Received February 5, 2022, accepted February 22, 2022, date of publication March 8, 2022, date of current version March 18, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3157617

Recent Generative Adversarial Approach in Face Aging and Dataset Review

HADY PRANOTO[1,2], YAYA HERYADI[1], HARCO LESLIE HENDRIC SPITS WARNARS[1], AND WIDODO BUDIHARTO[2]

[1]Computer Science Department, BINUS Graduate Program-Doctor of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia
[2]Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

Corresponding author: Hady Pranoto (hadypranoto@binus.ac.id)

This work was supported in part by the U.S. Department of Commerce under Grant BS123456.

ABSTRACT Many studies have been conducted in the field of face aging, from approaches that use pure image-processing algorithms, to those that use generative adversarial networks. In this study, we review a classic approach that uses a generative adversarial network. The structure, formulation, learning algorithm, challenges, advantages, and disadvantages of the algorithms contained in each proposed algorithm are discussed systematically. Generative Adversarial Networks are an approach that obtains the status of the art in the field of face aging by adding an aging module, paying special attention to the face part, and using an identity-preserving module to preserve identity. In this paper, we also discuss the database used for facial aging, along with its characteristics. The dataset used in the face aging process must have the following criteria: (1) a sufficiently large age group in the dataset, each age group must have a small range, (2) a balanced distribution of each age group, and (3) has enough number of face images.

INDEX TERMS Face recognition, image generation, image database, face aging dataset, deep generative approach, generative adversarial network.

I. INTRODUCTION

Face-aging has recently attracted the attention of the computer vision community, and a variety of approaches, ranging from pure algorithms in the field of computer graphics to approaches that use deep learning architectures have been proposed. Several successes have been reported, from approaches that use theory in anthropology to approaches that use deep learning.

Generally, approaches to age progression are classified into four categories: (1) modeling, (2) reconstruction, (3) prototyping, and (4) deep learning approaches. The first to third category approaches usually use a simulation of the aging process from facial features by use (a) adopting anthropometric knowledge [1], or (b) representing face geometry and appearance by setting it through conventional parameters. Examples of methods that use this approach are Active Appearance Models (AAMs), 3D Morphable Model (3DMM). Even with many studies that have obtained inspiring results, the representation of this method is still linear and has various limitations when modeling non-linear aging processes.

Face aging across ages becomes very challenging because faces changes over time are not linear. However, the linear method cannot solve this problem. However, a modern approach, the Deep Generative Method (DGM) for face modeling and mapping of the aging process, achieves a state-of-the-art face aging approach. Deep learning algorithms have a better ability to interpret and transfer nonlinear aging features. Several faces aging studies on how to produce superior synthetic image results as in [2]–[7], including the Generative Adversarial Networks method.

Inspired by a study that obtained state-of-the-art results, in this study, we aim to provide a review of the current Generative Adversarial Networks in face aging progression and the dataset used in that method. In this study, each method is discussed in terms of its structure and formulation. In this paper, we will cover, in general, the conventional approach and the Generative Adversarial Networks approach. Discuss

The associate editor coordinating the review of this manuscript and approving it for publication was Long Wang.

how they produced face synthetics and used a public dataset. The paper structure is built sequentially: discuss the dataset used in face aging and their characteristics, and the distributions, approaches used in face aging, and also discusses challenge and opportunity in face aging if we add more information to the dataset.

II. FACE AGING DATASET

Collecting a dataset related to face aging is challenging. Several criteria were satisfied in the collection process. First, each subject/identity in the dataset must have images of different ages and must cover a long age range. However, this is not an absolute requirement, because, in some research, we found some face aging approaches do not need sequence images of the same person at different ages and can still produce the aging pattern. It is important to discuss the dataset in Generative Adversarial Network because the success factor of this model in obtaining a robust model depends on the characteristics of the dataset. Currently, the face aging dataset is limited in terms of age labels and the number of available datasets.

The MORPH-Album 1 dataset [8] contained 1,690 grayscale portable gray maps (PGM) images from 515 individuals. Metadata this dataset metadata has information about subject identifier, date of birth, picture identifier, image date, race, facial hair flag, age differences, glasses flag, image filename, gender. The image age group distribution in this dataset was not balanced and was dominated by ages–18-to 29.

The MORPH-Album 2 dataset [8] consisted of 55,000 unique images from 13,000 subjects aged 16–77 years, with an average age of 22. Has information about subject identifier, race, picture identifier, date of birth, image date, and age difference. Image film number, race, gender, facial hair flag, and glass flag. The age group distribution is more balanced compared to the Morph Album 1 dataset; a balanced distribution on the dataset causes a face aging architecture easy to learn aging patterns from each age group, and the number of face images from each age group must sufficient to get the age group pattern.

FG-Net [9] is a dataset consisting of 1,002 face images from 82 subjects that contain information, such as Identity ID and age. The age distribution is not balanced, is not well separated, has a very small number of images, is dominated by the young age group, and is not easy to obtain an aging process pattern.

The AdienceFace dataset [10], [11] consisted of 26,580 face images from 2,984 subjects. Grouped into 8 age group labels (0-2, 4-6, 8-12, 15-20, 25-32, 38-43, 48-53, 60+), and has a gender and identity label. The age distribution of this dataset is dominated by young labels/groups.

The Cross-Age Celebrity Dataset (CACD) [12] consists of 163,446 facial images from 2000 celebrities from 2004-to 2013, the following information: name, id, year of birth, celebrity ranking, LBP features from 16 facial landmarks. The age groups distribution is slightly balanced but still dominates the age 20 to 60 years.

The IMDB-WIKI [13] dataset contains 523,051 images from 20,284 subjects obtained from the IMDB and Wikipedia. The dataset contains information on the date of birth, taken date, gender, face location, face score, second face score, celeb name, celeb id, age calculated using the date taken, and the date of birth. The distribution of age groups in this dataset is dominated by 20-30 and 30-40 age groups and unbalanced in the young and old age groups.

The AgeDB [14] dataset contained 16,488 images from 568 subjects and was annotated manually to ensure that the age labels were clean. The AgeDB is an in-the-wild dataset, and the average number of images per subject was 29. The dataset contains image information, including ID, subject name, and age. Domination using greyscale images. The age distribution is dominated by the age groups of 30-40, and 40-50, slightly balanced, and the number of images is sufficient to face the aging architecture learning the age pattern.

The UTKFace [4] dataset is a large-scale face dataset with a long age span (ranging from 0 to 116 years), containing more than 20,000 images with information about age, gender, and ethnicity. The UTKFace dataset is in the wild, with variations in pose, expression, occlusion, illumination, and resolution. The age distribution was quite balanced, dominated by the age group 10-19 years.

An unbalanced dataset distribution makes it difficult for the face aging architecture to identify good-aging patterns of each age group. The ability of the architecture to obtain aging patterns from each age group depended on the number of images in each group. An insufficient number of face images does not provide sufficient pattern information from age group images.

The characteristics of the dataset and the age distribution of several existing face aging datasets are summarized in Table 1 and Figure 1. The sample images for each dataset are shown in Figure 25.

III. FACE AGING METHOD

A. CONVENTIONAL APPROACH

1) MODEL-BASED APPROACH

Early research on age progression utilized an appearance model to represent the shape of the face structure and face texture in input face images. The aging process is represented by an aging function that applies the parameter sets of different age groups to the input image. Patterson *et al.* [15] and Lanitis *et al.* [16] used Active Appearance Models (AAMs) to simulate adult craniofacial aging in images using two approaches, (1) estimating age in an input image, and (2) shifting the active appearance model parameters in directional aging axis. Active Appearance Models (AAMs) presented an anthropological perspective on the active appearance model of facial aging and had an effect on face recognition for the first time.

TABLE 1. Properties of different face aging datasets in the wild, dataset collected from unconstrained environment condition.

Database	#Images	#Subject	Label Type	Aligment	Subject Type	Clean Label	Dataset Distribution
MORPH – Album 1[8]	1,690	628	Age	Frontal	Non-famous	Yes	<18(±157), 18-29(±985), 30-39(±415), 40-49(±111), 50+(±22)
MORPH – Album 2[8]	55,134	13,000	Age	Frontal	Non-famous	Yes	<20(±7,469), 20-29(±16.3225), 30-39(±15.357), 40-49(±12,050), 50+(±3.993)
FG-NET[9]	1,002	82	Age	In-the-wild	Non-famous	Yes	<20(±710), 20-29(±144), 30-39(±79), 40-49(±46), 50+(±23)
AdienceFaces[10][11]	26,580	2,284	Age groups	In-the-wild	Non-famous	Yes	0-2(±1,427), 4-6(±2,162), 8-13(±2,294), 15-20(±1,653), 25-32(±4,897), 38-43(±2,350), 48-53(±825), 60+(±869)
CACD[12]	163,446	2,000	Age	In-the-wild	Celebrities	No	0-10(0), 10-19(±7,057), 20-29(±39,069), 30-39(±43,104), 40-49(40,344), 50-59(±30,960), 60+(±2,912)
IMDB-WIKI[13]	523,051	20,284	Age	In-the-wild	Celebrities	No	0-10(±2), 11-20(±6), 21-30(±15), 31-40(±45), 41-50(±18), 51-60(±7), 61-70(±3), 71-80(±2), 81-90(1),91-100(1)
Age-DB[14]	16,488	568	Age	In-the-wild	Celebrities	Yes	0-10(±1), 11-20(±4), 21-30(±15), 31-40(±30), 41-50(±20), 51-60(±13), 61-70(±10), 71-80(±4),81-90(±2),91-100(1)
UTKFace[4]	>20,000	-	Age	In-the-wild	Non-famous	Yes	0-10(±3,300), 10-19(±1,600), 20-29(±7,400), 30-39(±4,500), 40-49(±2,200), 50-59(±2,300), 60+(±2,750)

A different study was proposed by Geng *et al.* [17] which introduced the Aging Pattern Subspace (AGES) approach. Gang *et al.* constructed a subspace representation of the aging pattern as a chronological sequence of face images using Principal Component Analysis (PCA). Finally, facial synthesis results in a certain age group by applying a variance aging effect on facial appearance. We used a face model transformation among all less than one-year-old subjects in the database, instead of using a cardioid strain transformation or global aging function. The main problem encountered in this architecture is the aging transformation, which is the difficulty in constructing a suitable training set and a sequential age progression from a different individual.

Suo *et al.* [18] introduced a dynamic compositional model of facial aging. each faces from each age group into a hierarchical "And-Or" graph model. The "And" node is used to decompose the facial image into several parts to define facial details such as hair, wrinkles, etc., a pattern that is crucial for defining the perception of age. This compositional and dynamic model seeks to build a perception of aging and at the same time maintain the identity of an input facial image by combining several parts of different faces with different aging effects. This method lacks individual facial sequences and requires retraining when additional facial sequences are included. In the study by Suo *et al.* [18], the model was not trained over a long lifespan. The resulting model still leaves ghosting artifacts on the resulting synthetic face which is caused by the difficulty in performing precise alignment between the model and the original image.

In the model-based approach, the face structure and texture changes, such as muscle changes, hair color, and wrinkles are

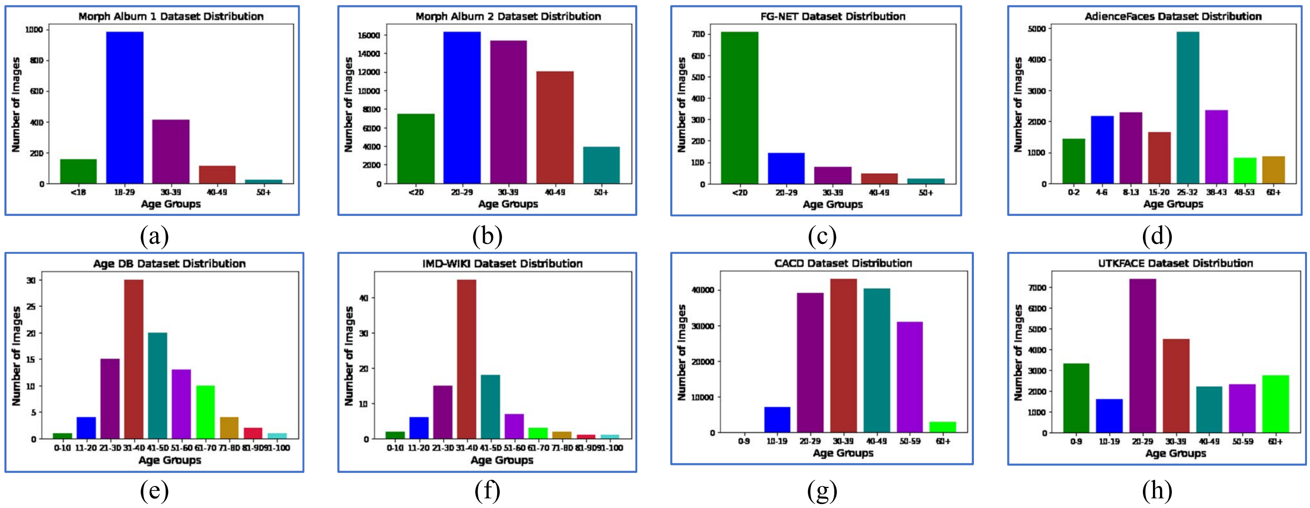


FIGURE 1. Dataset distribution (a) MORPH album 1; (b) MORPH album 2; (c) FG-NET; (d) AdienceFaces; (e) AGE-DB; (f) IMDB-WIKI (g) CACD (h) UTK-Face.

modeled using parameters. The general form of face aging modeling is difficult to determine, which makes it difficult to apply this facial aging model to certain faces, and there is a possibility that the resulting facial model is not suitable for certain faces. Determining this parameter requires a large number of training samples and need high computational. Mismatch and difficulty in performing precise alignment make difficult to produce realistic images of aging faces without losing identifying information.

2) PROTOTYPE APPROACH

The basic idea of the prototype approach [19] is to apply the differences between age groups to the input face image to produce a new image in the desired age group. A prototype of the average face estimation or mean face [20] for each age group is applied to the input face image to generate new faces in the desired age group. To produce a good synthetic image, a precise alignment is required to avoid producing a plausible synthetic image. This method produces a relightable age subspace, a novel technique for subspace-to-subspace alignment that can handle photo "in the wild," with a variant in illumination pose, and expression, and enhances realism for older subject output as a progressed image from an input image. This method used a collection of head and up torso images, which has the effort to get the image. The prototype approach to facial aging is limited to producing aging patterns and losing global understanding of the human face, such as personalized information and possibly facial expressions. A sharper average face was introduced in [21].

3) RECONSTRUCTION-BASE APPROACH

The reconstruction-based approach focuses on determining and combining the aging patterns of each age group. This dictionary was used to convert an input image into a synthetic faces image for the targeted age group.

Coupled Dictionary Learning (CDL) [22], a method that uses a reconstruction-based approach, models personalized aging patterns by preserving the personalized features of each individual by formulating a short-term aging photo. It is difficult to collect long-term dense-facing sequences. A person always has a dense short-term face aging photo but not a long-term aging photo, covering all age groups

Bi-level Dictionary Learning-based Personalized Age Progression (BDLPAP) method [23], automatically renders an aging face personally using short-term face-aging pairs. A person's face can be composed of a personalization layer and an aging layer during the face-aging process. Using the aging invariant pattern that was successfully obtained using Coupled Dictionary Learning (CDL), a dictionary captures the aging characteristics, which learns Personality-aware formulation and short-term coupled learning. individual characteristics such as a mole, birthmark, permanent scar, etc. are represented in face aging sequences $\{x_i^1, x_i^2, \ldots, x_i^G\}$ on their aging dictionaries.

