**Face Emotion Recognition**

**Problem Statement:** To identify face emotions in a video input, and predict the emotion of the person.

**Methodology:**

The paper “Real-time Convolutional Neural Networks for Emotion and Gender Classification” has been used as a reference in developing this model. Here we use a convolution neural network to train the emotion recognition model. We use a labelled dataset of faces. The input Image dataset contains multiple images of different people and its corresponding emotion label.

Initially dataset pre-processing is done. First the images are converted into a gray scale and then downscaled to 256\*256 pixels. This image is then flattened and converted into a single array of length 256\*256. Each image is labelled with a corresponding emotion. The emotions are represented by numbers – 1 = Happy, 2 = Angry, 3 = neutral, 4 = sad, 5 = surprised.

All the images are then stacked up upon each other with their corresponding labels which generates a new 2-D array of all images with its labels.

A neural network is created with 256\*256 neurons as input and a single output layer. The input is a single image and corresponding output is the expression.

The dataset used here is the FER-2013 emotion dataset, found on Kaggle.

The model was tested on IMBD gender dataset, which is again found on Kaggle.

Here we used the VGG-16 architecture to train our model. We also created our own model with three hidden layers, 2 Conv 2D layers and a softmax output layer, with RELU activation function.

Once the model is trained, we import the video and go frame by frame.

For every frame, we apply face detection and recognition by using histogram of oriented gradients. Using HOG, we identify faces and extract the features and create a region of interest. Once we have the Region of interests, we convert them into gray scale and crop them and pass this image of cropped face/faces through the model, which predicts the emotion.

The time taken to process a frame with 3 faces, of the video is approximately 1 second. Thus a 24 frames per second video of 20 seconds with minimum 3 faces per frame, will take 480 seconds or 8minutes to process. It also depends on the number of faces present in a video.

The excel sheet generated has frame number, number of faces for that frame, the face bounding box and prediction with accuracy. The frame number might seem to be repeated based on the number of faces found in frame, just to fill the excel sheet and not have blank spaces in between.