% after 15 epochs.The Fer2013 dataset used had less number of disgust images. So the model is unable to distinguish the disgust emotion.

**References-**

MachineLearningMastery

developers.google.com

www.evoketechnologies.com

Analytics Vidhya

GeeksforGeeks

Kaggle - <https://www.kaggle.com/datasets/msambare/fer2013>

GitHub