



# ChildGAN: Face aging and rejuvenation to find missing children

Praveen Kumar Chandaliya\*, Neeta Nain

*Malaviya National Institute of Technology Jaipur, India*



## ARTICLE INFO

### Article history:

Received 20 April 2021

Revised 25 April 2022

Accepted 28 April 2022

Available online 30 April 2022

### Keywords:

Child face aging and rejuvenation

Child datasets

Face recognition

Age estimation

Gender preservation

Child trafficking

## ABSTRACT

Child-face aging and rejuvenation have amassed considerable active research interest, owing to their immense impact on a broad range of social and security applications, e.g., digital entertainment, fashion and wellness, and searching for long-lost children using childhood photos. All current face aging approaches based on generative adversarial networks (GANs) focus on adult images or long-term aging. We present a new large-scale longitudinal Indian child (ICD) benchmark dataset to facilitate face age progression and regression, cross-age face recognition, age estimation, gender prediction, and kinship face recognition to alleviate these issues. Furthermore, we propose an automatic child-face age progression and regression model, namely, ChildGAN, that generates visually realistic images for enhanced face-identification accuracy while preserving the identity. Consequently, we have trained state-of-the-art (SOTA) face aging models on ICD for comprehensive qualitative and quantitative evaluations. We also present a multi-racial experiments dataset named Multi-Racial Child Dataset (MRCD) containing 64,965 child face images. The images are selected from publicly available datasets and web crawling. Finally, we investigate the generalization of ChildGAN by experimenting with White, Black, Asian, and Indian races. The experimental results suggest that the proposed ChildGAN and SOTA models can aid in reconnecting young children, who were lost at a young age as victims of child trafficking or abduction, with their families. The model and the MRCD web crawled images are available at [https://github.com/praveenkumarchandaliya/ChildGAN\\_Tamp1](https://github.com/praveenkumarchandaliya/ChildGAN_Tamp1).

© 2022 Elsevier Ltd. All rights reserved.

## 1. Introduction

According to a global report on human trafficking, most trafficking victims are women and children<sup>1,2</sup>. In 2016, 30% of the trafficked victims were children. The National Center for Missing and Exploited Children (NCMEC) reported that approximately 800,000 children disappear in the USA every year. The National Crime Records Bureau in India reported the total amount of human trafficking as 88,008, 6.9% more than the 2018 year. On average, the victim ratio is 1 : 6 for boys to girls [1]. Our work can help law-enforcement agencies to trace missing children because we use face images as a tool to identify long-lost children.

The face is usually the only biometric available to trace missing children, as parents and relatives are more likely to have images of the child, rather than other biometrics, such as fingerprints or

irises. Two girls,<sup>3</sup> lost at the ages of 3 and 6 in Delhi, India, were reunited with their family by the Child Welfare Committee, Kanpur at the ages of 9 and 12 (as shown in Fig. 1).

A boy<sup>4</sup> who was just two years old, when his mother ran off with him to Mexico, after attempting to kill his father with a milkshake laced with pesticide. Eight years later, she was arrested and returned to US authorities; however, the boy was not with her. Fig. 2 shows the age-progressed image of the 10-year-old boy (middle), created from a photo of him as a two-year-old, which is quite similar to the actual photo taken when he was found (right), primarily around the mouth.

A large number of child trafficking victims are very young. As time passes, their faces may appear to be quite different from the photos provided by their parents or relatives, and many times the face images are obstructed or blurry. Consequently, the face transfiguration makes it difficult to find the victims. Thus, our work can help law-enforcement agencies to trace missing children.

\* Corresponding author.

E-mail addresses: [\(P.K. Chandaliya\), \[\\(N. Nain\\).\]\(mailto:nain.cse@mnit.ac.in\)](mailto:2016rcp9511@mnit.ac.in)

<sup>1</sup> A human being below the age of 18 years unless, under the law applicable to the child, the majority has been attained earlier; <https://www.unicef-irc.org/portfolios/crc-optional-2.html>.

<sup>2</sup> <https://www.unodc.org>

<sup>3</sup> <https://www.haqcrc.org/news/six-year-journey-family-search-missing-daughters-kids/>

<sup>4</sup> <https://boingoing.net/2009/10/20/how-forensics-use-ph.html>



**Fig. 1.** Two missing girls. (a) Age 3 and age 6 with their parents before their disappearance. (b) After 6 years, (age 12) and (age 9) were reunited with their mother and uncle. To ensure data privacy, we intentionally cover the eyes.



**Fig. 2.** (Left) Two-year-old boy at the time of his disappearance; (middle) age-progressed image of the boy as a 10-year-old; (right) actual image of 10-year-old boy. To ensure data privacy, we intentionally cover the eyes.

In this study, the proposed model is based on a variational auto-encoder (VAE) [2] with a generative adversarial network (GAN) [3], which can simultaneously age and rejuvenate the face in the image space. The following are our contributions:

1. We present a large-scale longitudinal Indian child dataset (ICD) to address face aging, cross-age face recognition, age estimation, and gender preservation in children.
2. ChildGAN, a new method for age-progression and rejuvenation, which can automatically generate visually realistic face photos, while attaining enhanced face-recognition, age-estimation, and gender-preservation rates, compared to state-of-the-art methods.
3. We also investigate the generalization of ChildGAN by presenting Multi-Racial Child Dataset (MRCD) containing 64,965 face images of four races (Asian, Black, White, and Indian).
4. An encoder and discriminator architecture, inspired by the self-attention GAN, which exhibits a better balance between the ability to model long-range dependencies and computational and statistical efficiency.

ChildGAN simultaneously ages and rejuvenates child faces based on the given age and gender. We conducted extensive experiments using ICD, with low-resolution faces and obstructions like caps or scarves, on images of real missing children. We also investigate the generalization of ChildGAN using MRCD, on Asian, Black, White, and Indian race. A similarity comparison with their parents outperformed state-of-the-art techniques, in terms of the qualitative and quantitative effects of aging in children.

## 2. Related work

Several computerized age progression and regression approaches, from anthropological theories to deep generative models, have been presented by the research community. However,

the majority of existing approaches have aimed to explore aging in adult faces and ignored children's faces. In this section, related work is organized into three subsections. [Section 2.1](#) briefly reviews the GAN and its variants for image generation. In [Section 2.2](#), we enumerate the age-progression and regression-based work. [Section 2.3](#) discusses the available face-aging datasets.

### 2.1. Generative model

Recently, there have been important developments in generative models [2–5]. Among them, two encouraging ones stand out by generating photo-realistic images. The first is a variational autoencoder (VAE) [2], which controls the distribution of the latent vector  $z$  by using the concept of variational inference. The variational inference between two distributions is measured by the  $KL$  divergence, which measures the dissimilarity between two distributions  $KL(p||q)$ . VAEs are limited by their element-wise reconstruction loss function, which causes blurriness in images.

The second is the GAN approach [3], which comprises generator ( $G$ ) and discriminator ( $D$ ) networks, which are simultaneously trained based on the zero-sum game framework. GAN-generated images are noisy and incomprehensible. Radford et al. [6] improved the GAN's network architecture by introducing stride convolutions and deconvolutions. By incorporating conditions into the generation process, the conditional GAN (CGAN) [7] further enlarged the application scope. However, the GAN still suffers from unstable training and mode-collapse problems.

Single-image super-resolution-based models, e.g., SRGAN [8], generate good results. However, the loss function based on the feature space in SRGAN objectively sacrifices the high peak signal-to-noise ratio (PSNR). In addition, WGAN [4] and WGAN-GP [9] improved the training techniques by adding a Wasserstein (earth mover) distance-metric loss function to the DCGAN architecture to address the generator and discriminator training problems and the vanishing-gradient problem. They used the WGAN-GP objective, which stabilizes the training over a wide range of architectures with almost no hyperparameter tuning. Therefore, instead of applying weight-clipping, WGAN-GP penalizes the model if the gradient norm moves away from its target norm value of one.

### 2.2. Age progression and regression

Face-aging and de-aging methods can be technically classified into two categories: (i) conventional statistical learning-based approaches and (ii) GAN-based approaches.

#### 2.2.1. Conventional statistical learning-based approaches

These approaches can be further divided into physical model-based and prototyping model-based approaches. The physical approach represents the faces in each age group by a three-level *and-*

or graph [10]. The challenge of this approach is collecting face images of the same person over a long time, and age-related variations are often mixed with other variations, e.g., facial-growth parameters [11].

The prototype face approach creates a set of images for different age groups with an average face [12], i.e., a low-rank subspace [12] of each category, as its prototype. The prototype input face image is translated into another face image of the target age category. The main drawback of this approach is that a single prototype is designed for each age group; therefore, a similar aging pattern is applied across time, and no distinct face-aging patterns are observed in the different age groups because it produces similar aging faces.

### 2.2.2. GAN-based approaches

Simulating face-aging and de-aging processes is an arduous task, owing to the large and nonlinear deviations present in the infant, toddler, youth, adult, senior, and old-age stages of facial growth. GANs are better equipped to specifically address the essential problems and challenges present in the first approach. GAN-based models have shown higher visual-fidelity results in facial aging.

Antipov et al. [13] proposed the age-cGAN method to achieve high-quality face-aging results that retain the person's identity. Zhang et al. [14] proposed a conditional adversarial autoencoder (CAAE) approach for age progression and regression, based on the face-manifold learning concept, and added the age as a conditional. Jain et al. [15] applied a pyramid adversarial discriminator with multiple scales and added an identity-preserving objective. Although these methods show remarkable improvements, their results are over-smooth, compared with real images.

Li et al. [16] proposed a global and local consistent-age GAN to train global and local features by introducing a residual face. Same author group proposed a wavelet-based GAN to address the matching ambiguity between young and aged face images inherent in unpaired datasets. Wang et al. [17] proposed an identity-preserving conditional GAN that aimed to synthesize a face according to the age group, incorporating AlexNet-based perceptual-loss and age-classification terms into the objective function. Zhao et al. [18] proposed an age-invariant model (AIM) for cross-age face rejuvenation/aging and recognition.

Zhu et al. [30] introduced an attention conditional GAN (AcGAN) model for altering the regions particularly relevant to face aging. Chandaliya et al. [31] proposed a conditional perceptual adversarial variational autoencoder (CPAVAE), and introduced the VGG19 [38] perceptual term into the objective function to preserve the identity and texture features.

However, the state-of-the-art techniques shown in Table 1 do not address child-specific or short-span aging. They consider a facial image sequence over a long age span for each subject.

Considering age progression and regression for finding missing children, the face images are frequently dark or lightly illuminated, have low resolution, or are obstructed by scarves or caps. If the obstruction covers the child's face, the image cannot be used to properly progress and regress using state-of-the-art models. These challenges motivated us to create a child-specific dataset for training. Furthermore, we propose ChildGAN, with an identity-preservation perceptual-loss function to boost the performance in the cases of low-resolution and obstructed faces.

### 2.3. Face-aging datasets

Prior studies on child face aging are limited, owing to the lack of child-face datasets. Table 2 summarizes the face datasets that include children.

## 3. Proposed model architecture

Fig. 3 illustrates the ChildGAN model. It consists of three major components: (a) Encoder ( $E$ ), (b) Generator ( $G$ ), and (c) Attention Discriminator ( $AD_{img}$ ).

