Sammy Sasaki
Duc Anh Than

Multi-Label Facial Attribute Classification

Introduction

Facial attribute classification is a key area of research in computer vision and has various real-world applications including face recognition, emotion detection, and demographic studies. It involves predicting the presence of different attributes in a facial image, which could be anything from hair color, presence of glasses, baldness, attractiveness, and much more.

Our system is based on a Convolutional Neural Network (CNN) architecture, which takes an input image and outputs a set of attribute predictions. The CNN is trained on a large-scale dataset of labeled facial images, and we use transfer learning to fine-tune the pre-trained network for our specific task.

Previous Solutions

Previous solutions for facial attribute classification largely relied on hand-crafted features and traditional machine learning algorithms. These methods often failed to capture the complexity of human faces, and required substantial effort to design effective features. With the advent of deep learning, CNNs have been widely adopted for this task. They automatically learn relevant features directly from image data and have been proven to outperform traditional methods significantly.

Dataset

The dataset used in the project is the CelebA dataset, a large-scale face attributes dataset with more than 200K celebrity images, each with 40 attribute annotations. The images in this dataset cover large pose variations and background clutter. Each image is annotated with 40 binary attributes indicating the presence(1) or absence(-1) of specific facial features. The attributes in the set are:

5_o_Clock_Shadow, Arched_Eyebrows, Attractive, Bags_Under_Eyes, Bald, Bangs, Big_Lips, Big_Nose, Black_Hair, Blond_Hair, Blurry, Brown_Hair, Bushy_Eyebrows, Chubby, Double_Chin, Eyeglasses, Goatee, Gray_Hair, Heavy_Makeup, High_Cheekbones, Male, Mouth_Slightly_Open, Mustache, Narrow_Eyes, No_Beard, Oval_Face, Pale_Skin, Pointy_Nose, Receding_Hairline, Rosy_Cheeks, Sideburns, Smiling, Straight_Hair, Wavy_Hair, Wearing_Earrings, Wearing_Hat, Wearing_Lipstick, Wearing_Necklace, Wearing_Necktie, Young

We used the first 5000 photographs and corresponding annotations to create training, validation, and test datasets.

Proposed Method

The proposed solution builds upon the existing CNN architecture to perform multi-label facial attribute classification. It involves several preprocessing steps including image resizing, normalization, and splitting the dataset into training, validation, and testing sets. The CNN architecture is then chosen and adjusted for the task. The network is trained using an appropriate loss function and optimizer. Hyperparameters such as learning rate, number of layers, and number of filters are then tuned to improve performance on the validation set. Furthermore, different architectures including a simple CNN model, a CNN model for a single attribute, and a more complex model using EfficientNet are explored.

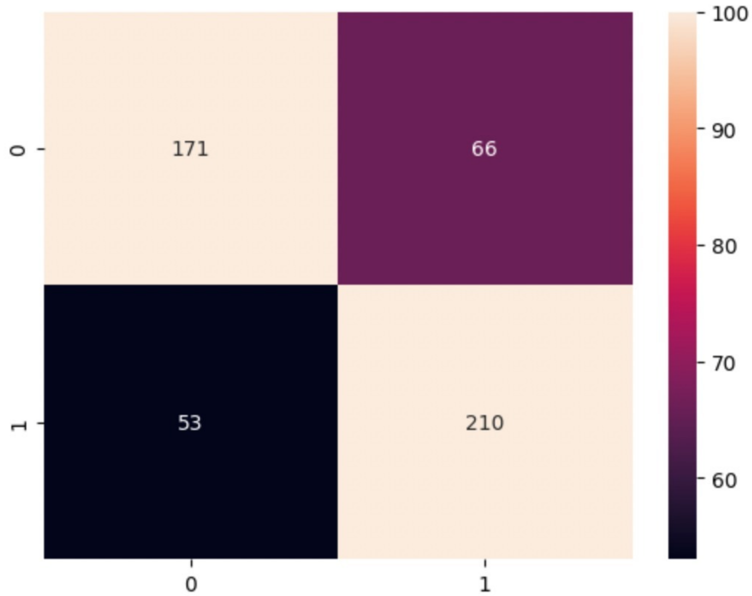
Evaluation Method

The performance of the CNN models is evaluated on the test set mainly using the metrics of test accuracy and test loss. We also use confusion matrices and heatmaps to visualize the performance of each model on each attribute.

