Images CycleGAN-based Face Aging Using Unpaired Data

Yanyi Su(36520221151855)[1]
Fanjie Xu(36520221151889) [1]
Yucheng Jin(36520221151807) [1]
Jiaqi Du(36520221151782)(AI)[1]
[1] Major in Materials Engineering, School of Chemistry and Chemical Engineering, Xiamen University

Abstract

The police may need to track down a fugitive or find a lost child, but may only have the photos of them from decades ago; Even with the web of facial recognition systems and cameras now in place, if we don't even know what the faces of the people, we need to find look like, we won't find them. Unless we can do face aging. Face aging and rejuvenation is to predict the face of a person at different ages. While tremendous progress has been made in this topic, there is a central problem remaining largely unsolved: the majority of prior works requires paired training data, which is very rare in real scenarios. To tackle these issues, in this paper, we develop a novel CycleGAN-based (Cycle Generative Adversarial Network) models, which enables face aging and rejuvenation to be trained from multiple sets of unpaired face images with different ages.

Introduction

I remember when Mr.Lu mentioned style transfer in class, he said that the current applications of style transfer are just cool, and sometimes the deep learning field needs these cool things. I didn't really believe it, so I went searching for some real-life applications of style transfer, and found the face aging task.

Face ageing is essentially style transfer, with the identity of the face as the 'content' of the style transfer and the age attribute as the 'style' of the style transfer. The process of ageing is actually the transfer of the 'style' of an older face to the 'content' of a younger face, i.e. the ageing of a younger face while keeping the identity features intact.

In the field of style transfer, tasks are divided into two main categories: those trained with paired data and those trained with unpaired data. In this context, paired data refers to training sets with different styles of the same content, such as images of the same person at different ages; similarly unpaired are training sets with different styles of different content, such as images of different people at different ages.

It can be easy to tell, paired images are very difficult to collect. The capture of day and night images of the same scene requires the camera to remain in a constant position. Images of the same person at different ages require the person collecting the images to follow the volunteers for decades or even centuries. This is why existing paired datasets are often online crawls of images of celebrities at different ages, which admittedly allow for a large amount of paired data to be collected quickly, but all of these images are from celebrities (e.g., Morph(Ricanek and Tesafaye 2006) , Celeb A(Liu et al. 2015), CACD (Chen, Chen, and Hsu 2015)) and the image content is not extensive. This is why the publication of the FGNET dataset (collected from ordinary people, with only a thousand images) has attracted attention and is often used as a test set, even though these huge numbers of celebrity datasets already exist. Of course, it is because the number of images in the FGnet(Lanitis, Taylor, and Cootes 2002) dataset is so small that it can only be used as a test set. Researchers are still eager for a universal paired dataset.

A large portion of the existing dataset is unpaired. Such images could not be used in research for a long time until Junyan Zhu proposed the unpaired style transfer model CycleGAN(Zhu et al. 2017). CycleGAN uses the idea of forward and then reverse translation in machine translation to achieve impressive results in unpaired style transfer tasks.

In real life applications, 1) face aging can be applied to short video software and camera software for entertainment items such as filters and special effects 2) face verification is currently used in all aspects of daily life, but face databases are not updated from time to time, e.g., ID cards are usually valid for ten years, and if you apply for an ID card at the age of 18, the information does not expire until you are 28. the face at 18 and the face at 25-28 often have some gap. This is why it is often the case that train station gates fail to recognize the face and then fail to pass. Face ageing, on the other hand, can predict ageing faces and thus improve the accuracy of face recognition. 3) The police may need to track down a fugitive who has been on the run for years, but may only have a photo of him from decades ago; they need to find a child who has been lost for over a decade, but may only have a family photo of him when he was a few years old. Even with the web of facial recognition systems and cameras now in place, if we don't even know what the faces of the people we need to find look like, we won't find them. But if we can do accurate face ageing, we can know exactly what they currently look like, and with the Skynet system, we can greatly improve the efficiency of the search.

