AgingGAN: Age Progression with CycleGAN

Jie Chen, Junwen Bu, Yu Zhao

{jiechen8, junwenbu, zhaoyu92}@stanford.edu

Abstract

In this study, we analyzed the feasibility of using CycleGAN for the problem of age progression. We explored using different datasets and using different components of a dataset, and then compared aging effect of these models. We also explored techniques like *hyperparameter tuning*, *fine tuning* and *transfer learning* to speed-up the CycleGAN training process. At the very end, we present the performance comparison of all our models quantitatively and qualitatively.

1. Introduction

The clock never stops, never waits. When we grow older, it is fun to take out the album (or various Photo APPs these days) and show your friends, families and co-workers what you look like when you are younger. Wouldn't it be awesome if you can do the opposite? Age progression, the process of aesthetically rendering a facial image with simulated effect of growing old, has attracted much attention from the *Deep Learning* and *Computer Vision* community, due to its wide range of applications in the entertainment industry and in forensic science, *e.g.*, generating contemporary portraits of individuals who went missing when they were young. However, it remains a challenging task because the patterns of aging we want to capture could be easily affected by the various conditions of the input image, such as facial expressions or photographic settings, not to mention the unexpected effect on people's appearance caused by physical environment that they grow up in. Further, the scarcity of paired data – two images of the same person taken at different time (20+ years apart) -- prevented existing solutions to achieve good performance. In this project, we proposed a simple, yet intuitive deep learning model based on CycleGAN [1] that can generate aging effects on people portrayed in images, without the need of paired data.

2. Related Work

Earlier attempts of age progression mostly employed bottom-up approaches, where people try to understand the effect of aging by studying one specific facial feature, such as the shape of head or wrinkles [2][3]. Recent works took more holistic approaches. Diederik *et al.* [4] used a flow-based generative model trained on high-resolution faces to synthesize realistic images, though the model itself is still complex and requires careful design. Wei *et al.* [5] proposed a recurrent face aging (RFA) framework that is based on a recurrent neural network (RNN). Their model can generate the fine-grained in-between faces across the aging process, yet one drawback of it being the need of many short-term faces of the same person for training. Many other researchers attempted to use *Generative Adversarial Nets (GAN)* [6] of various customization, such as *Conditional GAN* [7] that are better at preserving the identity of the original person, *Contextual GAN* [8] that are better at capturing the gradual changes in face's shape and texture across adjacent age groups, and *GANs with pyramid architecture* [9] that estimates high-level age-specific features at multiple scales. GAN approaches are significantly better than the earlier studies due to simplicity in design, implementation and lower requirements towards training sets.

3. Dataset and Features

We explored two datasets, the *IMDB-WIKI – 500k+ face images with age and gender labels* [10] dataset that contains 460k images from IMDb and 62k images from Wikipedia, and the *Cross-Age Celebrity Dataset (CACD)* [11] dataset that contains 163k+ images of 2,000 celebrities. CycleGAN requires two collections of images for training, therefore our dataset is divided into two groups, one group of people in their 20s, and another group of people in their 50s or 60s. We also limited the number of images between 3,000 to 5,000 for each group, following the setup of the original CycleGAN paper. All images will be re-sized to 256 x 256 by the CycleGAN model.

3.1 Data Processing

For the IMDB-WIKI dataset, we only used the 62k images from Wikipedia, which already provides ample enough images. Using the metadata of the dataset, we could calculate the age of the person in the image when the photo was taken. The dataset provided a face cropped version of all images and a detector score entry called *face_score* in the metadata. We enforced a minimum *face_score* for both groups to ensure the quality of the image (excluding images with no faces or blurry faces). Finally, we removed nearly all the grayscale images in the dataset, as CycleGAN was designed to take in RGB images. We end up with 5,003 images in the young people group and 2,779 images in the elder people group. Finally, we peeled off 5% of these images as test set, while keeping the rest as training set.

For the CACD dataset, we couldn't extract as much useful information from its metadata, so we did not perform much processing. The dataset does not seem to contain any grayscale images, and most of the images were face cropped as well. We randomly selected 2,200 images for each group and peeled off 5% of those as test set, leaving the rest as training set.

