Machine Learning Engineer Nanodegree Capstone Project Proposal

Facial Keypoint Recognition System

Daniel Vargas, 2019-10-19

Domain Background

Detecing facial keypoints (also called facial landmarks) on face images is a very challenging problem. Facial features vary greatly from one individual to another, and even for a single individual, there is a large amount of variation due to 3D pose, size, position, viewing angle, and illumination conditions. Computer vision research has come a long way in addressing these difficulties, but there remain many opportunities for improvement.

Solving this problem that can provide the building blocks for several applications, such as:

- tracking faces in images and video
- analysing facial expressions
- detecting dysmorphic facial signs for medical diagnosis
- biometrics / face recognition

I chose this specific challenge because I currently work in the medical diagnosis field. I expect this project to help me understand facial keypoints recognition in a deeper way.

Problem Statement

The objective of this project is to accurately predict the facial keypoints (facial landmarks) of a face image. My hypothesis is, that this prediction can be performed based on a training set containing accurate facial keypoints.

Datasets and Inputs

The data was acquired from this `Kaggle competition`.

Each predicted keypoint is specified by an (x,y) real-valued pair in the space of pixel indices. There are 15 keypoints, which represent the following elements of the face:

```
left_eye_center, right_eye_center, left_eye_inner_corner, left_eye_outer_corner,
right_eye_inner_corner, right_eye_outer_corner, left_eyebrow_inner_end,
left_eyebrow_outer_end, right_eyebrow_inner_end, right_eyebrow_outer_end, nose_tip,
mouth_left_corner, mouth_right_corner, mouth_center_top_lip, mouth_center_bottom_lip
```

- Left and right here refers to the point of view of the subject.
- In some examples, some of the target keypoint positions are misssing (encoded as missing entries in the csv, i.e., with nothing between two commas).
- The input image is given in the last field of the data files, and consists of a list of pixels (ordered by row), as integers in (0,255). The images are 96x96 pixels.

Data files

- `training.csv`: list of training 7049 images. Each row contains the (x,y) coordinates for 15 keypoints, and image data as row-ordered list of pixels.
- `test.csv`: list of 1783 test images. Each row contains ImageId and image data as row-ordered list of pixels

Solution Statement

A `Convolutional Neural Network` (`CNN`) will be applied to predict the facial keypoints. A `CNN` was chosen for this problem because:

- This is a computer vision problem that requires capturing features for prediction
- CNNs are very useful in capturing features in images

Benchmark Model

A simple `linear regression` will be used as a baseline model for comparison to confirm

Evaluation Metrics

`Root Mean Squared Error` (`RMSE`):

`RMSE` is very common and is a suitable general-purpose error metric in regression problems. Compared to the `Mean Absolute Error`, `RMSE` punishes large errors:

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}$$

Project Design

- Collect the `training.csv` and `test.csv` data.
- Explore and visualize the data
    - Detect faces Using a `Haar Cascade Classifier`
    - Add eye detection
- Train a `Convolutional Neural Network` (`CNN`) to detect facial keypoints
- Test the trained model against the `test dataset` throught `RMSE`