



## OPEN ACCESS

## EDITED BY

Naveen Aggarwal,  
UIET, Panjab University, India

## REVIEWED BY

Suhas M. V.,  
Manipal Institute of Technology, India  
Federica Marcolin,  
Polytechnic University of Turin, Italy

## \*CORRESPONDENCE

M. S. Sannidhan,  
✉ sannidhan@nitte.edu.in

RECEIVED 14 December 2023

ACCEPTED 17 May 2024

PUBLISHED 06 June 2024

## CITATION

Martis JE, Sannidhan MS, Pratheeksha Hegde N and Sadananda L (2024), Precision sketching with de-aging networks in forensics. *Front. Sig. Proc.* 4:1355573. doi: 10.3389/frsip.2024.1355573

## COPYRIGHT

© 2024 Martis, Sannidhan, Pratheeksha Hegde and Sadananda. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Precision sketching with de-aging networks in forensics

Jason Elroy Martis<sup>1</sup>, M. S. Sannidhan<sup>2\*</sup>, N. Pratheeksha Hegde<sup>1</sup> and L. Sadananda<sup>3</sup>

<sup>1</sup>Department of ISE, NMAM Institute of Technology (Nitte Deemed to be University), Udupi, India,

<sup>2</sup>Department of CSE, NMAM Institute of Technology (Nitte Deemed to be University), Udupi, India,

<sup>3</sup>Department of CSE, Shri Madhwa Vadiraja Institute of Technology and Management, Udupi, India

Addressing the intricacies of facial aging in forensic facial recognition, traditional sketch portraits often fall short in precision. This study introduces a pioneering system that seamlessly integrates a de-aging module and a sketch generator module to overcome the limitations inherent in existing methodologies. The de-aging module utilizes a deepfake-based neural network to rejuvenate facial features, while the sketch generator module leverages a pix2pix-based Generative Adversarial Network (GAN) for the generation of lifelike sketches. Comprehensive evaluations on the CUHK and AR datasets underscore the system's superior efficiency. Significantly, comprehensive testing reveals marked enhancements in realism during the training process, demonstrated by notable reductions in Frechet Inception Distance (FID) scores (41.7 for CUHK, 60.2 for AR), augmented Structural Similarity Index (SSIM) values (0.789 for CUHK, 0.692 for AR), and improved Peak Signal-to-Noise Ratio (PSNR) metrics (20.26 for CUHK, 19.42 for AR). These findings underscore substantial advancements in the accuracy and reliability of facial recognition applications. Importantly, the system, proficient in handling diverse facial characteristics across gender, race, and culture, produces both composite and hand-drawn sketches, surpassing the capabilities of current state-of-the-art methods. This research emphasizes the transformative potential arising from the integration of de-aging networks with sketch generation, particularly for age-invariant forensic applications, and highlights the ongoing necessity for innovative developments in de-aging technology with broader societal and technological implications.

## KEYWORDS

facial aging, de-aging networks, sketch generation, forensic facial recognition, deepfake-based neural network, generative adversarial network

## 1 Introduction

Facial aging is a multifaceted process influenced by internal factors like genes and hormones, as well as external factors such as Sun exposure and smoking. These elements impact the aging pace, resulting in varied outcomes among individuals. Facial changes involve alterations in skin elasticity, muscle tone, bone density, and fat distribution. Skin develops wrinkles, muscles lose tone, bones become less dense, and fat tissues shift, affecting facial shape (Bocheva et al., 2019). These transformations pose challenges for facial recognition systems, especially in applications like security and forensics, where accuracy is crucial. Understanding facial aging extends beyond individual health, holding broader implications for societal and technological advancements (Kyllonen and Monson, 2020; Donato et al., 2023).

Forensic agencies use sketch portraits, created from eyewitness descriptions, to aid in criminal apprehension. These depictions, crafted by forensic artists or software, employ techniques like composite sketches, facial approximation, or facial reconstruction. Sketch portraits help narrow down suspects, generate leads, and seek public assistance. Despite their utility, challenges arise in accurately identifying past criminals (Kokila et al., 2017a; Kokila et al., 2017b; Pallavi et al., 2018). Sketch portraits often lack precision, completeness, or fail to reflect changes over time, like aging, hairstyles, facial features, scars, tattoos, or cosmetic alterations, leading to potential inaccuracies in identification (Jin et al., 2018).

De-aging networks, a form of deep learning, generate realistic images with a younger appearance, aiding forensic investigations by enabling accurate face recognition from old sketch comparisons (Rafique et al., 2021; Atkale et al., 2022). For instance, if a crime suspect is identified only through an aged sketch, de-aging networks can produce a younger version for comparison, improving the chances of identification. These networks can also enhance the quality and resolution of old sketches for improved face recognition. Techniques include facial landmark detection, facial attribute manipulation, facial expression preservation, and age progression reversal. However, challenges like ethical concerns, privacy issues, data availability, and model robustness exist. De-aging accuracy depends on relevant data, as diseases or disorders may alter a person's appearance. Sketches simplify facial recognition by highlighting essential features, abstracting complex details, enhancing contrast, and preserving identity and expression, benefiting de-aging networks (Abate et al., 2007).

Given the significant challenges posed by facial aging in accurately identifying individuals in forensic contexts, there is a pressing need for innovative solutions that enhance the precision and reliability of facial recognition systems over time. Our proposed system, which integrates de-aging networks with sketch generation, offers a novel approach to address these challenges. Specifically, it holds promise for applications in forensic science, such as aiding in the identification of missing persons over time and revisiting cold cases with updated visual records, thereby potentially transforming forensic methodologies and outcomes. The de-aging module employs a deepfake-based neural network that can de-age faces of any gender, race, and culture, by eliminating blemishes and wrinkles, and enhancing facial features. The sketch generator module utilizes a pix2pix-based GAN that can convert photos into sketches, by extracting and preserving the structural and contrast details of the faces (Sannidhan et al., 2019). Our system can produce both hand-drawn and composite sketches, which are more realistic and consistent than existing methods. We assess our system on two standard datasets (Sannidhan et al., 2019), Chinese University of Hong Kong (CUHK) and Aleix Martinez and Robert Benavente (AR) and exhibit its efficiency and supremacy over other contemporary methods.

In view of the design of the proposed system, the following are the key contributions of this research article:

- 1) Integration of de-aging networks with sketch generation for advanced forensic facial recognition capabilities, especially for age-invariant applications.
- 2) Utilization of a pix2pix-based GAN for converting photos into enhanced sketches that are more realistic and consistent.
- 3) Demonstration of enhanced performance and accuracy on CUHK and AR Face Sketch Databases outperforming contemporary methods in terms of realism and accuracy during the training process.
- 4) To support diverse demographic attributes, ensuring broad applicability in forensic applications.

## 2 Related works

Recent developments in facial recognition technology showcase a broad range of interdisciplinary techniques, extending from the creation of innovative neural architectures to the implementation of advanced imaging methods that elucidate aspects of the aging process. Gupta and Nain (2023) offer an exhaustive analysis comparing single and multi-attribute learning models. Single-attribute models analyze specific features such as age or gender in isolation, while multi-attribute models amalgamate various characteristics to improve predictive accuracy. These models are particularly beneficial in complex application environments, such as demographic studies and security systems, where a nuanced understanding of the interrelationships among different facial attributes is essential. Furthermore, the authors introduce groundbreaking neural frameworks that significantly enhance the performance of existing facial recognition technologies, marking a progressive stride in the evolution of this field.

Pezzini et al. (2023) explore facial aging through a dermatological lens, employing non-invasive, high-resolution imaging techniques to scrutinize the subtle skin changes across different age groups. Their thorough analysis of 140 facial skin images spanning seven age groups establishes a detailed correlation between visible signs of aging and underlying imaging characteristics. This research not only enhances our understanding of dermatological aging but also pioneers new diagnostic tools and methodologies for early detection and intervention in skin-related conditions. By gaining a deeper understanding of these age-related changes, the study provides invaluable insights into the fields of dermatology, cosmetic science, and geriatric care, potentially revolutionizing skincare and anti-aging treatments.

Henry et al. (2023) integrates insights from psychology, neuroscience, and gerontology within the Social Cognitive Resource (SCoRe) framework to examine facial aging. This innovative approach facilitates a detailed exploration of how cognitive processes interact with physiological aspects of aging, influencing facial expressions and muscle dynamics over time. The holistic perspective provided by this research illuminates the complex dynamics of aging, offering a more comprehensive understanding of how aging impacts human facial features from a neurobiological standpoint.

Building on these thematic concerns, Chandaliya and Nain (2023) explore the technological advancement of facial recognition systems through the development of an aging framework that utilizes a wavelet-based Generative Adversarial Network (GAN). This model is meticulously designed to enhance the realism and precision of age progression in digital images. By integrating wavelet transformations, their framework adeptly captures intricate facial details and textures, which are crucial for

achieving realistic transformations and accurate age estimations. This advancement significantly boosts the performance of facial recognition systems, particularly in environments where precise age determination is critical.

In a related stripe, [Du et al. \(2019a\)](#) introduces the Cycle Age-Adversarial Model (CAAM), a novel neural network architecture designed to enhance cross-age face recognition. This model employs a cyclic adversarial learning strategy to minimize the perceptual differences between age-progressed and actual images of the same individual, thereby improving the model's ability to accurately recognize faces across significant age variations. The CAAM demonstrates superior performance on benchmark datasets, underscoring its potential to bolster security systems and other applications that require reliable age-invariant recognition capabilities.

