Proceedings of 12th IOE Graduate Conference
Peer Reviewed
ISSN: 2350-8914 (Online), 2350-8906 (Print)
Year: 2022 Month: October Volume: 12

Generating Progressed Face Ageing Image using Generative Adversarial Network

Niraj Kumar Gupta [a], Sharan Thapa [b], Bal Krishna Nyaupane [c], Prabin Nepali [d]

[a, b, c, d] *Department of Electronics & Computer Engineering, Paschimanchal Campus, IOE, Tribhuvan University, Nepal*
[a] guptaniraj2051@gmail.com, [b] sharant@ioepas.edu.np, [c] bkn@wrc.edu.np, [d] prabino.nt@gmail.com

Abstract

The aged version of own face image is a matter of curiosity that one would look in near future. Among various technique use for modeling progressed aged face image, Generative Adversarial Network (GAN) and its extension conditional GAN with regression has shown astonishing results. This research work aims to generate the progressed facial image using the proposed model. The model takes the input image of size 256 *256 and the target age in range 1 to 80 for the generation of aged image. The input is converted into intermediate eighteen different style of latent space by age encoder in StyleGan domain which is input to StyleGAN2 generator to produce the target aged image. The output; aged face image is passed to age predictor to estimate the age. The eighteen style control the feature of generated output images like pose, hair, face shape, eyes etc. The loss between estimated and the target age along with other losses is used to update the model to produce aged version of input face image which is of size 1024 *1024. UTKFace datasets has been used to train the model. The model is able to generate plausible progressed aged face image in the range of 1 to 80 for single front facing image.

Keywords

Generative Adversarial Networks, Pixel to Style to Pixel, StyleGAN, Face Aging, Latent Space

1. Introduction

Face aging is a sequential change in facial skin tone and structure in future appearance of individual, can be used in search of missing human individual, fugitive criminal, cinematography. Face aging image creation model has a high uncertainty and very much affected by individual facial expression, posture, illumination and resolution.The aging rate vary by person to person which seriously increase the problem of generalization of model.

Generative Adversarial Networks GANs are a type of deep neural network created by Ian Goodfellow in 2014 [1] that employs a generator and discriminator network that train each other via repeated cycles of generation and discrimination while attempting to mislead one another. The discriminator is trained to discern between fake and true data, while the generator is taught to generate bogus data.The structural diagram of the GANs is shown in figure 1 GANs provide a powerful framework for rapidly creating data from supplied data probability distributions, in contrast to CNN, which is incapable of retrieving finer details and usually creates fuzzy pictures. The GAN optimization process is a minmax problem, and it is finished at the Nash equilibrium point. In a strategy profile, each player's plan is the ideal response to all other player strategies, according to the Nash Equilibrium.

The GAN model has advanced with the use of conditional GANs (cGANs). They have shown to be superior to standard GANS since they enable the generation of images with certain criteria or qualities in advance.

Target age condition along with the input face image is given in the model, to generate the aged face image. The progressed aged face image is achieved by generating aged face image at various age and then concatenated to display face aging progression. Aging is an unavoidable and ongoing process. Face age progression is required in various situations such as child/person missing, fleeing criminal, entertainment, security checkpoint, and so on. Human aging may be divided into two stages: childhood to adulthood and adulthood to old age. Cranial development occurs

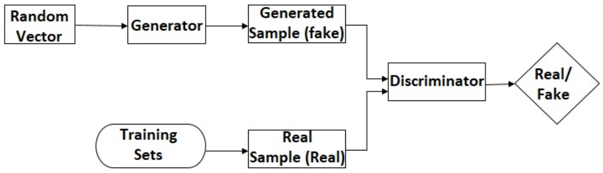


Figure 1: The Typical Structure of Generative Adversarial Networks (GANs).

from childhood through adulthood, and the latter is marked by changes in skin shape and texture. As a result, if a deep neural network model is developed that can imitate the given age version of the person at numerous ages, it can assist in identifying the person at the desired age. Similarly, in today's digital era, facial biometric security (at immigration, security checkpoints, and so on) has been deployed; these systems may be made more resilient using face aging modeling without the hassle of gathering facial data every time for cross verification. The cinematography makes extensive use of it since it depicts the characters' ages over a lengthy period of time, allowing for the easy synthesis of their many aged versions. This can be used to visualize own different aged version and to smile.

