Hanoi University of Science and Technology
School of Information and Communication Technology

Facial Beauty Prediction

IT3910E - 709162 Project I

Tran Quoc Lap
20194443
Data Science & AI 01 - K64

Hanoi - 2022

Contents

Introduction

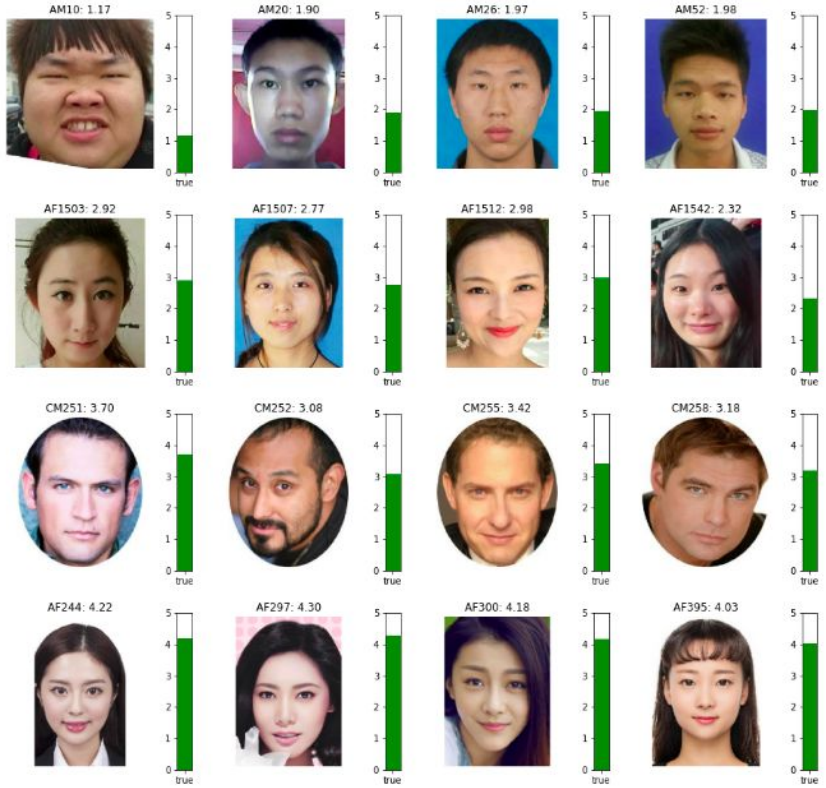


Figure 1. Images and beauty scores from the SCUT-FBP5500 dataset.

Assessing beauty is a natural behavior of humans. But it is also sensitive. So my main focus is how to solve the problem efficiently and effectively.

Applications

- Make-up suggestion
- Digital entertainment
- Content-based image retrieval

Related Work

Xie *et al.* 2015

- SCUT-FBP dataset of 500 Asian females.
- Neutral expressions, simple backgrounds, and minimal occlusion.
- 70 volunteers.

Xie *et al.* 2018

- SCUT-FBP5500 dataset of 5500 images.
- Male/female, Asian/Caucasian.
- Face landmarks
- Neutral expressions, simple backgrounds, and minimal occlusion.
- 60 volunteers.

