Prediction of face age progression with generative adversarial networks

Neha Sharma[1] · Reecha Sharma[1] · Neeru Jindal[2]

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Face age progression, goals to alter the individual's face from a given face image to predict the future appearance of that image. In today's world that demands more security and a touchless unique identification system, face aging attains tremendous attention. The existing face age progression approaches have the key problem of unnatural modifications of facial attributes due to insufficient prior knowledge of input images and nearly visual artifacts in the generated output. Research has been continuing in face aging to handle the challenge to generate aged faces accurately. So, to solve the issue, the proposed work focuses on the realistic face aging method using AttentionGAN and SRGAN. AttentionGAN uses two separate subnets in a generator. One subnet for generating multiple attention masks and the other for generating multiple content masks. Then attention mask is multiplied with the corresponding content mask along with an input image to finally achieve the desired results. Further, the regex filtering process is performed to separates the synthesized face images from the output of AttentionGAN. Then image sharpening with edge enhancement is done to give high-quality input to SRGAN, which further generates the super-resolution face aged images. Thus, presents more detailed information in an image because of its high quality. Moreover, the experimental results are obtained from five publicly available datasets: UTKFace, CACD, FGNET, IMDB-WIKI, and CelebA. The proposed work is evaluated with quantitative and qualitative methods, produces synthesized face aged images with a 0.001% error rate, and is also evaluated with the comparison to prior methods. The paper focuses on the various practical applications of super-resolution face aging using Generative Adversarial Networks (GANs).

Keywords Generative adversarial networks (GANs) · Face age progression · Face super-resolution · Age estimation

✉ Neha Sharma
  nehaleo_sharma@yahoo.com

Extended author information available on the last page of the article

1 Introduction

Face age progression is becoming a widely used technique in the modern era as it serves numerous applications. For an instance, in law enforcement where the face age progression helps to find the missing children or missing person with the help of their previous photo [32, 43, 61], in face recognition [21, 38, 41], and in facial analysis like e-commerce platforms [16, 41], etc. Further, the use of a biometric system is to identify a person based on some particular characteristics. Nowadays, the face of humans is the most leading biometric area [41]. Because it is unique for each individual and provides a guarantee of accuracy and security. However, face age progression remains a challenging task because of facial variations like illumination effects, many intrinsic and extrinsic effects on the face [38]. An abundance of research is present to resolve these practical issues. Face age synthesis or face age progression is a human face appearance at different ages though preserving their unique identity. It is indicated by face geometry, skin texture, skin color, etc. The facial characteristics of human changes with progress in life [40] as illustrated in Fig. 1. Figure 2 shows the synthesized super-resolution face age-progressed images.

First and foremost, the pandemic of COVID-19 has created the touchless face recognition system at the topmost. It is accepted globally to avoid the virus spread, leading from all the other biometric systems that require fingerprints, or any touch-based biometric services. The face of humans is used to provide immense authentication and security to the individual [42].

In law enforcement for keeping security high, the digital biometric passport of the person is used for face matching at the border check. The appearance of person changes as humans progresses in life in terms of both skin texture and geometric shape etc. So, face age progression is unique for humans and it can provide authentication for the specific person's image from a database of digital passports.

In the health sector, deep learning has made substantial developments. The need has been shown for regularly checking the patients, remote consulting, health insurance ID card, so face recognition can provide health and comfort to humans [46].

Further, the touchless system of face age progression using machine learning can be useful for several facial recognition systems like banking [12]. Also, a possibility in the future with the help of face age progression in face recognition system, can decrease the number of visits required for updating the person's photo to many service providers thus, delivers ease to the customers.

1.1 Face age progression with GANs

In various fields, GANs are robust performers. GANs have generated impressive results in many fields [13, 14, 55, 58]. They have shown wider applications in various fields such as image-to-image translation [19, 30, 66], text to speech generation [20, 47], and many more.
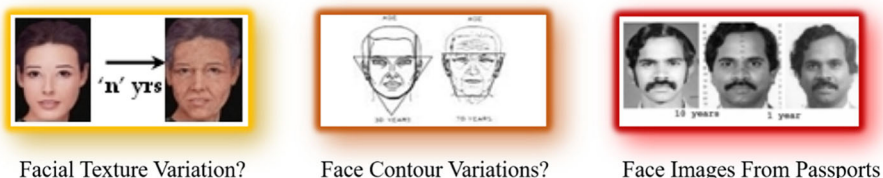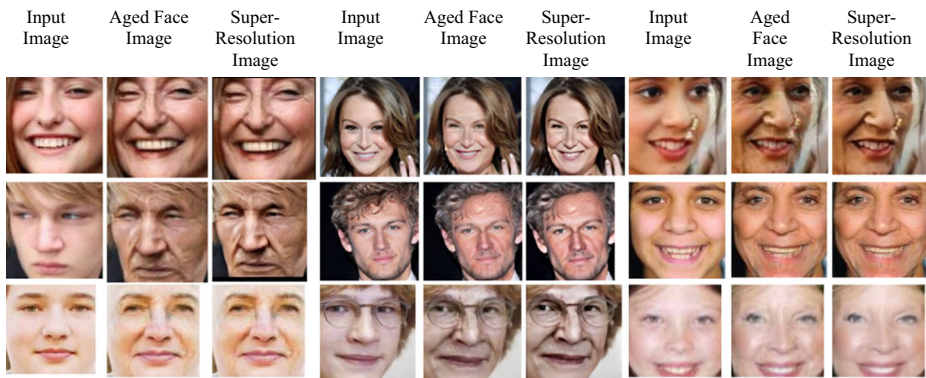

Facial Texture Variation?   Face Contour Variations?   Face Images From Passports

Fig. 1 Changes in the face with age progression [40]

Fig. 2 The synthesized super-resolution face age-progressed images. The diversity of the input images are shown in the image with age (child, youngster, adult), ethnicity, gender, facial expression

But the biggest disadvantage is that GANs are also used for generating fake media content and are the expertise behind Deep fakes [34, 54, 59]. Face age progression using GANs have attained a lot of attention in the facial verification system. GANs have produced remarkable results in face age progression [2, 5, 37, 63]. The proposed work focuses on the face aging with a super-resolution of an image to determine its practical possibility.

The main contributions of the proposed work:

1. The proposed work presents a combined approach for face age progression using AttentionGAN and SRGAN. To the best of the author's knowledge from literature, AttentionGAN is a primary attempt on face aging task in the proposed work.
2. In the proposed work, the use of a regex filter reduces the computational complexity as well as training time by selecting the synthesized face images.
3. The efficacy of the proposed method is assessed on the number of publicly existing datasets named UTKFace, CACD, and cross-dataset evaluation on IMDB-WIKI, CelebA, FGNET datasets.
4. The validation of the proposed work is verified by the various attacks like the presence of pose, expression, makeup, illumination. Moreover, simulation outcomes are also compared to existing approaches.

2 Related work

Previous work on face age progression had a focus on facial attributes like geometric growth of face [24, 44], wrinkles [4, 6, 35], face sub-regions [48, 49], and various techniques [45, 53, 62]. Some existing face age progression techniques are given in Table 1. The aging process was the cascade flow of mean faces from age clusters using their eigenfaces. Where the eigenfaces were termed as an appearance-based method that tried to capture the different collections of face images. After that, it took this information to code and compare the images of each face. Then, deep learning had gained tremendous attention in the computer vision field that had produced a remarkable result in many fields [3, 9, 15, 31, 33]. Now, the research on GAN is exploring the main aspect that is, an improvement in the training process and the second is positioning the GANs in real-world applications [17, 60]. The major goal of GAN is

Table 1 Some existing face age progression techniques

Method	Objective and approach	Dataset	Age groups	Evaluation metrics
Learning continuous pyramid of GANs, 2019 [63]	Pyramid adversarial discriminator at the various scale was used for aging effects and adversarial scheme for training one generator and multiple parallel discriminators.	MORPH, CACD, FGNET	31–40, 41–50, 50+	Quantitative analysis- Face++ verification method was used. Qualitative analysis- Comparison to prior work (user study).
Age-cGAN, 2017 [2]	First GAN to generate high-quality images. Optimization of a latent vector was used for identity preservation of original face image.	IMDB-WIKI cleaned dataset	0–18, 19–29, 30–39, 40–49, 50–59, 60+	Identity preservation (Open Face software).
Personalized face aging using aging dictionary 2015 [45]	The aging layer along with a personalized layer in linear combination generated automatic aging images of the face.	CACD, MORPH	0–5, 6–10, 11–15, 16–20, 21–30, 31–40, 41–50, 51–60, 61–80	Qualitative comparison - Ground truth. Quantitative comparison - User study with prior work.
Face age progression using exemplar face templates, 2020 [11]	Generation of face template based on the average face for given ethnicity and age. Target age was generated by applying an appropriate face template.	Manual collection to gather ethnicity-oriented dataset, FEI, MORPH, FGNET	20, 30, 40, 50, 60, 70, 80	Qualitative analysis - Comparison to prior work. Quantitative analysis - User study, CS (Cosine Similarity). SSIM (Structural Similarity Index) metric used, image map method (IMG-online).
CAAE, 2017 [65]	Face aging was achieved by traversing on a manifold($\mathcal{M}$).	CACD, MORPH, FGNET	0–5, 6–10, 11–15, 16–10, 21–30, 31–40, 41–50, 51–60, 61–70, 71–80	Qualitative and Quantitative analysis - Comparison to ground truth (user study). Comparison to prior work (user study).
RFA with hierarchical Autoregressive memory, 2018 [57]	In the RFA framework, series of face-aged images were generated automatically and hidden layers were connected autoregressively.	LFW, CACD, FGNET	0–5, 6–10, 11–15, 16–20, 21–30, 31–40, 41–50, 51–60, 61–80	Qualitative analysis with prior work (user study). Quantitative analysis - Confusion matrix, age accuracy, identity preservation.
IPcGAN, 2018 [56]	Pre-trained AlexNet was adopted for identity preservation in feature space.	CACD	11–20, 21–30, 31–40, 41–50, 50+	Qualitative comparison with prior work. Quantitative analysis - Image quality, Inception score, age classification, face verification (user study).

