Capstone Project 05

Live Class Monitoring System – Face Emotion Recognition

By: Prashant Gaikwad

(Individual Project)

Points for Discussion

- ❑ Introduction
- ❑ Problem Statement
- ❑ FER Dataset
- ❑ Data Pre-Processing
- ❑ Feature Extraction
- ❑ CNN Model Building
- ❑ Model Testing
- ❑ Confusion Matrix
- ❑ Conclusion

INTRODUCTION

- 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.

- 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.

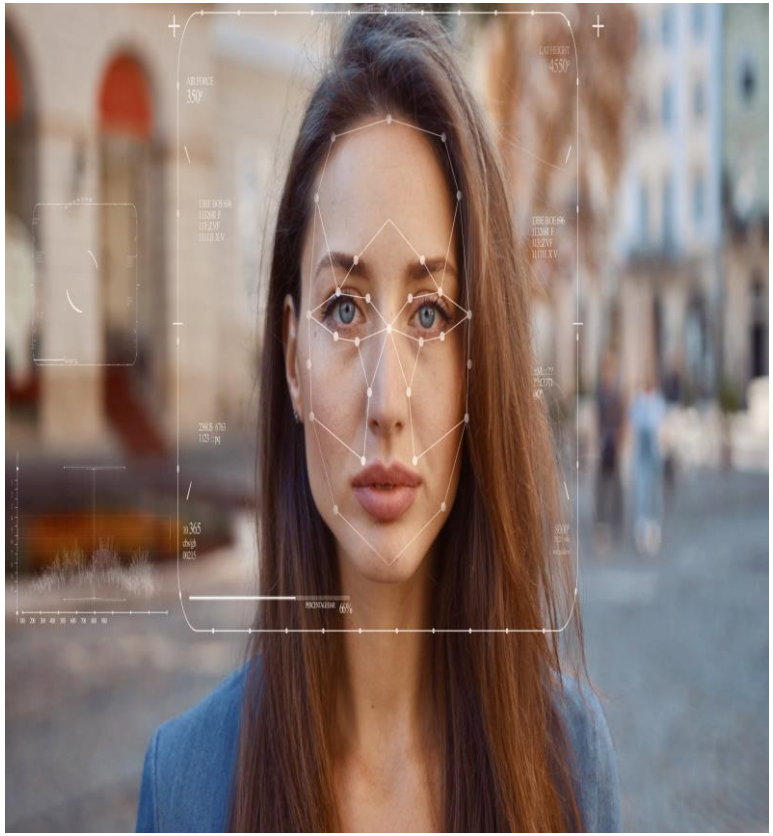
- In a physical classroom during a lecturing teacher can see the faces and assess the emotion of the class and tune their lecture accordingly, whether he is going fast or slow. He can identify students who need special attention.

- 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.

PROBLEM STATEMENT

- Deep learning backed system not only solves 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 this challenge by applying deep learning algorithms to live video data.



Face Emotion Recognition Dataset

FER-2013 Data:

The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centered and occupies about the same amount of space in each image.

The task is to categorize each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

The training set consists of 28,709 examples and the public test set consists of 3,589 examples.

Data Pre-Processing

It Contains several steps like :

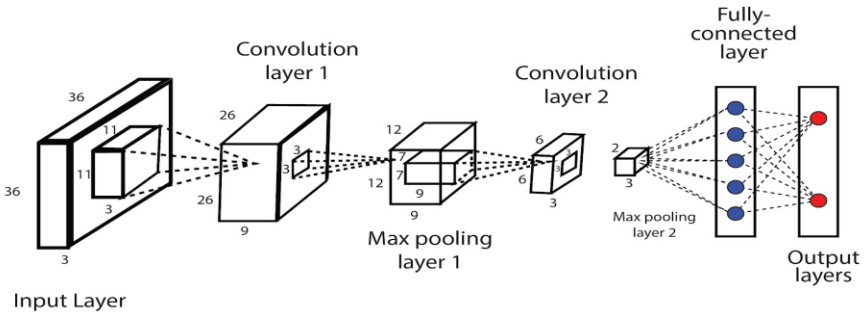
Data Collection
Data Import
Data Inspection
Data Splitting
Resizing image data

Feature Extraction

- In Traditional Machine Learning
- Input Image >> Feature Selections >> Classifier >> Result
- In Deep Learning
- Input Image >> Neural Network >> Result
- For CNN
- Input Layer >> Hidden Layers >> Output
- Here, input layer takes the input and output gives the desired output. That means hidden layer is doing some feature extractions. And if we extract the outputs of the hidden layers, then we will get different features.
- So basically, In ML we have to create our own feature vector, while for DL algorithms they extract features automatically in convolution layers.

Convolutional Neural Network Model (CNN):

- A CNN is composed of two basic parts of feature extraction and classification.
  - Feature extraction includes several convolution layers followed by max-pooling and an activation function.
- The classifier usually consists of fully connected layers. CNN automatically detects the important features without any human supervision And it makes efficient model which performs automatic feature extraction to achieve superhuman accuracy



CNN Model Building Details :

The main building block of CNN is the convolutional layer.

