Final Year Project

In-depth Study of Generative Adversarial Networks for Face Aging

Na Li (17210325)

A thesis submitted in part fulfilment of the degree of

MSc. (Master) in Computer Science

Supervisor: Prof. G. C. Silvestre

UCD School of Computer Science
University College Dublin

August 18, 2019

Table of Contents

Abstract

Human is curious about exploring the world, and they are also interested in to know what they will look like in the future. Face aging is a well-known task in science. The aim of the face aging task is synthesizing facial images by building a model of the facial aging process. Human faces contain tremendous information and each person has their unique face, and even their old faces have many differences comparing with their young faces. In recent years, many researchers are pursuing to design effective age models for generating high quality of realistic aging images. Noticeably, one challenge lies ahead for researchers in the field of face and gesture recognition, which is facial data collection and data processing as facial data is the foundation of a building model. If a large dataset can be fed into a model for facial aging, it is beneficial to train and robust a model. Another challenge is how to keep personal identities in synthetic faces for modeling the face process. With the advent of Generative Adversarial Nets (GANs), variable face aging models are designed based on GANs, mainly because generating high-quality images and own more effective training process.
Having deep investigated the three state-of-art face aging frameworks which are related GANs, that is, Identity-Preserved Conditional Generative Adversarial Networks (IPCGANs), Conditional Adversarial Autoencoder (CAAE) and Cycle-consistent Generative Adversarial Networks (CycleGANs), combing with age and gender estimator as evaluation tools, this paper aims to deep study, analyze and implement one of the models following by facial datasets collection and processing.

Keywords

Face aging, Gender and Age classification, Age and Gender estimation, Face Alignment, Generative Adversarial Nets (GANs), IPCGANs, CycleGANs, CAAE

1 Introduction

Face aging is a aging processing of facial images under a certain age, also is one of most popular technology in the computer version of deep learning area, as it can apply various applications for facial recognition, for example, seeking long-lost children [1] and entertainment such as Face APP. Traditional face aging approaches are more complex and high computing costs, which are separated prototype-based approaches [2] and physical model-based approaches [1, 3, 4]. The prototype-based approaches are to the computed average age in the same age groups, and to learn a transformation of each group, but it may lose many personal identities. The physical model-based approaches have to cost large computation to learn face features. Both categories methods are hard to build an effective model for face aging. On the other hand, there are many challenges for many researchers who devoted themselves of face aging technologies: Dataset collection and processing as training and testing because the facial expression, pose, illumination and occlusion may influence the identities of facial features leading to influence training results. The emergence of Generative Adversarial Nets (GANs) technology [5] has been witnessed their success in generating high quality of images. Therefore, many face aging methods have been developing dramatically depending on GANs. Identity-Preserved Conditional Generative Adversarial Networks (IPCGANs) [1], Conditional Adversarial Autoencoder (CAAE) [3] and Cycle-consistent Generative Adversarial Networks (CycleGANs) [6] are impressive for many researchers as the three models are the latest face aging methods based on the GANs model. Moreover, they employed different aspects and concepts to implement face transformation and the corresponding personal identity-preserved questions. The CAAE networks, the input images are encoded by an encoder, and then fed into a generator. The aim is to learn a face manifold, and stepping along the manifold to realize smooth age progression and regression; CycleGANs networks are a kind of image-to-image translation, two aged groups of a dataset are experienced two generators and discriminators processes, from transforming aged faces to reconstructing faces that are similar with the input. The aim to not only keep personal identities but also transform new features from each other; IPCGANs is the best effective networks for face aging comparing two others, which consist of identity-preserved module to learn and save specific identities so as to generate aged faces are more realistic, meanwhile injecting a pretrained age classifier is an effective way to enforce the generator to synthesize specified aged images.

This paper intends to detailed and in-depth study and analyze the three state-of-the-art face aging methods, IPCGANs, CAAE, and CycleGANs, based on the GANs, and involve gender and age estimators for evaluation of the three face aging methods, which were researched by evaluating around 29 academic articles according to the previous literature review. This paper's outline is as follow, the section 2, related work for face aging, GANs, and face age estimator; the section 3, data processing; the section 4, Introduce of the three networks for face aging; the section 5 experiments and evaluation; the paper ends with a comprehensive conclusion and discussion of future work.

2 Related Work

2.1 Face aging

The acknowledged two classes of traditional approaches for face aging are physical model-based methods and prototype-based methods, which are illustrated in many papers. Many researchers point that both exist the limitations and weaknesses, even though it can be built a model for face

aging. Firstly, the physical model-based [3, 4] focuses on the physical features such as muscle, wrinkle, hair color and so one. Obviously, they have to require a large amount of data for training and spend much time for training and the model seems too complex. Secondly, the prototype-based method [2] is to compute average age in the different age groups and to learn differences between age groups [3]. Nevertheless, it may be difficult to preserve personal identities. Nowadays, most state-of-the-art works in face aging apply and redecorate Generative Adversarial Networks (GANs) such as Conditional GANs [7], Deep Convolution GANs [8]. In GANs network, the input images are encoded into a latent space and rebuilt new images under the target ages. In addition, it is difficult having a fair solution to evaluate these face aging models quantitatively [9]. Some researchers invite volunteers to vote for each result from different face aging models. Another way, Palsson *et al* (2018) [2, 9] point that, is to compare the performance of a cross-age face verification system with generated faces under the target ages.

2.2 Generative Adversarial Networks (GANs)

In recent years, many researchers are exploring an excellent way to create a generative model from natural images. Currently, Generative Adversarial Networks (GANs) [5] have been proven that is the best feasible networks to solve the task, special for face aging task in computer vision. GANs have two models of deep neural networks, one is a generative model (G) and another one is a discriminated model (D). The main task of G is to encode the input images such as facial images and generate new images that are similar to inputs as much as possible. By contrast, the discriminated model continues to clarify the input images and generated images from G. In other words, during the training process, G tries to generate more realistic images to deceive D, at the same time, D learns to distinct both kinds of images. The training process is the adversarial process between G and D so that to generate more realistic images. The formula (1) of GANs as follow,

$$\min_{G}\max_{D} V(D,\ G)\ =\ E_{x\sim\ P_{data}(x)}\left[logD(x)\right] + E_{z\sim\ P_{z}(z)}\left[log(1-D(G(z)))\right]. \qquad (1)$$

x is input images, $p_{data}(x)$ is real image data distribution; z is a any noise vector, $p_{data}(z)$ is noise variables. G is function of generation to map z and get $G(z)$ as generated images. Then the $G(z)$ and x are fed into discriminator as $D(G(z))$ and $D(x)$. The discriminator tried to distinct $G(z)$ and x, whereas the generator tries to trick the discriminator by generating $G(x)$, which is more similar to x.

2.3 Face Age and Gender Estimation

Face age and gender estimation are used to evaluate generated facial images and real facial images in this project. According to the previous study in the literature review, Deep Expectation of Apparent Age (DEX) [10] is one of the most popular and latest age classifier, which is a deep convolution neural works based on VGG-16 architecture [11], but the pre-trained model is developed by Caffe framework which is not suitable this project. VGG Face model of face recognition is based on VGG-16 architecture, and face age and gender estimation are similar to face recognition technology. What's more, the pre-trained model of VGG Face recognition is publicly available. Therefore, combining the DEX architecture and VGG Face pre-trained model, face age and gender estimators can be implemented for the evaluation of the face aging GANs model in the Keras framework for this study.