**Encoder ( $E$ ):** The encoder network encodes input image  $x$  into identity vector  $z$ , using a stack of four convolution layers. Each layer uses a filter of size four, stride two, with padding size one, and a ReLU activation function. With each layer, the image size decreases by a factor of two, while the number of feature maps sequentially increases by two. A residual network, with nine residual blocks containing two  $3 \times 3$  convolution layers with batch normalization, and a self-attention block are incorporated after convolution blocks, respectively. A residual block is employed in the encoder to simulate the aging effect between age groups. The self-attention block (SAB) is inspired by the self-attention generative adversarial network (SAGAN) [41] reciprocal to convolutions, which helps with modeling long-range, multi-level dependencies across face-image regions [42].

As shown in Fig. 3, residual learning adds attention blocks to establish a relationship between the shallow and deep layers, which captures the face shape and texture information. In addition, the back-propagation path is shortened during training; hence, the network converges more easily. We obtain a  $512 \times 8 \times 8$  tensor, which is then converted into a fully connected layer with 32,768 features. Further, this vector is converted into a vector with 50 size vectors, each having a mean ( $\mu$ ) and covariance ( $\sigma$ ). As shown in Fig. 3, the latent vector  $z$  is sampled from  $\mu$  and  $\sigma$ , i.e.,  $z \sim N(\mu, \sigma)$  using a reparameterization trick, i.e.,  $z = \mu + \epsilon \otimes \sigma$ , where  $\epsilon \sim N(0, I)$ .

Meanwhile, the age (one-hot vector of five age groups) $\times 10$  and gender (one-hot vector of length one) $\times 25$  are combined to form a vector  $c$  is of size 75. This vector  $c$  is further appended with vector  $z$  of size 50, resulting in a vector of length 125 represented by  $zc$  (75). Consequently, while changing age ( $l$ ), we can progress and regress face images while preserving their identity features [14].

**Generator ( $G$ ):** The generator  $G$  acts as a decoder, corresponding to the deep-convolution encoder  $E$ . As shown in Fig. 3, the input of  $G$  is a concatenated vector of identity vector  $z$ , and condition vector  $c$ ; i.e.,  $zc$ , where it is converted into a 32,768 feature vector and reshaped to a tensor of size  $512 \times 8 \times 8$ . The remainder of the network is a stack of four nearest-neighbor convolution layers, where the tensor is up-sampled by a factor of two, and the convolution layer is applied with a filter size three, stride one, and padding one with a ReLU activation function. We use the tanh activation function to ensure that the generated images  $\bar{x}$  are in the range  $[-1, 1]$ .

**Attention discriminator ( $AD_{img}$ ):**  $AD_{img}$  is a stack of four stride-Convolution-ReLU-InstanceNorm layers, which downsize images by a factor of two and increase the number of channels by a factor of two. It is followed by a self-attention block, after one fully connected layer, and further followed by a ReLU and Sigmoid at the final output.  $AD_{img}$  forces  $G$  to generate a realistic face by adversarial learning. In addition, the one-hot condition vector  $c$  are encoded using fractionally-strided convolutions and concatenated at the second convolutional layer of  $AD_{img}$  (see in Fig. 3). This allows training interactions with  $x$  and  $\bar{x}$ , while discriminating among the generated faces consistent with the conditions. Furthermore, we incorporate a self-attention-based mechanism to help model long-range, multi-level dependencies across face image regions [42].  $x$  and  $\bar{x}$  are both passed into  $AD_{img}$  to calculate the discrepancy between the reconstructed and input images.

## 4. Loss functions

**Adversarial loss:** The generator  $G$  and discriminator  $AD_{img}$  stimulates generated results to be indistinguishable from real images.

**Table 1**  
Comparison between our model and existing state-of-the-art face-aging models.

GAN Model	Age Groups	Dataset	Evaluation Protocol
CAAE [14]	10 AG: 0–5, 6–10, 11–15, 16–20, 21–30, 31–40, 41–50, 51–60, 61–70, 71–80.	Morph [19], CACD [20], UTKFace [14] T: 13,600	Perceptual human study.
Age-cGAN [13]	6 AG: 0–18, 19–29, 30–39, 40–49, 50–59, 60 +	IMDBWiki [21] T: 120K	OpenFace using FR with age gap > 20 years.
PAG-GAN [15]	4 AG: 14–30, 31–40, 41–50, 51–60	Morph [19], CACD [20], FGNET [22] T: 44,029	FR and AE using Face+ [23].
C-GAN [24]	7 AG: 0–10, 11–18, 19–29, 30–39, 40–49, 50–59, 60+	CACD [20], FGNET [22], Morph [19], Adience [25], IMDB [21], LFW [27] T: 19077	Perceptual human study; Center loss-based [26] FR with age gap > 20 years.
GLCAGAN [16]	4 AG: 14–30, 31–40, 41–50, 51–62	Morph [19], CACD [20], FGNET [22] T: 153,106	Perceptual human study; [26] LightCNN-29 FIP.
IPCGAN [17]	5 AG: 11–20, 21–30, 31–40, 41–50, 51+	CACD [20] T: 146,794	Perceptual human study for AE & FR, IS [5] and VGG-Face [28] score
CANGAN [29]	4 AG: 14–30, 31–40, 41–50, 51+	MORPH [19] T: 35,486 CACD [20] T: 121,027	FR and AE using Face+ [23].
AcGAN [30]	5 AG: 10–20, 21–30, 31–40, 41–50, 51+	Morph [19] T: 54,539	FR and AE using Face+ [23].
CPAVAE [31]	4 AG: 0–5, 6–10, 11–15, 16–20	CLF [32] T: 10,752	IS [5] and FID [33] score
<b>ChildGAN</b>	5 AG: 2–5, 6–8, 9–11, 12–14, 15–19	Indian Child Dataset (ICD) T: 35,484 Multi-Racial Child Dataset	FAR on Children; FR & FI using FaceNet [34], PFE [35] & ArcFace [36];
	5 AG: 0–3, 4–8, 9–12, 13–16, 17–20	Asian: 17,211, Black: 13,354 White: 19,297, Indian: 15,103	AE & GE using SSRNet [37] GS on four races.

AG: Age group, T: Total images, FR: Face recognition, FI: Face identification, AE: Age estimation, FIP: Face-identity preservation, FAR: Face aging and rejuvenation, GE: Gender estimation, IS: Inception score, FID: Fréchet inception distance, and GS: Generalization Study.

**Table 2**  
Child face image datasets.

Dataset	Subjects	Images	Images/Subject	Age Range	Avg. Age	Gender (M: F)	Public
FGNet [22]	82	1002	6–18	0–69	16	415: 295	Yes
Adience [25]	2958	26,580	NA	0–60	NA	3,341: 4195	Yes
MorphII [19]	13,000	55,134	2–53 (avg. 4.2)	16–77	42	6,265: 1,105	Yes
CACD [20]	2000	163,446	22–139 (avg. 81.7)	16–62	31	1,928: 4,337	Yes
IMDBWiki [21]	20,284	523,051	avg. 25.79	0–100	38	12,778: 19,074	Yes
UTKFace [14]	NA	23,709	NA	0–116	33	2,114: 2,552	Yes
AgeDB [39]	568	16,488	avg. 29	1–101	50.3	229: 313	Yes
ITWCC [40]	754	7990	3–37 (avg. 10.7)	0–32	13	3,067: 4,444	Yes
CLF [32]	919	3682	2–6 (avg. 4.0)	2–18	8	6,471: 3,792	No
<b>ICD</b>	16,969	35,484	2–6 (avg. 2.7)	2–19	10	19,184: 16,299	No
<b>MRCD</b>	NA	64,965	NA	0–20	NA	10,068: 6063	Yes

Besides, generation fidelity and attribute immutability are also guaranteed by involving the attribute of input image as a one-hot conditional vector in improved adversarial training. To accomplish these two goals, we define the losses for training  $AD_{img}$  (see in Eq. (1)) and  $G$  (see in Eq. (2)) as:

$$\begin{aligned} \mathcal{L}_{AD_{img}} = & -\mathbb{E}_{x,c \sim P_{data}(x,c)} [AD_{img}(x,c)] \\ & + \mathbb{E}_{x,c \sim P_{data}(x,c)} [(AD_{img}(G(E(x),c)))] \\ & + \lambda_{gp} \mathbb{E}_{\hat{x} \sim P_{\hat{x}}} [(\|\nabla_{\hat{x}} AD_{img}(\hat{x})\|_2 - 1)^2] \end{aligned} \quad (1)$$

where  $\hat{x} \sim P_{\hat{x}}$  is uniformly sampled along straight lines between pairs of input ( $x$ ) and generated images  $G(E(x), c)$ ,  $\nabla$  is the gradient operator,  $P_{data}$  is the real data distribution,  $c$  is the concatenated conditional vector of age ( $I$ ) and (g),  $\lambda_{gp}$  is the penalty coefficient to penalize the gradients.

$$\mathcal{L}_G = -\mathbb{E}_{x,c \sim P_{data}(x,c)} [(AD_{img}(G(E(x),c)))] \quad (2)$$

**Perceptual loss:** The deep-feature consistency principle [43] is applied to ensure that VAE, with adversarial learning (output) and its corresponding input images, has a similar deep feature with a much clearer and more natural nose, eyes, teeth, and hair texture, with much fewer artifacts, even on obstructed and low-resolution

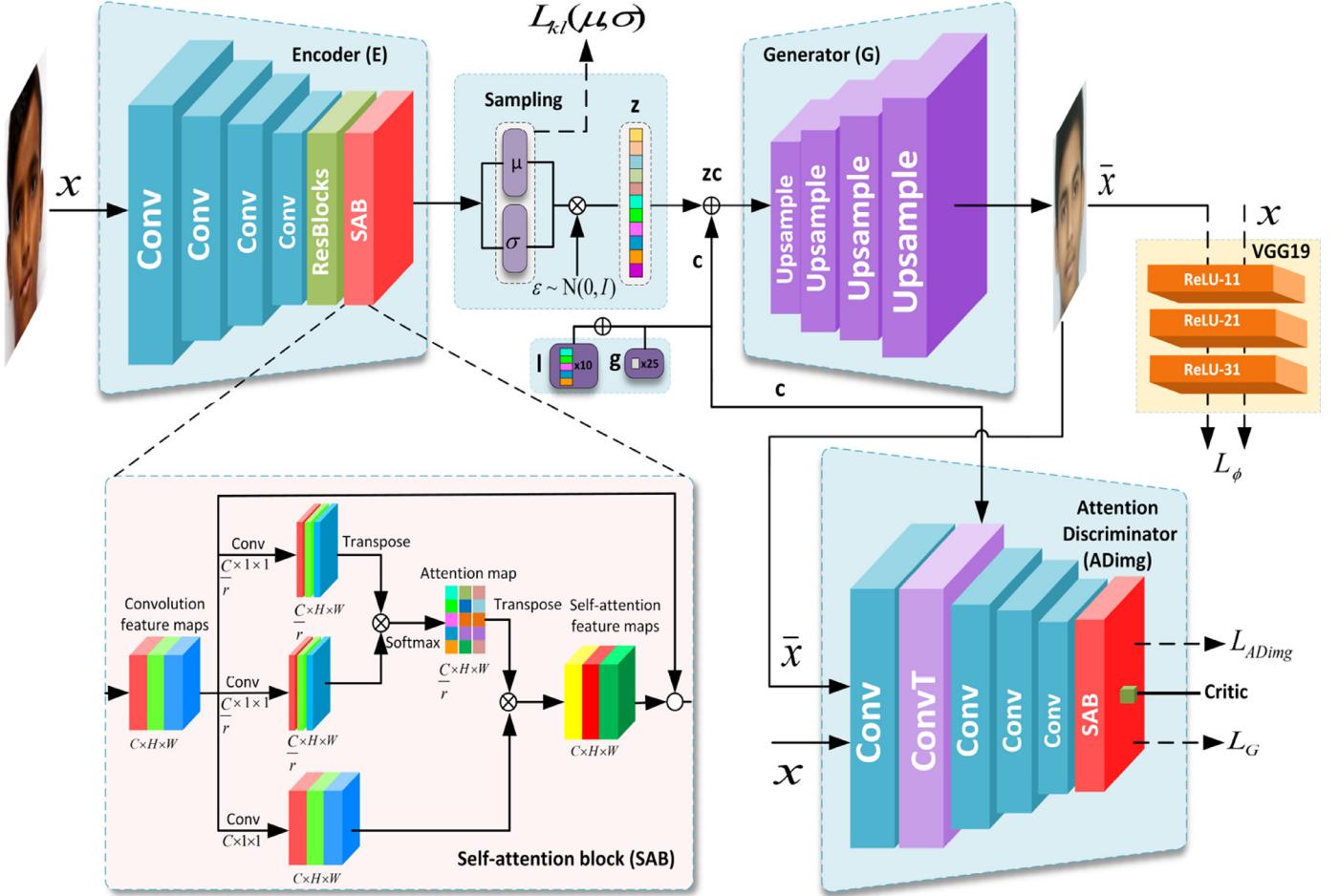
faces. To obtain the facial features with identity preservation, we introduce the perceptual loss using low-level feature maps (i.e., features are extracted from ‘ReLU-11’, ‘ReLU-12’, and ‘ReLU-13’) of the pre-trained VGG19 [38] network. It also effectively removes the checkerboard and ghosting artifacts.

