IJIRR

International Journal of Information Research and Review
Vol. 12, Issue, 03, pp.7963-7974, March, 2025

RESEARCH ARTICLE

A NOVEL CONDITIONAL GAN FRAMEWORK FOR REALISTIC AGED FACE SYNTHESIS (AFS)

Logeswari Saranya, R.[1*] and Umamaheswari, K. [2]

[1]Research Scholar, Department of Information Technology, PSG College of Technology, Coimbatore, India
[2]Professor, Department of Information Technology, PSG College of Technology, Coimbatore, India

ARTICLE INFO

Article History:
Received 18th December, 2024
Received in revised form
24th January, 2025
Accepted 25th February, 2025
Published online 17th March, 2025

Key words:

Age Classification, Face Image Synthesis, Pre-Processing, Multi Task Cascaded Convolutional Neural Network, Conditional Generative Adversarial Networks.

ABSTRACT

Modeling the human face's aging process is crucial for cross-age face verification and recognition, garnering increasing attention due to its diverse applications in areas such as cross- age recognition and entertainment. Potential uses include aiding in the identification of lost children or predicting future appearances. However, the scarcity of labeled facial data spanning long age ranges presents a significant challenge. Additionally, due to varying aging rates among individuals, many existing models focus on synthesizing faces within broader age groups rather than predicting a specific age. Most methods primarily highlight prominent changes, like wrinkles and facial shape, which often fail to preserve individual identity. This research proposes a novel approach that preserves facial identity while simulating aging using Conditional Generative Adversarial Networks. By employing this technique, more realistic and identity- consistent facial images are generated. The model is further evaluated using identity preservation metrics and age classification, supported by user studies on face verification and age estimation.

INTRODUCTION

This research focuses on developing a Generative Adversarial Network (GAN) capable of synthesizing an aged facial image of an individual within a specified age range based on their younger image. The primary goal is to preserve the individual's identity in the generated image, ensuring it can be accurately recognized by face recognition systems. Face aging, which involves predicting a person's future appearance, is important in fields such as information forensics and security, as facial features typically change with age. This research aims to leverage advancements in technology to predict the future appearance of individuals, addressing challenges like recognizing people after many years or distinguishing between individuals with similar facial features. An application that can predict a person's future appearance while maintaining their identity would be valuable in various domains, including cross-age face recognition, locating missing individuals, and entertainment. It could also be used in the film industry for generating aged appearances of celebrities or in face verification systems for official documents, reducing the need for frequent photo updates. Existing facial aging models often require large datasets for training, leading to high computational costs and time inefficiency, with outputs that are often blurry or fail to preserve the individual's identity. Additionally, many models do not capture the full facial region or produce sharp, realistic images. Entertainment applications, for instance, often add random wrinkles and hair color changes, but these do not accurately reflect the individual's identity. This research aims to address these shortcomings by developing a more efficient and identity-preserving facial aging model.

Literature Survey

The references stated below are used for this survey to select the best model for generation of aged facial image and to identify the best dataset for this research work. Traditional methods of modeling face age progression can be roughly divided into two categories: physical model-based and prototype-based methods. Physical model-based methods mechanically simulate the changes of facial appearance over time, such as muscles and facial skins, via a set of parameters. However, physical model-based methods are usually computationally expensive and do not generalize well due to the mechanical aging rules.

Corresponding author:** ***Logeswari Saranya, R.
Research Scholar, Department of Information Technology, PSG College of Technology, Coimbatore, India.

On the contrary, prototype-based methods compute the average faces of people in the same age group as the prototypes. As a result, the testing face can be aged by adding the differences between the prototypes of any two age groups. The main problem of prototype-based methods is that the personalized features cannot be preserved well due to the use of average faces. Lanitis presented a model for the growing face where the training set is estimated on a model space for face recognition. The face model thus formed contained 50 model parameters and is a combination of shape and intensity model. Then the model parameters are converted to new sets of parameters to be consistent to target age after age estimation using suitable aging function [1]. Park proposed a generic method that consists of a 3D aging model to improve the face recognition performance. They used pose correction steps and separate modeling for shape and texture [2]. Nowadays, some commercial and academic institutions designed distinct deep networks, such as FaceNet is designed by Google, Oxford research group designed VGGNet, DeepFace is architecture by Facebook and DeepID is designedby CUHK group. It initially utilizes deep network for feature extraction of faces and then performs classification. DeepFace applied an integrated deep neural network using face alignment in the preprocessing step. In the classification, DeepID uses a Joint Bayesian classifier to make the classification more vigorous. FaceNet exploits very deep networks to perform face recognition. It uses nearly 8 million images of 2 million people and applies the triple loss strategy to train the network. DeepFace model applies a network trained by 4 million images. Here we need to point out that face recognition in DeepFace is a two-step process [3][4]. Yogita Mahajan, Shanta Sondur in paper Ageing Face Recognition Using Deep Learning makes use of the VGGNet CNN model and mainly focuses only on the particular region of the face for feature extraction. The drawback of this model is High computation time and uses large dataset. Focuses only on certain features. It was trained on the FGNet dataset [5]. Xiao presented a novel method for face recognition using a combination of texture and shape descriptors, called the Biview face recognition algorithm. For texture feature subspace learning methods are used and graphs are constructed for shape topology for face images. Chi-square measure is used as a difference (dissimilarity) measure to calculate the distance between two histograms [6].

Ling proposed a discriminative method for face verification over progression of age. In this approach, GO (Gradient Orientation) and GOP (Gradient Orientation Pyramid) is used for feature description and SVM (Support Vector Machine) for a classification which results in images in two groups as intra subject and inter subject [7]. Luefei-Xu, Luu, Savvides, Bui, and Suen, proposed Aging face particular recognition where WLBP (Walsh-Hadamard Transform Encoded Binary Pattern) used for feature extraction only on preprocessed particular region and UDP (Unsupervised Discriminant Projection) is used on WLBP image to build subspaces. Particular region is a dense part of face; this WLBP feature remains constant for particular region [8]. Wang introduced a recurrent face aging (RFA) framework based on arecurrent neural network.