Furthermore, [Du et al. \(2019b\)](#) expand their research into age-invariant facial recognition with the development of the Cross-Age Face Recognition (CAFR) technique, which incorporates the Age Factor Removal Network (AFRN). This approach leverages adversarial and transfer learning strategies to isolate age-related features while preserving individual identity, demonstrating a high level of robustness and adaptability across diverse datasets. The effectiveness of this system in various practical applications signifies a significant advancement in the field, opening new avenues for the development of increasingly sophisticated facial recognition technologies.

The research contributions advance further with [Chandaliya and Nain \(2022\)](#), who introduced PlasticGAN, an advanced generative framework that utilizes a conditional Generative Adversarial Network (cGAN) combined with a deep residual variational autoencoder. This innovative model excels in generating post-surgery facial images with high authenticity and variability, proving particularly valuable in clinical decision-making, forensic science, and the entertainment industry.

Expanding the scope of generative models, [Atkale et al. \(2021\)](#) delve into facial aging through regression analysis and progression modeling using a sophisticated deep GAN architecture. Their research highlights the generation of high-quality images that portray realistic aging effects, aiming to improve the training and performance of age-progression models across various age groups and amidst dataset variability.

Contributing to technological diversity, [Olivier et al. \(2023\)](#) have developed FaceTuneGAN, a novel method for crafting 3D face models from textual descriptions. This model exhibits high versatility in transferring facial features across different age groups and proves particularly useful in digital entertainment and virtual reality applications, where customizable facial expressions and features are crucial.

Building on the advances in facial recognition and aging models discussed earlier, the developments in sketch conversion technologies, such as those by Jo and Park with SC-FEGAN (2019) and Yang et al. with S2FGAN (2022), represent further strides in generative modeling. These technologies enhance the ability to create realistic images from sketches by integrating sophisticated loss components and attribute mapping networks into their designs. This enhancement in sketch conversion parallels the improvements seen in age-progression and face modeling technologies, showing a consistent trend towards more

precise and adaptable image generation across various applications, from digital entertainment to forensic science.

Upon conducting an exhaustive literature study, targeting on the shortcomings of existing techniques, the proposed system builds upon previous advancements in facial recognition and aging technologies by integrating a de-aging module and a sketch generator module, aimed at overcoming the limitations of traditional forensic facial recognition. Utilizing a deepfake-based neural network and a pix2pix-based Generative Adversarial Network (GAN), this system significantly enhances the accuracy and realism of facial sketches across diverse datasets, demonstrating substantial improvements in key performance metrics. These innovations not only refine age-invariant recognition capabilities but also set new standards in forensic applications, highlighting the transformative potential of combining de-aging and sketch generation technologies.

## 3 Proposed system

To enhance the training performance of generated De-aged quality sketches, we have proposed a two-step approach. [Figure 1](#) presents the overall working of the model of our proposed methodology.

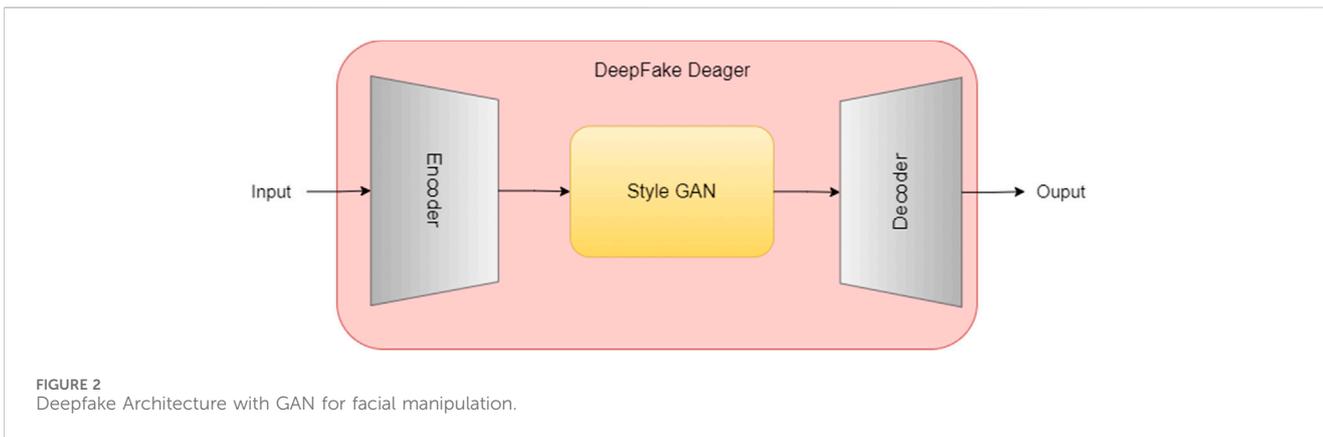
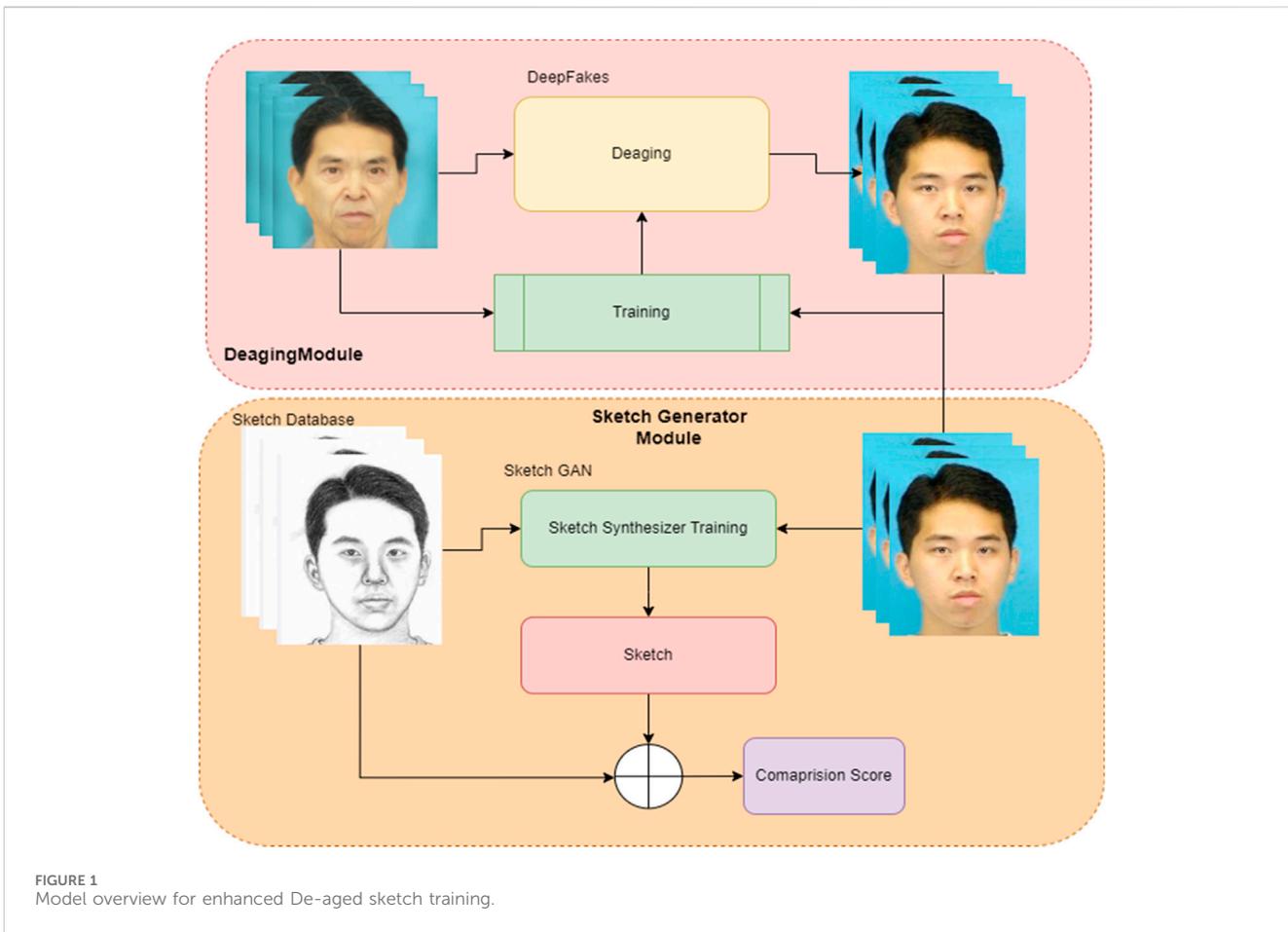
As presented in [Figure 1](#) our system is divided into two main modules 1. The De-Aging Module and 2. Sketch Generator Module. A thorough explanation of these modules is presented in the successive sections.

### 3.1 De-Aging module

The De-Aging module in the system focuses entirely on reverse aging the captured photographic evidence of the human face. This model is fine-tuned in such a way so that it de-ages only frontal faces captured as mugshots from aged suspects. This is a vital step in the system since profile and other views can distort the de-aging process. The model is a neural network that utilizes a deepfake trained system that can de-age faces of any gender, race, and culture as shown in [Figure 2](#) ([Zhang, 2022](#)).

Deepfakes consist of a GAN embedded within a Encoder decoder network. The Encoder-Decoder extracts noticeable abstract features from the face and passes the same towards the GAN. The GAN network manipulates the face accordingly based on the training provided to it ([Seow et al., 2022](#)). The decoder decodes and reconstructs the face accordingly. The Style GAN ([Richardson et al., 2021](#)) represented in [Figure 2](#) Is designed to remove blemishes and wrinkles. [Figure 3](#) displays the internal structure of the Style GAN used in our proposed system.