3 Dataset

In this experiment, according to the analysis of facial dataset collection in the literature review and the exist experiments from some research papers, four groups of datasets for face aging experiments and age/gender estimators are chosen. Firstly, two datasets are Cross Age Celebrity Dataset (CACD2000) [12] and UTKFace [3] as a training and validation dataset for the three face aging models (IPCGANs, CAAE, and CycleGANs). The CACD2000 contains 163,446 images from 2000 celebrities with 16 to 62 age ranges in the world. The UTKFace has 23,706 images with a large age range from 0 to 116. The two datasets not only have high-quality images but also have a large scale of age ranges. As for age and gender estimators, WIKI-IMDB dataset [13] is a typical and valuable dataset for training age and gender estimators. It contains 20,000+ images with two mat files that label precise information of images such as gender, date of birth and the year of taken photos for age calculation etc. The last dataset as a testing dataset is FGNET [14], which is 1002 images from 82 subjects with the age range from 0 to 69.

3.1 Data Processing

Before applying the four groups of datasets, although all datasets are face images, there are still many variations in illumination, expression, poss, and style [1] such as wearing glasses, have one more people in the same image, or just black image in CACD2000 and UTKFace dataset, therefore it is necessary to clean dataset by face alignment (including face detection), center cropping and re-size images in order to obtain the clear and aligned face images. What's more, CACD2000 without gender label, but training CAAE need to use gender label for face aging with different genders. Therefore, an age estimator is implemented to add the gender labels for the core training dataset CACD2000, which is one of the reasons for training an age estimator. And the second reason is for evaluation in section 5.

In this experiment, the imutils and dlib packages [15] are involved for face alignments. The figure 1 is from CACD2000 images, it is obvious that the face image is much clear after doing face alignment, re-size and center cropping. It is a benefit of training models as the main facial feature can be detected accurately.

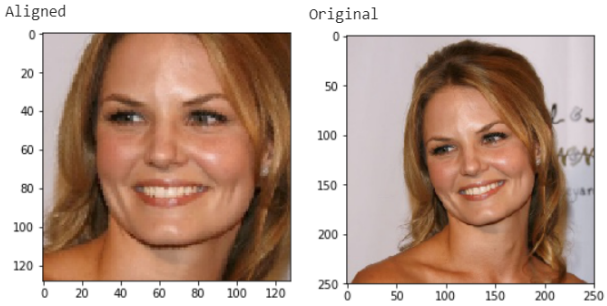

Figure 1: Compared two faces after face alignment

The detail amount of each dataset is in the table 1, and the WIKI-IMDB dataset is for training and testing age-gender estimators model, which do not need data processing as filtering dataset can be based on the labels information. Meanwhile, FGNET dataset keep the original dataset as test datset in order to verify the robustness of three GANs models.

Table 1: The amounts of each dataset

	CACD2000	**UTKFace**	**FGNET**	**WIKI-IMDB**
	Training-set	Validation-Set	Test-Set	Age-Gender Estimator
Data processing	158,661	20,168	1002	115,549
Original	163,446	23,706	1002	115,549

4 GANs for Face Aging

In this study, the three typical GANs are introduced and analyzed which are introduced the previous study in the literature review phase theoretically, that is, Identity-Preserved Conditional Generative Adversarial Networks (IPCGANs) [1], Age Progression/Regression by Conditional Adversarial Autoencoder (CAAE) [3] and Cycle-Consistent Adversarial Networks (CycleGANs) [6]. Firstly, IPCGANs is a type of conditional GANs with using the concept of "Identity-Preserved" because a synthesized face should include the same identities from the original face (sample data). Meanwhile, a pre-trained Alex classifier is applied in order to force the generated faces in the target age groups. Therefore, the synthesized faces may look realistic in the different age groups. Secondly, a face manifold is proposed in the model of CAAE. Zhang *et al* (2017) [3] present that the points of the input faces can be generated in a face manifold, and following along the face manifold, the synthesized faces of different age could be generated with preserving personality (*e.g.* age labels). Obviously, learning a face manifold is a core task of CAAE. Meanwhile, by controlling the age attribute, the CAAE can achieve age progression and age regression. However, Zhang *et al* (2017) [3] point that it is complex operation for creating a face manifold as it is high dimensional, then they use a convolution encoder to ensure the input face to map a latent vector, then the vector will be projected the face manifold with age labels in a generator. Lastly, CycleGANs can solve the unpaired image to image transfer topic effectively, which is a popular and well-known model. The main different of structure with IPCGANs and CAAE models is that CycleGANs contain two main processes which are the input image to translated image, and the translated image can translate back to the input image. The discriminator is trained between the sample image and the translated image. During the two generating processes and two discriminating processes, the image features can be learned and preserved. Therefore, it may be interesting and valuable to employ CycleGANs for facing image translation with different age groups. In this section, the Detail structures are illustrated for the three models, following by explaining experiments for each model.

4.1 Identity-Preserved Conditional Generative Adversarial Networks (IPCGANs)

IPCGANs is a kind of a conditional GANs [16] as a face aging approach. The conditional GANs (CGANs) is an extension of GAN models that it can conditionally identify image features. IPCGANs consist of a Conditional Least Squares Generative Adversarial Networks(LSGANs) as a generator, a pre-trained Alexnet classifier as an identity-preserved module, a pre-trained age classifier to identity the input faces from which age groups, and a discriminator. Wang *et al* (2018) employ the Conditional LSGANs [17] which is a special CGANs as generator for facial generation in the five target age groups (11-20, 21-30, 31-40, 41-50 and 50+). The main reason why they approach LSGANs is that LSGANs decrease the detection boundary so that the synthetic face and the input face are indistinguishable as much as possible in the discriminator. The pre-trained Alexnet classifier combined perceptual loss [18] can preserve the identical information such as color, texture, wrinkle, *etc.*. In addition, they argue that lower feature layers should be adopted

pre-trained Alexnet in order to the identity information and the aged faces can be balanced. At last, discriminator can ensure the consisting features between original faces in the certain age groups and the target age.

4.1.2. IPCGANs Network Architecture

IPCGANs model has three main neural networks, generator, discriminator and Alexnet age classifier.
Firstly, Generator neural network with six residual blocks (ReNet) refers to the network of style transfer [18] and the network of unpaired image-to-image translation [6] in figure 2, The source of RGB images can be any kinds pixels, in this experiment, the input size is 400x400x3, after it is centre cropped to 128x128x3 and concatenated conditional feature maps (age labels,128x128x5) as input sources to the first convolution layer. Mathematically, the generation task formula is,

$$\mathcal{L}_D = \frac{1}{2}\mathbb{E}_{x\sim\ p_x(x)}\left[(D(x|C_t) - 1)^2\right] + \frac{1}{2}\mathbb{E}_{y\sim\ p_d y(y)}\left[D(G(y|C_t))^2\right]. \tag{2}$$

$$\mathcal{L}_G = \frac{1}{2}\mathbb{E}_{y\sim\ p_y(y)}\left[(G(y|C_t) - 1)^2\right] \tag{3}$$

1. The two first 2D convolution layers process the output from the previous layer from size 128x128x32 to 64x64x64 followed a batch normalization (Batchnorm) (batch size:32) and a rectified linear operator (ReLU), kernel sizes are 7x7 and 3x3, strides are 1 and 2 respectively.
2. In the third layer, 2D convolution layer "Conv3" with kernel size (3x3) and strides is 2 processed the 64x64x64 output from "Conv2", and followed by an activation function ReLU. 32x32x18 is third layer output and entered to the six iterations of residual blocks as third layer. The residual block network has one convolution layer (kernel size:3x3, strides:1), followed a batch normalization and ReLU, the output 128x128x32 is input pixels of second 2D convolution layer(kernel size:3x3, strides:1) with a batch normalization.
3. The residual block (Resnet) output combined the "Conv3" output as two continue 2D deconvolution layers' input (64x64x64) for fifth and sixth layers, and followed by a batch normalization and ReLu for the fifth layer, the sixth layer only followed by a batch normalization (kernel size:3x3, strides:2).
4. The last layer is a 2D convolution (kernel size:7x7, strides:1) with an activation function Tanh, and the final output pixels are 128x128x3.