The problem of generating progressed face image and identifying aging accuracy is considered for this paper work. When developing a face age picture, visual integrity, aging correctness, and identity preservation are crucial factors. In most cases, the visual accuracy of a synthetic face image is determined in terms of human perception. A contemporary technique, like Frechet Inception Distance, defines the quantitative evaluation of visual quality. Whether a synthetic face image is within the intended age range or not can be determined by aging accuracy. Age estimator and user studies are the two quantitative evaluation methods that are employed. Identity preservation can be evaluated with three methods: automatic face verification [2], automatic face identification or user groups. Metrics for face identification are best calculated using ArcFace and FaceNet [3].

The main objective of this paper is to develop a methodology based on generative adversarial network for generating progressed aged face image.

2. Related Work

Face aging has been categorized broadly into two type viz. Traditional methods and Deep learning based methods. The traditional methods has been categorized into physical-model based and prototype based .Similarly, in deep learning based model, translational-based, sequence-based and conditional based are these three method been in used for the generating aged image of the face as shown in figure 2

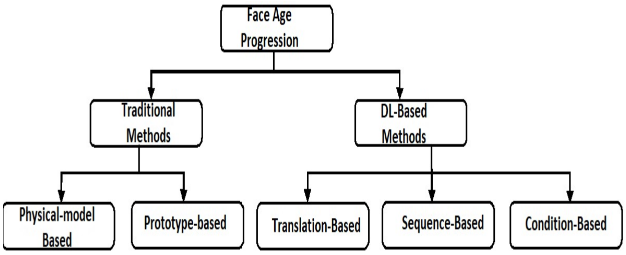


Figure 2: Face age progress and types.

For the first time, the extra age labels were incorporated into the network model by Zhang et al.[4]. The writers implements a Conditional Adversarial Auto encoder (CAAE) network, believing all images of the human face are represented by a high-dimensional manifold. To accomplish this, a convolutional encoder is used to map an input image of a face to the latent space. The encoded samples are moved in the direction of age change when the images are projected into the latent space by assessing the age label. A decoder network is then used to recreate the input facial image with the ageing effect. To construct Identity-preservation (LID), Yang et al.[5] took a deep face descriptor that has been pre-trained to extract identity-based feature vectors from both young and aged face images. The penalization of network for large identity difference is done by calculating the euclidean distance between the associated identity-related feature vectors. Similarly, L_{age} includes an age classifier that penalizes the difference between the age of the synthesized face image and the target age to prevent the synthesized face from straying from the target age. Following this logic, Wang et al. [6] proposed an Identity Preserving GAN (IPCGAN), which combines an identity-preserving component with a pre-trained CNN that functions as an age estimator. Shen et al. [7] describes InterFaceGAN, a model for manipulating facial features in a given face image. InterFaceGAN operates in the latent space of a previously developed facial image generating model, such as StyleGAN [8]. InterFace GAN leverages the well-structured latent space by looking for the linear boundaries that divide the latent space into two subspaces in terms of a binary semantic. Finally, an individual's age is continually determined by shifting a latent vector perpendicular to the border. However, the more the latent vector is moved in one direction, the greater the change in the identity of the original data is seen.

Vanilla GANs are effective at creating crisp images, but due to model stability, they are limited to small image dimensions. While Progressive growing GANs[9] is a reliable method for training GANs models which produces huge, high-quality images by gradually expanding the size of the model throughout the training procedure. In Progressive GAN, batch normalization is not used instead it uses other two technique mini-batch standard deviation and pixel-wise normalization. After each convolution layer, the generator does pixelwise normalization, which normalizes each pixel value in the activation throughout the channel.This is a type of activation limitation known more broadly as local response normalization. In this GAN, the bias for each layer is set to zero, and the model weights are set to a random Gaussian before being rescaled using the He weight normalization technique and model is optimized using Adam optimizer.

