Baby Face Generator with CycleGAN

Shihao Li(lsh0357@stanford.edu)

Abstract

In the study, we explored the possibility of generating baby face with parents pictures using CycleGAN. We tried different datasets and applied transfer learning to improve the performance of our models. We present the generated baby faces with our models at the very end.

1. Introduction

The project's idea comes from Face App, a recently popular application around the world. It generates highly realistic transformations of faces in photographs by using neural networks. This project explores a similar problem as Face App. It takes parents faces or an adult face and generate the baby style face of the morphed parents face or the single adult face. This project mainly focus on the baby face generator part. If the input is only one single face(man or woman), the model will consume the picture and generate a baby face. Else if the input are more than 1 faces, we firstly use http://www.morphthing.com/ to morph parents face, and then generate a baby style face with our model. The challenge part of the problem is that it's hard for us to find a datasets of high quality paired pictures of parents and baby. So the CycleGAN model shines in this situation as it consumes unpaired images. There are lots of high quality images that contains faces from a various of ages.

2. Related Work

We could find several applications in app store that could generate baby faces. But few of them generate customized baby faces. Most applications generate baby faces based on specific features such as hair, skin, eye colors instead of the whole input pictures. One advantage of our model using cycleCAN is that it considers every pixel of the input. It was hard to find baby face generation research papers, but we were able to find some researches about age progression[2]. In the earlier attempts of age progression mostly focused on a specific facial feature, such as the wrinkles[3]. Recent age progression recent starts to take care of the whole faces. A research proposed a recurrent face aging framework that is based on a RNN[4]. This RNN model could generate fine grained faces, yet it has the drawback that needs many short-term faces of the same person. Many other researchers explored to use GAN to solve age progression problems. Conditional GAN is good at preserving the identity of the original person[5]. Contextual GAN is good at capturing the gradual changes in face's shape and texture across adjacent age groups[6]. CycleGAN[1] uses unpaired images for training which solves the problem of finding highly restricted images.

3. Dataset and Data Preprocessing

We uses 2 datasets for training our model. The original one we were using is UTKFace dataset(https://susanqq.github.io/UTKFace/). It is a large-scale face dataset with long age span (range from 0 to 116 years old). The dataset consists of over 20,000 face images with annotations of age and covers gender and ethnicity. This is the origin dataset we were using to train our model. However, the images are not in high quality(many pictures are around 2KB) which leads to poor quality generated images. So we switch to use a face image dataset, IMDB-WIKI - 500k+ face images with age and gender labels, which contains high quality images in our latest training.

CycleGAN requires two collections of images for training. Therefore we write some helper functions to filter the images to form the 2 groups, group X of people in their 0-5 ages, group Y of people in their 27-35 ages. We were using 10000 images for training at first but it was too slow. So we restrict our training set to be around 2000 images. In specific, we use around 1000 images for group X and 1000 images for group Y. All images will be resized to 256 * 256 by the CycleGAN model.

In the original training, we were using the original pictures which contains many background noisy. The performance was pretty poor with these backgrounds such as clothes and gestures. So we extract faces only using OpenCV Face Detection Neural Network from the original images. (https://github.com/kb22/Create-Face-Data-from-Images) and train our model with the cropped faces.

In addition, the project is mainly focus on baby style face generator. So all the test sets take a male and a female images are morphed using http://www.morphthing.com/ manually. To morph two images, you need upload images and choose corresponding points by hand. Then the morphed image are fed into the model. The other test sets take only one image doesn't need to go through the website.