4. Methods and Implementations

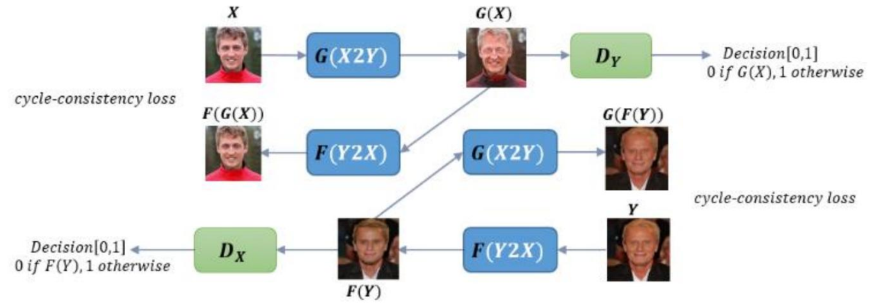

Figure 1. CycleGAN Model

CycleGAN [1] contains two generators $G(X)$, $F(Y)$ and two discriminators D_X, D_Y. Half of the model (G and D_Y) is trained with inputs from domain X, and the other half of the model (F and D_X) is trained with inputs from domain Y. The "cycle" part involves the newly generated $G(X)$ image (now in domain Y) is fed into the generator F and converted back into an image of domain X, and the same process also happen for $F(Y)$ image. Ensuring the generated "cyclic" image is close enough to the original input image guarantees a meaningful mapping is defined, without the need of paired dataset.

4.1 Implementation

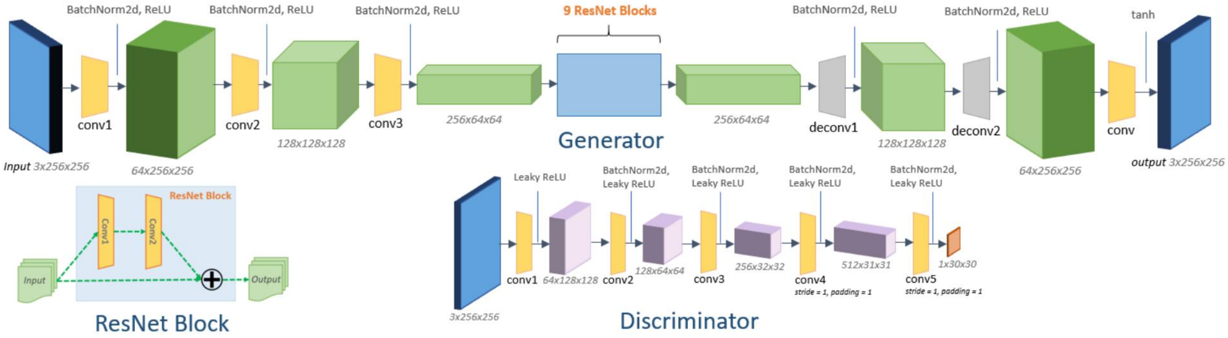

Figure 2. Generator Network, ResNet Block and Discriminator Network

We used the implementation provided by the original CycleGAN paper [12]. The generator network has an **Encoder** (several Conv layers), followed by a **Transformer** (9 ResNet Blocks) and finally a **Decoder** (several Conv Layers). The discriminator is simply a convolutional network contains 5 downsampling layers.

4.2 Formulation

Mappings $G: X \rightarrow Y$ and $F: Y \rightarrow X$, where X represents domain of images of young people and Y represents domain of images of old people.

Adversarial discriminators D_X and D_Y, where D_Y distinguishes between real images of old people $\{y\}$ and generated images of old people $\{G(x)\}$ and D_X distinguishes between real images of young people $\{x\}$ and generated images of young people $\{F(y)\}$.

Adversarial Losses to match distribution of generated images to data distribution.

$$L_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y\sim p_{data}(y)}[logD_Y(y)] + \mathbb{E}_{x\sim p_{data}(x)}[\log(1 - D_Y(G(x))],$$

$$L_{GAN}(F, D_X, Y, X) = \mathbb{E}_{x\sim p_{data}(x)}[logD_X(x)] + \mathbb{E}_{y\sim p_{data}(y)}[\log(1 - D_X(F(y))],$$

Where $x \sim p_{data}(x)$ and $y \sim p_{data}(y)$ represent data distribution of X and Y.