The internal structure portrayed in [Figure 3](#) Features a Style Mapper Network with fully connected layers serving as style mappers. These styles are embedded into the synthesizer, a generator network adding noise and style vectors from the encoder. A progressive GAN is employed, enhancing with each iteration, starting with  $4 \times 4$  images and progressively refining details until the discriminator can no longer diminish finer crisp details. The detailed generator architecture is explained in the subsequent sections.



### 3.1.1 Up sampling layer

The up-sampling layer is crucial in GANs for expanding image spatial dimensions, specifically width and height. This expansion is advantageous in GANs, where the generator starts with a compact input and gradually enlarges it to produce a full-sized output image. The Up-sampling technique (Amin et al., 2019) in this research article employs bilinear interpolation, as indicated by the mathematical framework in Eqs 1, 2. Bilinear interpolation involves a two-dimensional linear interpolation for resizing images, which enhances image quality by considering the

intensity values of the nearest  $2 \times 2$  neighborhood of known pixel values to estimate the unknown pixel values. This method provides a balance between computational efficiency and output quality, making it suitable for de-aging applications where maintaining facial details is critical.

Specifically, Eq 1 outlines the general up-sampling function, and Eq 2 details how pixel values are interpolated. The bilinear approach ensures that de-aged images retain smoother transitions in pixel values, which is crucial for preserving subtle facial features and expressions in forensic imagery.

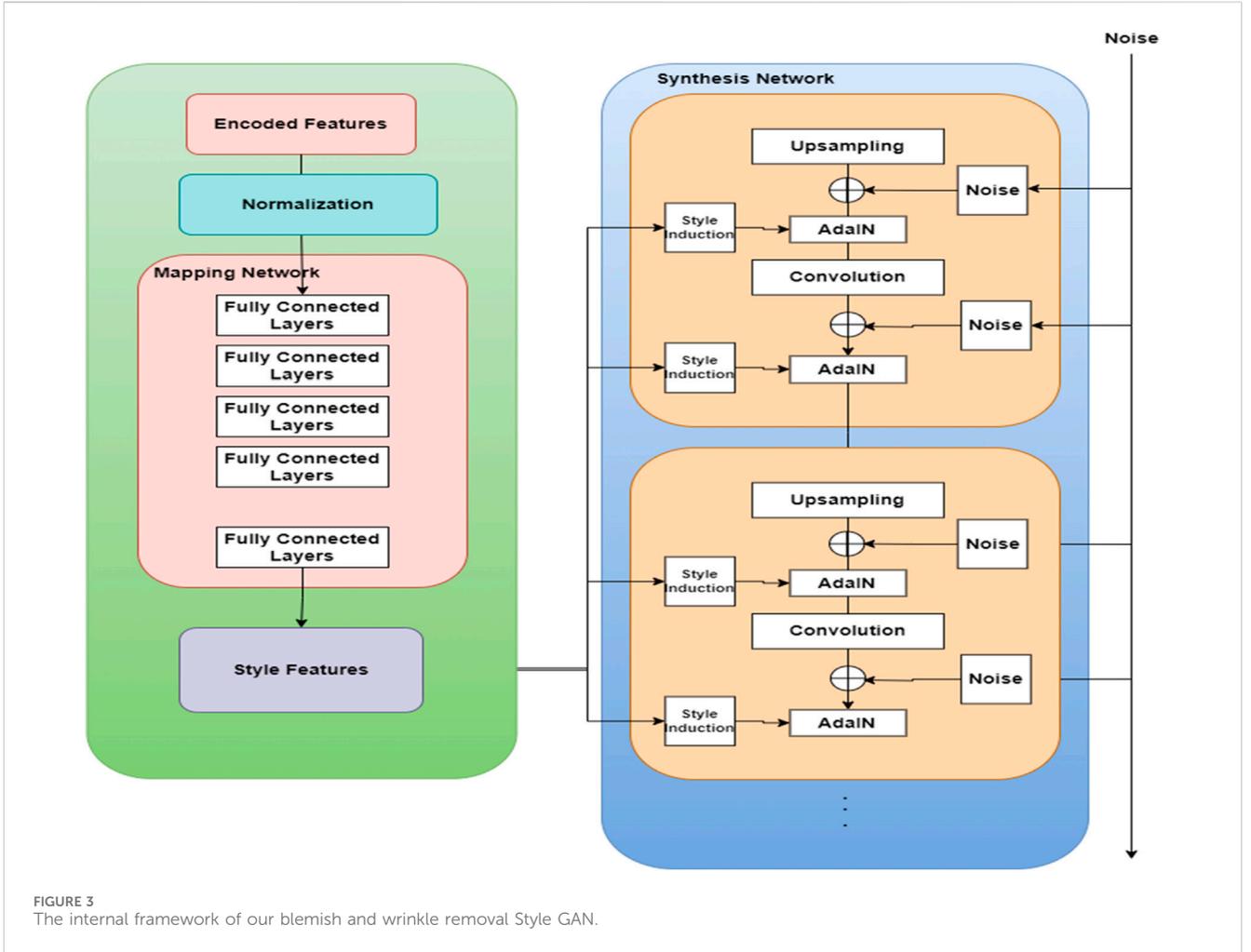


FIGURE 3 The internal framework of our blemish and wrinkle removal Style GAN.

$$Y = UpSample(X) \tag{1}$$

$$Y(i, j) = \frac{1}{4} \left( X \left( \left\lfloor \frac{i}{2} \right\rfloor, \left\lfloor \frac{j}{2} \right\rfloor \right) + X \left( \left\lfloor \frac{i}{2} \right\rfloor, \left\lceil \frac{j}{2} \right\rceil \right) + X \left( \left\lceil \frac{i}{2} \right\rceil, \left\lfloor \frac{j}{2} \right\rfloor \right) + X \left( \left\lceil \frac{i}{2} \right\rceil, \left\lceil \frac{j}{2} \right\rceil \right) \right) \tag{2}$$

Term  $X$  be the input feature map of size  $W \times H$ . The up-sampled output  $Y$  of size  $2W \times 2H$ , where  $UpSample(\cdot)$  is a function that interpolates additional rows and columns into  $X$ .

### 3.1.2 Adaptive Instance Normalization layer

Adaptive Instance Normalization (AdaIN) is pivotal in Style GAN, a variant of GANs. AdaIN plays a vital role in aligning the feature statistics of a content image with those of a style image, facilitating the transfer of style onto the content. In Style GAN, AdaIN allows manipulation of the latent space to control specific characteristics of the generated images, including style at various levels of detail (Gu and Ye, 2021). This is achieved by adjusting and shifting features in each channel of the content image based on the statistics of the corresponding channel in the style image. The result is a robust and flexible framework for image synthesis, capable of producing high-quality images with a diverse range of intricate styles. Eq. 3 reveals the mathematical formulation of the layer.

$$A = \sigma(S) \cdot \frac{C - \mu(C)}{\sigma(C)} + \mu(S) \tag{3}$$

In the equation, content feature map  $C$  and a style feature map  $S$ ,  $A$  is the AdaIN Output, where  $\mu(\cdot)$  and  $\sigma(\cdot)$  denote the mean and standard deviation operations, respectively.

### 3.1.3 Convolution layer

The convolutional layers of our de-aging network utilize a combination of  $3 \times 3$  and  $5 \times 5$  filter sizes. The  $3 \times 3$  filters adept at extracting high-resolution features from the facial images. As outlined by Eqs 4, 5 (Wang et al., 2021; Pehlivan et al., 2023; Zhao et al., 2023).

$$O = Conv(I, F) \tag{4}$$

$$O(i, j) = \sum_{u=0}^{m-1} \sum_{v=0}^{n-1} I(i+u, j+v) \times F(u, v) \tag{5}$$

Parameter  $O$  is the output feature map, and  $Conv(I, F)$  denotes the convolution operation.  $I$  denote the input image and  $F$  represents Convolutional filter or kernel. In Eq. 5,  $O(i, j)$  is value of the output feature map at position  $(i, j)$ . Parameter  $I(i+u, j+v)$  represents pixel value of the input image at the position offset by  $(u, v)$  from the position  $(i, j)$ .  $F$  is a filter corresponding to the offset and  $m, n$  denotes dimensions of the filter. The symbol  $\sum$  is used here to