It can be expected that in the future, unpaired face data will dominate because of the ease of collection. If satisfactory results can be achieved using unpaired faces trained based on CycleGAN, this could be of great significance for future face ageing tasks.

In the general task of training a network model using faces of the same person of different ages, there is no need to consider the problem of keeping identity features, face attribute features (gender, race, etc.) constant. In the unpaired task, where different faces are used, various methods have to be used to keep these attribute features intact as the face ages. The addition of a large number of constraints may also make it difficult for the model to converge, so using unpaired face images for the face ageing task based on CycleGAN is also a worthwhile and challenging task to explore.

Related work

Over the past decade, deep neural networks have achieved great success in representation learning and have also been widely applied to face aging. Wang(Wang et al. 2016) proposed a recurrent face aging framework that is able to smooth the aging process, but less identity information can be maintained. Duong et al. successfully applied Deep Restricted Boltzmann Machines (Nhan Duong et al. 2015, 2016; Duong et al. 2019a) and Deep Reinforcement Learning (Duong et al. 2019b,c) to face aging and have achieved an impressive performance. However, these approaches still need paired images with different ages of the same person for model training.

The advent of GANs ushered a vigorous development of face aging. Antipov et al.(Antipov, Baccouche, and Dugelay 2017) first deployed the conditional GAN for age transformation. Later, Zhang et al. (Zhang, Song, and Qi 2017) proposed a conditional adversarial-auto-encoder (CAAE) that leverages the high-level features of input images for the generation of target images. But their synthesized faces show little change on aging effect for different age conditions.

Another line of research in face aging is based on Unpaired Image-to-Image Translation, which models the age progression and regression with one or two models. For example, Song et al. (Song et al. 2018) proposed to use two conditional GANs (Dual cans) to model face progression and regression respectively.

Liu et al. (Liu, Li, and Sun 2019; Liu et al. 2021) combined multi-level discriminator with a wavelet packet transform (WPT) module for high-level age-related features extraction to improve the visual fidelity of generated images. Shao et al. (Shao and Bui 2022) further combined multi-level generator with a WPT module for improving the identity verification confidence in face aging. All three articles use a p-dimensional vector to represent face attributes. The p-dimensional attribute vector describing the input face image is embedded to both the generator and discriminator to reduce matching ambiguity inherent to unpaired training data. These three articles use cycle-consistent loss, which is derived from CycleGAN, and they do not use the CycleGAN framework but use the ideas of CycleGAN. However, the three articles still use paired datasets.

Sun et al. (Liu et al. 2017) used a transition pattern discriminative network to maintain identity information of adjacent age groups. Wang et al. (Wang et al. 2018b) used a h(x) (Alex Net) to extract face features and then compare the similarity of two faces, which in turn kept the face identity information unchanged. Wang et al. (Wang et al. 2018a) employ the deep Convolutional Neural Network model (Sun et al. 2014) for face verification. Palsson et al. (Palsson et al. 2018) completed face aging with CycleGAN, but it did not use p-dimensional vectors to represent face attributes, nor did it use unpaired datasets for training. To be precise, there is no model based on CycleGAN and trained using unpaired images.

Proposed Solution

Pre-training

So, I used CycleGAN as the basic model, and made some changes on the model to adapt it to the face aging task. And I only use unpaired datasets to complete the task of facial aging.

When using a paired dataset, each young image has at least one corresponding old image. When using unpaired images, however, the original dataset does not have this one-to-one mapping relationship, which the model has to find for itself.

CycleGAN is one of the best unpaired style migration models. there are two generators and two discriminators in CycleGAN, namely the aging generator, the rejuvenation generator, the young image discriminator, and the old image discriminator. The young image enters the ageing generator first to generate the ageing image. The ageing image needs to enter the ageing image discriminator to determine whether it is a real ageing image; in the second round, the generated ageing image needs to enter the rejuvenation generator to generate the recovered young image. The recovered young image needs to enter the young image discriminator to determine whether it is the real young image. In addition, the young recovered image needs to be as close to the young original image as possible.

In this way, CycleGAN can establish one-to-one mapping relationships in unpaired datasets.