Cycle Consistency Losses to prevent G and F from contradicting each other, in our case, to make sure generated "agedness" images still represent the original people.

$$L_{cyc}(G, F) = \mathbb{E}_{x\sim p_{data}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y\sim p_{data}(y)}[\|G(F(y)) - y\|_1],$$

Objective:

$$G^*, F^* = \arg \min_{G,F} \max_{D_X, D_Y} [L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + \lambda L_{cyc}(G, F)],$$

In practice, the log likelihood objective in LGAN was replaced by a ***least-squares*** loss for easier training. For instance, in $L_{GAN}(G, D, X, Y)$, G was trained to minimize $\mathbb{E}_{x\sim p_{data}(x)}[(D(G(x)) - 1)^2]$ and D was trained to minimize $\mathbb{E}_{y\sim p_{data}(y)}[(D(y) - 1)^2] + \mathbb{E}_{x\sim p_{data}(x)}[D(G(x))^2]$.

5. Experiments and Results

To improve the performance of the model and to reduce training time, we did experiments with different dataset, different data composition and employed deep learning techniques such as transfer learning, fine tuning and hyperparameter tuning (see Table 1).

#	Source	Mix	Epochs	Preloaded?	Freeze until	G Size	Max	Avg	10+	15+	20+
0	CACD	All	200	N/A	N/A	9 blocks	25.8	6.7	22%	5.5%	1.7%
1	WIKI	All	200	N/A	N/A	9 blocks	31.2	8.8	37%	14%	5.8%
2	WIKI	Female	200	N/A	N/A	9 blocks	19.5	4.6	7.1%	2.5%	0.0%
3	WIKI	Male	200	N/A	N/A	9 blocks	27.3	10.3	50%	19%	5.1%
4	WIKI	Male	200	N/A	N/A	6 blocks	N/A	N/A	N/A	N/A	N/A
5	WIKI	All	200	horse2zebra	8th block	9 blocks	27.4	11.0	55%	20%	6.3%
6	WIKI	All	200	summer2winter	8th block	9 blocks	25.0	8.9	36%	10%	1.7%
7	WIKI	All	200	monet2photo	8th block	9 blocks	20.1	6.6	15%	2.5%	0.4%
8	WIKI	Male	100	horse2zebra	N/A	9 blocks	25.8	9.9	46%	12%	1.3%
9	WIKI	Male	100	Model #2	N/A	9 blocks	32.8	10.3	51%	18%	6.0%

Table 1. List of all models we have explored in our study and their quantitative result. The columns are (from left to right): model number, data source, data composition, number of epochs trained, pre-trained network to initialize with, freeze until what layer of generator net, the size of the generator net, maximum age progression (years), average age progression, last 3 columns are the % of test cases where age progression is over 10 years, 15 years and 20 years. Results in model #4 is N/A due to model collapsing.

5.1 Determining the Optimum Data Source

We trained two basic models with the CACD dataset (model #0) and the IMDB-WIKI (model #1) dataset. We discovered that model #1 outperforms model #0 significantly (Figure 3). We manually examined some of the input images of the two datasets, and we think model #0's poor performance is due to the images in the CACD dataset are mostly taken under professional settings (e.g. with makeup, lighting), while those in the IMDB-WIKI dataset are mostly taken under less professional settings. With this finding, we decided to focus on the IMDB-WIKI dataset for the rest of our study.

5.2 Determining the Optimum Data Composition

When examining the test results of model #1, we discovered for some male in the age progressed images have grown feminine traits, such as redder lips (Figure 4), and similarly some females have grown masculine traits, such as mustache or beards. This prompt us into thinking whether we should separate the male dataset with the female dataset. We separated the training set by gender and then we trained two models respectively (model #2 and #3). Indeed, the gender biased traits disappear after the separation, and at the same time we see performance improvements, as the aging effect of model #3 is better that of model #1 (Figure 5).

5.3 Attempts to Speed-up Training Process

One major issue with the CycleGAN model is the training is incredibly slow. We are essentially training 4 separate networks with 28M parameters combined (11.4M for each generator and 2.8M for each discriminator), not to mention that the loss function has an extra cycle consistency part. With 4,630 images in the training set, it takes 15 minutes to finish one epoch of training on a Tesla V100 GPU. To speed-up training, we applied a few deep learning techniques and evaluated their performance.

