



PlasticGAN: Holistic generative adversarial network on face plastic and aesthetic surgery

Praveen Kumar Chandaliya¹ · Neeta Nain¹

Received: 12 November 2020 / Revised: 1 April 2021 / Accepted: 10 March 2022 /

Published online: 12 April 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

By embracing Generative Adversarial Networks (GAN), several face-related applications have significantly benefited and achieved unparalleled success. Inspired by the latest advancement in GAN, we propose the PlasticGAN which is a holistic framework for generating images of post-surgery faces as well as reconstruction of faces after surgery completion. This preliminary model works as a helping hand in assisting surgeons, biometric researchers, and practitioners in clinical decision-making by identifying patient cohorts that require building up of confidence with the help of vivid visualizations prior to treatment. It helps them better provide the tentative alternatives by simulating aging patterns. We used the face recognition system for evaluating the same individual with and without masks on surgery face, keeping the current trends in mind such as forensic and security application and recent worldwide COVID scenario. The experimental results suggested that plastic surgery-based synthetic cross-age face recognition (PSBSCAFR) is an arduous research challenge, and state-of-art face recognition systems can negatively affect face recognition performance. This can present a new dimension for the research community.

Keywords Plastics surgery · Generative adversarial networks · Cross-age face recognition · Face aging · Face masks

1 Introduction

Usually, facial plastic surgery reshapes the structure or improves the appearance of the face or neck. Procedures typically include the nose, ears, chin, cheekbones, and neckline. People seeking this surgery might be motivated by a desire to remove irregularities introduced in the face by an injury, a disease, congenital disabilities, or post-surgical deformities. The following is according to the recent statistics released by The American Society for Aesthetic

✉ Praveen Kumar Chandaliya
2016rcp9511@mnit.ac.in

Neeta Nain
nnain.cse@mnit.ac.in

¹ Malaviya National Institute of Technology, Jaipur, 302017, India

Plastic Surgery (ASPS)¹ for the year 2019 and International study on aesthetic procedures (ISAP)² for the year 2019.

1. According to the ASPS, almost 17.7 million people underwent surgically and minimally invasive cosmetic procedures, and more than 5.9 million reconstructive procedures happened in the United States in 2019.
2. In 2019, According to the ISAP, more than 2 million facial rejuvenation surgical procedures were performed, and the most common were Chemical Peel, Full-Field Ablative Resurfacing, Micro-Ablative Resurfacing, Dermabrasion, Microdermabrasion, Nonsurgical Skin Tightening, and Face Rejuvenation.
3. In 2019, 11.1% and 8.5% of face and head plastic surgery procedures were performed in Brazil and the USA, respectively. 10.5% and 18.3% of facial rejuvenation were constituted by USA and Japan, respectively.
4. Plastic surgery distribution by age: 0–18 years constitutes 4%, 19–34 years constitutes 43%, 35–50 years constitutes 36%, 51–64 years constitutes 14%, and 65 years and above constitute 3% of the total number of plastic surgery procedures.
5. 86.5% of women and 13.5% of men are now more affirmative toward plastic surgery procedure as per the data provided by the ISAP in 2019.

The statistics provided by ASPS and ISAP indicate the popularity of facial plastic surgery across all age categories, ethnicity, countries, and gender. In South Korea, every one in three women between 19 to 29 year had a cosmetic or plastic surgery.³

These surgical processes demonstrate ideal for individuals experiencing facial distortions or those who want to counter the aging process. Similarly, these procedures can be misused by individuals who are attempting to hide their personality with the goal to cause extortion or dodge law implementation [34]. These surgeries may permit anti-social elements to openly move around with no dread of being recognized by any automated face recognition framework. A considerable amount of research on plastic surgery has been performed and reported [33], cross-age face recognition(CAFR) [40], synthetic aging [5, 8], and synthetic face mask recognition [27]. Due to advances in technology in the medical field, variation in faces due to plastic/cosmetic surgery has also led to the emergence of co-variates [33] of face recognition. Furthermore, if we add synthetic aging to surgery faces then the cross-age face recognition task becomes arduous.

Facial plastic surgery is a discipline that requires years of training for a surgeon to gain the necessary experience, skill, and dexterity. As the demand of minimally invasive procedures (MIP) is increasing rapidly, the patient wants to know how the changes are reflected on their face after the surgery. But in these procedures, the surgeon's vision often relies on their own and the patient's imagination completely. Due to the lack of appropriate visualization techniques and technology, surgeons are bound to rely on their skills and imagination power while performing the surgery; this can make this task more challenging.

To attempt to alleviate these challenges, we propose the PlasticGAN framework which can generate diverse photo-realistic faces with respect to facial surgery; this can work as a middleware between surgeons and patients and aid clinical decisions with the help of vivid

¹<https://www.surgery.org>

²<https://www.isaps.org>

³<https://www.isaps.org/event/aesthetic-plastic-surgery-2018-korean-society-aesthetic-plastic-surgery>

visualizations. In this manuscript, we also focus on quantifying the effect of plastic surgery with aging and wearing face masks on the performance of face recognition systems.

Our key efforts are summarized as follows:

1. An effective Conditional Generative Adversarial Network (cGAN) based network, PlasticGAN, is proposed to solve the face aging and rejuvenation problem on faces that have undergone plastic surgery for the very first time. Specifically, age and gender are passed as conditional information into both the discriminator and generator to acquire more fine-grained information between input and synthesized results. Besides, BlockCNN-based residual blocks are adopted to remove the artifacts and improve convergence behavior.
2. PlasticGAN will work as a middleware between surgeons and patients in terms of motivation provider and confidence booster for the surgery by providing a better glimpse of post-surgery looks.
3. Our framework does not require pre- and post-plastic surgery faces in the training dataset. At the time of testing, our model synthesizes face aging, rejuvenation, and face completion on surgery faces.
4. We defined a new challenge in the face recognition field named plastic surgery-base synthetic cross-age face recognition (PSBSCAFR).
5. We evaluated the robustness of the PlasticGAN model. For this, we performed an excessive qualitative and quantitative comparison with faces with and without face masks that will contribute to the forensic and law enforcement field.

The primary aim of this paper is to add a new dimension to clinical decision-making between surgeon and patient as well as lend an impact on cross-age face recognition on faces that have undergone plastic surgery. The remainder of the paper is organized as follows: In Section 2 we present a detailed description of different types of plastic surgery. In Section 3, we provide the generative model-based related work on face images. In Section 4, we present the proposed PlasticGAN model in detail. Section 5 presents the overall objective functions used in the optimization of our model. Section 6 describes data collection, pre-processing on surgery, and mask-wearing face. Section 7 presents the qualitative study on different types of surgeries. Section 8 discusses qualitative study on mask-wearing faces. Section 9 presents extensive quantitative experiments to demonstrate the superiority and wide-range practical application of our proposed model. Section 10 presents the performance ablation study. Finally, in Section 11, we conclude the paper and discuss the challenges in the face recognition research domain.

2 Plastic surgery and face recognition

Primarily, plastic surgery constrained to the face can be characterized into two major categories: (1) local plastic surgery and (2) global plastic surgery. Local plastic surgery accounts for correcting defects and craniofacial anomalies and improving skin texture. Procedures for local plastic surgery include Rhinoplasty, Mentoplasty, Blepharoplasty, Browlift, Malar Augmentation, and Otoplasty. It is also used for cosmetic and aesthetic purposes [28]. Global plastic surgery remodels the overall facial structure. This plastic surgery procedure is recommended for cases of recovery from fatal burn, changing facial appearance and skin texture, and modifying facial geometries. Procedures for global plastic surgery include Rhytidectomy, Skin Peeling, Craniofacial, Dermabrasion, and Mole Removal [26, 32].

When global plastic surgery is performed on an individual, face components such as nose, lips, and eyelid geometries might be disturbed or modified. In parallel, we observed that face recognition accuracy was significantly improved by the commercial of the shelf (COTS) and open source deep face recognition system on face aging [5], plastic surgery [34], and disguise face [10].

