SYNTHETIC AGEING USING CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS WITH IDENTITY PRESERVATION

Rahul Radhakrishnan

Department of Electrical and Computer Engineering
University of Florida
r.radhakrishnan@ufl.edu

ABSTRACT

Synthetic Ageing or Age synthesis is the process of transforming a facial image of a person of a particular age to the desired target age. Generative Adversarial Networks(GAN) in the recent past has proven that synthetic images can be produced with high visual quality.In the previous works of synthetic ageing using GAN, the facial attributes modification was the key role and hence the original identity is most commonly lost in target age image.

In this project, i will be focusing on identity preservation of the persons original traits in the target age image. This is carried out by Identity preserved conditional Generative Adversarial Network. A identity loss is introduced into the objective for this purpose.A qualitative comparison between the state-of-art latent vector optimization for identity preservation cycleGAN and the IPCGAN is carried out to evaluate the performance of the system.

Index Terms— Synthetic Ageing, GAN, Deep Learning

1. INTRODUCTION

Synthetic ageing is the procedure of misleadingly incorporating an individual's face picture taken at an age to an objective age. Synthetic Aging is a hotly debated issue of research and has discovered its applications in different fields like the Forensic Art, Security control and Surveillance Monitoring, Entertainment and Cosmetology [3] and henceforth it is drawing in numerous scientists. For example, it can discover its application in the restorative medical procedures to furnish the specialists with a reference to disclose to their customer about the revived face structure already. As of now, there is a great deal of studies going on face maturing, however it has been a difficult undertaking in the field of computer vision in light of the fact that there is deficient preparing information of an individual at various age [4].

Conventional Synthetic ageing technique comprised of two strategies when all is said in done known as model-based methodology [15] and physical model-based methodology [1]. Model based methodology is actualized by processing a normal face for each scope of ages and afterward the age combined countenances are delivered with correlation of the objective age run normal face [5]. This technique sums up the maturing example of each individual and the individual personality of people are lost and subsequently the blended outcomes are ridiculous. Physical model-based methodology is executed utilizing a parametric model where it shows physical highlights of the face, for example, shape and texture.[1] But, such a usage requires exorbitant preparing tests and calculation power.

As of late, Generative Adversarial Network (GAN) have demonstrated enormous capacity to create pictures with top notch [6]. A substitute sort of GAN known as Conditional Generative Adversarial Network (cGAN) takes earlier condition and afterward creates a picture with the condition to deliver specific properties in the picture. Since my application requires a molding and being intrigued by the aftereffects of cGAN I will execute an Identity-Preserved Conditional Generative Adversarial Network (IPCGAN). Identity Preserved Conditional Generative Adversarial Network comprises of three units: a cGAN, a personality protection module, and an age classifier [1]. The model works so that the generator system of the cGAN takes the info picture and an objective age and processes a combined face picture with that age. This subsequent picture is sent through the discriminator system and is required to recognize the first picture in the objective age and the combined picture. I have build up a perceptual loss [8] in the objective to save the personality of the individual.

At last, to check that the synthesized image falls in the right age classification, we go it through a pre-trained age classifier and watch the outcomes. An age order misfortune is built up in the goal work.

2. LITERATURE SURVEY

2.1. Synthetic Ageing

Customarily, two strategies know as Prototype-based methodology [3] and physical model-based methodology [1] were used for Face Aging. Physical model-based methodology lies significance with respect to change in the physical highlights of the face, for example, the skin texture and facial muscle [9]. This technique is not exceptionally effective and doesn't deliver expected outcomes since it requires a lot of information across various age scope of people and computational force. The Prototype-based procedure contains figuring the normal face during a time run and consolidating it as the maturing design in the integrated picture of target age [3][5]. The combined pictures processed utilizing this strategy do not look practical on the grounds that the individual character of the individual is lost. To beat this, [10] has proposed a methodology utilizing hidden factor analysis joint sparse representation to consolidate the character highlights of a person in the incorporated face picture. Although the character properties of the face are protected, this strategy does not show a lot of promising outcomes on account of the creepy look in the incorporated face acquired because of the recreation procedure. In [2], they implement a conditional GAN for Face Aging. Their strategy was wasteful on the grounds that it figures a latent optimization vector for each picture, which requires huge computational force and handling time. Further, there are a few other execution and proposed strategy for face ageing however every usage has a downside or unpromising outcomes.