They employ a two-layer gated recurrent unit as the basic recurrent module whose bottom layer encodes a young face to a latent representation and the top layer decodes the representation to a corresponding older face. The framework first maps the faces into Eigen face subspace and then utilizes a recurrent neural network to model the transformation patterns across different ages smoothly. As one of many supervised- based face aging methods, it requires massive paired faces of the same subject over a long period for training, which is impractical. Generally, these techniques require sufficient age sequences as the training data, which limits these methods' practicality [9]. Face Ageing with Contextual GANs by Liu, S.; Sun, Y.; Zhu, D.; Bao, R.; Wang, W.; Shu, X.; Yan, S., used skip layers along with an age discriminator and transition pattern disc and worked on combining 5 datasets such as CACD, FGNET, MORPH, LFW, SUP. The main limitation was that only adjacent age groups can be predicted and large dataset is trained. This model was evaluated based on common people survey [10]. Age Progression/Regression by Conditional Adversarial Auto encoder (CAAE) by Zhang, Z.; Song, Y.; Qi, H., was trained on MORPH and CACD dataset where all the pooling layers in the convolutional layers were replaced withstrides. The setback of this methodology was that the images produced are very blurry and not very realistic [11]. Quang T. M. Pham, Janghoon Yang and Jitae Shin in Semi-Supervised Face GAN for Age Progression and Regression trained on UTKFace dataset and used a methodology similar to CAAE but replaces auto encoder with Unet architecture consisting of skip connections. The model was evaluated using Multitask Cascaded Convolutional Network with a Confidence score of 94%. But this model does not learn local aging features [12].

Face Ageing with Conditional Generative Adversarial Networks by Antipov, G.; Baccouche, M.; Dugelay, J. This paper proposed a methodology where they used FaceNet CNN which is a pretrained model as their base structure and the main parameter which they considered was the Euclidean distance. This model was trained on the IMDB Wiki Cleaned dataset with an accuracy of 82.9% which was measured using the openFace face recognition software. The limitation of this methodology was that it needed a large dataset with heavy preprocessing along with a high computation time [13]. Zhizhong Huang, Shouzhen Chen, Junping Zhang, Hongming Shan Progressive Face Ageing With GAN – was modeled using residual skip connections with binary gates in the Convolutional layers to pass only the relevant feature points to the upcoming layers. The dataset which they used was MORPH along with CACD and achieved a verification confidence of 92.57% in FACE++ software [14]. Above stated insights show that a lot of research has been done in the field of Aged Facial Image Synthesis but still, it demands a lot of efforts by developing new techniques or improving the existing technique. Many of the proposed models don't seem to be so successful in preserving the identity of the person and to generate clear images. On the other side GAN models have shown great improvement compared to the normal deep learning techniques.

Methodological Framework for Aged Face Synthesis (AFS): The primary objective of this research is to develop a model that synthesizes aged facial images based on younger images. Achieving this requires integrating insights from age classifier models and style transfer techniques for pre-processing. In the proposed method, the input image is passed through a discriminator with conditioning, while a noise vector with conditioning is fed into the generator. The generator produces a synthesized image, which is evaluated by an age classifier before being sent to the discriminator for further training.

The proposed method includes a detailed examination of the dataset used and the models developed for feature extraction. The architecture of the Convolutional Neural Network (CNN) for the age classifier, generator, and discriminator is discussed in depth, along with the selection of parameters such as loss functions and optimizers to provide clarity on the model's design. Face aging, also referred to as age synthesis or progression, and involves rendering facial images with natural aging or rejuvenation effects. While Generative Adversarial Networks (GANs) have shown promise in generating high-quality synthetic images, existing Conditional GAN (cGAN) models typically use a single network to learn aging effects across different age groups. These models often struggle to meet the three key requirements of face aging: high image quality, accurate aging effects, and preservation of identity, especially when there are large age gaps, resulting in ghosting artifacts. This research seeks to address these limitations by focusing on synthesizing faces whose target age falls within a specified age group, rather than aiming for an exact age. By grouping faces within the target age range, the model transfers aging patterns to the input face while preserving the individual's identity, ensuring compatibility with face recognition software. Let G (z, x age) represent the generator, where: z is a noise vector sampled from a distribution (e.g., Gaussian), x age is an image of a person from a particular age group. Let D(x) represent the discriminator, which receives an input x and outputs the probability of x being a real image.

The generator receives both the noise vector z and the image x age, and outputs the generated image G (z, x age), aiming to create a realistic aged image that preserves identity is given by the Equation (1).

$$G: (z, x\ age) \longrightarrow \hat{x}_{age}\ \text{(generated aged image)} \tag{1}$$

The generator's objective is to minimize its loss, which is a combination of: Adversarial loss L_{adv}, which ensures the generated image is indistinguishable from real images by the discriminator. Identity loss L_{id}, which ensures the identity of the person is preserved. Age classification loss L_{age}, which ensures the generated image falls into the correct age category is represented by the Equation (2).

$$L_G = L_{adv}(G) + \lambda_1 L_{id}(G) + \lambda_2 L_{age}(G) \tag{2}$$

Where λ1 and λ2 are hyper parameters that control the trade-off between identity preservation and age accuracy.

- The generator image $\hat{x}$ is passed through as age classifier C ($\hat{x}$), which outputs the predicated age group. If the predicated age group does not match the target, the generator is refined to minimize using the Equation (3).

$$L_{age} = \text{CrossEntropy}\,(C(\hat{x}_{age}), \text{target age group}) \tag{3}$$

- The discriminator receives two inputs: x_{real}, a real image from the same age group as the target, and $\hat{x}$ the generated image. The discriminator's loss L_D is minimized when it can correctly classify real images x_{real} as real and generated images $\hat{x}$ as fake is given in the Equation (4).

$$L_D = -\,[\log D(x_{real}) + \log(1 - D(\hat{x}_{age}))] \tag{4}$$

- The generator and discriminator are trained in an adversarial manner: The generator tries to fool the discriminator by generating images that look real. The discriminator tries to detect whether the image is real or generated. The training continues until the discriminator age cannot distinguish between real and generated images, $D(x_{real}) \sim D(\hat{x}) \sim 0.5$.

