Face Aging using Generative Adversarial Network

Guanghao Chen
A53276390
guc001@eng.ucsd.edu

Qimin Chen
A53284263
qic003@ucsd.edu

Yunfan Chen
A53287711
yuc014@ucsd.edu

Ke Xiao
A53283874
k1xiao@ucsd.edu

Abstract

Recent studies have shown remarkable success in image-to-image translation for two domains. We especially focused on using StarGAN and a proposed novel model named AE-cDCGAN. By exploring cDCGAN, we observed that it can generate elder and young faces with random noise vector. By plugging an encoder component to the generator and another MSE term to generator's loss function. Finally, compared with StarGAN, AE-cDCGAN can complete face aging task but look less clear. In otder to generate most reasonable face, we tuned α parameter as 10. For the training time, StarGAN takes 37 mins but AE-cDCGAN takes 26 mins for one epoch, which means our proposed models is a lightweight framework.

1 Introduction

In recent years, generative adversarial network achieve well-performed result. It can be applied into many fields. The initial focus for GAN is to generate data from a noise vector, such as image, music and etc. However, the scope of it is far larger than to be a generative model. For example, it can help a image to complete a domain transfer. Or it also can helps a robot to learn faster in reinforcement learning.

In this final report, our group plans to utilize conditional version of GAN to enable a face to be aging. The state of art is Star-GAN, which can be applied to complete image domain transformation. In this project, we firstly set Star-GAN[1] to complete face aging. Besides, we proposed a modified version of cDCGAN to accept an original image and transform it into elder faces. To be specific, we plugged an encoder component into the generator to accept original image and plugged another mean square error term into loss function of generator. Intuitively, what we expect with this model is to accept an image and preserve the identity.

2 Related Works

The deep generative networks have exhibited a remarkable capability in image generation and have also been investigated for age progression. These approaches render faces with more appealing aging effects and less ghosting artifacts compared to the previous conventional solutions. Specifically, these approaches truly aims to complete image domain translation instead of focusing on face aging. However, the essence of face aging should both preserve the identity and modify the age related characteristics simultaneously. In addition, some of the works works fine for more concrete age stage. For example, Yang[2] proposed a model to generate faces under different ages instead of just young and elder.

1

3 Methodology

3.1 Dataset

CelebFaces Attributes Dataset (CelebA) [3] is a large scale face attributes dataset with more than 200k celebrity images, each with 40 attributes. CelebA has large diversities, large quantities and rich annotations including 10,177 number of identities, 202,599 number of face images and 5 landmark locations, 40 binary attributes per image.

For the face aging task, we specifically focus on age attribute which will be the additional input to the generative adversarial network so that we can guide the network to generate images of the same identity in different ages. Due to the time limitation, we will not train the CDC-GAN and StarGAN on the whole CelebA dataset instead using a portion of it.

Figure 1 shows some sample images from CelebA dataset. We will crop the initial 218×178 size images to 178×178 then resize images as 128×128.

Figure 1: Sample images from CelebA dataset.