to acquire the distribution of generator p_a to come close to real data distribution p_b and cycle consistency loss tries to get back the original input from the synthesized image while it preserves the identity of an image. For each image, u from domain U, u to $G(u)$ to $F(G(u)) \approx u$, is a forward direction cycle consistency loss. U and V are two domains, u and v are the images present in the respective domains.

For each image u (domain U), u to $G(u)$ to $F(G(u)) \approx u$, signifies the forward direction cycle consistency as:

$$min_G max_D E_u \sim p_b \log[\mathrm{D}(\mathrm{u})] + E_v \sim p_a log[1 - D(v))\Big] \qquad (1)$$

$$L_{cyc}(\mathrm{G},\mathrm{F}) = E_{u \sim Q_{data(u)}} \Big[\| \mathrm{F}\Big(\mathrm{G}(\mathrm{u}) - \mathrm{u}\| \Big] + E_{v \sim Q_{data(v)}} \Big[\| G\Big(F(v) - v\| \Big] \qquad (2)$$

where G and F are generators, p_a is generator probability distribution data, p_b is real data distribution, u is real face image, v is the synthesized image, L_{cyc} is cyclic loss, and D is a discriminator.

The image-to-image conversion and pix2pix [19] used a paired dataset. This problem was eliminated by the two-domain image-to-image transformation model with an unpaired dataset, CycleGAN [66], and DiscoGAN [23] that had shown remarkable results in various domains like a horse to zebra conversion and many more. On the contrary, failure included not changing the shape of an object during the transition. In 2018, Spatial fusion GAN [64] for image synthesis combined the geometry and appearance synthesizer to attain realism in respective domains. Identity loss was introduced to preserves the features of an original face image. The MI-GAN framework worked on retinal images that generated the synthesized retinal images and their segmented masks. The synthesized images were realistic. The model learned the features from a small training dataset and outperformed in comparison to other methods [18]. Moreover, a contactless biometric system provided many benefits such as personal hygiene was maintained in a contactless system, it became more convenient, free from a contaminated surface. Thus, a multi-modal biometric system using face and palmprint can be used to provide more security and authentication to the system [7]. Palmprint has several unique features that can be used for the person's identification [39]. A novel contribution was shown by the researcher in the field of palmprint recognition. In multi-instance contactless palmprint recognition (fusion system), features of left and right hands were extracted using 2-dimensional DCT (Discrete Cosine Transform) to compose a dual-source space. The experiment performed had shown that the designed Dual Power Analysis (DPA) outperformed single-source dual power analysis [29]. Besides this, StarGAN [8] was a multi-domain image-to-image conversion having one generator and a discriminator. This GAN had shown robustness for handling more than two domains. Progressive growing GAN [22] described the new training method for GAN where generator and discriminator grow progressively and produced astonishing results. The novel approach towards face age progression and regression with template face, considered the average face for ethnicity and age. The template face helped to generate the target face image for age progression and regression. The method had achieved accuracy and efficiency [11]. Further, the Laplacian Pyramid of Adversarial Networks (LAPGAN) [10] introduced a cascading convolutional neural network in a Laplacian pyramid framework that generated coarse-to-fine style output images. From the literature survey for future directions, some important novel contributions are mentioned to

enhance security in the field of biometrics. To enhance the robustness against various security attacks, in a palmprint texture code. Because it suffers alignment issues while matching which was its obstacle to be directly adopted in biometric cryptosystems. Thus, 2DPHC (2DPalmhash code) based Fuzzy Vault had improved the key recovery rate [26]. Further, to enhance the privacy and security in the palmprint biometric system, a novel dual-key-binding scramble and 2D palmprint phasor algorithm was introduced. Thus, protect palmprint and information security. The existence of lacking in cancelability in the existing palmprint cryptosystem was overcome. This scheme could also be used in other biometric with some alterations. Also, the applications can be useful to further palmprint texture feature coding [27]. Palmprint authentication with the remote cancellable method based on multi-directional two-dimensional palm phasor (MTDPP) was proposed [28]. MTDPP was used as a cancellable palmprint template, provided biometric protection. Thus, the multi-model biometric with face and palmprint can be taken as the future scope in the contactless technology to provide more authenticity to an individual.
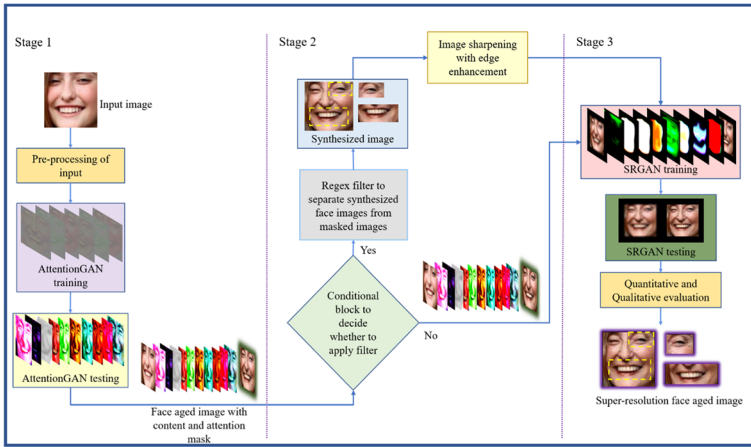
3 The proposed work

The suggested work translates an input face image to the required face aged image and produces high-resolution images with less computation time and storage space by the use of a filter process. Further, image sharpening with edge enhancement is used to provide better quality input to SRGAN.

For this aim, a three-stage learning framework is shown in Fig. 3. In this work, UTKFace and CACD large-scale publicly available datasets are used for training the network. The input images of the face are first pre-processed to attain only RGB images from UTKFace and CACD datasets. Then, manually separated into the four age groups 0–20, 21–40, 41–60, 60+ with good quality face images. Then training, test, and validation datasets with input and target images are prepared for experimental results. The images are resized and cropped to 100×100.

> **Stage 1:** The pre-processed images are fed to AttentionGAN generator G to perform the face age progression with an image-to-image conversion AttentionGAN (scheme-II) [51]. The generator acquires both background and foreground attention to generates a high-quality face image with its identity preservation. The unique property of AttentionGAN is that the generator focuses on the foreground of the required image and simultaneously preserves the background of an input image efficiently with the help of attention and content masks. However, the input image goes to a generator that has sub-module parametric sharing encoder G_E, content mask generator G_C, and attention mask generator G_A. The p-1 content masks are generated by generator G_C. Also, p-1 foreground attention masks $\{A_v^f\}_{f=1}^{n-1}$ and one background attention mask A_v^f are simultaneously generated by G_A. The corresponding attention and content mask get multiplied with an input face image as shown in eq. (3) and generate the target face aged image. The high intensity in the attention mask contributes to change in the facial attributes. The use of the various attention and content masks helps in generating the face aged output image. Mathematically, the output is expressed as shown in eq. (3):

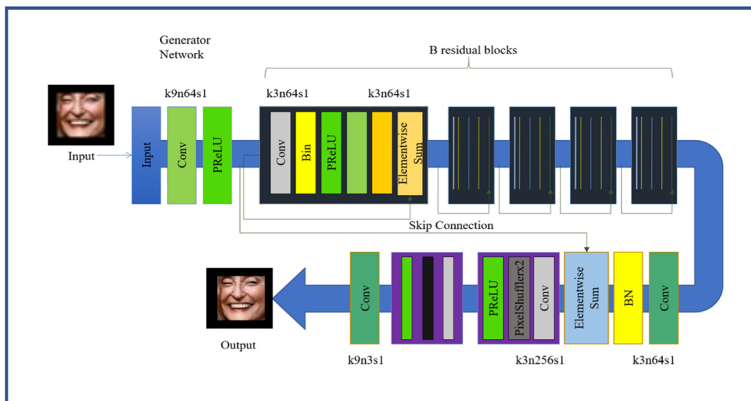$$G(u) = \sum_{f=1}^{n-1} \left(C_v^f * A_v^f \right) + u * A_v^f, \tag{3}$$

where $\left(\{A_v^f\}_{f=1}^{n-1}, A_v^f \right)$ are p attention mask, u is the input image, C_v^f is a content mask, $G(u)$ generated target face aged image, U and V are two domains and u and v are images in the respective domain.

