Machine Learning Engineer Nanodegree

Capstone Proposal

Karthik Ramesh May 1st, 2018

Proposal

Domain Background

Facial feature extraction and analysis have become one of the rapidly growing areas of exploration with the advent of deep learning and its advancement. The Convolution Neural Network (CNN) has been a revelation wherein it takes into account the spatial aspect of the images to come up with efficient analysis. The field has definitely grown leaps and bounds over the past few years and the applications where it is used is expanding constantly.

CNN has found to be very efficient in the field of visual computing and analysing images to identify useful patterns. This has eliminated the need for identifying feature vectors manualy and led to more accurate understanding of implicit features. But the real world datasets are not always taken in a well equipped enviroment. This means that the images that we obtain in real world will be of lesser quality, differnt positions, sizes and far from perfection. The CNN should be able to perform well with these datasets and only then it can be deemed fit for real world applications.

The academic paper that was also the motivation for the project is "Indian Movie Face Database: A Benchmark for Face Recognition Under Wide Variations" [2]. The challenges in the mere collection of this dataset and the appicability and exploration of potential of deep learning on such datasets would be really interesting.

Problem Statement

The problem is to identify the age of the actors/actresses from the image supplied to the network. This is a multi-class classification problem where the age lies in any of the following buckets - Young, Middle and Old. The input images are captured from videos and hence the quality of images vary widely. The CNN should be able to identify and extract the features from these images with high degree of variability in terms of scale, pose, expression, illumination, age, resolution, occlusion, makeup and categorize the images to proper age bucket.

Datasets and Inputs

The dataset for this problem is obtained from the ongoing datahack that is being conducted by Analytics Vidya (https://datahack.analyticsvidhya.com/contest/practice-problem-age-detection/ (https://datahack.analyticsvidhya.com/contest/practice-problem-age-detection/)). This is a subset of data obtained from ndian Movie Face database (IMFDB). Indian Movie Face database (IMFDB) is a large unconstrained face database consisting of 34512 images of 100 Indian actors collected from more than 100 videos. The images are manually selected and cropped from the videos that make it high degree of variability in terms described in the problem statement earlier. The IMFDB also provides detail annotation for every image in terms of age, pose, gender, expression and type of occlusion that may help other face related applications [2].

The images from the database widely vary in size and shape, brightness, quality and viewpoints. The images have to be resized so that we have same dimensions to deal with. Most of the image detection algorithms use normalization to achieve better standardization and resolve any bias during computation which is also a good technique to build better models. But these are only preliminary steps to standardize and will need to research and apply some other techniques along the way if it helps to improve the prediction.

The dataset that will be used for this project consists of 26742 cleaned and formatted images with 9906 images in train and 6636 images in test. The attributes of this dataset are the Unique ID of the image and the age bin - Young, Middle and Old. There are total of 10805 images in Middle class, 2397 images in Old class and 6707 images in Young class

Benchmark Model:

The benchmark model can consists of vanilla model involving only dense layers with the flattened input. The sample model can be as below:

Input Layer -> Flatten -> Dense Layer -> Output Layer.

This is a basic model wherein the input is flattened into an array and have one hidden dense layer that act on the flattened input. The output layer is the classification layer that provides multi-class output. The model that this project aims to develop should be able to outperform this benchmark model.

Solution Statement

I propose to use Convolution Neural Networks (CNN) to analyse the training set to better predict the age bin of the images in the test set. In some ways, this is similar to the dog prediction project that was performed as a part of deep neural networks wherein the breed of the dog was predicted based on the supplied images. But the challenges to accomplishing higher accuracy to this prediction comes in various forms as below:

```
  - High Degree of Variability in terms of pose, image quality and expression
 s.
  - Applying proper image filters to extract the information that is required
   to solve the problem at hand.
  - Research and apply transfer learning filters that can be used to make the
   feature extraction more effective.
```

Evaluation Metrics

I will be making use of Keras for the implementation of CNN and the accuracy metrics in keras will be used to evaluate the model's performance. The test data consits of close to 6500 images and can give pretty good idea of model's performance. Metrics - accuracy and f1 score. Accuracy to understand overall how well the algorithm is performing and F1 score metric is used to calculate the harmonic average of precision and recall which is considered one of the unbiased units for determining efficiency.

Project Design

The project workflow would be as below:

1. Explore the dataset and capture interesting findings.
2. Load the training dataset.

3. Use image cleansing, augmentation techniques if needed.
4. Build the convolution layer using Keras
5. Analyze the performance and tweak the layer based on the performance
6. Explore transfer learning to detect better facial features
7. Compare performance between various models
8. The solution from the model would be ID of the image and the predicted age bin.

There should certainly be formatting of images needed and brief overview of that was provided in the datasets section as well. The images vary in size and need to be resized and normalized. I will implement some of the techniques that were used during "Dog Breed Prediction" assignment in terms of formatting, implementation and visualization. The algorithm will make use of Conv2D, maxpooling layers to identify crucial edges in the images to better perform in prediction. I can make use of transfer learning from any of VGG-19, ResNet-50, Inception and Xception to see how the algorithm perfform to compare.

References

1. Felipe Moure C´ıcero, Ary Henrique M Oliveira, Glenda Michele Botelho (2016) Deep Learning and Convolutional Neural Networks in the Aid of the Classification of Melanoma
2. S. Setty et al., "Indian Movie Face Database: A benchmark for face recognition under wide variations," 2013 Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), Jodhpur, 2013, pp. 1-5.
3. National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2013.