# Machine Learning Engineer Nanodegree

## Facial Emotional Recognition

Karthik Balasubramanian
May 6th 2019

## Proposal

> Most of the content taken from the papers I read about Facial Emotional Recognition

### Domain Background

Facial emotions are important factors in human communication that help us understand the intentions of others. In general, people infer the emotional states of other people, such as joy, sadness, and anger, using facial expressions and vocal tone. According to different surveys, verbal components convey one-third of human communication, and nonverbal components convey two-thirds. Among several nonverbal components, by carrying emotional meaning, facial expressions are one of the main information channels in interpersonal communication. Interest in automatic facial emotion recognition (FER) has also been increasing recently with the rapid development of artificial intelligent techniques, including in human-computer interaction (HCI), virtual reality (VR), augment reality (AR), advanced driver assistant systems (ADASs), and entertainment. Although various sensors such as an electromyograph (EMG), electrocardiogram (ECG), electroencephalograph (EEG), and camera can be used for FER inputs, a camera is the most promising type of sensor because it provides the most informative clues for FER and does not need to be worn.

My journey to decide on this project was exciting. My motive was to prove the utility of Deep neural nets in the contemporary research. Facial emotional recoginition/ pattern recognintion had been in research since long. The following acaemic papers were very helpful in

1. Giving a historic overview of research in Facial Emotional Recognition
2. Deciding on a posed dataset with seven different emotions
3. Developing a baseline algorithm
4. Improving the Facial Emotions Recognition using Deep Convolutional Neuralnets

### Problem Statement

The objective of this project is to showcase two different solutions in solving the problem of Facial emotional recognition from a posed dataset. Both the solutions are based on the problem space of supervised learning. But the first solution I propose is more involved and has more human interference than the second solution which uses state of art artificial neuralnets. The goal is to compare the two approaches using a performance metric - i.e how well the supervised learning model detects the expression posed in a still image. The posed dataset has labels associated with it. The labels define the most probable emotion. After running our two supervised learning model, we use accuracy score as the performance metric to decide how well the model has performed.

accuracy score = A ratio of # of correctly predicted emotions in images / total number of images.

### Datasets and Inputs

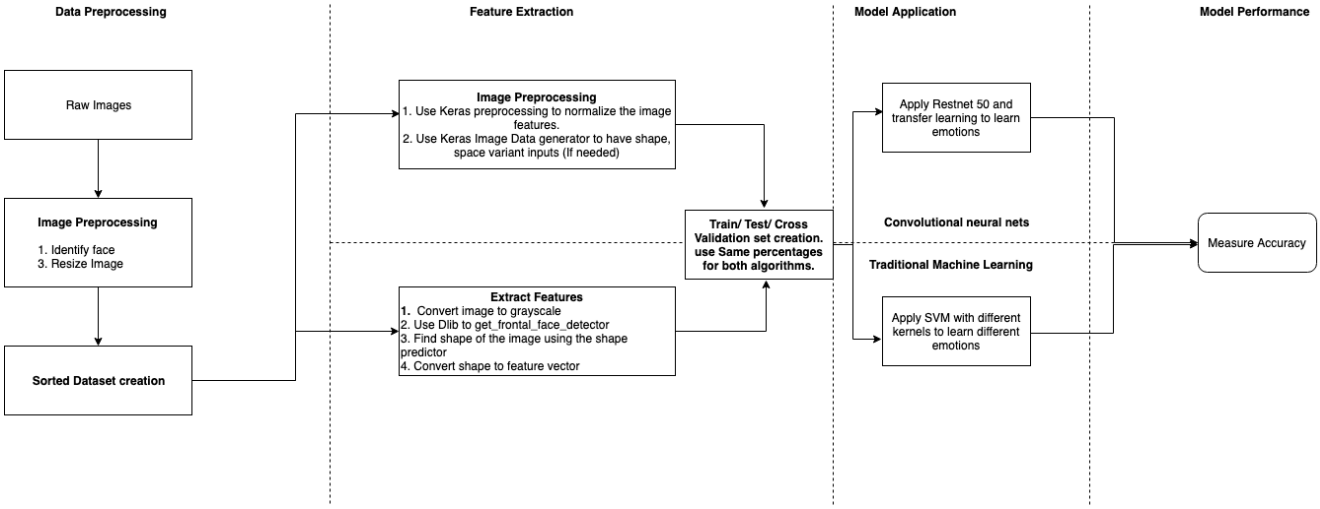
I use Cohn-Kanade dataset. This dataset has been introduced by Lucey et al. 210 persons, aged 18 to 50, have been recorded depicting emotions.Out of 210 people, only 123 subjects gave posed facial expression. This dataset contains the recordings of their emotions. Both female and male persons are present from different background. 81 % Euro-Americans and 13% are Afro-Americans. The images are of size 640 *490 pixels as well as 640* 480 pixels. They are both grayscale and colored. in total there are 593 emotion-labeled sequences. There are seven different emotions that are depicted. They are:

