Facial Emotion Recognition

CS337: Artificial Intelligence and Machine Learning

Nama Hari Krishna	Dileep Kumar	Poorvi Hebbar	Ashrut Vaddela
170050077	170050080	170050094	170050079

Description

A face emotion recognition system comprises of two step process i.e. face detection (bounded face) in image followed by emotion detection on the detected bounded face:

- Haar feature-based cascade classifiers: It detects frontal face in an image well. It is real time and faster in comparison to other face detectors.
- CNN Model : We will train a classification CNN model architecture which takes bounded face (48×48 pixels) as input and predicts probabilities of 7 emotions (happy, neutral, angry, disgust, scared, sad and surprise) in the output layer.

Data-set

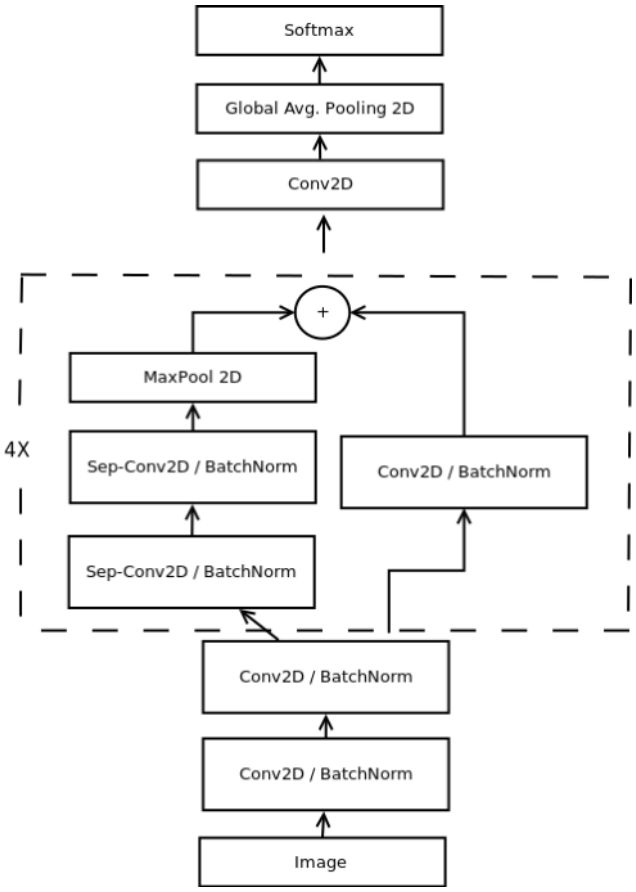
Data-set used is the facial expression recognition (FER) data-set from Kaggle challenge (link). The data consists of 48×48 pixel gray scale images of faces. The training set consists of 35,888 examples. train.csv contains two columns, "emotion" and "pixels".

Our project files

- **models:** Models corresponding to different epochs are saved in this folder. Name of model file indicates epoch number and accuracy achieved by this model.
- **classify-image.py:** It first loads data, images and the models. It then reads the frame to detect faces. The region of interest of the face is then used for classification via the CNN. The different emotion probabilities are then displayed.
- **cnn.py:** We define the class cnn-model here.
- **train.py:** Trains on the dataset using the cnn_model and saves all those models which result in a decrease in validation loss from previous model.
- **video-demo.py:** Similar to classify-image.py, it detects faces in each frame of the real time video.
- **datasets.py:** *_load_fer2013*: Reads the dataset (fer2013.csv) and returns faces and emotion labels.
- **interference.py:** Various functions in inference.py help in utilities like detecting faces and drawing bounding boxes.
- **preproessor.py:** *def preprocess_input*: scales images down to lie between -1 and 1.

Model

Deep-learning-based FER approaches highly reduce the dependence on face-physics-based models and other pre-processing techniques by enabling "end-to-end" learning to occur in the pipeline directly from the input images.
The final architecture is a fully-convolutional neural network that contains 4 residual depth-wise separable convolutions where each convolution is followed by a batch normalization operation and a ReLU activation function. The last layer applies a global average pooling and a soft-max activation function to produce a prediction.



This model has less number of parameters. Reducing the number of parameters overcomes two problems. First, the use of small CNNs alleviates from slow performances. Second, the reduction of parameters provides a better generalization under an Occam's razor framework. This model completely eliminates fully connected layers. This architecture has approximately 60, 000 parameters. Depth-wise separable convolutions are composed of two different layers: depth-wise convolutions and point-wise convolutions. The main purpose of these layers is to separate the spatial cross-correlations from the channel cross-correlations.

Results

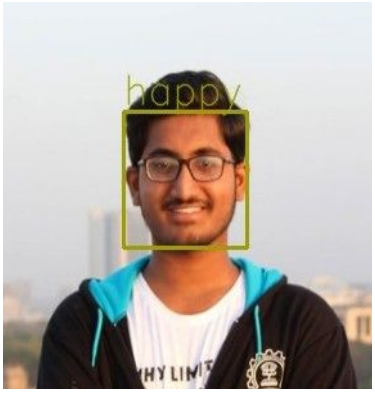
The results obtained by us on the model trained for 85 epochs, which gives 65% validation accuracy are as shown in the next page.