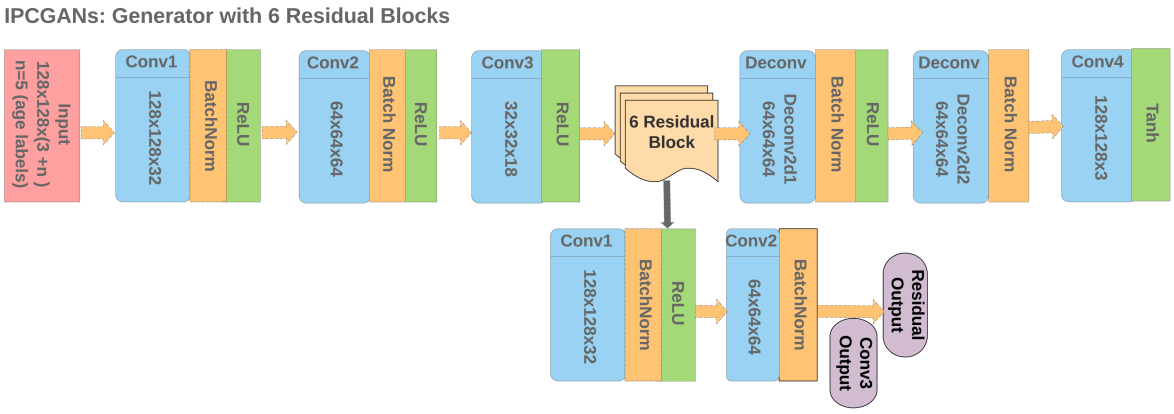

Figure 2: IPCGANs Generator Network Architecture

Second network is a patch discriminator neural network. The structure of discriminator is from conditional GANs [19] and unpaired image-to-image translation [6]. 5 layers with 128x128x3

input sources. In figure 3, the first four 2D convolution layers with LeakyReLU (slope:0.2), and the first 2D convolution layer does not insert Batchnorm but other three layers. After the first layer "Conv2d1", the conditions (features) (size: 64x64x5) are injected, then input and the conditions are combined together (Conv2d1:64x64x(64+5)) to the "Conv2d2". The other three 2D convolution layers with Batchnorm and LeakyReLU, from "Conv2d2":32x32x128, strides:2, "Conv2d3":16x16x25, strides:2 to "Conv2d4":16x16x512, strides:1, At the last 2D convolution layer is still 16x16x512, strides:1.

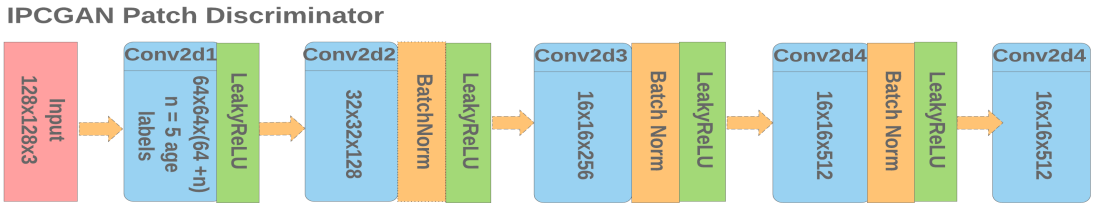

Figure 3: IPCGANs Discriminator Network Architecture

The last main network is age classification from Alexnet structure. In figure 4. The input source size is 227x227x3 for the first layer. From the Conv1 to Maxpooling5, the age classification has the same structure with Alex, after Maxpooling5, two fully connected layers with a dropout, and one last fully connected layers. The output shape is 5 which are the five age groups. Between the input layer and "Conv5" layers, the 2D Convolution layers ("Conv1" and "Conv2") with ReLU, Max Pooling (3x3, strides:2) and local response normalization (depth radius:2, alpha: beta:2e−05, bias:0.75); The third and fourth layers are 2D convolution layers with ReLU. The age classifier loss formula is,

$$\mathcal{L}_{age} = \Sigma_{x \in p_x(x)} \iota(G(x|C_t)). \tag{4}$$

The perceptual loss in the identity-preserved module is

$$\mathcal{L}_{age} = \Sigma_{x \in p_x(x)} \|h(x) - h(G(x|C_t))\|^2. \tag{5}$$

The overall loss function for generation of faces with the same identity is,

$$\mathcal{G}_{loss} = \lambda_1 \mathcal{L}_G + \lambda_2 \mathcal{L}_{identity} + \lambda_3 \mathcal{L}_{age}, \mathcal{D}_{loss} = \mathcal{L}_D. \tag{6}$$

where $\lambda_1 = 75$ is for controlling the aged input image, $\lambda_2 = 5e - 5$ and $\lambda_3 = 0$ in the target age group, both are controlled to keep the identity information and ensure the synthetic faces can drop into the correct age group. What's more, the λ_1 and λ_2 are initialized in the 11-20 age group, then the target group in the oldest age group 50+. In fact, λ_3 is dynamic parameter, if it grows, it shows the aging effect are more distinct. Finally, Generator and discriminator adopt the Adam optimization [20] with beta 0.5.

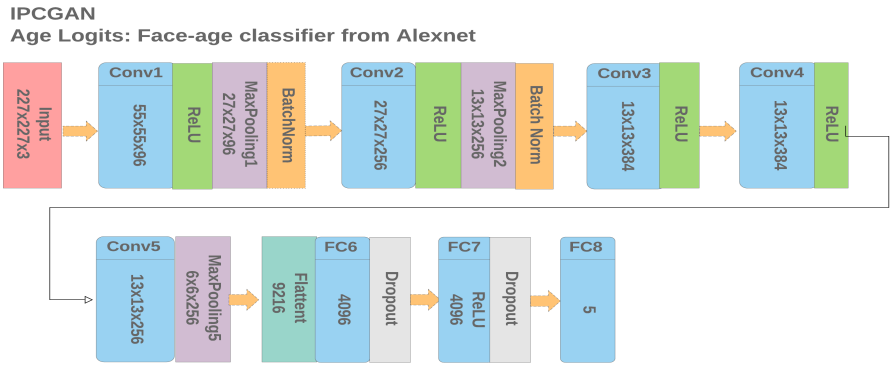

Figure 4: IPCGANs age classifier based on Alexnet

4.2 Conditional Adversarial Autoencoder (CAAE)

The most special part of CAAE is the concept of a high-dimensional manifold. The aim of CAAE network is learning the face manifold. Zhang *et al* (2017) [3] point that it is flexible to generate faces with different age groups by control the age attributes on the face manifold. However, it is complex to learn and build a high-dimensional manifold, a convolution encoder is first network with face images to map a latent vector. The vector will be the input in a deconvolution generator with the age attributes. Encoder and Generator are two adversarial networks for generating more realistic faces by ten age groups (0-5, 6-10, 11-15, 16-20, 21-30, 31-40, 41-50, 51-60, 61-70, and 71-80). What's more, two discriminator networks (D_z and D_{img}) are involved for Z which is generated from encoder network and real image with one-hot age label respectively. The loss function is

$$\begin{aligned} Loss = \min_{E,G} \max_{D_z, D_{img}} \; & \lambda \mathcal{L}(x, G(E(x))) + \gamma TV(G(E(x), l)) \\ & + E_{z^* \sim \; P_z} [\log D_z(z^*)] \\ & + E_{x \sim \; P_{data}}(x) \, [log(1 - D_z(E(x)))] \\ & + E_{x,l \sim \; P_{data}}(x, l) \, [log(D_{img}(x, l))] \\ & + E_{x,l \sim \; P_{data}}(x, l) \, [log(1 - D_{img}(G(E(x), l)))] \, . \end{aligned} \tag{7}$$

where TV(.) joined the sum of variation in order to remove the dirty information, and the coefficients λ and γ can improve the quality of resolution.