- After training, the discriminator D is discarded. The generator G produces the final images $\hat{x}_{final}$ that passes the age classifier and preserves the person's identity is given in the Equation (5).

$$\hat{x}_{final} = G(z, x_{age}) \tag{5}$$

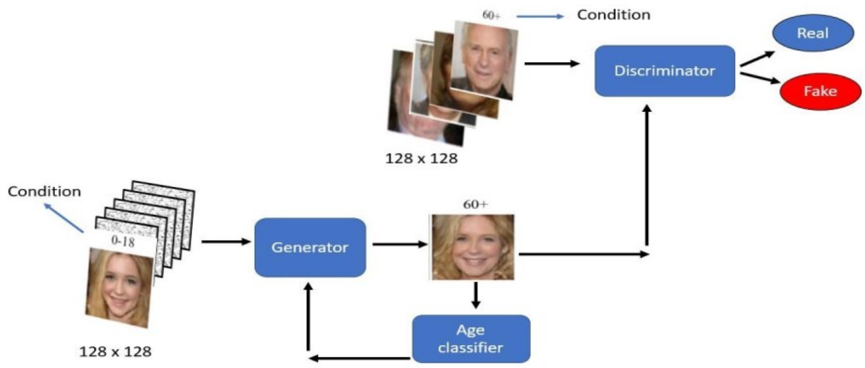

Fig. 1. Work Flow of Proposed AFS Model

The key mathematical components are the generator, discriminator, loss functions, and iterative refinement process that optimizes for realism, identity preservation, and age classification accuracy. The complete work flow is shown in Fig. 1.

Pre-processing Pipelines: Pre-processing is a critical step in most research work, as only a clean and well-prepared dataset can produce accurate and reliable results. Given the large size of the dataset, which includes numerous images with varying poses and lighting conditions, data augmentation is unnecessary. The primary pre-processing steps involve face detection, alignment, and center cropping. All these tasks were performed using the Multi-Task Cascaded Convolutional Neural Network (MTCNN), which is specifically used for face detection.

Let *I* be the input image. Let *B*= {(x1, y1), (x2, y2), (x3, y3), (x4, y4)} represents the coordinates of the bounding box around the detected face. Let *P*= {(x left eye, y left eye), (x right eye, y right eye), (x nose, y nose), (x left mouth corner, y left mouth corner), (x right mouth corner, y right mouth corner)} represents the 5 key facial landmarks detected.

Use the MTCNN algorithm to detect one or more faces in image *I*. Each detected face corresponds to a bounding box B_i and a set of key points P_i for each face *i.* Face Selection criteria is given as, if more than one face is detected, let C=(x_c, y_c) represents the center of image *I,* and let d_i represent the distance from the center to the face *i*, calculated as given in the Equation(6).

$$d_i = \sqrt{(x_{\text{center}} - x_i)^2 + (y_{\text{center}} - y_i)^2} \tag{6}$$

Select the face *i** is represented by the Equation (7).

$$i^* = \arg\min_i d_i \tag{7}$$

If no face ids detected, the image is discarded. Align the detected face using a similarity transformation based on the key points P_i. This can be represented as a transformation matrix T, ensuring proper face orientation by aligning the eyes and mouth. After alignment, the detected and aligned face image is resized to a fixed dimension is represented by the Equation (8).

$$I_{resized} = \text{resize}\,(I, 400 \times 400) \tag{8}$$

Novel Age Classification Framework: The architecture consists of five convolutional layers, three max-pooling layers, two fully connected layers, and a soft max output layer. This configuration was selected based on insights from a comprehensive literature review, as it is intended to serve as a pre-trained module within the final GAN model. The input size, output size and kernel numbers and window size are listed in Table 2. The Convolutional Neural Network (CNN) was trained using images from the CACD dataset.

Table 1. Parameters of CNN Layers in Age Classifier

LAYER	Input Size	Output Size	KERNEL and size
Conv1	227*227*3	55*55*96	96 - 11*11
MaxPooling	55*55*96	27*27*96	3*3
Conv2	27*27*96	27*27*256	256 - 5*5
MaxPooling	27*27*256	13*13*256	3*3
Conv3	13*13*256	13*13*384	384 - 3*3
Conv4	13*13*384	13*13*384	384 - 3*3
Conv5	13*13*384	13*13*256	256 - 3*3
MaxPooling	13*13*256	6*6*256	3*3
Linear FC1	6*6*256	4096	
Linear FC2	4096	4096	
Softmax output	4096	5	

The training process that combines the use of the Adam Optimizer, cross- entropy loss, and a pre-trained Age Classifier to enhance the performance of a GAN (Generative Adversarial Network).The Adam optimizer updates the model's parameters based on gradient descent using the first and second moments of the gradients is given in the Equation (9).

$$\theta_{t+1} = \theta_t - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \tag{9}$$

Where

- θ_t are the model parameters at time step t,
- α is the learning rate,
- $\hat{m}_t$ is the biased-corrected first moment estimate (mean of gradients),
- $\hat{v}_t$ is the biased-corrected second moment estimate (variance of gradient)
- ϵ is a small constant to prevent division by zero.

The Cross-entropy loss function, used to measure the difference between the predicted class probability distribution and the true class distribution, is defined in the Equation (10).

$$L_{\mathrm{CE}} = -\sum_i y_i \log(p_i) \tag{10}$$

Where

- y_i is the true label (1 for the correct class and 0 for the others),
- p_i is the predicted probability for class i.