3.2 cDCGAN

In the original work, GAN is able to gradually utilized to generate clear images. However, both vanilla GAN and DCGAN are not directed by the user whereas it generates the result randomly. In order to control the content generated, the model should be feed with some additional information. In this project, we are attempting to generate face images of a specific age based on the celebA training data set. Mirza et al.[4] introduced a novel way to train generative model known as conditional adversarial networks. Generative adversarial networks can be extended to a conditional model by conditioning both the generator and discriminator on some additional input, for example age in this task. Besides, by designing the generator and discriminator to be deep convolutional architecture extends the cGAN to conditional deep convolutional generative adversarial networks (cDCGAN). In our project, we designed the architecture as the following. Figure 2 and Figure 3 respectively show the architecture of generator and discriminator.
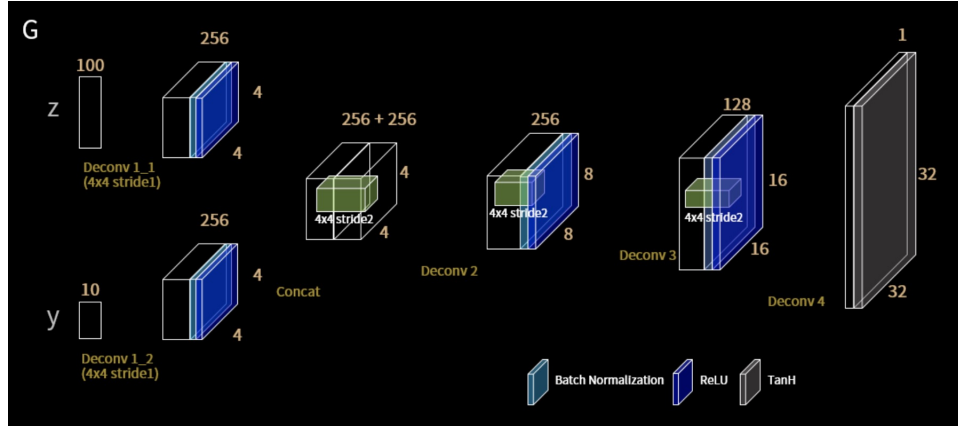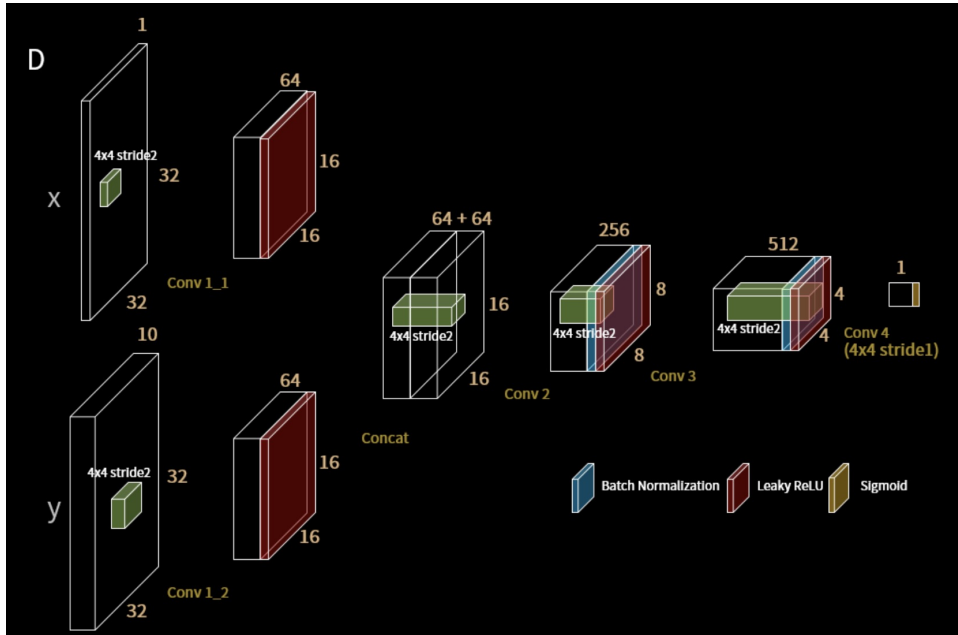
Figure 2: Generator of cDCGAN

Figure 3: Discriminator of cDCGAN

The objective function for GAN model is a minmax function described in Equation (1). We expect both D(x) converging to zero and D(G(z)) converging to one.

$$minmax(D, G) = E_{x\ P(data)}[log(D(x))] + E_{z\ p(z)}[log(1 - D(G(z)))] \qquad (1)$$

