Facial Attribute Classification

Image Processing, CNN with Transfer Learning, and Evaluation Methods

Phoebe Jeske &
Lila Schisgal

Data Processing

Initially:

- Loaded each image
- Standardized around training mean
- Reshaped to 1D array

Final model:

- Used three ImageDataGenerators for train, validation, and test
- Rescaled from 255 to 0-1
- These generated batches of images with specified attributes and labels

Building the Model

Model from Scratch

- Conv2D, Flatten, Dense, and Dropout Layers
- Fully trained: test accuracy of 0.0257, weighted average precision of 0.40, recall of 0.34, and f-1 score of 0.36.

Pros: More customizable, predictions centered around a similar threshold for each attribute

Transfer Learning

- Transfer Learning using ResNet50
- Froze the ResNet50 layers and then fine-tuned
- Ran for five epochs, fine-tuned for five epochs
- test accuracy of 0.009, weighted average precision of 0.63, recall of 0.30, and f-1 score of 0.39 (however, not descriptive because of variety in thresholds)

Pros: Less training time, higher accuracy, less data necessary

Evaluating the Model

Sklearn.classification_report

- We were able to see and compare the precision, f-1, and recall, as well as the difference between the weighted and unweighted averages

Confusion matrices

- We generated confusion matrices for each of the 40 attributes being looked at by our model. This allowed us to determine one of the issues with our accuracy was a tendency to predict all positive or all negative based on the attribute’s threshold

Evaluating the Models

Sklearn.classification_report

- We were able to see and compare the precision, f-1, and recall, as well as the difference between the weighted and unweighted averages – pretrained had more 0s.

Mustache	[illegible]	0.01	0.01	772
Narrow_Eyes	0.00	0.00	0.00	2968
No_Beard	0.85	0.81	0.83	17041
Oval_Face	0.28	0.02	0.04	5901
Pale_Skin	0.00	0.00	0.00	840
Pointy_Nose	0.26	0.03	0.05	5704
Receding_Hairline	0.00	0.00	0.00	1694
Rosy_Cheeks	0.06	0.00	0.00	1432
Sideburns	0.04	0.01	0.02	926
Smiling	0.51	0.44	0.47	9987
Straight_Hair	0.00	0.00	0.00	4190
Wavy_Hair	0.36	0.17	0.24	7267
Wearing_Earrings	0.22	0.06	0.09	4125
Wearing_Hat	0.03	0.00	0.01	839
Wearing_Lipstick	0.52	0.46	0.49	10418
Wearing_Necklace	1.00	0.00	0.00	2753
Wearing_Necktie	0.07	0.02	0.03	1399
Young	0.76	0.84	0.80	15114
micro avg	0.52	0.32	0.40	184653
macro avg	0.26	0.14	0.16	184653
weighted avg	0.41	0.32	0.34	184653
samples avg	0.54	0.33	0.38	184653

Confusion matrices

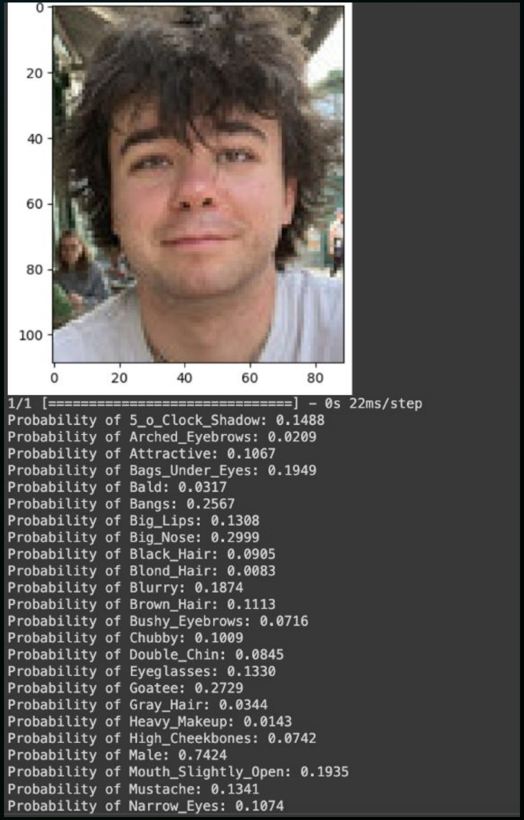
- We generated confusion matrices for each of the 40 attributes being looked at by our model. This allowed us to determine one of the issues with our accuracy was a tendency to predict all positive or all negative based on the attribute's threshold

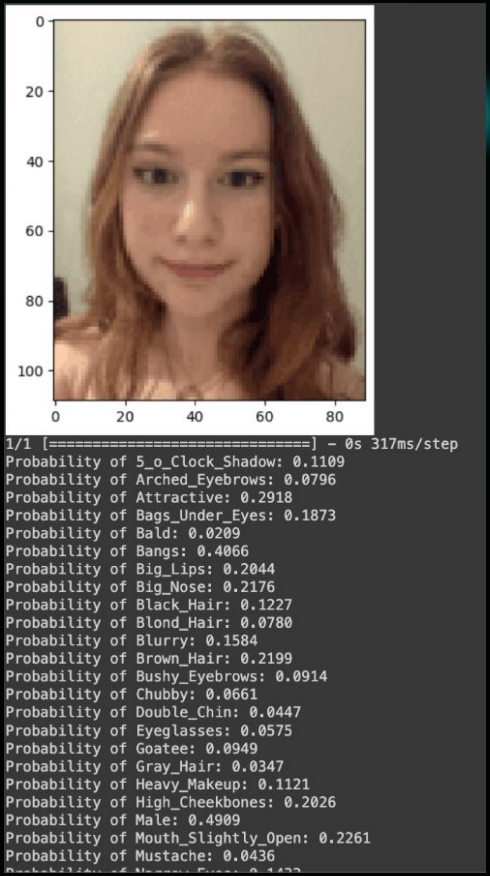
Mustache	0.00	0.00	0.00	39
Narrow_Eyes	0.00	0.00	0.00	146
No_Beard	0.87	1.00	0.93	867
Oval_Face	0.00	0.00	0.00	286
Pale_Skin	0.00	0.00	0.00	36
Pointy_Nose	0.28	1.00	0.43	277
Receding_Hairline	0.00	0.00	0.00	83
Rosy_Cheeks	0.00	0.00	0.00	83
Sideburns	0.00	0.00	0.00	39
Smiling	0.52	0.95	0.67	512
Straight_Hair	0.12	0.00	0.01	213
Wavy_Hair	0.00	0.00	0.00	362
Wearing_Earrings	0.00	0.00	0.00	214
Wearing_Hat	0.00	0.00	0.00	42
Wearing_Lipstick	0.51	1.00	0.67	506
Wearing_Necklace	0.00	0.00	0.00	131
Wearing_Necktie	0.00	0.00	0.00	61
Young	0.74	0.45	0.56	761
micro avg	0.35	0.32	0.33	9328
macro avg	0.12	0.21	0.10	9328
weighted avg	0.29	0.32	0.24	9328
samples avg	0.35	0.34	0.33	9328

Evaluating the Model (Part 2)

Custom Images:

By using custom images, we were able to get a better visual sense of how our model was performing, as well as the different thresholds. For example, aspects such as hair color always had low probability, whereas aspects like male always had high probability.





Discussion

If we had more time and resources:

We would run each of our models to full completion and potentially increase the number of layers.

We would try out other pretrained models to find one better suited to our data.