Machine Learning Engineer Nanodegree

Capstone Project Proposal

Image Detection for Facial Attribute - Smiling

Jyothi Mudur

December 28, 2019

Proposal

Domain Background

Machine Learning has varied uses. One such is in the field of Image Recognition to detect facial attributes in the images of people. When it comes to images, Convolutional Neural Network (CNN) models are preferred to Multi Layer Perceptrons (MLPs). After the basic preprocessing of the data (if required) is done, the images can be directly fed into the CNNs. This is however, not the case with MLPs wherein it is required to convert the multi-dimensional array (images) to a single dimensional vector. Also, the number of parameters to train would be enormously huge in the case of MLPs.

In this project, I would be creating a model with acceptable accuracy (as compared to the benchmark model) to detect if the person in the given image is smiling or not. CalibA [1] dataset is used here which contains over 200,000 celebrity images, each with 40 attribute annotations - 'Smiling' being one such attribute.

Problem Statement

The goal here is to implement a Convolutional Neural Network (CNN) model to be used on a subset of the CalibA dataset and successfully identify (with acceptable accuracy) if the person in the image is Smiling or not. Initially a simple CNN model would be created from scratch (this will be considered as a benchmark model). The goal would then be to solve the problem using a better model with improved accuracy. This would be achieved by either using pre-trained models (Transfer Learning) and/or augmentation techniques.

Datasets and Input

CelebFaces Attributes Dataset (CelebA) is a large scale face attributes dataset with more than 200,000 celebrity images, each with 40 attribute annotations. The images here cover large pose variations and background clutter. This data was originally collected by the researchers at MMLAB, The Chinese University of Hong Kong. Description of the dataset from Kaggle: https://www.kaggle.com/jessicali9530/celeba-dataset

Content

Overall

- 202,599 number of face images of various celebrities
- 10,177 unique identities, but names of identities are not given
- 40 binary attribute annotations per image
- 5 landmark locations

Data Files

- img_align_celeba.zip: All the face images, cropped and aligned
- list_eval_partition.csv: Recommended partitioning of images into training, validation, testing sets. Images 1-162770 are training, 162771-182637 are validation, 182638-202599 are testing
- list_bbox_celeba.csv: Bounding box information for each image. "x_1" and "y_1" represent the upper left point coordinate of bounding box. "width" and "height" represent the width and height of bounding box
- list_landmarks_align_celeba.csv: Image landmarks and their respective coordinates. There are 5 landmarks: left eye, right eye, nose, left mouth, right mouth
- list_attr_celeba.csv: Attribute labels for each image. There are 40 attributes. "1" represents positive while "-1" represents negative

In this project, a comparatively smaller subset of the dataset is being considered.
Eg. Training - 10,000
Validation - 2500
Test - 2500
All the images here are of the dimension 178 x 218. These images need to be rescaled by dividing every pixel in every image by 255.Each is a color image and hence interpreted by the computer as an array with a depth of 3.

Benchmark Model

A simple CNN model from scratch would be developed. This would have a few iterations of conv2D layers along with MaxPooling, Dropout layers. The final output layer will have few fully connected dense layers with a softmax function. The accuracy of this model would be used as a benchmark. The solution model would aim to achieve an accuracy greater than this.

Solution Statement

Usage of available pre-trained models like ResNet-50, VGG-19, Inception-V3 (Transfer learning) and/or using Image Augmentation methods would help in increasing the accuracy of the model. I will try a few pre-trained models and choose the one that gives good result. Using these, the CNN model thus trained would be able to satisfactorily detect if the person in the given image is Smiling or not with acceptable accuracy.

Evaluation Metrics

Accuracy of the model can be considered as a good evaluation metric to quantify the performance of both the benchmark model and the solution model that would be worked upon. Accuracy is a good metric when classifiers are considered.
This can be defined as:
Accuracy = (Correctly classified points) / Total Number of Points i.e.
Accuracy = (Number of True Positives + Number of True Negatives) / (Total number of pictures)

Project Design

Inorder to achieve a significant improvement in the model performance, the following need to be considered:

1. Transfer Learning: Any suitable pre-trained model like ResNet-50 / VGG-19 / Inception-V3 can be included to improve accuracy. These pre-trained models have learnt over comparatively bigger datasets and the lower layers can easily identify colors, edges and similar other features from a given image. However, the last few layers are of no use to us and hence would be removed. This would then include few more Conv2D and/or MaxPooling/GAP layers with final fully connected layer(s) with appropriate nodes (here 2) and softmax activation function. These would depend on the training to achieve a good accuracy.
2. Data Augmentation: This increases the robustness of the model and thereby accuracy.
3. Fine tuning the hyper parameters.

With the above, the model should be compiled and fit after which it is evaluated (model is trained for an optimum number of epochs ensuring that it is not resulting in overfitting). After this, the model is evaluated to get the accuracy.

References

1. http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

Citation:

```
@inproceedings{liu2015faceattributes,
 title = {Deep Learning Face Attributes in the Wild},
 author = {Liu, Ziwei and Luo, Ping and Wang, Xiaogang and Tang, Xiaoou},
 booktitle = {Proceedings of International Conference on Computer Vision (ICCV)},
 month = {December},
 year = {2015}
}
```


Acknowledgements

Original data and banner image source came from http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html
As mentioned on the website, the CelebA dataset is available for non-commercial research purposes only. For specifics please refer to the website.

The creators of this dataset wrote the following paper employing CelebA for face detection:

S. Yang, P. Luo, C. C. Loy, and X. Tang, "From Facial Parts Responses to Face Detection: A Deep Learning Approach", in IEEE International Conference on Computer Vision (ICCV), 2015