(a)
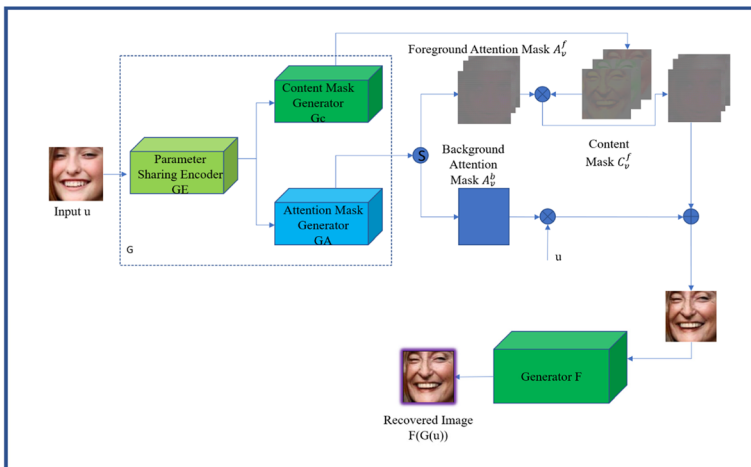
(b)

(c)

Fig. 3 (a) The proposed workflow comprises three stages, (b) Stage 1 process, (c) Stage 3 process

For the cycle consistency loss, the generated aged image is given to another generator F. Thus, the F generator similarly generates the content and attention masks for the foreground along with the background image. Therefore, combines them to generates the recovered face image. Again, in the generator F, various masks help to preserve the image information and get back the input image with minimum losses, thus preserving the identity of an image. The reconstruction of generated image G(u) to the original input image u is expressed mathematically in eq. (4):

$$F(G(u)) = \sum_{f-1}^{n-1} \left(C_u^f * A_u^f \right) + G(u) * A_u^b, \tag{4}$$

where $F(G(u))$ is the reconstructed image that should be very close to the original image u. A_u^f, A_u^b, C_u^f, G (u) are foreground attention mask, background attention mask, content mask, and generated image respectively. F is a generator similar to generator G which also consists of three subnets parametric sharing encoder F_E, attention mask generator F_A, content mask generator F_C. Where F_C tries to generate p-1 content mask and F_A tries to generate p attention mask for foreground $\{A_u^b\}_{f=1}^{n-1}$ and a background image A_u^b. Then, the attention and content mask with the generated face image get multiplied according to eq. (4) to generate a reconstructed image. Mathematically, the objective of optimization in AttentionGAN Scheme II is expressed as shown in eq. (5):

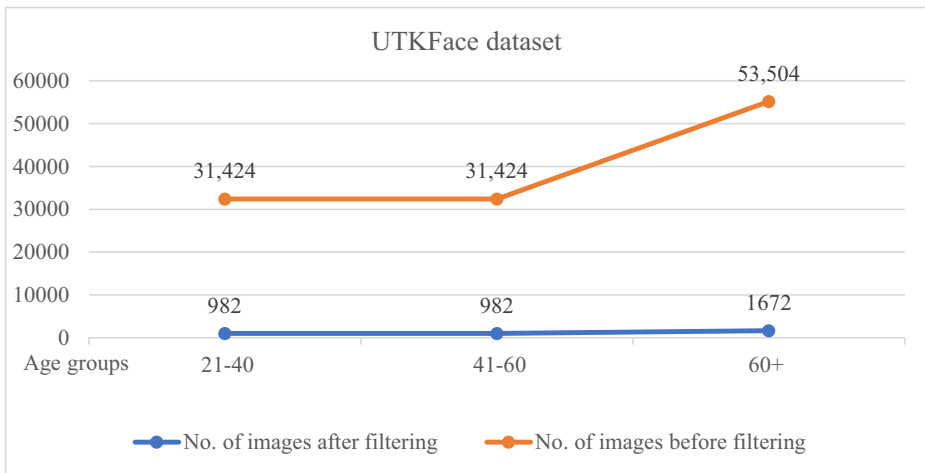$$L = L_{GAN} + \lambda_{cycle} * L_{cycle} + \lambda_{id} * L_{id}, \tag{5}$$

where L_{GAN} is GAN loss, L_{cycle} is cyclic loss and L_{id} is identity preserving loss. Further, λ_{cycle} and λ_{id} are the parameters to control each term relation.

Stage 2: Output of AttentionGAN is fed to a conditional block where it is decided whether to apply the regex filter or not.
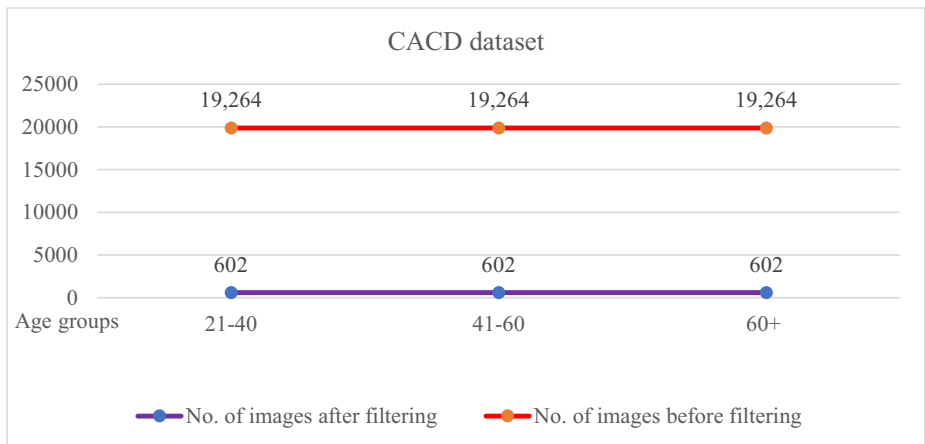
If conditional block output is yes, then a regex filter selects the synthesized face aged images from AttentionGAN. Because, AttentionGAN output consists of the synthesized face images, attention masks and content masks images. Thus, the use of the regex filter process aids to reduces the computation time required for further SRGAN training. The filtered synthesized face images are approximately 3% in comparison to the total output images of AttentionGAN for each age group in UTKFace and CACD datasets as shown in Figs. 4 (a) (b). Then, image sharpening using edge enhancement is performed on the filtered face aged images which provide better input to SRGAN [25] training. Also, SRGAN primarily learns the shape, texture, and color of the object and amend the output images with a few sharp edges [50]. So, image sharpening with edge enhancement is used to sharpen the edges of face aged images. Thus, concentrating on the edges of the image as the rest of the image is unchanged, thus provides sharpening at the edge. The process of filtering and image sharpening is described in Algorithm 1. The limited dataset with good quality images is now fed to SRGAN. The reduced SRGAN training time after the filtering process for each age group is approximately 2 h.

If conditional block output is no, then the entire output of AttentionGAN is directly fed to SRGAN training. But, with this method, the complete training of SRGAN took approximately 26 h, due to the presence of various unwanted images in SRGAN training. The unwanted images are content and attention masks of aged faces which are not required for SRGAN training. As only face aged images are required for the final output.

UTKFace dataset

Age groups	21-40	41-60	60+
No. of images before filtering	31,424	31,424	53,504
No. of images after filtering	982	982	1672

(a)

CACD dataset

Age groups	21-40	41-60	60+
No. of images before filtering	19,264	19,264	19,264
No. of images after filtering	602	602	602

(b)

Fig. 4 Number of images before and after filtering for (a) UTKFace and (b) CACD datasets

Stage 3: SRGAN training is performed in stage 3 to get the final output image. Now, when the image sharpening output is fed to SRGAN training. In this case, SRGAN training is done on high-quality synthesized images. Then, testing generates the super-resolution images at the output. This process reduces the computation complexity as well as training time. On the other side, when the face aged images with content and attention masks are directly fed to SRGAN training, it increases the complexity and training time.

In SRGAN, where the residual blocks help to build the base model and perceptual loss for the optimization. Thus, enhance the general visual quality of a face image. The generator network along with batch normalization layer (BN) and dense skip connections combining the various facial features at different levels produce the super-resolution image at the output. The generator network

in Fig. 3 (c) shows the details with a corresponding number of feature (n, kernel size (k), and stride (s) in each convolutional layer. While training, SRGAN tries to synthesize a super-resolution image by down sampling the input high-resolution face image to a low-resolution face image. Then, a discriminator attempts to differentiate among synthesized super-resolution images from a real high-resolution images.