1. 0=Neutral
2. 1=Anger
3. 2=Contempt
4. 3=Disgust
5. 4=Fear
6. 5=Happy
7. 6=Sadness
8. 7=Surprise

The images within each subfolder may have an image sequence of the subject. The first image in the sequence starts with a neutral face and the final image in the sub folder has the actual emotion. So from each subfolder ( image sequence), I have to extract two

images, the neutral face and final image with an emotion. ONLY 327 of the 593 sequences have emotion sequences. This is because these are the only ones the fit the prototypic definition. Also all these files are only one single emotion file. I have to preprocess this dataset to make it as an uniform input. I will make sure the images are all of same size and atmost it has one face depicting the emotion for now. After detecting the face in the image, I will convert the image to grayscale image, crop it and save it. I will use OpenCV to automate face finding process. OpenCv comes up with 4 different pre-trained classifiers. I will use all of them to find the face in the image and abort the process when the face is identified. These identified, cropped, resize image becomes input feature. The emotion labels are the output.

## Solution Statement



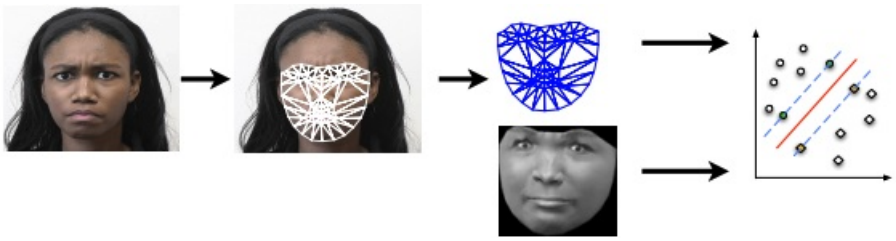

As suggested earlier, the goal of this project is to prove the utility of Deep neural nets and to showcase how it avoids careful human involvement. The procedure has 4 different phases as described in the image.

1. Data preprocessing - Steps done to create a clean dataset
2. Feature Extraction and Normalization methods - The supervised learning approach uses some hand crafted shape prediction methods while the deep neuralnets have data normalization steps. Both these methods are followed by creation of train, valid test set splits. Both the methods use the same proportion of train, valid and test splits.
3. Model Application - Model train and test phase.
4. Model Performance - Performance of the models are compared. The metric used is Accuracy.

## Benchmark Model

The baseline method I use to identify emotions is a very strong contender. There has been years of research involved in this approach.

Images Courtesy - Cohn-Kanade+ paper



The following method has 2 different phases.

1. Finding the face - Libraries like dlib has handy functions like `get_frontal_face_detector` which is handy to identify the face region
2. Extracting the features in the face - This is where most of the research in the past has gone into. It has been done so far by realizing through manual interference. One of the method is called Facial Action Coding System (FACS) which describes Facial expression using Action Units (AU). An Action Unit is a facial action like "raising the Inner eyebrow". Multiple Activation units when combined expresses the emotion in the underlying face. An example is provided below.

