A Project Report

On

"Age Progression and Regression using xAI-CPAVAE with explainability"

by

Tisha Choksi ET21BTAI007

Jashn Jain ET21BTAI021

Taashna Jariwala ET21BTAI026

Pranavi Mehta ET21BTAI041

under the mentorship of

Prof (Dr.) Dhruti Sharma

Associate Professor, Information Technology Department

October 2024

DEPARTMENT OF INFORMATION TECHNOLOGY/ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

SARVAJANIK UNIVERSITY

SR. R K DESAI MARG, OPP. MISSION HOSPITAL, ATHWALINES, ATHWA, SURAT, GUJARAT 395001

CERTIFICATE

This is to certify that the Project Report submitted by **Tisha Choksi (ET21BTAI007), Jashn Jain (ET21BTAI021), Taashna Jariwala (ET21BTAI026), and Pranavi Mehta (ET21BTAI041)** to the **SARVAJANIK UNIVERSITY** for the partial fulfillment of the subject credit requirements is a bonafide work carried out by the student.

This is to further certify that I have been supervising the Project of **Tisha Choksi (ET21BTAI007), Jashn Jain (ET21BTAI021), Taashna Jariwala (ET21BTAI026), and Pranavi Mehta (ET21BTAI041).**

The contents of this report, in full or in parts, have not been submitted to any other Institute or University for award of any degree, diploma or titles.

Sign of Faculty Mentor :

Name of Faculty Mentor: Prof (Dr.) Dhruti Sharma

Date:

ACKNOWLEDGEMENT

First and foremost, we would like to express our deepest gratitude to **Professor Dhruti Sharma** for her exceptional guidance and mentorship throughout the course of this project. Her unwavering support and insightful advice have been instrumental in navigating the complexities of our research.

We would also like to extend our heartfelt appreciation to **Professor Praveen K. Chandaliya** for sharing his profound expertise and knowledge with us. His extensive experience and invaluable guidance have played a crucial role in shaping the direction and outcome of our project. Professor Chandaliya's dedication to teaching and his commitment to our success have been a source of great inspiration.

Their combined efforts and mentorship have not only enhanced the quality of our work but have also empowered us with the confidence and skills necessary to excel in this field. We sincerely thank them for their substantial contributions, continuous encouragement, and for believing in our potential.

Name of Student : Tisha Choksi, Jashn Jain, Taashna Jariwala, and Pranavi Mehta

Enrollment No : ET21BTAI007, ET21BTAI021, ET21BTAI026, and ET21BTAI041

ABSTRACT

The project aims to develop an advanced model for age progression and rejuvenation using Explainable Artificial Intelligence (xAI) within the Conditional Perceptual Adversarial Variational Autoencoder (CPAVAE) framework. Traditional methods for facial aging, such as Generative Adversarial Networks (GANs) and Autoencoders, often face challenges like mode collapse, interpretability issues, and high data dependency. This report proposes integrating xAI-based feedback into CPAVAE to enhance both the quality and transparency of the generated facial transformations. By utilizing xAI techniques like LIME and saliency maps, the model provides interpretable feedback, making it more efficient and reliable for tasks such as missing child identification and forensic investigations. The proposed hybrid framework ensures high-quality, identity-preserving transformations, while reducing computational costs and improving data efficiency. The results aim to contribute to applications in security, healthcare, and entertainment, where both accurate aging predictions and explainable models are critical.

List of Figures

Table of Contents

1. Introduction

Generative models play a critical role in computer vision, enabling the creation of realistic images, transformations, and data augmentations for various applications. Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) are two of the most widely used frameworks in generative modeling. GANs generate new data by learning the distribution of input data, while VAEs focus on modeling data by mapping inputs to a latent space, which helps with both image reconstruction and sampling. Each framework brings unique strengths, but combining these architectures in a meaningful way remains a challenge. Conditional Perceptual Adversarial Variational Autoencoder (CPAVAE) is an innovative approach that bridges these techniques, integrating VAE-based sampling with GAN-style adversarial learning to achieve high-quality conditional transformations [1]. However, a significant limitation of CPAVAE is its black-box nature, which limits transparency in training and restricts insight into how transformations occur.