2.2. Generative Adversarial Network

Generative Adversarial Network [6] in any case basically known as GAN has demonstrated enormous outcomes in image generation. It comprises of two modules, to be specific Generator and Discriminator. The Generator takes a random vector created from the normal distribution or a uniform distribution and unique picture x as the input and synthesizes an image x'. The Discriminator takes the synthesized image (x') and the original image (x) as the input distinguishes between the original and fake image. The objective function is represented in 1 and the generator tries to minimize the function while the discriminator maximizes the function. The generator is prepared in such a manner to trick the discriminator into misclassification of the original and the synthesized images.

$$\begin{aligned} min_G max_D V(D,G) = E_{x\sim P_{data}(x)}[logD(x)] + \\ E_{z\sim P_Z(Z)}[log(1-D(G(Z)))] \end{aligned} \tag{1}$$

In [7], a substitute variant of GAN is presented known as the Conditional Generative Adversarial Networks (cGAN). The cGAN is an alternate variant of the GAN were the generator and discriminator are conditioned on a discretionary data y. A random vector from a normal distribution or uniform distribution and a condition data y is given as the input to the Generator and it delivers a synthesized image with the incorporated conditioning. The subsequent synthesized image is sent through the discriminator with the original image and the conditional information y to recognize the synthesized image and the original image. The objective function is represented in 2 and the generator tries to minimize the function while the discriminator maximizes the function.

$$\begin{aligned} min_G max_D V(D,G) = E_{x\sim P_{data}(x)}[logD(x|y)] + \\ E_{z\sim P_z(z)}[log(1-D(G(z|y)))] \end{aligned} \tag{2}$$

2.3. Style Transfer

In [11], such as fusing an objective age in a picture, the idea of Style transfer is jotted. Given an input and an artistic style image the substance of the input image are replicated in the subsequent picture with the style of artistic style image. To accomplish the guaranteed outcomes a perceptual loss function for training a feed-forward network is optimized[11]. The neural system extricates progressively perceptual features when contrasted with raw-pixel features. Unmistakable to style move, which moves style of one image to another, in age synthesis the features of the target age group are to be incorporated in the synthesized image. In this way, Style move isn't applied straightforwardly for age synthesis.

3. IDENTITY-PRESERVED CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS (IPCGAN)

3.1. Synopsis

The face images are first divided into 5 different age ranges. The face images within these 5 ranges are aged 10-19, 20-29. 30-39, 40-49 and 50+. When an input face image x is given into the system, a code C_s is generated to represent the age group on the input face image. C_s has a shape h*w*n, where h and w represent the height and width f the filter and n represent the age group the input face belongs to. One feature map is set to all ones depending on the age group and all the others are set to zero. Synthetic ageing aims at computing a target age image x' of the input face image.

Thus, it is important for the synthesized image to look realistic, x' and x have the same identity and the synthesized image x' is correctly classified into the right age group. To incorporate all the above said details in the system, an IPCGAN framework is implemented.

3.2. Identity-Preserved Conditional Generative Adversarial Networks

Figure 1 represents the complete pipeline of the Identity-Preserved Conditional Generative Adversarial Networks (IPCGAN). The IPCGAN consists of a Conditional Generative Adversarial Networks which is responsible to generate a synthesized face image x' within the target age group C_t, an identity preservation module which is responsible to retain the identity of x in x', an age classifier which ensure that the synthesized image lies in the correct age group as intended.