Further, a perceptual loss is a weighted sum of the content loss and adversarial loss as shown in eq. (6):

$$L_p = l_c + 10\text{–}3 \; l_{adv}, \tag{6}$$

L_p is perceptual loss, l_c is content loss, and $10^{-3} \; l_{adv}$ adversarial loss. Content loss comprises VGG loss and MSE loss. MSE loss is a pixel-wise error between the super-resolution generated image and the original image. VGG loss is a feature map generated by nth convolution before mth maxpool layer within the VGG19 network. It is denoted with the symbol φ (m, n).

The adversarial loss is the probabilities of discriminator general training samples. It is expressed as shown in eq. (7):

$$l_{adv} = \Sigma_{q=1}^{Q} -\log D_{\alpha_D} \left(G_{\alpha_G} \left(\text{I lr} \right) \right), \tag{7}$$

where I ^{lr} is the low-resolution input image, q = 1, . . ., Q training samples, $D_{\alpha_D} \left(G_{\alpha_G} \left(I^{lr} \right) \right)$ is a probability that the reconstructed image is a natural super-resolution image.

Algorithm 1 To give high-quality input to SRGAN from the output of AttentionGAN.

Input: G(u) - Synthesized image; F(G(u) - Recovered image; C_u^f - Content mask in domain V; A_u^f −Foreground attention mask in domain V; A_u^b - Background attention mask in domain V; C_v^f - Content mask in domain U; A_v^f - Foreground attention mask in domain U; A_v^b −Background attention mask in domain U.

Output: Synthesized face images with image sharpening.

1. Extraction of synthesized images from source path (output of AttentionGAN) using regex of Power shell programming, $filter=[regex] "fake_[A-Z] \. (jpg|png)".
2. Move the synthesized face images into the destination path (data file of SRGAN).
3. Image sharpening with edge enhancement using the cv2 library is performed on the synthesized images (data file) to get sharpened images in train file of SRGAN.
4. Split the train, dev, test ratio in 70 (train): 30 (dev/test).
5. Begin training of SRGAN model using the dev file and training file.
6. Begin the testing of SRGAN.
7. Finally, get a super-resolution face aged image.

---

4 Simulation results

Extensive experiment evaluations have been performed to validate the proposed work for producing realistic and super-resolution face aged images. The qualitative and quantitative outcomes are described in subsequent subsections to validate the efficiency of the suggested work.

4.1 Face aging datasets

The experiment is conducted using two benchmark datasets: UTKFace and CACD (the cross-age celebrity dataset). The UTKFace dataset has an age range from 0 to 116 years. Only the UTKFace dataset provides images from zero to five years (babies), six to fifteen years (offspring), above seventy-five (elder people). However, images in CACD have an age range from 16 to 62 years. Few images in CACD have wrong labels or mismatches between the face image and its annotation which makes the dataset very challenging. Some images from UTKFace and CACD datasets are shown in Fig. 5.

For cross-dataset evaluation, three datasets are used: FGNET (The Face and Gesture Recognition Research Network), CelebA, and IMDB-WIKI. FGNET contains a total of 1002 images and is widely used for testing purposes. CelebA provides a face images dataset on a large scale, images are in-wild similar to the CACD dataset. Whereas, IMDB-WIKI dataset has 500 k + face images along with their gender and age annotation. The age range of all face images is from 0 to 100 years.

4.2 Training and implementation scheme

The training process details are illustrated in Fig. 6. Individual age groups have been trained with 200 epochs and batch size of 4 for the aging process, on GTX 1660 Ti with a GPU, Windows 10, and operating system of 64-bit, i7 processor. CACD and UTKFace datasets run for 200 epochs for AttentionGAN training. For SRGAN training, CACD and UTKFace datasets run for 500 epochs. Moreover, the input image takes an RGB face image with a crop size of 100 and generates the p attention masks and p-1 content masks, multiplying the corresponding mask and input image produces the target face aged image. Thus, the least square loss is used for stabilizing the training of the network with a learning rate of 0.0002. In the cross-dataset evaluation, a total of 1000 images are randomly sampled from CelebA, IMDB-WIKI, and FGNET datasets. To avoid data imbalance in CACD, because age range is from 16 to 62 years old thus, limiting the 60+ years age images and no children images. So, the

| 6 | 14 | 20 | 23 | 31 |

| 40 | 48 | 54 | 60 | 62 |
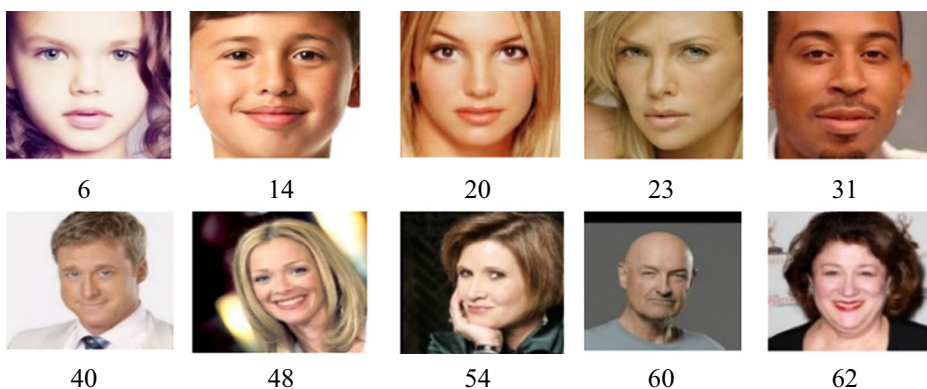
Fig. 5 Ten images from the UTKFace dataset (first row) and the CACD dataset (second row) with their age annotation
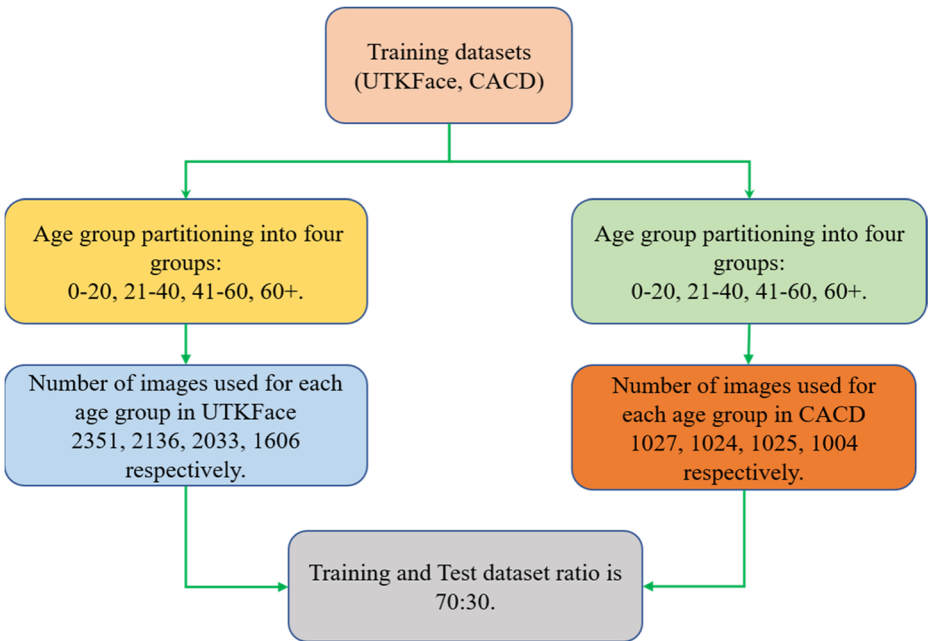
Fig. 6 Training datasets partitioning for the proposed work

number of training images selected with a range of 1024 approximately, to create a balance between each age group evaluation.

Also, a comparison with other GAN-based methods for the execution of training time is shown in Table 2. Training time depends upon various factors such as system architecture, number of images used, quality of images etc.

4.3 Face aging results

The qualitative and quantitative assessment is performed to show the effectiveness of the proposed work. Figure 2 in section 1 shows the super-resolution face aged images. It clearly shows that the proposed work has achieved convincing results. Figure 7, Fig. 8 shows the

Table 2 Analysis of training time for GAN-based methods

Database -No. of images used	GPU used	Method	Training time	Objective	Author
CACD - average 8172 MORPH – 51,699	GTX 1080Ti	Continuous pyramid of GANs	~25 h	Face age progression	Yang et al. 2019 [63]
UTKFace - 5000	NVIDIA Titan X GPU	CycleGAN	~ 20 h	Face age progression	Thengane et al. 2018 [52]
CACD - average 4080 UTKFace - average 8126	NVIDIA GEFORCE GTX 1660Ti	AttentionGAN+SRGAN	~16 hours	Super-resolution face age progression	The Proposed work

Input Image 0-20	Output Image 21-40	Attention Mask	Input Image 0-20	Output Image 41-60	Attention Mask	Input Image 0-20	Output Image 60+	Attention Mask