B. DEEP GENERATIVE MODEL FOR FACE AGING APPROACH

1) TEMPORAL RESTRICTED BOLTZMAN MACHINE-BASED MODEL (TRBM)

Temporal Restricted Boltzman Machine-based model utilizing embedding of temporal relationships between sequences of facial images. Duong *et al.* [2] proposed an Age Progression using the log-likelihood objective function and ignored the l_2 reconstruction error during the training process. This model could efficiently capture the nonlinear aging process and automatically produce the sequential development of each age group in greater detail. A combination of the TRBM and RBM produced a model to simulate the age variation model and transform the embedding. Using this approach, linear and nonlinear interacting structures can be used to exploit the facial images. Facial wrinkles increased with geometric constraints during post-processing for more consistent results. However, incorrect results could still be obtained using this method.

2) RECURRENT NEURAL NETWORK-BASED MODEL

Wang *et al.* [3] proposed a Recurrent Neural Network (RNN) and utilized two-layer gated-recurrent-units(GRU) to model aging sequences. A recurrent connection between hidden units can efficiently utilize information from the previous layer and the "memory" is obtained from the previous layer through recurrent connection to create smooth transitions between age groups in the process of synthesizing new faces.

This approach is superior to the classic approach, which performs a fixed reconstruction and always results in a blurry and unclear image. The combination of facial structure and facial aging modeling into a single unit using RBMS creates a nonlinear change that can be interpreted properly and efficiently and produces a wrinkle, texture model for each age group with a more consistent output. However, this method requires a large amount of training data to produce a robust, general model.

Temporal Non-Volume Preserving (TNVP) was introduced by Duong *et al.* [24]. This method uses embedding feature transformation between faces in consecutive stages and compares the CNN structure. This method uses an empirical balance threshold and Restricted Boltzmann Machine, which guarantees that the architecture is intractable model density with the ability to exact inferences between faces in consecutive stages. The TNVP model has advantages in terms of architecture to improve image quality and highly nonlinear feature generation. The objective function of the TNVP can be formulated as:

$$\begin{aligned} z^{t-1} &= \mathcal{F}_1\left(x^{t-1};\theta_1\right) \\ z^t &= \mathcal{H}\left(z^{t-1};x^{t-1};\theta_2,\theta_3\right) = \mathcal{G}\left(z^{t-1};\theta_3\right) + \mathcal{F}_2\left(x^t;\theta_2\right) \end{aligned} \tag{1}$$

where $\mathcal{F}_1$, $\mathcal{F}_2$ represent the bijection function mapping x^{t-1} and x^t to their latent variables z^{t-1}, z^t, respectively. The aging transformation between the latent variables was performed by embedding $\mathcal{G}$. The framework of Temporal Non-Volume Preserving can be shown in Figure 2.

3) GENERATIVE ADVERSARIAL NETWORKS APPROACH

Goodfellow *et al.* [25] proposed a new architecture called Generative Adversarial Network (GAN). Borrow the idea of a pair game between the generator and discriminator. The generator attempts to generate sample data that can deceive the discriminator to consider the data as original data, and the discriminator learns to discriminate between real and fake data produced by the generator. The process continued until the discriminator could distinguish the generated data were real or fake.

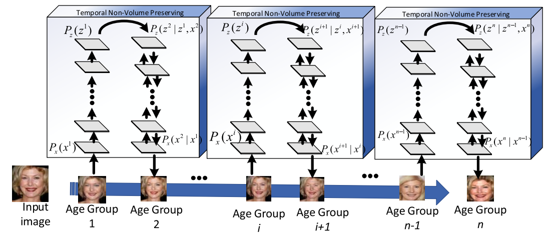


FIGURE 2. Temporal non-volume preserving framework.

A GAN can produce an object image from a given noise, depending on the given training data. To generate an artificial face image, a GAN must be trained using the face images. A GAN can produce certain characteristics [26] of the resulting object by assigning conditions to architecture [27] and applying a residual block, to change the aging pattern [28]. A GAN can also create a new synthetic image from the input image, not from scratch or noise, to accelerate the generation process [6].

From step by step aging process, GAN is transformed into a direct age process [4], [27], [29]. The process directly adds an aging pattern such as sideburn, wrinkles, eye bags, and gray hair, and structure change such as the structure of the head, enlarged eyes, shrunken chin, are directly implemented when the aging or rejuvenation process happens.

GANs can be categorized into three types: translation-based, sequence-based, and condition-based. The translation-based method is based on transferring style from one set domain of an image into another set domain of the image.

CycleGAN [30] is a translation-based method that captures style characteristics from a set of images to be implemented in another set of images. CycleGAN does not require paired domain sets. This advantage can be utilized in face aging to translate images from one age group into another. However, CycleGAN can only translate two age groups into pairs, which is a limitation in its architecture. In figure 2, the author illustrated how CycleGAN worked to generate an image $\hat{y}$ in domain Y from image x, and the cycle consistency process tries to restore image $\hat{y}$ to image x which is in domain X. Forward process and back-forward process, and cycle consistency loss from this process enable architecture to get style map from both domains. A clear understanding of the translation-based framework is presented in Figure 2.

The sequence-based model was implemented in a step-by-step process, and each model was trained independently, to produce a sequential translation between two neighboring age groups. In this model, the translation process is performed sequentially and each resulting model is combined into a complete network. The output from the ith current network is employed as the input in the next network $i+1$. A sequence-based method was used to produce faces for each age group and stage. The challenge in this method is the production

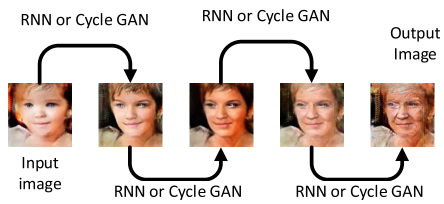


FIGURE 3. Translation-based approach.

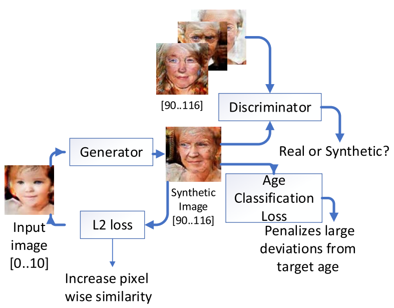


FIGURE 4. Sequence-based approach.

of sequential and complete images of an individual. The framework is illustrated in Fig. 3.

The conditional-based method [31] uses conditional architecture to produce an artificial face image in a certain age group. The conditional is in the form of a label that is converted into a one-hot encoder tensor. This one-hot code tensor was used to force the network to generate a synthetic face image at the desired age. The location of this one-hot encoder in a varied architecture, some are placed on the generator and discriminator [6], or there is also an architecture that only puts this one-hot encoder on the generator section [32].

The concept of the conditional-based method is to drive the generator by providing extra information on the age target that you want to produce, which has a significant advantage and high efficiency compared to translation-based and sequence-based approaches. The conditional-based method framework is illustrated in Figure 5.

The one-step approach to facial aging is still a top priority, and there are still many challenges to producing synthetic faces for certain age groups using only one training course [33], and the current face aging method used "one-shot" and achieves state-of-the-art [34]–[36]. Much research on current generative face aging categories is presented in Appendix C.

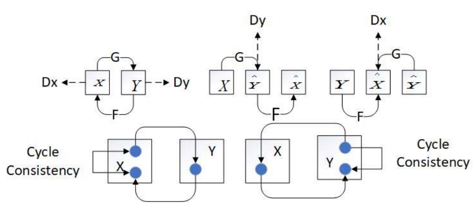


FIGURE 5. Conditional-based approach.

C. GENERATIVE ADVERSARIAL NETWORKS APPROACH

1) TRANSLATION-BASED APPROACH

a: GENERATIVE ADVERSARIAL STYLE TRANSFER NETWORKS FOR FACE AGING

CycleGAN [30] implements a style transfer architecture [37], utilizing cyclic consistency between the input image and the generated image, which can maintain the identity of the generated image still the same as the input face. The aging effect is produced by utilizing the mapping functions of this style transfer architecture to minimize the values of the age loss and cycle consistency loss in Equation 2:

$$\begin{aligned}\mathcal{L}_{age}\left(G_{+}, G_{-}\right) = {} & \mathbb{E}_{k\ P(k)}\mathbb{E}_{x_o\ P_{data}(x_o)}\left[\left|DEX\left(G_{+}\left(x_o\right), k\right)\right.\right.\\ & - DEX\left(x_o\right) - k|\\ & + \left|DEX\left(G_{\mp}\left(x_o\right), k\right) - DEX\left(x_o\right) + k\right|]\\ \mathcal{L}_{cyc}\left(G_{+}, G_{-}\right) = {} & \mathbb{E}_{k\ P(k)}\mathbb{E}_{x_o\ P_{data}(x_o)}\left[\left|\left|G_{+}\left(\hat{x}_{-k}\right) - x_o\right|\right|_1\right.\\ & \left. + \left|\left|D\left(G_{-}\left(\hat{x}_k\right), k\right) - x_o\right|\right|_1\right] \qquad (2)\end{aligned}$$

Training stability is produced using LSGAN loss which is formulated in Equation 3.

$$\begin{aligned}\mathcal{L}_{LSGAN}\left(G, D\right) = {} & \mathbb{E}_{x_o\ P_{data}(x_o)}\left[\left(D\left(x_o\right) - 1\right)^2\right]\\ & + \mathbb{E}_{x_o\ P_{data}(x_o)}\left[D(F\left(x_o\right))^2\right] \qquad (3)\end{aligned}$$

To create the aging effect and preserve the identity, the final objective function was set, as shown in Equation 4:

$$\begin{aligned}\mathcal{L}\left(G_{+}, G_{-}, D\right) = {} & \mathcal{L}_{LSGAN}\left(G_{+}, G_{-}, D\right) + \mathcal{L}_{age}\left(G_{+}, G_{-}\right)\\ & + \mathcal{L}_{cyc}\left(G_{+}, G_{-}\right) + \mathcal{L}_{ac}\left(G_{+}, G_{-}\right) \qquad (4)\end{aligned}$$

The style transfer method contained in cycleGAN utilizes pairwise training between age groups to create artificial face images at the desired age. The CycleGAN framework is illustrated in Figure 2.

2) TRIPLE-GAN: PROGRESSIVE FACE AGING WITH TRIPLE

Triple-translation loss was utilized in Triple Generative Adversarial Network(Triple-GAN) by Fang *et al.* [36] to model the strong interrelationship between the aging patterns in different age groups. The ability to learn the mapping between labels offered by multiple training pairs uses triple translation loss, as follows:

$$\mathrm{T}\mathcal{L}_{triple} = \left|\left|G\left(x, L_t\right) - G\left(G\left(x, L_f\right), L_t\right)\right|\right|_2^2 \qquad (5)$$

Generate three kinds of face images $G\left(x, L_t\right)$, $G\left(x, L_f\right)$ and $G\left(G\left(x, L_f\right), L_t\right)$ and all synthesized faces used in identity preservation and age classification

The final objective function can be formulated as:

$$\begin{aligned}&\mathcal{L}_G = \alpha\mathcal{L}_g + \beta\mathcal{L}_{identity} + \gamma\mathcal{L}_{age} + \lambda\mathcal{L}_{triple}\\ &\mathcal{L}_D = \alpha\mathcal{L}_d \qquad (6)\end{aligned}$$

where α, β, γ and λ are values that control the four objectives. The triple transformation loss reduces the distance between synthetic faces of the same age target by producing images on different paths. The triple-GAN framework is shown in Fig. 6.

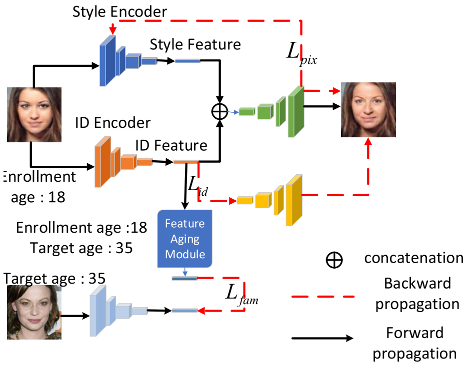


FIGURE 6. **Child age progression framework.**

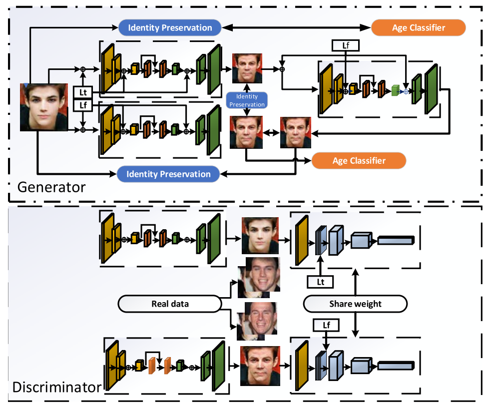


FIGURE 7. **TRIPLE GAN framework.**

3) CHILD FACE AGE-PROGRESSION VIA DEEP FEATURE AGING

Deb *et al.* [38] proposed a feature aging module that can simulate the age progress deep face feature output using a face matcher to guide age progression in image spaces. Synthesized aged faces enhanced longitudinal face recognition without requiring explicit training. This model can increase close-set face recognition over a 10-year time-lapse and enhance the ability to identify young children who are possible victims of child trafficking or abduction. Instead of separating an age-related component from an identity feature, they want to automatically learn a projection within latent space.

Let us Assume $x \in X$ and $y \in Y$, X and Y are two face domains when images are acquired at ages e and t, where e is source age and t is age target. Domains X and Y exhibited differences in aging, noise, quality, and pose. This architecture simplifies the F modeling transformation in a deep feature using the $\mathcal{F}'$ operator, and is formulated as:

$$\hat{y} = \mathcal{F}'(\psi(x), e, t) = W \times (\psi(x) \oplus e \oplus t) + b \tag{7}$$

where function $\psi(x)$ is a function that encodes features in the latent space. Function $\mathcal{F}'$ learns the feature space projection and generates an image x in X with Y age features from source age e to age target t. The representation lies in d-dimensional Euclidean space, where $\mathcal{Z}$ is highly linear. The output of $\mathcal{F}'$ is a linear shift in deep space, $W \in \mathbb{R}^{dxd}$ and $b \in \mathbb{R}^{d}$, learned the parameter of $\mathcal{F}'$ and $\oplus$ is concatenation in the layer. The scale parameter permits the feature to be scaled directly from the registration source age and target age because the feature does not change extremely during the aging process, such as wrinkle, or color of the eye. This architecture has the advantage of projecting the face of aging in young people or children and can be used to find children who are victims of human trafficking. The framework is illustrated in Figure 7.

4) SEQUENCE-BASED APPROACH

a: FACEFEAT GAN, TWO STAGES A TWO-STAGE APPROACH FOR IDENTITY-PRESERVING FACE SYNTHESIS

FaceFeat GAN [39] solve problems in terms of preserving identity with two stages of synthesis, namely feature generation and feature to image rendering. The first stage works in the feature domain to produce a synthesis of various facial features, and the second stage works in the image domain to render realistic photos with high diversity and to maintain identity information.

To reconstruct the input image and produce a more accurate pixel-wise image, the extractor $\{E_i\}_i^k$ implements the real features $\{f_i^r\}_{i=1}^k$ of the input image x^r face and recognition face x^r, which extracts the identity feature f_{id}. The identity feature was used as the input to the G^I generator to make x^{rec} generate an x^r image and become the D^I input discriminator as a synthetic image. The generator G^I attempts to learn to map from the feature space to the image space under identity-preserving constraints. The real feature f_i^r extracted by E_i is used by the discriminator D_i^f to force the G_i^f generator to produce realistic features.