However, the input of cDCGAN is a random noise vector and the output is both the elder faces and young faces in this situation, which is different from what we plans to complete in this project. Therefore, we proposed a modified version of cDCGAN, named AE-cDCGAN. The architecture of AE-cDCGAN shows in Figure 4. The motivation of this architecture is to utilize an encoder to encode an input image. Besides, one of the problem for this proposed model is identity preservation, which means the generated face should be the same identity as the original person. In order to achieve this goal, another auto encoder loss function term was plugged into the Generator loss function like Equation 2.
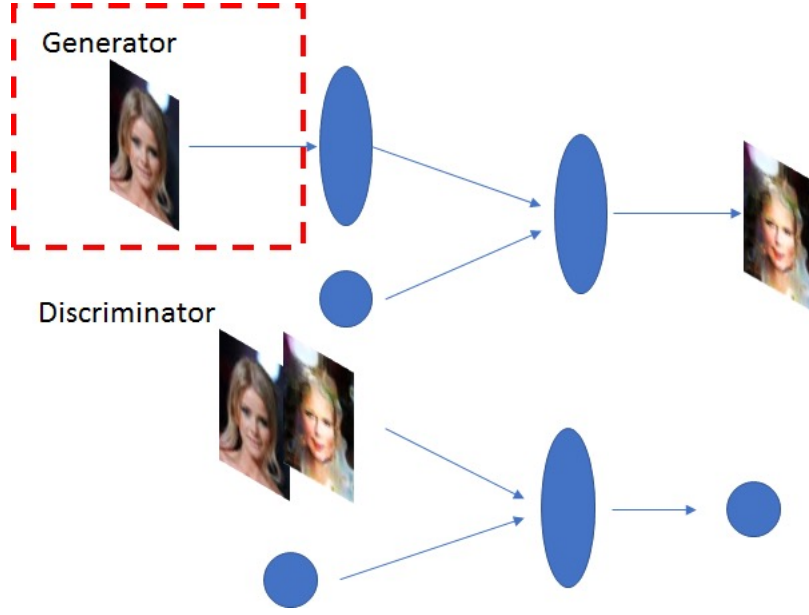
3

Figure 4: Architecture of AE-cDCGAN

$$L = L_G + L_D$$
$$L_D = BCELoss(D(x_T, y), 1) + BCELoss(D(G(x, y), y), 0) \quad (2)$$
$$L_G = BCELoss(D(G(x, y), y), 1) + \alpha MSE(x, G(x, y))$$

3.3 StarGAN

Y. Choi et al. [1] introduce a novel and scalable approach that can perform image-to-image translations for multiple domains using only a single model.

Figure 5 shows the overall structure of StarGAN which consists of two modules, a discriminator D and a generator G. Basically, **(a)** D learns to distinguish the real images from fake one and classify the real images to its domain. **(b)** Both target domain and input images are fed into G to generate the fake image. **(c)** G tries to reconstruct the original images by using the fake image and the original domain of real image. textbf(d) G tries to generate images that are indistinguishable from the real images and the domain classified by D.
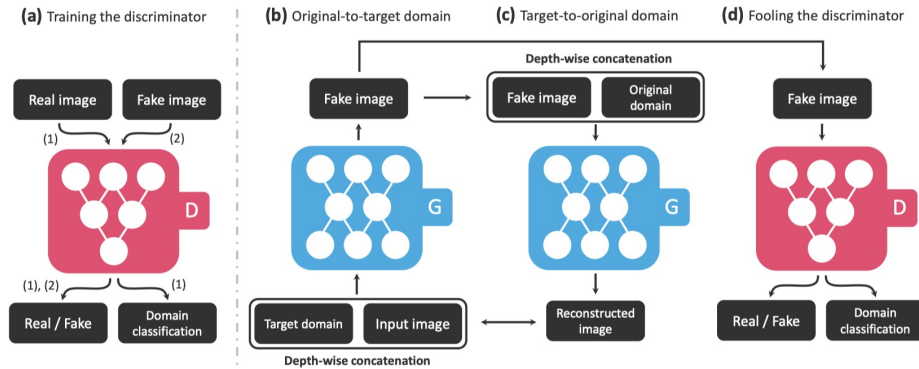

Figure 5: Overview of StarGAN.

4

4 Experiments and Results

In this section, we will first show the result of cDCGAN and Star-GAN respectively. Further, we will design some experiments to compare the performance of the two models.