Given a VGG19 network, the perceptual loss defined by the feature map extracted from layer ReLU- $\Upsilon$  can be computed as follows:

$$\mathcal{L}_{\Phi} = \frac{1}{C_{\Upsilon} W_{\Upsilon} H_{\Upsilon}} \|(\Phi_{\Upsilon}(x) - \Phi_{\Upsilon}(\bar{x}))\|_2^2 \quad (3)$$

where  $\Phi_{\Upsilon}(x)$  is a feature map of dimension  $C_{\Upsilon} W_{\Upsilon} H_{\Upsilon}$  for the input image  $x$ .  $\Phi_{\Upsilon}(\bar{x})$  is the feature map for the generated image  $\bar{x}$ . Here,  $\Upsilon$  denote above mentioned ReLU layers of VGG19.

**Kullback-Leibler divergence loss:** We add  $\mathcal{L}_{kl}(\mu, \sigma)$  to regularize the  $z$  vector with an approximate prior  $P(z) \sim N(0, I)$ . For input image ( $x$ ), the  $E$  network output the mean ( $\mu$ ) and covariance ( $\sigma$ ). To calculate the  $P(z|x)$ , we sample a random  $\epsilon$  gaussian distribution, where  $\epsilon \sim N(0, I)$ . Similar to VAE [2], we sample the feature vector using  $z = \mu + \sigma \otimes \epsilon$ , where  $\otimes$  represent the element-wise



**Fig. 3.** ChildGAN has three major components: an encoder  $E$ , generator  $G$ , and attention discriminator  $AD_{img}$ . The model is trained under the  $\mathcal{L}_{kl}(\mu, \sigma)$  represent the KL loss,  $\mathcal{L}_\phi$  represent the perceptual loss,  $\mathcal{L}_G$  and  $\mathcal{L}_{AD_{img}}$  represent generator and discriminator adversarial loss functions. For simplicity, we have omitted the total variation loss  $\mathcal{L}_{tv}$ .  $r$  is a constant value to reduce the number of channels, which we set it equals with 8 just like SAGAN [41].  $\otimes$  denotes element-wise matrix multiplication,  $\oplus$  represents concatenation and  $\circ$  represent element-wise summation.

multiplication.  $\mathcal{L}_{kl}(\mu, \sigma)$  is shown in Eq. (4).

$$\mathcal{L}_{kl}(\mu, \sigma) = -\frac{1}{2} \sum_k (1 + \log(\sigma_k) - \mu_k^2 - \sigma_k) \quad (4)$$

where  $k$  denotes indexes over the dimensions of latent vector. it also use to reduce the gap between the distribution modeled by the  $E$  i.e.,  $P(z|x)$  and prior distribution i.e.,  $P(z)$ .

**Total variational loss:** The total variational ( $\mathcal{L}_{tv}$ ) is used to ensure measurable continuity and smoothness in the generated image to avoid noisy and overly pixelated results. The  $\mathcal{L}_{tv}$  is the sum of the absolute differences for adjacent pixel values in the generated image  $\bar{x}$ . Eq. (5) for the  $\mathcal{L}_{tv}$  loss calculation is:

$$\mathcal{L}_{tv} = \sum_{c=1}^C \sum_{w=1}^W \sum_{h=1}^H |(\bar{x})_{w+1,h} - (\bar{x})_{w,h}|^2 + |(\bar{x})_{w,h+1} - (\bar{x})_{w,h}|^2 \quad (5)$$

**Full objective functions:** Finally, the goal of ChildGAN is to minimize the objective function for the  $AD_{img}$  and  $E$  as well as  $G$  which are shown in Eqs. (6) and (7), respectively, as:

$$\min_{||AD_{img}||_L \leq 1} \mathcal{L}_{AD_{img}} = \lambda_{adv1} \mathcal{L}_{AD_{img}} \quad (6)$$

where  $||AD_{img}||_L \leq 1$  represents the set of 1-Lipschitz [9] constraint on  $AD_{img}$ .

$$\min_{E,G} \mathcal{L}_{EG} = \lambda_p \mathcal{L}_\phi + \lambda_{kl} \mathcal{L}_{kl}(\mu, \sigma) + \lambda_{adv2} \mathcal{L}_G + \lambda_{tv} \mathcal{L}_{tv} \quad (7)$$

In our experiments, we used  $\lambda_p$ ,  $\lambda_{kl}$ ,  $\lambda_{tv}$ , and  $\lambda_{adv1}$  and  $\lambda_{adv2}$  as hyperparameters, which are set as 1.0, 1.0, 1.0, 1.0, and 0.0001, respectively.

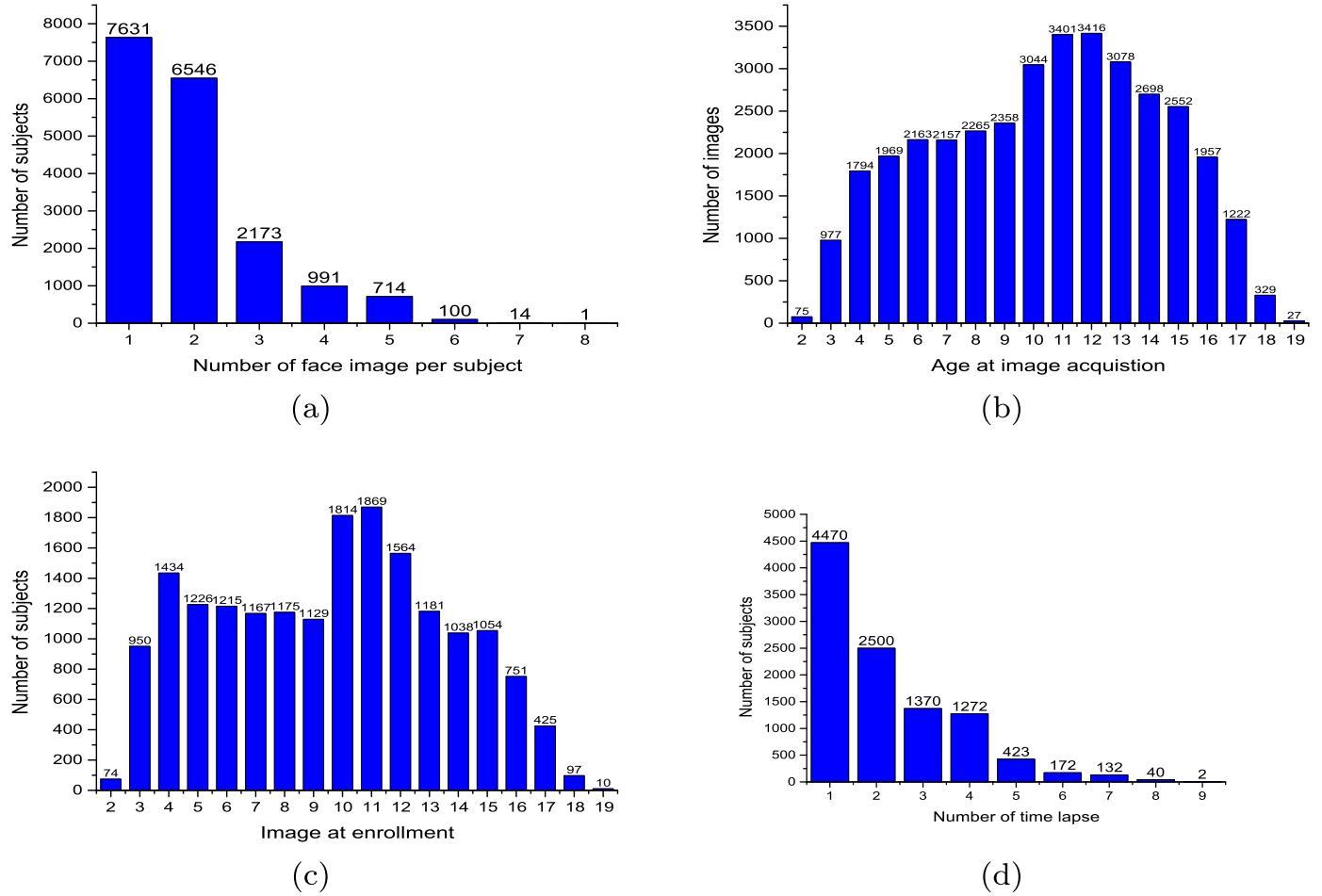
## 5. Child dataset

Previous state-of-the-art models CAAE [14], PAG-GAN [15], IPC-GAN [17], AgeGAN [13], C-GANs [24], and GLCA-GAN [16] on the facial aging and de-aging of children are limited by the lack of publicly available face datasets that include children. To alleviate this, we present two large-scale “Indian Child Dataset (ICD)” and “Multi-Racial Child Dataset (MRCD)”.

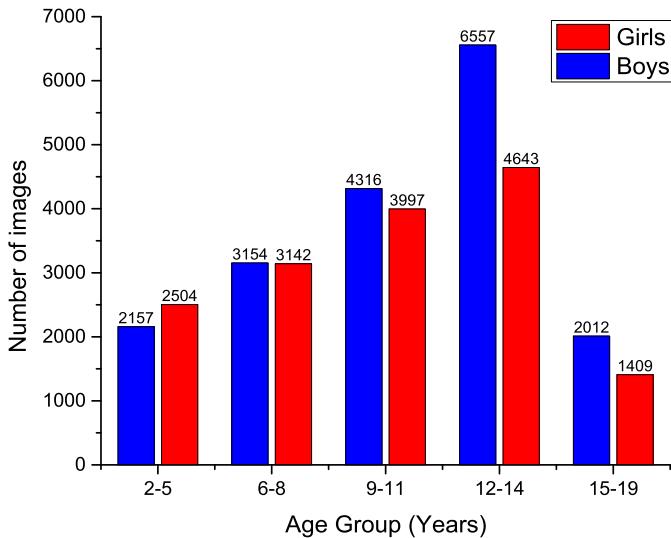
### 5.1. ICD dataset

ICD contains 35,484 child face images from 9475 paired subjects and 7494 single (not paired) subjects, in the age group of [2 : 19] years, annotated with age and gender. By paired images, we mean longitudinal images of the same subject at different ages. Single images comprise only one image (for a subject). The demographic makeup of the ICD dataset consists of 19,184(54%) boys and 16,299(46%) girls. The dataset statistics are shown in Fig. 4.