The generator G^I and discriminator D^I are trained using the following function:

$$\min_{\ominus_{G^I}} \mathcal{L}_{G^I} = \emptyset^I\left(x^{rec}\right) + \emptyset_{rec}\left(x^r, x^{rec}\right) + \lambda_1 \emptyset_{id}\left(x^{rec}\right) + \lambda_2 \emptyset_{id}\left(x^s\right) \tag{8}$$

$$\min_{\ominus_{D^I}} \mathcal{L}_{D^I} = \emptyset^I\left(x^r\right) - \lambda_3 \emptyset^I\left(x^s\right) - \lambda_4 \emptyset^I\left(x^{rec}\right) \tag{9}$$

where $\emptyset_{rec}(x^r, x^{rec}) = ||x^r - x^{rec}||_1$, l_1 reconstruction loss, and $\emptyset_{id}(.)$ is a loss function used to measure identity-preserving quality. $\emptyset_i^f$ is an energy function to determine facial features face is real or fake. $\emptyset^I(.)$ to determine whether the generated face is real or fake. Coefficient $\lambda_i^f, \lambda_1, \lambda_2, \lambda_3$ and λ_4 is the value of strength a different term.

Face Feat utilizes Identity and 3DMM features. The identity feature f_{id} from the input image x^r uses the face recognition module as a classification task with a cross-entropy loss. The 3DMM feature is a 3D morphable model used to model 2D images in 3D with a basic set of shapes A_{id}, a set of basic expressions A_{exp} uses a two-stage generator, where the output of the first generator is the input for the second generator. By applying these two generator levels, the competition between the two generators to produce synthetic images yields diverse patterns in the output image. The Face Feat framework is illustrated in Figure 8.

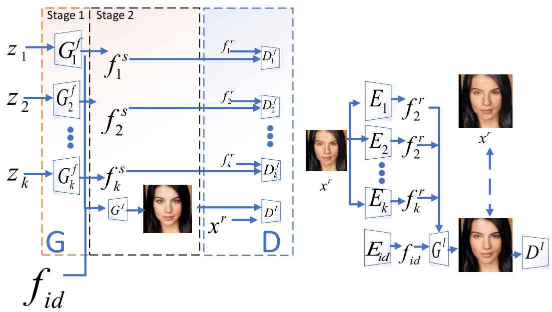


FIGURE 8. Face feat GAN framework.

5) SUBJECT-DEPENDENT DEEP AGING PATH (SDAP) MODEL

An aging controller was used in the TNVP structure in [7] based on the hypothesis that each individual has facial development. Rather than simply embedding aging transformations into pairs that are linked between successive age groups, the Subject Dependent Aging Policy(SDAP) structure studies age transformations within the entire facial sequence to produce better synthetic ages. The SDAP network is an architecture that ensures the availability of an appropriate planned aging path to produce a face-aging controller related to the subject features. It is important to note that SDAP is a pioneer in the IRL framework concerning age progression.

SDAP uses $\varsigma_i = \left\{x_i^1, a_i^1, \ldots, x_i^T\right\}$ as the age sequence of the $i - th$ subject, where $\left\{x_i^1, \ldots, x_i^T\right\}$ are faces sequences representative of the face development of the $i - th$ subject and as a control variable for how much the aging effect will be added to an image x_i^j to became x_i^{j+1}. The probability from ς_i can be formulated using the energy function $Er\left(\varsigma_i\right)$:

$$P\left(\varsigma_i\right) = \frac{1}{Z}\exp\left(-Er\left(\varsigma_i\right)\right) \tag{10}$$

where Z is the partition function and is similar to the joint distribution between the variables of the RBM. The goal is to predict a_i^j for each x_i^j synthesizing image. The SDAP objective function can be formulated as:

$$\Gamma^* = \arg\max \mathcal{L}(\varsigma_i \Gamma) = \frac{1}{M}\log\prod_{\varsigma_i \in \varsigma} P\left(\varsigma_i\right) \tag{11}$$

SDAP optimizes the trackable look-like hoot objective function with a convolutional neural network based on a deep neural network, providing appropriate facial aging development for individual subjects by optimizing the reward-aging process. This method allows multiple-age images to be used as inputs. By considering all information from subjects of various ages, seeks an optimal aging path for a given subject to produce an efficient face aging process by utilizing the power of the generative probabilistic model under the IRL approach in an advanced neural network. For a clear understanding of the Subject-dependent Deep aging path (SDAP) model, the framework is illustrated in Figure 9.

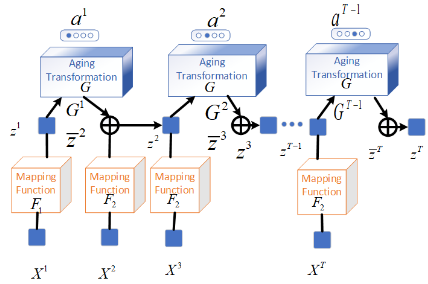


FIGURE 9. Subject-dependent deep aging path (SDAP) framework.

6) CONDITIONAL-BASED APPROACH

a: CONDITIONAL ADVERSARIAL AUTOENCODER(CAAE)

A conditional adversarial Autoencoder(CAAE) [4], which adopts the conditional GAN approach, adds an age label as a conditional encode to enable GAN to produce a synthetic face at a certain age from an input image. CAAE changes were applied to the objective functions in the original GAN to:

$$\begin{aligned}\min_{E,G}\max_{D_x D_{img}} &\lambda\mathcal{L}\left(x, G\left(E(x), l\right)\right)\\ &+\gamma TV\left(G\left(E(x), l\right)\right) + \mathbb{E}_{z*\sim p(z)}\left[\log D_z\left(z^*\right)\right]\\ &+\mathbb{E}_{x\sim Pdata(x)}[\log\left(1 - D_z\left(E(x)\right)\right)]\\ &+\mathbb{E}_{x,l\sim Pdata(x,l)}\left[\log D_{img}\left(x, l\right)\right]\\ &+\mathbb{E}_{x,l\sim Pdata(x,l)}\left[\log\left(1 - D_{img}\left(G\left(E(x), l\right)\right)\right)\right]\end{aligned} \tag{12}$$

where l is the vector representing the age level, z is the latent feature vector, and E is the decoder function, for example, $E(x) = z$. $\mathcal{L}$ (., .) and TV a (.) the ℓ_2 norm and total variation function which is effective for removing ghosting artifacts. The coefficients λ and γ are intended for a balance between smoothness and high resolution. The latent vector was used to personalize face features and age conditions to control progression by studying the learning manifold. Make the age progression/regression more flexible and easy to manipulate.

Framework of. Conditional Adversarial Autoencoder (CAAE) is shown in Figure 10.

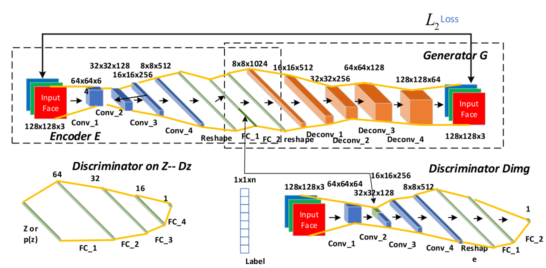


FIGURE 10. Conditional adversarial autoencoder(CAAE).

b: AGE CONDITIONAL GENERATIVE ADVERSARIAL NEURAL NETWORK MODEL

Conditional GAN(cGAN) [26] produces one-to-one image translation by implementing certain characteristics by embedding a condition. Antipov, Baccouche, and Dugelay [27] proposed an age Conditional Generative Adversarial Network (acGAN) by implementing cGAN, which is capable of generating a synthetic face image in the required age category. This method reconstructs an input into a new face for a certain age group and preserves the identity of the face image using latent-vector optimization. Euclidean distance between input image x, and reconstructed image $\bar{x}$ by minimized Euclidean distance embedding from $FR(x)$ and $FR(\bar{x})$ used to preserve identity. AcGAN creates a new face with high quality, and the resulting image is optimized using a latent vector along with the adversarial process of acGAN.

With input face image x with age y_0, an optimal latent vector z^* find to allow reconstructed face $\bar{x} = G(z^*, y_0)$ generated as close as possible to the initial one, with a given target age y_{target}, a new synthetic face image generated by $x_{target} = G\left(z^*, y_{target}\right)$ with switching the age at the input generator. The objective function in the acGAN is as follows:

$$\begin{aligned}\min_{\theta_G}\max_{\theta_D} v\left(\theta_G, \theta_D\right) = &\ \mathbb{E}_{x,y\sim p_{data}}\left[\log D\left(x, y\right)\right]\\ &+\mathbb{E}_{z\sim p_z(z),\tilde{y}\sim p_y}\\ &\times\left[\log\left(1 - D(G\left(z,\tilde{y}\right)\tilde{y})\right)\right]\end{aligned} \tag{13}$$

The placement of conditionals in this architecture will ensure that the architecture can produce changes in the aging pattern on the resulting synthetic facial and make the modeling process more focused and convergent. This framework is illustrated in Figure 11.

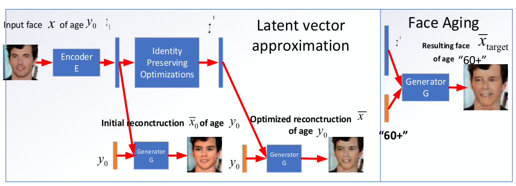


FIGURE 11. Age conditional generative adversarial neural network model.

c: CONTEXTUAL GENERATIVE ADVERSARIAL NETS (C-GANs)

Contextual Generative Adversarial Nets(C-GANs) [28], consist of a conditional transformation network and two discriminative networks. This conditional imitates the aging procedure with several specially designed residual blocks. The pattern discrimination in this architecture is the best part to distinguish real transition patterns from fake ones, guiding the synthetic face to fit the real conditional distribution and extra regularization for the conditional transformation network, ensuring that the image pairs fit the real transition pattern distribution.

The transition pattern discriminative network $D_t(x_y, x_{y+1}, y, y+1)$ has a task of transition patterns between x_y with age y and the image x_{y+1} in the next age group $y+1$. D_t, the task is to distinguish the real joint distribution x_y, x_{y+1}, y $P_{data}\left(x_y, x_{y+1}, y\right)$ from the fake one and force the generator to obey the real transition pattern distribution when generating fake pairs $\left\{x_y, G\left(x_y, y+1\right)\right\}$. By considering the loss in the conditional transformation in generator G, age discriminative D_a, and transition pattern networkD_t, the objective of the function is formulated as:

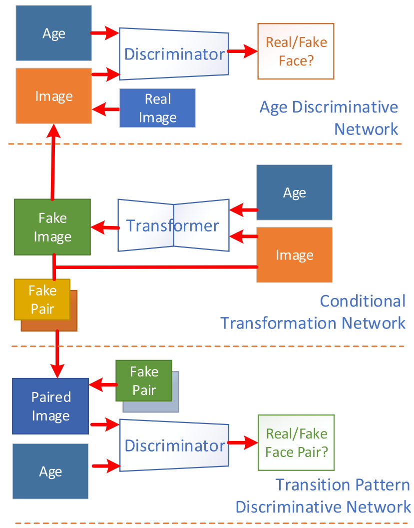


FIGURE 12. Contextual generative adversarial nets(C-GANs).

$$\begin{aligned}&\min_{G}\max_{D_a}\max_{D_t} E\left(\theta_G, \theta_{D_a}, \theta_{D_t}\right)\\ &= E_a + E_t + \lambda TV = E_{x_y,y\sim P_{data}(x_y,y)}\left[\log D_a\left(x_y, y\right)\right]\\ &\quad + E_{x\sim P_x\tilde{y}\sim P_y}\left[\log\left(1 - D_a\left(G\left(x,\tilde{y}\right)\tilde{y}\right)\right)\right]\\ &\quad + E_{x_y,x_{y+1},y\sim P_{data}(x_y,x_{y+1},y)}\left[\log D_t\left(x_y, x_{y+1}, y\right)\right]\\ &\quad + \frac{1}{2}E_{x_y,y\sim P_{data}(x_y,y)}\left[\log\left(1 - D_t\left(x_y, G\left(x_y, y+1\right), y\right)\right)\right]\\ &\quad + \frac{1}{2}E_{x_y,y\sim P_{data}(x_y,y)}\left[\log\left(1 - D_t\left(G\left(x_y, y-1\right), x_y, y-1\right)\right)\right]\\ &\quad + \lambda\left(TV\left(G\left(x_y, y-1\right)\right) + TV\left(G\left(x_y, y+1\right)\right) + TV\left(G\left(x,\tilde{y}\right)\right)\right)\end{aligned} \tag{14}$$

d: DUAL CONDITIONAL GAN FOR FACE AGING AND REJUVENATION

Song *et al.* [40] proposed a novel dual conditional GAN mechanism that enables face aging and rejuvenation by training multiple sets without identities of different ages. This framework contains two conditional GANs: the primary(primal) GAN transforms a face into other ages based on age conditions and the dual GAN or second network learns how to invert a task. Construction error identity can be preserved using a loss function. The generator learns transition patterns, such as the shape and texture, between groups to create a realistic face photo. With multiple training images $F_1, F_2, \ldots, F_N$ of different ages, the network learns the face aging or rejuvenation model G_t and face reconstruction model G_s with a facial image X_i age of C_s and target age condition C_t. This model can predict the face $G_t(X_i, C_t)$ of a person of different ages, and the identity is preserved by $X_s' = G_s(G_t(X_i, C_t), C_s)$.

This framework uses adversarial loss (L_{GAN}) and generation loss (L_{gene}) to match the data distribution and similarity between synthetic and original images, with this mechanism framework, can produce a specific synthetic image at a certain age. Reconstruction loss (L_{recons}) was used to evaluate the consistency between synthetic and original images, and the identity was preserved. L_{GAN}, L_{gene} and L_{recons} were formulated as follows:

$$\begin{aligned} L_{GAN} &= E_{x,\ c\ Pdata}\left[\log(1 - D_t(G_t\left(X_i, C_t\right), C_t))\right] \\ &\quad + E_{x\ Pdata}\left[\log(D_t(X_t, C_t))\right] \\ &\quad + E_{x,c\ Pdata}\left[\log(1 - D_s(G_s\left(X_j, C_s\right), C_s))\right] \\ &\quad + E_{x\ Pdata}\left[\log(D_s(X_s, C_s))\right] \\ L_{recons} &= \left\|G_s\left(G_t\left(X_i, C_t\right), C_s\right) - X_i\right\| \\ &\quad + \left\|G_t\left(G_s\left(X_j, C_s\right), C_t\right) - X_j\right\| \\ L_{gene} &= \left\|G_t\left(X_i, C_t\right) - X_i\right\| + \left\|G_s\left(X_j, C_s\right) - X_j\right\| \end{aligned} \tag{15}$$

And final objective:

$$\begin{aligned} L\left(G_s, G_t, D_s, D_t\right) &= L_{GAN}\left(G_t, D_t, G_s, D_s\right) \\ &\quad + \alpha L_{recons}\left(G_t, G_s\right) + \beta L_{gene}\left(G_t, G_s\right) \end{aligned} \tag{16}$$

where α and β are hyperparameters to balance the objective function.

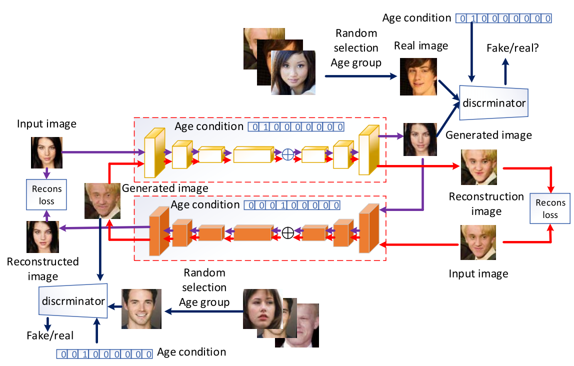


FIGURE 13. **Dual conditional GAN for face aging and rejuvenation.**

For a clear understanding of the dual conditional GAN framework, we illustrate it in Figure 13.

e: IDENTITY-PRESERVED CONDITIONAL GENERATIVE ADVERSARIAL NETWORK (IPCGANs)

IPCGANs [6] has three modules consisting of a conditional GAN (cGAN) module to generate synthetic faces according to the expected age target C_t and produce ($\ddot{x}$) photos that look real as a result of syntheses that successfully model face aging at age stages in a short period, and the two preserved-identity modules which ensure $\ddot{x}$ has the same identity as x, and a classifier module [41] that pushes an image $\ddot{x}$ to the desired target age. The IPCGAN conditional module adopts the derived cGAN architecture proposed by Isola *et al.* [26].