4. Methods

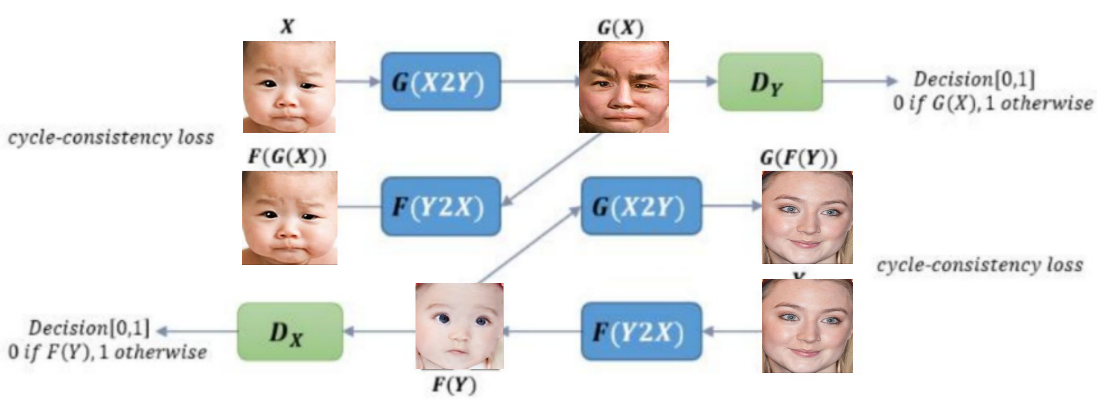


Figure 1 CycleGAN Model

As show in Figure 1, CycleGAN[1] contains two generators G: X -> Y and F : Y->X, and two Discriminators Dx and D_Y . Dx aims to distinguish between images {x} and translated images {F(y)}, in the same way, D_Y aims to distinguish between images {y} and translated images {F(x)}. G tries to generate images from domain X(baby faces dataset) that look similar to images from Domain Y(adult faces dataset), while F tries to generate images from domain Y that look similar to images from X

4.1 Formulation

The goal is to learn mapping functions between two domains X and Y given training sets X and Y. In specific in our project, the goal is to train generator F so as to generate a baby face from an adult face.

Adversarial Loss G aims to minimize the objective against D tries to maximize it.

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

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

Cycle Consistency Loss implies that generators should be able to bring x or y back to the original image, so as to generate a desired output instead if any random permutation of images.

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

Full Objective is

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, Y, X) + \lambda\mathcal{L}_{cyc}(G, F).$$

where λ controls the relative importance of the two objectives. And we aim to solve

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

4.2 Implementation

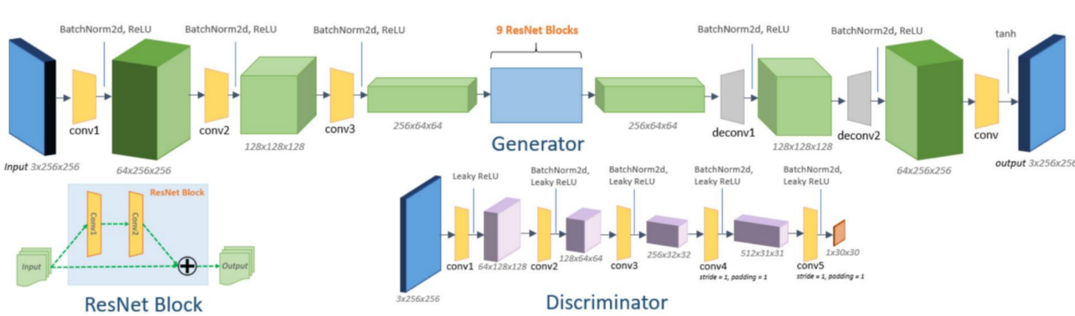


Figure 2: Generator Network, ResNet Block and Discriminator Network[2]

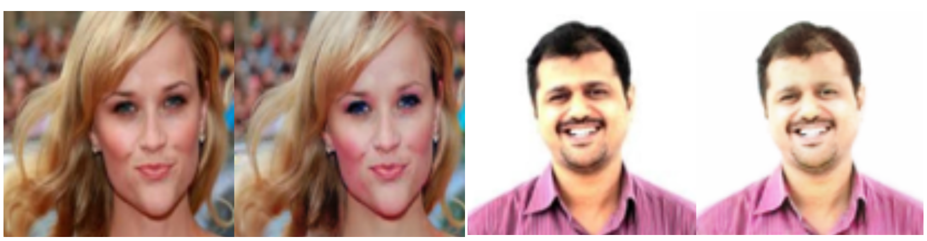
We use the implementation provided by the original CycleGAN paper[1]. The generator has an Encoder followed by a Transformer and finally a Decoder. The discriminator is a CNN contains 5 downsampling layers[2].