Pixel2Style2pixel[10] introduces a technique that transform the input image into intermediate z and extended latent vector w+ that can be used with StyleGAN generator for easy manipulation of facial attribute traversing the extended latent vector w+. The architecture for encoder is used same as in paper[10] to achieve the objective of this paper.

StyleGAN1[11] is an advancement of the progressive growing GAN for generating high resolution images.The StyleGAN generator no longer accepts a latent space point as input; instead, two new randomness sources are employed to build a synthetic image: a solo mapping network and noise layers.The mapping network produces a vector that defines the styles and connects them at each point in the generator model via a new layer called adaptive instance normalization. Control over the style of the resulting image is provided by using this style vector. The output image of this has blob like aritifact. The adaptive instance normalization in broken down into modulation and demodulation process in styleGAN2[12] that empirically prove resolving of blob like artifact produced in styleGAN1.

Or et. al[13] introduces Lifespan age transformation synthesis scheme that generate aging image by interpolating between age group using the latent vector which is trained on FFHQageing data. Similarly, Yao et. al.[14] explains the generation of high resolution aged image, the model of which is trained on FFHQ.

3. Methodology

The main framework of this paper is as in figure 3

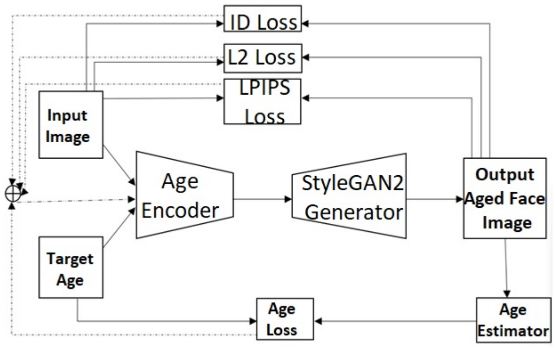


Figure 3: Model architecture.

3.1 Age Encoder

Age encoder is based on Pixel-to-Style-Pixel encoder architecture, the input to age encoder is four channel. The input image is added with input target age that is randomly sampled, as a constant value. The age encoder extracts features maps at three spatial level, the fine, medium and coarse style groups of StyleGAN. From these three level of spatial style, maptostyle convolutional neural network block change it to 18 different style of latent vector codes. I_{target_age} is the 4 channel input encompassing 3 channel of image and 1 channel of age. The age vector is stacked to image array tensor using vector broadcasting.

$$Age\_vector\_tensor = target\_age * (1, image\_width, image\_height) \quad (1)$$

I ≡ Input image

$I_{target\_age} = concatenate(I, target\_age)$

The structure of encoder is such that the input is fed to conv2d block with channel 4, filter size 64, kernel size (3, 3) stride (1, 1) and padding (1, 1). Followed by batch normalization and then parametric rectified linear activation . The output from convolution block fed to subsequent 24 ResNet-block. Each resNet block is characterized by maxpool2d followed by batch normalization followed by convolution then PReLU, followed by conv2d and then batchnorm2d.

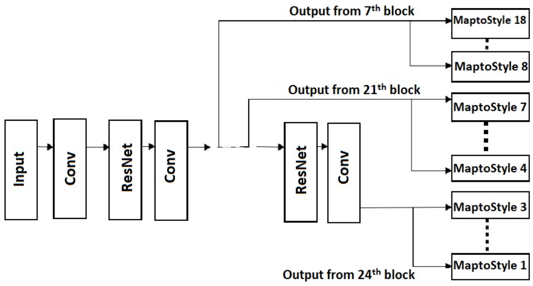


Figure 4: Age Encoder

.

Again the output is inputted to next block that is also a convolution block which is characterized by average pooling then conv2d, then Relu activation again conv2d and finally sigmoid activation function.

There are 24 resNet+convolution block, the three early style namely middle, coarse and fine are taken from 7, 21, 24 block , which are then fed to map-to-style block. The map-to-style block is a convolution neural network block . The three style is changed to 18 * 512 latent vector code. The fine style form 1-3 latent vector code, middle form 4-7 and coarse form 8-18 latent vector code out of 18 different style code of size 512 each.