In this paper, PlasticGAN based on Generative Adversarial model is synthesized with reference to faces that have undergone plastic surgery with and without mask considering face aging and rejuvenation effect. Subsequently, we performed face verification using Face++ app on faces generated by PlasticGAN. Hence, this creates a significant challenge for the face recognition system as it produces extensive changes in facial features. This challenge, namely plastic surgery-based synthetic cross-age face recognition (PSBSCAFR) can become a new dimension for upcoming researches.

3 Related work

3.1 Generative adversarial networks

GAN was proposed by Ian Goodfellow et al. which incorporates two networks, a generator which produces new instance data by accepting noise, a discriminator of the genuineness of the produced images by generator. GAN has gained interest due to its high performance in a wide range of practical applications such as facial attribute alteration [3, 19], finding missing children [4, 5], transferring and removing makeup style [7], super-resolution techniques [22], anti-forensic [11, 37] and law enforcement [1, 21], Age-Invariant Face Recognition [40], etc. Additionally, GAN can explicitly control the features or latent vector in a manner that in its given class includes categorical description of text [29], landmark-type conditions, and background and foreground details to generate images; these conditions make it a conditional GAN (cGAN) [25] model.

However, GAN still has the disadvantage of unstable training and mode collapse problems. GAN-based Single Image Super-Resolution (SISR) models such as SRGAN [22] generate photo-realistic results with respect to these problems. However, the loss function on the feature space makes it to sacrifice the issue of its. Simultaneously, Wasserstein GAN (WGAN) [2] and Wasserstein GAN-Gradient Penalty (WGAN-GP) [15] improved training techniques by adding a Wasserstein (Earth Mover) distance metric loss to address the issues of generator and discriminator training and gradient vanishing. They stabilize their training over an extensive range of architectures with almost no hyper-parameter tuning. It does not rely on weight clipping techniques but penalizes the model if the gradient norm moves away from its target norm value 1. The adversarial loss proposed by Gulrajani et al. [15] moved the distribution of the generated images to the distribution of the real images. Especially, we seek to generate photo-realistic as well as less blurry images of post-surgery faces. To accomplish this, we employ deep feature consistent principle [17] to generate comprehensible face images with natural eyes, teeth, hair texture, and nose. In parallel, we also use improved GAN training mechanism to generate images that lie in the manifold of natural images.

All these principles and improved GAN techniques motivate us to work in facial plastic and aesthetic surgery fields, benefiting the society with its needs and considering current trends such as forensic and security application and the recent worldwide COVID-19 face mask scenario.

3.2 Face aging and rejuvenation

Face age progression is the prediction of future looks, and rejuvenation is the estimation of younger faces also referred to as facial age regression. It significantly impacts a wide range of applications in various domains. Generative models [14] are progressively used to perform face aging and de-aging due to their unquestionable and plausible generation of natural face images with an adversarial training approach. For e.g. Zhang et al. proposed CAAE [38] for face aging and de-aging framework that learns a face manifold. Yang et al. [36] designed pyramidal adversarial discriminator for high-level face representations in a detailed manner. Wang et al. [35] presented an identity permanence conditional GAN and used pre-trained age classification loss to estimate the age correctly. Attention-cGAN [41] used the attention-based mechanism, which is a modification of only the facial regions relevant to aging. Recently, Praveen et al. [5] proposed ChildFace specific to child-based face aging and de-aging by introducing the gender and age-aware condition vector to preserve the identity in a small age span.

The aforementioned generative models used for beautification and face rejuvenation inspired us to propose PlasticGAN that integrates both face aging and rejuvenation. Moreover, it does not require any training on the dataset with before and after plastic surgery faces. We leverage the conditional GAN-based architecture integrated with adversarial, perceptual-based identity, and reconstruction loss function. The proposed model is designed as an innovative method for vivid visualization of realistic post-surgery faces to help surgeons in building up confidence and acquiring the patient's acceptance for surgery.

4 Network architecture

We propose a PlasticGAN framework. The architecture of this framework evolved from the deep feature consistency principle [18], adversarial auto encoder [23], and BlockCNN base residual block [24]. As depicted in Fig. 1, the PlasticGAN system consists of four deep networks: a) deep residual variational autoencoder including a probabilistic encoder $E(x)$, b) probabilistic generator $G(z, l)$, c) pre-trained VGG19 (Φ) [31] for identity preservation, and d) deep residual critic discriminator $D_{img}(x, \bar{x})$. This model is based on the principle of WGAN-GP [15] to improve the accuracy results for face aging and rejuvenation in terms of generating natural and realistic images.

The encoder (E) compresses the input image x having the size $128 \times 128 \times 3$ through two fully connected layers (for mean μ and variance σ), and then they are added to sample latent vector z . Furthermore, this vector z appends with identity feature maps that combine two vectors: 12 times of age vector (a) and 25 times gender vector (g). Then, the output from previous z passes as an input to the generator (G) for generating image $\bar{x}$.

Then in the next step, VGGNet takes both x and $\bar{x}$ as input and extracts deep feature from these images and then constructs the perceptual loss. Meanwhile, the generated images and the real image are conveyed to the Discriminator D_{img} . The idea is to take these two as inputs to perform the adversarial min-max game policy and to calculate the discrepancy between the generated and input images. In addition to this information, D_{img} also considers the identity feature map l (e.g. age and gender vector) in its first hidden layer as depicted in Fig. 1; these vector values are used as a conditional setting to obtain more fine-grained age and gender information between x and $\bar{x}$.

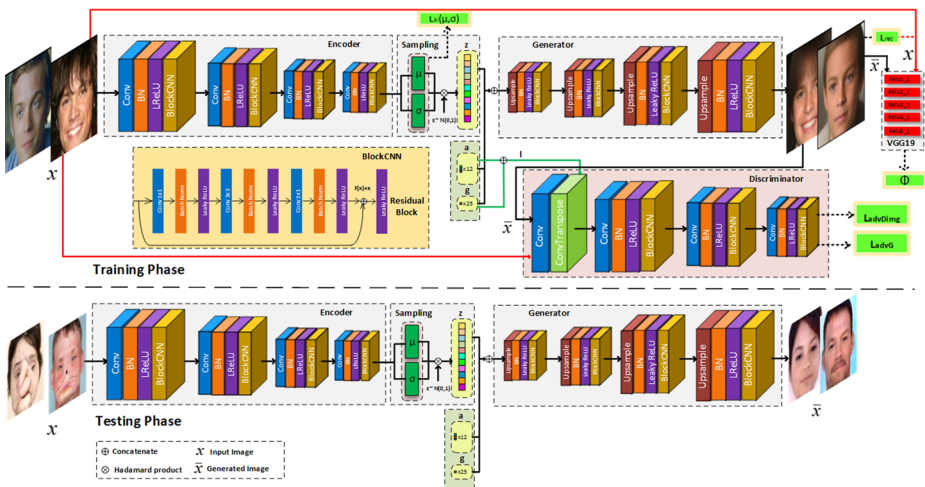


Fig. 1 Overview of training and testing phase of the proposed PlasticGAN model. Encoder (E), Generator (G), and Discriminator (D_{img}) are used for reconstruction and aesthetic surgery. $\mathcal{L}_{KL}(\mu, \sigma)$ represent the KL loss, $\mathcal{L}_{rec}$ represent the reconstruction loss, Φ represent the perceptual loss, $\mathcal{L}_{advG}$ and $\mathcal{L}_{advD_{img}}$ represent generator and discriminator adversarial loss. For simplicity, we have omitted the total variation loss $\mathcal{L}_{TV}$

E , G , and D_{img} have BlockCNN-based residual blocks after each convolution and deconvolution layer except the first one of D_{img} . These blocks are used to improve convergence behavior and remove the compression artifact [24]. The spatial differences of pre-trained VGG19 network are calculated in the middle of the layer architecture and then are combined to find total perceptual loss (Φ). This loss network is based on the principle of deep feature consistency [17] and is used to capture the most prominent image features. Our model generates age-progressed and regressed plastic surgery images with comparatively better aesthetic results in terms of the reconstruction of damaged facial parts such as nose, teeth, lips, mouth, and ear textures.