Fig. 7 Synthesized face image results for the UTKFace dataset

results from UTKFace and CACD datasets with their corresponding attention mask. It is shown in the results that the proposed work generates realistic images. The proposed work also shows significant results with various face attacks such as the presence of pose change, expression, make-up, illumination, and spectacles. Figure 9, Fig. 10 shows the continuous transition of input face image to corresponding age groups 21–40, 41–60, 60+ for UTKFace and CACD datasets. It is observed that the results on the UTKFace dataset outperform the CACD dataset results. After the manual examination of some input images from two datasets. It is observed that the low performance on the CACD dataset is because the images are taken under high professional settings such as make-up, lighting. While images in the UTKFace dataset are taken under low professional settings. Thus, the more natural the input image better it produces the synthesized results. The cross-dataset evaluation results are taken from only the UTKFace pre-trained model with the age group 60+. The results are shown in Fig. 11 for FGNET, CelebA, and IMDB-WIKI datasets along with their corresponding content mask. It is observed that in cross-dataset evaluation FGNET has better results
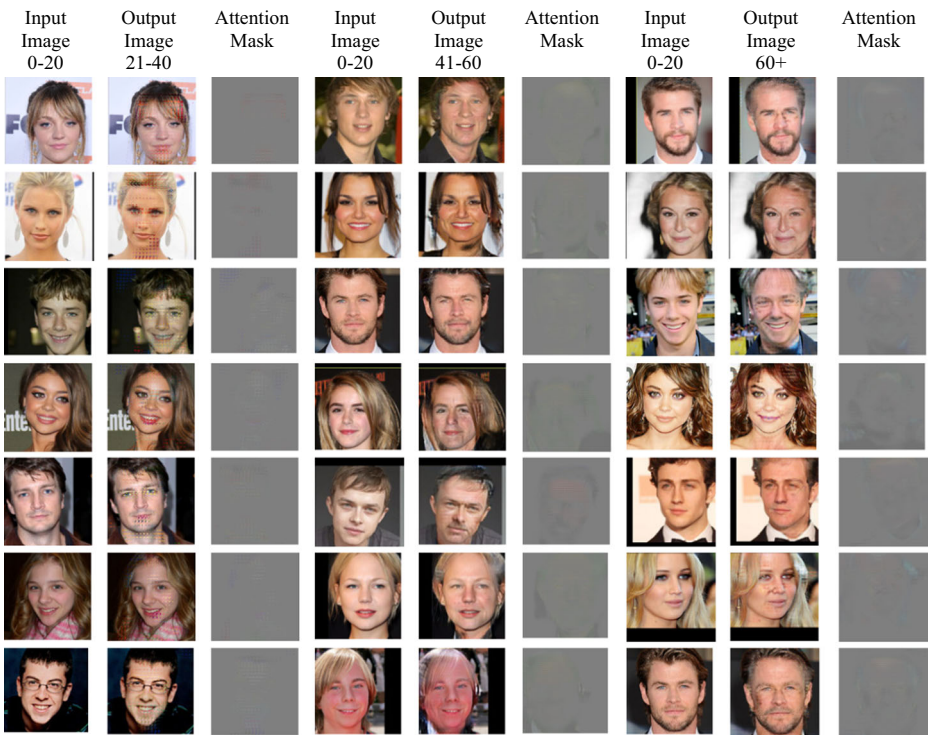
Fig. 8 Synthesized face image results for the CACD dataset

than IMDB-WIKI and CelebA. It has been observed that aging is different among females and males [1]. As the face of a female tends to age faster in comparison to the face of a male. That's why some male images show poor performance. From the synthesized images results, it is observed that output images with high contrast in the mask
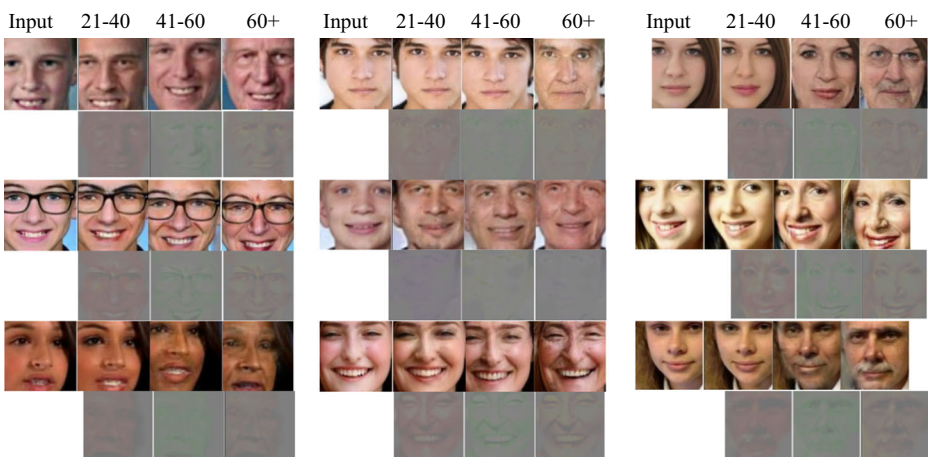

Fig. 9 Continuous face age progression images obtained from the UTKFace dataset
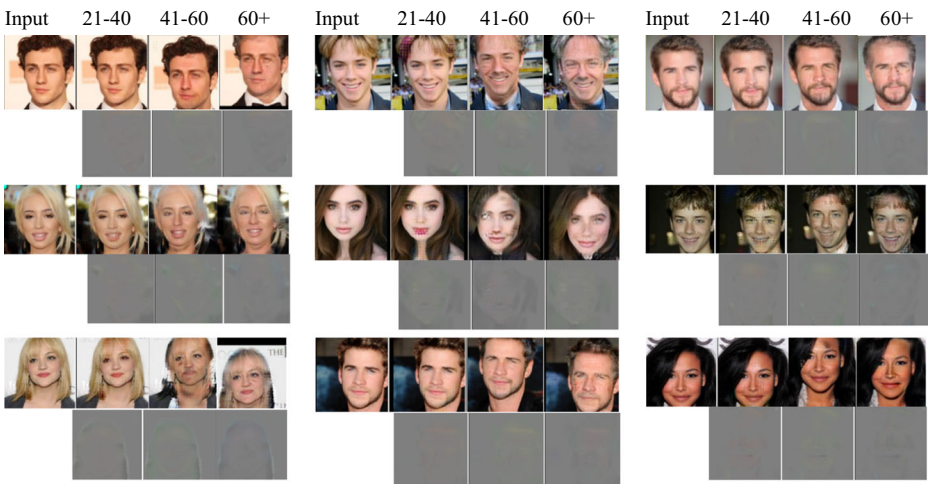
Fig. 10 Continuous face age progression images obtained from the CACD dataset

attain rich information as compared to low contrast in the mask. Further, Fig. 12 shows the face build for men and women to show the aging details changes with age progression and observed that with age progression longer and deeper laugh lines, thinner lip and forehead wrinkles are getting deeper on the face, while the identity of the face is well preserved.
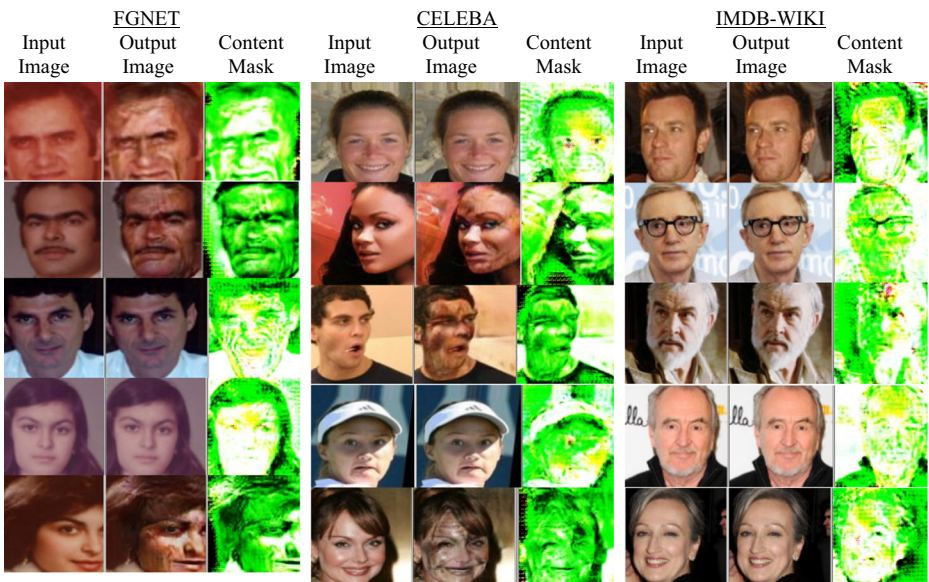