Data

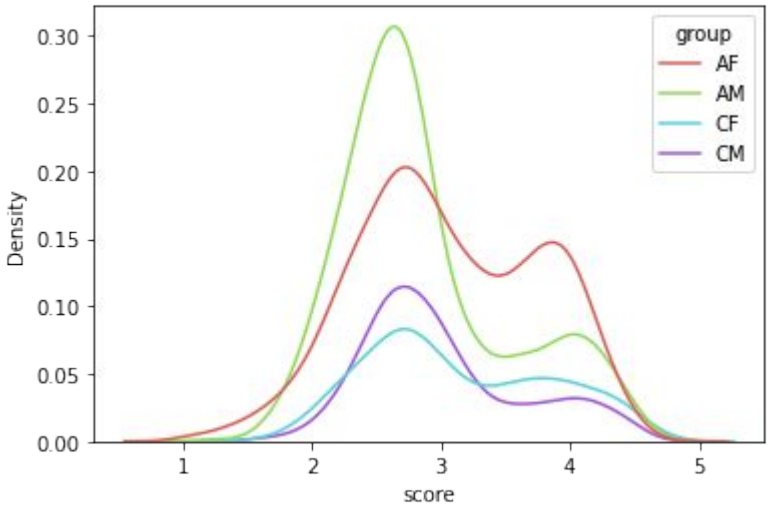


Figure 2. Distribution of Asian females (AF), Asian males (AM), Caucasian females (CF) and Caucasian males (CM).

The SCUT-FBP5500 dataset

- 5500 images.
- Male/female, Asian/Caucasian.
- Face landmarks
- Neutral expressions, simple backgrounds, and minimal occlusion.
- 60 volunteers.
- Image size 350x350.
- Beauty score ranges from 1 to 5.

Ground truth labels might have some noise itself. While labeling, volunteers might be affected by personal taste, culture, momentary emotion, age, etc.

Method

General approach

- General network architecture: famous architecture, eg. MobileNet or GoogleNet.
- Input: RGB color image.
- Train set/test set: 4400/1100.
- Batch-size: 32.
- Loss function: MAE.
- Hyperparameter tuning: search in coarse ranges, then gradually narrow the range depend on where the best results are turning up.
- Reduce the learning rate by tenth if the training loss is on plateau.
- Prevent overfitting: Early stopping and image augmentation.
- Use drop-out to take effect of ensemble learning.

Experiments

Notes

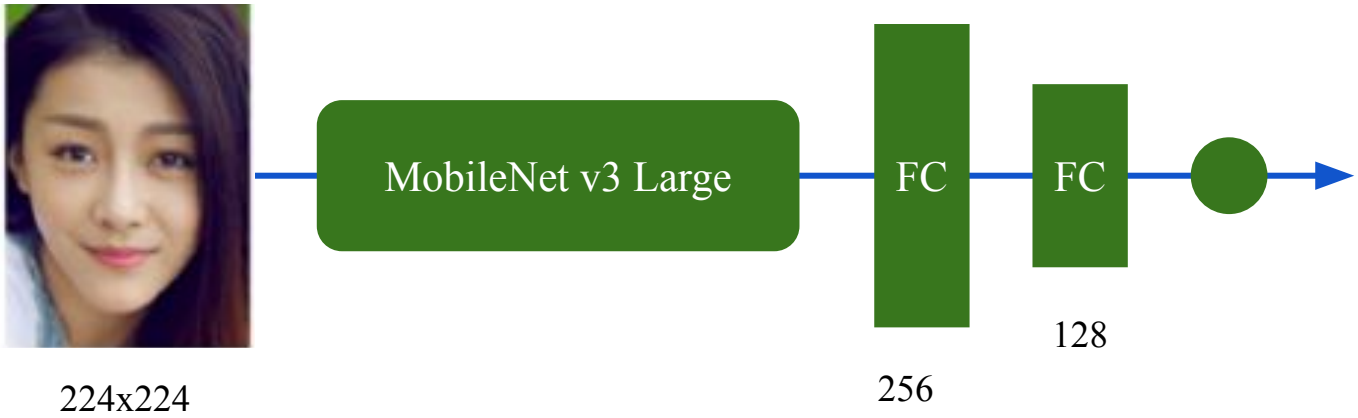
For the first approach, many optimizing algorithms have been used: SGD, Adam, AdaDelta, RMSprop. For each optimizer, the learning rate is initialized with varying values. Generally, there is no significance difference between these optimizers - all of them have the CNN model converge to the same training loss.