5 Objective function

The adversarial training of PlasticGAN can be considered as a two-player min-max game in which the team of probabilistic residual encoders and generators is trained against the team of residual adversarial discriminators. Both teams have to minimize five losses: 1) KL divergence loss ($\mathcal{L}_{KL}(\mu, \sigma)$) to regularize the feature vector z with the prior distribution $P(z) \sim N(0, 1)$; 2) the reconstruction loss ($\mathcal{L}_{rec}$) between input and generated images is adopted so that sparse aging outcomes are produced and the image background is preserved; 3) $\mathcal{L}_{\Phi}$ perceptual loss is computed by pre-trained high-performing CNN as VGGNet [31] the loss network captures the spatial correlation between x and $\bar{x}$ face images; 4) the aim of $\mathcal{L}_{adv}$ adversarial by incorporating the WGAN-GP into PlasticGAN is to improve the perceptual quality of the output images; 5) The total variation loss ($\mathcal{L}_{TV}$) function goal is to regularize the total variation in the generated images.

KL divergence loss $\mathcal{L}_{KL}(\mu, \sigma)$ helps the residual encoder network learn better feature space representations. For input face x image, the E network $E(x) = (\mu(x), \sigma(x))$, output

the mean μ and variance σ of the approximate posterior. To calculate the feature vector z for given x , we sample a random ϵ Gaussian distribution where $\epsilon \sim \mathcal{N}(0, 1)$. We sample the feature vector using $z = \mu + \sigma \odot \epsilon$, where $\odot$ represents element-wise multiplication. $\mathcal{L}_{KL}(\mu, \sigma)$ as shown in (1).

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

where k denotes the indexes over the dimensions of the latent vector.

Reconstruction Loss $\mathcal{L}_{rec}$ ensures the generated image preserves the low-level image content between its input x . For this, we incorporated a mean square base reconstruction loss between x and $\bar{x}$ in the image space which could be written as (2).

$$\mathcal{L}_{rec} = \|x - (G(E(x), z)l)\|_2^2 \quad (2)$$

Where G is taken in the latent vector z generated by $E(x)$ and the identity feature map (l) concatenated with z and passed as zl with input image x .

Perceptual loss Perceptual loss calculates the spatial difference between the layers of VGG19 [31] and effectively minimizes the perceptual distance between the synthesized $\bar{x}$ and input image x . Here, we denote perceptual loss by $\Phi(x)^l$ with l as the layer. Here, we exploit the intermediate activation layer feature map denoted as *relu1_1*, *relu2_1*, *relu3_1*, *relu4_1*, *relu5_1* of VGG as VGG19.

$$\Phi = \Sigma_{\gamma} \Phi^{\gamma} \quad (3)$$

In order to calculate Φ^{γ} at layer l , we use Euclidean distance between the activation map of module l for input image x and generated image $\bar{x}$. Here C , W , and H denote the number of filters, width, and height of each feature map, respectively. Φ^{γ} denotes the perceptual loss of a single layer γ .

$$\Phi^{\gamma} = \frac{1}{2 \times C^{\gamma} W^{\gamma} H^{\gamma}} \sum_{c=1}^{C^{\gamma}} \sum_{w=1}^{W^{\gamma}} \sum_{h=1}^{H^{\gamma}} (\Phi(x)_{c,w,h}^{\gamma} - \Phi(\bar{x})_{c,w,h}^{\gamma})^2 \quad (4)$$

Adversarial loss The adversarial training between the generator G and discriminator D_{img} stimulates the generated results to be realistic and identical to real ones. Besides, image generation quality and attribute immutability is also guaranteed by including the attribute of input face images as a conditional vector in adversarial training. To accomplish these two goals, our discriminator D_{img} is designed to take the input and generated images with their corresponding attributes after the first convolution block. D_{img} calculates the improved adversarial loss by discriminating between input image x and the image generated by G . Formally, the objective function for training the discriminator adversarial loss ($\mathcal{L}_{advD_{img}}$) is shown in (5).

$$\begin{aligned} \mathcal{L}_{advD_{img}} = & \mathbb{E}_{x,l \sim P_{data}(x,l)} [D_{img}(x, l)] \\ & - \mathbb{E}_{x,l \sim P_{data}(x,l)} [(D_{img}(G(E(x), l)))] \\ & - \lambda_{gp} \mathbb{E}_{\hat{x} \sim P_{\hat{x}}} [(\|\nabla_{\hat{x}} D_{img}(\bar{x})\|_2 - 1)^2] \end{aligned} \quad (5)$$

Where $\hat{x} \sim P_{\hat{x}}$ is uniformly sampled along straight lines between pairs of input x and generated $\bar{x}$ images, and λ_{gp} is the penalty coefficient to penalize the gradients $n_{critic}=5$.

The generator network G is trained to confuse D_{img} with visually conceivable synthetic images and the objective function is shown in (6).

$$\mathcal{L}_{advG} = -\mathbb{E}_{x,l \sim P_{data}(x,l)} [D_{img}(G(E(x), l))] \quad (6)$$

Total Variational Loss ($\mathcal{L}_{TV}$) Total Variational (TV) loss is used for assuring measurable continuity and smoothness in the generated image to avoid noise and sudden changes in high-frequency pixel intensities. The TV loss is the sum of the absolute differences for adjacent pixel values in the generated image. (7) shows TV loss.

$$\mathcal{L}_{TV} = \sum_{c=1}^C \sum_{w=1}^W \sum_{h=1}^H |(\bar{x})_{w+1,h} - (\bar{x})_{w,h}|^2 + |(\bar{x})_{w,h+1} - (\bar{x})_{w,h}|^2 \quad (7)$$

Overall objective To generate realistic faces while also preserving the identity corresponding to input. The final objective function for the discriminator D_{img} is shown in (8).

$$\max_{\|D_{img}\|_L \leq 1} \mathcal{L}_{D_{img}} = \lambda_{advD_{img}} \mathcal{L}_{advD_{img}} \quad (8)$$

where $\|D_{img}\|_L \leq 1$ represents the set of 1-Lipschitz constraint on D_{img} . The final objective function for the generator G is shown in (9).

$$\min_{E,G} \mathcal{L}_G = \lambda_{kl} \mathcal{L}_{KL}(\mu, \sigma) + \lambda_{rec} \mathcal{L}_{rec} + \lambda_{per} \Phi + \lambda_{advG} \mathcal{L}_G + \lambda_{tv} \mathcal{L}_{TV} \quad (9)$$

where $\lambda_{kl}, \lambda_{rec}, \lambda_{per}, \lambda_{advD_{img}}, \lambda_{advG}, \lambda_{tv}$ are the hyper-parameters that tune the weight of the above-mentioned loss function. In our model, we used $\lambda_{kl}, \lambda_{rec}, \lambda_{per}, \lambda_{tv}, \lambda_{advD_{img}}$ as 1 and λ_{advG} as 0.0001.

6 Experimental results

The primary objectives of facial plastic surgery are to reconstruct faces, remove defects, and improve the appearance of the patient or preserve the facial personality. In this section, we demonstrate the power of publicly available facial dataset on a large scale. This section is further divided into 3 subsections: (1) Description of dataset which merges the different publicly available datasets. (2) Dataset preprocessing. (3) Processing of mask-to-face mapping on surgery faces.