The datasets were selected as small unpaired datasets: AGFW-v2(Duong et al. 2019b) consists of 36,299 images divided into 11 age groups spanning 5 years. In contrast to other age-related databases, most of the subjects in AGFW-v2 are not public figures and are unlikely to have significant make-up or facial grooming, which helps embed accurate ageing effects in the learning process. More importantly AGFW-v2 is an unpaired dataset, where younger faces are not related to older faces in any capacity. In this paper, we only do two age groups for the ageing and rejuvenation tasks. If you need to do multi-age ageing, then training once between each two age groups is sufficient.

The CycleGAN model looks for a one-to-one mapping relationship between the two data domains. If the two data domains differ by a large order of magnitude, the model will become unconvergent. In this paper, the images of men aged 10-29 years in the AGFW-v2 dataset are selected as the young dataset, with 4997 images. The images of males aged 45-60 years were used as the older dataset, with a total of

4058 images. Training set: test set = 9:1. The images were read in and scaled from 0 255 to -1 1 pixel values.

The face ageing task is, in essence, a style migration task. So, for the pre-training, we directly used the original CycleGAN to do the training on the face images of the two age groups. After 100 epochs of training, the results obtained are shown in Figure 1.

Figure 1: The first row shows ageing, with the 16 young originals on the left and the 16 ageing images on the right; the second row shows rejuvenation, with the 16 old originals on the left and the 16 rejuvenated images on the right.

As can be seen, not only does the ageing and rejuvenation have little effect, but the generated images show significant Checkerboard Artifacts. as shown in the Figure 2, the generator loss values are not stable during training, no matter how low the learning rate is set, and the results are sometimes good and sometimes bad. The model was difficult to converge and the loss function did not reflect the good or bad training.

Checkerboard Artifacts are the result of an “Uneven overlap” due to deconvolution. This causes one part of the image to be darker than the rest.

When we use deep learning to generate an image, it is often based on a low-resolution image with high level semantics. This will allow deep learning to fill in the details on this low-resolution image. In general, to perform the conversion from a low-resolution image to a high-resolution image, we often have to perform deconvolution. In simple terms, the deconvolution layer allows the model to draw a square on the high-resolution image through each point. The Checkerboard Artifacts are generated in relation to the stride, kernel size of the deconvolution. If the kernel size is not divisible by the stride, the output of the deconvolution will be uneven overlap.

Solution

Previous solutions, where multiple layers of repeated deconvolution would make the board more complex than effective. This problem is well solved by using a stride of 1 for the deconvolution, but a stride of 1 prevents up-sampling. Deconvolution takes a kernel size that is divisible by stride, but still produces Checkerboard Artifacts if our kernel is unevenly learned.

Therefore, this paper does not use deconvolution. When up-sampling is required, the feature image is first increased to the required size using linear interpolation, and then the number of channels is reduced by a 1x1 convolution layer, thus achieving the up-sampling effect of the original CycleGAN network with deconvolution.

At the same time, WGAN (Arjovsky, Chintala, and Bottou 2017) and WGAN-GP(Gulrajani et al. 2017) were introduced to replace parts of the original GAN(Goodfellow et al. 2020) to make the model training more stable and reasonable. After several training sessions, it was also found that changing the instance normalization to batch normalization gave the model normal results.

Experiments

During training, the weight of the penalty term in WGAN-GP is set to 10. Batch size was set to 16 and the initial learning rate was 0.0002. Training set: test set = 9:1. The model was trained using Adam optimizer. GAN loss weights were set to 2 and cycle loss to 1. cycle loss weights were raised to 5 when training steps were greater than 3000. When steps are greater than 5000, the learning rate is 0.0001 and the cycle loss weight is increased to 10. for steps greater than 7000, the learning rate is 0.00001 and the cycle loss weighting is increased to 20. for steps greater than 10,000, the cycle loss weighting is increased to 20. The learning rate mentioned above refers to the learning rate of the generator, and the learning rate of the two discriminators is only 1/3 of that of the generator.

When training, a total of four networks needs to be trained, and while one network is calculating the gradient and updating the parameters, the weights of the remaining three networks need to be frozen.