Results and Discussion

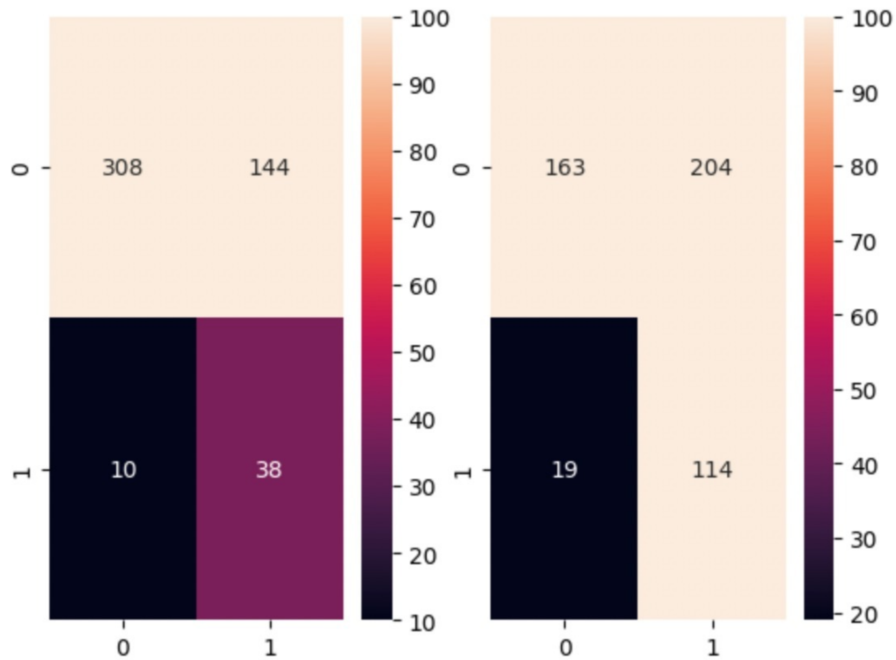
The results of the experiment show that the CNN models are capable of learning to identify various facial attributes with reasonable accuracy. Our first model was trained and tested on only 1 attribute: Attractive. This was chosen because our set had a reasonable split in the data between positive and negative values for this trait. The single-attribute model performed well on the attractiveness attribute, giving a test accuracy of 0.762 and giving the below confusion matrix.

```
Test loss: 0.5017677545547485
Test accuracy: 0.7620000243186951
<Axes: >
```



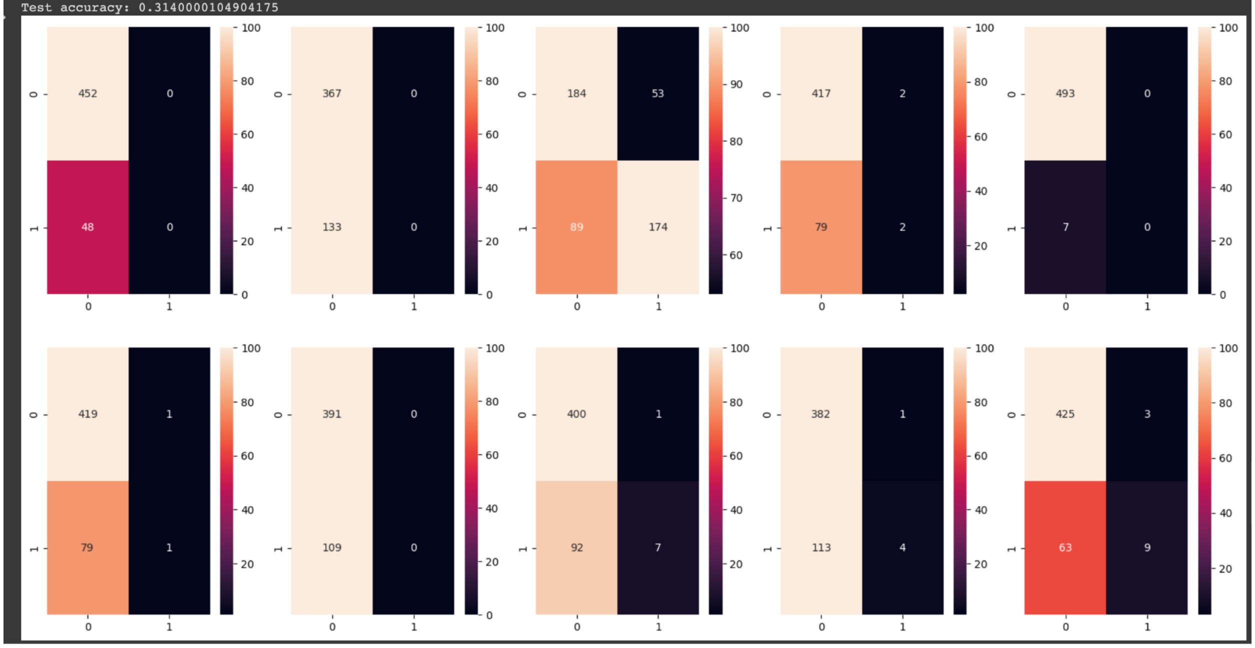
We also trained a model on two attributes, this time using 5_o_Clock_Shadow and Arched_Eyebrows. This model performed slightly worse than the 1-attribute model, giving a test accuracy of 0.559 and producing the following confusion matrix.

```
16/16 [==============================] - 0s 18ms/step - loss: 0.4729 - accuracy: 0.5560
Test loss: 0.4728894829750061
Test accuracy: 0.5559999942779541
<Axes: >
```