6.1 Dataset

To train a relevant population of diverse facial plastic surgery synthesis models, one of the key elements is to generate plausible images and reasonably aged face images from different ethnicities. Thus, we have selected [1–40] year age range images from publicly available Cross-Age Celebrity Dataset 35,450(CACD) [6], 5822 UTKFace [38], 35,484 CLF [4, 9], and 1113 Adience [12]. In total, we have used 77,889 images of size 128×128 pixels and divided the dataset into 4 age groups as [1–10], [11–20], [21–30], [31–40] as shown in Fig. 2. To test our model on plastic and aesthetic surgery face images, we have web-crawled real-world pre- and post-surgery face images. In total, there are 24 paired before and after surgery face images which are referred to as plastic-surgery testing images. The test dataset contains various types of face surgeries such as Otoplasty (ear surgery), Skin

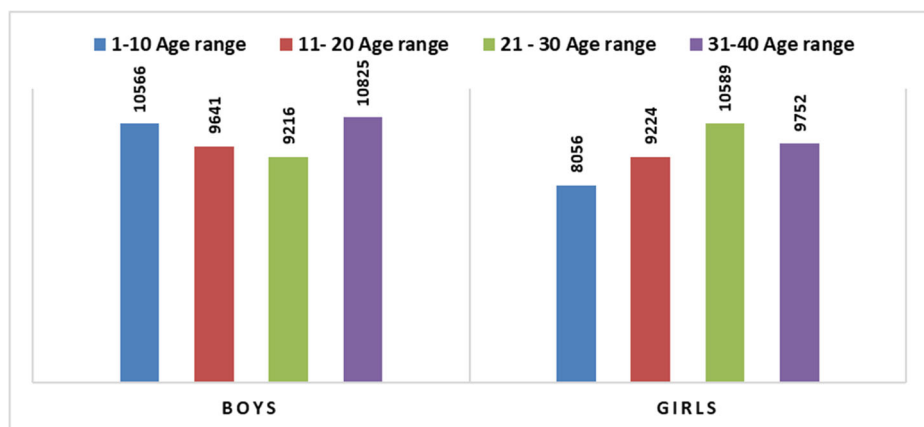


Fig. 2 Training dataset group formation: shows the distribution of face images in given age ranges to train PlasticGAN model

Resurfacing (skin peeling), Lip Augmentation, Oral Surgery (teeth surgery), Craniofacial, and Dermabrasion. The testing dataset of plastic surgery includes pairs of before and after plastic and aesthetic surgery faces as shown in Fig. 3.

6.2 Preprocessing on dataset

For training the proposed model to reconstruct damaged areas in the correct orientation, preprocessing the dataset is necessary. As the dataset images have improper alignment and different resolutions, we have used MTCNN [39] to detect the five landmarks points (two eyes, nose, and two mouth corners) used for proper alignment and for cropping the images to a resolution of 128×128 pixels as shown in Fig. 3.

6.3 Processing on Mask-to-face mapping

To test our proposed model on the mask-wearing face, we have to cover pre-surgery images with a mask; we used MTCNN [39] to crop and align pre-surgery faces. Besides, an image of 12 key points has been manually annotated on the reference mask image as shown in Fig. 4. In the final stage, it has used the face-to-mask mapping on cropped and aligned images.

6.4 Implementation details

We have trained PlasticGAN model on 77,889 images and divided these into 4 equal age categories, i.e., 1–10, 11–20, 21–30, 31–40. The architecture of our model is presented in Fig. 1. In the training phase, all components are trained with batch size 24 using ADAM [20] with hyper-parameter $\alpha = 0.0001$ and $\beta = (0.5, 0.999)$. The output of Generator (G) is restricted to $[-1, 1]$ using the tanh activation function. After 20,000 iterations, we received competent results. In the testing phase, we included plastic surgery images with and without masks, E and G are responsible for generating age-progressed and regressed facial images. The model was trained from scratch with a learning rate of E , G , and D_{imag} as 0.0001. We have optimized D_{imag} every 5 iterations and G is updated at every iteration.



Fig. 3 First and third row are examples of pre- and post-surgery images respectively where images were acquired using web-crawling. Second and fourth show the cropping and alignment effect of MTCNN[39] face detector. The types of plastic surgery procedures are given below the fourth row

7 Qualitative evaluation of plastic surgery face

In order to extensively evaluate the performance of our proposed PlasticGAN framework, we use state-of-the-art work CAAE⁴, AcGAN⁵, AIM (Age-Invariant Model)⁶ and IPCGAN⁷.

The following presents important common observations related to CAAE, AcGAN, AIM, and IPCGAN on various surgery faces. As observed, CAAE does not perform well in the aging effect and even produces artifacts and blurry results due to pixel-wise loss between the input and generated images. AcGAN utilizes an attention mechanism that only modifies the regions relevant to the aging effect. Hence, AcGAN performs poorly on plastic-surgery testing images in terms of reconstruction (teeth, face, lips), aging effect, and generate wired faces. Age-Invariant Model (AIM) though addresses the challenges of face recognition with large age variations but is not capable of generating photorealistic surgery faces with the desired aging effects. IPCGAN uses Image-to-Image translation-based generator network component so it cannot properly structure the different types of plastic surgery faces into a

⁴<https://github.com/ZZUTK/Face-Aging-CAAE>

⁵<https://github.com/JensonZhu14/AcGAN>

⁶<https://github.com/ZhaoJ9014/High-Performance-Face-Recognition/tree/master/src/>

⁷<https://github.com/dawei6875797/Face-Aging-with-Identity-Preserved-Conditional-Generative-Adversarial-Networks>

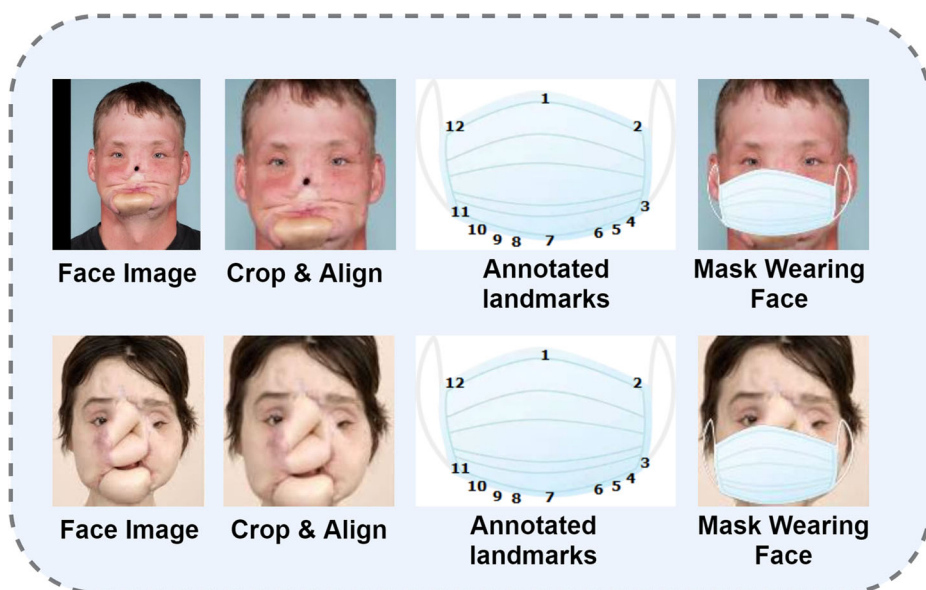


Fig. 4 First column presents the examples of pre-surgery face images where images were acquired using web-crawling. The second column shows the cropping and alignment effect of MTCNN [39] face detector. Third column references the mask images. Fourth column presents the mask-wearing face

realistic face. Compared to state-of-the-art aging frameworks, the age progress and regress images of the PlasticGAN model are photo-realistic, and it rejuvenates identity-preserved faces on plastic-surgery faces.

7.1 Evaluation I: Teeth surgery

As is evident in each dotted box 6th column of Fig. 5, the pre-surgery image is improperly aligned (crooked, missing teeth). Our model has generated a perfect-looking set of teeth and the real image for the same which is shown in post-surgery row. In 6th column, missing teeth are generated in the mouth. As seen in all age range observations, the aesthetics of the face improve with age. In addition, the respective skin texture is preserved. At the beginning of this Section 7, we have mentioned logical reasons and demonstrated why the face-aging state-of-the-art models do not perform well on surgery faces in terms of post-surgery looks.

7.2 Evaluation II: Face surgery

As shown in Fig. 6, the pre-surgery image has no nose and mouth. However, PlasticGAN generated both the missing components perfectly by adhering to structure and texture of the face and even improved the appearance of eyes, producing a youthful appearance. AIM addresses the challenges of cross-age face recognition with large age, however, it is not capable of generating visually appealing faces with the desired aging effect. IPCGAN and AcGAN generator network components translate the input to image space and reconstruct from this space; hence, they cannot properly structure a before-surgery face into an after-surgery face as shown in Fig. 6. Due to this reason, these frameworks are not very helpful to clinical decision-making between doctor and patients.