4.2.1 CAAE Network Architecture

Four main network components constitute CAAE network architecture. The first network is an encoder E, Zhang *et al* (2017) [3] present that encoder E is beneficial to generate a face with the specific features from original faces, but the general GANs choose samples randomly. In the figure 5, it is Encode E network architecture with 5 layers. The input source (128x128x3) feed into the continue 2D convolution layers with stride with the size 2 and the activation function ReLU. The outputs sizes from "E_Conv0" to "E_Conv4" are 64x64x64, 32x32x128, 16x16x256, 8x8x512, 4x4x1024 receptively. And then Fully-connection layer is the last layer (output size:50) with one-hot vector of 10 age categories and two gender labels as the input source (output size:150) for Generator network which is the second network of CAAE, refer to figure 6, the first layer is also a fully-connection with output is 4x4x1024. There are six continue 2D deconvolution layers followed by ReLU, kernel size is 5 and stride is 1, excluding the last 2D deconvolution layer followed by Tanh (kernel size:5, stride:1).

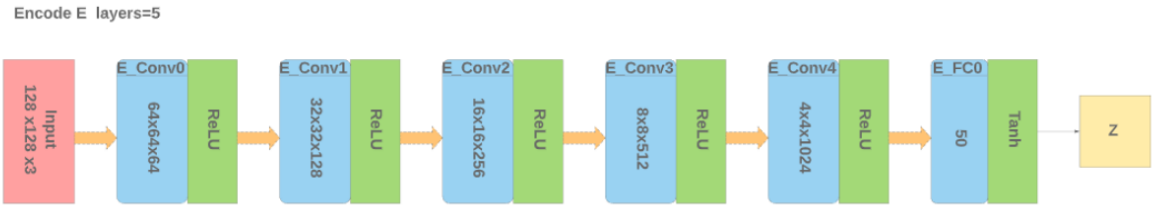

Figure 5: CAAE Encoder Network Architecture

The third network is discriminator (D_z) on z which is generated by Encoder E. In the figure 7, the first top network is the discriminator on z, obviously, it is combined four fully-connections. In the middle of three layers, Zhang *et al* (2017) [3] designed a fully-connection with BatchNorm and ReLU, and the last layer is a Sigmoid activation function and the output size of each layer is 64, 32, 16 and 1.
As for the last network is a discriminator on generating faces and input faces D_{img} . There are

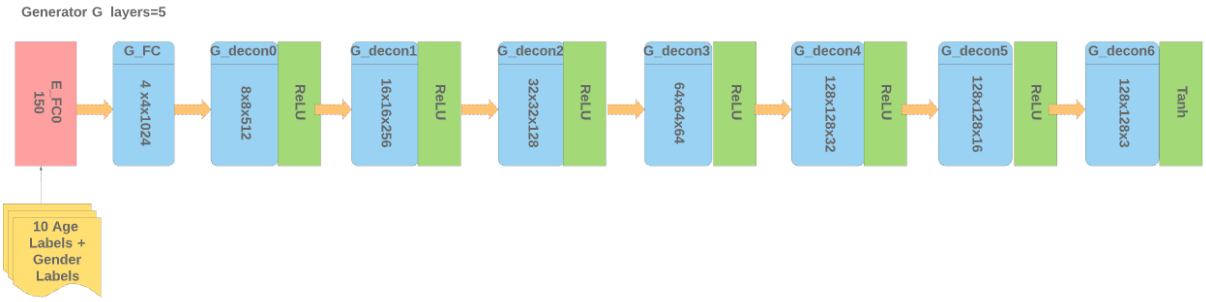

Figure 6: CAAE Generator network Architecture

four 2D convolution layers and two fully-connections (second network in the figure 7). The four 2D convolution layers (kernel size:5, stride:2) followed by the BatchNorm and ReLU. Specially, The age labels and gender labels are injected into the first layer's output for discriminating against fake faces at a certain age. Zhang *et al* (2017) compared both with and without discriminator on the image. Using D_{img} can obtain more realistic generated faces, rather than randomly choosing a age label during training. The last-second Fully-connection layer (F1) followed by BatchNorm and ReLU, the output 1024 feeds into the last fully-connection with Sigmoid, and the output is 1. The output of each layer refers to the bottom one in figure 7.

Figure 7: CAAE Two Discriminators Network Architecture

4.3 Cycle-consistent Generative Adversarial Networks (CycleGANs)

CycleGANs, in essence, generating high quality of fake face that is more similar to the original face by applying generator twice. It not only preserve the original faces features but transfer the special personalities or features from the other paired or unpaired group dataset. In the first CycleGANs paper [6], (G, D_x) and (G, D_y) are two paired networks in the CycleGANs. In other word, each group of dataset are not only input images but also target images, they tried to generate the images with the same features as target images in the process of $G(x)$ or $G(y)$ and also they preserve their own features by the processing of $G(D_y)$ or $G(D_x)$. The cycle-consistency loss can be described a forward cycle-consistency loss: $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ and a backward

cycle-consistency loss $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$. For face aging topic, Palsson *et al* (2018) [9] present the concept of "Group-GAN" that the original face dataset are separated the different age groups such as 20s and 50s, and one CycleGANs model can be trained each paired age groups images. The loss function is

$$\mathcal{L}_{cyc}(G,\ F) = \mathbb{E}_{x\sim\ p_{data}(x)}\left[\|F(G(x)) - x\|_1\right] + E_{y\sim\ p_{data}(y)}\left[\|G(F(y)) - y\|_1\right]. \quad (8)$$

The full objective of CycleGANs is

$$\mathcal{L}(G,\ F,\ D_x,\ D_y) = \mathcal{L}_{LSGAN}(G,\ D_Y,\ X,\ Y) + \mathcal{L}_{LSGAN}(F,\ D_x,\ Y,\ X) + \lambda\mathcal{L}_{cyc}(G,\ F). \quad (9)$$

4.3.1 Network Architecture

In the CycleGAN overview network architecture in figure 8, it refers to the Zhu *et al* (2017) [6] structure for face aging task. In the pre-divided paired group age datasets, the first process is at input A, each image (Input_A) in the group A will be mapped to some of images (Input_B) in the group B. Once the progress of pairing has done, the image A from the domain A to the domain B will bring the shared features. In other words, Input A is fed in a generator (A-B) and transformed to generated_B with the share features in the domain B, and then the generated_B is trained into a generator (B-A) in order to force preserving the input A features with the special features from domain B. In a similar way, input_B will be matched to some images from dataset B and generator(B-A) will train on image B to generate Generated_A, and apply generator(A-B) to force the Generated_A to return to the image B. The job of discriminator is to distinct whatever images or generated images, so it has an ability to defy the generator and even reject the generated images. However, the two generators of CycleGANs try to generate the image which is hard to distinct by discriminators. In the CycleGAN architecture has two discriminators are designed. The inputs (A and B) are fed into each discriminator and the two generated_A and generated_B are fed into a discriminator.