After training for 100 epochs, the training process was very stable. The loss of generators is shown in Figure 3. The generated images are shown in Figure 4. Thanks to the use of CycleGAN, we also obtained a rejuvenation generator, and Figure 5 shows the effect of rejuvenation.

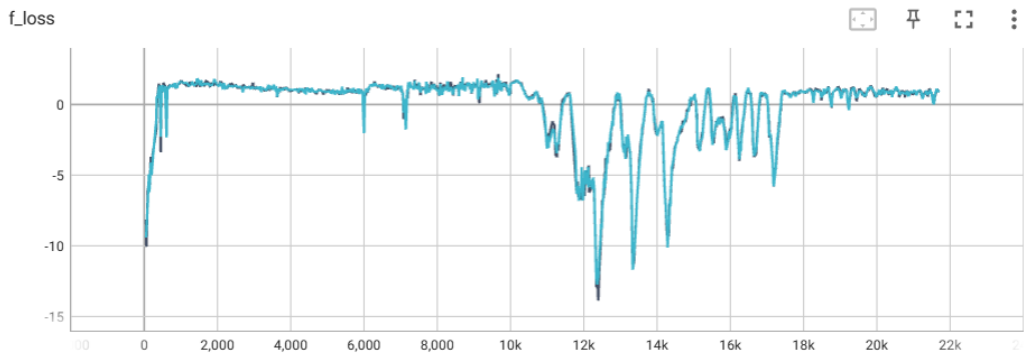

Figure 2: The generator loss values are not stable during training

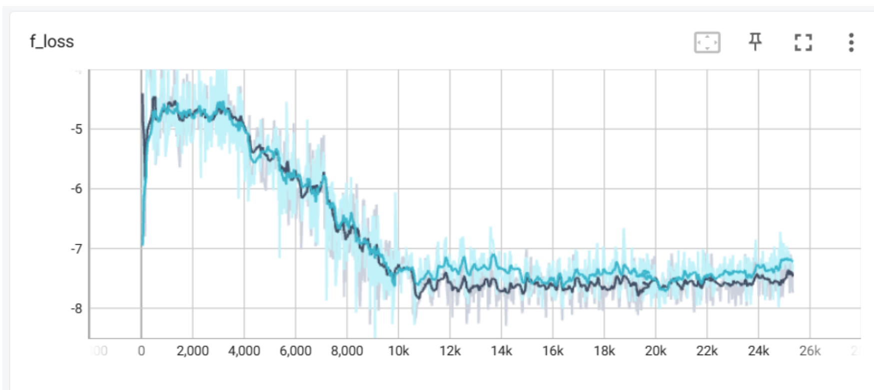

Figure 3: The loss of generators

Figure 4: The left-hand image in each column of the figure is the young original image and the right-hand image is the old generated image.

Figure 5: The left-hand image in each column of the figure is the old original image, the right-hand image is the young generated image.

	FG-net	aging task	rejuvenation task
FACE++	77.243	75.305	79.786
Baidu AI	91.980	91.604	93.459

Table 1: The results of the comparison

From a subjective point of view, the generated image is completely free of Checkerboard Artifacts and the image clarity is comparable to that of the original image. To the naked eye one can easily see the signs and effects of ageing and rejuvenation, and one can also see that it belongs to the same person as the original image, i.e., that the identity features remain the same.

For the face ageing task, an evaluation of the age accuracy and identity retention of the generated images is required. Here we use MEGVII's FACE++(2020) and Baidu AI face comparison and face attribute recognition functions. Face attribute recognition is still very difficult task and the recognition accuracy is very low even with the most advanced models. FACE++ and Baidu AI are tested here using the FG-net dataset, a paired training set with 818 images, and a test set with 170 images. The dataset contains photos of 82 people at different ages, and also provides information on 68 key points of faces in each image.