So the final configuration for all approaches in my experiments is:

- SGD with Nesterov momentum 0.9.
- Initial learning rate: 5e-3 or 1e-2.

Experiments - The first approach

Network architecture



Experiments - The first approach

Figure 3. Variation of face position and conditions.

The model should be able to handle face of different scales, faces with background clutters, or faces with different viewpoints.

However, original images in SCUT-FBP5500 are faces already cropped and centered, so image augmentations such as zoom or translation are highly important.

Experiments - The first approach

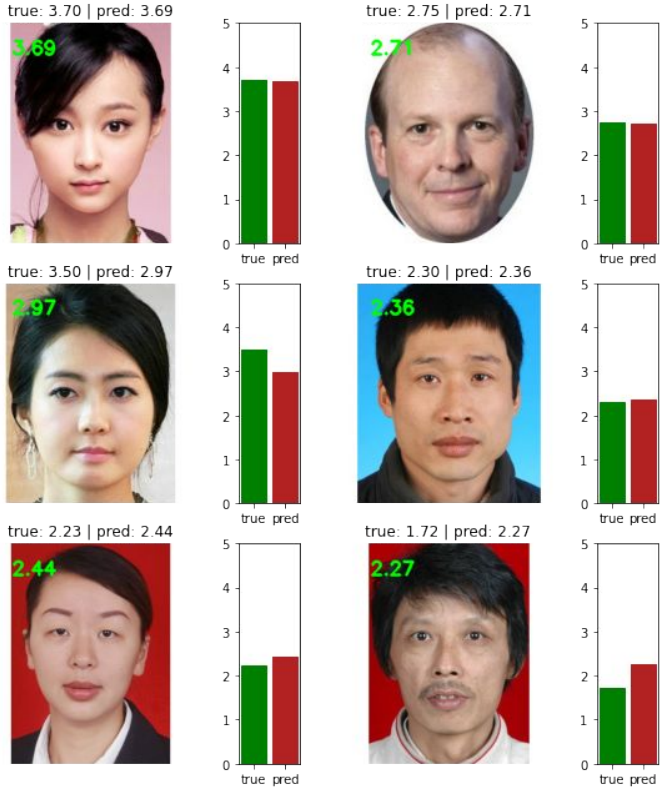


Figure 6. The first approach's prediction on test set.

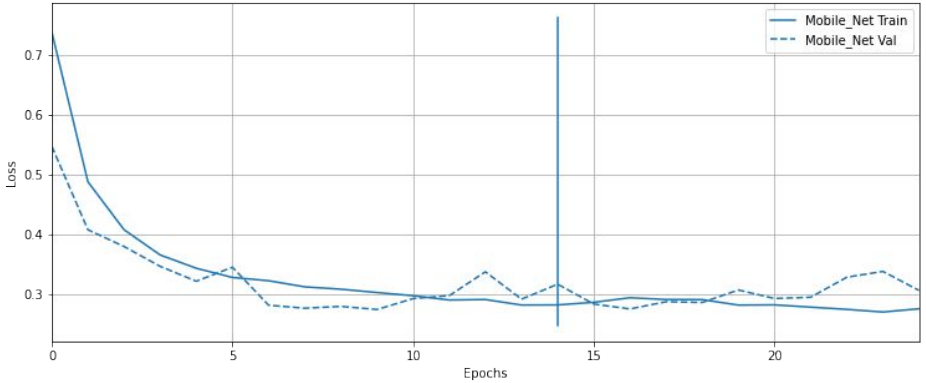


Figure 5. Learning curve of the first model with original image input.

Pearson correlation: 0.87

Experiments - The first approach

- What has the model learnt?
- What does the output score represent?
- That score is a beauty representation of face only or the background?
- Is the output score affected by hair style or emotion expression?

Face is the brightest, so it is in focus. But some patches around the face also get attention. And, image with no face still have a beauty score.