Fig. 11 Cross-dataset evaluation results from the UTKFace pre-trained model

(A) Forehead and front hairline reduction,
(B) Eye deepening, (C) Laugh line and
mouth region aging, (D) Facial hair growth,
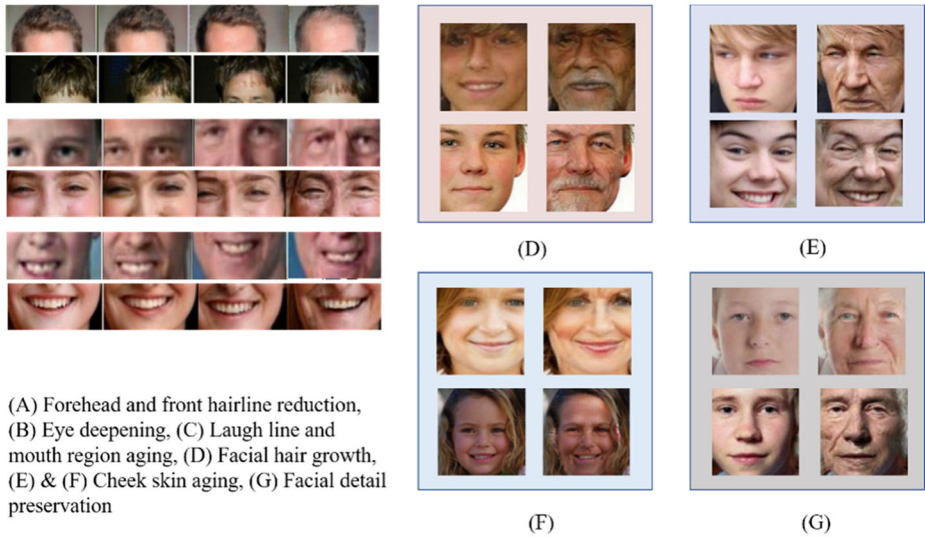(E) & (F) Cheek skin aging, (G) Facial detail
preservation

Fig. 12 Illustrate the visual comparison of face build in men and women

4.3.1 Super-resolution visual assessment

Super-resolution face aged images are shown in Fig. 2. Stage 3 generates the super-resolution results on the face aged image. The details of the image are highly preserved during processing time. Thus, the proposed work generates a super-resolution face aged image that retains rich information in an image as given in Fig. 13.

4.4 Quantitative evaluation

4.4.1 Age estimation evaluation

The age estimation on synthesized face aged images and real face images is performed using Face++ [36] online tool. The proposed work follows the evaluation method used by Yang H. et.al in [63]. The synthesized face images from CACD and UTKFace datasets are used for evaluation. For evaluation, 20 images are selected from real face images and aged face images. The 20 images are evaluated by mean and its standard deviation for three age groups 21–40,
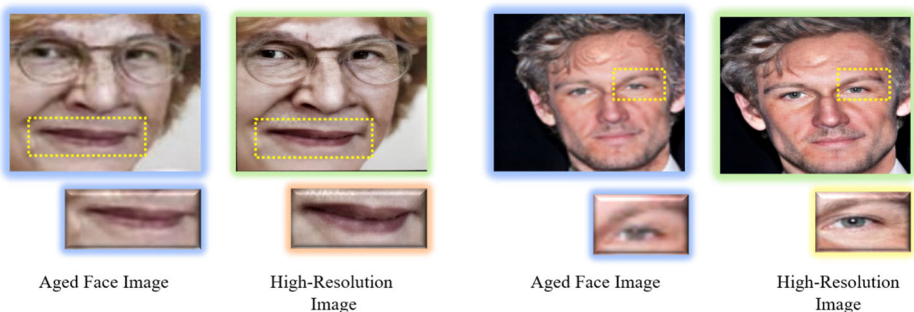

Aged Face Image   High-Resolution Image   Aged Face Image   High-Resolution Image

Fig. 13 Visual evaluation of generated super-resolution images

Table 3 Age estimation results on UTKFace and CACD datasets

UTKFace				CACD		
Age Group	21–40	41–60	60+	21–40	41–60	60+
Synthesized images	29.65 ± 5.07	53.85 ± 6.63	70.75 ± 6.79	36.65 ± 5.34	46.7 ± 5.31	57.4 ± 5.78
Real images	31.4 ± 4.38	52.3 ± 8.17	72.85 ± 9.155	34.3 ± 6.59	56.1 ± 12.25	64.25 ± 8.05

41–60, 60+. The values presented in Table 3, depict that the estimated age of the generated face images is approximately near to the estimated age of the real face images. The graphical representation is shown in Figs.14 (a) (b) and Figs.15 (a) (b) for UTKFace and CACD datasets respectively. It clearly shows that face age progression is a unique process for individuals. Thus, various factors such as external and internal factors affect the appearance of a person. Make-up also has an important role which affects the appearance of the human face.

In addition to that, the confusion matrix evaluation is used for age estimation performance. The confusion matrix is a good technique to depict the performance of the classification problem. The matrix shows the visual and quantitative idea of the right predictions and the types of errors present in the classification problem. Figure 16 shows the generalized confusion matrix labeled block used for evaluation.

Figure 17 (a) shows the confusion matrix of the UTKFace dataset has good evaluation results for the age groups 21–40 and 60+. Similarly, the CACD dataset in Fig. 17 (b) is showing good results for 41–60 and 21–40 age groups.

4.4.2 Identity preservation

Identity preservation is the most important parameter for the evaluation in the face aging method. As it provides authentication of the same person. Thus, the proposed work is evaluated with a face verification score. It illustrates the similarity between the two face images. If the confidence score value exceeds the threshold value, then the similarity scores very high, and images are considered to be similar. Table 4 shows the values of confidence score by Face++. All the values have a high score and beyond the threshold level. Figures 18 (a) and (b) show the graphical representation of the face verification score for UTKFace and CACD datasets.

(a)   (b)

Fig. 14 Age estimation graphical representation for the UTKFace (a) synthesized face images (b) real face images

Fig. 15 Age estimation graphical representation for CACD (a) synthesized face images (b) real face images

Predicted values					
		A1	B1	C1	D1	Total
Actual values	A	True A	False B	False C	False D	A
	B	False A	True B	False C	False D	B
	C	False A	False B	True C	False D	C
	D	False A	False B	False C	True D	D
	Total	A1	B1	C1	D1	

Fig. 16 Generalized confusion matrix with labeled blocks

UTKFace		Synthesized images			
		0-20	21-40	41-60	60+
Real images	0-20	20	0	0	0
	21-40	0	20	0	0
	41-60	0	0	15	5
	60+	0	0	0	20

(a)

| CACD		Synthesized images			
		0-20	21-40	41-60	60+
Real images	0-20	20	0	0	0
	21-40	0	17	3	0
	41-60	0	0	20	0
	60+	0	0	11	9

(b)

Fig. 17 The confusion matrix for age estimation using Face++ tool (a) UTKFace (b) CACD

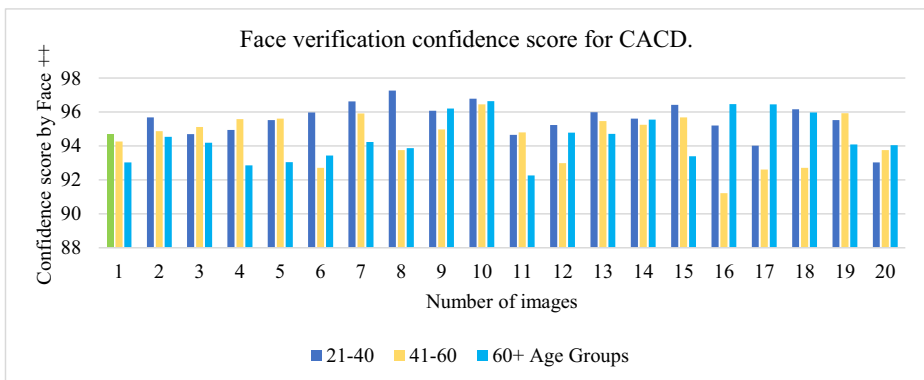**Table 4** Face verification confidence score on UTKFace and CACD datasets

UTKFace				CACD		
Age Group	21–40	41–60	60+	21–40	41–60	60+
Confidence Score	84.92 ± 4.70	87.69 ± 3.64	82.08 ± 4.31	95.51 ± 0.99	94.49 ± 1.41	94.49 ± 1.33
Face verification threshold:1e-5 = 73.975 Error rate 0.001%				Face verification threshold:1e-5 = 73.975 Error rate 0.001%		

4.4.3 PSNR and SSIM evaluation for super-resolution

PSNR and SSIM have been mostly used evaluation metrics for super-resolution. However, the PSNR (peak signal to noise ratio) incline to result in extreme smoothness and outcome can differ in large, among almost identical images. Further, the SSIM (structural similarity index) evaluates in terms of contrast, brightness, and structure. It measures the value between 0 and 1. The value 1 means two images are identical and as the value goes lower means greater the difference. The use of image sharpening and then SRGAN training produces better results for
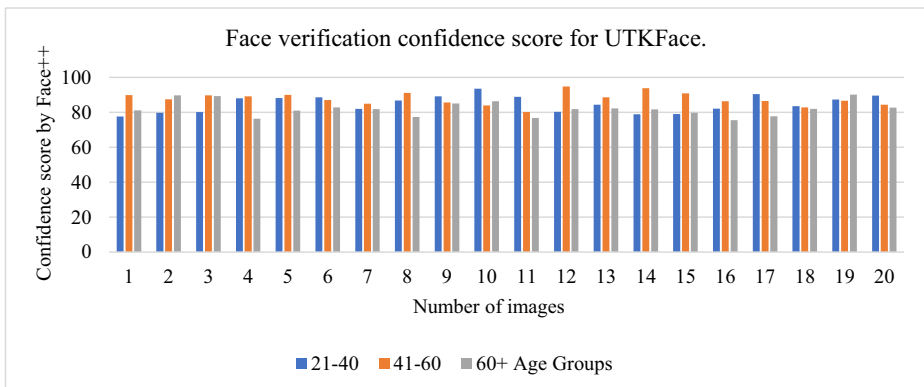
(a)

(b)

Fig. 18 Face verification confidence score (a) UTKFace (b) CACD

Table 5 Quantitative analysis of super-resolution face images

Evaluation metric	PSNR			PSNR		
Dataset	**UTKFace**			**CACD**		
Age group	**21–40**	**41–60**	**60+**	**21–40**	**41–60**	**60+**
The proposed method	30.2084	4.8726	36.0997	24.7096	9.7867	30.0154
Evaluation metric	**SSIM**			**SSIM**		
Dataset	**UTKFace**			**CACD**		
Age group	**21–40**	**41–60**	**60+**	**21–40**	**41–60**	**60+**
The proposed method	0.9354	0.2876	0.9560	0.8280	0.3426	0.9098

super-resolution final output. The quantitative evaluation with PSNR and SSIM for super-resolution face aged images after image sharpening and SRGAN training are shown in Table 5.