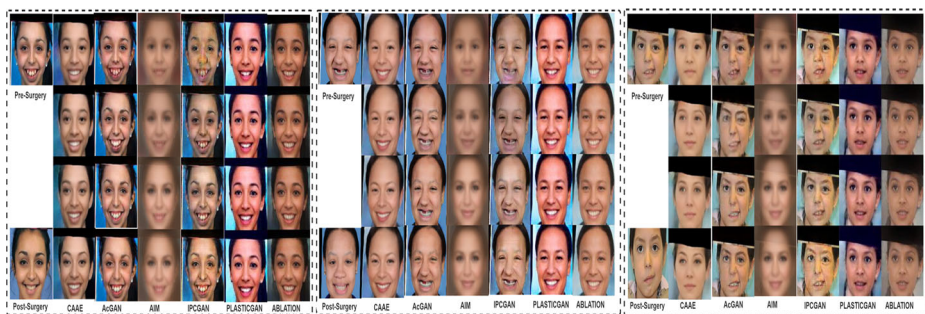


Fig. 5 Teeth Surgery: Each dotted box denotes one person's pre- and post-teeth surgery images. In each box, from second column left to right, are images generated by CAAE, AcGAN, AIM, IPCGAN, PlasticGAN, and ablation study, respectively

7.3 Evaluation III: Ear surgery

In Fig. 7, the pre-surgery image shows an ear lunging outwards. In PlasticGAN, the generated images are aligned perfectly to the normal settings. In addition, PlasticGAN constructs the internal structure of the ear compared to the state-of-the-art models. Therefore, the post-surgery images resembles the age progress and regress images. In the case of IPCGAN and AcGAN, only the face region is altered. Therefore, the age progress and regress ear surgery images do not properly align.

7.4 Evaluation IV: Lips surgery

In Fig. 8, the pre-surgery image contains a partial lip and a deformed face structure. PlasticGAN input this image and completed the lip as well as produced open eyes, depicting how the child will look in the future. In addition, PlasticGAN performed age translation and beautifies the entire face, which enhances the rejuvenation effects. Compared to our framework, CAAE produces over-smoothed surgery images with subtle changes of appearance.

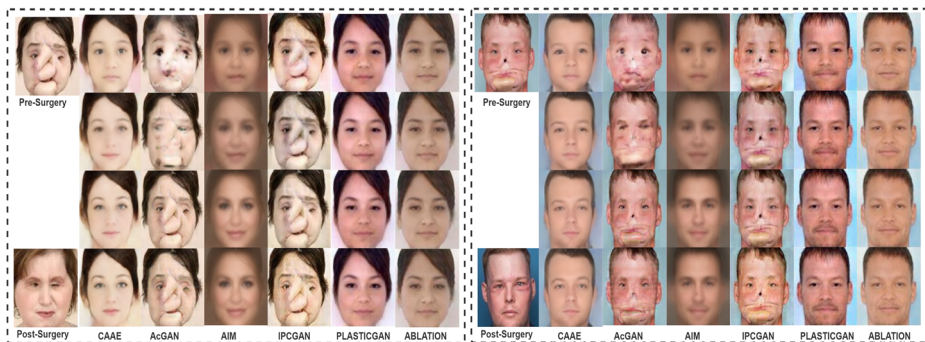


Fig. 6 Face Surgery: Each dotted box denotes one person's pre- and post-surgery facial images. In each box, from second column left to right, are images generated by CAAE, AcGAN, AIM, IPCGAN, PlasticGAN, and ablation study, respectively

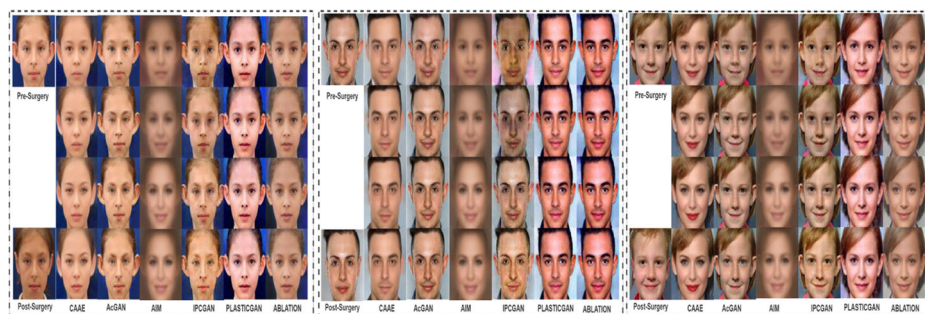


Fig. 7 Ear Surgery: Each dotted box denotes one person's pre- and post-ear surgery images. In each box, from the second column left to right, are images generated by CAAE, AcGAN, AIM, IPCGAN, PlasticGAN, and ablation study, respectively

As for IPCGAN and AcGAN, due to their incapability of face completion, faces generated by these models are deformed as evident in Fig. 8.

8 Model robustness on mask wearing face

To check the robustness of the proposed model, we covered the nose and mouth areas with synthetic masks and checked the aging effect on overall plastic surgery face. As shown in Fig. 9, IPCGAN and AcGAN models could not remove the face mask, could not complete surgery tasks, and could not show the aging effect. In case of CAAE, the face mask region can be seen with a little transparency which causes artifact in generated images. These effect due to pixel-wise loss. As evident from the results for PlasticGAN, it performed well overall in the context of various parameters such as skin tone, hair color, open eyes, reconstruction, and lighter to darker beard appearance. In addition to these, it generated better facial structures with the aging effect.



Fig. 8 Lip Surgery: Each dotted box denotes one person's pre- and post-lip surgery images. In each box, from second column left to right, are images generated by CAAE, AcGAN, AIM, IPCGAN, PlasticGAN, and ablation study, respectively

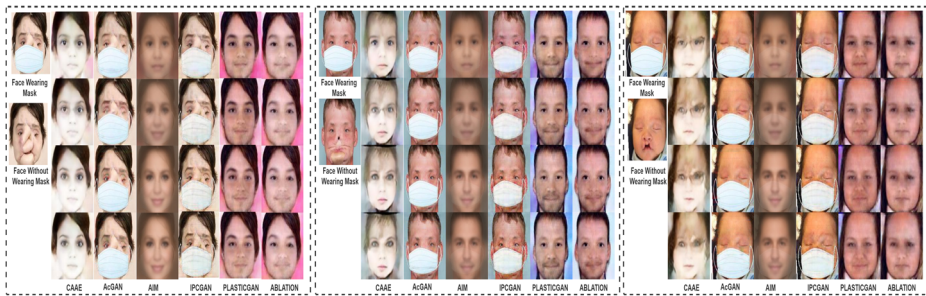


Fig. 9 Mask Wearing face: Each dotted box denotes the same individual's mask-wearing and pre-surgery images. In each box, from second column left to right, are images generated by CAAE, AcGAN, AIM, IPCGAN, PlasticGAN, and ablation study, respectively

9 Quantitative evaluation

Most of the existing age estimation and face verification approaches have primarily focused on unconstrained face recognition and no endeavor has been made to examine their effect on synthesized face aging and rejuvenation on local and global plastic surgery faces as well as mask-wearing surgery faces. As surgery-based aging and rejuvenation procedures becomes more and more prevalent, face verification framework fails to recognize individual's faces after surgery. In this section, we explore age estimation and verification aspect on synthesized surgery faces.

We have evaluated the aging and identity permanence accuracy on age progress and regress on plastic surgery face with/without the mask. For this, we have generated all age range [1–10], [11–20], [21–30], [31–40] faces from pre-surgery faces with/without the mask. Then, we used the online face analysis tool known as Face++ API [13] to estimate the age distribution and face verification scores. We considered twenty-four test faces and the following protocol used for our comparison:

Face++ test: [test face, progress-face1], [test face, progress -face2], [test face, progress-face3], [test face, progress-face3], [test face, progress-face4]. (where test face is pre-plastic surgery faces with/without mask).

9.1 Age estimation

Identically, age estimation was conducted to measure aging and deaging accuracy.