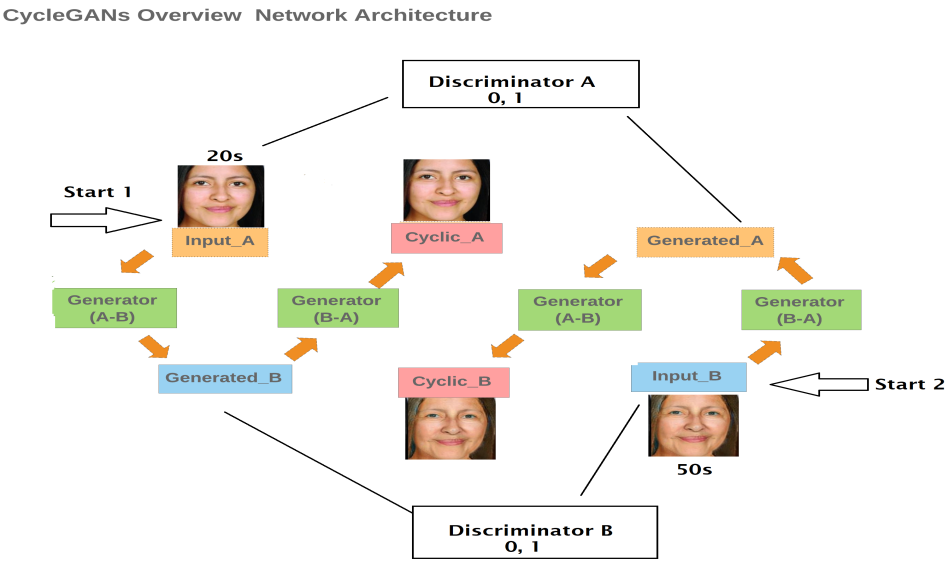

Figure 8: CycleGAN Overview Network Architecture

As for the single Generator Architecture. In figure 9, it consists of three components which are encoding, transformation and decoding. In the encoding, three 2D convolution layers with Instance Normalization [21] which the performance of a deep neural network can be improved obviously for generating image issues and ReLU, the first convolution layer followed by stride 1 and kernel 7. And the last two 2D convolution layers followed by stride 2 and kernel 2. The image size is 256x256x3, and the output size at the last layer is 64x64x256; The transformation has 6 or

9 Resnet blocks. Zhu *et al* (2017) [6] present 6 Resnet blocks which is enough. Some researchers prefer to apply 9 Residual blocks for face aging generation. In the CycleGAN, the main purpose of the residual block is saving properties of previous layers and available in the later layer of the input layer. Moreover, the Residual block is the same as in the IPCGANs which is mentioned in section IPCGANs. In this network, each residual block includes two convolution layers with kernel 3 and stride 1. The input of the residual is added to the output of the residual layer for each residual layer.

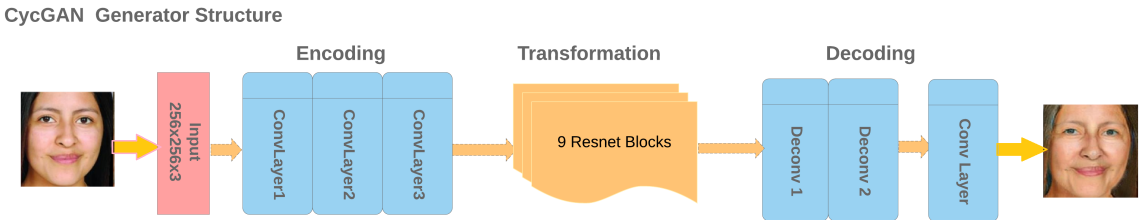

Figure 9: CycleGAN Generator network

As for the discriminator architecture, In figure 10, a discriminator is composed by four convolution layers followed by Instance Normalization and LeakyReLU except the first layer "Conv2d1" only followed by LeakyReLu and the last layer is a simple convolution layer. The image size during feeding in the discriminator is 256x256x3 as input to 32x32x1 as output.

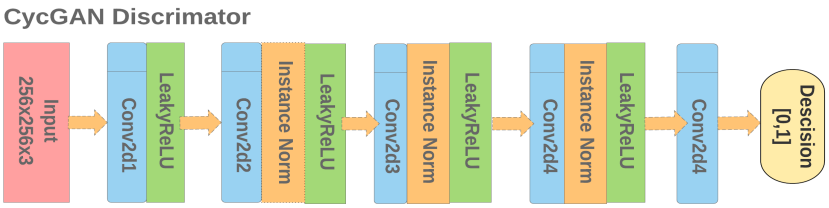

Figure 10: CycleGAN Discriminator network

5 Experiments and Evaluation

In this experiment of three models, in implementing of IPCGANs model and CAAE model, the authors provided programmes with a Tensorflow framework in the Git-Hub [22,23] with the detail instructions about the train from scratch, and IPCGANs programme package also included the pre-trained model. The CAAE programme package has an initial model and dataset link which they used for training. For those two existing programmes of IPCGANs and CAAE, three main tasks in this experiment for study and analysis,
Firstly, upgrading programmes using the latest TensorFlow-GPU 1.14.0, python 3.6.8 and so on. Modify old version function with parameters to match the new development and test environment. Secondly, the IPCGANs programme was developed and tested by TensorFlow-GPU 1.4.1 and python 2.7.x; CAAE was implemented and test by TensorFlow-1.7.0 and python 2.7.x. It is necessary to replace some functions that have deprecated in the latest 3rd-package dependencies of python, *e.g.* the function "imsave" has been deprecated in spicy 1.0.0, and another majority of changes for running programmes on TensorFlow-GPU smoothly.
Thirdly, implementing some tool for checking information during training *e.g.* output log files in the IPCGANs and CAAE as there is no independent GPU environment, training model should be created a job on the server cluster. It is not flexible to control all the information properly. On the other hand, some files have to be prepared before training from scratch, for example, IPCGANs

has more complex programme structure, the data source files with age labels (including image file name and the exact age label) and five train/test age labels files have to be prepared, then the programme can read sourced images correctly.
As for implementing of CycleGANs for face aging, there is no programme in the paper "Generative Adversarial Style Transfer Networks for Face Aging" which is studied in the literature review phase related to the CycleGANs model for face aging (*e.g.* Group-GANs and FA-GANs [9]), therefore, the CycGANs is implemented from scratch using Keras framework on Google Colab and pyCharm.
The Google Colab is a powerful platform for preliminary investigation and experiments of this research project, as the Google Colab platform supplies available and free GPU environment (Tesla K80) and it can be continued running 12 hours each session.

5.1 Training the IPCGANs

1.Hyperparameter
The final hyperparameter used for training 170,000+ samples are shown on the table 2 what's

Table 2: IPCGANs Hyperparameters

Hyperparameter	Value
Learning rate	1e-3
Batch size	32
Max epochs	500,000
Beta1 for Adam	0.5
GAN Loss Weight	75
Feature Loss Weight	5e-5
Age Loss weight	30
Face loss weight	None
Discriminator iteration	1
Generator iteration	1

more, according to Wang [1] (2018) present the age labels which inject into different layers can affect the generated face dramatically in the Generator. Figure 11 shows that the input age belongs to in 11-20 age groups and the target is the 50+ age group, so when the age labels set into "Conv5" has a great result. So this experiment set the layer is also "Conv5" by default. "no" in this figure 11 means that remove identity-preserving module.

