SYNOPSIS

- **Topic**: CV: Facial recognition and emoji generation
- **Problem statement** : Given a dataset of faces displaying various emotions, classify the emotion.
- **Image pre-processing:**

Pre-processing is the term for operations on images at the lowest level of abstraction.

There are 4 different types of Image Pre-Processing techniques and they are listed below.

1. Pixel brightness transformations/ Brightness corrections
2. Geometric Transformations
3. Reduce the noise
4. Convert The Image To Binary/Grayscale.

Pixel brightness transformations(PBT)

Brightness transformations modify pixel brightness and the transformation depends on the properties of a pixel itself.

Geometric Transformations

Geometric transforms permit the elimination of geometric distortion that occurs when an image is captured. The normal Geometric transformation operations are rotation, scaling and distortion (or undistortion!) of images.

There are two basic steps in geometric transformations:

1. Spatial transformation of the physical rearrangement of pixels in the image

2. Grey level interpolation, which assigns grey levels to the transformed image

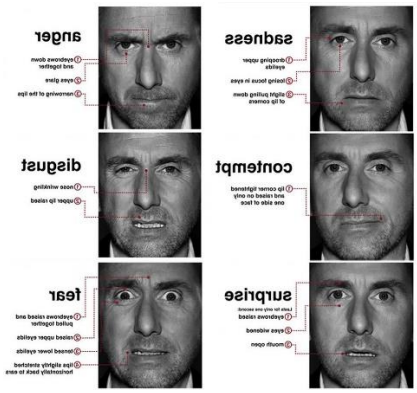
Wiener filter

The wiener filter [18] method take out noise from each pixel in an image using statistical approach. It performs an optimal transaction between inverse filtering and noise smoothing. It is a best filter in the process of noise smoothing to reduce the overall MSE. It tries to build an image by imposing a MSE constraint between the estimated and the original image. Wiener filters uses frequency domain for the analysis and some times are unable to reconstruct frequency component which have been degraded by noise.

minimised error is given by

$$e^2 = \mathrm{E}\left\{\mathrm{J}(\mathrm{x},\mathrm{y}) - \hat{J}(\mathrm{x},\mathrm{y})\right\}$$

- **Techniques used for classification :**
  - Face Registration : Face Registration is a computer technology being used in a variety of applications that identifies human faces in digital images. In this face registration step, faces are first located in the image using some set of landmark points called “face localization” or “face detection”. These detected faces are then geometrically normalized to match some template image in a process called “faceregistration”.
  - Facial Feature Extraction : Facial Features extraction is an important step in face recognition and is defined as the process of locating specific regions, points, landmarks, or curves/contours in a given 2-D image or a 3D range image. In this feature extraction step, a numerical feature vector is generated from the resulting registered image. Common features that can be extracted are a. Lips b. Eyes c. Eyebrows d. Nose tip
  - Emotion Classification : In the third step, of classification, the algorithm attempts to classify the given faces portraying one of the six basic emotions. Paul Ekman is an American psychologist at the University of California who is a pioneer in the study of emotions and their relation to facial expressions. He has created an "atlas of emotions" with more than ten thousand facial expression.



ALGORITHM :

Step 1 :Collection of a data set of images.

Step 2 :Pre-processing of images.

Step 3 :Detection of a face from each image.

Step 4 :The cropped face is converted into grayscale images.

Step 5 : The pipeline ensures every image can be fed into the input layer as a (1, 48, 48) numpy array.
Step 6:The numpy array gets passed into the Convolution2D layer.

Step 7 :Convolution generates feature maps.

Step 8 :Pooling method called MaxPooling2D that uses (2, 2) windows across the feature map only keeping the maximum pixel value.

Step 9 :During training, Neural network Forward propagation and Backward propagation performed on the pixel values.

Step10:The Softmax function presents itself as a probability for each emotion class.

Neural Network Approach : The neural network contained a hidden layer with neurons. The approach is based on the assumption that a neutral face image corresponding to each image is available to the system. Each neural network is trained independently with the use of on-line back propagation.

Deep learning is a popular technique used in computer vision.

Input Layer : The input layer has pre-determined, fixed dimensions, so the image must be pre-processed before it can be fed into the layer. We used OpenCV, a computer vision library, for face detection in the image. The haarcascade_frontalface_default.xml in OpenCV contains pre-trained filters and uses Adaboost to quickly find and crop the face. The cropped face is then converted into grayscale using cv2.cvtColor and resized to 48-by-48 pixels with cv2.resize. This step greatly reduces the dimensions compared to the original RGB format with three color dimensions

Convolutional Layers : The numpy array gets passed into the Convolution2D layer where we specify the number of filters as one of the hyperparameters. The set of filters are unique with randomly generated weights. Each filter slides across the original image with shared weights to create a feature map. Convolution generates feature maps that represent how pixel values are enhanced .Other filters are applied one after another creating a set of feature maps. chose Convolutional Neural Network (CNN) layers as building blocks to create our model architecture. CNNs are known to imitate how the human brain works when analyzing visuals. Pooling is a dimension reduction technique usually applied after one or several convolutional layers.

Dense Layers : The dense layer takes a large number of input features and transform features through layers connected with trainable weights. These weights are trained by forward propagation of training data then backward propagation of its errors. Back propagation starts from evaluating the difference

between prediction and true value, and back calculates the weight adjustment needed to every layer before. We can control the training speed and the complexity of the architecture by tuning the hyper-parameters, such as learning rate and network density. As we feed in more data, the network is able to gradually make adjustments until errors are minimized.Essentially, the more layers/nodes we add to the network the better it can pick up signals.

Output Layer : Instead of using sigmoid activation function, we used softmax at the output layer. This output presents itself as a probability for each emotion class.Therefore, the model is able to show the detail probability composition of the emotions in the face. As later on, you will see that it is not efficient to classify human facial expression as only a single emotion.

A typical architecture of a convolutional neural network contain an input layer, some convolutional layers, some dense layers (aka. fully-connected layers), and an output layer . These are linearly stacked layers ordered in sequence

CONCLUSION

Its built-in convolutional layer reduces the high dimensionality of images without losing its information. That is why CNNs are especially suited for image classification over any other algorithm . in the case of deep neural networks each neuron in a given layer is fully connected to all the neurons in the previous layer. Because of these large number of connections the number of parameters to be learned increases. As the number of parameters increases the network becomes more complex. This more complexity of the network leads to overfitting. In CNN, the input image data will be subjected to set of convolution operations such as filtration and max pooling. Then, the resultant data which will be of lesser dimension compared to the original image data will be subjected to Fully connected layers.