Age recognition was performed using FACE++ and Baidu AI on images of men aged 10-29 and 45-60 in FG-Net, respectively. For FACE++, the accuracy was 77.59% and the 1-off accuracy was 94.83%. For Baidu AI, the accuracy was 45.65% and the 1-off accuracy was 97.83%. The 1-off accuracy rate here refers to the fact that recognition of adjacent age groups is also considered correct, which is a common accuracy rate in the field of face age recognition.

It can be seen that FACE++ has better recognition.

When measured using FACE++, the ageing accuracy was 93.55% and the rejuvenation accuracy was 81.25%. When using Baidu AI, the ageing accuracy was 83.87% and the rejuvenation accuracy was 62.5%. It can be seen that the ageing accuracy of the model is much higher than that of rejuvenation.

It is also necessary to do an assessment of identity retention, using the face matching function of FACE++ and Baidu AI. Here FG-net is used for comparison. This is because FG-net is an image of the same person at different ages, which is the most realistic of real-life ageing situations. The results of the comparison are shown in Table 1.

FACE++ and Baidu AI use different comparison algorithms, and the same FG-net only has an identity verification confidence of 77.243 in FACE++, while in Baidu AI it has an identity verification confidence of 91.98. When using FACE++, the identity verification confidence for the old images generated by the ageing task and the original young images was 75.305, while the identity verification confidence for the rejuvenation task was 79.7855. The identity verification confidence of the ageing model was very close to that of FG-net. Similarly, when using Baidu AI, the identity verification confidence for the old images generated by the ageing task and the original young images was 91.604 and 93.459

ours	75.305
FG-net	77.243
PFA-GAN	85.870
Wavelet-GAN	87.530
Wavelet-multi-GAN	94.370

Table 2: 32 sample results compared with PFA-GAN(Huang et al. 2020), Wavelet-GAN(Liu, Li, and Sun 2019), Wavelet-multi-GAN(Shao and Bui 2022)

for the rejuvenation task. Similarly, when using Baidu AI, the identity verification confidence for the older images generated by the ageing task and the original younger images was 91.604 and 93.459 for the rejuvenation task. The identity verification confidence for the ageing model is also very close to that of FG-net.

This means that the degree of identity retention of the ageing images generated by the method in this paper is comparable to that of real-life ageing.

Of course, it is not possible to say that the closer the identity retention is, the better; this metric has to be evaluated in combination with the ageing accuracy. Otherwise, if the model adds no ageing features to the picture at all and outputs the original picture exactly as input, the degree of identity retention would get a good score of 100. As demonstrated above, the degree of identity retention is not as good with ageing as with rejuvenation, but the ageing accuracy is higher with ageing than with rejuvenation.

Table 2 compares the Identity Verification Confidence assessment using FACE++ with other models of facial ageing.

There is a gap between the degree of identity retention of models trained using unpaired data based on CycleGAN and mainstream models. However, there is no universally accepted uniform measurement method for face ageing tasks, so many papers require questionnaires or online voting to obtain subjective scores. The FG-net, which is the best representation of real-world reality, only scored 77.243 on Identity Verification Confidence in FACE++, so the FACE++ Identity Verification Confidence is not necessarily a complete reflection of the strengths and weaknesses of the ageing results.

Conclusion

In this paper, based on CycleGAN, we discarded deconvolution and used bilinear interpolation with 1x1 convolution to accomplish the face ageing task using unpaired data. Even though the final generated images are lower than the mainstream face ageing models using paired data in all measures. However, this is the first time that non-paired data is used for the face ageing task.

References

Antipov, G.; Baccouche, M.; and Dugelay, J.-L. 2017. Face aging with conditional generative adversarial networks. In *2017 IEEE international conference on image processing (ICIP)*, 2089–2093. IEEE.

Arjovsky, M.; Chintala, S.; and Bottou, L. 2017. Wasserstein generative adversarial networks. In *International conference on machine learning*, 214–223. PMLR.

Chen, B.-C.; Chen, C.-S.; and Hsu, W. H. 2015. Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. *IEEE Transactions on Multimedia*, 17(6): 804–815.

Duong, C. N.; Luu, K.; Quach, K. G.; and Bui, T. D. 2019a. Deep appearance models: A deep boltzmann machine approach for face modeling. *International Journal of Computer Vision*, 127(5): 437–455.