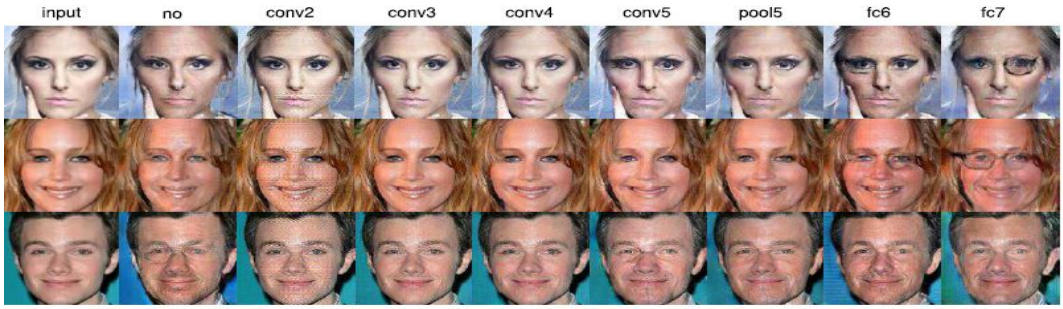

Figure 11: The aging effect with different feature layers [1]

2. Training Process [Environment/Strategy/Duration time]
In this experiment, the IPCGANs is trained on the GPU: Tesla V100-PCIE-32GB with allocated 20 CPUs on campus Sonic HPC Cluster. The total numbers of the training process are around 4+ times on the GPU, and each time has to continue 40 hours at least in order to investigate and verify the default hyperparameters whether the trained model can obtain the best result.

Meanwhile, testing on the processed data is better than the original data. The final training model spent around four days with one GPU.

3. Training Result (Train from the scratch)
The training result in the final epoch and the last row result is from the middle of epoch (around 299,999) which has evident age change in figure 12, the images are from training set which the programme chose randomly at each time.

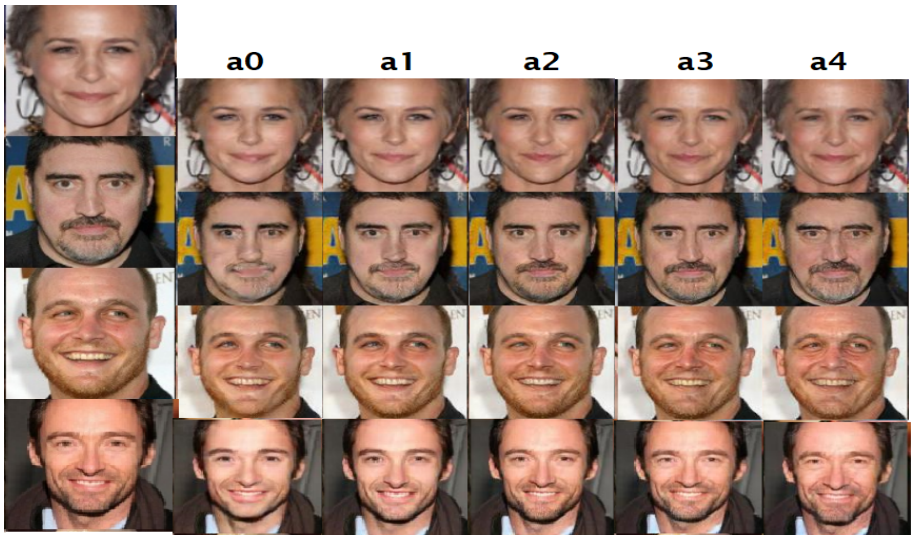

Figure 12: IPCGANs Result at the last epoch

5.2 Training the CAAE

1. Hyperparameter
The final Hyperparameter is shown on the table 3,

Table 3: CAAE Hyperparameters

Hyperparameter	Value
Learning rate	2e-4
Batch size	100
Epochs	50
Beta1 for Adam	0.5
Generating image loss	5e-5
Encoder on z loss	0
TV loss	0

2. Training Process [Environment/Strategy/Duration time]
In training of CAAE, the environment is the same as training IPCGANs (GPU: Tesla V100-PCIE-32GB with allocated 20 CPUs on campus Sonic HPC Cluster). There are two strategies to train and study this model, which are using the checkpoint model they provided to initialize a new model before the start of training, and training from the scratch without loading the initial model. In this experiment, first of all, following the instruction on the Git-Hub [22] the UTKFace dataset is used as a training set and loading the initial model and duration time is around 3 hours. Secondly, training on the CACD2000 (Training set) with a pre-trained model (epochs:50) during 40 hours. Lastly, training from scratch without a pre-trained model to initial new model weights (epoch:50) in around 40 hours.

3. Training result
Figure 13 is the result from the first strategy. Apparently, ten sample images are chosen randomly from the UTKFace dataset, and each sample is shown the age changing in ten age groups from the top 71-80 age group to the bottom 0-5 age group. Because of UTKFace dataset is a part of the dataset for the training of the pre-trained model, the result is the best comparing another strategy.

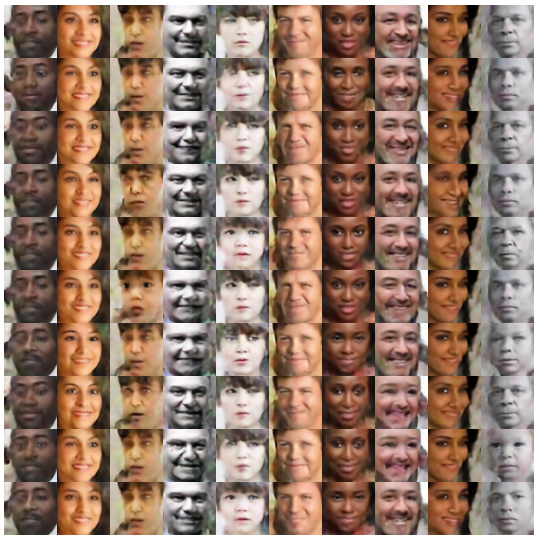

Figure 13: CAAE Result with UTKFace dataset and initial Model

For the last two approaches, the generated faces are fuzzy and anamorphic, the model did not preserve sample features during training properly. The weights of the provided pre-trained model are not compatible CACD2000 dataset. As for the last training approach (without the pre-trained model and using CACD2000), the result is the same as the second approach, the generated faces are not clear in figure 17 bottom one.

5.3 Training the CycleGANs

1. Hyperparameter The hyperparameter of CycleGANs for face aging is shown table 4

Table 4: CycleGANs Hyperparameters

Hyperparameter	Value
Learning rate_Generator	2e-4
Learning rate_Discriminator	2e-4
Batch size	1
Epochs	200
Cyclic loss weight(A2B)	10
Cyclic loss weight(A2B)	10
Identity A loss weight	1
Identity B loss weight	1
Discriminator loss weight	1
Beta_1(Adam Optimizer)	0.5
Beta_2(Adam Optimizer)	0.999

2. Training Process [Environment/Strategy/Duration time]
The environment for developing and unit test programme is on the Google Colab platform with

GPU, and the training CycleGAN is on the campus GPU (GPU: Tesla V100-PCIE-32GB with allocated 20 CPUs on campus Sonic HPC Cluster). The programme, at first phase, reproduces CycGANs model with Keras framework from CycGAN s paper *et al* and refers to their code on Git-Hub [24]. The sample data used "HorseToZebra" which is supplied by Zhu *et al* (2017) [6]. In the second phase, Palsson *et al* (2018) [9] raise Group-GANs based on CycleGANs framework for face aging where the face images are separated different groups and then trained between each paired groups. They presented that the performance is poor when the target age distance is small. The CycleGANs model was retrained with a training dataset of face images, two age groups for training are Train_A (10s-20) and Train_B (50s-70),

3. Training result

The whole training process is smooth using "HorseToZebra", the best result is at epoch 34 in figure 14 left, finally the horse and zebra images not only reserve the original images identities (e.g. position, lights, background) but also own the opposite features.