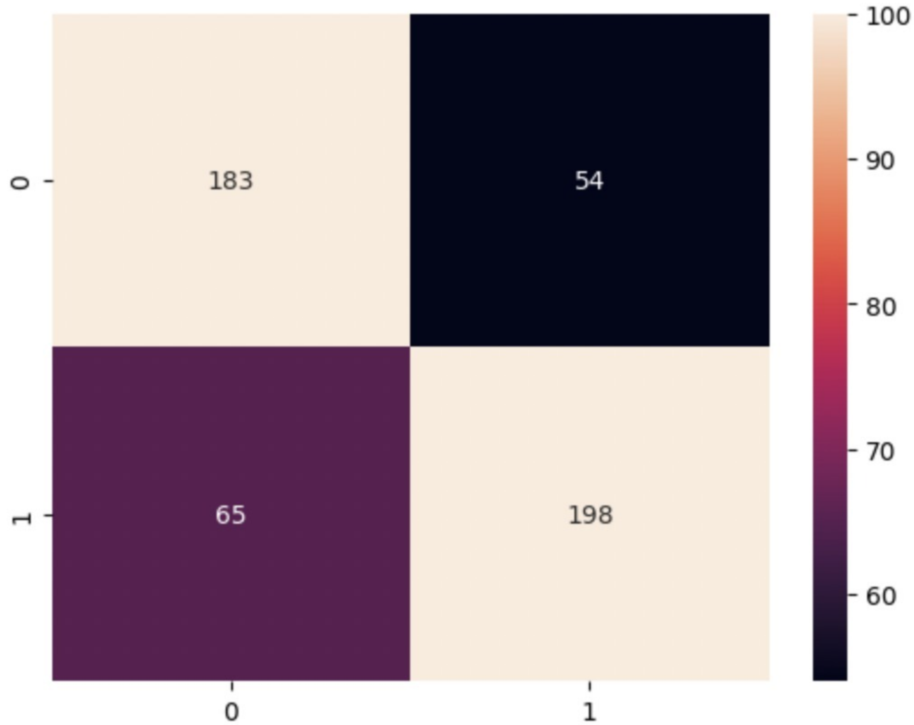
Lastly, we trained a model on 10 attributes, this time simply selecting the first 10 attributes listed by the CelebA dataset. This model did not perform very well, giving a test accuracy of only 0.314. However, this can be expected, since it is attempting to predict 10 traits simultaneously.

```
Test loss: 0.3458256125450134
Test accuracy: 0.3140000104904175
```



We also explored the use of the MobileNetV2 model on 1 attribute, which showed fairly similar results to the CNN models. The architecture of MobileNetV2 consists of an initial fully convolution layer followed by 17 residual bottleneck layers and a final convolution layer. It also uses depthwise separable convolutions, similar to MobileNetV1, which significantly reduces the model size and computation compared to standard convolutions. (Sandler et al)

`<Axes: >`



Work Cited

Sandler, Mark, et al. https://arxiv.org/abs/1801.04381.