The Age Classifier is used to assign a loss based on whether the generated face is classified correctly into the desired age group (Group C or Group 2 Category: 31-40). The loss, denoted as L_{age}, is small if the generated face belongs to the correct age group and large otherwise is represented in Equation (11).

$$L_{\mathrm{age}} = -\log(p_{\mathrm{age\ group\ C}}) \tag{11}$$

Where

$p_{\mathrm{age\ group\ C}}$ is the probability of the generated face being classified in Group C. The overall loss function for training the GAN combines the adversarial loss L_{GAN} and the age classification loss L_{age} is given in the Equation (12).

$$L_{\mathrm{total}} = L_{\mathrm{GAN}} + \lambda \cdot L_{\mathrm{age}} \tag{12}$$

Where

- L_{GAN} is the standard GAN loss (either a minimization of binary cross-entropy for the discriminator or generator),
- L_{age} is the age classification loss,
- λ is a hyperparameter controlling the contribution of the age classification loss.

This total loss function guides the GAN's training, ensuring that the generated faces align with both the desired age group and the overall goal of fooling the discriminator. The model was evaluated using different training and test data split ratios, including 70:30, 80:20, and 90:10, with the highest accuracy observed at a 90:10 split, utilizing a batch size of 512 and a learning rate of 0.0001. The loss value decreased significantly from 1.603 to 0.116, while accuracy improved from 26.1% to 91.5%. To ensure that generated faces fall within the target age group (Group 2)-31 to 40

Generative Adversarial Network (GAN) Module for Enhanced Data Synthesis

Novel Generator Module: The Generator module is composed of four convolutional layers, each followed by batch normalization, and two de-convolutional layers, as detailed in Table 3. Table 3 also provides a comprehensive overview of the layer configuration, including kernel sizes, input, and output dimensions. In the generator module, we utilize a mechanism analogous to a one-hot encoding scheme.

Specifically, $\mathbf{F} \in \mathbb{R}^{h \times w \times c}$ let represent the feature maps, where h and w denote the height and width, respectively, and c indicates the number of feature maps. Each feature map F_k is defined in the Equation (13).

$$\mathbf{F}_k = \begin{cases} 1, & \text{if } k = \text{condition index} \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

$\mathbf{I} \in \mathbb{R}^{h \times w \times d}$ Consequently, for k = 1, 2, ….., *c,* only one features map Fk is populated with ones, while all other maps are set zero. Prior to the first convolutional layer, we introduce conditions by concatenating the input images $\mathbf{I} \in \mathbb{R}^{h \times w \times d}$with the condition feature maps F. The concatenation is represented in the Equation (14).

$$C = \text{concat}(I, F) € R^{h \times w \times (d+c)} \tag{14}$$

where d is the number of input channels. The combined feature maps C are subsequently processed through the first convolutional layer defined by a set of convolutional $\mathbf{O} = \text{Conv}(\mathbf{C}, \mathbf{K}) \in \mathbb{R}^{h' \times w' \times m}$filters, where kh and kw are the height and width of the filters, respectively, and m denote the number of output channels. The convolution operator can be mathematically expressed in the Equation (15).

$$\mathbf{O} = \text{Conv}(\mathbf{C}, \mathbf{K}) \in \mathbb{R}^{h' \times w' \times m} \tag{15}$$

where $h^{/}$ and $w^{/}$ are the dimensions of the output after applying the convolution, and Conv(.) denotes the convolution operation followed by any applicable activation function.

Table 2. Parameters of CNN Layers in Generator

LAYER	In channel	Out channel	KERNEL size	Stride	Padding
Conv1	8	32	7	1	0
Batch Norm	32				
Conv2	32	64	3	2	0
Batch Norm	64				
Conv3	64	128	3	2	0
Batch Norm	128				
6 blocks each of 2 conv layer of in=128, out=128, stride=1, padding=1					
deConv1	128	64	3	2	1
Batch Norm	64				
deConv2	64	32	3	2	0 [out=1]
Batch Norm	32				
Conv4	32	3	7	1	0

Optimizing Discriminator Modules: The discriminator is a critical component in GANs, responsible for guiding model improvement based on the computed loss values. The discriminator module is comprised of five convolutional layers, each with batch normalization, except for the first and last layers. A kernel size of 4 and a stride of 2 are consistently applied across all convolutional layers, as outlined in Table 4. This architecture allows the discriminator to effectively evaluate the generated data and provide feedback for refining the overall model.

Table 3. Parameters of CNN Layers in Discriminator

LAYER	In channel	Out channel	KERNEL size	Stride
Conv1	3	64	4	2
Adding condition vector (64+5=69)				
Conv2	69	12	4	2
Batch Norm	128			
Conv3	128	256	4	2
Batch Norm	256			
Conv4	256	512	4	2
Batch Norm	512			
Conv5	512	512	4	2

The discriminator D is a classifier that distinguishes between real images x and generated images G (z) (where z is a latent variable). The objective of D is to maximize the probability of assigning the correct label (real or fake) to both real and generated images. The loss function for the discriminator is typically represented by the Equation (16).

$$L_D = -\mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] - \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \quad (16)$$

Where

- $p_{data}(x)$ is the distribution of real images.
- $p_z(z)$ is the distribution of the latent variable z.
- $D(x)$ is the probability that x is a real image.