The first three style/ feature are c1, c2 and c3 of dimension (128, 64, 64), (256, 32, 32) and (512, 16, 16) respectively. The feature from c2 is up sampled and is added to c3 and thus the resulting output is of dimension (512, 32, 32) and say this is style p2. Similarly, the output from c1 is again up sampled and is added to p2 resulting p1 of dimension (512, 64, 64). The output of c3 fine style is fed to map2Style block (1-3), which generates the extended latent vector of dimension 512, similarly the output of p2 is fed to map2Style block (4-7) this also produce extended latent vector of size 512 and the output of p1 is fed to map2style block from (8-18) which too produces 512 dimension latent vector. Thus age encoder in total produces 18*512 extended latent vector. This latent vector is called extended because it is in styleGAN domain.

3.2 StyleGAN Generator

The StyleGAN generator is a pre-trained styleGAN2 generator. The input to this is 18 style latent vector from the map to style block. W space from map to style is separated from the image space, where the factor of variation is more linear in nature. In the StyleGAN2, at first synthesis block, it is fed with 512*4*4 constant input. The received input is then convolved with kernel size of 3*3. The resulting is passed to (toRGB) block that convert it into RGB channel followed by up sample block that increases the dimension by factor 2. The up sample is done every twice convolution block and thus 18 synthesis block. Thus, the constant input is thus progressively convolved and up sample from 4*4 to 8*8, to 16*16 to 32*32 to 64*64 to 128*128 to 256*256 to 512*512 and finally outputs 1024*1024.

In figure 5, A denotes the linear layer and B denotes broadcast and scaling operation , noise in the single channel. StyleGAN2 has better generator network, the AdaIN operation has been replaced with weight modulation and demodulation step. According to creator of styleGAN2, it improve droplet artifact from the image generated by styleGAN generator, the earlier version which was brought by normalization step from adaptive instance normalization. Style vector code per layer is computed as from equation 2. The convolution weights w are computed as below for modulation as in equation 3.

$$s_i = f_{Ai}(W_i) \tag{2}$$

$$W^{'}_{i,j,k} = s_i \times w_{i,j,k} \tag{3}$$

Similarly, convolution weight is demodulated as follows where i is the input channel, j is output channel and k is kernel index.

$$W^{''}_{i,j,k} = \frac{W^{'}_{i,j,k}}{\sqrt{\sum_{i.j} W^{'}_{i,j,k} + \in}} \tag{4}$$

Also, another feature called path length regularization has been introduced that motivates a constant step in W+ to get in a non-zero that is a shift of constant magnitude in image generated by generator.

StyleGAN2 leverage the use of residual connections with down-sampling in the discriminator and skip connections in the generator with up-sampling. At the beginning of training time period, the contribution of low-resolution layers is large and subsequently the high –resolution layers take over.

As a result, the generator starts with a learning constant and then proceeds through a sequence of