4.1 Result of cDCGAN

In the model, we confronted several issues during training the model. First all, in celebA dataset, the fraction of elder faces and young faces are significantly imbalanced, which leads to the generator hard to generate clear faces. To be specific, there are 40k elder faces and 150k younger faces in the celebA dataset. In this situation, the generated images looks fuzzy like Figure 6.
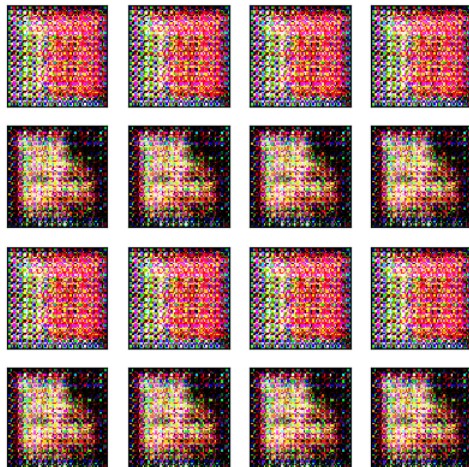

Figure 6: Imbalanced elder faces and young faces lead the generator hard to train.

In order to enable the training set to be balanced and simultaneously utilized as more original as possible to generate clear images. Therefore, we solved this issue with the following procedure showing in Algorithm 1. Figure 11 shows the loss change of cDCGAN. From the figure, we can find

---

Algorithm 1 Training Process

1: randomly choose 40k young faces and combine it with 40k elder faces as one data loader
2: Similarly, we utilized the 40k elder faces again and choose another 40k young faces. Repeat the same procedure again. And finally, we obtained three data loaders which contain 40k elder faces and 40k older faces.
3: During the training, those three data loaders are iterated once in one epoch

---

that the loss of discriminator doesn't change much compared with the generator. At the $epoch = 1$, there's a large drop of the generator's loss, and so the loss of discriminator increases a little bit, which can be explained by the mutual inhibition relationship between them. When the model becomes converge, the loss of both generator and discriminator will not change.
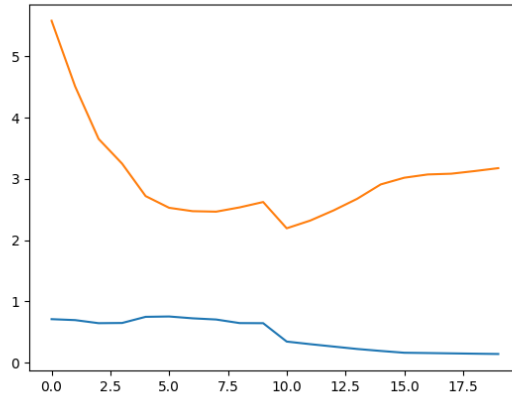
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323

Figure 7: Loss change of cDCGAN over 20 epochs. The orange line shows the loss of generator. The blue line shows the loss of discriminator.

4.1.1 Variation across the epoch

Figure 8 shows the generated young faces and elder faces under different epoch. All of the images in this figure are generated by a fixed random noise. Therefore, all of the faces in the figure are similar. From left to right, the face gradually becomes clear and then becomes blurred, which matches the generator loss curve in Figure 11.
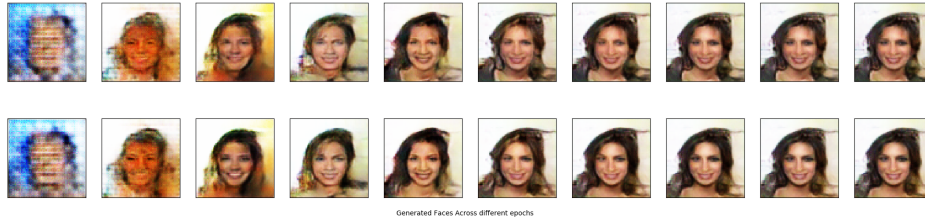

Figure 8: Generated faces under different epochs. The first row represents the result of elder faces under different epochs. The second row represents the generated young faces under different epochs.