For training our model and state-of-the-art face-aging models, we used 33,891 images. Chellappa et al. [11,44] proposed a craniofacial growth and face-recognition model over the years. Inspired by this study, we divided the training data into five non-



**Fig. 4.** ICD statistics: (a) number of face images per subject; (b) age at image acquisition in years; (c) enrollment age of a subject in years; (d) time span between enrollment and latest image acquisition for each subject in years. (Zoom in for a better view.)



**Fig. 5.** ICD training dataset group formation shows the number of child face images in each age group.

overlapping age groups, i.e., [2–5], [6–8], [9–11], [12–14], and [15–19], as shown in Fig. 5. The number of training images corresponding to the age groups are 4661, 6296, 8313, 11200, and 3421, respectively. We mainly used ICD for training.

For testing the face verification, identification, age estimation, and gender preservation, we created a separate dataset, namely, the ICD test dataset, which includes 277 pairs of the youngest and oldest images of the same subject with a time lapse of [5–8] years.

To inspect the robustness of ChildGAN, we also experimented on the publicly available UTKFace [14] and FG-NET [22] datasets. UTKFace contains 23k+ face images in the wild, with ages ranging from [0–117] years old. For age estimation, gender preservation, and image-quality performance, we selected 328 images in the age range [3–19] with Indian ethnicity. FGNET [22] contains 1,002 face images from 82 subjects, with ages ranging from 0 to 69 years old. To evaluate the performance, we selected six subjects in the [3–19] age range with Indian ethnicity.

## 5.2. MRCD dataset

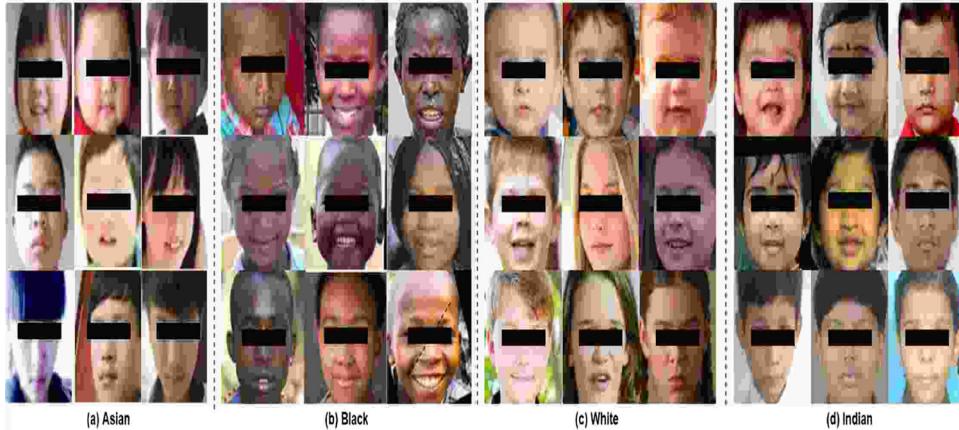
We present Multi-Racial Child Dataset (MRCD) containing 64,965 face images consisting of four race groups (Asian, Black, White, and Indian). These images are selected from publicly available datasets CACD [20], MORPH-II [19], FFHQ [45], UTKFace [14], IMDB [21], MegaAsian [46], and web crawling.

CACD, FFHQ, IMDB, and MegaAsian datasets; some of these datasets do not contain all the required annotated attributes like age, race or gender etc. Thus, we estimated such missing attributes using newly released commercial-off-the-shelf (COTS) FaceVACS-DBScan ID 5.6. After collecting images in the age range [0–20] from these public datasets, we observed that Black and Asian ethnicity in the age range [0–12] are very few. To alleviate this is-

**Table 3**  
MRCD: Multi Racial Child Faces in-the-wild dataset statistics.

Dataset	Images	Age Range	Annotated Attribute	Estimated Attribute	Race/Ethnicity				Age Range [0 – 20]
					Asian	Black	White	Indian	
CACD	163,446	16–62	A,G	R	NA	NA	6137	NA	6137
MorphII	55,134	16–77	A,G,R	NA	NA	3330	NA	NA	3330
FFHQ	70,000	NA	NA	A,G,R	NA	3,996	1322	7934	13,252
UTKFace	23,709	0–116	A,G,R	NA	1,197	410	2251	716	4,574
IMDB	461,871	0–100	A,G	R	NA	NA	1482	NA	1,482
MegaAsian	41,941	0–70	A	G,R	6,641	NA	NA	NA	6,641
ICD	35,484	2–19	A,G,R	NA	NA	NA	NA	13,561	13,561
Web crawling	16,168	NA	NA	A,G,R	5557	8292	1493	826	16168
MRCD	64,965	0–20	A,G,R	NA	17,211	13,354	19,297	15,103	64,965

[\*] A: Age, G: Gender, R: Race, NA: Not Applicable



**Fig. 6.** Exemplar blurred face images of MRCD dataset. (a–d) Asian, Black, White, Indian race. To ensure data privacy, we intentionally degrade the quality of the exemplar facial images.

sue, we further collect such images using web crawling with key words as: ‘African child’, ‘East Asian child’ etc. In addition, we have manually refined and annotated gender, ethnicity, and age. **Table 3** shows the detailed statistics of the MRCD dataset and **Fig. 6** shows some sample images from MRCD. From MRCD dataset, we create independent training and test sets according to the four races and divide them into 5 age groups as [0–3], [4–5], [9–12], [13–16], and [17–20]. Asian race contains 17,211 face images with training and test images as 16,350 and 861, respectively. Black race contains 13,354 face images and similarly as Asian race, with training and test images as 12,686 and 668, respectively. White race contains 19,297 face images with training and test images as 18,332 and 965, respectively. Indian ethnicity contains 15,103 face image with training and test images as 14,348 and 755, respectively.

## 6. Experimental setup

In this section, we first introduce the preprocessing of the ICD, MRCD, UTKFace, and FGNET datasets and then present our model implementation details.

### 6.1. Preprocessing of dataset

For data preprocessing, we used a multitask cascaded convolutional network (MTCNN) [47] to detect the five landmark points (two eyes, the nose, and two mouth corners). Via a similarity transformation, the face images are properly aligned and cropped to a resolution of  $128 \times 128$  pixels. Before being passed as input, each pixel in the RGB image is normalized by subtracting the mean of 127.5 and dividing by the variance of 128, as shown in **Fig. 7**. The ICD test dataset images were resized (as per the model requirement) to  $160 \times 160$ ,  $96 \times 112$ , and  $112 \times 112$  for FaceNet [34],

probabilistic face embeddings (PFE) [35], and ArcFace [36], respectively.

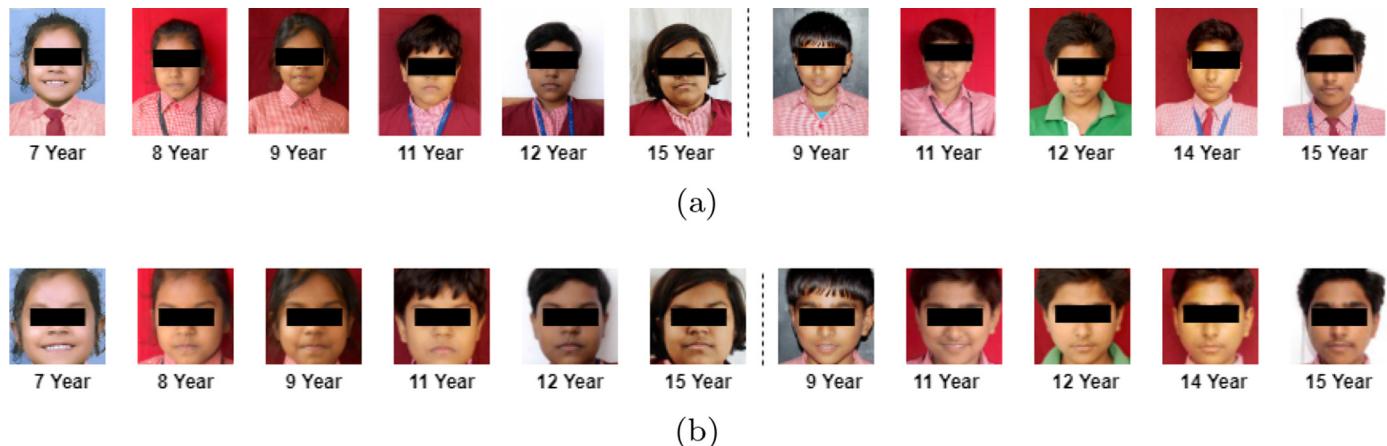
### 6.2. ChildGAN training

For ChildGAN, all components are trained with a batch size of 32 using the Adam optimizer with hyperparameters  $\alpha = 0.0001$  and  $\beta = (0.5, 0.999)$ . The model was trained from scratch with a learning rate of  $E$  at 0.0001, and for  $G$  and  $AD_{imag}$  at 0.0002. We apply a perceptual-level critic after every five iterations, and  $G$  is updated at every iteration. After 20,000 iterations, we were able to achieve competent results. For testing, only encoder  $E$  and generator  $G$  are needed to generate age-progressed and regressed face images.

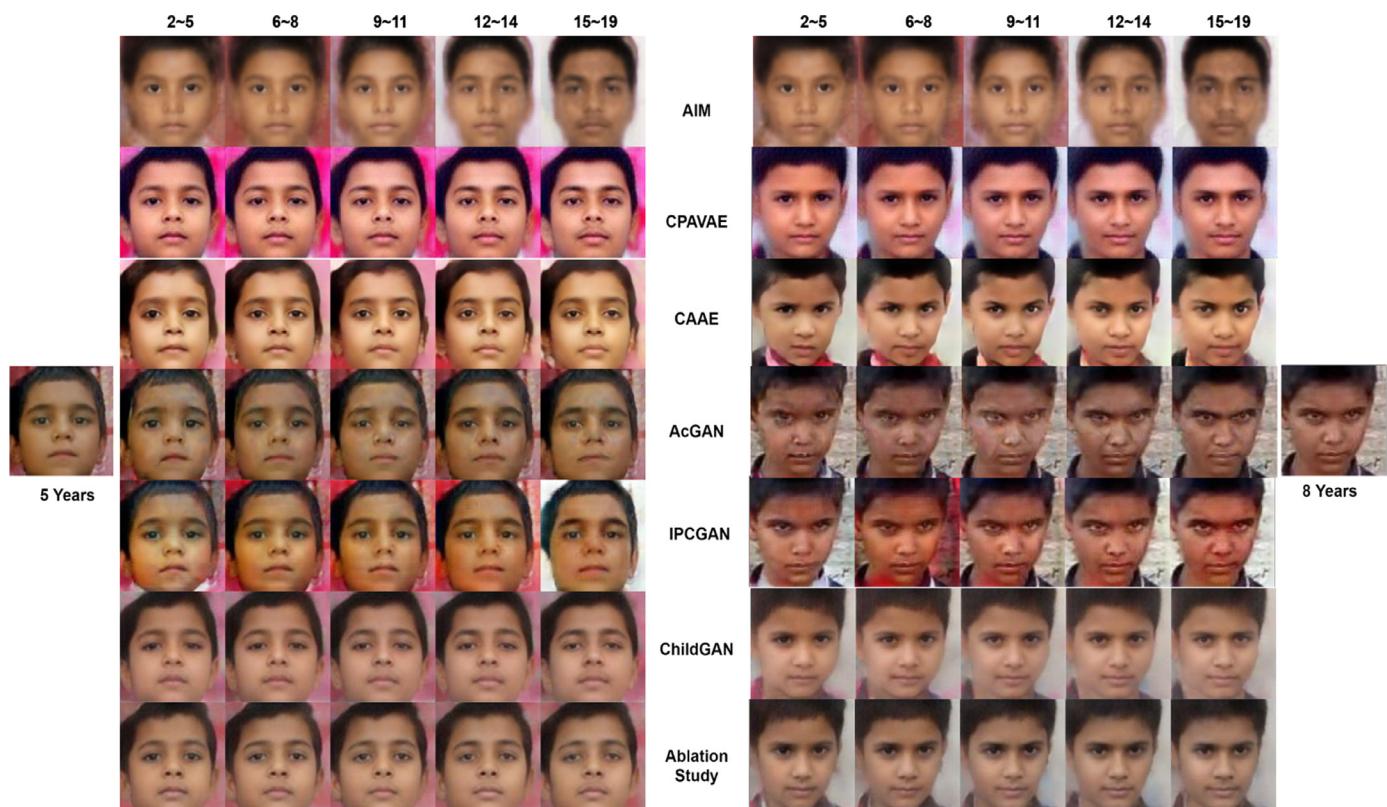
## 7. Qualitative evaluation

The main aim of ChildGAN model and ICD dataset are to help law-enforcement agencies find missing children. It may be noted that the most missing-children are from rural families, where the family often does not have photos of their missing child. Even if they have, the quality of photos is extremely bad, low resolution, obscure, and degraded from multiple handlings and scans. It is an arduous task to trace such missing children by their photos. Keeping these challenges in mind, we have tested the qualitative performance on the following two criteria:

1. Performance on low resolution faces of missing children; i.e., whether the synthesized images are photo-realistic.
2. Performance on faces with obstructions; i.e., whether the synthesized face removes the obstructions while rendering obvious aging effects.



**Fig. 7.** MTCNN [47] face detector: (a) Examples of longitudinal face data in the ICD dataset of two subjects where images were acquired annually. The age at image acquisition is given below each image; (b) Cropping and alignment effect of MTCNN.



**Fig. 8.** Comparison of missing and reunited children's faces: AIM, CPAVAE, CAAE, AcGAN, IPCGAN, and our model. The leftmost column represents the missing child's image and the rightmost column represents the subsequent reunited child's image; the ground-truth age is given below each image. The remaining columns represent the age-progressed and regressed images, along with their age categories. (Zoom in for a better view of the aging details.)

We compared ChildGAN with the following state-of-the-art models: CAAE<sup>5</sup>, CPAVAE [31], AcGAN<sup>6</sup>, age-invariant model (AIM)<sup>7</sup>, and IPCGAN<sup>8</sup>.

### 7.1. Evaluation on low resolution faces

The results of the state-of-the-art face aging models, ChildGAN, and ablation study models are shown in Fig. 8. It is easy to observe that the results generated by AIM are unrealistic, showing

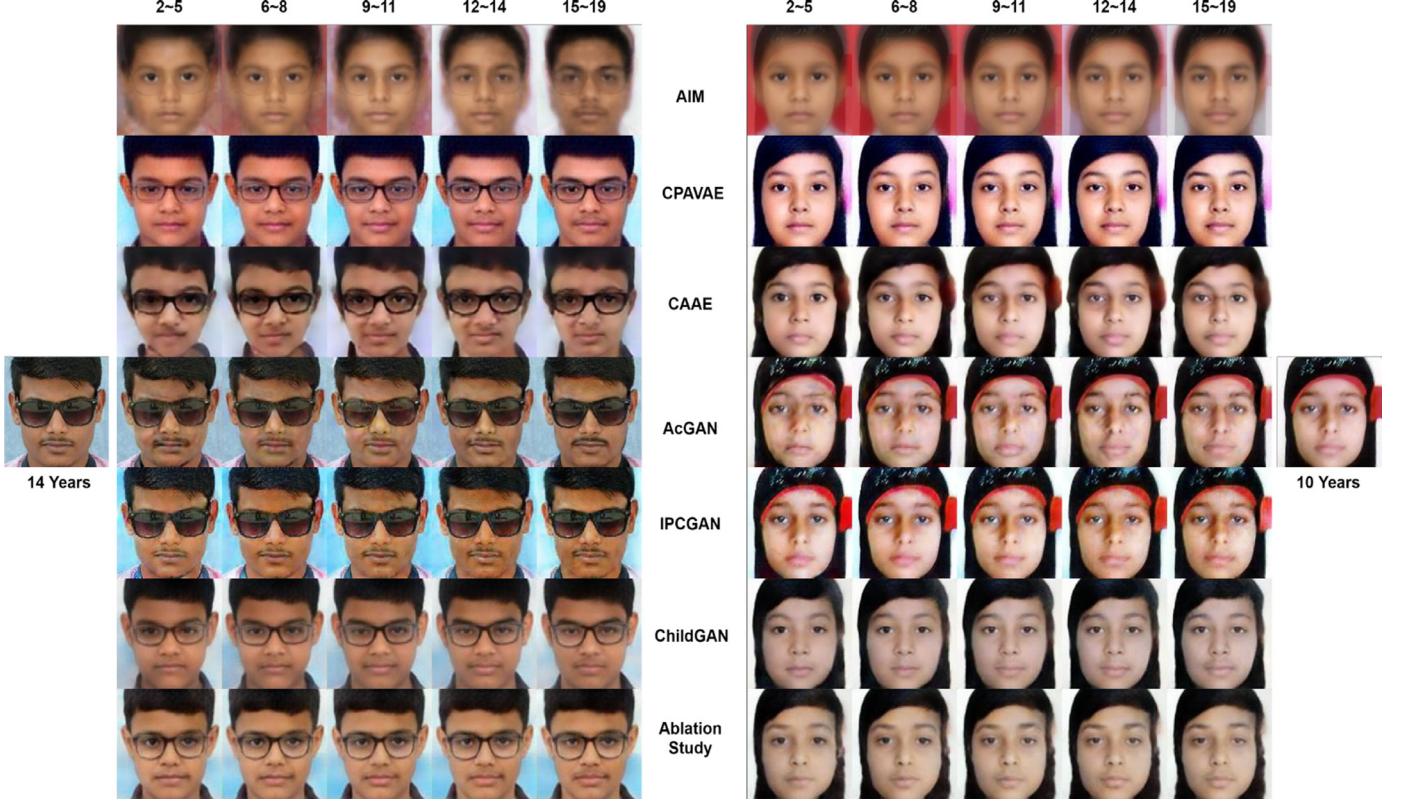
similar generated faces in different age groups, and observing very little age progression. CPAVAE beautifies images and does not preserve the skin color. In addition, it is clearly visible that the images generated by CAAE are blurry and the aging effect is not obvious. CAAE's reconstruction loss cannot capture the aging patterns to progress/regress faces. The aging changes are also not evident in AcGAN because of its attention mechanism, which only alters certain specific features, e.g., eyes and nose for face aging.

<sup>5</sup> <https://github.com/ZZUTK/Face-Aging-CAAE>.

<sup>6</sup> <https://github.com/JensonZhu14/AcGAN>.

<sup>7</sup> <https://github.com/ZhaoJ9014/High-Performance-Face-Recognition>.

<sup>8</sup> <https://github.com/dawei6875797/Face-Aging-with-Identity-Preserved-Conditional-Generative-Adversarial-Networks>.



**Fig. 9.** Comparison with prior work on obstructions (sunglasses, or head covering): AIM, CPAVAE, CAAE, AcGAN, IPCGAN, and our model. The extreme-left column represents the missing child's image and the extreme-right column represents the subsequent reunited child's image; the ground-truth age is given below each image. The remaining columns represent the age-progressed and regressed images, along with their age categories. (Zoom in for a better view of the aging details.)

IPCGAN synthesizes satisfying aging/de-aging effects and preserves the identity of the input images. However, the generated images have noticeable artifacts near the hair and mouth; they also suffer from added colored artifacts. It is easy to note that, in an ablation study (without  $\mathcal{L}_{kl}(\mu, \sigma)$ ), the hair color becomes light, with visible ghosting artifacts near the hair, and facial features, e.g., eyes, nose, and mouth, are poorly visible. Using our model, we generate visually appealing aging and rejuvenation results for a range of coherent and diverse cases, with the right shape structure for the eyes, forehead, mouth, ears, nose, hair, and skin color; therefore, it maintains identity. Further, it maintains uniformity in the background, hair regions, and face boundaries and shows rendering-enhanced aging details on blurry and noisy images.

## 7.2. Evaluation on faces with obstructions

We chose face images with obstructions, e.g., a bandage, sunglasses, or head covering, as shown in Fig. 9. CAAE cannot generate visually plausible and convincing results for face aging in such cases because of pixel-wise reconstruction loss. AcGAN, which is based on the attention mechanism, preserves the background well; however, it is unable to show prominent aging features, e.g., mustaches on male faces in the [15–19] age group, and only focuses on face appearances without hair. Furthermore, the AIM addresses the challenges of cross-age face recognition with significant age variations; however, it cannot generate photo-realistic faces with the desired aging effects.

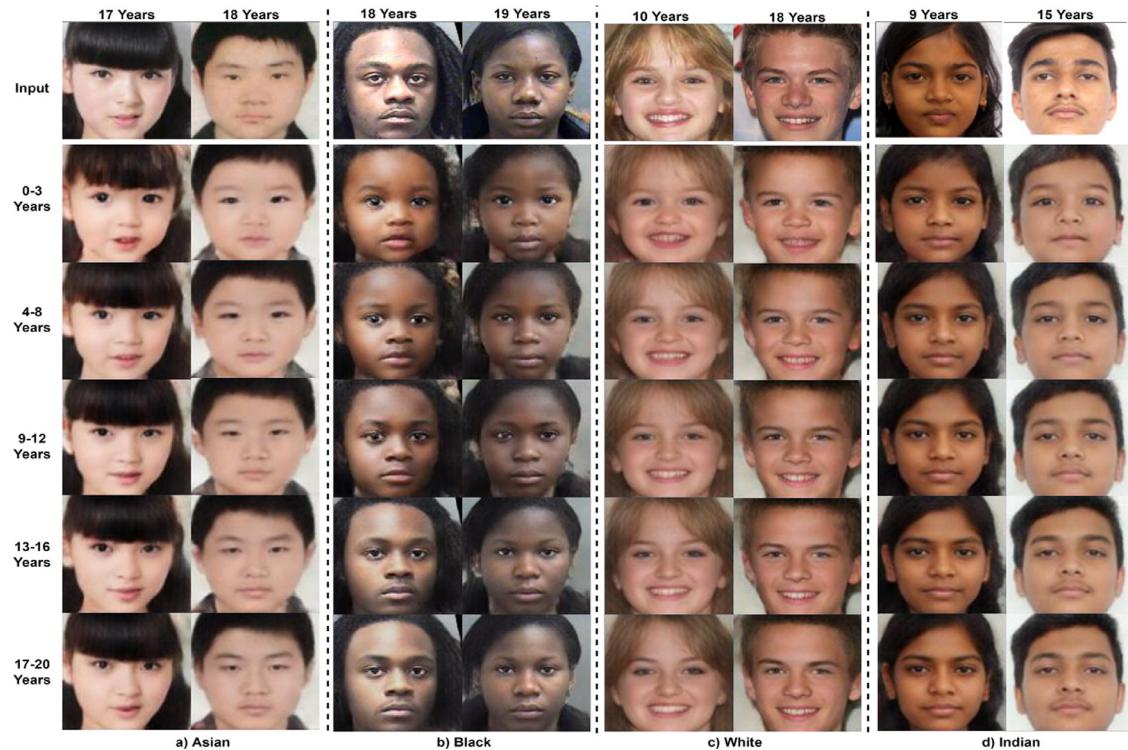
IPCGAN uses an image-to-image translation-based generator network component; hence, it cannot properly structure a partially covered face into a realistic face. CPAVAE achieves age translation and beautifies the entire face, enhancing the rejuvenation effects; however, it does not preserve the skin color very well.