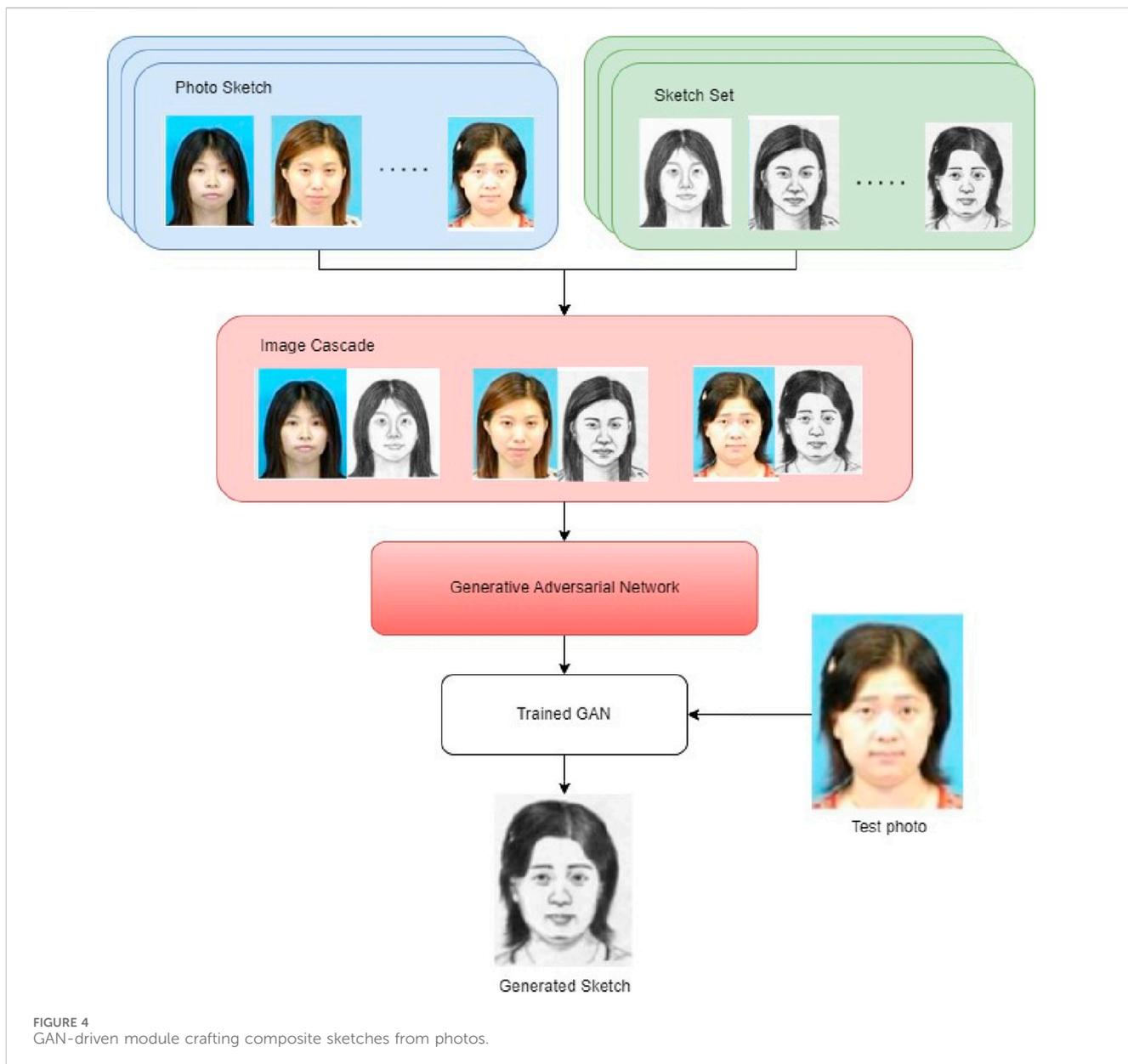


FIGURE 4  
GAN-driven module crafting composite sketches from photos.

indicate that the output pixel value is the weighted sum of the input pixel values covered by the filter, with weights given by the filter values.

To complement the finer details captured by  $3 \times 3$  filters, the system also employs  $5 \times 5$  filters which are designed to grasp broader features like overall face shape, contour changes, and major wrinkles. These larger filters help in understanding the macro changes in facial structure that occur with aging, providing a balance between detailed textural representation and global facial morphology. The dual approach of using both  $3 \times 3$  and  $5 \times 5$  filters allows our system to comprehensively model facial aging by effectively capturing both micro-detail enhancements and significant age-related transformations. This is achieved by layering these filters in a manner that each type contributes uniquely: the  $3 \times 3$  filters maintain sharpness and detail, crucial for the realistic portrayal

of young features, while the  $5 \times 5$  filters aid in the smooth transition of these features as aging progresses, as supported by the deployment strategies discussed in Eqs 6, 7.

$$Width(0) = \left\lfloor \frac{W - m + 2p}{s} + 1 \right\rfloor \tag{6}$$

$$Height(0) = \left\lfloor \frac{H - n + 2p}{s} + 1 \right\rfloor \tag{7}$$

In Eqn 6, parameter  $Width(0)$  denotes width of the output feature map after the convolution,  $W$  represents the input feature map, parameter  $m$  is width of the filter,  $p$  corresponds to the padding applied to the input and  $s$  is the step size at which the filter is applied across the input feature map. Furthermore, Eqn 7 reveals height of the output feature map after the convolution denoted by  $Height(0)$ , parameters  $H$  &  $n$  represents the height of the feature map and filter respectively. Parameters  $p$  &  $s$  represents as per the description presented in Eq. 6.

TABLE 1 Architectural breakdown of patch GAN for image comparison.

Layer name	Size	Description
Input Layer	(256,256,3)	The input image to be compared
Input Layer	(256,256,3)	The Target image to be compared
Concatenation Layer	(256,256,6)	Concatenation of input and Target
Down sampling Layer	(128,128,64)	Compress the spatial extent of the feature matrices
	(64,64,128)	
	(32,32,256)	
Padding Layer	(34,34,256)	Prevent the shrinking of the input image and to preserve the information presented in the corners
Convolutional Layer	(31,31,512)	Dynamically acquire and adjust the layered structure of spatial features from the given data
Normalization Layer	(31,31,512)	
ReLU	(31,31,512)	It is a piecewise linear function that outputs the input directly if it is positive, otherwise, it outputs zero
Padding Layer	(33,33,512)	
Convolutional Layer	(30,30,1)	Final Target Layer

### 3.1.4 Encoder and decoder layer

Deepfake technology employs an encoder-decoder architecture. The encoder transforms input data, like an image or video frame, into a condensed latent space, preserving critical attributes and eliminating unnecessary details (Akram and Khan, 2023). The decoder reconstructs the original data from this latent space. In Deepfake, this process facilitates face swaps by extracting facial features from both individuals. The decoder then superimposes the age feature onto the others, creating a convincing young appearance. The formulation of this layer is represented by Eqs 8–11.

$$(L, F) = E(I) \quad (8)$$

$$(L_1, F_1) = E(I_1) \quad \text{and} \quad (L_2, F_2) = E(I_2) \quad (9)$$

$$O = D(L, F') \quad (10)$$

$$O_1 = D(L_1, F_2) \quad \text{and} \quad O_2 = D(L_2, F_1) \quad (11)$$

As represented in the equations, encoder  $E$  transforms the input image  $I$  into a latent representation  $L$  and a feature vector  $F$ . This equation shows the decoder  $D$  reconstructing the output image  $O$  using the latent representation  $L$  and a feature vector  $F'$  from another image. These represent the encoder processing two different images  $I_1$  and  $I_2$ , and the decoder reconstructing images  $O_1$  and  $O_2$  with swapped features.

## 3.2 Sketch generator module

The Sketch generator module in the proposed technique operates through a GAN that takes a photo as input and produces a detailed sketch of a person. Given the stark differences between sketches and photos, accomplishing the entire objective using a single GAN becomes a challenging task. Consequently, the procedure is bifurcated into a two-fold approach. The GAN has been enhanced to generate not only hand-drawn sketches but also composite versions. Figure 4 elucidates the overall structure of the GAN training crafted for composite sketch generation.

Employing the standard Conditional GAN pix2pix model, the GAN associates photos with sketches. The training process involves initially establishing associations between photo-sketch pairs, binding them together through image cascading as outlined in Eq. (12).

$$I_{merged}(x, y) = \sum_{(i+m, j+n) \in [I_{sketch}, I_{photo}]} [I_{photo}(i, j) + I_{sketch}(m, n)] \quad (12)$$

Here in Eq. 12,  $I_{merged}(x, y)$  represents the merged image created from the pixels  $I_{photo}(i, j)$  and  $I_{sketch}(m, n)$  respectively. The internal structure of our GAN comprises of two main modules, i.e. 1) The Decimator and 2) The Generator. Detailed architecture of the same are explained in the subsequent sections.

### 3.2.1 The discriminator

The discriminator model in the pix2pix architecture utilizes the traditional patch GAN approach, concentrating on  $N \times N$  image patches and evaluating them individually. The scrutiny intensifies with progressing training iterations, enhancing overall clarity. Table 1 provides a breakdown of the Patch GAN network's architecture.

As depicted in Table 1, it offers insights into the discriminator layer's process of examining the input image and the target for comparison. Notably, the pooling layer is employed to reduce the size of numerous pixels to a coarser number. The Patch GAN (Henry et al., 2021) configuration includes two convolutional layers, with one positioned in between and the other at the end. Additionally, the down sampling layers incorporate multiple convolutional layers, as mathematically outlined in the provided Eq. 13.

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim P_{data}(x)} [\log D(x | y)] + \mathbb{E}_{z \sim P_z(z)} [\log (1 - D(G(z | y)))] \quad (13)$$

In this context,  $x$  denotes actual data,  $z$  is identified as the noise vector,  $y$  symbolizes the condition,  $G$  stands for the generator, and  $D$

TABLE 2 Architectural overview of Pix2Pix GAN with U-Net modification.

Layer name	Size	Description
Input Layer	(256,256,3)	Input Layer
Down sampling Layer 1	(128,128,64)	A Convolutional Layer with a Normalizing Layer
Down sampling Layer 2	(64,64,128)	
Down sampling Layer 3	(32,32,256)	
Down sampling Layer 4	(16,16,512)	
Down sampling Layer 5	(88,512)	
Down sampling Layer 6	(44,512)	
Down sampling Layer 7	(22,512)	
Down sampling Layer 8	(11,512)	
Concatenate	(22,512)	
Down sampling Layer 10	(2,21,024)	
Concatenate	(44,512)	
Down sampling Layer	(88,512)	
Concatenate	(8,81,024)	
Down sampling	(16,16,512)	
Concatenate	(16,16,1024)	
Down sampling	(32,32,256)	
Concatenate	(32,32,512)	
Down sampling	(64,64,128)	
Concatenate	(64,64,256)	
Down sampling	(128,128,64)	
Concatenate	(128,128,128)	
Lambda Layer	(256,256,3)	Output Layer
Transpose Layer		

signifies the discriminator. The initial part of the equation is indicative of the discriminator's expected log-likelihood of accurately identifying real data as real, contingent on the condition. Conversely, the latter part of the equation reflects the expected log-likelihood of the discriminator accurately recognizing artificial data, which is produced by the generator, as fake, also based on the condition.