9.1.1 Plastic surgery face without mask

Following are the observations on plastic surgery face without mask shown in Table 1. IPCGAN and AcGAN introduce an identity-preserved loss and an age classification loss with the Image-to-Image translation-based generator network. However, if the classification error value is high then the gradient for small age range is not accurate. Therefore, these models' age estimation accuracy is lower compared to PlasticGAN in most age ranges. AIM generated similar-looking and aged faces as depicted in Figs. 5, 6, 7, and 8. Therefore, this model's age estimation standard deviation values are high in all age ranges. In case of CAAE, the generated images are blurry owing to the fact that the generated faces in the age

Table 1 Estimated Age Distribution (years) on Plastic-surgery testing images by PlasticGAN, ablation, and state-of-the-art models. For simplicity, we only address the mean and standard deviation of age estimation error computed over all age ranges

Age group	1-10	11-20	21-30	31-40
CAAE [38]	14.41 ± 7.58	20.50 ± 5.18	27.22 ± 5.18	28.58 ± 6.34
AcGAN [41]	16.13 ± 14.13	34.96 ± 17.00	32.83 ± 19.33	41.26 ± 18.55
AIM [40]	24.00 ± 10.40	40.44 ± 7.42	36.67 ± 9.15	42.11 ± 7.42
IPCGAN [35]	13.81 ± 14.45	17.24 ± 9.01	22.14 ± 13.01	12.56 ± 7.66
PlasticGAN	16.08 ± 5.16	24.36 ± 8.68	32.33 ± 18.66	28.80 ± 7.83
Ablation	15.79 ± 9.11	17.63 ± 9.70	20.96 ± 11.25	23.00 ± 12.44

ranges (1–10, 11–20) are aged. PlasticGAN model provides better age estimation results in three age ranges out of four compared to other state-of-the-art models.

9.1.2 Plastic surgery face with mask

CAAE uses the mean square-based reconstructed loss, and AcGAN and IPCGAN generator architecture are based on image-to-image translation network. AIM disentangles the age and identity attributes. Consequently, the progress and regress images with face mask are completely deformed as they detect local and global face attributes e.g, nose, lips, mouth, eyes completely deformed. Therefore, state-of-the-art model generated the aged face in age ranges (1–10, 11–20, 21–30) compared to PlasticGAN as illustrated in Table 2.

Note: Face images generated by AcGAN and IPCGAN are not detected by Face++ app because due to the fact that they did not remove the face mask region and due to unstructured eye construction.

From the result in Tables 1 and 2, we have the following observation.

1. The age estimation results of no mask compared to with mask are better.
2. IPCGAN and AcGAN do not unmask the surgery face as shown in Fig. 9. Therefore, the aging pattern only reflects the periocular region and forehead. Thus, the age estimation value is not properly distributed over the age ranges as shown in Table 2.

Table 2 Estimated Age Distribution (years) on Plastic surgery with mask, testing images using PlasticGAN, ablation, and state-of-the-art models. For simplicity, we only address the mean and standard deviation of age estimation error computed over all age ranges

Age group	1-10	11-20	21-30	31-40
CAAE [38]	30.52 ± 10.56	35.19 ± 9.62	38.42 ± 8.52	41.14 ± 9.49
AcGAN [§] [41]	25.00 ± 6.21	24.42 ± 3.99	26.84 ± 4.65	29.89 ± 5.67
AIM [40]	22.31 ± 7.90	40.77 ± 4.50	40.59 ± 5.60	45.50 ± 5.06
IPCGAN [#] [35]	26.75 ± 5.03	29.25 ± 7.23	30.58 ± 9.87	17.00 ± 8.00
PlasticGAN	20.50 ± 6.99	25.95 ± 8.13	30.66 ± 8.72	28.63 ± 6.55
Ablation	17.50 ± 9.49	20.77 ± 9.81	24.77 ± 9.65	27.50 ± 10.34

[#]: 5 face is not detected, [§]: 12 face is not detected

9.2 Face verification on surgery faces with and without mask

For identity permanence, face verification rates are reported along with threshold set to $76.5 @ FAR = 10^{-5}$ experimented according to the Face++ API [13]. A confidence score is then obtained for each comparison, demonstrating the similarity between two faces. The confidence range lies under [0–100]. A higher confidence score indicates a higher probability that two faces (real and generated) are from the same subject.

9.2.1 Plastic surgery face without mask

With plastic surgery, it is evident that PlasticGAN outperforms over AcGAN and AIM. Although, CAAE in the context of identity information generates surgery faces with ghosting artifacts. PlasticGAN, AIM, and CAAE aging models generate age progress and regress face by disentangling the age and personality features from the latent vector, due to which identity is also altered with age. Thus, face verification accuracy is low compared to IPCGAN and AcGAN.

9.2.2 Plastic surgery face with wearing mask

To verify the robustness and stability of the proposed model, even if the mask covers a region of the plastic surgery face, the features of the upper half of the face, such as eyes, eyebrow, and forehead can still be used to improve the masked cross-face recognition (MCFR). The experiment results are shown in Fig. 9; PlasticGAN can still progress and reconstruct a complete face. In AcGAN and IPCGAN, the network component is unable to unmask the masked face. These models alter the regions particularly relevant to face aging. This is why the face verification accuracy is high compared to other models as shown in Table 4. Objectively speaking, a few progress faces have distortion. Thus, the Face++ APP is unable to detect. Following are general observations from Tables 3 and 4.

1. Without wearing mask cross-age face verification score is improved with wearing face mask in all age ranges and face aging model.
2. With increasing the age gap, the face verification score is decreased in all age ranges.
3. The face verification score of CAAE and AIM models are lower than PlasticGAN.
4. IPCGAN and AcGAN do not unmask the surgery face as shown in Fig. 9. Owing this, the face verification accuracy is better of these models with the no-mask surgery face.

Table 3 Face verification results (in %) on Plastic surgery testing images by PlasticGAN and other state-of-the-art models

	1-10	11-20	21-30	31-40
Verification rate (Threshold=76.5 FAR=1e-5)				
CAAE [38]	69.22 ± 14.42	60.96 ± 12.18	54.56 ± 11.88	49.24 ± 12.22
AcGAN [41]	94.47 ± 4.09	94.02 ± 2.81	92.14 ± 6.56	88.98 ± 10.49
AIM [40]	65.45 ± 11.20	57.96 ± 12.18	48.75 ± 14.07	46.56 ± 13.83
IPCGAN [35]	91.63 ± 0.32	91.66 ± 0.49	95.01 ± 0.60	93.55 ± 1.63
PlasticGAN	73.63 ± 8.64	69.27 ± 8.87	66.53 ± 9.33	63.05 ± 9.68
Ablation	71.28 ± 10.97	67.13 ± 12.01	65.95 ± 12.30	64.55 ± 9.01

Table 4 Face verification results (in %) on plastic surgery with mask; testing images by PlasticGAN and other state-of-the-art models

	1-10	11-20	21-30	31-40
Verification rate (Threshold=76.5 FAR=1e-5)				
CAAE [38]	61.52 ± 9.19	58.83 ± 9.36	55.81 ± 9.40	52.12 ± 9.37
AcGAN [§] [41]	89.16 ± 1.27	88.30 ± 1.49	87.29 ± 2.28	84.61 ± 3.51
AIM [40]	72.03 ± 7.20	67.79 ± 8.54	58.93 ± 11.36	57.49 ± 11.55
IPCGAN [#] [35]	93.88 ± 0.76	93.48 ± 0.78	92.03 ± 3.57	90.30 ± 7.79
PlasticGAN	62.53 ± 8.11	61.03 ± 8.87	60.53 ± 9.13	58.05 ± 9.68
Ablation	60.26 ± 9.48	58.23 ± 9.81	55.11 ± 9.49	53.41 ± 10.04

[#]: 12 face is not detected, [§]: 5 face is not detected

Despite the potential of our proposed model, we can conclude that the task of cross-age face recognition on these synthetic faces after applying surgery becomes challenging as it degrades the verification score. This challenge, namely PSBSCAFR, can become a new dimension for upcoming research regarding how to improve the recognition accuracy of synthetic surgery faces generated by GAN.

9.3 Comparison between surgery face with and without mask

Traditional CNN-based face recognition systems trained on existing datasets are almost ineffective on faces that have undergone surgery or that are wearing a mask. Simultaneously, new challenges create new opportunities and research direction in this field. The one which we want to include in this study is how plastic surgery face and facial mask face can be correlated based on the idea of and considering face verification of same individuals. During our experiment, we observed a few significant aspects which can also be seen in the Fig. 10. We have conducted this study by Face++ App.