4.5 Comparison with existing approaches

The proposed work generates the super-resolution face aged image which helps in identifying the details of aging signs precisely. The proposed work is compared to previous methods [11] [57]. The qualitative evaluation of the proposed work is performed using the FGNET dataset as given in Fig. 19.

4.5.1 User study evaluation

As state-of-the-art papers are shown in Table 1, a user study evaluation was conducted from 10 observers. It was asked to evaluate a pair-wise image comparison to the existing methods [11, 57]. Total 36 paired images of 18 persons were used from the available work and viewers were asked to check the pair-wise images comparison, to evaluate the super-resolution face age-progressed images. Among 360 votes, 60% prefer the proposed work is better, 30% say prior work is better, and 10% indicate that they are even. In the prior work [11, 57] cropped faces were used for face aging. So, lack in showing aging details on faces. The proposed work generates plausible texture changes on faces in old age such as wrinkles, forehead, and front

Input Image						
[63]						
[53]				*.	*.	*. Face image results are not available in the reference paper.
Proposed Work						

Fig. 19 Comparison with existing approaches

hairline dropping and preserves the identity well, and produces super-resolution face aged images as shown in Fig. 12 and Fig. 13. Thus, the proposed work is showing better performance.

5 Conclusion

In this paper, AttentionGAN with super-resolution GAN is used to get super-resolution face aged images. The proposed work produces plausible super-resolution face aged images. The robustness and efficacy of the proposed work are shown with qualitative comparison and quantitative evaluation using age estimation, identity preservation analysis. The generalization ability of the model is shown with three cross-dataset evaluations. The suggested work retains the age-progressed face images to real images with an error rate of 0.001%. However, the future scope is open for face age progression implementation on real-time applications.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Albert AM, Jr KR, Patterson E (2007) A review of the literature on the aging adult skull and face: implications for forensic science research and applications. Elsevier Forensic Science international 172(1):1–9
2. Antipov G, Baccouche M, Dugelay J-L (2017) Face aging with conditional generative adversarial networks. IEEE International Conference on Image Processing (ICIP). https://doi.org/10.1109/ICIP.2017.8296650
3. Arjovsky M, Chintala S, Bottou L (2017) Wasserstein GAN. arXiv: 1701.07875v3, pages 1-29
4. Bando Y, Kuratate T, Nishita T (2002) A simple method for modelling wrinkles on human skin. IEEE 10th Pacific conference on computer graphics and applications proceedings, pages 166–175.
5. Bessinger Z, Jacobs N (2019) A generative model of worldwide facial appearance. IEEE winter conference on applications of computer vision (WACV). https://doi.org/10.1109/WACV.2019.00172
6. Boissieux L, Kiss G, Thalmann NM, Kalra P (2000) Simulation of skin aging and wrinkles with cosmetics insight. Springer, Computer Animation and Simulation, pages 15–27, doi: https://doi.org/10.1007/978-3-7091-6344-3_2
7. Bokade GU, Kanphade RD (2019) Secure multimodal biometric authentication using face, palmprint and ear: a feature level fusion approach. 10th conference on computing, communication, and networking technologies (ICCCNT), doi: https://doi.org/10.1109/ICCCNT45670.2019.8944755
8. Choi Y, Uh Y, Yoo J, Ha J-W (2020) Stargan v2: diverse image synthesis for multiple domains. arXiv: 1912.01865v2(cs)
9. Cui Y, Liu Q, Gao CY, Su Z (2018) FashionGAN: display your fashion design using conditional generative adversarial nets. Computer Graphics Forum 37(7):109–119 https://doi.org/10.1111/cgf.13552
10. Denton E, Chintala S, Szlam A, Fergus R (2015) Deep generative image models using a Laplacian Pyramid of Adversarial Networks. arXiv:1506.05751v1(cs), pages 1–10
11. Elmahmudi A, Ugail H (2020) A framework for facial age progression and regression using exemplar face templates. The visual computer. 37, pages 2023-2038, springer. 37:2023–2038. https://doi.org/10.1007/s00371-020-01960-z
12. Gaussian R, Jain H, Pratap S (2018) Enhancing bank security system using face recognition, Iris scanner and palm vein technology. 2018 3rd international conference on internet of things: smart innovation and usages (IoT-SIU), doi: https://doi.org/10.1109/IoT-SIU.2018.8519850.
13. Gonog L, Zhou Y (2019) A review: generative adversarial networks. 14th IEEE conference on industrial electronics and applications (ICIEA), pages 505-510, doi: https://doi.org/10.1109/ICIEA.2019.8833686.
14. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. arXiv:1406.2661v1, pages 1-9.

15. Gulrajani I, Ahmend F, Arjovsky M, Dumoulin V, Courville A (2017) Improved training of Wasserstein GANs. Adv Neural Inf Proces Syst 30:1–10

16. Huang Y, Chai Y, Liu Y, Shen J (2019) Architecture of next-generation e-commerce platform. Tsinghua Sci Technol 24(1):18–29

17. Hwang U, Jung D, Yoon S (2019) HexaGAN: generative adversarial nets for real-world classification. arXiv:1902.09913v2

18. Iqbal T, Ali H (2018) Generative adversarial network for medical images (MI-GAN). Journal of medical systems, 42:231, arXiv:1810.00551v1, pages 1-19

19. Isola P, Zhu J-Y, Zhou T, Efros, AA (2017) Image-to image translation with conditional adversarial networks. 2017 IEEE conference on computer vision and pattern recognition, doi: https://doi.org/10.1109/CVPR.2017.632.

20. Jia Y, Zhang Y, Weiss RJ, Wang Q et.al (2019) Transfer Learning from speaker verification to multi-speaker text-to-speech synthesis. 32nd Conference on Neural Information Processing Systems, Canada, pages 1–11

21. JSD R, Moctezuma D, Conde C et al (2016) Automated border control e-gates and facial recognition system Elsevier. Computers and Security vol 62:49–72

22. Karras T, Aila T, Laine S, Lehtinen J (2018) Progressive growing of GANS for improved quality, stability and variation. arXiv.1710.10196v3, pages 1-26

23. Kim T, Cha M, Kim H, Lee JK, Kim J (2017) Learning to discover cross-domain relations with generative adversarial networks. arXiv:1703.05192v2(cs)

24. Koudelova J, Hoffmannova E, Dupej J, Velemı'nska J (2019) Simulation of facial growth based on longitudinal data: age progression and age regression between 7 and 17 years of age using 3D surface data. PLoS One 14(2):e0212618. https://doi.org/10.1371/journal.pone.0212618

25. Ledig C, Theis L, Huszar F et al (2017) Photo-realistic single image super-resolution using a generative adversarial network. arXiv:1609.04802v5(cs).

26. Leng L, Teoh ABJ (2015) Alignment-free row-co-occurrence cancelable palmprint fuzzy vault. Elsevier, Pattern recognition 48(7):2290–2303. https://doi.org/10.1016/j.patcog.2015.01.021

27. Leng L, Zhang J (2011) Dual-key binding cancelable palmprint cryptosystem for palmprint protection and information security. Elsevier Journal of network and computer applications 34(6):1979–1989

28. Leng L, Teoh ABJ, Li M, Khan MK (2014) A remote cancellable palmprint authentication protocol based on multi-directional two-dimensional PalmPhasor-fusion. Security and communication networks, 7(11), pages 1860–1871, https://doi.org/10.1002/sec.900

29. Leng L, Li M, Kim C, Bi X (2015) Dual-source discriminator power analysis for multi-instance contactless palmprint recognition. Multimedia tools and applications,76(1), pages 333-354. 76:333–354. https://doi.org/10.1007/s11042-015-3058-7

30. Liu M-Y, Breuel T, Kautz J (2017) Unsupervised image-to-image translation networks. 31st conference on neural information processing systems, pages 1-9

31. Liu L, Zhang H, Ji Y, Wu QM (2019) Toward AI fashion design: an attribute-GAN model for garment match. Neurocomputing 341:156–167. https://doi.org/10.1016/j.neucom.2019.03.011

32. Ma M (2014) Automated age-progression software lets you see how a child will age. https://www.washington.edu/news/2014/04/09/see-what-a-child-will-look-like-using-automated-age-progression-software.