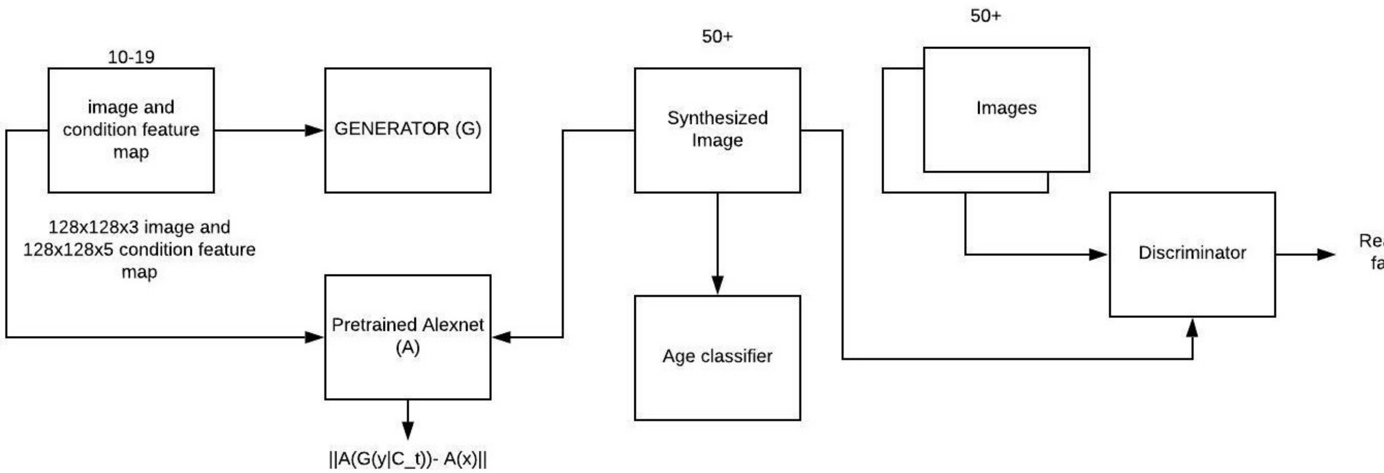

Fig. 1. The framework of IPCGAN for face ageing

CGAN face generation module Since the aim of the project is to synthesize a face image at a target age, I am using a conditional Generative Adversarial Network (cGAN) for the purposes of face generation. x represents the input image and y represents the real faces in the target age group having the distributions $P_x x$ and $P_y y$ respectively. The Discriminator should be able to distinguish between real image and the fake image generated by the generator. The probability that the real face is the real face i.e. $log(D(x|C_t)$ should be high. Also, the discriminator is responsible to stack the synthesized image with the target age, that brings us to an objective function as follows:

$$\begin{aligned} min_G max_D V(D,G) = E_{x\sim P_{data}(x)}[logD(x|C_t)] + \\ E_{y\sim P_y(y)}[log(1 - D(G(y|C_t)))] \end{aligned} \tag{3}$$

But, like [6] the optimization of the objective function of cGAN represented in 3 is also unstable and hence results in poor results. Thus, in [12] a Least square Generative Adversarial Networks(LSGAN) is proposed. In this paper, it is shown that the objective of the GAN[6] gets stuck in a small loss for the synthesized image because the discriminator can distinguish the two images easily. In LSGAN, it enforces the real face image and the synthesized image towards decision boundary and hence makes it identical. Thus, I am using a conditional LSGAN which is an alternate variant of cGAN. The conditional Least Square Generative Adversarial Network is represented as follows:

$$\begin{aligned} L_D = \frac{1}{2}E_{x\sim P_x(x)}[(D(x|C_t - 1))^2] + \\ \frac{1}{2}E_{y\sim P_y(y)}[(D(G(y|C_t)))^2] \\ L_G = \frac{1}{2}E_{y\sim P_y(y)}[(D(G(y|C_t) - 1)))^2] \end{aligned} \tag{4}$$

The discriminator proposed in [13] is used for the optimization of LSGAN since it has shown promising results for incorporating condition in the synthesis.

Identity-preservation module For synthetic ageing it is of utmost importance to preserve the identity of the individual in the synthesized image. The adversarial loss will not guarantee identity preservation as it only tunes the generator to follow the targets distribution and hence can produce an image which is like anyone in the target age group. I have introduced a perceptual loss in the objective function for identity preservation.

$$L_{Identity} = \sum_{x\sim P_x(x)} \|A(x) - A(G(x|C_t))\|^2 \tag{5}$$

In 5 A(.) represents the extracted features in a pre-trained neural network layer. The mean squared loss is not used in this case because the input image x and the synthesized image $G(x|C_t)$ do not appear same in the pixel space. This is because the synthesized image varies in the age group and hence contains changes such as wrinkles, facial hair, hair colour etc. The Mean Squared Error forces $G(x|C_t)$ to be exactly like x. Whereas, the perceptual loss forces the synthesized Image to retain the features of the input face image in the feature space.