4.1.2 Generated faces across different identities

Figure 9 shows the generated faces across different fixed noises which stands for different identities. The first row represents the elder faces and the second stands for the young faces.
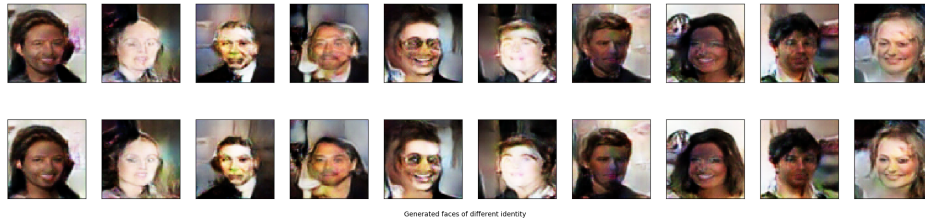

Figure 9: Generated faces under different epochs. The first row represents the result of young faces under different epochs. The second row represents the generated elder faces under different epochs.

6

From Figure 9, we can find several characteristics in our generated elder and young pairs. The first is that the eyes are drooping in elder faces whereas the eyes are big in young faces. Other characteristics we observed is that the skin for the elder is wrinkled and hairs are gray.

4.2 Result of AE-cDCGAN

In this section, we will show the result of AE-cDCGAN. Figure 11 shows the loss curve of AE-cDCGAN. The best-performed hyper parameter in this model determined was shown in Table 1.
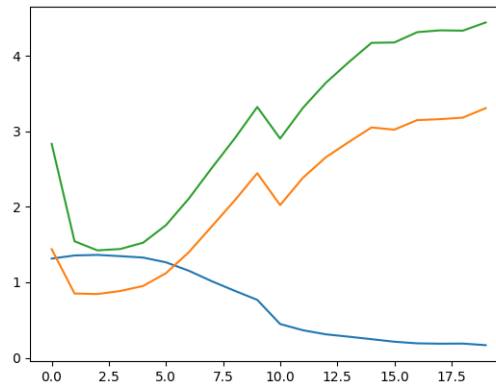

Figure 10: Loss change of AE-cDCGAN over 20 epochs. The orange line shows the loss of generator. The blue line shows the loss of discriminator. The green line shows the auto-encoder loss across the epoch.

Table 1: Hyper Parameter of AE-cDCGAN

Parameter	Value
Batch size	128
Alpha α	10
Learning rate	0.0001

4.2.1 Variation across difference significance of α for AE component

In this part, we will explore the difference of parameter α. Intuitively, when α grows up, the identity will be preserved more, which means the generated faces is more likely to be the same person of the original image. Through the experiment, it is noticeable that the generated faces are more alike with the original image when increasing value of $alpha$. However, when increasing the $alpha$, it is more difficult to make the model converge.

Figure 11: Generated elder faces with different $alpha$ value. The left most image is the original image, and in the right side, they respectively represent $alpha = 11, 12, 15$.

4.2.2 Aged Faces across different identities

Figure 12 shows the aged faces using AE-cDCGAN model. The first row are original faces and the second row are aged faces. Basing this model, it is evident that the identity has been preserved even though the background was not recovered. Besides, we can observe the difference between elder faces and the origin faces. For example, one of the most apparent characteristic is that the skin is wrinkled. Besides, the eyes look more drooped.

Figure 12: Generated elder faces based on AE-cDCGAN. The first row is original faces. The second row is the aging faces.