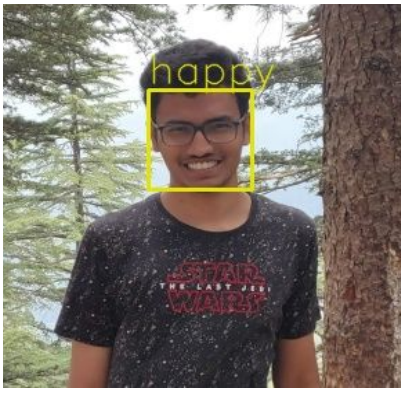
References

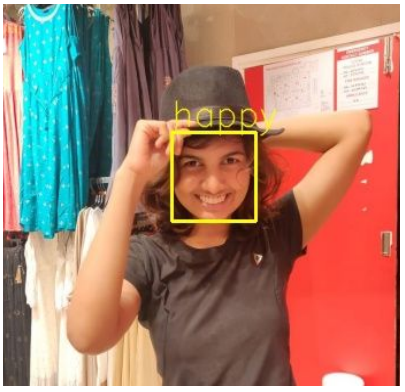
- A Brief Review of Facial Emotion Recognition Based on Visual Information: link
- Real-time Convolutional Neural Networks for Emotion and Gender Classification: link to pdf
- Face Classification: github repo

Emotion labels for test images:

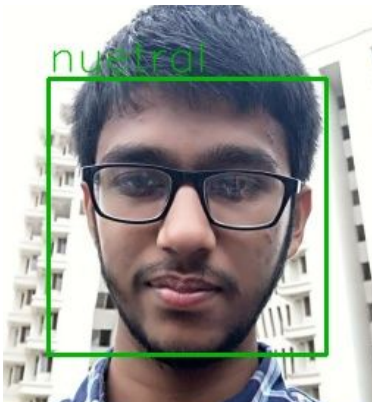
Happy:

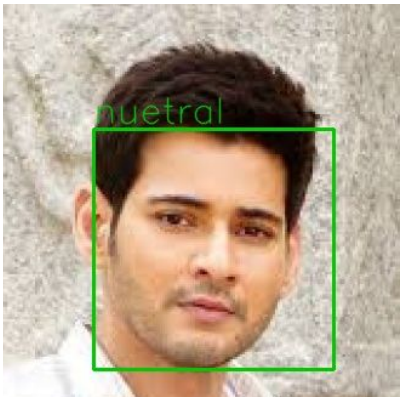




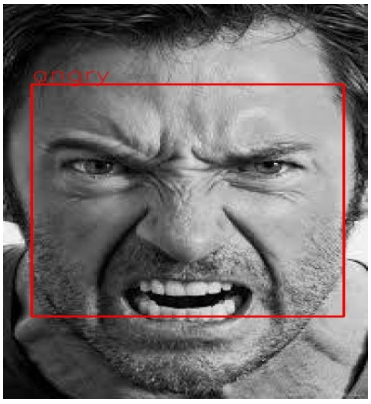


Neutral:

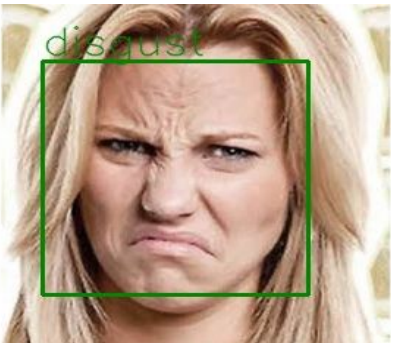


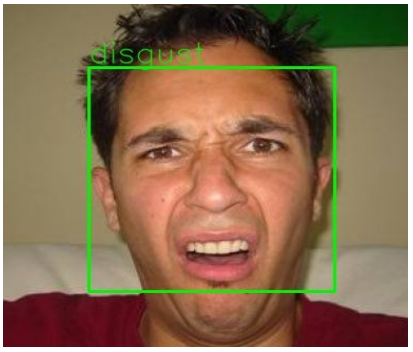


Angry:

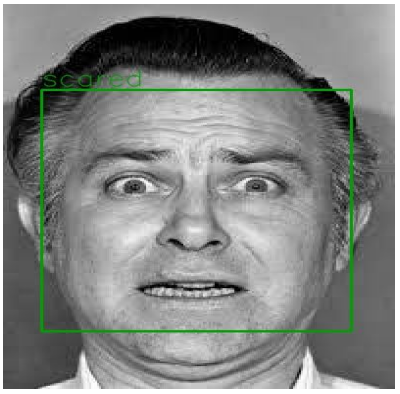


Disgust:

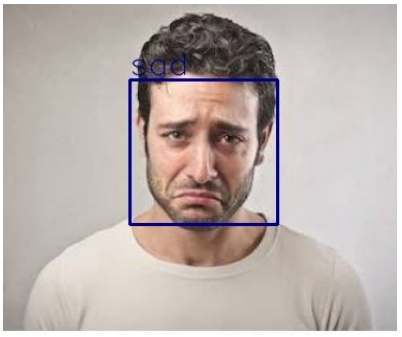




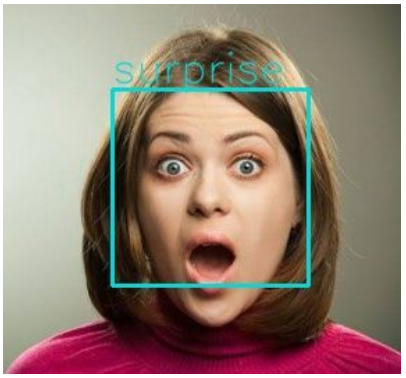
Scared:



Sad:



Surprise:



THANK YOU