In our case the convolution is applied on the input data using a convolution filter to produce a feature map.

After a convolution operation we usually perform pooling to reduce the dimensionality.

This enables us to reduce the number of parameters, which both shortens the training time and combats overfitting.

Pooling layers down sample each feature map independently, reducing the height and width.

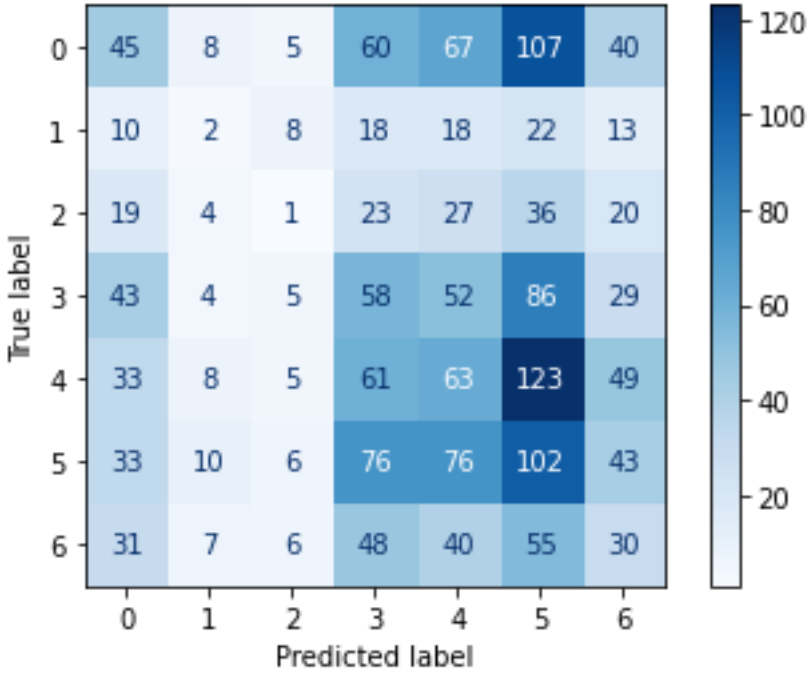
The output of both convolution and pooling layers are 3D volumes, but a fully connected layer expects a 1D vector of numbers.

So we flatten the output of the final pooling layer to a vector and that becomes the input to the fully connected layer.

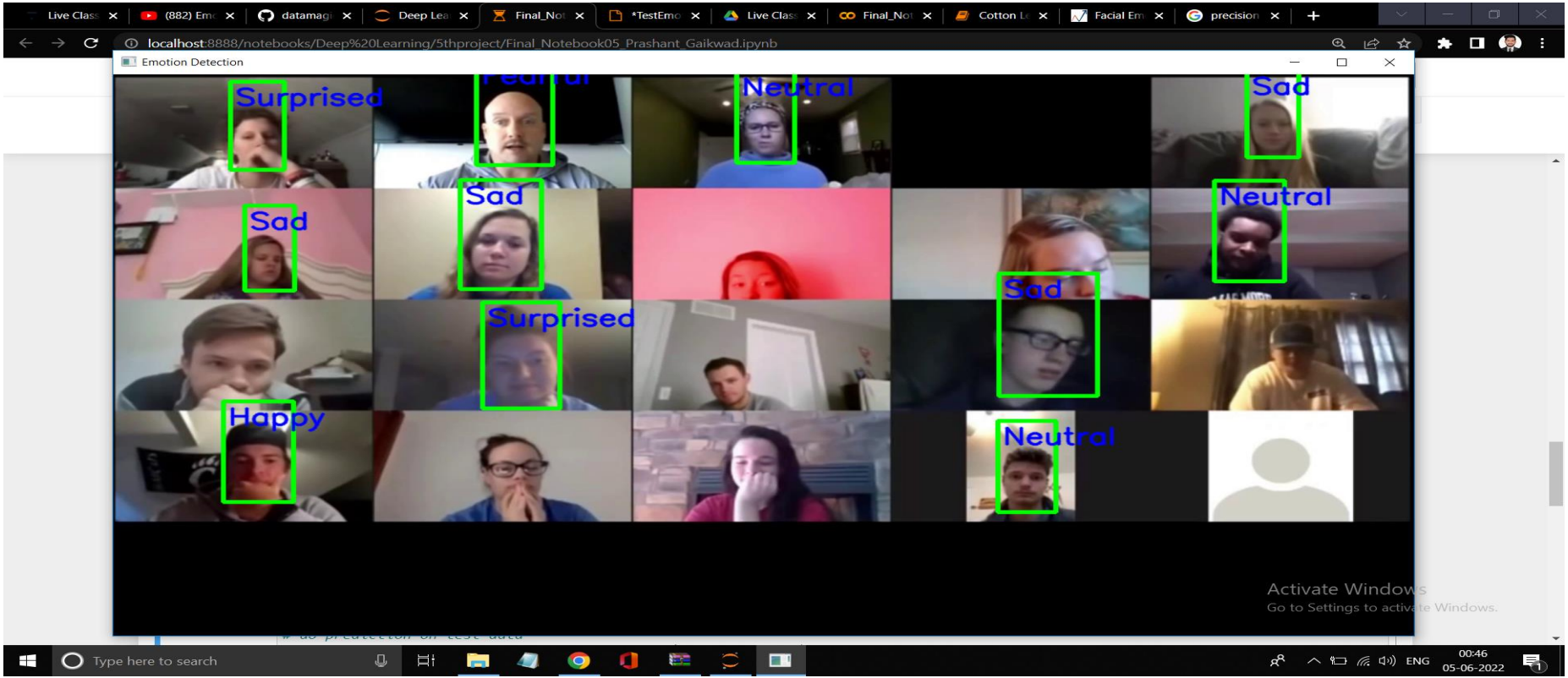
Flattening is simply arranging the 3D volume of numbers into a 1D vector.