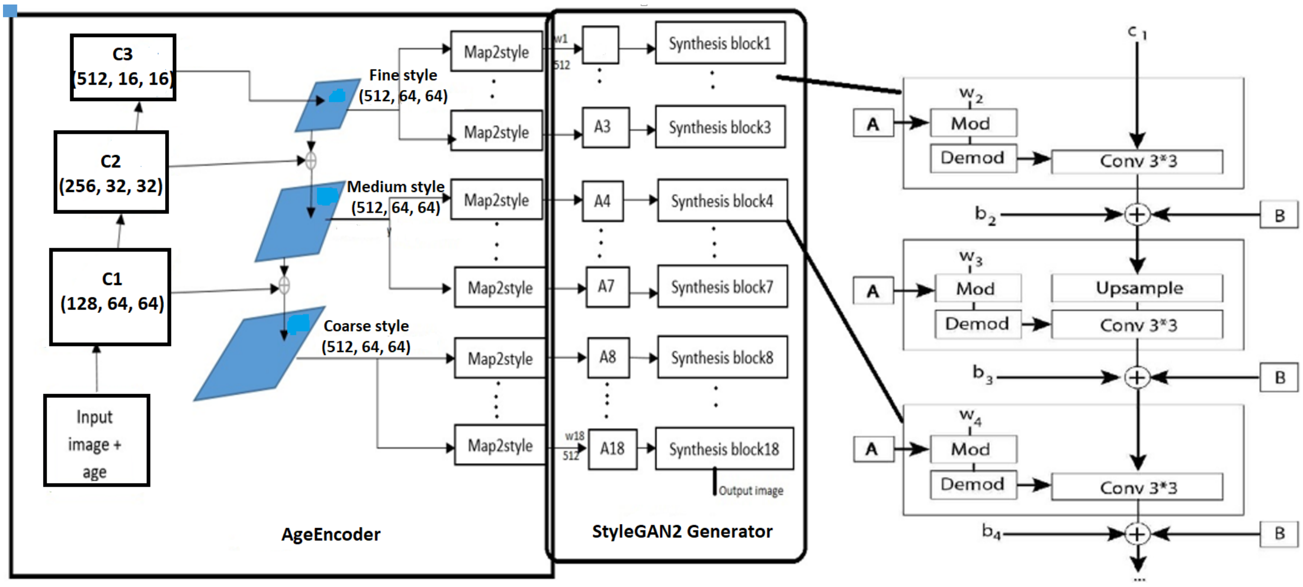


Figure 5: Age Encoder details

blocks, with the feature map being doubled at each block. Each block generates an RGB picture, which is then scaled and summed to give the final full resolution RGB image.

The 18 style help us to control the feature of the generated the output image. The coarse style helps to control pose, hair, face shape, similarly middle feature helps control feature such as eyes and fine styles helps to control color scheme of face. Latent vector 8, 9 is used to control the hair style and color of the source image given target image.

The problem of face aging is solved by using conditional GAN regression model. Here, the condition is target age that the model try to convert the input image to the aged face image of target age. This model takes the estimated age of the image generated by the generator estimated by state of art predtrained model. The L2 loss between the target age and the age predicted by age predictor is used too used train the encoder network of the face aging model. Mathematically,

$$Aging\_loss = \| target\_age - AP(Net(I_{target\_age}) \|_2 \quad (5)$$

In vanilla GAN, the discriminator compare the generated image(Fake) with the target image(True) to compute BCE/MAE/MSE loss. Here, the age predictor use the generated image to predict estimated age. Then L2 lossloss is calculated by comparing the estimated age with target age, the so calculated loss with other losses together is fed to age encoder network to update the weight of the model. The process repeat until the achievable loss saturates/ converge. Hence, the approach for discriminator network to make generator produce more realistic data is done by age predictor, thus this is regression in GAN[15].

3.3 Training Objective function

The training objective for the proposed GAN network is the sum of forward loss and cyclic loss which in aggregate need to be minimized.

3.4 L2 loss

L2 loss is the MSE loss to learn similarities at pixel level between the input image and the target aged face image. As age grow, the shape of face increase, thus this encourage to put higher weight for this loss.

$$L_2(I_{target\_age}) = \| I - Net(I_{target\_age}) \|_2 \quad (6)$$

3.5 Cropped L2 loss

In this, l2 loss is calculated for the cropped part of face image to give more significance to center face.The cropped has been as taken as image[13:227,15:229,:] that is cropped image consider 13 to 227 rows of image and 15 to 229 column for all 3 channel of image.The training image and its cropped part is discussed in datasets section.

$$L_2(I_{target\_age})_{cropped} = \| I - Net(I_{target\_age})_{cropped} \|_2 \quad (7)$$

3.6 Learned Perceptual Image Patch Similarity Loss

High score of LPIPS indicate the image patches are perceptually dissimilar. The perceptual feature of image is calculated by VGG pertained network.

$$L_{LPIPS}(I_{target\_age}) = \| FE(I) - FE(Net(I_{target\_age}) \|_2 \quad (8)$$