Duong, C. N.; Luu, K.; Quach, K. G.; Nguyen, N.; Patterson, E.; Bui, T. D.; and Le, N. 2019b. Automatic face aging in videos via deep reinforcement learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10013–10022.

Duong, C. N.; Quach, K. G.; Luu, K.; Le, T.; Savvides, M.; and Bui, T. D. 2019c. Learning from longitudinal face demonstration—where tractable deep modeling meets inverse reinforcement learning. *International Journal of Computer Vision*, 127(6): 957–971.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2020. Generative adversarial networks. *Communications of the ACM*, 63(11): 139–144.

Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; and Courville, A. C. 2017. Improved training of wasserstein gans. *Advances in neural information processing systems*, 30.

Huang, Z.; Chen, S.; Zhang, J.; and Shan, H. 2020. PFA-GAN: Progressive face aging with generative adversarial network. *IEEE Transactions on Information Forensics and Security*, 16: 2031–2045.

Lanitis, A.; Taylor, C. J.; and Cootes, T. F. 2002. Toward automatic simulation of aging effects on face images. *IEEE Transactions on pattern Analysis and machine Intelligence*, 24(4): 442–455.

Liu, S.; Sun, Y.; Zhu, D.; Bao, R.; Wang, W.; Shu, X.; and Yan, S. 2017. Face aging with contextual generative adversarial nets. In *Proceedings of the 25th ACM international conference on Multimedia*, 82–90.

Liu, Y.; Li, Q.; and Sun, Z. 2019. Attribute-aware face aging with wavelet-based generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11877–11886.

Liu, Y.; Li, Q.; Sun, Z.; and Tan, T. 2021. A 3 GAN: an attribute-aware attentive generative adversarial network for face aging. *IEEE Transactions on Information Forensics and Security*, 16: 2776–2790.

Liu, Z.; Luo, P.; Wang, X.; and Tang, X. 2015. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, 3730–3738.

Nhan Duong, C.; Luu, K.; Gia Quach, K.; and Bui, T. D. 2015. Beyond principal components: Deep boltzmann machines for face modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4786–4794.

Nhan Duong, C.; Luu, K.; Gia Quach, K.; and Bui, T. D. 2016. Longitudinal face modeling via temporal deep restricted boltzmann machines. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5772–5780.

Palsson, S.; Agustsson, E.; Timofte, R.; and Van Gool, L. 2018. Generative adversarial style transfer networks for face aging. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2084–2092.

Ricanek, K.; and Tesafaye, T. 2006. Morph: A longitudinal image database of normal adult age-progression. In *7th international conference on automatic face and gesture recognition (FGR06)*, 341–345. IEEE.

Shao, J.; and Bui, T. D. 2022. Wavelet-based multi-level generative adversarial networks for face aging. *Computer Vision and Image Understanding*, 223: 103524.

Song, J.; Zhang, J.; Gao, L.; Liu, X.; and Shen, H. T. 2018. Dual Conditional GANs for Face Aging and Rejuvenation. In *IJCAI*, 899–905.

Sun, Y.; Chen, Y.; Wang, X.; and Tang, X. 2014. Deep learning face representation by joint identification-verification. *Advances in neural information processing systems*, 27.

Wang, W.; Cui, Z.; Yan, Y.; Feng, J.; Yan, S.; Shu, X.; and Sebe, N. 2016. Recurrent face aging. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2378–2386.

Wang, W.; Yan, Y.; Cui, Z.; Feng, J.; Yan, S.; and Sebe, N. 2018a. Recurrent face aging with hierarchical autoregressive memory. *IEEE transactions on pattern analysis and machine intelligence*, 41(3): 654–668.

Wang, Z.; Tang, X.; Luo, W.; and Gao, S. 2018b. Face aging with identity-preserved conditional generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7939–7947.

Zhang, Z.; Song, Y.; and Qi, H. 2017. Age progression/regression by conditional adversarial autoencoder. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5810–5818.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, 2223–2232.