33. Mao X, Li Q, Xie H, Lau R, Wang Z, Smolley S (2017) Least squares generative adversarial networks. IEEE international conference on computer vision (ICCV), pages 2813–2821, doi: https://doi.org/10.1109/ICCV.2017.304

34. Marra F, Gragnaniello D, Cozzolino D, Verdoliva L (2018) Detection of GAN-generated fake images over social networks. IEEE conference on multimedia information processing and retrieval, pages 384-389, doi: https://doi.org/10.1109/MIPR.2018.00084

35. Martin V, Seguier R, Porcheron A, Morizot F (2018) Face aging simulation with a new wrinkle oriented active appearance model. Springer. Multimedia Tools Application 78:6309–6327

36. Megvii Inc. (2013) Face++ research toolkit. http://www.faceplusplus.com/.

37. Or-El R, Sengupta S, Fried O, Shechtman E, Kemelmacher-Shlizerman I (2020) Lifespan age transformation synthesis. arXiv:2003.09764v2, pages 1-27

38. Osman AM, Viriri S (2018) Face verification across age progression: a survey of the state-of-the-art. ICTAS conference on information communications technology and society. https://doi.org/10.1109/ICTAS.2018.8368755

39. Rahmi Z, Ahmad MI, Isa MNM, Khalib ZIA (2019) Matching score level fusion for face and palmprint recognition system on spatial domain. IEEE 9th international conference on system engineering and technology (ICSET), doi: https://doi.org/10.1109/ICSEngT.2019.8906321

40. Ramanathan N, Chellappa R (2006) Face verification across age progression. IEEE Trans Image Process 15(11):3349–3361

41. Ring T (2020) Simplified arrival biometrics land at key US airports. ScienceDirect. Biometric technology today volume 2020, issue 10, page 3, https://doi.org/10.1016/S0969-4765(20)30127-2
42. Ring T (2020) Face ID firms battle Covid-19 as users shun fingerprinting Elsevier, vol 2020, issue 4, pages 1–2, ISSN 0969–4765. Biometric technology today, doi: https://doi.org/10.1016/S0969-4765(20)30042-4.
43. Seiver S (2015) What-would-you-look-like-20-years-from-now. https://www.themarshallproject.org/2015/06/25/what-would-you-look-like-20-years-from-now.
44. Sharma R, Patterh MS (2015) Face recognition using face alignment and PCA techniques: a literature survey. IOSR Journal of Computer Engineering (IOSR-JCE) 17(4):17–30
45. Shu X, Tang J, Lai H, Liu L, Yan S (2015) Personalized age progression with aging dictionary. IEEE international conference on computer vision, pages 3970–3978. doi: https://doi.org/10.1109/ICCV.2015.452
46. Sumathi S, Malini RR (2010) Face recognition system to enhance E health. 2010 international conference on E-health networking digital ecosystem and technology (EDT), pages 195-198
47. Sun H, Tan X, Gan J-W, Liu H et al. (2019) Token-level ensemble distillation for grapheme-to-phoneme conversion. arXiv: 1904.03446v3
48. Suo J, Zhu S-C, Shan S, Chen X (2010) A compositional and dynamic model for face aging. IEEE Trans Pattern Anal Mach Intell 32(3):385–401
49. Suo J, Chen X, Shan S, Gao W, Dai Q (2012) A concatenational graph evolution aging model. IEEE Trans Pattern Anal Mach Intell 34(11):2083–2096
50. Takano N, Alaghband G (2019) SRGAN: Training dataset matters. arXiv:1903.09922v1, pages 1-7.
51. Tang H, Liu H, Xu D, Torr P HS, Sebe N (2020) AttentionGAN: unpaired image-to-image translation using attention-guided generative adversarial networks. arXiv:1911.11897v4. Pages 1-13
52. Thengane VG, Gawande MB, Dudhane AA, Gonde AB (2018) Cycle face aging generative adversarial networks. IEEE 13th international conference on industrial and information systems (ICIIS), doi: https://doi.org/10.1109/ICIINFS.2018.8721435
53. Tiddeman B, Burt M, Perrett D (2001) Prototyping and transforming facial textures for perception research. IEEE Comput Graph Appl 21(5):42–50
54. Tolosana R, Vera-Rodriguez R, Fierrez J, Morales A, Ortega-Garcia J (2020) DeepFakes and beyond: a survey of face manipulation and fake detection. arXiv:2001.00179v3, pages 1-23
55. Uricar M, Krizek P, Hurych D, Sobh I, Yogamani S, Denny P (2020) Yes, we GAN: applying adversarial techniques for autonomous driving. Computer vision and pattern recognition, arXiv:1902.03442v2, pages 1–16.
56. Wang Z, Tang X et al (2018) Face aging with identity-preserved conditional generative adversarial networks. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2018.00828
57. Wang W, Yan Y, Cui Z, Feng J, Yan S, Sebe N (2019) Recurrent face aging with hierarchical autoregressive memory. IEEE Transactions Pattern Analysis and Machine Intelligence 41(3):654–668. https://doi.org/10.1109/TPAMI.2018.2803166
58. Wang X, Wang K, Lian S (2020) A survey on face data augmentation for the training of deep neural networks. Springer. Neural Comput & Applic 32:15503–15531. https://doi.org/10.1007/s00521-020-04748-3
59. Wang R, Juefei-Xu F, Ma L, Xie X, Huang Y, Wang J, Liu Y (2020) FakeSpotter: a simple yet robust baseline for spotting AI-synthesized fake faces. arXiv:1909.06122v3
60. Wang Z, She Q, Ward TE (2020) Generative adversarial networks in computer vision: a survey and taxonomy. arXiv:1906.01529v6, pages 1-41
61. Wilson S, Foster L, Mogor J (2017) How-experts-use-age-progression-technology-to-help-find-missing-kids. https://www.wusa9.com/article/news/local/outreach/bring-them-home/65-430201809.
62. Yang H, Huang D, Wang Y, Wang H, Tang Y (2016) Face aging effect simulation using hidden factor analysis joint sparse representation. IEEE Trans Image Process 25(6):2493–2507. https://doi.org/10.1109/TIP.2016.2547587
63. Yang H, Huang D, Wang Y, Jain AK (2019) Learning continuous face age progression: a pyramid architecture of GANs. IEEE transactions on pattern analysis and machine intelligence, vol 43(2):499–515. https://doi.org/10.1109/TPAMI.2019.2930985
64. Zhan F, Zhu H, Lu S (2018) Spatial fusion GAN for image synthesis. arXiv: 1812.05840v3, pages 1-10
65. Zhang Z, Song Y, Qi H (2017) Age progression/regression by conditional adversarial autoencoder. IEEE Conference on Computer Vision and Pattern Recognition, pages:4352–4360. https://doi.org/10.1109/CVPR.2017.463
66. Zhu J-Y, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. IEEE International Conference on Computer Vision. https://doi.org/10.1109/ICCV.2017.244

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Neha Sharma received B. Tech degree in 2007 and M. TECH degree in 2010 in the field of Electronics and Communication Engineering from Punjab Technical University, Jalandhar. Currently, she is a Ph.D. research scholar at Punjabi University, Patiala. Her research area includes Image Processing, Generative Adversarial Networks.

Dr. Reecha Sharma did her graduation in Electronics and Communication Engineering in 2003 from MMEC, Mullana (Ambala), and Post-Graduation in the field of Electronics and Instrumentation Engineering in 2005 from TIET, Patiala. She had completed her Ph.D. in 2015 from Punjabi University, Patiala. Her area of expertise is Image Processing. Presently, she is working in Punjabi University, Patiala as an Assistant Professor in Electronics and Communication Department.

Dr. Neeru Jindal has received the B.Tech and M. Tech degree in Electronics and Communication Engineering in 2003 and 2007 respectively. She is Ph.D. in Electronics and Communication Engineering from TIET, Patiala. Her current research interests include Image and Video Processing, Machine Learning, Deep Learning, Generative Adversarial Networks. She is the reviewer of many reputed journals. She is working as an Assistant Professor in Electronics and Communication Engineering in TIET, Patiala.

Affiliations

Neha Sharma[1] · **Reecha Sharma**[1] · **Neeru Jindal**[2]

    Reecha Sharma
    reecha@pbi.ac.in

    Neeru Jindal
    neeru.jindal@thapar.edu

[1]   Department of Electronics and Communication Engineering, Punjabi University, Patiala, Punjab 147001, India

[2]   Department of Electronics and Communication Engineering, Thapar Institute of Engineering and Technology, Patiala, Punjab 147001, India