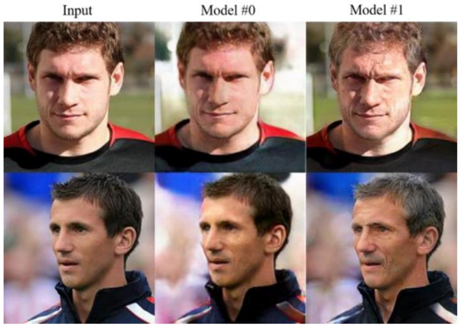

Figure 3. Model #0 and #1 results

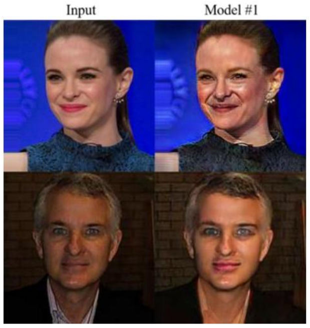

Figure 4. Model #1's gender bias

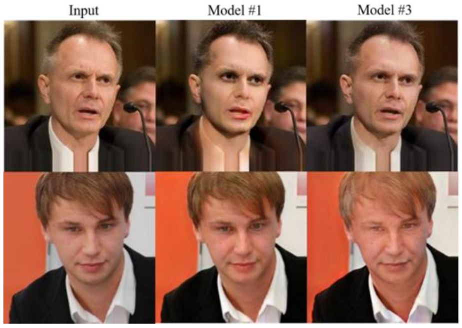

Figure 5. Model #1 and #3 results

We tried ***tuning hyperparameter*** to reduce the network size. The original CycleGAN implementation [12] has 9 ResNet blocks for each of the generator networks. We reduced the number of ResNet blocks to 6, which in turn reduced the number of parameters from 11.4M to 7.8M. With smaller networks, we can increase the mini-batch size to speed-up training. The modified model, however, has poor performance. Reducing the number of ResNet blocks has caused the network to collapse and failed to yield any reasonable result. We discussed about increasing the number of epochs for better result but did not do it because it is against our original goal of speed-up training.

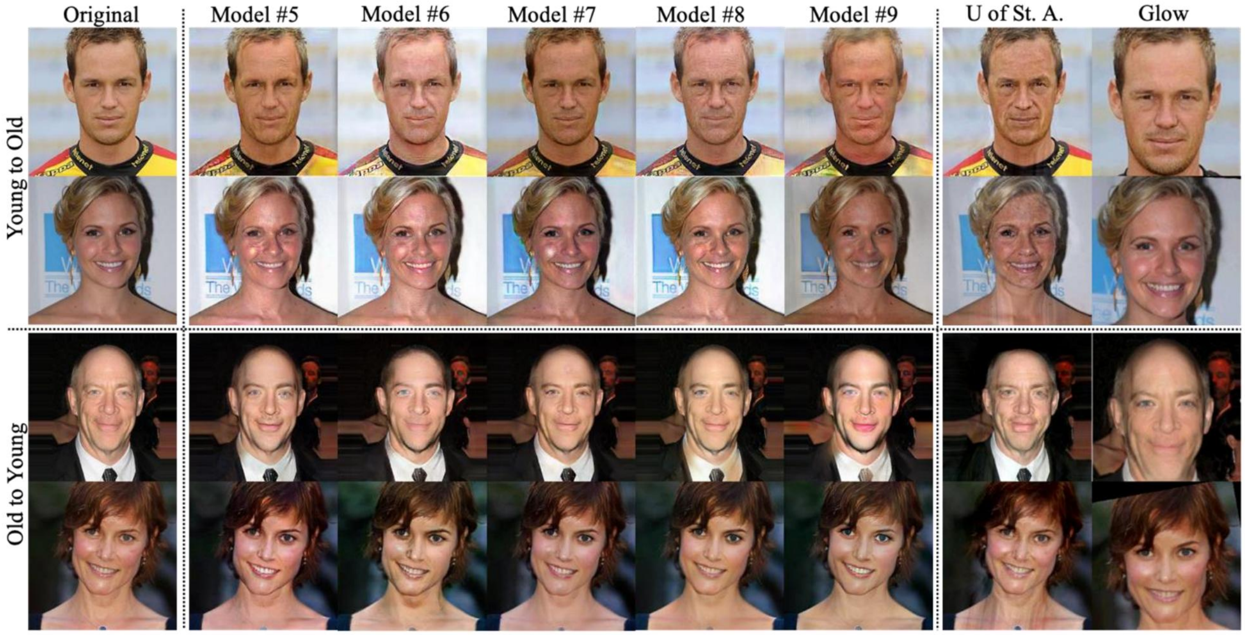

Figure 6. Comparison among results generated by our transfer learning and fine tuning models. Last 2 columns are age progression images generated by models of other academic works.