Using a generator, a synthetic image was generated from the input image with the age condition C_t. Discriminator D ensured that the generated images are real or fake. Generator G increases the probability that generated image is mistaken for the original image $D\left(x|C_t\right)$ by discriminator D. Discriminator tasks to align C_t label input to the generated image. To perform this task successfully, the objective function is formulated as:

$$\begin{aligned} \min_G \max_D V\left(D, G\right) &= \mathbb{E}_{x\ P_x(x)}[\log D(x|C_t)] \\ &\quad + \mathbb{E}_{y\ P_y(y)}[1 - \log D(G\left(y|C_t\right))] \end{aligned} \tag{17}$$

IPCGAN uses Least Square Generative Adversarial Network (LSGAN) [42] in discriminator force generated images that look real and difficult to distinguish. The LSGAN conditional can be formulated as follows:

$$\begin{aligned} L_D = {} & \frac{1}{2}\mathbb{E}_{x\ P_x(x)}\left[\left(D\left(x|C_t\right) - 1\right)^2\right] \\ & + \frac{1}{2}\mathbb{E}_{y\ P_y(y)}\left[D\left(G\left(y|C_t\right)\right)^2\right] \end{aligned} \tag{18}$$

at generator loss (L_G):

$$L_G = \frac{1}{2}\mathbb{E}_{yP_y(y)}\left[\left(D(G\left(y\right)|C_t)\right) - 1\right)^2\right] \tag{19}$$

To preserve identity IPCGAN uses perceptual loss instead of adversarial loss. The adversarial loss generates sample data following desired distribution, the resulting image can be any person within the age target. Perceptual loss identity Loss ($L_{identity}$) is formulated as:

$$L_{identity} = \sum_{x \in P_x(x)} ||h(x) - h\left(G\left(x|\ C_t\right)\right)||^2 \tag{20}$$

where $h(.)$ is related to a feature extracted by a specific layer in a pre-trained neural network.

This function loss does not use Mean Square Error (MSE) to calculate losses between image x and generated age face $G\left(x|\ C_t\right)$ in pixel space because the generated face has a change in hair color, side-burn, wrinkle, gray hair, etc., which causes a large difference between the input images x and the generated face. The MSE loss forces generated face $G\left(x|\ C_t\right)$ to be identical to image x. IPCGAN uses proper layer $h(.)$ to preserve identity, an experiment on style transfer showed that a lower feature layer is good for preserving face content (identity), and a higher layer for preserving styles such as face texture and wrinkles; identity information certainly should not change.

In the IPCGAN, age classification is used to generate an aging face for the desired age target. If the resulting generated face is in the correct category, the architecture provides a small penalty and vice versa. The age classification loss was used for this and formulated as

$$L_{age} = \sum_{x\ \in\ P_x(x)} \ell\left(G\left(x|C_t\right), C_t\right) \tag{21}$$

where $\ell(-)$ is the Softmax loss. In backpropagation, the age classification loss $\left(L_{age}\right)$ forces the parameter to change and ensures the generated face is in the right age or age group.

To produce a new face at the same age and identity using the conditional GAN (cGAN), the final objective function is formulated as follows:

$$G_{loss} = \lambda_1 L_G + \lambda_2 L_{identity} + \lambda_3 L_{age} \tag{22}$$

where λ_1 is the desired control as long as the input image is old. In addition, λ_2 and λ_3 have control over the extent to which we want to store identity information, and the resulting image falls to the appropriate age group.

The advantage of this method is that the architecture is trained in one go to be able to produce a model that can synthesize new faces in many groups in one model. The IPCGAN framework is illustrated in Figure 14:

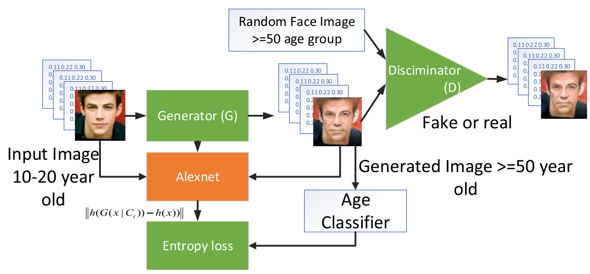


FIGURE 14. **IPCGAN framework.**

f: AGE PROGRESSION AND REGRESSION WITH SPATIAL ATTENTION MODULES

Spatial attention mechanisms were exploited by Li *et al.* [43], [44]. to restrict image modification to areas closely related to age changes, giving images high visual fidelity when synthesized in wild cases. This model uses adversarial loss as follows:

$$\begin{aligned}\mathcal{L}_{GAN} = {} & \mathbb{E}_{I_y}\left[\left(D_p^I\left(G_p\left(I_y,\alpha_o\right)\right)-1\right)^2\right] \\ & +\mathbb{E}_{I_o}\left[\left(D_p^I\left(I_o\right)-1\right)^2\right]+\mathbb{E}_{I_y}\left[D_p^I\left(G_p\left(I_y,\alpha_o\right)\right)^2\right] \\ & +\mathbb{E}_{I_o}\left[\left(D_r^I\left(G_r\left(I_o,\alpha_y\right)\right)-1\right)^2\right] \\ & +\mathbb{E}_{I_o}\left[\left(D_r^I\left(I_y\right)-1\right)^2\right]+\mathbb{E}_{I_y}\left[D_r^I\left(G_r\left(I_o,\alpha_y\right)\right)^2\right]\end{aligned} \tag{23}$$

To penalize the difference between the input images uses reconstruction loss is formulated as:

$$\begin{aligned}\mathcal{L}_{recons} = {} & \mathbb{E}_{I_y}\left[\left\|G_r\left(G_p\left(I_y,\alpha_o\right),\alpha_y\right)-I_y\right\|_2\right] \\ & +\mathbb{E}_{I_o}\left[\left\|G_p\left(G_r\left(I_o,\alpha_y\right),\alpha_o\right)-I_{oy}\right\|_2\right]\end{aligned} \tag{24}$$

Activation loss uses the total activation of the attention mask, formulated as:

$$\mathcal{L}_{actv}=\mathbb{E}_{I_y}\left[\left\||G_p^A\left(I_y,\alpha_o\right)|\right\|_2\right]+\mathbb{E}_{I_o}\left[\left\||G_r^A\left(I_o,\alpha_y\right)|\right\|_2\right] \tag{25}$$

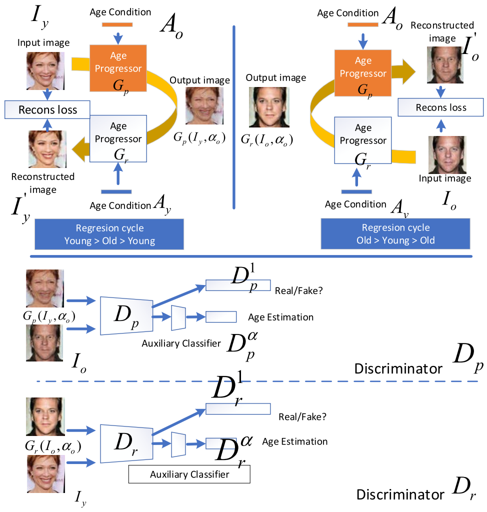


FIGURE 15. **Age progression and regression with spatial attention modules.**

To force the generator to reduce the error between the estimated age and target age used Age Regression Loss formulated as:

$$\begin{aligned}\mathcal{L}_{reg} = {} & \mathbb{E}_{I_y}\left[\left\|D_p^\alpha\left(G_p\left(I_y,\alpha_o\right)\right)-\alpha_o\right\|_2\right] \\ & +\mathbb{E}_{I_y}\left[\left\|D_p^\alpha\left(I_y\right)-\alpha_y\right\|_2\right] \\ & +\mathbb{E}_{I_o}\left[\left\|D_r^\alpha\left(G_r\left(I_o,\alpha_y\right)\right)-\alpha_y\right\|_2\right] \\ & +\mathbb{E}_{I_o}\left[\left\|D_r^\alpha\left(I_o\right)-\alpha_o\right\|_2\right]\end{aligned} \tag{26}$$

By optimizing this equation, the auxiliary regression network D^α gains age estimation ability and generator G encourages the production of a fake face at the desired age target.

The final loss function in this model is a linear combination of all defined losses. Formulated as follow:

$$\mathcal{L} = \mathcal{L}_{GAN} + \lambda_{recon}\mathcal{L}_{recon} + \lambda_{actv}\mathcal{L}_{actv} + \lambda_{reg}\mathcal{L}_{reg} \tag{27}$$

where λ_{recon}, λ_{actv}, and λ_{reg} are coefficients to balance each loss.

By adding a spatial attention mechanism to the architecture, the learning process in training focuses only on the focus of attention, strengthening the attention component has a significant impact on the aging patterns formed. The effects of aging and rejuvenation were more pronounced in the areas of concern.

Age Progression and Regression with Spatial Attention Modules framework illustrated as:

g: S2GAN: SHARE AGING FACTOR ACROSS AGE AND SHARE AGING TRENDS AMONG INDIVIDUAL

He *et al.* [45] proposed continuous face aging with favorable accuracy, identity preservation, and fidelity using the interpretation (coefficient) of any pair of adjacent groups. The architecture consists of three parts: (1) personalized aging, (2) transformation-based age representation, and (3) representation of the age face decoder.

The personalized aging basis processes each individual dominated by a personalized aging factor using a neural network encoder E. maps input image to personalized $B_i = [b_{i1}, b_{i2}, \ldots, b_{im}]$ using $B_i = E(X_i)$. Given an aging basis, B_i can obtain age representation using an age-specific transform, formulated by:

$$r_i^k = \sum_{j=1}^{m} W_{kj} b_{ij} = B_i w_k \tag{28}$$

To achieve this framework aims, they use the following objective: Age loss for accurate face aging, which is formulated as:

$$l_i^{age} = -\sum_k \log(C_k(\hat{x_i^k})) \tag{29}$$

where $C_k(.)$ denote the probability that a sample falls into *k-th* age group, predicted by the classifier *C*.

L_1 Loss for identity preservation is formulated as:

$$l_i^{L1} = \sum_k \delta(y_i = k) ||x_i - \hat{x_i^k}||_1 \tag{30}$$

Adversarial Loss for image fidelity is formulated as:

$$l_i^{adv-d} = \max(1 - D(x_i, y_i), 0) + \sum_k \max\left(1 + D\left(\hat{x_i^k}\right), 0\right)$$
$$l_i^{adv-g} = \sum_k -D(\hat{X_i^k}, k) \tag{31}$$

where *D* is discriminator real, fake regulator.

Finally, the objective function for this framework can be formulated as:

$$\min_{E,G,\{w_k\}} \sum_i \lambda_1 l_i^{age} + \lambda_2 l_i^{L1} + l_i^{adv-g} \min_{E,G,\{w_k\}} \sum_i l_i^{adv-g} \tag{32}$$

where λ_1 and λ_2 are hyperparameters to balance the losses and optimize two objectives. S2GAN framework is illustrated in Figure 16.

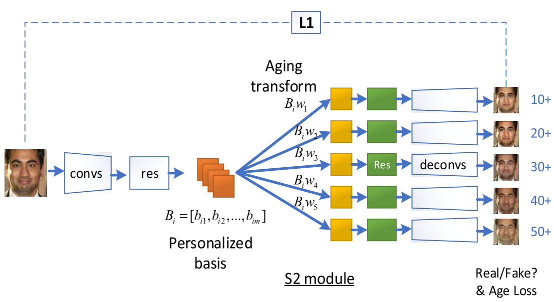


FIGURE 16. S2GAN: Share aging factor across age and share aging trends among individual framework.

h: AUTOMATIC FACE AGING IN VIDEO VIA DEEP ENFORCEMENT LEARNING

Duong *et al.* [46] proposed a novel approach for the automatic synthesis of age-progressed facial images in video sequences using deep enforcement learning. Modifying the face structure and longitudinal face aging process of a given subject across video frames. Deep enforcement learning preserves Garantie's visual identity from an input face.

The embedding function $\mathcal{F}_1$, maps X_y^t into latent representation $\mathcal{F}(X_y^t)$, a high-quality synthesis image, which has two main properties: (1) linearity separability and (2) detail preservation. Age progression can be interpreted as linear transversal from the younger region $\mathcal{F}(X_y^t)$ toward the older region $\mathcal{F}(X_o^t)$ using the formula:

$$\mathcal{F}_1\left(X_o^t\right) = \mathcal{M}\left(\mathcal{F}_1\left(x_y^t\right); X^{1:t-1}\right) = \mathcal{F}_1\left(X_y^t\right) + \alpha \Delta^{x^t|X^{1:t-1}} \tag{33}$$

where $\Delta^{x^t|X^{1:t-1}}$ learning from neighbors containing only aging effect only without the presence of other factors, i.e identity, pose, etc., estimated by:

$$\Delta^{x^t|X^{1:t-1}} = \frac{1}{K}\left[\sum_{x \epsilon \mathcal{N}_o^t} \mathcal{F}_1(\mathcal{A}(x, x_y^t)) - \sum_{x \epsilon \mathcal{N}_y^t} \mathcal{F}_1(\mathcal{A}(x, x_y^t))\right] \tag{34}$$

Framework Automatic face aging in video via deep enforcement shown Figure 17:

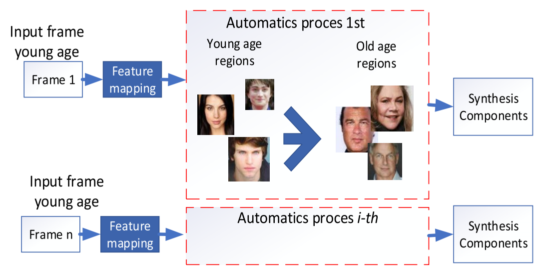


FIGURE 17. Automatic face aging in video via deep inforcement learning.

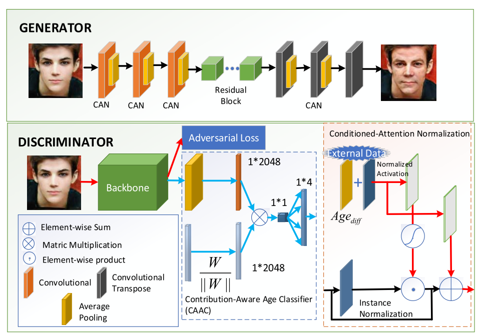


FIGURE 18. Conditioned-attention normalization GAN (CAN-GAN).

i: CONDITIONED-ATTENTION NORMALIZATION GAN (CAN-GAN)

Shi *et al.*, [47] introduced a Conditioned-Attention Normalization GAN (CAN-GAN), an architecture for age synthesis from input face images by leveraging the aging differences between two age groups, to capture a face aging region with different attention factors. This architecture can freely translate an input face into an aging face in certain age groups with strong identity preservation that satisfies the aging effect and provides authentic visualization. The CAN-GAN layer was designed to increase age-related information on the face when smoothing unrelated information by using an attention map.