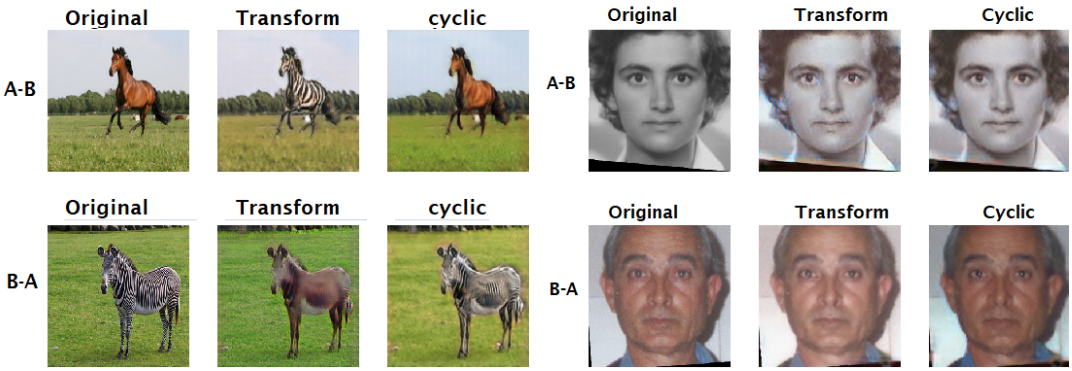

Figure 14: CycleGAN with HorseToZebra and Human dataset

As for training on face images the result is figure 14 right. After epoch 100, it still do not have face aging effect. According to Zhu *et al* (2018) CycleGAN tutorial blog [25], they present that this model is not fit for the shape of object, they built a CycleGAN model to transform genders using celebA dataset, the reconstructed images are distorted. However, CycleGANs for face aging is still worth further study.

5.4 Evaluation

Intuitively, the evaluation of face aging results mainly is to verify whether the generated face images in the target age group. There are divers evaluation approaches in related face aging papers. In this section, general and effective evaluation approaches for results of face aging are collected and the specific evaluation approach: age and gender estimator model will be illustrated.

5.4.1 Evaluation Approaches

Qualitative comparison and Quantitative comparison are two directions [1, 3] for the comparison with volunteers images and comparison with prior works.

Comparison with real images, The ground truth images under an age group and the generated images with the same age group are compared to confirm the performance of the trained model, which belongs to the qualitative comparison. Meanwhile, a certain number of volunteers are invited to vote whether the generated images are similar to the truth images under the same age groups. For example, Zhang *et al* (2017) [3] received 3208 votes totally, which 48.38% showing that the generated image is similar the real image. Wang *et al* (2018) [1] demonstrated three main methods of evaluation, first one is "Image quality" which volunteers voted the quality

of generated images; second one is "Age classification", volunteers estimated which age group belongs to for the generated faces; last one is "Face verification" that volunteers were asked whether the randomly selected generated face in group 4 is different from other generated faces from the same input face in the other groups. Those approaches indicated the quantitative performance directly.

Comparison with prior works, It is necessary for comparison to precisely retrain some state-of-the-art works including the same dataset and even the training environment, but some models do not provide source code, implementation of original work may be necessary [1]. Especially, Zhang *et al* (2017) [3] evaluate their model's robustness by examining the toleration to different poses, facial expression, and occlusion, which is also an effective and impressive way for model evaluation. What's more, IPCGANs consist of several components such as age classifier, identity-preserved components (Alexnet), so Wang *et al* (2018) evaluated each component, which is a comprehensive evaluation.

5.4.2 Age and Gender Estimator Model

Age and gender estimators are implemented with the Keras framework, and it is similar to DEX age classifier [10]. However, Rothe *et al* (2015) [10] used the Caffe framework for the implementation. In this research project, due to time constraint, the estimator uses pre-trained initial weights for VGG Face model based on the VGG-Very-Deep-16 CNN architecture [11]. Moreover, Oxford visual geometry group shared the systematical source code using the Keras framework and instructions on the Git-Hub [26]. Figure 15 [27] is the VGG Face model. In implementing the estimators, all layers are fixed excluding the last three convolution layers and modified the output units from 2622 to 101 for age estimator as the age range is 0 to 100, and to 2 for gender estimator. Meanwhile, refers to Antipov *et al* (2017) work [28], IMDB-Wiki_cleaned dataset [10] will be chosen as training set.

Figure 15: VGG Face Architecture [27]

Training the age/gender Estimators, due to GPU source limitation, data processing of IMDB-wiki dataset, building age/gender estimator models and evaluating this model were implemented on the Google Colab platform. The age/gender estimators are trained separately and spent around 5 hours. It is worth mention that the training process was smooth and stable from the loss plots in figure 16. As for hyperpermators, epochs are 250, the learning rate is 0.001 and the batch size is 256. The accurate rate for gender estimator is 97% on test set, the confusion matrix is ([[1781, 146], [55, 4660]]). The accurate rate for age estimator is not high as it should be converted classification task to regression (the predicted age is each softmax out multiply by the related labels.), Finally, the Mean Absolute Error (MAE) is 6.30 which is slightly high, which means the predicted ages has $\pm$6.3 error averagely.

5.4.3 Evaluation GANs model

As for CycleGANs, the performance is not satisfactory in the training section, it may have a slight change between the input images and generated images. In this section, it is valuable to evaluate the results by IPCGANs and CAAE. In the training phase, IPCGANs has a more high quality of

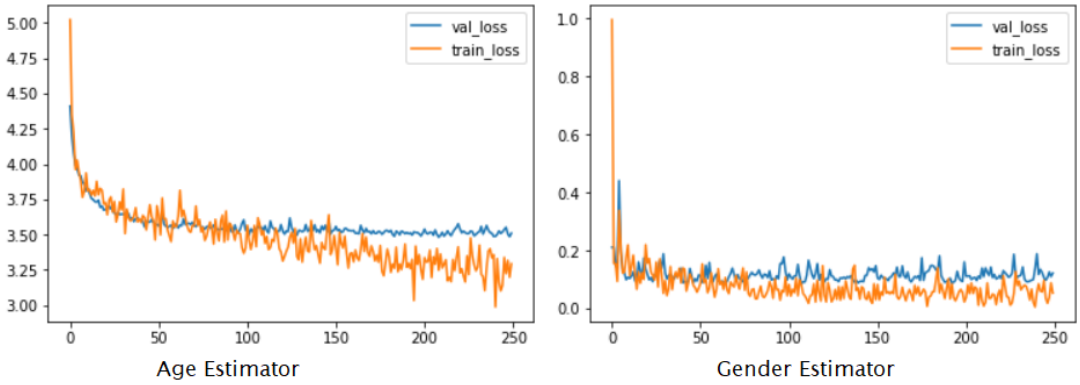

Figure 16: Training Loss of age/gender estimators

generated faces by using the training set. However, the results by CAAE is unrealistic, even it seems that CAAE losses more personalities on the original images. It is the same evaluation from Wang *et al* (2018) [1], the age effect is not strong and evident because of the usage of pixel loss between the original images and the related generated images. It is necessary to evaluate the two face aging GANs (IPCGANs, CAAE) by generating synthetic faces in the test dataset (FGNET). Randomly choosing the test images in the young age group, figure 17 top one is the test results with gender and age labels from IPCGAN, and predicted the generated faces by the pre-trained age estimator which is developed in this experiment, and figure 17 bottom one is the result from testing CAAE. In fact, intuitively, IPCGANs' results are much better CAAE and the synthesized images by IPCGANs have a higher image quality and fewer unrealistic features comparing with CAAE, the reason may be that IPCGANs detects and preserves a large number of individual features and flexibly change the setting age labels to affect the feature layers which is effective ways to robust model performance. As for using the two estimators, the results of IPCGANs match the target age groups as the age estimator has around ± 6.30 error.