The generator G aims to produce images $G(z)$ that the discriminator cannot distinguish from real images. The generator tries to minimize the probability of the discriminator correctly classifying generated images as fake.The loss function for the generator is given by the Equation (17).

$$L_G = -\mathbb{E}_{z\sim p_z(z)}[\log D(G(z))] \quad (17)$$

This loss function encourages the generator to improve it output such that D (G (z)) is a close as possible to 1, meaning the discriminator is "fooled". The GAN model involves a mini max game between the generator and discriminator is given by the Equation (18).

$$\min_G \max_D V(G,D) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \quad (18)$$

The discriminator maximizes this objective, while the generator minimizes it. The iterative training process continues until the discriminator D reaches an optimal point where D (x) = D (G (z)) = ½, meaning it classifies both real and generated images as equally real, achieving equilibrium. A key factor in preserving a person's identity in this model is not just the optimal selection of the CNN architecture and its parameters, but also the design of the loss functions, which play a crucial role. Adversarial loss alone is insufficient to maintain identity, as it only ensures that the generator produces samples that match the target data distribution, potentially generating images of any individual within the target age group. Therefore, adversarial loss cannot guarantee that the identity information of the original subject is retained. To address this, a perceptual loss is introduced into the Aged Face GAN model, ensuring that the synthesized facial images resemble the original input and maintain key identity features. The key components of the perceptual loss function, feature selection from CNN layers, and the overall loss function. This is achieved by selecting the most relevant features from the appropriate CNN layers, a critical step in preserving identity, particularly for passing face recognition tests. While common aging effects like wrinkles, hair color changes, and hair thinning may alter appearance, the individual's identity should remain unchanged. Based on insights from the literature, lower CNN layers are effective at preserving content, whereas higher layers capture style-related attributes such as color and texture. Thus, the perceptual loss function is formulated as the sum of the age loss from the age classifier, feature loss from the classifier, and the GAN loss from both the generator and discriminator, calculated using Mean Squared Error (MSE). This relationship continuously updates the GAN parameters during each iteration, resulting in a stable and efficient model that successfully preserves the individual's identity.

Experimental Results and Performance Evaluation: This chapter is presented with the results obtained from each process of the experiment along with the overall view with the working.

Dataset Used: Several face image datasets are commonly utilized in the field of facial image synthesis and recognition research, including the IMDB-Wiki Dataset, FG-Net, MORPH, Cross-Age Celebrity Dataset (CACD), Labeled Faces in the Wild (LFW), UTK Face Dataset, and KAN Face. For this study, the Cross-Age Celebrity Dataset (CACD) was selected as the primary dataset, sample image as illustrated in Fig. 2.

Fig. 2. Sample Dataset Image

Table 4. Age Category Split

Age Range	11-20	21-30	31-40	41-50	50+
No.of images	25,098	36,662	38,736	35,768	26,974

The Cross-Age Celebrity Dataset (CACD) contains 163,446 facial images of approximately 2,000 celebrities, spanning an age range from 16 to 62. This dataset was created by collecting images through Google Image Search, with certain controlled conditions applied. As a result, CACD exhibits significant variations in pose, illumination, and expression, making it more challenging compared to other datasets. Each image is labeled with the age, celebrity's name, and a unique photo number. The distribution of images across different age categories is shown in Table 1, illustrating that the dataset is relatively evenly distributed across all age groups.

Pre-Processing Result: Pre-processing was carried out using a Multi-Task Cascaded Convolutional Neural Network (MTCNN) to detect, align, and resize the facial images. Fig. 3 illustrates the CACD dataset, which consists of 163,446 raw images before pre-processing, alongside an example of a pre-processed image resized to 400 x 400 pixels. Fig. 3. Performance Comparison with and without Pre-processing.

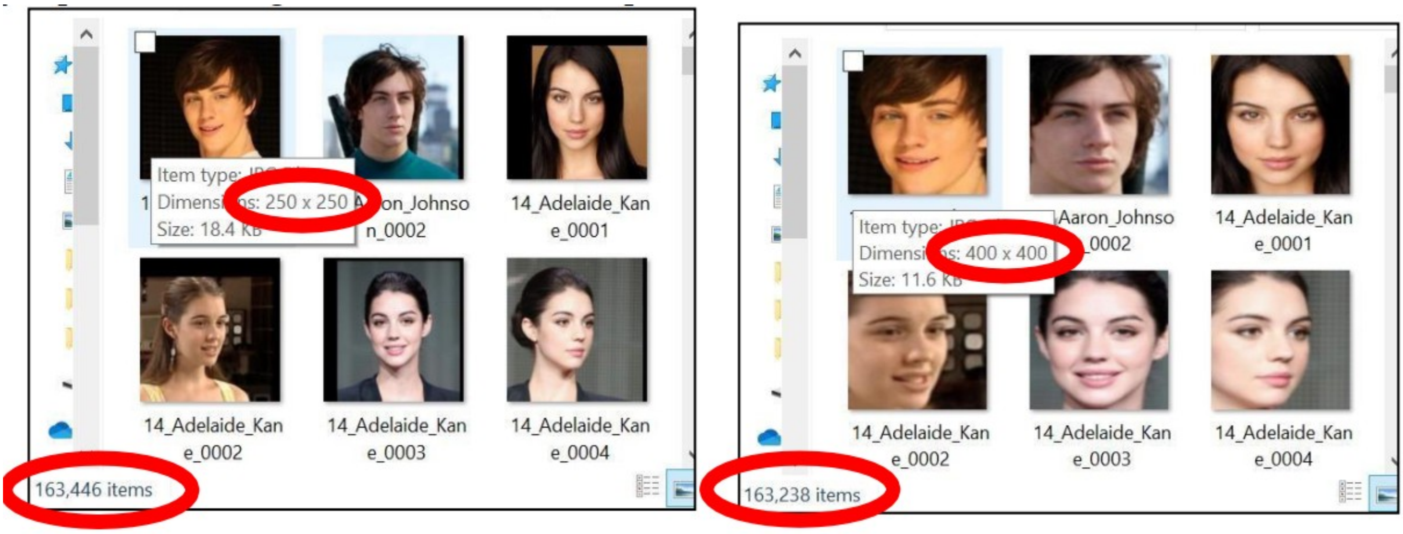

Fig. 3. Performance Comparison with and without Pre-processing

As illustrated, the face region is zoomed in while the background is minimized, and the total image count has been reduced to 163,238 due to the removal of high-noise images and those with undetectable faces. The pre-processed CACD dataset is divided into 90% for training and 10% for testing.