In training CAN-GAN uses the following formulation for adversarial loss:

$$L_{adv} = E_x [D_{adv}(x)] - E_{x,age_{diff}} [D_{adv}(G(x, age_{diff}))] \\ -\lambda_{gp} E_{\tilde{x}}[(||\nabla_{\tilde{x}} D_{adv}(x)||_2 - 1)^2] \quad (35)$$

where x denotes the real image, and (x) is sampled between pairs of real and synthetic images. Construction loss used to construct an image in age target formulated as:

$$L_{rec} = ||x - G(x, age_{diff} = 0)||_1 \quad (36)$$

And optimized generator G by L^f_{cls} for optimizing D_{cls} and G when generating synthetics images to target a group by

$$L^r_{cls} = E_{x,c} \left[-\log D_{cls} (c|x) \right]; \\ L^f_{cls} = E_{x,c} \left[-\log D_{cls} \left(c'|G(x, age_{diff}) \right) \right]; \quad (37)$$

where c and c' denote to original age class label and age target class label. And final losses are formulated as:

$$L_D = L_{adv} + \lambda_1 L^r_{cls} \\ L_G = L_{adv} + \lambda_1 L^f_{cls} + \lambda_2 L_{rec} \quad (38)$$

where λ_1 and λ_2 is trade-off parameters.

Focusing on the important parts of the face that characterize changes in age, this architecture is more assertive in defining the attributes of the changes that occur at the desired age.

j: HIERARCHICAL FACE AGING THROUGH DISENTANGLED LATENT CHARACTERISTICS

Lie *et al.* [48] proposed a disentangled adversarial autoencoder (DAAE) to disentangle face images into three independent factors: age, identity, and extraneous information. A hierarchical conditional generator passes the disentangled identity age embedded into the high and low layers with class conditional batch normalization to prevent the loss of identity and age information used in this architecture. The disentangled adversarial learning mechanism boosted the quality progression.

By employing age distribution, DAAE can create face synthesis with the effect of face aging at an arbitrary age. The architecture learns from an input image, learns the age-progression distribution, and is treated as an age estimator. DAAE can estimate the age distribution efficiently and accurately.

DAAE consists of two components, the inference network E and the hierarchical generative network G, X_i, X_r and X_s denote the real sample, reconstruction sample, and new

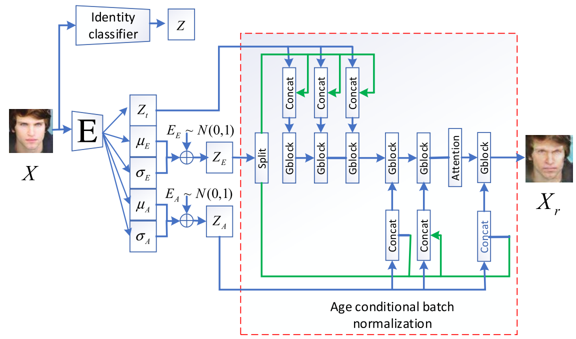


FIGURE 19. Hierarchical face aging through disentangled latent characteristics framework.

sample This framework, based on the original variational autoencoder (VAE) [49] and inspired by IntroVAE [50], aims to estimate the accuracy, identity preservation, and image quality. To preserve identity and age accurately, two regulations were used for generator G and formulated as follows:

$$\mathcal{L}^{(age)}_{reg} = \frac{1}{C_A} \sum_{i=1}^{C_A} ||Z'^i_A - Z_A|| + \frac{1}{C_A} \sum_{i=1}^{C_A} ||Z''^i_A - \hat{Z}_A|| \\ \mathcal{L}^{(id)}_{reg} = \frac{1}{C_I} \sum_{i=1}^{C_I} ||Z'^i_I - Z_I|| + \frac{1}{C_I} \sum_{i=1}^{C_I} ||Z''^i_I - \hat{Z}_I|| \quad (39)$$

where Z'_A and Z'_I are inferred representation generated image X_s, and Z''_A and Z''_I are inferred to generate an image $\hat{X}_s$.

To avoid a blurry image in VAE they use an inference network E and generator adversarial and formulate as:

$$\mathcal{L}^{(adv)}_E = \mathcal{L}^{(ext)}_{kl} (\mu_E, \sigma_E) + \propto \left[m - \mathcal{L}^{(ext)}_{kl} \left(\mu'_E, \sigma'_E \right) \right]^+ \\ + \left[m - \mathcal{L}^{(ext)}_{kl} \left(\mu''_E, \sigma''_E \right) \right]^+ \\ \mathcal{L}^{(adv)}_G = \mathcal{L}^{(ext)}_{kl} \left(\mu'_E, \sigma'_E \right) + \mathcal{L}^{(ext)}_{kl} \left(\mu''_E, \sigma''_E \right) \quad (40)$$

where m is positive margin, $\propto$ is weight coefficient, and (μ_E, σ_E), $\left(\mu'_E, \sigma'_E\right)$, and $\left(\mu''_E, \sigma''_E\right)$ computed from real data X_s, reconstruction sample X_r and new sample X_s.

The final objective of this framework are:

$$\mathcal{L}_E = \mathcal{L}_{rec} + \lambda_1 \mathcal{L}^{(age)}_{kl} + \lambda_2 \mathcal{L}^{(id)}_{kd} + \lambda_3 \mathcal{L}^{(adv)}_E \\ \mathcal{L}_G = \mathcal{L}_{rec} + \lambda_4 \mathcal{L}^{(age)}_{reg} + \lambda_5 \mathcal{L}^{(id)}_{reg} + \lambda_6 \mathcal{L}^{(adv)}_G \quad (41)$$

k: LIFESPAN AGE TRANSFORMATION SYNTHESIS

Or-El *et al.* [51] proposed a framework to address the problem of single-photo age progression and regression. A prediction of how a person in the future or looks in the past, a novel multidomain image-to-image generation method using GAN. The framework learns the latent-space mode, which is a continuous bidirectional aging process trained to approximate a continuous age transformation from 0 years to 70-year-olds, modifying the shape and texture with six anchorage classes.

This framework, which is based on a GAN, contains a conditional generator and a single discriminator. This conditional generator was responsible for the transition among the age groups. It consists of three parts: an identity encoder, mapping network, and decoder. This framework preserves identity by separately encoding identity and age.

A single generator consisting of an identity encoder, a latent mapping network, and a decoder was used for all ages. The training process uses an age encoder to embed both the real and generated images into the age latent space. The decoder uses an age with identity features and generates an output image with style using a convolutional block. In general, the generator mapping from the input image, the target age vector z, into output image z, is formulated as:

$$y = G(X, Z_t) = F(E_{id}(X), M(Z_t)) \tag{42}$$

The age encoder mapping of the input image X into the correct location vector space Z, produces an age vector corresponding to the target age group.

The discriminator-style GAN with a minibatch standard deviation discriminates the last layer, to force the synthesized image to produce an age target t.

In the training scheme, they processed source image clusters s into a target cluster t, where $s \neq t$, and performed three forward phases:

$$\begin{aligned} Y_{gen} &= G(X, Z_t) \\ Y_{rec} &= G(X, Z_s) \\ Y_{cyc} &= G(Y_{gen}, Z_s) \end{aligned} \tag{43}$$

where Y_{gen} is the generated image at target age *t*, $Y_{rec}$is the reconstructed image at source age *s*, and Y_{cyc} is the cycle-reconstructed image at source age *s*, generated from the generated image at age target *t*.

They used a conditional adversarial loss $\mathcal{L}_{adv}$ to generate an image at the target age t; self-reconstruct loss $\mathcal{L}_{rec}$ and cycle consistency loss $\mathcal{L}_{cyc}$ to force the generator to learn identity translation, and they formulated as:

$$\begin{aligned} \mathcal{L}_{adv}(G, D) &= E_{x,s}\left[\log D_s(x)\right] + E_{x,t}\left[1 - \log D_t\left(Y_{gen}\right)\right] \\ \mathcal{L}_{rec}(G) &= ||x - Y_{rec}||_1 \\ \mathcal{L}_{cyc}(G) &= ||x - Y_{cyc}||_1 \end{aligned} \tag{44}$$

And to keep identity preserving they use:

$$\mathcal{L}_{id}(G) = ||E_{id}(X) - E_{id}(Y_{gen})||_1 \tag{45}$$

Age vector loss to correct embedding of a real and generated image, by penalizing distance between age encoder output with age vector Z_s and Z_t which generated a sample by the generator. Age vector loss formulates as:

$$\mathcal{L}_{age}(G) = ||E_{age}(X) - Z_s)||_1 + ||E_{age}(Y_{gen}) - Z_t)||_1 \tag{46}$$

Framework optimized by:

$$\begin{aligned} \min_G \max_D \mathcal{L}_{adv}(G, D) + \lambda_{rec}\mathcal{L}_{rec}(G) + \lambda_{cyc}\mathcal{L}_{cyc}(G) \\ + \lambda_{id}\mathcal{L}_{id}(G) + \lambda_{age}\mathcal{L}_{age}(G) \end{aligned} \tag{47}$$

How the framework worked is shown in Figure 20.

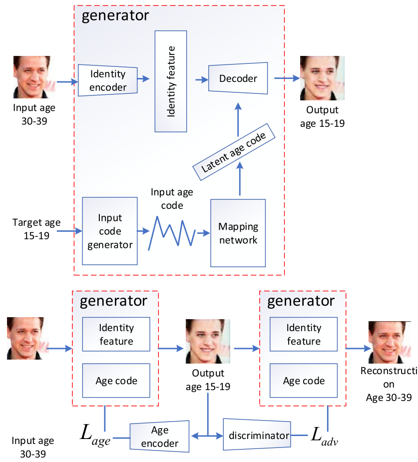


FIGURE 20. Lifespan age transformation synthesis.

I: PROGRESSIVE FACE AGING WITH GENERATIVE ADVERSARIAL NETWORK (PFA-GAN)

Huang *et al.* [52] proposed progressive face aging using a Generative Adversarial Network (PFA-GAN) to remove ghost artifacts that arise when the distance between age groups increases. The network was trained to reduce accumulated artifacts and blurriness. consists of several subnetworks whose job is to imitate the aging process from young to old, or vice versa. Learning was similar to the specific aging effect in two adjacent age groups. Age estimation loss was used to increase the aging accuracy, and Pearson's correlation was used as an evaluation metric to calculate aging smoothness.

The architecture comprises four generator sub-networks. Each subnetwork G_i is responsible for generating aging faces from i to $i+1$, consisting of a residual skip connection, binary gate λ_i, and subnetwork G_i. Binary gate λ_i can control the aging flow and determine which aging mapping of the subnetwork should be involved. Each subnetwork can be formulated as $X_t = G([X_s; C_t])$, and the progressive aging framework can be formulated as:

$$X_t = G_{t-1} \circ G_{t-2} \circ \cdots \circ G_s(X_s) \tag{48}$$

where the symbol $\circ$ is the function composition.

To prevent the architecture from producing the same image as the input image, a residual skip connection was added to each subnetwork to produce an identity mapping from the input to the output. By adding this connection and binary gate into the subnetwork the change from age group i to $i+1$ can be rewritten as:

$$X_{t+i} = G_i(X_i) = X + \lambda_i G_i(X_i) \tag{49}$$

where $\lambda_i \in \{0, 1\}$ is the binary gate that controls if the subnetwork G_i elaborate on the path to age groups. $\lambda_i = 1$ if the subnetwork G_i is among source age groups s and target age group t, i.e $s \leq i < t$ otherwise $\lambda_i = 0$. Tensor C used in cGAN based method converted into a binary gate vector $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{N-1})$ regulatory aging flow in the age

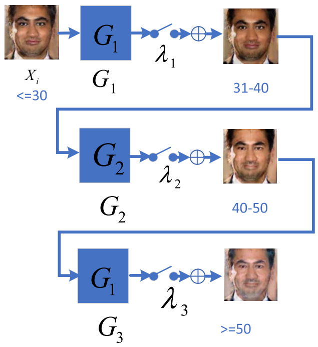


FIGURE 21. PFA-GAN framework.

progression framework in figure x and expressed as:

$$\begin{aligned} X_4 &= X_3 + \lambda_3 G_3(X_3) \\ &= X_2 + \lambda_2 G_2(X_2) + \lambda_3 G_3(X_3) \\ &= X_1 + \lambda_1 G_1(X_1) + \lambda_2 G_2(X_2) + \lambda_3 G_3(X_3) \end{aligned} \tag{50}$$

The network is highly elastic in modeling the age progression between two different age groups using this framework. Finally, by providing an image of a young X_s face from the source age group s, the aging process for the target age group t can be formulated as follows:

$$X_t = G(X_s; \lambda_{s:t}) \tag{51}$$

The least-square loss function for adversarial loss used in generator G formulate as:

$$\mathcal{L}_{adv} = \frac{1}{2} E_{x_s}[D([(X_s, \lambda_{s:t}); C_t 0]) - 1]^2 \tag{52}$$

And age estimation loss for progression face aging can formulate as:

$$\mathcal{L}_{age} = \frac{1}{2} E_{x_s}[\left||y \hat{-} y|\right|_2 + \ell(A(X)W, C_t)] \tag{53}$$

For identity consistency to keep the identity and discard unrelated information, they adapt mixed identity loss a Structural Similarity (SSIM) and formulated into three losses function $\mathcal{L}_{pix}$, $\mathcal{L}_{ssim}$, and $\mathcal{L}_{fea}$.

Final loss for generative loss and discriminative loss formulate as equation 50.

$$\begin{aligned} \mathcal{L}_G &= \lambda_{adv}\mathcal{L}_{adv} + \lambda_{age}\mathcal{L}_{age} + \lambda_{ide}\mathcal{L}_{ide} \\ \mathcal{L}_D &= \frac{1}{2} E_x[D([X; C]) - 1]^2 + \frac{1}{2} E_{x_s}[D([G(X_s, \lambda_{s:t}); C_t)] \end{aligned} \tag{54}$$

And Progressive Face Aging with Generative Adversarial Network (PFA-GAN) framework can be illustrated as:

m: AGE FLOW: CONDITIONAL AGE PROGRESSION AND REGRESSION WITH NORMALIZING FLOW

Huang *et al.* [53] proposed a framework that integrates the advantages of a flow-based method model and GAN. It Consists of three parts: (1) an encoder that maps a given face image into a latent space. Through an invertible conditional translation module (ICTM) that translates the source latent vector to another target vector, a decoder reconstructs the resulting face of the target latent vector with the same encoder, all of which are invertible, which can achieve bijective aging mapping.

The novelty of ICTM is its ability to manipulate the direction of change in age progression. While keeping the other attributes unchanged, to keep the changes insignificant.

Second, they used a latent space to ensure that the resulting latent vector was indistinguishable from the original. The experimental results demonstrated superior performance. The flow-based encoder G maps the input images into latent spaces and decoder that inverse using function G^{-1} formulated the encoding process $z = G(I)$ and generated reserve procedure $I = G(z)$, where z is are gaussian distribution $p(z)$.

Generator G optimized by

$$\mathcal{L}_G = -E_{z\ p\theta(z)}[\log p\theta(z) + \log|det \frac{dG}{dI}|] \tag{55}$$

where G is the extracted latent vector and G^{-1} is the inverse of G.

The invertible conditional translation module is given a latent vector z_s from the source age group C_s, and with age, the progression translates z_s to z_t (age target C_t) using the ICTM. Cycle consistency was achieved by the reversibility of the ICTM. Each flow contained two convolutional networks with a channel attention module to focus the model on the required latency.

The discriminator used a simple multilayer perceptron (MLP) with 512 neurons, followed by spectral normalization [54], leaky ReLU as activations, negative slope 0.2, bottom layer with 1 neuron, and others for age classification to improve the age accuracy in the framework.