To improve the training process of the discriminator, Binary Cross-Entropy (BCE) loss function is utilized. The BCE loss function is mathematically represented as presented in Eq. 14

$$L_D = -\frac{1}{N} \sum_{i=1}^N [y_i \cdot \log(D(x_i)) + (1 - y_i) \cdot \log(1 - D(x_i))] \quad (14)$$

$L_D$  denotes the loss for discriminator,  $N$  represents number of training samples. Parameter  $y_i$  is either 0 or 1 and  $D(x_i)$  is the discriminator's probability output.

The loss function in Eq. 14 ensures that the discriminator is trained to correctly classify real and generated images as accurately as possible. To further prevent the discriminator from overfitting, we

can implement dropout regularization in the discriminator layers. Dropout randomly sets a fraction of input units to 0 at each update during training time, which helps to prevent overfitting by reducing the reliance on any individual neuron. The dropout rate typically ranges from 20% to 50%. This technique not only helps in regularization but also encourages the discriminator to learn more robust features. The mathematical function for dropout regularization during training is described as depicted in Eq. 15

$$R(x) = x \odot M \quad (15)$$

In the equation,  $R(x)$  represents the output on applying dropout,  $x$  is the input layer vector and  $M$  is the mask vector as described according to the Bernoulli distribution and the symbol  $\odot$  denotes element-wise multiplication.

### 3.2.2 The generator

The Pix2Pix GAN leverages a generator derived from a tweaked U-Net architecture, akin to an autoencoder. This U-Net structure comprises an encoder for down-sampling and a decoder for up-

TABLE 3 A detailed examination of CUHK and AR face sketch databases.

Dataset	Source	Number of images	Number of subjects	Image conditions
CUHK	CUHK Campus	606	Varies by database	Frontal pose, normal lighting condition, neutral expression
AR	CVC at UA.B.	Over 4,000	126 (56 women and 70 men)	Faces in frontal view showcasing varied expressions, under different lighting conditions and with obstructions

TABLE 4 FID scores illustrating face sketch generation performance across epochs and datasets.

Epochs	Datasets used and their FID score (↓)	
	CUHK	AR
100	50	75
200	46	70
300	44	65
400	43	63
500	41	58
600	40	56
700	39	55
800	38	54
900	38	53
1,000	38	53

sampling, as detailed in (Basu et al., 2020; Zhao et al., 2023). Within this architecture, the encoder is tasked with capturing the image's context, and the decoder is designed for detailed localization. A notable characteristic of our U-Net is the inclusion of skip connections between matching layers of the encoder and decoder, which are instrumental in retrieving detailed features lost during down-sampling. In the Pix2Pix system, the generator's function is to take an image as input and generate a comparable output image, aiming to trick the discriminator into classifying this output as real. The full architecture of our generator is depicted in Table 2.

Defining the Generator loss function involves utilizing the equation, as elucidated in (16). This equation plays a fundamental role in guiding the optimization process for the Generator within the given context.

$$L_G = E_{x,y} [(y - G(x))^2] \quad (16)$$

In the equation,  $x$  is the input image,  $y$  is the target image,  $G(x)$  is the generated image.  $E_{x,y}$  defines the expected value generated from the loss difference. The lambda layer used below is a custom layer defined to clarify sketches as the output. The entire structure of the lambda layer is explained below.

### 3.2.3 Lambda layer

In the proposed system, the Lambda layer serves dual purposes: rendering and gamma adjustment. For rendering, it conducts

operations on a color image, initially transforming it into grayscale to reduce color information for better sketching suitability. Following this, a Gaussian blur with a  $21 \times 21$  kernel size and zero standard deviation in both  $x$  and  $y$  directions is applied to the grayscale image, effectively smoothing noise and creating a blurred version. Finally, the original grayscale image is divided by the blurred image with a scale factor of 256, intensifying edges and contrast. This division enhances the darkness of darker pixels and the brightness of lighter pixels, resulting in an image resembling a pencil-drawn sketch. The choice of a  $21 \times 21$  kernel size for Gaussian blur is driven by the need to achieve a balance between smoothing effectiveness and computational efficiency. Further the scale factor 256 ( $2^8$ ) corresponds to the maximum value for an 8-bit channel, thereby normalizing the image to use the full range of possible intensity values, enhancing the depth and realism of the final image. Mathematical Eqs 17–21 underpin the operational workings of these processes.

$$I(x, y) = [(0.2126 * Red) + (0.7152 * Green) + (0.0722 * Blue)] \quad (17)$$

$$G(x, y) = K(x, y) * I(x, y) \quad (18)$$

$$K(x, y) = \frac{1}{2\pi\sigma^2} \times e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (19)$$

$$\beta(x, y) = I(x, y) / G(x, y) * 256 \quad (20)$$

$$A(x, y) = \text{table}[\beta(x, y)], \text{ where table}[i] = (i / 255)^{\frac{1}{\text{gamma}}} \text{ for } i \text{ in } [0, 255] \quad (21)$$

In the equations,  $G(x, y)$  represents the blurred grayscale image,  $I(x, y)$  is the original grayscale image, and  $K(x, y)$  represents the Gaussian kernel with a standard deviation  $\sigma$  controlling the blur intensity. Larger sigma values result in more blur. The kernel size is selected to cover at least three  $\sigma$  in both directions. Gaussian blur is achieved through convolution, where each output pixel is a weighted average of its neighbors based on the kernel values. The convolution can be applied by multiplying each pixel with the entire kernel or separately for each axis using a one-dimensional kernel.  $\beta(x, y)$  is the output image, and the division operation enhances edges and contrast by adjusting pixel values based on intensity. The gamma adjustment operation further modifies brightness, mapping each pixel value to a new value with a lookup table. The gamma parameter controls brightness, with lower values darkening the image and higher values brightening it.  $A(x, y)$  is the output,  $I(x, y)$  is the input, gamma is the gamma parameter, and table is the lookup table with 256 mappings. This operation is a non-linear transformation that alters contrast and dynamic range in the image.

TABLE 5 Quality band for referential metrics.

Metric	Good	Medium	Bad
PSNR ↑	>40 dB	30–40 dB	<30 dB
SSIM ↑	0.8–1	0.5–0.8	<0.5
MSE ↓	0–10	10–50	>50
NIQE ↓	0–5	5–10	>10
BRISQUE ↓	0–20	20–40	>40

## 4 Results and discussions

The results and discussions topic for the research exertion shown is a section that presents and analyzes the findings of the experiment on sketch generation from photos. The system was run on Tesla P100 GPU has 12GB of HBM2 memory, which provides high bandwidth and low latency for data-intensive workloads. The Tesla P100 GPU also supports NVIDIA NV Link, which is a high-speed interconnect that enables multi-GPU scalability and performance. We have also performed in-depth analysis of various parameters for detailed understanding of the modules. The analysis is shown in the subsequent sections.

### 4.1 Dataset description

The CUHK Face Sketch Database and the AR Face Sketch Database are significant datasets employed for experimental inquiries in computer vision. The subsequent Table 3 provides a comprehensive overview of these datasets, detailing their characteristics and key attributes.

### 4.2 Evaluating sketch generation with GANs

In assessing the performance and quality of the sketch generation GAN, metrics are utilized to measure the realism, diversity, and similarity of the generated images to real ones. The subsequent sections provide a detailed discussion of the various evaluation metrics adopted for this purpose (Fernandes and Bala, 2016). Further, the generator's learning rate is set at 0.0002, contrasting with the discriminator's rate of 0.0001 to maintain balance and prevent either network from overpowering prematurely. The batch size of 64 is set, which ensures stable learning by smoothing the gradient updates. The generator features a U-Net-like architecture with encoder-decoder paths and skip connections, essential for preserving important facial details.

#### 4.2.1 Inception score or FID score

This is an objective and quantitative metric that involves using a pre-trained classifier network (such as Inception) to classify the generated images into different categories, and then calculating a score based on both the diversity (entropy) and quality (confidence) of the classifications. A higher score indicates better generated images. Eq. 22 shows mathematical form of the inception score.

$$\text{Score} = e^{(E_x(KL(p(y|x)||p(y))))} \quad (22)$$

TABLE 6 Image quality metrics comparison for CUHK and AR datasets across age groups.

Evaluation metrics			CUHK	AR dataset
Average	20–30	PSNR↑	0.78	0.692
		SSIM↑	0.865	0.902
		MSE ↓	6.24	5.36
		NIQE ↓	3.24	2.36
		BRISQUE ↓	17.45	12.454
	30–40	PSNR↑	0.802	0.705
		SSIM↑	0.879	0.915
		MSE ↓	6.05	5.25
		NIQE ↓	3.15	2.45
		BRISQUE ↓	17.20	12.35
	40–50	PSNR↑	0.872	0.910
		SSIM↑	0.865	0.902
		MSE ↓	6.15	5.45
		NIQE ↓	3.20	2.55
		BRISQUE ↓	17.35	1,250
	50–60	PSNR↑	0.790	0.690
		SSIM↑	0.865	0.905
		MSE ↓	6.30	5.55
		NIQE ↓	3.30	2.65
		BRISQUE ↓	17.50	12.65
60–70	PSNR↑	0.785	0.685	
	SSIM↑	0.860	0.900	
	MSE ↓	6.40	5.65	
	NIQE ↓	3.24	2.36	
	BRISQUE ↓	17.45	12.454	

Score is the inception score,  $E_x$  is the expectation over all generated images  $x$ .  $KL$  represents the Kullback-Leibler divergence, a metric that quantifies the disparity between two distinct probability distributions.  $p(y|x)$  is the conditional distribution of the class label  $y$  given the image  $x$ , computed by the classifier network.  $p(y)$  the marginal distribution of the class label  $y$ , computed by averaging  $p(y|x)$  over all generated images  $x$ .