By contrast, the ChildGAN synthesizes faces without removing such obstructions and can render apparent aging effects and high visual-fidelity face images with a consistent background. It probably uses a VGG19-based perceptual-loss function, paving the way for robust performance when tracing missing children.

## 8. Generalization study

We train our model on these four race groups with the same hyperparameter settings and test on the MRCD test dataset to achieve better generalization. Two tests are performed to check (a) whether race peculiarities and (b) skin tones are preserved in age-progressed faces. Fig. 10 shows the qualitative results for the first experiment. Results show that our model has successfully preserved the peculiarities of the four ethnicities while aging the faces. Below are some of the distinct characteristics of the different races, which can be observed in our model, which demonstrates its generalization ability to different races.

1. White and Black intercanthal widths are identical, whereas Asian groups show greater intercanthal widths and smaller eye openings. A narrower nasal base and larger tip projection are evident in White compared to Asians and Black [48].
2. Asians typically have a less wide mouth, elongated intercanthal width, and wider lower nasal margins. Furthermore, Asian facial structure is similar to that of an infant, including a broader and rounder face, with higher eyebrows, fuller upper lid, the lower nasal bridge with horizontally placed flared ala, flatter malar prominence and midface, fuller and more protuberant lips, and a more receded chin [48]. These characteristics are very distinct in the age progression faces too.



**Fig. 10.** Age progressed faces on four different races (a) Asian, (b) Black, (c) White (d) Indian. The first row shows the original face image, and the second to six row shows the age progression results of ChildGAN model.

**Table 4**  
Skin tone analysis of the four races.

Race	Real Image	Age progressed groups (years)					
		[0 – 3]	[4 – 8]	[9 – 12]	[13 – 16]	[17 – 20]	Average
Asian	183.73	182.46	176.01	174.38	173.31	174.86	176.20
Black	145.90	123.45	122.07	122.49	137.39	137.39	130.71
White	190.71	185.35	177.63	176.08	178.46	180.43	179.59
Indian	175.29	155.51	157.90	162.17	162.61	163.76	160.39

3. Black ethnic groups typically have a broad nasal base, with decreased nasal projection, bimaxillary protrusion, prominent lips, and increased facial convexity [48].
4. Indian ethnic groups compared to Asians typically possess eyelids that are on a more highly exposed platform, have a well-developed nasal bridge with tip projection, and have comparatively darker and more uneven skin tones [48]. Also, Indians tend to have fuller lips and higher cheekbones with more buccal fat, which gives the lower cheek a more rounded contour in aged faces.

We select 56 images from the MRCD test dataset for the second experiment. There are frontal faces with no hair covering the forehead in these images, and uniform illumination is used. We analyze the skin tone of the four races by selecting a  $3 \times 3$  patch from the forehead of the subject and averaging the patch values as a skin tone indicator. The average skin tone indicator for different races in the MRCD dataset is shown in Table 4. The white skin tones are lighter than Asian, Indian, and Black skin, which are also preserved in the age-progressed groups, and the average Black skin tone is darker than Asian, Indian, and White skin races.

## 9. Quantitative evaluation

The following are three quantitative criteria for age progression and regression:

1. Identity permanence is tested by comparing the source and synthesized images.
2. We evaluate whether the transformation of the face has been achieved for all age categories with gender preservation.
3. The quality of the generated image is evaluated.

To evaluate state-of-the-art face-aging models and our model's performance, we employed the following publicly available face matchers. FaceNet<sup>9</sup> [34] was trained on the VGGFace2 [28] dataset, using the softmax loss function. Probabilistic face embeddings (PFE)<sup>10</sup> [35] was trained using 64 CNN layers on the MS-Celeb-1M [49] dataset, and ArcFace<sup>11</sup> [36] was trained on the MS-ArcFace dataset [36].

### 9.1. Face verification

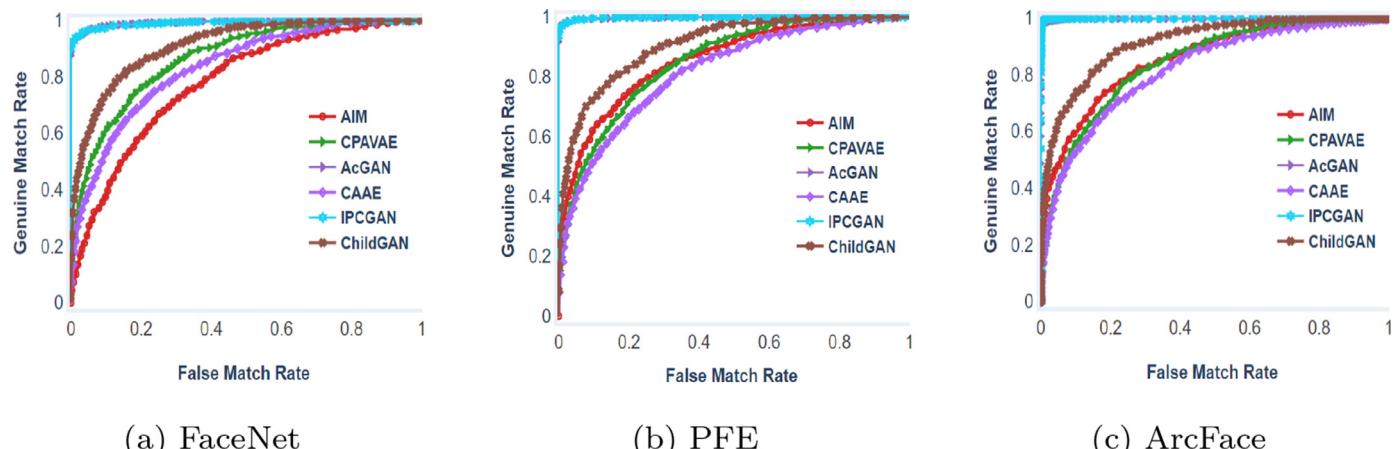
We tested the models on the ICD test dataset by finding impostor/genuine image pairs. The protocols used for a genuine (positive) and impostor (negative) pair comparison are as follows:

1. The youngest-aged (ground-truth) image is compared with an artificially aged image of the same age group.

<sup>9</sup> <https://github.com/davidsandberg/facenet>.

<sup>10</sup> <https://github.com/seasonSH/Probabilistic-Face-Embeddings>.

<sup>11</sup> <https://github.com/deepinsight/insightface>.



**Fig. 11.** (a) FaceNet [34], (b) PFE [35], and (c) ArcFace [36] face-verification results.

**Table 5**

Face-identification performance (%) of different face-aging models using CNN-based face recognition on the ICD test dataset.

Model	FaceNet [34] Rank-1%	PFE [35] Rank-1%	ArcFace [36] Rank-1%
AIM [18]	27.79	53.07	57.04
CPAVAE [31]	20.93	23.46	
ACGAN [30]	<b>77.61</b>	<b>93.86</b>	97.47
CAAE [14]	13.00	14.80	23.46
IPCGAN [17]	76.89	93.50	<b>98.19</b>
ChildGAN	34.29	34.29	40.43
<b>Ablation</b>	32.49	33.93	37.18

2. The oldest-aged (ground-truth) image is compared with an artificially aged image of the same age group, generated by the proposed model.
3. The cross-subject impostor (negative) pairs are created by pairing the progressed or regressed images with the ground truth.

The ICD test dataset includes 277 pairs of the youngest and oldest images of the same subject with a time lapse of [5–8] years. In total, there are 554 genuine matches and 7086 impostor matches.

The receiver operating characteristic (ROC) curves are illustrated in Fig. 11. We can see that the face verification by ChildGAN has a better area under the curve (AUC) compared to CAAE, AIM, and CPAVAE. IPCGAN and AcGAN perform better because they are inspired by image-to-image translation, which only changes the local facial features, e.g., eyes, nose, and mouth. However, in the case of obstructed, blurry, or low-resolution faces, they do not age the face very well, as shown in Figs. 8, and 9, and 12.

## 9.2. Face identification

In this study, the performance is reported in terms of rank-1 closed-set identification accuracy (the recovered child is in the gallery) under the youngest (gallery) to oldest (probe) protocol. For all of our experiments, the gallery set has 277 age-progressed/regressed faces from different subjects in the age range of 15–19, with 277 (oldest) real images in the probe set. The aforementioned face matchers were employed as face-identification models in this experiment. The results are shown in Table 5. We find that our model achieves the third-best true-match search accuracy, compared to IPCGAN [17] and AcGAN [30].

## 9.3. Age estimation

Along with the face aging and de-aging, the estimated age should increase and decrease. Identically, the objective age is estimated to measure the aging and de-aging accuracy [50]. We employed a state-of-the-art publicly available CNN-based age-estimation model, namely, the soft stagewise regression network (SSR-Net) [37]<sup>12</sup>, which is trained on the MegaAge-Asian [46] dataset. We separately estimated the ages of young-to-old and old-to-young faces of the same subject to obtain more insight on aging. We employed the aforementioned age-estimation CNN model to estimate the ages of synthesized faces produced by ChildGAN and compared them with state-of-the-art models.

Note that AIM generates similar-looking faces, as depicted in Figs. 8, 9, and 12. Thus, for a fair comparison with our model and various state-of-the-art models, we cannot consider this model for age and gender estimation.

IPCGAN and AcGAN use the age labels as conditions in an image-to-image translation-based generator network and incorporate an age-classification loss function. However, because the classification error rate is high, the gradient for the small age range is not accurate. As a result, the estimation accuracy of these models is lower than that of the ChildGAN and CPAVAE models for the majority of the age groups, as shown in Tables 6 and 7.

ChildGAN performs well for both the youngest-to-oldest and oldest-to-youngest age-estimation results, compared to CAAE. CPAVAE and ChildGAN both achieve similar accuracy because they both use the age range and gender with a discriminator network.

We observe that IPCGAN and AcGAN overly age the high-level feature region (nose and eye areas) for face aging. Hence, these models create wrinkles around these features on the face, as shown in Figs. 8, 9, 12, and 13. As a result, these faces appear overaged. This is also observed in the values of the mean and standard deviation for the age estimation of these models with larger variations. ChildGAN is nearly as good as the ablation study, which also demonstrates that the age-related representation is well disentangled from the age-irrelevant representation.

## 9.4. Gender preservation

The gender preservation, referring to the percentage of the original gender preserved, is calculated by comparing the estimated gender with the original ICD test annotations. ChildGAN uses gender labels as conditions in adversarial learning. Therefore, the pro-

<sup>12</sup> [https://github.com/wayen820/gender\\_age\\_estimation\\_mxnet](https://github.com/wayen820/gender_age_estimation_mxnet).



**Fig. 12.** Comparison with prior work on UTKFace and FGNET: AIM, CPAVAE, CAAE, AcGAN, IPCGAN, and ChildGAN. In the leftmost and rightmost columns, we provide the input faces, with their ground-truth ages below each image. The middle columns show the generated face images, along with their categories. (Zoom in for a better view of the aging details.)

**Table 6**

Age estimation (years) evaluated by SSR-Net [37]. Owing to limited space, we only list the mean and standard deviation of the age estimation computed over young-to-old faces.