To make this framework work the loss function in this framework contained Attribute-aware Knowledge Distillation Loss ($\mathcal{L}_{akd}$), Adversarial loss ($\mathcal{L}_{adv}$), Age Classification Loss ($\mathcal{L}_{age}$), and Consistency Loss ($\mathcal{L}_{cl}$). Formulate as:

$$\begin{aligned} \mathcal{L}_{akd} &= E_{z_s}|z'_t - (z_s + s \times (\bar{z}_{t,\alpha_i} - \bar{z}_{t,\alpha_i}))| \\ \mathcal{L}_{adv} &= \frac{1}{2} E_{z_s}[(D(Z'_t) - 1)^2] \\ \mathcal{L}_{age} &= E_{z_s}[\ell(A(Z'_t), t)] \\ \mathcal{L}_{cl} &= -E_{z_s\ p\theta(z)}[\log p\theta(\mu'_s, \sigma'_s, z_s)] \end{aligned} \tag{56}$$

And final loss formulates as:

$$\mathcal{L}_T = \lambda_{akd}\mathcal{L}_{akd} + \lambda_{adv}\mathcal{L}_{adv} + \lambda_{age}\mathcal{L}_{age} + \lambda_{cl}\mathcal{L}_{cl} \tag{57}$$

The framework of age flow can see in as:

n: DISENTANGLED LIFESPAN FACE SYNTHESIS

He *et al.* [54] proposed a lifespan face synthesis (LFS) model to generate a set of photo-realistic face images of an entire person throughout life, with only one image as a reference.

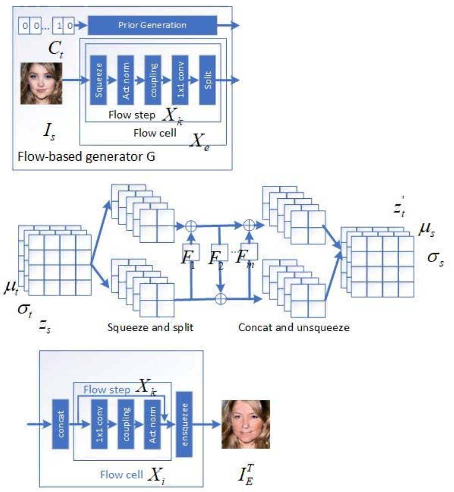


FIGURE 22. Age Flow: Conditional age progression and regression with normalizing flow.

The generated face image on a target of a given age provides a nonlinear transformation of the structure and texture.

The LFS is a GAN conditional on a latent face representation. LFS was proposed to disentangle face synthesis, where structure, identity, and age transformation can be modeled effectively.

The shape, texture, and identity features were extracted separately from the encoder. There are two transformer modules, one conditional-convolutional-based and the other channel attention-based, to model the nonlinear shape and texture. Accommodate distinct aging processes and ensure the synthesis image has age sensitivity and is identity-preserved.

Three distinct set features: shape $f(s)$, texture $f(t)$, and identity $f(id)$ are extracted from the encoder. And formulated as:

$$\begin{aligned} f(s) &= R_s(\varepsilon_m(I_r)) \\ f(t) &= \mathcal{T}(\varepsilon_d(I_r)) \\ f(id) &= \mathcal{T}_d(\varepsilon_d(I_r)) \end{aligned} \tag{58}$$

where R_s is a residual block used to extract shape information and I_r is a convolutional projection module for extracting texture information and pooling it into a vector identity. Identity features were extracted using another convolutional projection module $_f(id)$.

$$f_t(Z_t) = \mathcal{T}_t(f_t, Z_t) = f_t \circ P_t(A_\varepsilon(Z_t)) \tag{59}$$

where $\circ$ is element-wise multiplication, and P_t is linear projection layer. Image generation generator G: $I_t = G(f_s(Z_t), f_t(Z_t))$ transforming reference face images into older age groups with the same shape information or $\mathcal{L}_s = ||R_s(\mathcal{E}_m(I_{r_e}) - R_s(\mathcal{E}_m(I_{r_e}))||^2$ the difference is minimized. To make the framework achieve the aim they use the identity loss function $\mathcal{L}_{id}$, cycle consistency loss $\mathcal{L}_{cyc}$, reconstruction loss $\mathcal{L}_r$, conditional adversarial loss $\mathcal{L}_{adv}$ and

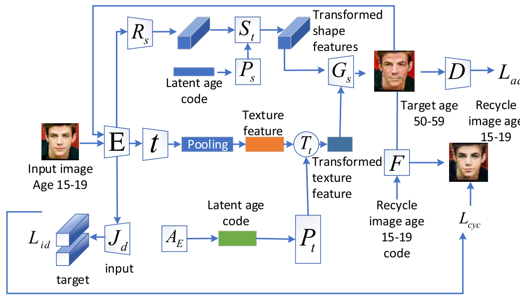


FIGURE 23. Disentangled lifespan face synthesis.

formulate as:

$$\begin{aligned} \mathcal{L}_{id} &= ||ID(\mathcal{E}_d(I_r)) - ID(\mathcal{E}_d(I_t))||^2 \\ \mathcal{L}_{cyc} &= ||I_r - f(I_t, Z_r)||^2 \\ \mathcal{L}_r &= ||I_r - G(f_s(Z_r), f_t(Z_r))||^2 \\ \mathcal{L}_{adv} &= E_{I^r_m}\, P(data(I^r_m))[\log(D(I^r_m|Z)] \\ &\quad + E_{I^g_m}\, P(data(I^g_m))[1 - \log(D(I^g_m|Z) \end{aligned} \tag{60}$$

Overall training objective formulated as:

$$\mathcal{L} = \lambda_{adv}\mathcal{L}_{adv} + \lambda_r\mathcal{L}_r + \lambda_{cyc}\mathcal{L}_{cyc} + \lambda_{id}\mathcal{L}_{id} + \lambda_s\mathcal{L}_s \tag{61}$$

where λ_{adv}, λ_r, λ_{cyc}, λ_{id}, and λ_s are hyperparameters to balance the objective function. The framework for Disentangled Lifespan face synthesis can see in Figure 24

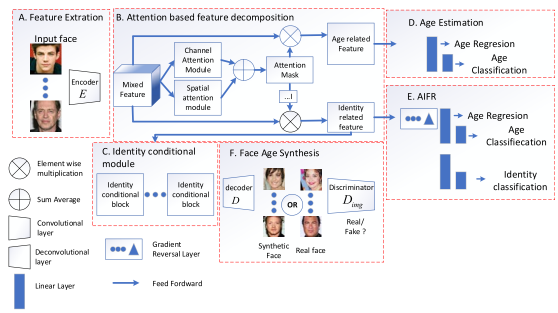


FIGURE 24. Age-invariant face recognition and age synthesis: Multitask learning framework.

o: AGE-INVARIANT FACE RECOGNITION AND AGE SYNTHESIS: MULTITASK LEARNING FRAMEWORK

Zhizong *et al.* [55] proposed a multitask framework to handle two tasks, termed MTL-Face, which can learn identity representation and generate pleasing face synthesis. Two unrelated components were decomposed: identity and age-related features through an attention mechanism.

To make this framework optimize, they conduct the following process:

(1) Age-invariant face recognition task for preserving identity formulated as:

$$\begin{aligned} \mathcal{L}^{AIFR} &= \mathcal{L}_{cosface}(L(X_{id}), Y_{id}) + \lambda^{AIFR}_{age}\mathcal{L}_{AE}(A_E(X_{age})) \\ &\quad + \lambda^{AIFR}_{id}\mathcal{L}_{AE}(GLR(X_{id})) \end{aligned} \tag{62}$$

(2) Face synthesis task generated synthesis image at age target use age label

$$I_t = D(\{E_l(I)\}_{l=1}^{3}, ICM\ (X_{id}, t)) \tag{63}$$

(3) Improve and stable training process using Least Square GAN(LSGAN):

$$\mathcal{L}_{adv}^{FAS} = \frac{1}{2} E_I[D_{img}([I_t; C_t])] \tag{64}$$

(4) Face aging and rejuvenation in holistic formulated as:

$$\begin{aligned} X_{age}^t, X_{id}^t &= AFD(E\ (I_t)) \\ \mathcal{L}_{age}^{FAS} &= l_{CE}(X_{age}^t, t) \\ \mathcal{L}_{id}^{FAS} &= E_{X_s} ||X_{id}^t - X_{id}||_F^2 \end{aligned} \tag{65}$$

where ||.|| represents the Frobeneius norm.

(5) Final loss processed by:

$$\mathcal{L}^{FAS} = \lambda_{adv}^{FAS} \mathcal{L}_{adv}^{FAS} + \lambda_{id}^{FAS} \mathcal{L}_{id}^{FAS} + \lambda_{age}^{FAS} \mathcal{L}_{age}^{FAS} \tag{66}$$

Using a weight-sharing strategy, this framework succeeded in increasing the smoothness of synthesizing new faces in the age group generated in the desired age group under wild conditions.

The age-invariant face recognition and age-synthesis multitask learning frameworks are as follows:

Age-Invariant Face Recognition and Age Synthesis: Multitask Learning Framework illustrated:

IV. CHALLENGE IN FACE AGING DATASET

The challenge in a deep neural network is the quality and number of datasets, and the success of a network is based on the availability of data that allows it to be properly trained to produce models. Currently, dataset availability relies on public datasets, and datasets related to children are limited. Most of the datasets used in face aging research related to children are private data. Obtaining a child dataset is difficult because of the several legal regulations that protect children's rights to privacy.

Ethnicity is also a challenge in determining facial aging patterns. The face aging pattern is nonlinear. If there are diverse datasets, the algorithm can be retrained to produce aging patterns across races.

Face-aging is a challenging dataset with a large age range in the age group dataset, which makes it difficult to find a pattern.

High-resolution generated faces are more attractive, and producing a high-resolution image is challenging in terms of the algorithms used to build high-resolution images difficult to find.

The incorrect label is also a challenge, to give the correct age label to the dataset is not easy, until now there is no clear benchmark in determining the age label of an image in the dataset, many datasets are collected with incorrect labels.

External factors, such as the aging pattern of each individual, are not linear and are influenced by lifestyle, nutrition, the environment, and disease. A dataset containing this information could provide a new direction for future research.

A. METHOD

A model-based approach using an anthropological perspective in the active appearance model is represented by an aging function that applies the mean face parameter of different age groups [15], [16] or a variant aging effect [17]. The difficulty of this method is the construction of a suitable training set where sequential age progression from different individuals is required to create face model transformation. Suo *et al.* [18] used a compositional and dynamic model for face aging, representing the face in each age group with a hierarchal "And-Or" graph. This compositional and dynamic model can build an aging effect perception and preserve identity from an input image in high but has difficulty developing a model without image sequences, cannot train for a long age range, and requires precise alignment to produce fine synthetic images without losing information.

Prototype-based approaches can produce reliable age subspaces and alignments that can handle wild photographs with numerous variations in poses, expression, and lighting. This method uses a collection of head and torso images to obtain synthetic images for a certain age. The prototype model approach requires precise image alignment to avoid plausible facial images.

Reconstruction-based approaches such as coupled Dictionary Learning (CDL) model aging patterns maintain the personalization of each person's face, use the aging base or basic pattern of each age group, and combine it into the input image. Producing short-term or neighborly patterns of age change requires a complete dataset of individuals, which is a difficult task.

A deep neural network approach, such as the TRBM and Recurrent Neural Network is a better approach to face aging. This method uses past information to identify soft transition patterns between age groups. Using a single unit model, interpretation of facial changes can be achieved The facial structure and changes in face aging can be interpreted in one training, this makes it easier for the architecture to form a robust and simple model. However, this approach requires a long computation time and a large number of datasets to produce a robust model.

The Generative Adversarial Network (GAN) is another approach for face aging. The translation method is based on how style is transferred from one set of group images to another set of a group images. For example, CycleGAN, which uses this method, attempts to capture the style of one age group from other age groups. CycleGAN has the advantage that it does not require consecutive photos of the same individual in each age group domain and only requires that each age group in the dataset has a sufficient number. The sequential-based approach seeks to identify the facial aging patterns in each adjacent age group. In contrast to other approaches that seek to obtain a direct model, this approach is more concerned with a sequential approach and requires

TABLE 2. Score, categories, and method face aging GAN.