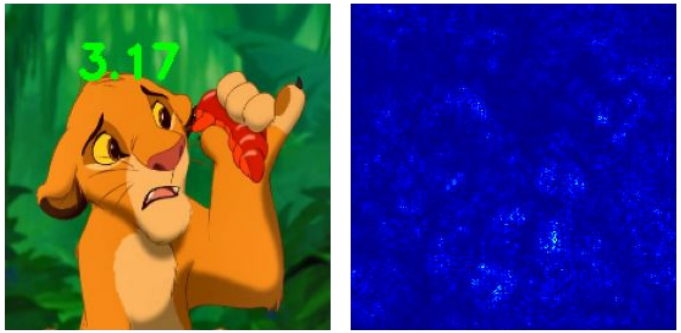


Figure 8. Prediction on a non-face image.

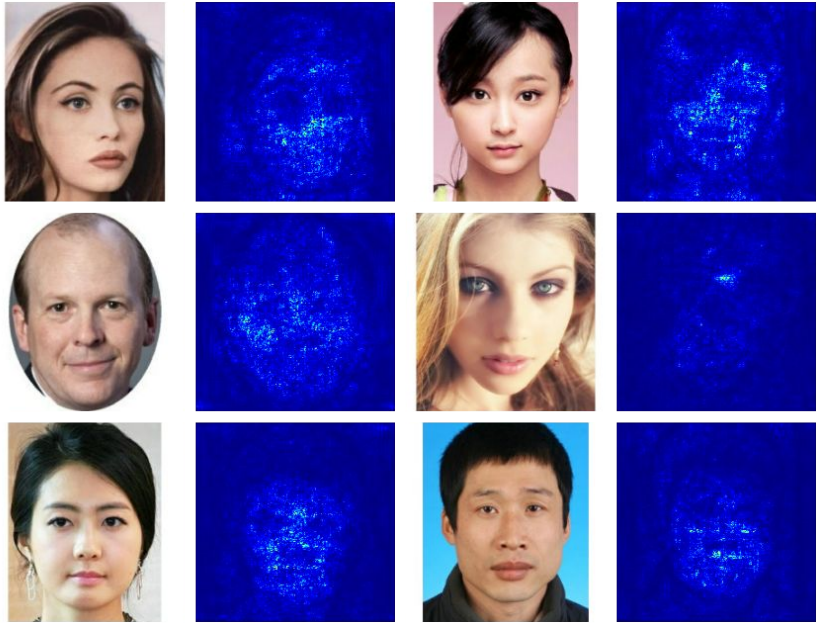


Figure 7. Saliency map on test images. The brighter pixel is, the more attention it receives from the model.

Experiments - The second approach

350x350

Face extractor

Pretrained MTCNN

224x224

MobileNet v3 Large

FC

256

FC

128

Face is extracted before feeding into the network using a pretrained Multi-task Cascade Neural Network, so image with no face is not evaluated.

Pearson correlation: 0.89

Experiments - The second approach

The second approach has achieved better performance. But another problem arises:
When test with different images of the same person, the beauty score varies significantly.

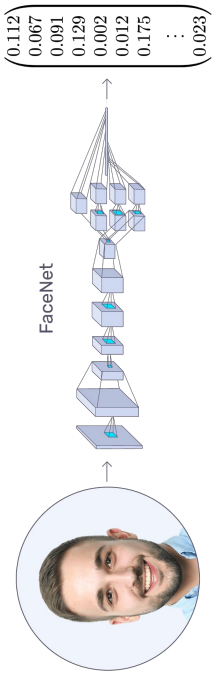
Figure 11. Beauty score prediction of the second approach on the same person.
From left to right: 3.07, 3.11, 2.87, 3.11, 2.52.

The model is not stable. In the dataset, there is no pair of images of the same person.

How can the model learn the consistent characteristics of a face, given only one image per individual available in the training set?