The FID score for sketch generation from faces was calculated across various epochs using two standard datasets. The obtained results are depicted in Table 4.

With increasing epochs, the generative model generates increasingly realistic images, evident in the declining FID scores for both datasets. Notably, the model consistently achieves lower FID scores on CUHK compared to AR at each epoch, implying superior performance on CUHK. This performance difference may be attributed to variations in dataset features, including size, diversity, and quality. The FID scores plateau beyond 800 epochs likely due to the

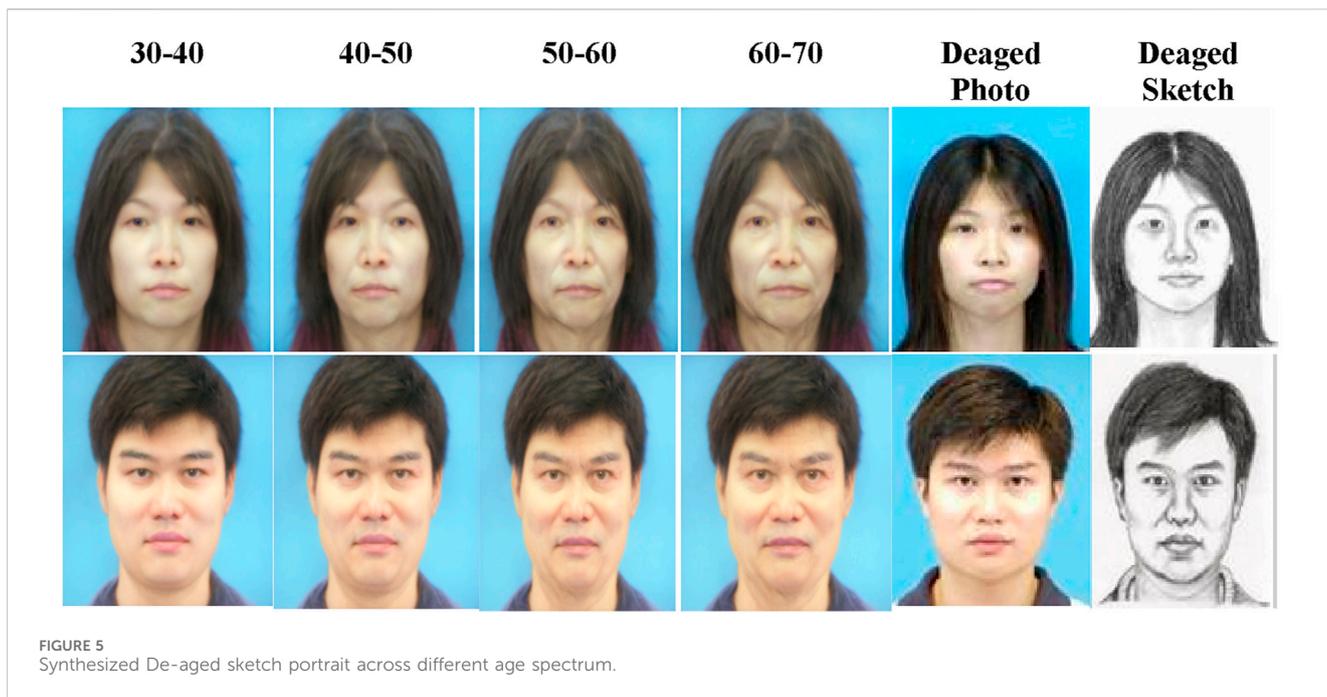


FIGURE 5  
Synthesized De-aged sketch portrait across different age spectrum.

GAN reaching its capacity for improvement, with both the generator and discriminator achieving equilibrium in their performance.

#### 4.2.2 Referential metrics

In the realm of sketch comparison, three primary metrics are employed to quantitatively assess the similarity and quality of sketches compared to their original images: MSE (Mean Squared Error), SSIM (Structural Similarity Index Measure) and PSNR (Peak Signal-to-Noise Ratio).

##### 4.2.2.1 PSNR (peak signal-to-noise ratio)

PSNR stands for Peak Signal-to-Noise Ratio, which evaluates the maximum error between the original and reconstructed images. In this context,  $MAX_I$  denotes the highest possible pixel value in the image. PSNR values are typically conveyed using the logarithmic decibel scale, where higher PSNR figures suggest superior image quality. This concept is mathematically formulated as shown in Eq. 23.

$$PSNR = 20 \cdot \log_{10} \left( \frac{MAX_I}{\sqrt{MSE}} \right) \quad (23)$$

##### 4.2.2.2 SSIM (structural similarity index measure)

SSIM is used for measuring the similarity between two images. It considers changes in structural information, luminance, and contrast. The values of SSIM range from  $-1$  to  $1$ , where  $1$  indicates perfect similarity. Eq. 24 depicts the mathematical procedure for the same.

$$SSIM(x, y) = \frac{(2\mu_x + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (24)$$

In the formulation,  $SSIM(x, y)$  represents the Structural Similarity Index between two images  $x$  and  $y$ .  $\mu_x$  and  $\mu_y$  are the average values of images  $x$  and  $y$ , respectively.  $\sigma_x$  and  $\sigma_y$  are the variance of images  $x$  and  $y$ .  $\sigma_{xy}$  represents the covariance of

images  $x$  and  $y$ .  $c_1$  and  $c_2$  are constants used to stabilize the division with weak denominators. These constants are typically small values.

##### 4.2.2.3 MSE (mean squared error)

MSE measures the average difference of squares between the original image and the reconstructed image. It is a probability function, corresponding to the expected value of the error loss. Lower MSE values indicate better quality. Computation of the metrics is depicted as per Eq. 25.

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (25)$$

MSE represents the Mean Squared Error.  $m$  and  $n$  are the dimensions of the images, representing the number of rows and columns, respectively.  $I(i, j)$  is the pixel value at position  $(i, j)$  in the initial image.  $K(i, j)$  is the pixel value at position  $(i, j)$  in the recreated or compared image. The double summation  $\sum_{i=0}^{m-1} \sum_{j=0}^{n-1}$  iterates over all pixel positions in the images.

#### 4.2.3 Non-referential metrics

In sketch comparison, non-referential metrics like BRISQUE (Blind/Reference less Image Spatial Quality Evaluator), PIQE (Perception based Image Quality Evaluator), and NIQE (Natural Image Quality Evaluator) play a crucial role (Fernandes and Bala, 2018; Sain et al., 2021). These metrics evaluate image quality without requiring a reference image. Each of these metrics is examined in detail in the trailing sections.

##### 4.2.3.1 NIQE (natural image quality evaluator)

NIQE is a completely blind image quality assessment model that works by assessing the statistical deviations from natural image

TABLE 7 Comparing facial aging techniques using the CUHK dataset.

Technique	Age band	FID ↓	PSNR↑	SSIM↑	MSE ↓	NIQE ↓	BRISQUE ↓
Wavelet-Based GAN (Chandaliya and Nain, 2023)	20–30	57	0.47	0.74	9.25	5.52	24.24
	30–40	57	0.52	0.73	9.30	5.48	24.52
	40–50	59	0.55	0.72	9.45	5.30	24.71
	50–60	61	0.53	0.73	9.25	5.40	24.70
	60–70	60	0.52	0.75	9.24	5.5	24.68
Cycle Age-Adversarial Model (Du et al., 2019a)	20–30	45	0.66	0.79	7.45	4.85	21.23
	30–40	48	0.66	0.80	7.25	4.82	21.23
	40–50	48	0.69	0.80	7.20	4.50	21.05
	50–60	45	0.65	0.81	7.25	4.82	21.18
	60–70	46	0.65	0.79	7.52	4.97	21.18
Age Factor Removal Network (AFRN) (Du et al., 2019b)	20–30	42	0.63	0.76	7.115	4.58	20.54
	30–40	42	0.61	0.75	7.120	4.52	20.64
	40–50	42	0.70	0.74	7.130	4.50	20.81
	50–60	42	0.69	0.75	7.125	4.53	20.94
	60–70	43	0.70	0.765	7.21	4.57	20.72
cGAN with Deep Residual VAE (Chandaliya and Nain, 2022)	20–30	51	0.67	0.76	8.25	5.24	21.044
	30–40	55	0.71	0.75	8.30	5.23	21.034
	40–50	52	0.74	0.74	8.35	5.21	21.11
	50–60	54	0.72	0.75	8.25	5.25	21.081
	60–70	54	0.69	0.74	8.30	5.23	21.12
DeepGAN for aging Progression (Atkale et al., 2021)	20–30	44	0.71	0.75	6.97	4.78	18.99
	30–40	48	0.71	0.74	6.99	4.65	19.05
	40–50	44	0.74	0.73	6.98	4.95	19.21
	50–60	47	0.72	0.74	7.012	4.60	19.42
	60–70	45	0.70	0.73	7.12	4.25	19.76
<b>Proposed System</b>	20–30	38	0.78	0.865	6.24	3.24	17.45
	30–40	38	0.80	0.879	6.05	3.25	17.20
	40–50	39	0.87	0.865	6.15	3.20	17.35
	50–60	38	0.79	0.865	6.30	3.30	17.50
	60–70	39	0.785	0.860	6.40	3.24	17.45

properties. NIQE's calculation involves creating a quality-aware feature set based on natural scene statistics and fitting a multivariate Gaussian model to this feature set. Mathematically NIQE is defined as depicted in Equation 26.

$$Nscore = \frac{I_{\text{factor}} - \mu}{\sigma} + 1 \quad (26)$$

Parameter Nscore represents the naturalness score.  $I_{\text{factor}}$  is the intensity result of the image under test.  $\mu$  and  $\sigma$  are the mean and standard deviation of the image pixels, correspondingly. The final

value Nscore of is a nonnegative scalar, with lower values indicating higher image quality.