Category	Method	Advanced	Criteria	Accuracy or MAE score
Conditional-based	Conditional Adversarial Autoencoder(CAAE) [4]	Smooth age regression and progression, producing more photorealistic.	Preserving Identity	Same person as ground truth 43.38%, Not same as ground truth 29.58%, Not Sure 22.04%
			Age Progression	From 235 paired images 79 subjects, 47 respondents 1508 votes 52.77 votes CAAE is better, 28.99% prior work better, 18.24% they equal
Conditional-based	Age Conditional Generative Adversarial Neural Network Model[27]	Preserving identity optimizations	Face recognition	Initial reconstruction 53.2% accuracy, "Pixelwise" Optimization 59.8% accuracy, Identity-Preserving Optimizations 82.9% accuracy
Conditional-based	Contextual Generative Adversarial Nets (C-GANs) [28]	Correctly synthesize real age progression	Face recognition	Identity preserving is a good result while input age and target age are contiguous, with different ranges maximum of 8 years. The majority have 90% accuracy
			Cross-Age Face Verification	Equal error rate (EER) on cGANs Synthetic Pair 8.72%, CAAE Synthetic Pair 11.05%, and Original Pair 17.41%
Conditional-based	Identity-Preserved Conditional Generative Adversarial Network (IPCGANs) [6]	Realistic face synthesis, good preserving identity, and age consistency	Face Verification	96.90%
			Image Quality	71.74%
			Age Classification	31.74%
			VGG-Face Score	36.33±1.85
			Time Cost	0.28s
Translation-based	Subject-dependent Deep Aging Path (SDAP) [7]	Efficient in synthesizing in the wild aging faces.	Face identification (SF+SDAP)	64.4%
			Age Estimation	SDAP's synthesized faces MAEs 3.94
Sequence-based	Two-Stage Approach for Identity-Preserving Face Synthesis [39]	Photo-realistic image, with high preserving identity quality	Identity Preserving	97.62%±0.78
			Similarity+Image Quality	Similarity 0.693, User Score 22.4%
Translation-based	Child Face Age-Progression via Deep Feature Aging[38]	Ability to identify young children	FACENET	With Feature Aging Module, CFA-Rank-1 55.30%, CFA-Rank-1 @1% FAR 53.58%, ITWCC-Rank-1 21.44%, ITWCC-Rank-1 @1% FAR 19.96%
			COSFACE	With Feature Aging Module, Face-identifaction CFA-Rank-1 94.24%, CFA-Rank-1 @1% FAR 94.24%, ITWCC-Rank-1 66.12%, ITWCC-Rank-1 @1% FAR 25.04%, Face recognition: 95.91% FGNET dataset, 99.58% CACD Dataset
Conditional-based	Age Progression and Regression with Spatial Attention Modules[43]	Synthesizing lifelike face image at desired age with personalized feature and keeping un age-irrelevant unchanged	Age Estimation Error	1.53±6.50 (MORPH), 1.78±7.53(CACD), 4.77±10.59(UTKFace)
			Verifaction Rate	100%(MORPH), 99.92%(CACD), 98.10%(UTKFace)
Translation-based	Triple-GAN: Progressive Face Aging with Triple Translation Loss[36]	Identity classification and age classfication	Age estimation using face++	Average 94.01% (MORPH), Average 93.38% (CACD Dataset)
			Age classification accuracy	Average 71.26% (MORPH), Average 69.43% (CACD)
Translation-based	Generative Adversarial Style Transfer Networks for Face Aging[37]	Aging effect	Age Progression (Likert scale:1-5)	3.17±0.016
			Age Regression (Likert scale:1-5)	3.28±0.075
Conditional-based	Conditioned-Attention Normalization GAN (CAN-GAN) [47][53]	Aging relevant information	Age estimation	Synthetic Face with CAAC: MORPH-(30-40 year) 36.01%±6.58 (41-50 year) 46.43%±5.89 (51-77 year) 58.52%±5.93 CACD: (30-40 year) 39.62% ± 8.35 (41-50 year) 47.73% ± 7.22 (51-77 year) 57.04% ± 8.45 Synthetic Face w/o CAAC:MORPH (30-40 year) 36.15% ± 6.54 (41-50 year) 47.38% ± 6.27 (51-77 year) 64.23% ± 5.96 CACD
				Real face:MORPH (16-30) 27.86% ± 6.12 (30-40 year) 39.11% ± 7.43 (41-50 year) 48.34%± 8.37 (51-77 year) 58.09% ± 8.90 CACD (16-30) 30.93% ± 7.80 (30-40 year) 38.40 ± 9.03 (41-50 year) 46.63 ± 10.34 (51-77 year) 53.99 ± 10.93
			Face Verification	Test face: MORPH Synthetics face1 96.45% Synthetics face2 95.04% Synthetics face3 90.09% CACD: Synthetics face1 95.58% Synthetics face2 93.72% Synthetics face3 88.14%; Synthetic face1: MORPH Synthetics face2 96.32% Synthetics face3: 92.20% CACD: Synthetics face2 95.65% Synthetics face3 89.03% Synthetic face2: MORPH Synthetics face3: 95.28% CACD: Synthetics face3 91.48%
Conditional -based	Progressive Face Aging with Generative Adversarial Network (PFA-GAN)[52]	Aging estimation loss and Pearson correlation for aging smoothness	Face Verification Rate	MORPH: 31-40(100%), 41-50(100%), 50+(99.7%) CACD:31-40(99.97%), 41-50(99.89), 50+(99.69)
Conditional -based	Age Flow: Conditional Age Progression and Regression with Normalizing Flow[53]	attribute-aware knowledge distillation to learn the manipulation direction of age progression while keeping other unrelated attributes unchanged, alleviating unexpected changes in facial attributes.	Age Accuracy Cosine Similarity gender Race	Age Accuracy: MORPH, 30-(87.19%), 31-40(91.85%), 41-50(90.68%), 50+(90.71%), CACD, 30-(86.49%), 31-40(78.06%),41-50(85.91%), 50+(88.15%) Cosine similarity: MORPH, 30-(0.897),31-40(0.923), 41-50(0.903), 50+(0.767), CACD, 30-(0.905), 31-40(0.926), 41-50(0.919), 50+(0.847) Gender: MORPH, 31-40(98.19%), 41-50(99.26%),51+(98.21), CACD, 31-40(98.63%), 41-50(98.02%), 51+(98.49%) Race: MORPH, 31-40(98.80%), 51+(97.72%)
Conditional -based	Disentangled Lifespan face synthesis[54]	lifespan face synthesis (LFS) generate a set of photo-realistic face images of a person's	Identity preservation	Identity preservation(3.07±0.19), Shape transformation(3.18±0.35), Teture transformation(3.30±0.21), Reconfiguration(4.07±0.27), Age error(3.53±2.81), Age Accuracy(65.6%)
Conditional -based	Hierarchical Face Aging through Disentangled Latent Characteristics[48]	Accurate estimated age distribution for synthetics face.	Aging Accuracy	MORPH, 31-40(99.48%), 41-50(99.36%), 50+(99.36) CACD, 31-40(99.24%), 41-50(99.19), 50+(99.19%)
Conditional -based	Lifespan Age Transformation Synthesis[51]	multi-domain image-to-image generative adversarial network architecture,	Identity Accuracy Age Accuracy	From 50 images, Same Identity: 15-19(50) , 30-39(45), 50-59(41), All(136) Age difference, 15-19(12.7%),, 30-39(11.6%), 50-59(9.%), All(11.3)
Conditional -based	Age-Invariant Face Recognition and Age Synthesis: Multitask Learning Framework[55]	Weight-sharing strategy improved smoothness of synthesis face	Accuracy	A96.23%(AgeDB-30), 95.62%(CALFW), 99.55%(CACD-VS), 94.78%(FG-NET)
Conditional -based	S2GAN: Share Aging Factor Across Age and Share Aging Trends Among individual[45]	The personalized aging basis for the synthesis of aging factors with age-specific transform	Age Accuracy	MORPH:Average of All Pairs (99.69%), Hardest Pair-(11-20,50+)(96.08%), Easiest Pair (11-20-21-30):100% CACD: Average of All Pairs (98.91%), Hardest Pair-(11-20,50+)(94.08%), Easiest Pair (11-20-21-30):99.96%
Conditional -based	Automatic Face Aging In Video via Deep Enforcement Learning[46]	Age-progressed faces, temporal smoothness, and cross-age face verification	Aging Consistency Temporal Smoothness Matching Accuracy	Aging Consistency: 245.64 Temporal Smoothness: 61.80 Matching Accuracy: 83.67%
Conditional -based	Dual conditional GANs for Face Aging and Rejuvenations[40]	Photo realistic face	Score of images	4+→10+(0.86), 8+→30+(0.82), 10+→20+(0.81), 10+→40+(0.80), 20+→60+(0.80), 30+→60+(0.72), 40+→60+(0.84), Average(0.81)

TABLE 3. Score, categories, and method face aging GAN.

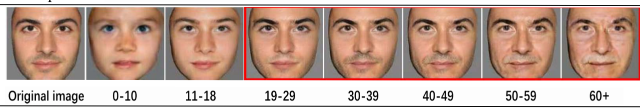
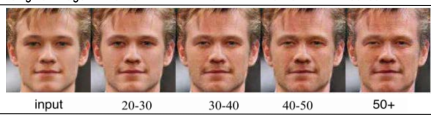
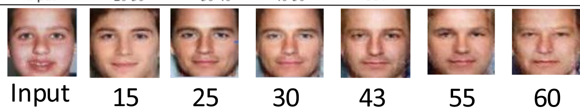
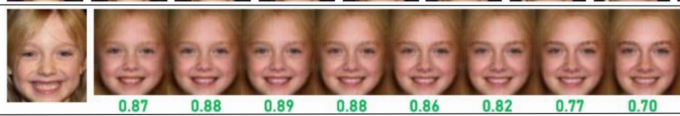
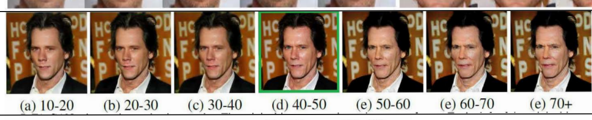
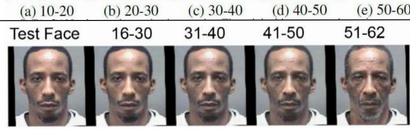
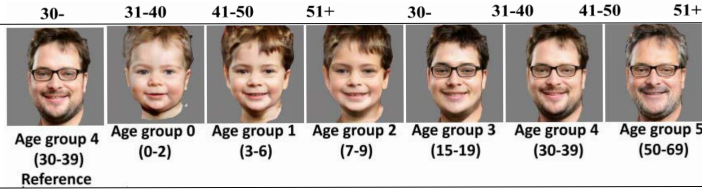
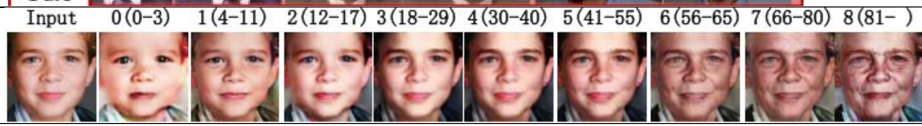
Category	Method	Sample Image
Conditional-based	Conditional Adversarial Autoencoder(CAAE) [4]	
Conditional-based	Age Conditional Generative Adversarial Neural Network Model[27]	
Conditional-based	Contextual Generative Adversarial Nets (C-GANs) [28]	
Conditional-based	Identity-Preserved Conditional Generative Adversarial Network (IPCGANs)[6]	
Translation-based	Subject-dependent Deep Aging Path (SDAP) [7]	
Sequence-based	Two-Stage Approach for Identity-Preserving Face Synthesis[39]	
Translation-based	Child Face Age-Progression via Deep Feature Aging[38]	
Conditional-based	Age Progression and Regression with Spatial Attention Modules[43]	
Translation-based	Triple-GAN: Progressive Face Aging with Triple Translation Loss[36]	
Translation-based	Generative Adversarial Style Transfer Networks for Face Aging[37]	
Conditional-based	Conditioned-Attention Normalization GAN (CAN-GAN) [47][53]	
Conditional -based	Progressive Face Aging with Generative Adversarial Network (PFA-GAN)[52]	
Conditional -based	Age Flow: Conditional Age Progression and Regression with Normalizing Flow[53]	
Conditional -based	Disentangled Lifespan face synthesis[54]	

TABLE 3. ***(Continued.)*** **Score, categories, and method face aging GAN.**

Conditional -based	Hierarchical Face Aging through Disentangled Latent Characteristics[48]	
Conditional -based	Lifespan Age Transformation Synthesis[51]	
Conditional -based	Age-Invariant Face Recognition and Age Synthesis: Multitask Learning Framework[55]	
Conditional -based	S2GAN: Share Aging Factor Across Age and Share Aging Trends Among individual[45]	
Conditional -based	Automatic Face Aging In Video via Deep Enforcement Learning[46]	
Conditional -based	Dual conditional GANs for Face Aging and Rejuvenations[40]	

a sequential dataset for each age group. Training will be conducted gradually from the young age group to the older age group, and vice versa. This transition was performed sequentially and each stage was trained independently or separately to complete the aging process. The conditional-based approach uses a label as one-hot code in the architecture embedded in the generator or discriminator. This method required a dataset with clear (correct) labels. A GAN is a convolutional neural network, and its architecture requires a long training time and high computational cost. The advantage of this approach is that it can produce a single model that can be applied to all age groups, and the resulting model can make a smooth and real transition from one age group to another, leaving only a few artifacts. The attention mechanism can enhance this architecture and emphasize parts that produce an aging pattern.

Following figure 6, we describe the timeline for the development of face aging using a generative approach. To describe the performance of each method, we provide the results in Table 2.

V. CONCLUSION

The dataset used in the facial aging process also plays an important role in the success of identifying aging patterns in each age group. Smaller age range better than larger age range, 5-year range better than 10-year range, because 5 years range has fewer differences than 10 years range. makes it an aging facial architecture, which makes it easy to identify an aging pattern. Teenagers aged 20 years have large differences from adults aged 30 years. They cannot be grouped into the same age range, suggesting that the age range must be narrowed down to 5 years. A young age range must be added to increase the ability of the architecture to generate young aging patterns. The dataset label must be clean and incorrect in the image(a photo must be correctly grouped). The dataset distribution of each age group must be balanced to prevent bias in the training process, and the number of images for each age group must be sufficient such that the architecture can produce a good model. The diversity of the dataset can improve the quality of the observed face aging pattern.

Based on the comparison of the Mean Absolute Error (MAE) and the accuracy of each method in Appendix B, the face aging architecture that uses the conditional GAN approach is still superior. The current approach applies state-of-the-art models to mobile and edge computers.

The synthetic face image quality during the face aging process depended on the algorithm used. In a model-based approach, it is difficult to find a general model for a certain age group, and a bad-alignment model with a face image produces blurry plausible images and identity information loss. Using a dynamic model, the resulting synthetic photovariations were found to be significant using a dynamic model. This prototyping approach eliminates the global identity information of an individual in synthetic facial images and produces image-loss identity information. A reconstruction approach in which CDL and identity information can be maintained, even though the reconstruction process must be sequenced from one aging group to another, generates synthetic faces with considerable variation. The Generative Adversarial approach is still the best, with the approach method, by finding the minimum value of the mean square

FIGURE 25. Dataset sample image.

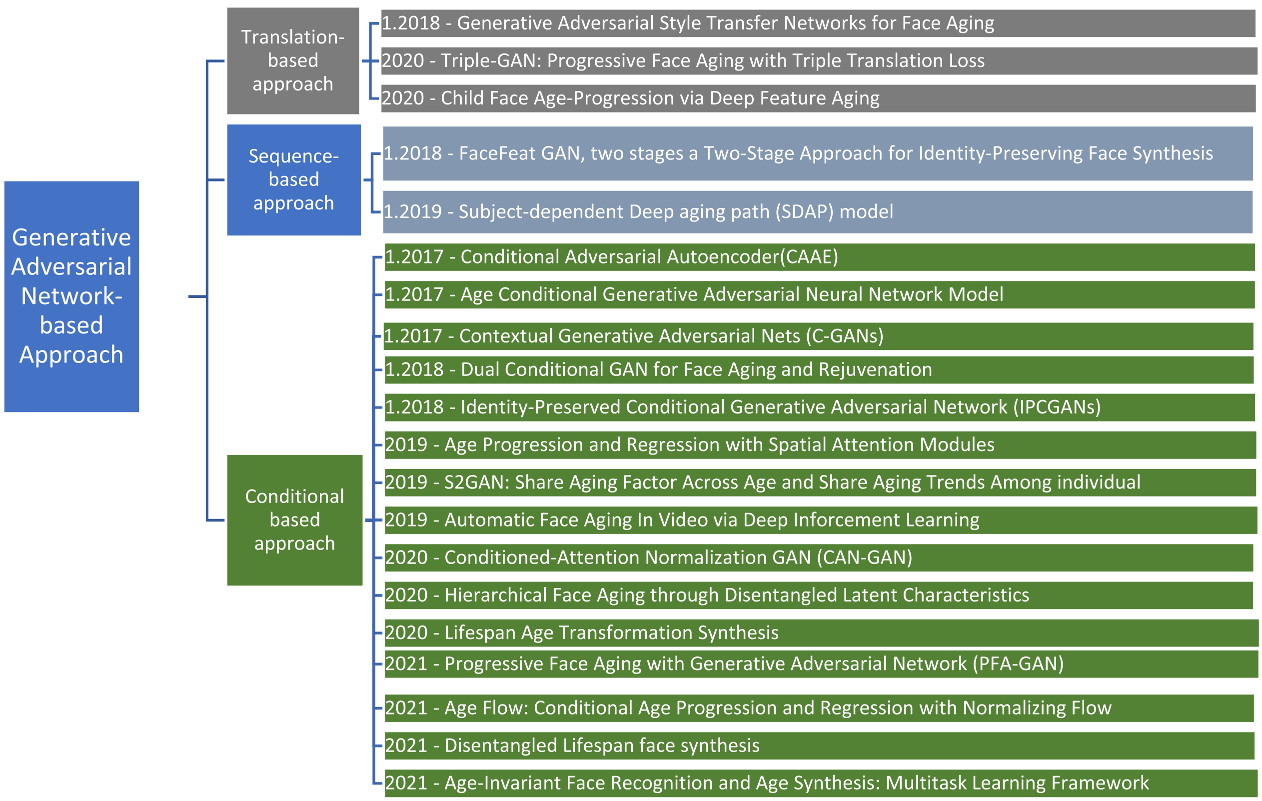


FIGURE 26. Generative adversarial network approach categorygenerative adversarial network approach category. A translation-based method is a method that captures style characteristics from a set of an image to be implemented in another set of images. The sequence-based method is a method in which is model is implemented step by step process, and each model is trained independently, to produce a sequential translation between two neighboring age groups. The conditional-based method uses a conditional in its architecture to produce an artificial face image in a certain age group, usually the conditional is in the form of a label that is made into a one-hot encoder tensor.

error for images of a certain age group, the pattern of age groups can be found, and by finding the minimum value of perceptual loss with the original image, the aging pattern can be implemented into input face images and identity information is maintained. The alternative method uses cyclic consistency in Fang *et al.* 's research [36] to maintain its identity or uses two stages of synthesis (feature generation and feature to image rendering) in GAN. Feature generation tasks to synthesize various facial features render them photorealistic in the image domain with high diversity but preserve their identity.

For future research, face aging at high resolution is a challenge, because processing high-resolution images make the architecture require a high computation process, and has challenged finding algorithms or methods that have a lower requirement and are faster. Improving the quality of datasets in the context of diversity is a challenge. A dataset with a variety of races, living environments, nutrition, and lifestyles can open the opportunity for research on the effects of nutrition, living environment, nutrition, lifestyle, and face aging on Asian people. The effects of disease and sickness on facial aging have been investigated.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

APPENDIX
A SCORE, CATEGORIES, AND METHOD FACE AGING GAN

See Table 2.