Performance Measures Analysis

In this subsection, a brief note on the metrics to evaluate the performance of face aging methods are discussed

- **Aging Accuracy:** Refers to the accuracy achieved by the classifier in guiding the generator to synthesize facial images that correspond to the target age category.
- **Execution Time:** Measures the time required for the model to generate four new images across different age categories.
- **Inception Score:** Provides a quantitative evaluation of the image quality.
- **Identity Preservation:** Assesses whether the individual's identity is maintained during the face aging process, as verified by face recognition software.

Despite the diverse range of input faces, covering variations in gender, pose, makeup, and expression, the generated aged faces exhibit photorealism, with natural details such as skin texture, muscle definition, and wrinkles. While hair color typically turns white as the face ages, this varies between individuals and is influenced by race and the training data, which accounts for some generated faces showing fewer aging effects. For the analysis, 100 images were randomly selected, and four aged images were generated for each input, resulting in a total of 400 aged images. These generated images were subsequently used for evaluation. For face verification, recognition, and identity preservation, the Face++ API was employed. Face++ is an AI open platform that provides computer vision technologies through APIs and SDKs. It computes a confidence score between two images—one being the real image and the other the generated aged facial image of the same individual, as demonstrated in Fig. 4.

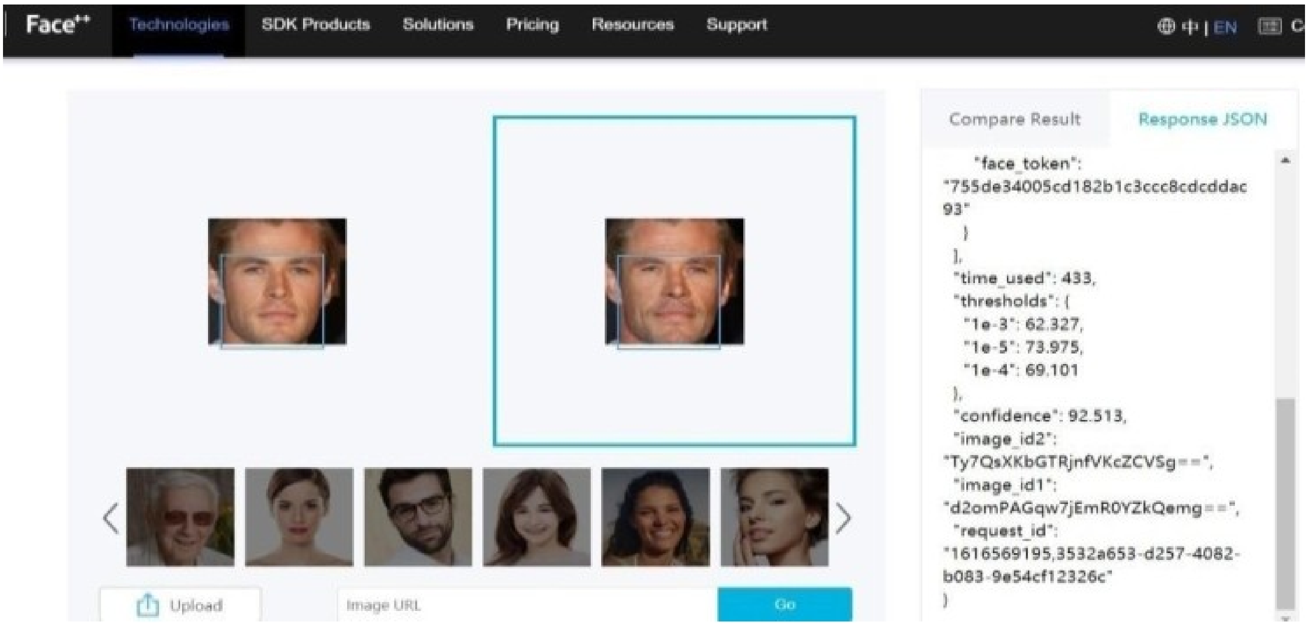

Fig. 4. Result of Face Recognition Using Face++

The Inception Score is another metric commonly used for evaluating the quality of generated images. However, since the Inception network trained on Image Net focuses on 1000 object classes that do not include human faces or related categories, applying this metric to facial images is inappropriate. Instead, the pre-trained VGG network for faces can be used as an alternative for evaluation. As proposed in the referenced paper, OpenAI's source code was utilized to compute this score, referred to as the VGG-Face Score.

Performance Evaluation of Proposed Model: A trial-and-error approach was employed to determine the optimal dataset split ratio for model development. The model was trained using three different split ratios: 70:30, 80:20, and 90:10. Accuracy and loss values for each ratio were recorded, as detailed in Table 5. Based on these results, the 90:10 split ratio was chosen for its superior accuracy as shown in Fig. 5. Consequently, 90% of the data was allocated for training, while 10% was reserved for testing.
Table 5 Training and Testing Accuracy of Age Classifier

Table 5 Training and Testing Accuracy of Age Classifier

Training Data	Testing Data	Training Accuracy	Testing Accuracy	Loss
70%	30%	0.921	0.893	0.201
80%	20%	0.934	0.907	0.169
90%	10%	0.939	0.915	0.116

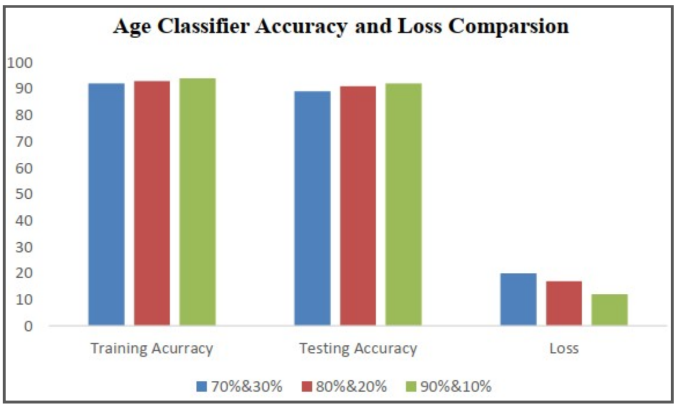

Fig. 5. Accuracy Level obtained based on Training and Testing Split Up

Table 6. Dataset Split for Training and Testing

Dataset		Age Category				
Category	Split	0 (11-20)	1 (21-30)	2 (31-40)	3 (41-50)	4 (50+)
Training	90%	22588	32996	34863	32192	24276
Testing	10%	2510	3666	3873	3576	2698
Total	100%	25098	36662	38736	35768	26974

The pre-processed dataset was divided into 90% for training and 10% for testing. This split was carefully managed to ensure even distribution across all age categories, thereby avoiding any bias towards a specific age group. The distribution of the dataset for training and testing is detailed in Table 6.