We tried ***Transfer Learning*** to reduce the number of trainable parameters. For *Transfer Learning*, we first initialize our AgingGAN model with parameters of a pre-trained model, then we freeze all parameters before the 8th ResNet block of the generator networks so that the number of trainable parameters was reduced from 11.4M to 2.7M. When choosing the pre-trained model, we experimented three models in the original CycleGAN paper: *horse2zebra* (model #5), *summer2winter* (model #6) and *monet2photo* (model #7). The models perform differently as expected. Model #5 has the best aging effect, followed by model #6, and then model #7 (Figure 6). We think this is because the *horse2zebra* and the *summer2winter* models captures the texture and shape differences between the image groups, while the monet2photo model failed to do so. Since predicting the skin texture and wrinkles are crucial parts of the aging process, it is expected that model #5 has the best performance.

We also tried ***Fine Tuning*** to reduce the number of epochs. The original CycleGAN implementation [12] requires 200 epochs of learning. We explored reducing this number to 100 while at the same time preserving the same level of aging effect through initialize the AgingGAN model with some pre-trained model. Model #8 was initialized with the *horse2zebra* model and model #9 was initialized with model #3. Both models yield decent results (Figure 6). We see that the parameters pre-trained with images from another domain (*horse2zebra*) do have positive effect in our AgingGAN model, and we can see clear aging effects with only 100 epochs of further training. Model #9 also generated good results, especially the aging effect of female, meaning the parameters of model #3 (pre-trained with male images) are helping the age progression of females in model #9.

5.4 Analysis of Loss Curve

Looking at the loss curve of model #1 (Figure 7), the generator loss and discriminator loss flatten quickly, which is expected as GAN attempts to strike a balance between their losses. However, we can see that the **Cycle Consistency** loss kept decreasing as the epochs increase. In the first 100 iterations, we can see a steady drop in the cycle consistency loss. In the latter 100 iterations, such decrease slows down significantly, but the trend-line is still downwards. Combined with the fact that aging quality improves as epochs increase, we suspect that the aging effect can be predicted with the cycle consistency cost, and the model will cease improving itself when such cost flattens.

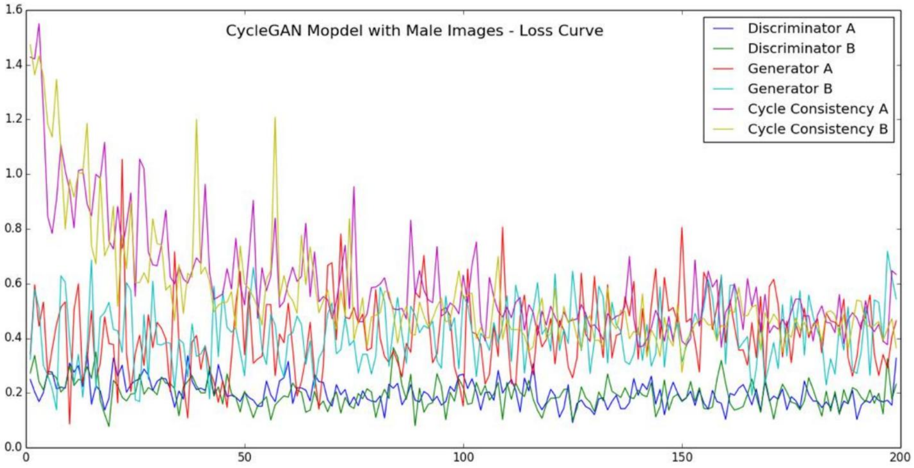

Figure 7. Loss Curve of Model #3

5.5 Results

We evaluated our models both quantitatively and qualitatively. Quantitatively, we generated the estimated age of all images in the test set (last 5 columns of Table 1) using a Keras implementation [13] of the DEX network by Rasmus *et al.* [14]. We can see that model #3, #5 and #9 all have 50+% of people in test cases "growing" more than 10 years and 20% "growing" more than 15 years. Moreover, the maximum aging effect produced by model #9 is a stunning 32.8 years. Qualitatively, we compared our results with two existing baseline models: Face Transformer by University of St. Andrews [15] and Glow [4] (last two columns of Figure 6). For some of the test images, our model achieves better aging effect than baseline models, as the images generated by our model are more realistic.