3.7 Cropped Learned Perceptual Image Patch Similarity Loss

This focus on the center region of the face image to calculate the cropped perceptual image patch similarity loss.

$$L_{LPIPS}(I_{target\_age})_{cropped} = \| FE(I) - FE(Net(I_{target\_age})_{cropped} \|_2 \quad (9)$$

3.8 Identity loss

Preserving identity of the face is the key point when a face is transformed during training. Thus identity loss is calculated by using cosine similarity of the source image and the input image. Also the identity preservation is less for the large age difference and more for the less age difference between target and source image.

$$L_{ID}(I_{target\_age}) = A*(1-[R(I), R(Net(I_{target\_age})]) \quad (10)$$

Where R is the ArcFace pretrained model. The weight function A(.) is defined by

$$A = 0.25*\cos(\pi*(|source\_age - target\_age)/80.|) + 0.75 \quad (11)$$

The value of A is minimum when difference between source and target age is high and vice versa.

3.9 Aging loss

The aging loss is characterized by the L2 loss between the target age supplied and the age predicted by the pretrained age predictor AP.

$$Aging\_loss = \| target\_age - AP(Net(I_{target\_age}) \|_2 \quad (12)$$

3.10 Cyclic Loss

Cyclic loss is calculated for the robustness of the network because the network must be able to generate the source image if the image generated by the network with predicted age by age predictor is passed as input to the network.

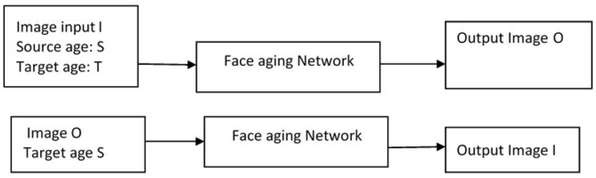


Figure 6: Cycle Loss

.

Loss Forward = J* L2 loss + K* cropped L2 Loss + L* LPIPS loss + M* cropped LPIPS loss + O * identity loss + P* aging loss

Where J, K, L, M, N, O, P are weights related to losses.

4. Datasets

The datasets for this paper has been taken from UTKFace kaggle repository. This repository has two folder UTKFace and crop_part1 with 23708 face image and 9780 images respectively. The filename for image is in convention such as age_gender_X_serialno. In this paper, for training the model 3200 images are taken and 800 images are taken for validation of the model. Sample images from dataset with their cropped one is displayed together.

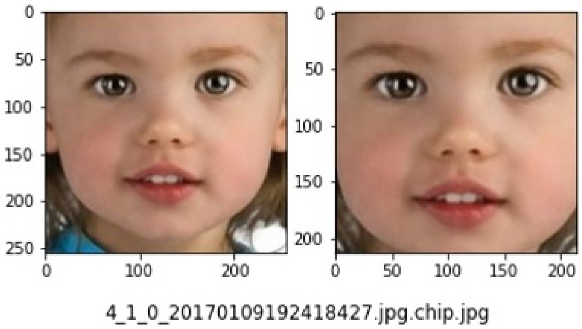


Figure 7: Sample datset image with cropped image side by

.

5. Results and Discussion

5.1 Implementation Details

The model is able to generate progressed face image given input image of age below 80. The obtained image reflect aged version of given input. The output image are in age group from 1-10,11-20, 21-30, 31-40, 41-50, 51-60, 61-70 and 70-80. The result are obtained by implementing following hyper parameter:

Input image size: 256*256, Output image size: 256*256, Image Batch size = 2, Number of input

Channel = 4, Learning rate for Ranger Optimizer = 0.0001, Alpha for ranger = 0.5, Number of batch for ranger optimizer = 6, Iteration = 16,000, L2 lambda = 0.25, L2 lambda crop = 1, LPIPS lambda = 0.1, LPIPS lambda crop = 0.6, ID lambda = 0.1, L2 aging lambda = 7, Cycle lambda = 1