During the training process, validation was performed on the generator for each epoch to assess the quality of the newly synthesized images. These validation images illustrate the progression of the GAN model's learning pattern. Fig. 6 displays the source batch of images used for validation, along with the transition between different age categories within the same epoch. This visualization demonstrates how the model adjusts its parameters to handle various age groups. The images shown are from epoch 0, and comparisons across different epochs reveal significant changes in the model's learning and training progress. The final model was saved and subsequently visualized using Tensor Board, which provides a graphical representation of the model's architecture. Tensor Board allows users to examine each sub-module in detail and observe how the input data transforms into the output.

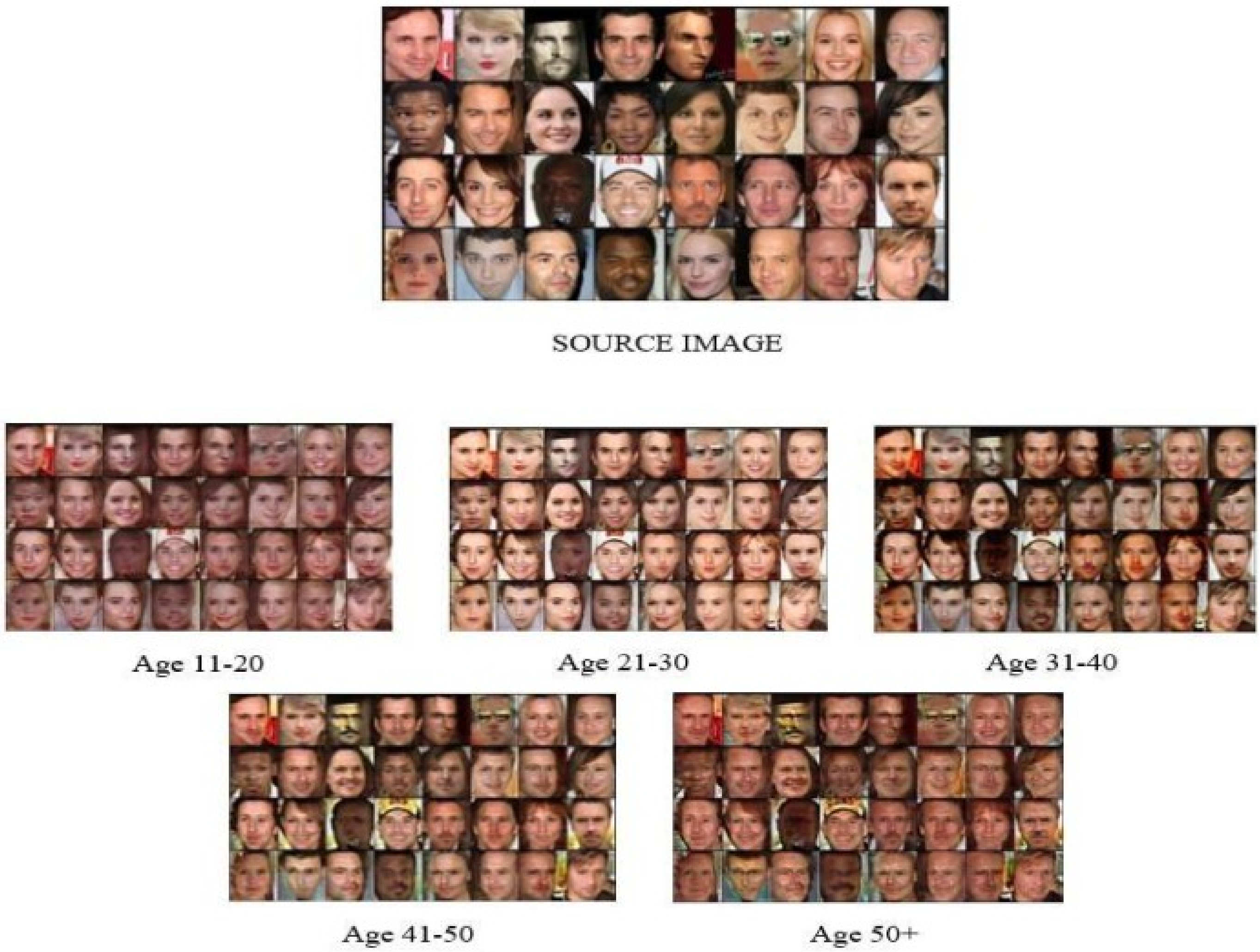

Fig. 6. Epoch 0 idx_7000 Validation Images

Similarly, the loss values for both the discriminator and the generator were recorded and visualized as graphs to monitor their gradual changes and eventual saturation over time. Fig. 7 presents these loss values, with pink representing the discriminator loss and blue representing the generator loss. The x-axis denotes the iteration steps, while the y-axis shows the loss values.

Fig. 7. Tensor Board Graph for Loss Values

The graph illustrates a steady decline in the discriminator loss and a gradual improvement in the generator loss. Upon zooming in, the graph reveals that both loss values reach a saturation point towards the end of the iterations, indicating an optimal stopping point for the model's training. The GAN model was trained for over 3 hours to achieve this level of saturation.

The model's evaluation was conducted based on the attributes outlined in the previous section, which are also detailed in Table 7 for comparison with the IPCGANs reference paper. The proposed model demonstrates significant improvements across all metrics and exhibits greater stability compared to IPCGANs is shown in Fig. 8.