6. Conclusion

A few things were learned in our age progression project with CycleGAN. The first and foremost finding is that CycleGAN can generate quality age progression images after decent amount of training. The choice of dataset can affect the performance of the model (IMDB-WIKI vs. CACD), so could the composition of the dataset (male vs. female vs. mixed). As the number of training epochs increases, the aging effects increases and the cycle consistency cost drops, but such effect become less and less apparent and the cost flattens in the end. Finally, *Transfer Learning* and *Fine Tuning* with pre-trained models (e.g. *horse2zebra*) can accelerate training process, though come with a slight compromise on the aging effect. Trying to speed-up training by reducing the size of the generator network is futile.

7. Contributions

Jie set up the AWS environment, created the Git repository, processed the data and trained some models. Junwen focused on network modeling, explored the original PyTorch implementation and performed baseline comparison. Yu did the literature review and focused on dataset processing, performed transfer learning and fine tuning. All team members worked together to develop model's architecture, tune hyperparameter, analyze results and write final report.

8. Repository

https://github.com/jiechen2358/FaceAging-by-cycleGAN

9. Acknowledgement

We would like to thank Ahmadreza Momeni for his helpful feedback during office hours. We would also like to thank the CS230 teaching team for their prompt answers to our questions on the Piazza forum. Finally, we would like to than Amazon AWS for providing us with credit and waiving charges that exceeded our credit limit.

10. Reference

[1] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. in ICCV, 2017.

[2] J. T. Todd, L. S. Mark, R. E. Shaw, and J. B. Pittenger. The perception of human growth. Scientific American, 242(2):132, 1980.

[3] Y. Wu, N. M. Thalmann, and D. Thalmann. A plastic-visco-elastic model for wrinkles in facial animation and skin aging. In PG, pages 201–214, 1994.

[4] D. P. Kingma, P. Dhariwal. Glow: Generative Flow with Invertible 1×1 Convolutions. arXiv preprint arXiv:1807.03039, 2018.

[5] W. Wang, Z. Cui, Y. Yan, J. Feng, S. Yan, X. Shu, and N. Sebe. Recurrent face aging. In CVPR, pages 2378–2386, Jun. 2016.

[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In NIPS, pages 2672–2680, Dec. 2014. 1, 3

[7] G. Antipov, M. Baccouche, and J. L. Dugelay. Face aging with conditional generative adversarial networks. arXiv preprint arXiv:1702.01983, 2017.

[8] S. Liu, Y. Sun, D. Zhu, R. Bao, W. Wang, X. Shu, and S. Yan. Face aging with contextual generative adversarial nets. In ACM MM, 2017.

[9] H. Yang, D. Huang, Y. Wang, and A. K. Jain. Learning face age progression: A pyramid architecture of GANs. arXiv preprint arXiv:1711.10352, 2017.

[10] R. Rothe, R. Timofte, L. V. Gool. IMDB-WIKI – 500k+ face images with age and gender labels. https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki.

[11] B.-C. Chen, C.-S. Chen, W. Hsu. Cross Age Reference Coding for Age-Invariant Face Recognition and Retrieval. http://bcsiriuschen.github.io/CARC.

[12] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. CycleGAN and pix2pix in PyTorch. https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix.

[13] Keras implementation of a CNN network for age and gender estimation. https://github.com/yu4u/age-gender-estimation.

[14] R. Rothe, R. Timofte, and L. V. Gool. Dex: Deep expectation of apparent age from a single image. In IEEE International Conference on Computer Vision Workshops (ICCVW), December 2015.

[15] University of St. Andrews: Face Transformer. http://cherry.dcs.aber.ac.uk/Transformer/index.html.