#### 4.2.3.2 BRISQUE (blind/reference less image spatial quality evaluator)

The BRISQUE (Blind/Reference less Image Spatial Quality Evaluator) method is a unique image quality assessment tool that operates without the need for a reference image. This approach is especially advantageous in situations where an original image is unavailable for comparison. BRISQUE

TABLE 8 Comparing facial aging techniques using the AR dataset.

Technique	Age band	FID ↓	PSNR↑	SSIM↑	MSE ↓	NIQE ↓	BRISQUE ↓
Wavelet-Based GAN (Chandaliya and Nain, 2023)	20–30	79	0.43	0.71	8.99	4.91	17.21
	30–40	75	0.48	0.73	8.91	4.84	17.23
	40–50	81	0.52	0.70	8.84	4.89	17.125
	50–60	82	0.52	0.71	8.87	4.75	17.264
	60–70	80	0.51	0.71	8.89	4.79	17.115
Cycle Age-Adversarial Model (Du et al., 2019)	20–30	68	0.59	0.81	7.01	4.49	15.012
	30–40	69	0.51	0.76	7.09	4.53	15.154
	40–50	72	0.55	0.82	6.98	4.5	15.324
	50–60	71	0.62	0.83	6.99	4.51	14.932
	60–70	68	0.59	0.81	7.01	4.62	15.014
Age Factor Removal Network (AFRN) (Du et al., 2019b)	20–30	63	0.61	0.72	6.81	4.31	15.25
	30–40	61	0.69	0.72	6.89	4.21	15.114
	40–50	59	0.69	0.79	6.74	4.21	15.094
	50–60	63	0.71	0.78	6.82	4.11	14.845
	60–70	68	0.65	0.81	6.81	4.05	14.886
cGAN with Deep Residual VAE (Chandaliya and Nain, 2022)	20–30	76	0.48	0.77	7.42	4.59	16.312
	30–40	73	0.48	0.75	7.47	4.59	16.221
	40–50	75	0.61	0.76	7.31	4.62	16.123
	50–60	74	0.64	0.77	7.52	4.63	15.954
	60–70	75	0.62	0.805	7.51	4.62	16.122
DeepGAN for aging Progression (Atkale et al., 2021)	20–30	68	0.51	0.69	7.52	4.33	15.011
	30–40	65	0.59	0.61	7.77	4.41	14.715
	40–50	67	0.73	0.72	7.41	4.41	14.712
	50–60	61	0.55	0.71	7.72	4.34	14.312
	60–70	64	0.47	0.79	7.61	4.51	13.991
Proposed System	20–30	54	0.692	0.902	5.36	2.36	12.45
	30–40	55	0.705	0.705	5.25	2.45	12.35
	40–50	53	0.970	0.902	5.45	2.55	12.50
	50–60	53	0.690	0.905	5.55	2.65	12.65
	60–70	53	0.685	0.90	5.65	2.36	12.45

stands out for its ability to appraise image quality independently. Formulation of BRISQUE evaluation is mathematically portrayed as according to Eq. 27

$$I_{\text{norm}}(\mathbf{x}, \mathbf{y}) = \frac{I(\mathbf{x}, \mathbf{y}) - \mu_{\text{local}}(\mathbf{x}, \mathbf{y})}{\sigma_{\text{local}}(\mathbf{x}, \mathbf{y}) + 1} \quad (27)$$

In the formulation,  $I_{\text{norm}}(x, y)$  represents the normalized luminance at a given pixel  $(x, y)$ .  $I(x, y)$  is the original luminance at the pixel.  $\sigma_{\text{local}}(x, y)$  and  $\mu_{\text{local}}(x, y)$  and are the local standard deviation and mean and of the luminance in a neighborhood around the pixel, providing the local normalization.

#### 4.2.4 Quality assessment criteria for sketch comparisons

Table 5 provides a general guideline to categorize the quality of sketch comparisons into ‘Good’, ‘Medium’, and ‘Bad’ based on the values of PSNR, SSIM, and MSE.

#### 4.2.5 Experimental evaluation of generated sketches

Extensive experimental investigations were conducted on two prominent datasets, CUHK and AR, emphasizing referential metrics across diverse age groups: 20–30, 30–40, 40–50, 50–60, and 60–70 years. The results of these thorough analyses are

systematically outlined in Table 6. This table not only illustrates the outcomes of experiments within these age brackets but also illuminates the effectiveness and precision of the methodologies when applied to these distinct datasets. Consequently, it provides a nuanced understanding of performance across various age groups.

In a comprehensive analysis of the CUHK and AR datasets across diverse age groups (20–30, 30–40, 40–50, 50–60, 60–70) and benchmarked against established quality standards, notable trends emerge in Table 6. PSNR values consistently fall within the ‘Good’ category, indicating significant potential for signal clarity. In contrast, SSIM scores consistently rank high in the ‘Good’ range, showcasing the dataset’s effectiveness in preserving structural integrity in image comparisons. MSE values hover in the ‘Medium’ range, suggesting a moderate level of error and room for accuracy improvement. NIQE scores vary between ‘Good’ and ‘Medium’ across age groups, reflecting naturalness in image quality with some variability. Lastly, BRISQUE scores predominantly fall into ‘Good’ and ‘Medium’ categories, indicating satisfactory spatial quality with minimal distortion. These findings provide comprehensive insights into dataset performance, emphasizing strengths in structural similarity and spatial quality, and identifying areas for improvement, particularly in enhancing signal quality and reducing error rates.

### 4.3 Visual representation of the experimental outcomes

The presented gallery in Figure 5 showcases age-progressed portraits and their de-aged output rendered through the proposed system. To age the person, we have used the Age Progression Manipulator and the SAM (Style-based Age Manipulation) system, both open-source tools. The former employs image morphing algorithms, while the latter utilizes a style-based regression model to illustrate the aging process across the age spectrum of 30–70 years.

### 4.4 Analysis of proposed system with existing techniques

This section offers a detailed comparative analysis of various facial aging techniques, assessing their effectiveness across a spectrum of age groups from 20 to 70 years. We compare selected state-of-the-art aging progression techniques with our proposed system using established metrics discussed in Section 4.2. The outcomes are documented in Table 7 for the CUHK dataset and Table 8 for the AR dataset.

The proposed system excels in both the CUHK, and AR, datasets, showcasing superior metrics across all evaluated categories. In the CUHK, dataset, it achieves FID, scores as low as 38, PSNR, up to 0.87, and SSIM, as high as 0.865. For the AR, dataset, it maintains strong performance with FID, scores around 53–55, PSNR, peaking at 0.97, and SSIM, reaching up to 0.905. These results demonstrate the system’s effectiveness in producing highly realistic and accurate facial aging effects, highlighting its robustness

and potential as a leading solution in facial aging technology for diverse applications.

## 5 Conclusion and future scope

In conclusion, this study showcases the transformative potential of de-aging networks integrated with sketch generation to enhance forensic facial recognition. The proposed system effectively counters the challenges of aging in facial recognition, skillfully de-aging and sketching faces to boost identification accuracy and reliability. The system, tested with the CUHK and AR Face Sketch Databases, demonstrates significant improvements in realism as training progresses. The de-aging module’s integration marks a notable advancement, evident in the lower FID scores (41.7 for CUHK, 60.2 for AR), higher SSIM (0.789 for CUHK, 0.692 for AR), and improved PSNR (20.26 for CUHK, 19.42 for AR). Additionally, the system comprises a deepfake-based neural network for de-aging, adept at handling faces of any gender, race, and culture, and a pix2pix-based GAN for sketch generation that preserves structural and contrast details. This dual-module approach produces both hand-drawn and composite sketches with enhanced realism and consistency, outperforming state-of-the-art methods. The research underlines the necessity for continual innovation in de-aging technology, vital for societal and age invariant forensic applications.

The study presented in this research article opens several avenues for future research, particularly in refining and expanding the capabilities of AI-driven facial transformation technologies. The rapid evolution of deep learning techniques presents opportunities to enhance the accuracy and efficiency of de-aging and sketch generation processes. For instance, future studies could explore the integration of reinforcement learning to dynamically adjust model parameters in real-time, optimizing the quality of generated images based on iterative feedback. Additionally, incorporating adversarial training methods could further refine the realism of age-transformed faces by encouraging the model to generate features that are indistinguishable from real human attributes. Another promising direction could involve the utilization of multi-modal data inputs, such as combining visual data with biological age indicators, to enrich the model’s understanding and handling of age-related facial changes. This could lead to more personalized and context-aware de-aging applications, enhancing their utility in forensic and entertainment industries. Finally, research could also focus on reducing computational demands and improving the scalability of these technologies to facilitate their adoption in mobile devices and low-resource settings, making advanced de-aging tools more accessible to a broader audience.

### Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

JM: Conceptualization, Methodology, Project administration, Writing—original draft, Writing—review and editing. MS: Conceptualization, Methodology, Writing—original draft, Writing—review and editing. NP: Formal Analysis, Writing—original draft, Writing—review and editing. LS: Resources, Supervision, Writing—review and editing.