1. In many cases, masked faces are not even detected when the eyes are closed and face is not properly aligned. (In block 2 and 4 of Fig. 10).
2. Different types of face surgery with mask leads to many differences in the impact on face recognition reliability. When it is the case of the mask covering the lip and half nose region, the confidence score is upgraded because eyes are an important consideration at the time of verification. This score is also affected by same mask covering. When it is the case of the above-mask region, the score is degraded due to the fact that the recognition system does not find any similarity points while verification except the mask. (In block 1 and 3 of Fig. 10).

9.4 Inception and Fréchet inception distance

The image quality with the diversity of the generated data is assessed in terms of the inception score (IS) [30] and the Fréchet inception distance (FID) [16]. In Table 5, the PlasticGAN achieves best IS and FID scores on plastic surgery test datasets compared to the state-of-the-art models. Meanwhile, high IS and low FID score indicate that our framework generates more realistic faces.

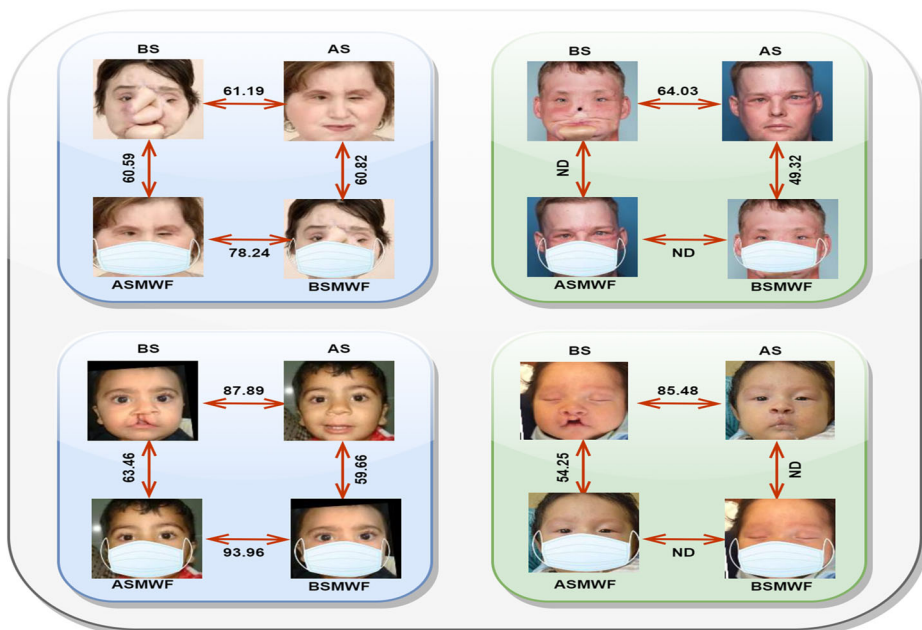


Fig. 10 Each block is an example of pre- and post-surgery face with and without mask. We have evaluated the confidence score with the help of Face++ app. ND represents the cases of no detection because the face is covered with mask and showing non similarity

9.5 Beauty Score and Gender Prediction

We evaluate beauty score and gender prediction experiment of PlasticGAN and state-of-the-art models on plastic surgery test images. For fair comparison, Face++ [13] as is used as a face analysis tool to evaluate the beauty score and gender prediction of pre-plastic surgery face and corresponding age progress in all age ranges as shown in Table 6.

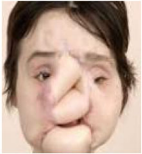
General observations are as follows:

1. CAAE generates images introducing grain-like artifact, which deteriorates image quality. Due to this reason, wrong gender prediction is shown in red color.
2. IPCGAN and AcGAN use an image-to-image translation-based generator network component. Hence, it cannot properly restructure a partially covered face into a realistic

Table 5 Comparing IS and FID on PlasticGAN and its variant with other state-of-the-art models

Model	IS	FID
CAAE [38]	2.10 ± 0.25	144.04
AcGAN [41]	2.38 ± 0.42	72.28
AIM [40]	1.41 ± 0.09	154.02
IPCGAN [35]	2.03 ± 0.34	74.43
PlasticGAN	2.40 ± 0.19	71.66
Ablation	2.18 ± 0.14	107.05

Table 6 Beauty score and gender prediction: The second column contains pre-surgery images followed by beautification score with gender prediction values. In column 4, the state-of-the-art model, PlasticGAN, and Ablation study are shown

S.No	Pre-Surgery Image	PBSGP	Model	GBSGP 1-10	GBSGP 11-20	GBSGP 21-30	GBSGP 31-40
1		74.21(M)	CAAE [38] AcGAN [41] AIM [40] IPCGAN [35] PlasticGAN Ablation	74.86(F) 76.93(M) 79.818(F) 74.87(M) 71.39(M) 70.30(M)	76.11(F) ND 81.86(F) 79.59(M) 71.46(M) 68.87(M)	73.32(M) 74.51(M) 82.12(F) 74.78(M) 69.19(M) 69.26(M)	72.50(M) 75.25(M) 84.07(M) 69.00(M) 70.23(M) 68.67(M)
2		73.27(M)	CAAE [38] AcGAN [41] AIM [40] IPCGAN [35] PlasticGAN Ablation	79.10(F) 75.17(M) 75.17(M) 67.51(M) 76.48(M) 74.73(M)	78.22(F) 77.12(F) 77.13(F) 73.34(M) 76.17(M) 74.92(M)	77.09(F) 71.82(M) 71.83(M) 70.90(M) 74.99(M) 73.79(M)	77.94(F) 68.26(F) 68.26(F) 66.45(M) 73.43(M) 73.79(M)
3		72.56(F)	CAAE [38] AcGAN [41] AIM [40] IPCGAN [35] PlasticGAN Ablation	76.49(F) 76.94(M) 77.64(F) 72.45(F) 75.49(F) 78.68(M)	78.04(F) 71.02(F) 79.52(F) 74.95(F) 77.38(F) 79.15(M)	79.44(F) 72.09(F) 77.16(F) 70.89(F) 79.81(F) 79.14(M)	77.69(F) 68.52(F) 75.49(F) 70.10(F) 77.62(F) 76.10(M)
4		74.83(M)	CAAE [38] AcGAN [41] AIM [40] IPCGAN [35] PlasticGAN Ablation	79.66(F) 70.37(M) 80.38(F) 68.58(M) 76.70(M) 78.12(M)	81.08(F) 78.73(M) 80.57(F) 70.83(M) 80.77(M) 79.56(M)	82.56(F) 72.08(M) 83.71(M) 71.23(M) 78.29(M) 76.07(M)	82.22(F) 74.36(M) 84.03(M) 69.72(M) 75.59(M) 76.16(M)
5		89.61(F)	CAAE [38] AcGAN [41] AIM [40] IPCGAN [35] PlasticGAN Ablation	84.23(F) 89.86(F) 79.85(F) 84.23(F) 90.49(F) 81.98(F)	80.99(F) 88.83(F) 83.30(F) 80.99(F) 90.39(F) 82.16(F)	81.31(F) 88.92(F) 82.47(F) 81.31(F) 87.58(F) 80.95(F)	78.91(F) 85.09(F) 81.51(F) 78.91(F) 87.01(F) 78.93(F)

This is followed by generated beautification score with gender of all age ranges from column 5 to 8

Note: PBSGP: pre-surgery beauty score and gender prediction, GBSGP: generated beauty score and gender prediction, ND: Not Detected face, M: male, F: female, highlighted incorrectly predicted gender in bold

- face. Owing to this, the destroyed face is not detected by Face++ App as mentioned (Not Detected) **ND** in Table 6.
- PlasticGAN is better at overcoming ghosting artifacts and color distortions. Further, it maintains uniformity in the background and face boundaries as well as shows comparative the beauty score with other state-of-the-art models.
- The pre-surgery test images' gender prediction value is either male or female. However, the generated images' gender prediction value is changed from the ground truth. Compare to PlasticGAN, state-of-the-art models are predicting incorrectly.

5. In ablation study (Without KL loss), the skin color is shown lighter compared to PlasticGAN as shown in Figs. 5, 6, 7 and 8. Due to this effect, the beauty score value is low compared to PlasticGAN.