5.2 Result and Discussions

The loss of the encoder is high in the beginning because at the beginning of the training, the model has not seen enough data. As the training progress, the encoder learns the data resulting in gradually decrease in the loss. The loss decreases exponentially but sudden peaks in the loss graph is visible. This sudden rise in the graph came due to the reason that model fail to construct the target age nearly. During training, the target age for construction is uniformly chosen between 1 to 80. After 14000 iteration the loss remains almost constant with some variance. The model models the progressed aged face image only with the single image as input. Also as people get old, the color of the hair also get changed. The proposed model work face but it can also be used to manipulate color of the hair. The hair color of the source face image can be changed to hair color of target image with this same trained model. For this, eighteen style latent vector of the source and target is evaluated and later the 8,9 style latent vector of source is replaced by target image. Thus the resultant latent vector is passed to generator for head hair color change.

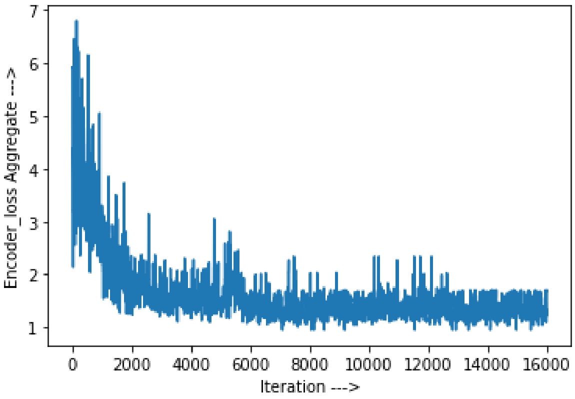
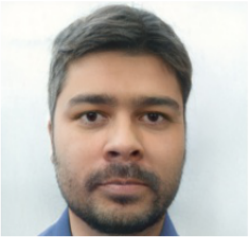


Figure 8: Encoder Loss

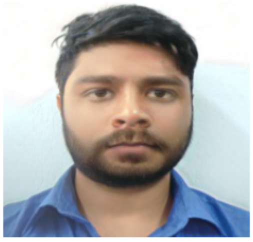
5.3 Testimony Images

The following images on test datasets are generated by trained model.

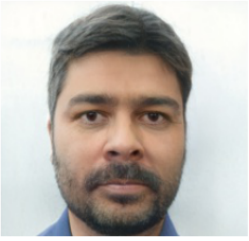
Target age : 30
recovered age: 35
Proposed Model



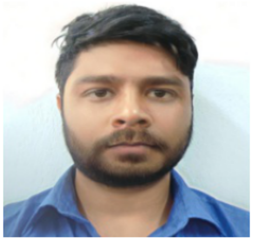
Target age : 30
recovered age: 27
HRFAE



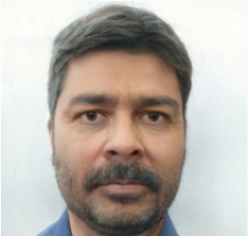
Target age : 40
recovered age: 44
Proposed Model



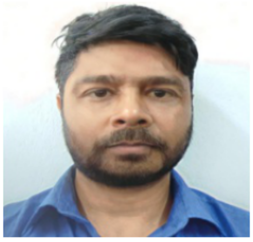
Target age : 40
recovered age: 38
HRFAE



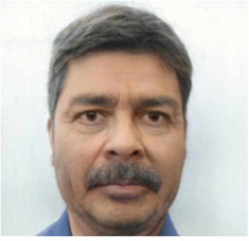
Target age : 50
recovered age: 50
Proposed Model



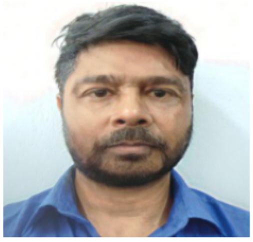
Target age : 50
recovered age: 52
HRFAE



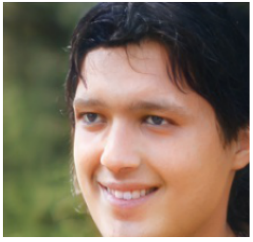
Target age : 60
recovered age: 57
Proposed Model