4.3 Result of StarGAN

Figure 13 shows the loss of discriminator and generator. The StarGAN [1] introduces an auxiliary classifier that helps a single discriminator to control multiple domains in that it decomposes the subjective into two term: one is the domain classification loss for fake images to optimize the generator and the other is the domain classification loss for real images to optimize the discriminator so that both the generator and discriminator are capable of convergence.
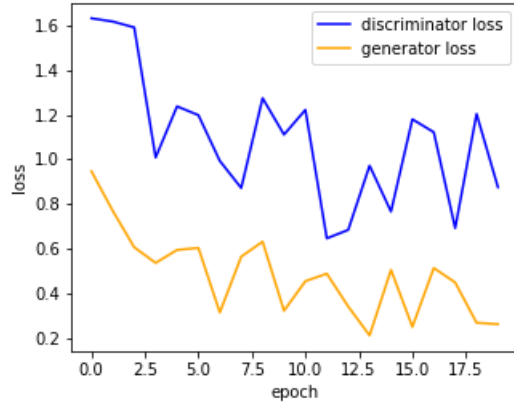
8

Figure 13: Generated faces across different identities. The first row is the original input images, the second row is the elders generation of StarGAN.

4.3.1 Variation across the epoch

Figure 14 shows the old faces under different epochs. The first image is the original input image. It is apparent that the elders images generated by StarGAN are much clearer than cDCGAN. Also the translation from youth to old is stable and natural compared to cDCGAN.

Figure 14: Generated faces under different epochs. First image is the original input.

4.3.2 Generated faces across different identities

Figure 15 shows the generated faces across different input identities. It is obvious that all the old faces have the same characteristics of the elder, including wrinkles, eye morphogenesis and drooping cheek. It can also be seen from the figure that the hair color of the elder is slightly whiter than the young in some identities.

9

Figure 15: Generated faces across different identities. The first row is the original input images, the second row is the elders generation of StarGAN.

4.4 Comparison

Table 2 shows the comparison of StarGAN and AE-cDCGAN on training time, number of images used for training and the quality of output. It takes roughly 37 minutes for StarGAN to train per epoch while AE-cDCGAN takes roughly 26 minutes to train per epoch. Because the imbalanced problem of the images with young/old attribute, we extract roughly $50,000$ young faces randomly to combine with the $50,000$ elder faces, similarly repeat 3 folds hence the AE-cDCGAN takes more resources for training. As for the quality for the generations, the aging faces generated by StarGAN are much clear, natural and realistic while aging faces generated by AE-cDCGAN are less clear and have oil-painting style.

model	training time / epoch	# images	quality
StarGAN	$\sim$ 37 mins	$202,599$	realistic / clear
AE-cDCGAN	$\sim$ 26 mins	$275,190$	oil-painting / less clear

Table 2: comparison between StarGAN and AE-cDCGAN on training time, number of images used for training and the quality of output.

5 Discussion and Conclusion

AE-cDCGAN could achieve the goal of preserve the identity and compared with StarGAN, AE-cDCGAN is more lightweight and could complete the face aging process. However AE-cDCGAN generates less clear and oil-painting style faces because it is hard to converge since the auto encoder is not powerful enough to preserve the identity information as much as possible. In the future work, fine-tuning the model elaborately is an effective approach to improve the performance of AE-cDCGAN. We also want to amplify the signal of additional information so that the disparity of young and old attributes representation could be increased to help the convergence. Besides, cropping the image to remove background noise and only keep the face would also help improve the quality of the images generation.

6 Contribution

Guanghao Chen   Guanghao re implemented cDCGAN based on CelebA dataset. Furtherly, he proposed AE-cDCGAN model and implement it to complete face aging. Finally, he finished part of the report.

Qimin Chen   Qimin replicated the experiment of StarGAN and wrote the corresponding part of report. We discussed every details of the project together.

Yunfan Chen   Yunfan mainly focused on running the experiments and test the model. Yunfan also plotted the results needed for report and presentation.

Ke Xiao   Ke Xiao mainly focused on data pre-processing and made the powperpoint of presentation. We discussed every details of the project together.

References

[1] Y. Choi, M. Choi, M. Kim, J. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," *CoRR*, vol. abs/1711.09020, 2017.

[2] H. Yang, H. Di, Y. Wang, and A. K. Jain, "Learning face age progression: A pyramid architecture of gans," 2017.

[3] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of International Conference on Computer Vision (ICCV)*, 2015.

[4] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," *arXiv e-prints*, p. arXiv:1411.1784, Nov 2014.