The choice of layer to extract features is a very important step in identity preservation module. In [11], it has been indicated that the higher feature layers helps in retaining aesthetic features such as color, texture etc, whereas the lower feature layer help in retaining the facial identity information like the face structure, eye position etc. Hence, we adopt lower layer features of pretrained neural network as A(.). A(.) was set to features of conv5 after looking through all the layers in Alexnet pretrained with Imagenet[1]. Qualitative evaluation of the resulting synthesizes face images show promising results for convolutional layer5 when compared to features of another layers.

Age Classification Module In this project,I will be using a pre-trained age classifier to ensure that the synthesized image falls in correct age group. When training the IPCGAN the parameters of the classifier is fixed, and a small compensation is given if the synthesized image is classified in the correct age group else a huge compensation is given. For this purpose, an age classification loss is introduced in to the objective function. It is represented as L_{Age}. l(.) represents the softmax loss.

$$L_{age} = \sum_{x \sim P_x(x)} l(G(x|C_t), C_t) \tag{6}$$

Objective Function considering all the factors discussed above to obtain a target age synthesized face image we arrive at the following objective function [1].

$$\begin{aligned} G_{Loss} = \lambda_1 L_G + \lambda_2 L_{Identity} + \lambda_3 L_{age} \\ D_{loss} = L_D \end{aligned} \tag{7}$$

λ_1,λ_2,λ_3 controls the extent of aging of the input image, the extent of preservation of identity and to classify the synthesized image into the right age group respectively.

3.3. Network Architecture

Generator network- The concept of style transfer introduced in [11] and image-to-image translation in [14] has shown very impressive results. Therefore the generator network is equivalent to the one executed in [14] besides that in the first convolutional layer my framework gets a 128x128x3 images and 128x128x5 feature maps as input to the framework and consequently we have 6 residual blocks.One feature map is set to all ones depending on the age group and all the others are set to zero. The condition is integrated into the system before the first convolutional layer. The image and condition feature map is joined together and the resulting feature map is passed into the first convolutional layer.

Discriminator network-The network architecture is the same as that in invertible conditional GAN [15]. The condition incorporated into this network has previously shown promising result and hence I am using this implementation. Like represented in [1] I will be using the naming convention shown in [11]. $conv_k$ represents a 4x4 convolutional layer with batchnorm leakyReLu activation with a stride of 2 and the output being k channels. Thus, the architecture can be represented as $conv_{64}$–$conv_{128}$–$conv_{256}$–$conv_{512}$–$conv_{512}$. We do not apply batch norm $conv_6 4$ and integrae the condition after this layer as suggested in [15]. The condition feature map introduced to the generator is reduced from 128 to 64 because after the first convolution layer the feature maps are 64x64. The feature maps of $conv_6 4$ is combined with the condition feature map and is injected into the next layer.

Age classification network- The classification network architecture is drawn from Alexnet. The classifier is the same as the Alexnet from pool1 to conv5 layer. Two fully connected and one softmax layer in added in the end and dropout is introduced to avoid overfitting.

4. EVALUATION

In this section I will be introducing the dataset use for the project and the image processing required to clean the dataset. Also, I will be evaluating the performance of the IPCGAN qualitatively in comparison with a cycleGAN.

4.1. Dataset

For my project I have use the Cross-Age Celebrity Dataset (CACD) [4]. It contains more than 1,60,000 face images of 2000 celebrities from age ranging from 16-60+. The face images in the dataset are of various pose, expression etc and hence has to be processed before it is used. The dataset is preprocessed to remove or mask unreliable data and then center cropping and resizing the image to 400x400. After cleaning and preprocessing we have a total of 1, 63,000 images in CACD data set.The dataset is then split as 90% training data and 10% test data set. Initially during the project proposal I had chosen a subset of IMDB-Wiki dataset which had about 500k images. But, though the dataset was large, when closely looked into the dataset the images are of very poor resolution and there is excess of unreliable data which could eventually lead to the downfall of the system. Hence, I have used the CACD dataset.

