**CSE- 454**

Data Mining Sessional Project

Gender Determination From Facial Image

**Group Members**

Sanzida Mojib Luna – 201614019

Maliha Sultana - 201614024

**Table of Contents**

Background3

Dataset3

Method3

Performance4

Summery4

**Background :**

As uses of technology is increasing every day, the importance of using robot and computer vision is high in demand. Computer takes image as input or its vision and does necessary analysis for performing required task. Also image processing and gaining data from image is used in a lot of sectors like biometrics database, area based security and restriction purpose etc.

On the other hand, machine learning is being used for gaining more accuracy and teaching a system how to be more efficient over time. Using the concept of machine learning “Convolution Neural Network (CNN)” has been introduced, which works a lot like the neural system of human brain.

Using these two methods several systems have been developed like – face detection, person recognition from image, facial expression detection, age detection etc. Gender detection from image can be classified into the mentioned category but the approach is different.

Gender detection from image can be used in entrance control, security purpose, gaining information from image database, understanding marketing demography and in many more aspects.

**Dataset Used :**

The dataset that is used here are from Wikipedia which is one of the largest database available online. 8000 images are selected for training purpose where meta data of those images are saved in a Matlab file.

**Method:**

Following steps have been conducted for developing the model -

1. Data Pre-processing : OpenCV Face Detector and HaarCascade has been used to find out faces in the image. Images not containing a face or blurry images are left out.

2. Data Cleaning : From the selected data Matlab file has been edited as such that it only contains metadata related to the preprocessed data. Scipy has been used for that.

3. Preparing Model : Following features have been selected as the basis of training from the face images -

* Eyebrows
* Eyes
* Nose tip
* Lip contouring

The CNN model consisting of the following layers have been built for training the model –

* Convolution layer 1 : The first convolutional layer has 96 nodes of kernel size 7.
* Convolution layer 2 : The second convolutional layer has 256 nodes with kernel size 5.
* Convolution layer 3 : The third convolutional layer has 384 nodes with kernel size 3.
* The two fully connected layers have 512 nodes each.

4. Training : The dataset from earlier has been used for training purpose. It took almost seven hours for the training to be finished and the result is model weight that is in CaffeNet form.

5. Testing : Another python code has been developed for combining the weighted data and the CNN model , so that it can be used in testing. For testing total 50 random images have been selected form internet keeping the diversity.

**Performance:**

As the performance measurement, the model calculates how much the features of testing data matches with calculated features of trained data. For testing, the model runs on only one image at a time. It calculate the confidence and shows output in an image form, where a square box drawn around the face and on top left corner the output “Male/ Female” is written.

**Summery:**

The model gives correct output for more than 70% of its test data which is quite satisfactory for a small scale model like this. It can gain for efficiency if it is trained on a larger data set.