Experiments - The third approach

In 2015, Schroff *et al.* proposed FaceNet that directly learns a mapping from face images to an embedding space where distances between embedding vectors directly correspond to a measure of face similarity. Faces of the same person have small distances, faces of distinct people have large distances.



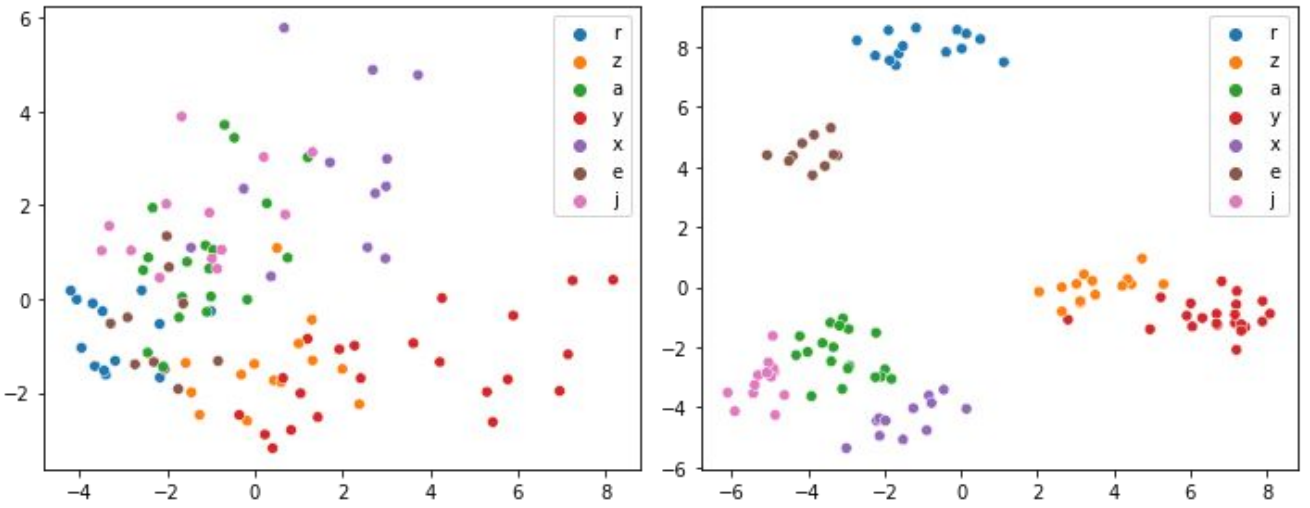


Figure 13. Embedding vectors of face compressed into 2D using PCA. Dots with the same color represent the same person. *Left*: using the last flattened feature map of the second model. *Right*: using FaceNet.

Experiments - The third approach

350x350

Face extractor

Pretrained
MTCNN

160x160

FaceNet

Pretrained,
freezed

128

FC

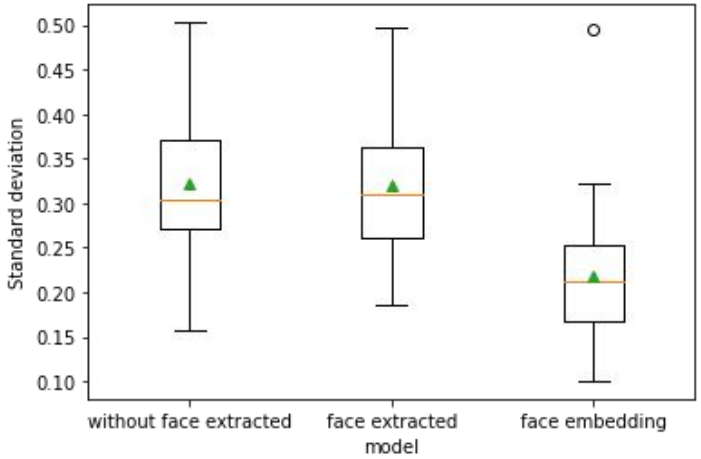
256

Pearson correlation: 0.86

The MAE validation loss (0.26) is not as good as the second model (0.24).