10 Ablation study

To comprehend the effect of $\mathcal{L}_{KL}(\mu, \sigma)$ over our proposed model, we conducted an experiment on variants of the PlasticGAN model by removing $\mathcal{L}_{KL}(\mu, \sigma)$ (sampling block). This effect can be easily seen in Figs. 5, 7, 6, and 8 for visual comparison. We have observed throughout the process that PlasticGAN produces artifacts-free age-progressed and regressed faces and applies some effects of beautification as well. As shown in the Tables 1, 2, 3, 4, 5, and 6, the KL loss helps in face verification, age estimation, fidelity generation, and gender preservation with beautification score. This further elucidates our objective functions and network components that are designed well for face aging and rejuvenation based on social and forensic applications.

11 Conclusions and future research work

The advancement of generative models in beautification and rejuvenation has inspired and motivated us to propose the robust and general PlasticsGAN framework. This model integrates face aging and rejuvenation, face recognition, and face completion which relies on plastic and aesthetic facial surgery cases. This can contribute to a wide range of applications such as surgeon and patient consultancy, forensics and security, digital entertainment, and even the fashion and wellness industry.

Furthermore, PlasticGAN unmask the mask wearing face and properly structures it with aging/deaging effect. Moreover, the PlasticGAN framework does not require pre- and post-plastic surgery faces as a paired dataset during training. In the testing phase, our model parallelly synthesized face aging, rejuvenation, and face completion on faces that had undergone surgery. From the concluded qualitative and quantitative experiments and from the comparison with state-of-the-art face aging architectures on various plastic surgery faces (teeth, face, ear, lips), it was found that our model is robust and has diverse applications especially in the case of aging and rejuvenation with face completion.

As future work, we would like to enhance the framework's performance by analyzing face aging and rejuvenation entailed in plastic surgery. This can further degrade a commercial and publicly available face recognition systems performance. When these co-occur with other factors, e.g different types of mask-wearing and synthetic surgery face. This can be a new dimension for future work.

Acknowledgements 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 would like to thank Anjali Vijayvargiya for her insight and helpful comments.

Declarations

Disclosure of potential conflicts of interest We declare that we have no financial and personal relationships with other people or organization that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed

as influencing the position presented in, or the review of, the manuscript entitled. “PlasticGAN: Holistic Generative Adversarial Network on Face Plastic and Aesthetic Surgery”.

References

1. Abozaid A, Haggag A, Kasban H (2019) Multimodal biometric scheme for human authentication technique based on voice and face recognition fusion. In: *Multimedia tools and applications*
2. Arjovsky M, Chintala S, Bottou L (2017) Wasserstein generative adversarial networks. In: *ICML*, vol 70, pp 214–223
3. Chandaliya PK, Kumar V, Harjani M, Nain N (2020) Scdae: Ethnicity and gender alteration on clf and utkface dataset. In: *Computer vision and image processing*, pp 294–306. Springer singapore
4. Chandaliya PK, Nain N (2019) Conditional perceptual adversarial variational autoencoder for age progression and regression on child face. In: *12Th international conference on biometrics (ICB)*, pp 1–8
5. Chandaliya PK, Sinha A, Nain N (2020) Childface: Gender aware child face aging. In: *BIOSIG 2020*, pp 255–263
6. Chen BC, Chen CS, Hsu WH (2014) Cross-age reference coding for age-invariant face recognition and retrieval. In: *(ECCV)*
7. Chen HJ, Hui KM, Wang SY, Tsao LW, Shuai HH, Cheng WH (2019) Beautyglow: On-demand makeup transfer framework with reversible generative network. In: *IEEE CVPR*
8. Deb D, Aggarwal D, Jain K (2020) A.: Identifying missing children: Face age-progression via deep feature aging. In: *ICPR*
9. Deb D, Nain N, Jain AK (2018) Longitudinal study of child face recognition. In: *2018 International conference on biometrics, ICB*, pp 225–232
10. Deng J, Zaferiouri S (2019) Arcface for disguised face recognition. In: *(ICCV)*
11. Di X, Patel VM (2018) Face synthesis from visual attributes via sketch using conditional vaes and gans
12. Eidinger E, Enbar R, Hassner T (2014) Age and gender estimation of unfiltered faces. In: *IEEE-TIFS*, vol 9, pp 2170–2179
13. Face++ research toolkit (2013) <https://www.faceplusplus.com>
14. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley EA (2014) Generative adversarial nets. In: *NIPS*, pp 2672–2680
15. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville AC (2017) Improved training of wasserstein gans 30
16. Heusel M, Ramsauer H, Unterthiner T et al (2017) Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: *NIPS*, pp 6629–6640
17. Hou X, Shen L, Sun K, Qiu G (2017) Deep feature consistent variational autoencoder. In: *WACV*, pp 1133–1141
18. Hou X, Sun K, Shen L, Qiu G (2019) Improving variational autoencoder with deep feature consistent and generative adversarial training. *Neurocomputing* 341:183–194
19. Karras T, Laine S, Aila T (2019) A style-based generator architecture for generative adversarial networks. In: *CVPR*, pp 4396–4405
20. Kingma DP, Ba J (2015) Adam a method for stochastic optimization. In: *ICLR*
21. Kushwaha R, Nain N (2019) Person-verification using geometric and haralick features of footprint biometric. In: *Multimedia tools and applications*
22. Ledig C, Theis L, Huszar F, Caballero J et al (2017) Photo realistic single image super resolution using a generative adversarial network. In: *CVPR*, pp 105–114
23. Makhzani A, Shlens J, Jaitly N, Goodfellow IJ (2015) Adversarial autoencoders
24. Maleki D, Nadalian S, Mahdi D (2018) e.a.: Blockcnn : A deep network for artifact removal and image compression. In: *IEEE CVPRW*
25. Mirza M, Osindero S (2014) Conditional generative adversarial nets. *arXiv:1411.1784*
26. Nappi M, Ricciardi S, Tistarelli M (2016) Deceiving faces: When plastic surgery challenges face recognition. *Image Vis Comput* 54:71–82
27. Ngan M, Grother P, Hanaoka K (2021) Ongoing face recognition vendor test (frvt) part 6b: Face recognition accuracy with face masks using post-covid-19 algorithms. In: *NIST Report*
28. Rathgeb C, Dantcheva A, Busch C (2019) Impact and detection of facial beautification in face recognition: An overview. *IEEE Access* 7:152667–152678
29. Reed SE, Akata Z, Yan X, Logeswaran L et al (2016) Generative adversarial text to image synthesis 48, pp 1060–1069

30. Salimans T, Ian J, Goodfellow WZ, Cheung V, Radford A, Chen X (2016) Improved techniques for training gans 29
31. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: ICLR
32. Singh R, Vatsa M, Bhatt HS, Bharadwaj S, Noore A, Nooreydzan SS (2010) Plastic surgery: a new dimension to face recognition. *IEEE Transactions on Information Forensics and Security* 5(3):441–448
33. Singh R, Vatsa M, Noore A (2009) Effect of plastic surgery on face recognition: a preliminary study. In: *CVPRW*, pp 72–77
34. Suri S, Sankaran A, Vatsa M, Singh R (2018) On matching faces with alterations due to plastic surgery and disguise. In: *BTAS*, pp 1–7
35. Wang Z, Tang X (2018) W.L., Gao, S.: Face aging with identity-preserved conditional generative adversarial networks (*CVPR*)
36. Yang H, Huang D, Wang Y, Jain AK (2018) Learning face age progression: a pyramid architecture of gans. In: *CVPR*, pp 31–39. *IEEE Computer Society*
37. Yang L, Song Q, Wu Y (2021) Attacks on state-of-the-art face recognition using attentional adversarial attack generative network. In: *Multimedia tools and applications*
38. Zhang Z, Song Y, Qi H (2017) Age progression/regression by conditional adversarial autoencoder. In: *CVPR*, pp 4352–4360. *IEEE Computer Society*
39. Zhang K, Zhang Z, Li Z, Qiao Y (2016) Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process Lett* 23(10):1499–1503
40. Zhao J, Cheng Y, Cheng Y (2019) Yang.: Look across elapse: Disentangled representation learning and photorealistic cross-age face synthesis for age-invariant face recognition AAAI
41. Zhu H, Huang Z, Shan H, Zhang J (2020) Look globally age locally: Face aging with an attention mechanism

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