5. Experiments and Results

model	source	Epochs	Cropped	Pretrained
1	UTKFace	155	No	No
2	UTKFace	10	Yes	No
3	IMDB-WIKI	100	Yes	No
4	IMDB-WIKI	155	Yes	Yes

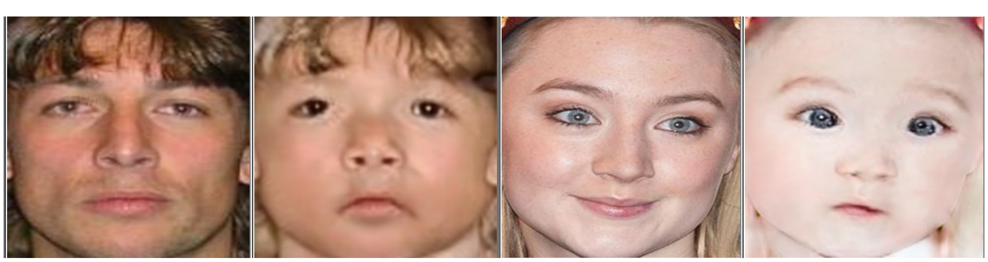
Table 1. Different datasets and other features we've explored.

Model 1 is the base line model we've tried firstly. The performance of the model is pretty poor. We can see that only the background color changes in the test sets. We analyze that this could because of the background noises such as clothes.



Model 2 then is fed with cropped images. But we found that the image quality in UTKFace is not high enough. Especially when we crop the image to only contain faces, the faces is too vague. So we stop the training early.

In **Model 3**, we find the high quality images from IMDB-WIKI dataset and trained 100 epoch for it. In **Model 4**, we use pre trained old2young model[2] and get a satisfactory result. Here are some of the images from the test set and from morphed faces.
Adult face to baby face:



Morphed face to baby face:

6. Conclusion

We were able to learn that CycleGAN could work well in face baby generation after decent amount of training. The quality of dataset and data preprocessing could influence the performance of a machine learning model significantly. Transfer learning from similar model could help with training.

7. Contribution

Shihao Li set up the AWS environment, chose datasets, preprocessed the data and trained the model.

8. Repository

https://github.com/lsh0357/Baby-Face-Generator-with-CycleGAN

9. Acknowledgement

We would like to thank Hao Sheng for providing feedbacks and the useful face extraction tool during TA meetings. And we would also like to thanks all the CS230 teaching staffs for such a fantastic course.

Reference

[1] Jun-Yan Zhu*, Taesung Park*, Phillip Isola, and Alexei A. Efros. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", in IEEE International Conference on Computer Vision (ICCV), 2017.
[2] Jie Chen, Junwen Bu, Yu Zhao. AgingGAN: Age Progression with CycleGAN.
[3] Yin Wu, Nadia Magnenat Thalmann and Daniel. A Plastic-visco-elastic Model For Wrinkles in Facial Animation and Skin Aging.
[4] W.Wang, Z.Cui, Y. Yan, J. Feng, S. Yan, X. Shu and N. Sebe. Recurrent face aging.
[5] G. Antipov, M. Baccouche and J.L. Dugelay. Face aging with conditional generative adversarial networks.
[6] S. Liu, Y. Sun, D. Zhu, R. Bao, W. Wang, X. Shu and S. Yan. Face aging with contextual generative adversarial nets.