Explanation: FaceNet is not trainable, it constrains the embedding space and forces the rest of the predictor to minimize the loss under constraint. In the meantime, all the weights of the second model are trainable and freely optimized to minimize the loss with no constraint.

Experiments - The third approach



Model pair	T-test p-value	Reject
('without face extracted', 'face extracted')	0.867	False
('face extracted', 'face embedding')	1.12e−6	True
('without face extracted', 'face embedding')	9.95e−7	True

Table 1. T-test result of models' standard deviation.

Collect images of 34 individuals. For each of them, take a sample of 30 images with different poses, emotion expressions, under different lighting conditions.

For each model, and for each individual's image sample, compute the standard deviation of beauty score.

Do a T-test to check if there is any difference between standard deviation of 3 models.

The approach based on face embedding is more stable than the 2 previous ones.

Experiments - Comparison with previous work

Method	PC	MAE
Geometric feature + Gaussian regression (Liang *et al.* [2])	0.67	0.39
ResNeXt-50 (Liang *et al.* [2])	0.88	0.25
Without face extracted (mine)	0.87	0.27
Face extracted (mine)	0.89	0.24
Face embedding (mine)	0.86	0.26

Table 1. Comparison between my approaches and previous works.

Notes

There were many attempts to solve the problem of facial attractiveness. Indeed, the first step in a project is to investigate related works to see state-of-the-art achievements in the field. However, this is my first experience in a specialized project, so I had actually done the research on previous works too late - when I almost finalized everything. At the time of writing, there is no more time to research.

Summary - Conclusion

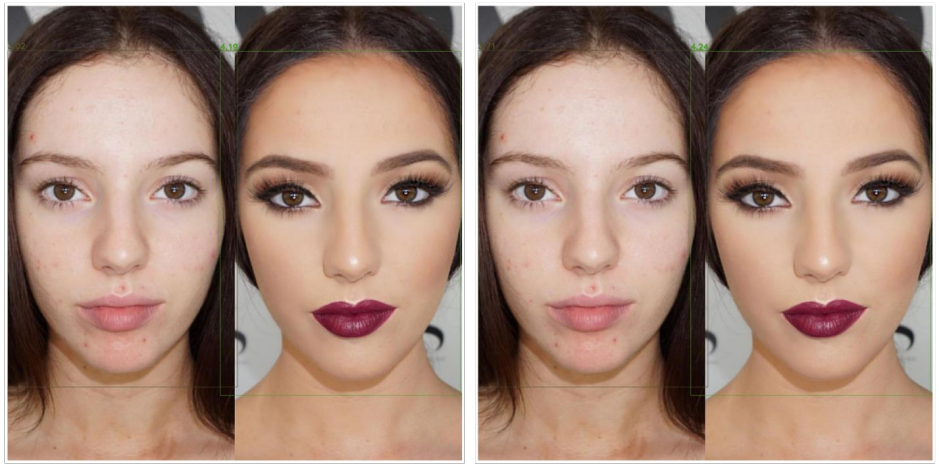
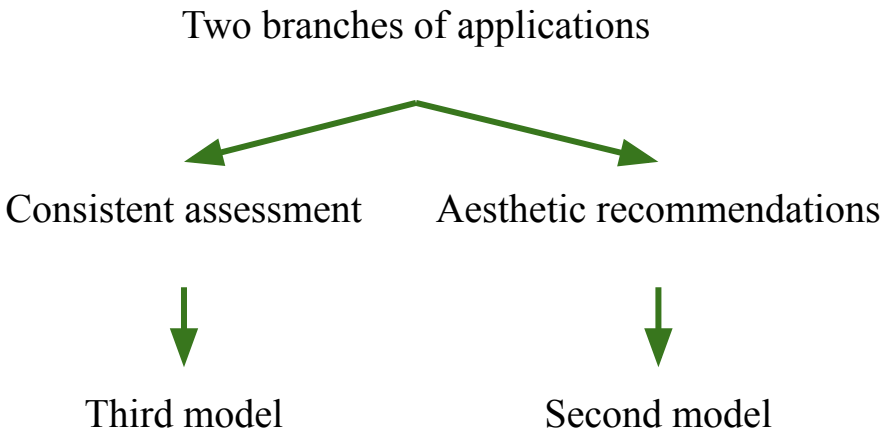