Age group	Age Estimation (years) of young-to-old faces (age progression).				
	2-5	6-8	9-11	12-14	15-19
CPVAE [31]	4.05 ± 2.31	5.72 ± 3.04	8.25 ± 4.19	12.71 ± 5.78	15.39 ± 6.19
AcGAN [30]	2.63 ± 1.76	8.77 ± 5.01	12.86 ± 5.99	18.06 ± 6.09	24.96 ± 5.23
CAAE [14]	5.93 ± 3.40	8.48 ± 4.58	10.85 ± 5.18	14.09 ± 5.54	16.90 ± 5.16
IPCGAN [17]	4.57 ± 3.64	6.40 ± 4.46	9.75 ± 6.08	14.30 ± 7.82	14.34 ± 8.06
ChildGAN	4.56 ± 2.58	6.83 ± 3.30	8.85 ± 4.38	11.92 ± 5.42	14.16 ± 5.87
Ablation	4.76 ± 2.98	6.25 ± 3.91	7.98 ± 4.61	10.40 ± 5.58	12.95 ± 6.23

**Table 7**

Age estimation (years) evaluated by SSR-Net [37]. Owing to limited space, we only address the mean and standard deviation of the age estimation computed over old-to-young faces.

Age group	Age Estimation (years) of old-to-young faces (age regression).				
	2-5	6-8	9-11	12-14	15-19
CPVAE [31]	12.17 ± 5.24	15.50 ± 5.52	18.17 ± 4.81	21.56 ± 3.75	22.41 ± 3.28
AcGAN [30]	12.14 ± 7.44	17.06 ± 6.60	19.85 ± 5.99	23.64 ± 4.49	27.22 ± 3.64
CAAE [14]	12.10 ± 5.55	16.27 ± 5.40	18.79 ± 5.01	21.34 ± 4.07	22.55 ± 3.50
IPCGAN [17]	9.05 ± 6.79	14.97 ± 7.45	21.43 ± 6.47	23.95 ± 6.41	23.06 ± 6.09
ChildGAN	11.37 ± 4.77	14.53 ± 5.13	18.12 ± 4.89	20.30 ± 3.98	21.85 ± 3.35
Ablation	13.39 ± 5.38	15.84 ± 5.62	18.77 ± 5.16	20.64 ± 4.57	22.06 ± 4.01

posed method achieves significant improvements in terms of gender preservation, as is clearly demonstrated in Tables 8 and 9.

#### 9.5. Generation fidelity: inception score and FID

The image quality with the diversity of the generated data is assessed in terms of the inception score (IS) [5] and the Fréchet inception distance (FID) [33]. As can be seen in Table 10, the proposed model achieves comparative IS and FID scores on the ICD

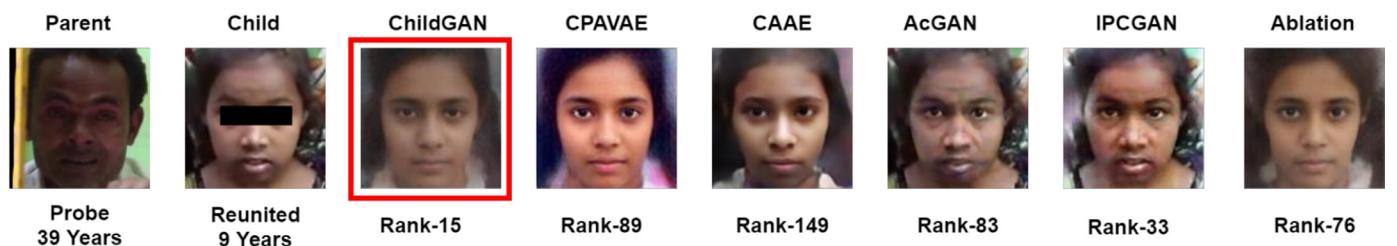
and UTKFace test datasets, compared to the state-of-the-art models.

#### 10. Additional comparison results with UTKFace and FGNET

ChildGAN model is trained on the ICD dataset. Additional comparison results with state-of-the-art face-aging models on unconstrained UTKFace and FGNET are shown in Fig. 12 and Table 11. CPAVAE beautifies the face images, as the model is based on



(a) A case study (mother and daughter1).



(b) A case study (father and daughter2).

**Fig. 13.** Column 1: Face images of parents, mother and father, obtained from the “These Are Our Kids” video<sup>13</sup>, enrolled in the probe. Column 2: Face images of reunited children. Columns 3–8: Age-progressed images in the 15–19-year age range with the rank identities correctly retrieved by ArcFace [36]. (The best result is highlighted in the red box.)

**Table 8**

Gender-preservation rate (%) computed for young-to-old faces by SSR-Net.

Age group	Gender-preservation rate (%) of young-to-old faces.				
	2–5	6–8	9–11	12–14	15–19
CPVAE [31]	76.90%	75.81%	74.37%	74.37%	75.09%
AcGAN [30]	64.62%	71.84%	69.68%	68.95%	70.04%
CAAE [14]	65.34%	66.06%	70.40%	68.23%	68.23%
IPCGAN [17]	62.09%	67.15%	67.51%	68.59%	69.31%
ChildGAN	78.70%	78.70%	75.09%	74.04%	77.26%
Ablation	72.56%	76.17%	76.90%	74.01%	75.45%

**Table 9**

Gender-preservation rate (%) computed for old-to-young faces by SSR-Net.

Age group	Gender-preservation rate (%) of old-to-young faces.				
	2–5	6–8	9–11	12–14	15–19
CPVAE [31]	79.42%	85.56%	83.39%	74.73%	72.56%
AcGAN [30]	56.70%	69.31%	68.59%	72.20%	71.48%
CAAE [14]	61.37%	68.59%	70.76%	67.87%	66.79%
IPCGAN [17]	57.40%	67.51%	71.12%	69.68%	67.87%
ChildGAN	83.03%	84.12%	77.98%	76.17%	75.09%
Ablation	82.67%	83.39%	82.67%	79.5%	77.98%

**Table 10**

Comparing the IS and FID of our model with state-of-the-art models. A higher score is better for IS, and a lower score is better for the FID score.

Model	ICD Test		UTKFace	
	IS	FID	IS	FID
CPVAE [31]	$1.59 \pm 0.02$	46.20	$1.81 \pm 0.22$	69.96
CAAE [14]	$1.54 \pm 0.03$	41.01	$1.67 \pm 0.22$	66.93
AcGAN [30]	$1.64 \pm 0.03$	13.62	$2.28 \pm 0.11$	29.86
IPCGAN [17]	$1.69 \pm 0.43$	26.25	$2.26 \pm 0.13$	37.84
ChildGAN	$1.63 \pm 0.04$	30.42	$1.86 \pm 0.20$	56.59
Ablation	$1.54 \pm 0.03$	33.77	$1.83 \pm 0.21$	62.57

the deep-feature consistency principle [43]. Where normalization, with  $\mu = [0.485, 0.456, 0.406]$  and  $\sigma = [0.229, 0.224, 0.225]$ , is performed on the input which beautifies the face. As a result, the faces in the [2–5] age range look younger.

IPCGAN and AcGAN generate artifacts on the face because their standard deviation values (of the generated population) are high, as shown in Table 11. As depicted in Fig. 12, the generated face images of ChildGAN and the ablation study (without  $\mathcal{L}_{kl}(\mu, \sigma)$ ) do not have eyeglasses for the [2–8] age range, whereas they are present for the [9–19] age-range images. This implies that the proposed model can better learn facial attributes. It has learned the correlation between wearing eyeglasses and age.

## 11. Case studies of reunited children and parents

For this study, we used six face-aging models (ChildGAN, CPVAE [31], CAAE [14], AcGAN [30], IPCGAN [17], and the ablation study) to construct a gallery of missing children, consisting of 290 age-progressed images in the age range of 15–19 years from the ICD test dataset.

Fig. 13 shows the images of reunited children, daughter1 and daughter2, aged 12 and 9 years, respectively, at the time of the reunion. The face images of the reunited children's parents, mother (35 years) and father (39 years), were used as the probe. ArcFace [36], a state-of-the-art CNN-based face-matcher, was able to retrieve the probe's true mate in a gallery set that was generated by the aforementioned aging models. The proposed ChildGAN and ArcFace were able to retrieve the true mate at Rank-31 and Rank-15, respectively. These case studies show the suitability of our ChildGAN model and the ArcFace matcher for face matching to help reunite missing children with their parents [51].

## 12. Ablation study

In this section, we conduct an ablation study on the low-resolution or obstructed faces of children and UTKFace, and FGNET

**Table 11**

Age estimation (years) evaluated by SSR-Net [37] on the UTKFace and FGNET test datasets. Owing to limited space, we only address the mean and standard deviation of the age estimation computed over the [2–19] age range.

Age group	Age Estimation (years) on UTKFace & FGNET.				
	2–5	6–8	9–11	12–14	15–19
CVAE [31]	8.84 ± 5.02	13.53 ± 6.59	18.96 ± 6.62	23.00 ± 5.46	23.99 ± 4.74
AcGAN [30]	9.01 ± 9.27	15.26 ± 9.73	18.28 ± 9.01	24.23 ± 7.49	28.34 ± 7.56
CAAE [14]	10.19 ± 6.18	14.44 ± 6.94	17.99 ± 6.78	20.74 ± 6.09	22.47 ± 5.59
IPCGAN [17]	12.12 ± 10.24	13.53 ± 10.63	16.75 ± 10.68	20.28 ± 10.94	20.67 ± 10.58
ChildGAN	9.47 ± 5.62	13.09 ± 6.39	18.17 ± 6.09	19.99 ± 5.56	22.12 ± 5.15
Ablation	10.36 ± 6.05	14.18 ± 7.02	18.88 ± 6.73	21.06 ± 5.99	22.83 ± 5.83

**Table 12**

Gender-preservation rate (%) computed over UTKFace and FGNET by SSR-Net [37].

Age group	Gender preservation rate (%) of UTKFace & FGNET.				
	2–5	6–8	9–11	12–14	15–19
CVAE [31]	79.79%	78.18%	74.70%	63.80%	62.80%
AcGAN [30]	77.44%	77.13%	76.83%	67.68%	72.56%
CAAE [14]	74.09%	67.68%	64.33%	60.98%	65.24%
IPCGAN [17]	77.74%	76.52%	75.61%	74.70%	69.51%
ChildGAN	80.57%	78.78%	75.12%	69.16%	68.02%
Ablation	78.72%	73.12%	71.67%	67.17%	64.43%

datasets to evaluate the introspective  $\mathcal{L}_{kl}(\mu, \sigma)$  loss term. One assumption of ChildGAN is that introducing the  $\mathcal{L}_{kl}$  loss function advances the generation of more subtle aging patterns. In Figs. 8, 9, and 12 present a visual comparison between ChildGAN and the ablation study (without  $\mathcal{L}_{kl}(\mu, \sigma)$ ). We observed that the ChildGAN age-progressed and regressed faces were artifact-free, with a clear skin texture and photo-realistic results, compared to the ablation study.

In Tables 5–10 show that ChildGAN outperforms the ablation study in identification, age estimation, gender preservation, and inception score. Furthermore, the age-estimation and gender-preservation accuracy demonstrate that ChildGAN effectively disentangled age-related information, e.g., craniofacial growth, and age-irrelevant information, e.g., gender and identity, in the latent space, compared to state-of-the-art models. The visualization and quantitative results demonstrate the relevance of the  $\mathcal{L}_{kl}(\mu, \sigma)$  in ChildGAN.