4.2. Evaluation details

I have compared my implementation with another face aging model which is an implementation developed on the concept of [15] i.e. it is a latent vector optimized identity preserved cycleGAN. The tensoflow implementation of both the models was taken from the internet and fine changes were made according to my need and trained on the same dataset. The cycleGAN uses gender information for the purpose of identity preservation but this has been omitted to make a fair comparison. I have used a pre-trained age classifier provided by the autor in [1] where a Alexnet is finetuned on CACD dataset for 2,00,000 steps. The network is trained with a learning rate=0.01 and a exponential decay every 15 epochs and learning decay rate=0.0005 with a batch size=64. The IPCGAN is trained by fixing the learning rate=0.001 with a batch size of 50. The whole training process gets completedin about 3,49,000 steps.

4.3. Qualitative Analysis

In this section I will compare both models described above based on the on the quality of the Synthesized image or visual aesthetics of the image. For this purpose we chose 5 random images from the test dataset from 10-19 age group

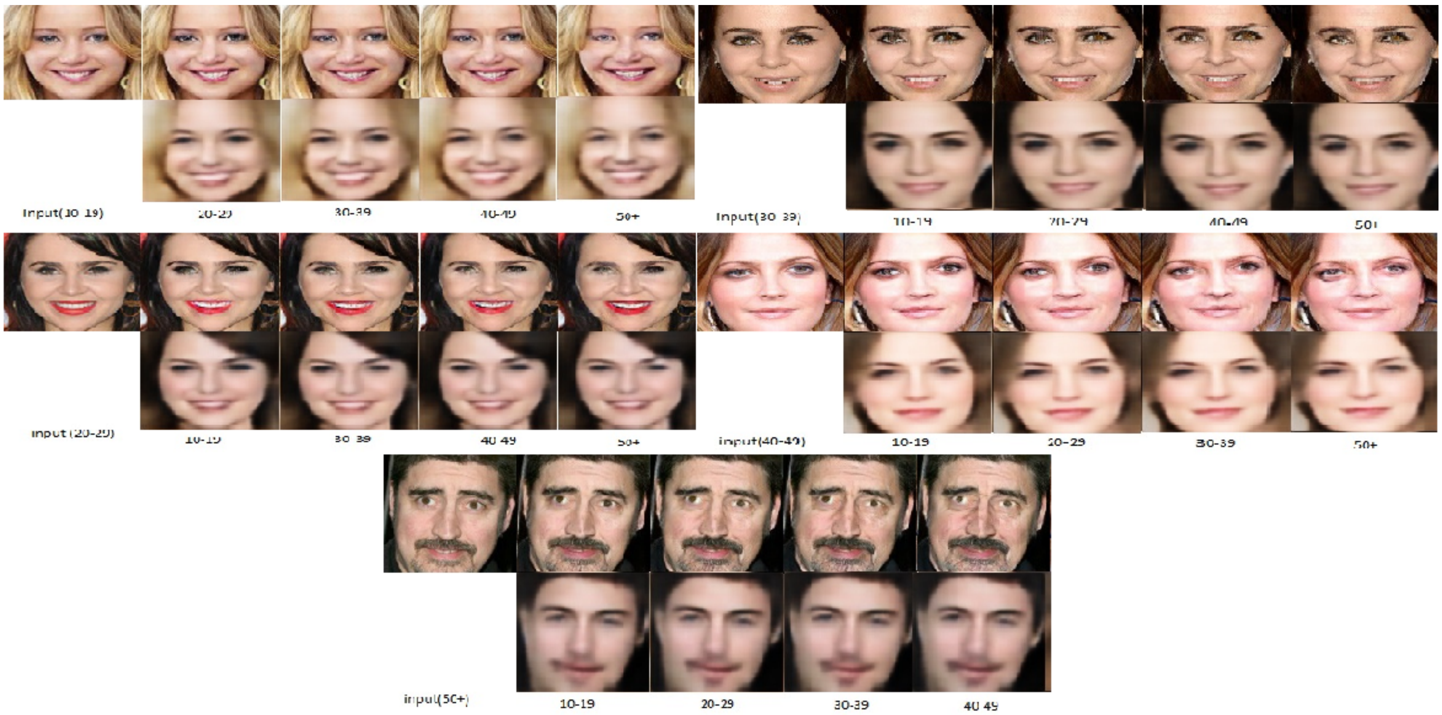

Fig. 2. Evaluation results for age synthesis for IPCGAN and Latent vector optimised cycleGAN