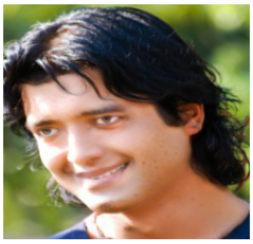
Target age : 60
recovered age: 64
HRFAE



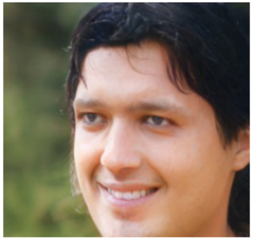
Target age : 21
recovered age: 19
Proposed Model



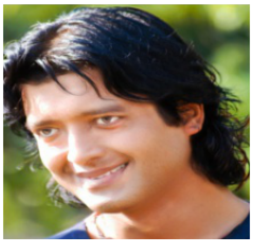
Target age : 21
recovered age: 20
HRFAE



Target age : 30
recovered age: 28
Proposed Model



Target age : 30
recovered age: 27
HRFAE



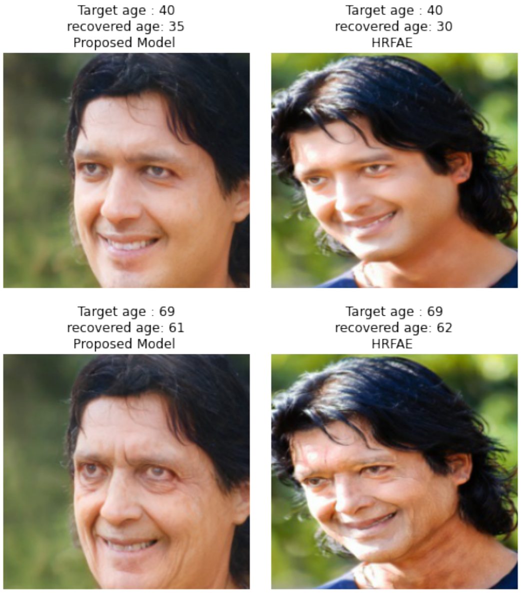


Figure 9: Output image generated by proposed model on left column and on right by HRFAE

6. Conclusion

The proposed model is based on GAN framework that synthesis the progressed aged face image. The input image is converted into intermediate extended latent vector code by age encoder to feed into generator to generate aged image of face. The result generated by proposed model and that by state of art model HRFAE [14] is comparable as seen from recovered age accuracy. The model is able to generate plausible progressed aged face image in the range of 1 to 80 for single front facing image. Moreover, the proposed model can be further improved by preserving the background of face image, and generating multiple aged face image from group photos is a part of future work.

References

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[2] Yunfan Liu, Qi Li, and Zhenan Sun. Attribute-aware face aging with wavelet-based generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11877–11886, 2019.

[3] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[4] Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5810–5818, 2017.

[5] Hongyu Yang, Di Huang, Yunhong Wang, and Anil K Jain. Learning continuous face age progression: A pyramid of gans. *IEEE transactions on pattern analysis and machine intelligence*, 43(2):499–515, 2019.

[6] Brandon Amos, Bartosz Ludwiczuk, Mahadev Satyanarayanan, et al. Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science*, 6(2):20, 2016.

[7] Yujun Shen, Jinjin Gu, Xiaoou Tang, and Bolei Zhou. Interpreting the latent space of gans for semantic face editing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9243–9252, 2020.

[8] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.

[9] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

[10] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. Encoding in style: a stylegan encoder for image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2287–2296, 2021.

[11] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.

[12] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020.

[13] Roy Or-El, Soumyadip Sengupta, Ohad Fried, Eli Shechtman, and Ira Kemelmacher-Shlizerman. Lifespan age transformation synthesis. In *European Conference on Computer Vision*, pages 739–755. Springer, 2020.

[14] Xu Yao, Gilles Puy, Alasdair Newson, Yann Gousseau, and Pierre Hellier. High resolution face age editing. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 8624–8631. IEEE, 2021.

[15] Lucy Chai, Jonas Wulff, and Phillip Isola. Using latent space regression to analyze and leverage compositionality in gans. *arXiv preprint arXiv:2103.10426*, 2021.