### 13. Conclusions

In this paper, we present two novel large-scale child face datasets as (a) Indian child dataset (ICD) and (b) Multi-Racial Child Dataset (MRCD) to spark progress in child face-aging, face recognition, skin tone analysis, age estimation, gender preservation, and kinship face identification. ChildGAN model and benchmark datasets should drive face-aging and cross-age face recognition research toward finding missing children, as well as numerous real-world applications, e.g., border control, passport verification, access control, forensic science, and underage driver's licenses. The proposed model achieved state-of-the-art performance in age estimation with flawless gender preservation. ChildGAN demonstrates adaptation to a wide range of ethnicities, such as Asians, Blacks, Whites, and Indians, while preserving the distinctive characteristics of each race. IPCGAN, AcGAN, ChildGAN models, along with the ICD and MRCD datasets, make a significant contribution to society by reuniting missing children with their loved ones. We wish to experiment with face aging with ethnicity bias and 2D-to-3D face aging in the future.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

This research is based upon work supported by the Ministry of Electronics and Information Technology (MeitY), Government of India, under Grant No. 4 (13)/2019-ITEA. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the TITAN V GPU used for this research. We are also grateful to Cognitec with the donation of the FaceVACS-DBScanID5.6 software for this work. We are grateful to anonymous reviewers for their critical comments for improved the manuscript.

### References

- [1] D. Deb, D. Aggarwal, A.K. Jain, Identifying missing children: face age-progression via deep feature aging, in: Proceedings of the ICPR, 2021, pp. 10540–10547.
- [2] D.P. Kingma, M. Welling, Auto-encoding variational bayes, in: Proceedings of the ICLR, 2014, pp. 1–14.
- [3] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, et al, Generative adversarial nets, in: Proceedings of the NIPS, 2014, pp. 2672–2680.
- [4] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks, Proceedings of the ICML (2017) 214–223.
- [5] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training gans, in: Proceedings of the NIPS, 2016, pp. 2234–2242.
- [6] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, Proceeding of the ICLR, 2016, pp. 1–16.
- [7] M. Mirza, S. Osindero, Conditional generative adversarial nets, CoRR (2014) 1–7.
- [8] C. Ledig, L. Theis, F. Huszar, J. Caballero, et al, Photo-realistic single image super-resolution using a generative adversarial network, in: Proceedings of the CVPR, 2017, pp. 105–114.
- [9] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A.C. Courville, Improved training of Wasserstein gans, in: Proceedings of the NIPS, 2017, pp. 5769–5779.
- [10] J. Suo, S. Zhu, S. Shan, X. Chen, A compositional and dynamic model for face aging, IEEE Trans. Patt. Anal. Mach. Intell. 32 (3) (2010) 385–401.
- [11] N. Ramanathan, R. Chellappa, S. Biswas, Computational methods for modeling facial aging: a survey, J. Vis. Lang. Comput. 20 (3) (2009) 131–144.
- [12] I.K. Shlizerman, S. Suwajanakorn, S.M. Seitz, Illumination-aware age progression, in: Proceedings of the CVPR, 2014, pp. 3334–3341.
- [13] G. Antipov, M. Baccouche, J. Dugelay, Face aging with conditional generative adversarial networks, in: Proceedings of the ICIP, 2017, pp. 2089–2093.
- [14] Z. Zhang, Y. Song, H. Qi, Age progression/regression by conditional adversarial autoencoder, in: Proceedings of the CVPR, 2017, pp. 4352–4360.
- [15] H. Yang, D. Huang, Y. Wang, A.K. Jain, Learning face age progression: a pyramid architecture of gans, in: Proceedings of the CVPR, 2018, pp. 31–39.
- [16] P. Li, Y. Hu, Q. Li, R. He, Z. Sun, Global and local consistent age generative adversarial networks, in: Proceedings of the ICPR, 2018, pp. 1073–1078.
- [17] Z. Wang, X. Tang, W. Luo, S. Gao, Face aging with identity-preserved conditional generative adversarial networks, in: Proceedings of the CVPR, 2018, pp. 7940–7947.
- [18] J. Zhao, Y. Cheng, Y. Cheng, et al, Look across elapse: disentangled representation learning and photorealistic cross-age face synthesis for age-invariant face recognition, in: Proceedings of the AAAI, 2019, pp. 9251–9258.

- [19] K. Ricanek, T. Tesafaye, Morph: a longitudinal image database of normal adult age-progression, in: Proceedings of the Automatic Face and Gesture Recognition, 2006, pp. 341–345.
- [20] B. Chen, C. Chen, W.H. Hsu, Cross-age reference coding for age-invariant face recognition and retrieval, in: Proceedings of the ECCV, 2014, pp. 768–783.
- [21] R. Rothe, R. Timofte, L.V. Gool, Dex: Deep expectation of apparent age from a single image, in: Proceedings of the ICCVW, 2015, pp. 252–257.
- [22] G. Panis, A. Lanitis, N. Tsapatsoulis, et al, Overview of research on facial ageing using the FG-NET ageing database, IET Biometr. (2016) 37–46.
- [23] Megvii Inc, Facial recognition API and SDKs, 2011 [www.faceplusplus.com](http://www.faceplusplus.com).
- [24] S. Liu, Y. Sun, D. Zhu, R. Bao, Face aging with contextual generative adversarial nets, in: Proceedings of the International Conference on Multimedia, 2017, pp. 82–90.
- [25] E. Eran Eidinger, R. Enbar, T. Hassner, Age and gender estimation of unfiltered faces, in: Proceedings of the IEEE-TIFS, volume 9, 2014, pp. 2170–2179.
- [26] Yandong Wen, Kaipeng Zhang, Zhifeng Li, Yu Qiao, A discriminative feature learning approach for deep face recognition, in: Proceedings of the ECCV, Springer International Publishing, 2016, pp. 499–515.
- [27] G.B. Huang, M. Ramesh, T. Berg, E. Learned-miller, Labeled faces in the wild : a database for studying face recognition in unconstrained environments, in: Proceedings of the Advances in Face Detection and Facial Image Analysis, IEEE, 2007, pp. 2170–2179.
- [28] Q. Cao, L. Shen, W. Xie, O.M. Parkhi, A. Zisserman, VGGFace2: a dataset for recognising faces across pose and age, in: Proceedings of the International Conference on Automatic Face and Gesture Recognition, 2018, pp. 1–11.
- [29] C. Shi, J. Zhang, Y. Yao, et al, CAN-GAN: conditioned-attention normalized gan for face age synthesis, Patt. Recogn. Lett. (2020) 520–526.
- [30] H. Zhu, Z. Huang, H. Shan, J. Zhang, Look globally, age locally: face aging with an attention mechanism, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2020, pp. 1963–1967.
- [31] P.K. Chandaliya, N. Nain, Conditional perceptual adversarial variational autoencoder for age progression and regression on child face, in: Proceedings of the ICB, 2019, pp. 1–8.
- [32] D. Deb, N. Nain, A.K. Jain, Longitudinal study of child face recognition, in: Proceedings of the ICB, 2018, pp. 225–232.
- [33] M. Heusel, H. Ramsauer, T. Unterthiner, et al, Gans trained by a two time-scale update rule converge to a local nash equilibrium, in: Proceedings of the NIPS, 2017, pp. 6629–6640.
- [34] F. Schroff, D. Kalenichenko, J. Philbin, Facenet: a unified embedding for face recognition and clustering, in: Proceedings of the CVPR, 2015, pp. 815–823.
- [35] Y. Shi, A. Jain, Probabilistic face embeddings, in: Proceedings of the ICCV, 2019, pp. 6901–6910.
- [36] J. Deng, J. Guo, X. Niannan, S. Zafeiriou, ArcFace: additive angular margin loss for deep face recognition, in: Proceedings of the CVPR, 2019, pp. 4685–4694.
- [37] T.-Y. Yang, Y.-H. Huang, Y.-Y. Lin, P.-C. Hsiu, Y.-Y. Chuang, SSR net: a compact soft stagewise regression network for age estimation, in: Proceedings of the IJCAI, 2018, pp. 1078–1084.
- [38] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Proceedings of the ICLR, 2015, pp. 1–14.
- [39] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, S. Zafeiriou, Agedb: the first manually collected, in-the-wild age database, in: Proceedings of the CVPRW, 2017, pp. 1997–2005.
- [40] N. Srinivas, K. Ricanek, D. Michalski, D.S. Bolme, M. King, Face recognition algorithm bias: performance differences on images of children and adults, in: Proceedings of the CVPRW, 2019, pp. 2269–2277.
- [41] H. Zhang, I. Goodfellow, D. Metaxas, et al, Self-attention generative adversarial networks, Mach. Learn. Res. (2019) 7354–7363.
- [42] G. Baykal, F. Ozcelik, G. Unal, Exploring deshufflegans in self-supervised generative adversarial networks, Patt. Recogn. (2022) 108244.
- [43] X. Hou, L. Shen, K. Sun, G. Qiu, Deep feature consistent variational autoencoder, in: Proceedings of the WACV, 2017, pp. 1133–1141. IEEE Computer Society.
- [44] H. Zhou, K.-M. Lam, Age-invariant face recognition based on identity inference from appearance age, Patt. Recogn. (2018) 191–202.
- [45] T. Karras, S. Laine, T. Aila, A style-based generator architecture for generative adversarial networks, in: Proceedings of the CVPR, 2019, pp. 1–8.
- [46] Y. Zhang, L. Liu, C. Li, C.C. Loy, Quantifying facial age by posterior of age comparisons, in: Proceedings of the BMVC, 2017, pp. 1–14.
- [47] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, Joint face detection and alignment using multitask cascaded convolutional networks, IEEE Signal Process. Lett. 23 (2016) 1499–1503.
- [48] N.A. Vashi, de Castro Maymone MB, R.V. Kundu, Aging differences in ethnic skin, J. Clin. Aesthet. Dermatol. (2016) 31–38.
- [49] Y. Guo, L. Zhang, Y. Hu, X. He, J. Gao, MS-Celeb-1M: a dataset and benchmark for large-scale face recognition, in: Proceedings of the ECCV, 2016, pp. 1–14.
- [50] G. Antipov, M. Baccouche, S.-A. Berrani, J.-L. Dugelay, Effective training of convolutional neural networks for face-based gender and age prediction, Patt. Recogn. (2017) 15–26.
- [51] X. Shu, J. Tang, H. Lai, Z. Niu, S. Yan, Kinship-guided age progression, Patt. Recogn. (2016) 156–167.

**Praveen Kumar** Chandaliya received his M.Tech degree at Malaviya National Institute of Technology Jaipur, in 2014. He is currently pursuing his Ph.D. degree at the Malaviya National Institute of Technology Jaipur. He is also working as a Senior Research Fellow on the research project sponsored by MeitY, Government of India. From 2019 to 2020, he was a visiting scholar at Michigan State University, USA under the supervision of distinguished Prof. Anil K. Jain. His current research interests include age-invariant face recognition, face aging synthesis, and he has a keen interest in application-oriented projects.

**Neeta Nain Ph.D.**, presently working as Associate Professor, Department of Computer Science and Engineering, Malaviya National Institute of Technology Jaipur, has a teaching experience of over 25 years. Her research area is Pattern Recognition, Machine Learning, Deep Learning and Biometrics. She has published more than 75 papers on these topics for various International Journals and conferences like Elsevier Journal of Visual Comm. Information Representation, ACM Transactions on Asian Language Information Processing, IETE Journal of Research, Springer Multi-media Tools and Applications etc. She has acted as Program Chair, of The 13th IEEE International Conference on SIGNAL IMAGE TECHNOLOGY & INTERNET BASED SYSTEMS (SITIS2017) and IAPR endorsed Fourth International Conference on Computer Vision and Image Processing, CVIP2019.