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## References

- Abate, A. F., Nappi, M., Riccio, D., and Sabatino, G. (2007). 2D and 3D face recognition: a survey. *Pattern Recognit. Lett.* 28 (14), 1885–1906. doi:10.1016/j.patrec.2006.12.018
- Akram, A., and Khan, N. (2023). US-GAN: on the importance of ultimate skip connection for facial expression synthesis. *Multimedia Tools Appl.* 83, 7231–7247. doi:10.1007/s11042-023-15268-2
- Amin, J., Sharif, M., Yasmin, M., Saba, T., Anjum, M. A., and Fernandes, S. L. (2019). A new approach for brain tumor segmentation and classification based on score level fusion using transfer learning. *J. Med. Syst.* 43, 1–16. doi:10.1007/s10916-019-1453-8
- Atkale, D. V., Pawar, M. M., Deshpande, S. C., and Yadav, D. M. (2021). “Residual network for face progression and regression,” in *Techno-societal 2020: proceedings of the 3rd international conference on advanced technologies for societal applications* (Cham: Springer International Publishing), 257–267. doi:10.1007/978-3-030-69921-5\_27
- Atkale, D. V., Pawar, M. M., Deshpande, S. C., and Yadav, D. M. (2022). Multi-scale feature fusion model followed by residual network for generation of face aging and de-aging. *Signal, Image Video Process.* 16 (3), 753–761. doi:10.1007/s11760-021-02015-z
- Basu, A., Mondal, R., Bhowmik, S., and Sarkar, R. (2020). U-Net versus Pix2Pix: a comparative study on degraded document image binarization. *J. Electron. Imaging* 29 (6), 063019. doi:10.1117/1.jei.29.6.063019
- Bocheva, G., Slominski, R. M., and Slominski, A. T. (2019). Neuroendocrine aspects of skin aging. *Int. J. Mol. Sci.* 20 (11), 2798. doi:10.3390/ijms20112798
- Chandaliya, P. K., and Nain, N. (2022). PlasticGAN: holistic generative adversarial network on face plastic and aesthetic surgery. *Multimedia Tools Appl.* 81 (22), 32139–32160. doi:10.1007/s11042-022-12865-5
- Chandaliya, P. K., and Nain, N. (2023). AW-GAN: face aging and rejuvenation using attention with wavelet GAN. *Neural Comput. Appl.* 35 (3), 2811–2825. doi:10.1007/s00521-022-07721-4
- Donato, L., Cecchi, R., Dagoli, S., Treglia, M., Pallocci, M., Zanovello, C., et al. (2023). Facial age progression: review of scientific literature and value for missing person identification in forensic medicine. *J. Forensic Leg. Med.* 100, 102614. doi:10.1016/j.jflm.2023.102614
- Du, L., Hu, H., and Wu, Y. (2019). Cycle age-adversarial model based on identity preserving network and transfer learning for cross-age face recognition. *IEEE Trans. Inf. Forensics Secur.* 15, 2241–2252. doi:10.1109/tifs.2019.2960585
- Du, L., Hu, H., and Wu, Y. (2019). Age factor removal network based on transfer learning and adversarial learning for cross-age face recognition. *IEEE Trans. Circuits Syst. Video Technol.* 30 (9), 2830–2842. doi:10.1109/tcsvt.2019.2923262
- Fernandes, S. L., and Josemin Bala, G. (2016). “Image quality assessment-based approach to estimate the age of pencil sketch,” in *Proceedings of fifth international conference on soft computing for problem solving: SocProS 2015* (Singapore: Springer), 1, 633–642. doi:10.1007/978-981-10-0448-3\_52
- Fernandes, S. L., and Josemin Bala, G. (2018). Matching images captured from unmanned aerial vehicle. *Int. J. Syst. Assur. Eng. Manag.* 9, 26–32. doi:10.1007/s13198-016-0431-5
- Gu, J., and Ye, J. C. (2021). AdaIN-based tunable CycleGAN for efficient unsupervised low-dose CT denoising. *IEEE Trans. Comput. Imaging* 7, 73–85. doi:10.1109/tci.2021.3050266

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Gupta, S. K., and Nain, N. (2023). Review: single attribute and multi attribute facial gender and age estimation. *Multimedia Tools Appl.* 82 (1), 1289–1311. doi:10.1007/s11042-022-12678-6

Henry, J., Natalie, T., and Madsen, D. (2021) *Pix2Pix GAN for image-to-image translation*. San Mateo, United States: Research Gate Publication, 1–5.

Henry, J. D., Grainger, S. A., and von Hippel, W. (2023). Determinants of social cognitive aging: predicting resilience and risk. *Annu. Rev. Psychol.* 74, 167–192. doi:10.1146/annurev-psych-033020-121832

Jin, B., Xu, S., and Geng, W. (2018). Learning to sketch human facial portraits using personal styles by case-based reasoning. *Multimedia Tools Appl.* 77, 5417–5441. doi:10.1007/s11042-017-4457-8

Jo, Y., and Park, J. (2019). “Sc-fegan: face editing generative adversarial network with user’s sketch and color,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 1745–1753.

Kokila, R., Sannidhan, M. S., and Bhandary, A. (2017). “A study and analysis of various techniques to match sketches to mugshot photos,” in *2017 international conference on inventive communication and computational technologies (ICICCT)* (IEEE), 41–44.

Kokila, R., Sannidhan, M. S., and Bhandary, A. (2017). “A novel approach for matching composite sketches to mugshot photos using the fusion of SIFT and SURF feature descriptor,” in *2017 international conference on advances in computing, communications and informatics (ICACCI)* (IEEE), 1458–1464.

Kyllonen, K. M., and Monson, K. L. (2020). Depiction of ethnic facial aging by forensic artists and preliminary assessment of the applicability of facial averages. *Forensic Sci. Int.* 313, 110353. doi:10.1016/j.forsciint.2020.110353

Martis, J. E., Shetty, S. M., Pradhan, M. R., Desai, U., and Acharya, B. (2023). Text-to-Sketch synthesis via adversarial network. *Comput. Mater. Continua* 76 (1), 915–938. doi:10.32604/cmc.2023.038847

Olivier, N., Baert, K., Danieau, F., Multon, F., and Avril, Q. (2023). Facetunegan: face autoencoder for convolutional expression transfer using neural generative adversarial networks. *Comput. Graph.* 110, 69–85. doi:10.1016/j.cag.2022.12.004

Pallavi, S., Sannidhan, M. S., Sudeepa, K. B., and Bhandary, A. (2018). “A novel approach for generating composite sketches from mugshot photographs,” in *2018 international conference on advances in computing, communications and informatics (ICACCI)* (IEEE), 460–465.

Pehlivan, H., Dalva, Y., and Dundar, A. (2023). “Styleres: transforming the residuals for real image editing with stylegan,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1828–1837.

Pezzini, C., Ciardo, S., Guida, S., Kaleci, S., Chester, J., Casari, A., et al. (2023). Skin ageing: clinical aspects and *in vivo* microscopic patterns observed with reflectance confocal microscopy and optical coherence tomography. *Exp. Dermatol.* 32 (4), 348–358. doi:10.1111/exd.14708

Rafique, R., Nawaz, M., Kibriya, H., and Masood, M. (2021). “DeepFake detection using error level analysis and deep learning,” in *4th international conference on computing and information sciences (ICIS)* (IEEE), 1–4.

Richardson, E., Alaluf, Y., Patashnik, O., Nitzan, Y., Azar, Y., Shapiro, S., et al. (2021). “Encoding in style: a stylegan encoder for image-to-image translation,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2287–2296.

- Sain, A., Bhunia, A. K., Yang, Y., Xiang, T., and Song, Y. Z. (2021). "Stylemeup: towards style-agnostic sketch-based image retrieval," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8504–8513.
- Sannidhan, M. S., Prabhu, G. A., Chaitra, K. M., and Mohanty, J. R. (2023). Performance enhancement of generative adversarial network for photograph-sketch identification. *Soft Comput.* 27 (1), 435–452. doi:10.1007/s00500-021-05700-w
- Sannidhan, M. S., Prabhu, G. A., Robbins, D. E., and Shasky, C. (2019). Evaluating the performance of face sketch generation using generative adversarial networks. *Pattern Recognit. Lett.* 128, 452–458. doi:10.1016/j.patrec.2019.10.010
- Seow, J. W., Lim, M. K., Phan, R. C., and Liu, J. K. (2022). A comprehensive overview of Deepfake: generation, detection, datasets, and opportunities. *Neurocomputing* 513, 351–371. doi:10.1016/j.neucom.2022.09.135
- Wang, S. H., Fernandes, S. L., Zhu, Z., and Zhang, Y. D. (2021). AVNC: attention-based VGG-style network for COVID-19 diagnosis by CBAM. *IEEE Sensors J.* 22 (18), 17431–17438. doi:10.1109/jsen.2021.3062442
- Yang, Y., Hossain, M. Z., Gedeon, T., and Rahman, S. (2022). "S2FGAN: semantically aware interactive sketch-to-face translation," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 1269–1278.
- Zhang, T. (2022). Deepfake generation and detection, a survey. *Multimedia Tools Appl.* 81 (5), 6259–6276. doi:10.1007/s11042-021-11733-y
- Zhao, B., Cheng, T., Zhang, X., Wang, J., Zhu, H., Zhao, R., et al. (2023). CT synthesis from MR in the pelvic area using Residual Transformer Conditional GAN. *Comput. Med. Imaging Graph.* 103, 102150. doi:10.1016/j.compmedimag.2022.102150
- Zhao, X., Yu, H., and Bian, H. (2023). Image to image translation based on differential image Pix2Pix model. *Comput. Mater. Continua* 77 (1), 181–198. doi:10.32604/cmc.2023.041479