Fig. 7. Tensor Board Graph for Loss Values

Parameters	IPCGANS	Synthesis of Aged Facial Image and Analysis
Face Verification (%)	88.90	91.82
Age Classification (%)	85.24	91.5
VGG-Face Score	36.33±1.85	47.26±0.25
Execution Time (s)	0.28	0.26

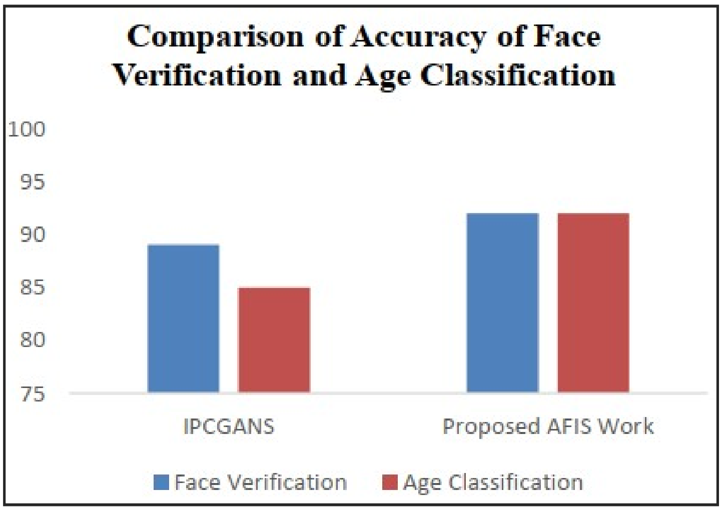

Fig. 8.Comparison of Proposed AFIS Work with Existing Method

CONCLUSION

In conclusion, an efficient Conditional Generative Adversarial Network (CGAN) model for synthesizing aged facial images was developed using a CNN-based architecture, combined with a pre-trained age classifier that achieved approximately 92% accuracy. The model's performance, particularly in face recognition of the synthesized aged images, was analyzed and found to deliver a high level of accuracy. Since the research utilized a celebrity dataset, the model performs optimally on faces with makeup. However, it can be retrained on standard facial images for broader applicability. The CACD dataset, chosen for this study, includes a wide variety of images with different poses and lighting conditions, making it suitable for this research. The accuracy of the model is closely tied to the dataset size, and increasing the dataset can further improve accuracy. Future work will focus on refining the model to enhance accuracy and reduce training time. Currently, even with GPU assistance, the model requires a substantial amount of time for training. This can be mitigated by optimizing the code for TPU compatibility or reducing the training dataset size. Additionally, the model could be adapted for age regression, enabling it to generate younger facial images from older input images.

REFERENCES

Lanitis, "Modeling the process of ageing in face images", Seventh IEEE International conference on Computer Vision,Kerkyra, volume 1, 31-136,1999, DOI: 10.1109/ICCV.1999.791208.

Park, U. "Age-Invariant Face Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 5, 947-954, 2010, DOI: 10.1109/TPAMI.2010.14.

Schroff, F. "Facenet: A Unified Embedding for Face", 815-823, 2015, DOI: 10.1109/CVPR.2015.7298682.

Taigman, Y. "DeepFace: Closing the Gap to Human-Level Performance in Face Verification", International Conference on Computer Vision and Pattern Recognition (CVPR), 2014, DOI: 10.1109/CVPR.2014.220.

Yogita Mahajan; Shanta Sondur, "Aging Face Recognition Using Deep Learning", International Journal of Engineering and Applied Sciences (IJEAS) ISSN: 2394-3661, Volume-5, Issue-8, August 2018, DOI: 10.31873/ijeas.5.8.17.

Xiao, "Biview Face Recognition in the Shape–Texture Domain", Elsevier Journal on Pattern Recognition, vol. 46, no. 7, 1906-1919, 2013, DOI: https://doi.org/10.1016/j.patcog.2012.12.009.

Ling, H. "Face Verification Across Age Progression Using Discriminative Methods", IEEE Transactions on Information Forensics and Security, vol.5, no. 1, 82-91, 2010, DOI: 10.1109/TIFS.2009.2038751.

Luu, "Investigating Age Invariant Face Recognition Based on Periocular Biometrics", IEEE Joint Conference on Biometrcs, 2011, DOI: 10.1109/IJCB.2011.6117600.

Wang,W.; Cui, Z.; Yan, Y.; Feng, J.; Yan, S.; Shu, X.; Sebe, N., "Recurrent Face Aging", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA,pp. 2378–2386, 27–30 June 2016.

Liu, S.; Sun, Y.; Zhu, D.; Bao, R.; Wang, W.; Shu, X.; Yan, S., "Face Aging with Contextual Generative Adversarial Nets", In Proceedings of the ACM Multimedia, Mountain View, CA, USA, 23–27 October 2017.

Zhang, Z.; Song, Y.; Qi, H., "Age Progression/Regression by Conditional Adversarial Autoencoder", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017 DOI: https://doi.org/10.48550/arXiv.1702.08423.

Quang T. M. Pham; Janghoon Yang; Jitae Shin, "Semi- Supervised FaceGAN for Face-Age Progression and Regression with Synthesized Paired Images", MDPI – Electronics and Artificial Intelligence, 2020, DOI: https://doi.org/10.3390/electronics9040603.

Antipov, G.; Baccouche, M.; Dugelay, J., "Face aging with conditional generative adversarial networks", In Proceedings of the IEEE International Conference on Image Processing (ICIP), Beijing, China, pp. 2089– 2093, 17–20 September 2017, DOI: https://doi.org/10.48550/arXiv.1702.01983.

Zhizhong Huang; Shouzhen Chen; Junping Zhang; Hongming Shan, "PFA-GAN: Progressive Face Aging with Generative Adversarial Network", IEEE Transactions on Information Forensics and Security, 2020, DOI: https://doi.org/10.48550/arXiv.2012.03459.

****