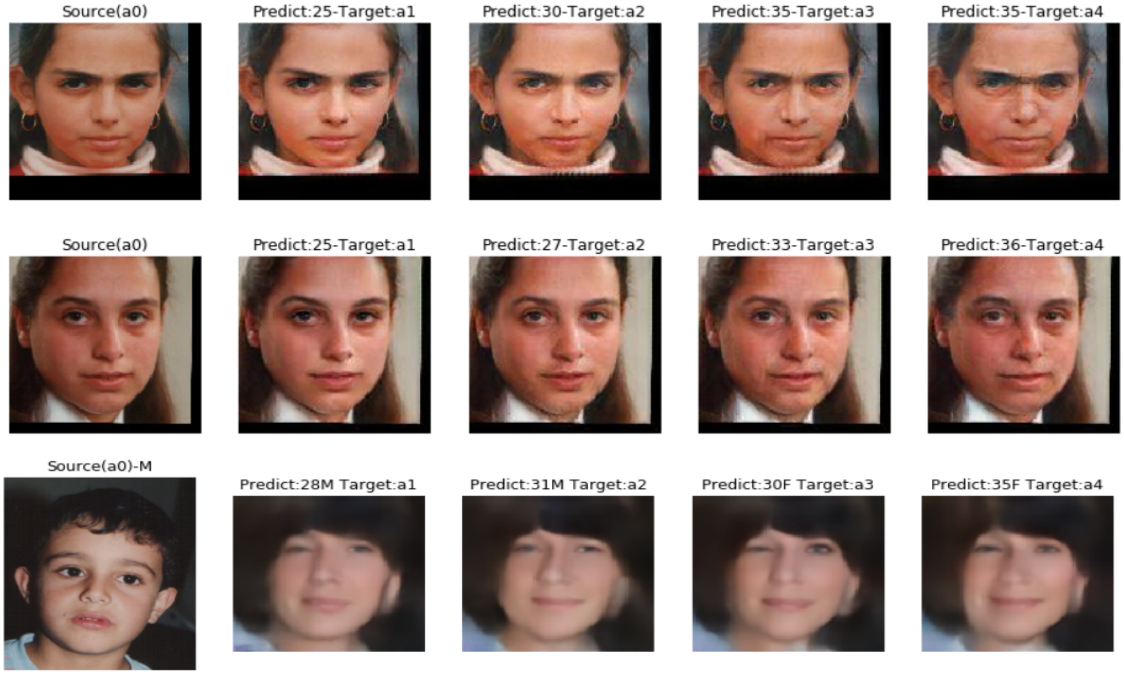

Figure 17: IPCGANs and CAAE test results by age/gender estimators

6 Conclusion

In conclusion, this paper has mainly a comprehensive study and analysis of three state-of-the-art GANs models for face aging (IPCGANs, CAAE, and CycleGAN) and two estimators applying pre-trained VGGFace model are built and trained as a type of effective evaluation way to implement. Due to researchers did not publish the official CycleGANs source code for face aging. The CycleGANs for face aging programme has been implemented based on the original paper using the Keras framework in this research project. As for training models, this experiment is more intensive mainly because of the limitation of GPU devices and large data volume which is slightly different from the corresponding original papers for the training model. Therefore, developing and training the two estimators employed on the Google Colab platform, which is a useful platform for the research project. Meanwhile, data processing is a significant preparation task before training and testing any kinds of models or classifiers. Human face images are special and complex mainly because of full of features on each person's face, so face detection, center cropping, and face alignment are three remarkable ways for data processing, then the models can identify and save more personality features in order to improve the performance of face aging GANs during the process of training. Apparently, by the results of the training and testing, IPCGANs are the most effective models for face aging. It is not denied that IPCGANs own the highest performance in terms of designing the model, because of its flexible model architecture and advanced identity-persevered model.

One aspect of future work is to investigate more accuracy and efficiency of the age estimation model for evaluation of the GANs model such as deep random forests for age estimation [29]. Another aspect is to collect face images of divers nationalities, it is noticeable that the three datasets are involved in this experiment, a few of the face images are from Asia and Africa. Lastly, it would be interesting to develop a real-time face aging model as it may be not only more entertaining but also could collect more face images to robust face aging models for more specific applications and to better serve in public.

Bibliography

[1] Z. Wang, X. Tang, W. Luo, and S. Gao, "Face aging with identity-preserved conditional generative adversarial networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[2] W. Wang, Z. Cui, Y. Yan, J. Feng, S. Yan, X. Shu, and N. Sebe, "Recurrent face aging," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[3] S. Y. Zhang, Zhifei and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017.

[4] J. Suo, S. Zhu, S. Shan, and X. Chen, "A compositional and dynamic model for face aging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 385–401, March 2010.

[5] 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 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 2672–2680. [Online]. Available: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

[6] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[7] G. Antipov, M. Baccouche, and J.-L. Dugelay, "Face aging with conditional generative adverarial networks," in *IEEE International Conference on Image Processing (ICIP)*, June 2017.

[8] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Computer Vision and Pattern Recognition*, 2016.

[9] S. Palsson, E. Agustsson, R. Timofte, and L. Van Gool, "Generative adversarial style transfer networks for face aging," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

[10] 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.

[11] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015.

[12] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Cross-age reference coding for age-invariant face recognition and retrieval," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 768–783.

[13] R. Rothe, R. Timofte, and L. V. Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," *International Journal of Computer Vision (IJCV)*, July 2016.

[14] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442–455, April 2002.

[15] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1867–1874, 2014.

[16] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *CoRR*, vol. abs/1411.1784, 2014. [Online]. Available: http://arxiv.org/abs/1411.1784

[17] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, and Z. Wang, "Least squares generative adversarial networks," *CoRR*, vol. abs/1611.04076, 2016. [Online]. Available: http://arxiv.org/abs/1611.04076

[18] J. Johnson, A. Alahi, and F. Li, "Perceptual losses for real-time style transfer and super-resolution," *CoRR*, vol. abs/1603.08155, 2016. [Online]. Available: http://arxiv.org/abs/1603.08155

[19] G. Perarnau, J. van de Weijer, B. Raducanu, and J. M. Álvarez, "Invertible conditional gans for image editing," *CoRR*, vol. abs/1611.06355, 2016. [Online]. Available: http://arxiv.org/abs/1611.06355

[20] J. B. Diederik P. Kingma, "Adam: A method for stochastic optimization," 2015. [Online]. Available: https://arxiv.org/abs/1412.6980

[21] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *CoRR*, vol. abs/1607.08022, 2016. [Online]. Available: http://arxiv.org/abs/1607.08022

[22] S. Y. Zhang, Zhifei and H. Qi. [Online]. Available: https://github.com/ZZUTK/Face-Aging-CAAE.git

[23] Z. Wang, X. Tang, W. Luo, and S. Gao. [Online]. Available: https://github.com/dawei6875797/Face-Aging-with-Identity-Preserved-Conditional-Generative-Adversarial-Networks.git

[24] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. [Online]. Available: https://github.com/hardikbansal/CycleGAN.git

[25] A. R. Hardik Bansal. [Online]. Available: https://hardikbansal.github.io/CycleGANBlog/

[26] A. Z. Omkar M. Parkhi, Andrea Vedaldi. [Online]. Available: https://github.com/rcmalli/keras-vggface.git

[27] Z. Parkhi, Vedaldi. [Online]. Available: http://www.robots.ox.ac.uk/~vgg/publications/2015/Parkhi15/poster.pdf

[28] G. Antipov, M. Baccouche, S.-A. Berrani, and J.-L. Dugelay, "Effective training of convolutional neural networks for face-based gender and age prediction," *Pattern Recognition*, vol. 72, pp. 15 – 26, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320317302534

[29] W. Shen, Y. Guo, Y. Wang, K. Zhao, B. Wang, and A. L. Yuille, "Deep differentiable random forests for age estimation," *CoRR*, vol. abs/1907.10665, 2019. [Online]. Available: http://arxiv.org/abs/1907.10665