Confusion matrix

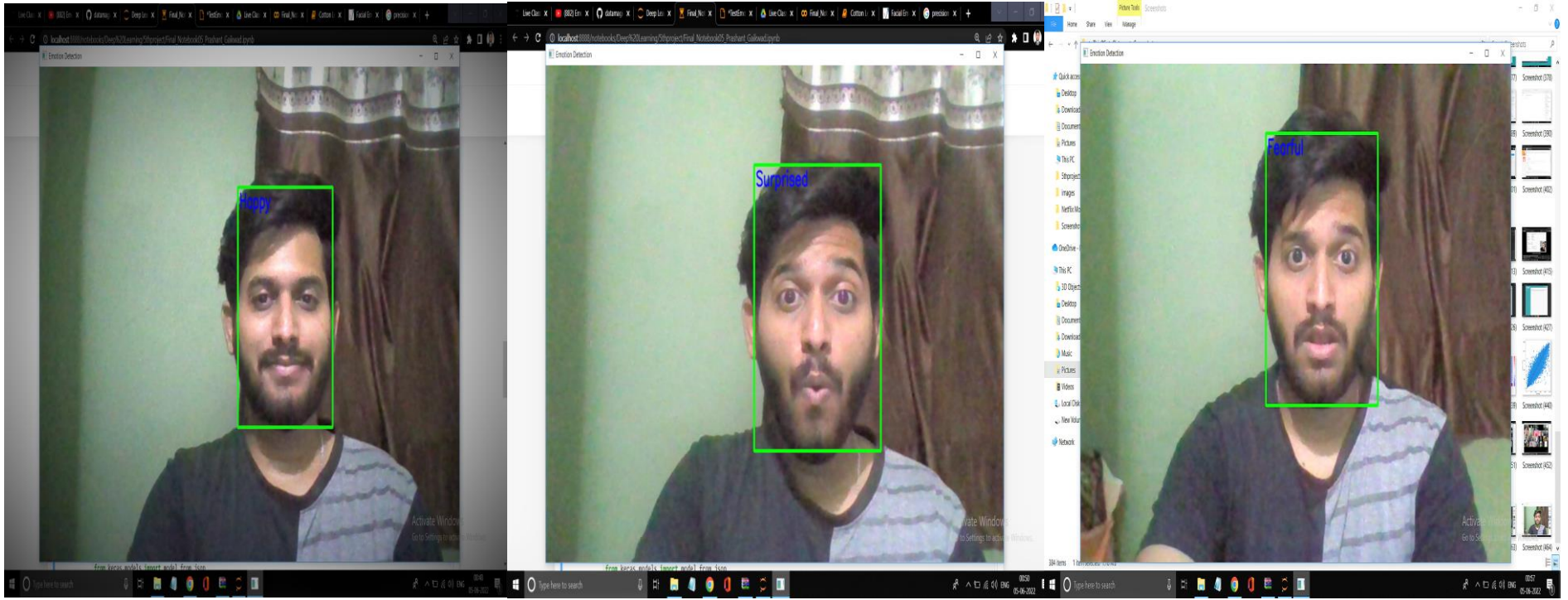
- The confusion matrix clearly shows that our model is doing good job on the class Happy, Neutral, Sad, Surprised, Angry, but it's performance is low on Fearful and Disgusted class.
- One of the reason for this could be the fact that these two classes have less data.
- Some images from these two classes are even hard for a human to tell whether the person is sad or neutral.
- Facial expression depends on individual as well. Some person's neutral face looks more like sad.



Live Class Monitoring System:



Live Webcam Face Recognition:



CONCLUSION

- In this Project, we have seen how to preprocess image data, design a network that is capable of classifying the emotions, and then use OpenCV for the detection of the faces and then pass it for prediction.
- The Project aims to create a system that automatically supports teachers and related education.
- Aims to create a system that automatically supports teachers and skills related to monitoring student behavior.
- The system will serve as an important factor in decision-making processes.
- The results of the emotion detection algorithm gave average accuracy up to 96% for Convolutional Neural Network model.
- The confusion matrix clearly shows that our model is doing good job on the class Happy, Neutral, Sad, Surprised, Angry, but it's performance is low on Fearful and Disgusted class.

THANK YOU