| AU | *Name* | N | AU | *Name* | N | AU | *Name* | N |
|---|---|---|---|---|---|---|---|---|
| 1 | *Inner Brow Raiser* | 173 | 13 | *Cheek Puller* | 2 | 25 | *Lips Part* | 287 |
| 2 | *Outer Brow Raiser* | 116 | 14 | *Dimpler* | 29 | 26 | *Jaw Drop* | 48 |
| 4 | *Brow Lowerer* | 191 | 15 | *Lip Corner Depressor* | 89 | 27 | *Mouth Stretch* | 81 |
| 5 | *Upper Lip Raiser* | 102 | 16 | *Lower Lip Depressor* | 24 | 28 | *Lip Suck* | 1 |
| 6 | *Cheek Raiser* | 122 | 17 | *Chin Raiser* | 196 | 29 | *Jaw Thrust* | 1 |
| 7 | *Lid Tightener* | 119 | 18 | *Lip Puckerer* | 9 | 31 | *Jaw Clencher* | 3 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 9 | *Nose Wrinkler* | 74 | 20 | *Lip Stretcher* | 77 | 34 | *Cheek Puff* | 1 |
| 10 | *Upper Lip Raiser* | 21 | 21 | *Neck Tightener* | 3 | 38 | *Nostril Dilator* | 29 |
| 11 | *Nasolabial Deepener* | 33 | 23 | *Lip Tightener* | 59 | 39 | *Nostril Compressor* | 16 |
| 12 | *Lip Corner Puller* | 111 | 24 | *Lip Pressor* | 57 | 43 | *Eyes Closed* | 9 |

Table 1. *Frequency of the AUs coded by manual FACS coders on the CK+ database for the peak frames.*

| Emotion | Criteria |
|---|---|
| Angry | AU23 and AU24 must be present in the AU combination |
| Disgust | Either AU9 or AU10 must be present |
| Fear | AU combination of AU1+2+4 must be present, unless AU5 is of intensity E then AU4 can be absent |
| Happy | AU12 must be present |
| Sadness | Either AU1+4+15 or 11 must be present. An exception is AU6+15 |
| Surprise | Either AU1+2 or 5 must be present and the intensity of AU5 must not be stronger than B |
| Contempt | AU14 must be present (either unilateral or bilateral) |

Table 2. *Emotion description in terms of facial action units.*

I use dlib's `shape_predictor` and its learned landmark predictor `shape_predictor_68_face_landmarks.bat` to extract AUs. Then as suggested in the model pipeline I implement SVM to compare distances between the extracted facial landmarks for different emotions.

The defintion of Activation units and its combinations to define emotions are very helpful. But handcrafting them and assuming that all the people express the same emotions with a combinations of same action units is not completely correct. This is where we can use Artificial Neural nets an uniform function approximater with no bias :)

### Evaluation Metrics

the evaluation metric I propose for comparing performances between Baseline and Neuralnets is accuracy. An ideal model would correctly classify the emotion to its true label.

Here is the definition of Accuracy

### Project Design

The Project workflow is designed the following way. I tend to keep the structure same but the content may vary.

1. Data Exploration and Preprocessing
   - Ensuring data without face images are filtered
   - Converting the image directory into an assigned subfolder based on its emotion
   - Resize/ Rescale images
   - Quantifying the emotion samples we have in the data via visualizations
1. Feature extraction
   - Visualizing action units
   - Deciding on Data normalization on Features
1. Creating a Train/ Test/ Validation sets
1. Deciding Algorithms
   - Using SVM for baseline models
   - Using different predefined neural net models
1. Hyper parameter tuning
   - Working with different gradient optimization methods for deep neural nets
   - Changing kernel for SVM models
1. Performance comparison
   - Providing confusion matrix for performances of both the models (baseline vs neural nets)
1. Retrospective
   - Learnings from the project
   - Paths for improvement

---

## References

I have provided most of my references in my write up. I want to thank Paul Vangent on helping me understand FACS system clearly.

Papers referenced:

1. Ko BC. A Brief Review of Facial Emotion Recognition Based on Visual Information. Sensors (Basel). 2018;18(2):401. Published

2018 Jan 30. doi:10.3390/s18020401
2. Lucey, Patrick & Cohn, Jeffrey & Kanade, Takeo & Saragih, Jason & Ambadar, Zara & Matthews, Iain. (2010). The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010. 94 - 101. 10.1109/CVPRW.2010.5543262.
3. Vanita Jain, Pratiksha Aggarwal, Tarun Kumar and Vaibhav Taneja. Emotion Detection from Facial Expression using Support Vector Machine. International Journal of Computer Applications 167(8):25-28, June 2017
4. DeXpression: Deep Convolutional Neural Network for Expression Recognition arXiv:1509.05371 [cs.CV]