and face age the images using both implementation. This is represented in figure (2). It can be clearly seen that the aged face images generated using the latent vector optimized identity preserved cycleGAN does not show any vivid change in different age groups and the images are blurry and not clear for visualization. The results of IPCGAN can be observed to have the same identity in all age groups and the face ageing pattern in each of the age groups are evident. The age classifier network forces the synthesized face into the target age and hence the face ageing pattern in each of the age groups are evident. The identity preservation module plays a significant role in this magnificent result, the feature layer chosen to retain the properties is the key to get good results. The conv5 layer was chosen after careful evaluation with other datasets.

5. CONCLUSION

The Generative Adversarial Network (GAN) has shown immense results in image generation. For the purpose of the project I have used an alternative version of Condition Generative Adversarial Network (CGAN). The identity preservation module in the IPCGAN ensures that the synthesized image should have all the high level features of the input image and the age classifier network guarantees that the synthesized image is enforced in the correct target age and hence making ageing pattern evident. This results in the IPCGAN producing promising results for synthetic ageing.

6. FUTURE WORK

It has been noticed during the process that images with spectacle and shades in the input image produces a synthesized aged image in which the ageing pattern is not very visually pleasing, and the image does not look realistic. Hence, this an open research problem and

7. REFERENCES

[1] Tang, X., Wang, Z., Luo, W. and Gao, S., 2018. Face Aging with Identity-Preserved Conditional Generative Adversarial Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition,.

[2] G. Antipov, M. Baccouche, and J.-L. Dugelay. Face aging with conditional generative adversarial networks. In IEEE International Conference on Image Processing, 2017.

[3] Y. Fu, G. Guo, and T. S. Huang. Age synthesis and estimation via faces: A survey. IEEE transactions on pattern analysis and machine intelligence, 32(11):1955–1976, 2010.

[4] B.-C. Chen, C.-S. Chen, and W. H. Hsu. Cross-age reference coding for age-invariant face recognition and retrieval. In Proceedings of the European Conference on Computer Vision (ECCV), 2014. pp.1955-1976.

[5] I. Kemelmacher-Shlizerman, S. Suwajanakorn, and S. M. Seitz. Illumination-aware age progression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3334–3341, 2014.

[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D.Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Advances in neural information processing systems, pages 2672–2680, 2014.

[7] Mirza, M. and Osindero, S., 2014. Conditional Generative Adversarial Nets.

[8] A. Dosovitskiy and T. Brox. Generating images with perceptual similarity metrics based on deep networks. In Advances in Neural Information Processing Systems, pages 658–666, 2016.

[9] J. Suo, S.-C. Zhu, S. Shan, and X. Chen. A compositional and dynamic model for face aging. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(3):385–401, 2010.

[10] H. Yang, D. Huang, Y. Wang, H. Wang, and Y. Tang. Face aging effect simulation using hidden factor analysis joint sparse representation. IEEE Transactions on Image Process- ing, 25(6):2493–2507, 2016.

[11] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In European Conference on Computer Vision, pages 694–711. Springer, 2016.

[12] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley. Least squares generative adversarial networks. arXiv preprint ArXiv:1611.04076, 2016.

[13] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee. Generative adversarial text to image synthesis. In Proceedings of The 33rd International Conference on Ma- chine Learning, volume 3, 2016.

[14] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired imageto- image translation using cycle-consistent adversarial networks. arXiv preprint arXiv:1703.10593, 2017.

[15] Antipov, G., Baccouche, M., Berrani, S. and Dugelay, J.,2020. Apparent age estimation from face images combining general and children-specialized deep learning-models. Proceedings of Computer Vision and Pattern-Recognition Workshops

[16] GitHub. 2020. Deatherving/Face-Aging-With-Cyclegan. [online] Available at: $< https : //github.com/deatherving/Face - Aging - with - cycleGAN/blob/master/cycle_gan.py >$ $[Accessed 29 April 2020]$.