APPENDIX B
SAMPE IMAGES FROM EACH GAN METHOD

See Table 3.

APPENDIX C
DATASET SAMPLE IMAGE

See Figure 25.

APPENDIX D
GENERATIVE ADVERSARIAL NETWORK APPROACH CATEGORY

See Figure 26.

ACKNOWLEDGMENT

No funding for this research.

REFERENCES

[1] N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 2006, pp. 387–394, doi: 10.1109/CVPR.2006.187.

[2] C. N. Duong, K. Luu, K. G. Quach, and T. D. Bui, "Longitudinal face modeling via temporal deep restricted Boltzmann machines," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5772–5780, doi: 10.1109/CVPR.2016.622.

[3] W. Wang, Z. Cui, Y. Yan, J. Feng, S. Yan, X. Shu, and N. Sebe, "Recurrent face aging," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2378–2386, doi: 10.1109/CVPR.2016.261.

[4] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4352–4360, doi: 10.1109/CVPR.2017.463.

[5] C. N. Duong, K. G. Quach, K. Luu, T. Hoang Ngan Le, and M. Savvides, "Temporal non-volume preserving approach to facial age-progression and age-invariant face recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 3735–3743.

[6] X. Tang, Z. Wang, W. Luo, and S. Gao, "Face aging with identity-preserved conditional generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, vol. 9, no. 1, pp. 7939–7947, doi: 10.1109/CVPR.2018.00828.

[7] C. N. Duong, K. G. Quach, K. Luu, T. H. N. Le, M. Savvides, and T. D. Bui, "Learning from longitudinal face demonstration—Where tractable deep modeling meets inverse reinforcement learning," 2017, *arXiv:1711.10520*.

[8] K. Ricanek, Jr., and T. Tesafaye, "MORPH: A longitudinal image age-progression, of normal adult," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit.*, 2006, pp. 1–4.

[9] A. Lanitis, (Oct. 2010). *FG-NET Aging Database*. [Online]. Available: http://www.fgnet.rsunit.com

[10] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 12, pp. 2170–2179, Sep. 2014, doi: 10.1109/TIFS.2014.2359646.

[11] G. Levi and T. Hassncer, "Age and gender classification using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 34–42, doi: 10.1109/CVPRW.2015.7301352.

[12] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Cross-age reference coding for age-invariant face recognition and retrieval," in *Proc. ECCV*, 2014, vol. 16, no. 7, pp. 768–783.

[13] R. Rothe, R. Timofte, and L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," *Int. J. Comput. Vis.*, vol. 126, pp. 144–157, Apr. 2018, doi: 10.1007/s11263-016-0940-3.

[14] S. Moschoglou, C. Sagonas, and I. Kotsia, "AgeDB: The first manually collected, in-the-wild age database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2006, pp. 51–59.

[15] E. Patterson, K. Ricanek, M. Albert, and E. Boone, "Automatic representation of adult aging in facial images," in *Proc. Int. Conf. Vis. Imag., Image Process.*, 2006, pp. 1–6.

[16] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 442–455, Apr. 2002, doi: 10.1109/34.993553.

[17] X. Geng, Y. Fu, and K. S. Miles, "Automatic facial age estimation," in *Proc. 11th Pacific Rim Int. Artif. Intell.*, Aug. 2010, pp. 1–130.

[18] J. Suo, S.-C. Zhu, S. Shan, and X. Chen, "A compositional and dynamic model for face aging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 385–401, Mar. 2010.

[19] I. Kemelmacher-Shlizerman, S. Suwajanakorn, and S. M. Seitz, "Illumination-aware age progression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3334–3341, doi: 10.1109/CVPR.2014.426.

[20] D. A. Rowland and D. I. Perrett, "Manipulating facial appearance through shape and color," *IEEE Comput. Graph. Appl.*, vol. 15, no. 5, pp. 70–76, Sep. 1995, doi: 10.1109/38.403830.

[21] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman, "Sift flow: Dense correspondence across different scenes," in *Proc. ECCV*, 2008, pp. 28–42, doi: 10.1007/978-3-540-88690-7_3.

[22] X. Shu, J. Tang, H. Lai, L. Liu, and S. Yan, "Personalized age progression with aging dictionary," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3970–3978, doi: 10.1109/ICCV.2015.452.

[23] X. Shu, J. Tang, Z. Li, H. Lai, L. Zhang, and S. Yan, "Personalized age progression with bi-level aging dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 905–917, Apr. 2018, doi: 10.1109/TPAMI.2017.2705122.

[24] C. N. Duong, K. G. Quach, K. Luu, T. H. N. Le, and M. Savvides, "Temporal non-volume preserving approach to facial age-progression and age-invariant face recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3755–3763, doi: 10.1109/ICCV.2017.403.

[25] B. Green, T. Horel, and A. V. Papachristos, "Modeling contagion through social networks to explain and predict gunshot violence in Chicago, 2006 to 2014," *JAMA Intern Med.*, vol. 177, no. 3, pp. 326–333, 2017, doi: 10.1001/jamainternmed.2016.8245.

[26] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5967–5976, doi: 10.1109/CVPR.2017.632.

[27] G. Antipov, M. Baccouche, and J.-L. Dugelay, "Face aging with conditional generative adversarial networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2089–2093, doi: 10.1109/ICIP.2017.8296650.

[28] S. Liu, Y. Sun, D. Zhu, R. Bao, W. Wang, X. Shu, and S. Yan, "Face aging with contextual generative adversarial nets," in *Proc. 25th ACM Int. Conf. Multimedia*, Oct. 2017, pp. 82–90, doi: 10.1145/3123266.3123431.

[29] P. Li, Y. Hu, Q. Li, R. He, and Z. Sun, "Global and local consistent age generative adversarial networks," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 1073–1078, doi: 10.1109/ICPR.2018.8545119.

[30] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251, doi: 10.1109/ICCV.2017.244.

[31] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.

[32] X. Yao, G. Puy, A. Newson, Y. Gousseau, and P. Hellier, "High resolution face age editing," 2020, *arXiv:2005.04410*.

[33] W. Wang, Y. Yan, Z. Cui, J. Feng, S. Yan, and N. Sebe, "Recurrent face aging with hierarchical AutoRegressive memory," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 3, pp. 654–668, Mar. 2019, doi: 10.1109/TPAMI.2018.2803166.

[34] J. Despois, F. Flament, and M. Perrot, "AgingMapGAN (AMGAN): High-resolution controllable face aging with spatially-aware conditional GANs," in *Proc. ECCV*, in Lecture Notes in Computer Science, vol. 12537, 2020, pp. 613–628, doi: 10.1007/978-3-030-67070-2_37.

[35] Y. Liu, Q. Li, and Z. Sun, "Attribute-aware face aging with wavelet-based generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11869–11878, doi: 10.1109/CVPR.2019.01215.

[36] H. Fang, W. Deng, Y. Zhong, and J. Hu, "Triple-GAN: Progressive face aging with triple translation loss," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 3500–3509, doi: 10.1109/CVPRW50498.2020.00410.

[37] S. Palsson, E. Agustsson, R. Timofte, and L. Van Gool, "Generative adversarial style transfer networks for face aging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 2165–2173, doi: 10.1109/CVPRW.2018.00282.

[38] D. Deb, D. Aggarwal, and A. K. Jain, "Child face age-progression via deep feature aging," 2020, *arXiv:2003.08788*.

[39] Y. Shen, B. Zhou, P. Luo, and X. Tang, "FaceFeat-GAN: A two-stage approach for identity-preserving face synthesis," 2018, *arXiv:1812.01288*.

[40] J. Song, J. Zhang, L. Gao, X. Liu, and H. T. Shen, "Dual conditional GANs for face aging and rejuvenation," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 899–905.

[41] H. Yang, D. Huang, Y. Wang, and A. K. Jain, "Learning face age progression: A pyramid architecture of GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 31–39, doi: 10.1109/CVPR.2018.00011.

[42] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2813–2821, doi: 10.1109/ICCV.2017.304.

[43] Q. Li, Y. Liu, and Z. Sun, "Age progression and regression with spatial attention modules," 2019, *arXiv:1903.02133*.

[44] H. Zhu, Z. Huang, H. Shan, and J. Zhang, "Look globally, age locally: Face aging with an attention mechanism," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1963–1967.

[45] Z. He, M. Kan, S. Shan, and X. Chen, "S2GAN: Share aging factors across ages and share aging trends among individuals," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9439–9448, doi: 10.1109/ICCV.2019.00953.

[46] C. N. Duong, K. Luu, K. G. Quach, N. Nguyen, E. Patterson, T. D. Bui, and N. Le, "Automatic face aging in videos via deep reinforcement learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 10005–10014, doi: 10.1109/CVPR.2019.01025.

[47] C. Shi, J. Zhang, Y. Yao, Y. Sun, H. Rao, and X. Shu, "CAN-GAN: Conditioned-attention normalized GAN for face age synthesis," *Pattern Recognit. Lett.*, vol. 138, pp. 520–526, Oct. 2020, doi: 10.1016/j.patrec.2020.08.021.

[48] P. Li, H. Huang, Y. Hu, X. Wu, R. He, and Z. Sun, "Hierarchical face aging through disentangled latent characteristics," in *Proc. ECCV*, in Lecture Notes in Computer Science, vol. 12348, pp. 86–101, 2020, doi: 10.1007/978-3-030-58580-8_6.

[49] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. 2nd Int. Conf. Learn. Represent.*, 2014, pp. 1–14.

[50] H. Huang, Z. Li, R. He, Z. Sun, and T. Tan, "Introvae: Introspective variational autoencoders for photographic image synthesis," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 52–63.

[51] R. Or-El, S. Sengupta, O. Fried, E. Shechtman, and I. Kemelmacher-Shlizerman, "Lifespan age transformation synthesis," in *Proc. ECCV*, in Lecture Notes in Computer Science, vol. 12351, 2020, pp. 739–755, doi: 10.1007/978-3-030-58539-6_44.

[52] Z. Huang, S. Chen, J. Zhang, and H. Shan, "PFA-GAN: Progressive face aging with generative adversarial network," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 2031–2045, 2021, doi: 10.1109/TIFS.2020.3047753.

[53] Z. Huang, S. Chen, J. Zhang, and H. Shan, "AgeFlow: Conditional age progression and regression with normalizing flows," in *Proc. 13th Int. Joint Conf. Artif. Intell.*, Aug. 2021, pp. 743–750, doi: 10.24963/ijcai.2021/103.

[54] S. He, W. Liao, M. Ying Yang, Y.-Z. Song, B. Rosenhahn, and T. Xiang, "Disentangled lifespan face synthesis," 2021, *arXiv:2108.02874*.

[55] Z. Huang, J. Zhang, and H. Shan, "When age-invariant face recognition meets face age synthesis: A multi-task learning framework," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7278–7287, doi: 10.1109/cvpr46437.2021.00720.

HADY PRANOTO received the B.S. and M.S. degrees in information management from Bina Nusantara University, Jakarta, Indonesia, in 1998 and 2015, respectively, where he is currently pursuing the Ph.D. degree with the BINUS Graduate Program-Doctor of Computer Science. Since 1998, he has been a Lecturer with Teknik Informatika, Bina Nusantara University, where he has been an Interactive Multimedia Concentration Content Coordinator with the Computer Science Department, School of Computer Science, Bina Nusantara University, since 2016. His research interests include computer vision, machine learning, multimedia, augmented reality, and virtual reality.

YAYA HERYADI received the bachelor's degree in statistics and computation from Bogor Agricultural University (IPB), Bogor, Indonesia, in 1984, the Master of Science degree in computer science from Indiana University at Bloomington, IN, USA, in 1989, and the Ph.D. degree in computer science from Universitas Indonesia, Depok, Indonesia, in 2014. In 2013, he was a Visiting Researcher with Michigan State University, East Lansing, MI, USA. He is currently a Researcher, a Faculty Member, and a Research Coordinator with the Doctor of Computer Science Program-Binus Graduate Program, Bina Nusantara University. His research interests include big data analytics, machine/deep learning, remote sensing data analytics, and natural language processing.

HARCO LESLIE HENDRIC SPITS WARNARS received the bachelor's degree in computer science in the information systems field from STMIK Budi Luhur (http://fti.budiluhur.ac.id/id/), Jakarta Selatan, Indonesia, in 1995, with the title S. Kom (Sarjana Komputer) and the bachelor's thesis topic about information systems using the Budi Luhur Scholarship, the master's degree in computer science with a major field information technology from University Indonesia, in 2006, with degree title Magister Teknologi Informasi (M.T.I.) and master's thesis topic about Datawarehouse, which was scholarship by Budi Luhur University, and the Ph.D. degree in computer science from Manchester Metropolitan University, Manchester, U.K., in 2012, with a Ph.D. thesis topic about data mining.

His Ph.D. degree was funded by the Directorate General of Higher Education, Ministry of Education and Culture, Republic of Indonesia (DIKTI) Scholarship. He is currently the Head of Concentration of Information Systems with the Doctor of Computer Science (DCS), Bina Nusantara University (http://dcs.binus.ac.id) and a Supervisor of some Ph.D. students in computer science. He has been teaching computer science subjects, since 1995. He has been an Indonesian National Lecturer Degree Lektor Kepala (550), since 2007, which is recognized as an Associate Professor. He had awarded some research grants, such as Program of Research Incentive of National Innovation System (SINAS) from the Ministry of Research, Technology and Higher Education of the Republic of Indonesia and Incentives article in the international journal from the directorate of research and community service, Ministry of Research, Technology and Higher Education of the Republic of Indonesia.

Dr. Warnars has been a member of some professional membership, such as The Society of Digital Information and Wireless Communication (SDIWC) member ID:3518 (www.SDIWC.net), since March 2014; the International Association of Engineers (IAENG) member number: 140849 (www.IAENG.org), since April 2014; a Senior Member of the International Association of Computer Science and Information Technology (IACSIT; www.IACSIT.org), since January 2014; and the Institute for Systems and Technologies of Information, Control and Communication (INSTICC) member number 5279, since June 2014. He is active as a reviewer/a program committee for some international journals or conferences and acts as the general chair, the program chair, and the general committee for some international conferences , active as an advisory and the editorial board for six journals. His review activities can be seen at (https://publons.com/researcher/1620795/harco-leslie-hendric-spits-warnars/metrics/). He had success running as the General Chair for two international conferences, such as 2017 cyberneticsCOM and 2018 Indonesian Association for Pattern Recognition International (INAPR) Conference. He is the Technical Program Chair for 2019 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE). His research publications can be reached at (https://www.researchgate.net/profile/Harco_Leslie_Hendric_Spits_Warnars2 Or https://scholar.google.co.id/citations?user=pplO3mEAAAAJ&hl=id or https://www.scopus.com/authid/detail.uri?authorId=57219696428).

WIDODO BUDIHARTO received the bachelor's degree in physics from the University of Indonesia, Jakarta, Indonesia, the master's degree in information technology from STT Benarif, Jakarta, Indonesia, and the Ph.D. degree in electrical engineering from the Institute of Technology Sepuluh Nopember, Surabaya, Indonesia. He took the Ph.D. Sandwich Program in robotics with Kumamoto University, Japan, and conducted Postdoctoral Researcher work in robotics and artificial intelligence with Hosei University, Japan. He worked as a Visiting Professor with the Erasmus Mundus French Indonesian Consortium (FICEM), France, Hosei University, and the Erasmus Mundus Scholar with the EU Universite de Bourgogne, France, in 2017, 2016, and 2007, respectively. He is currently a Professor in artificial intelligence with the School of Computer Science, Bina Nusantara University, Jakarta. His research interests include intelligent systems, data science, robot vision, and computational intelligence.