Figure 16. Model prediction after make-up. *Left* (Second model): 3.02 and 4.19. *Right* (Third model): 3.71 and 4.24.



Summary - Improvement

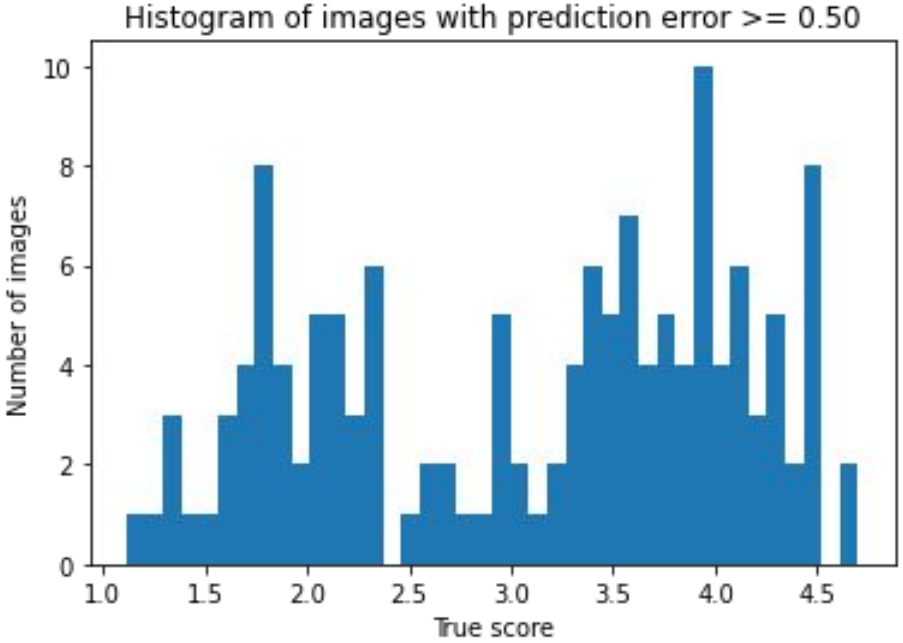


Figure 17. Histogram of prediction error larger than 0.5

- FaceNet was pre-trained mostly on American and British actors. → train further on Asian.
- 60 labelers might exist some bias (age, culture, emotion) → gather more volunteers.
- Use generative models to generate more facial expressions or condition variation of the same person → an alternative for the third approach using FaceNet.
- Do error analysis. Most error occurs around 2 or 4. → collecting more data of these cases, or manually check whether they are wrongly labeled.

References

[1] Andrew Howard et al. “Searching for MobileNetV3”.In: (2019).

[2] Lingyu Liang et al. “SCUT-FBP5500: A Diverse Benchmark Dataset for Multi-Paradigm Facial Beauty Prediction”. In: (2018).

[3] Florian Schroff, Dmitry Kalenichenko, and James Philbin. “FaceNet: A Unified Embedding for Face Recognition and Clustering”. In: (2015).

[4] Duorui Xie et al. “SCUT-FBP: A Benchmark Dataset for Facial Beauty Perception”. In: (2015).[5] Kaipeng Zhang et al. “Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Net-works”. In: (2016).

Thanks for your attention

IT3910E - 709162 Project I
Tran Quoc Lap
20194443
Data Science & AI 01 - K64