Recent advancements in explainable AI (xAI) techniques aim to tackle the challenge of GAN interpretability, making generative models more transparent. One such example is the Interpretable Face Aging (ICAAE) framework, which introduces LIME-based perturbations into the Conditional Adversarial Autoencoder (CAAE). LIME generates explanations by modifying the latent space and observing the effect of these perturbations on aging transformations, helping the model focus on the most important facial regions. The corrective feedback provided by LIME guides the adversarial training, improving the quality of generated images by highlighting discrepancies between real and synthetic faces. This is highly relevant to the proposed integration of CPAVAE and xAI-GAN, where similar xAI techniques can improve both interpretability and the effectiveness of facial transformations [3].

Working of CPAVAE. The CPAVAE architecture consists of several components, including an encoder, a generator (decoder), two discriminators, and a perceptual loss network. The process begins by encoding the input image into a latent vector zzz, which captures essential attributes of the input (such as facial features) in a compact form. The latent vector is then combined with conditional age labels to guide the generator in producing transformed images corresponding to different stages of age progression or regression. For example, given a photo of a child, CPAVAE can generate both older and younger versions by altering the age conditioning vector while preserving the child's identity [1]. This identity-preserving transformation is achieved through the use of perceptual loss, which ensures that the generated image maintains stylistic consistency with the input.

The two discriminators in CPAVAE serve distinct roles:

1. Discriminator D_z regularises the latent space to ensure it follows a smooth and meaningful distribution (often Gaussian), which improves the model's ability to generate diverse and realistic outputs.

2. Discriminator D_{img} ensures that the generated image appears realistic by distinguishing between real and synthetic samples.

Together, these components allow CPAVAE to perform tasks such as age progression and regression while overcoming common issues like mode collapse and blurry outputs. However, the model's adversarial nature and reliance on perceptual loss introduce complexity in training. Moreover, CPAVAE lacks interpretability—users cannot easily understand why certain transformations are successful or how the model learns, which limits its practical utility in high-stakes applications like forensic investigations [1].

Integration of xAI-GAN and CPAVAE. While CPAVAE excels at generating conditional transformations, it struggles with transparency and training efficiency, particularly in scenarios with limited labeled data. This is where xAI-GAN provides a complementary solution. xAI-GAN introduces interpretability into the training process by leveraging Explainable AI (xAI) systems to generate feedback that identifies the most important features influencing the discriminator's decisions [2]. In a traditional GAN, the generator only receives scalar loss feedback from the discriminator (i.e., whether the sample is classified as real or fake). However, xAI-GAN enhances this feedback by producing feature-level explanations using tools such as:

- LIME (Local Interpretable Model-Agnostic Explanations) [18]: Provides local feature importance by perturbing inputs and analyzing how the model's predictions change [2].
- SHAP (SHapley Additive exPlanations) [19]: Measures the contribution of individual features to the output [2].
- Saliency Maps: Highlight the most influential pixels in the image that affect the discriminator's decisions [2].

By incorporating feature-level feedback into the gradient update process, xAI-GAN improves training efficiency and ensures that the generator focuses on the most important aspects of the data. This is particularly useful for tasks like age progression, where specific facial features (e.g., wrinkles, facial structure) are crucial for realistic transformations. Moreover, xAI-GAN makes the training process interpretable by explaining why a particular transformation succeeded or failed, providing users with actionable insights to fine-tune the model.

Together, CPAVAE and xAI-GAN offer a powerful hybrid framework:

- CPAVAE provides conditional generative capabilities with identity-preserving transformations, making it ideal for tasks such as age prediction and face recognition across ages.
- xAI-GAN enhances training by offering interpretability and data efficiency, ensuring that the generator learns more effectively, even with limited datasets.

This paper proposes a new hybrid framework that integrates CPAVAE with xAI-GAN to deliver both high-quality conditional outputs and interpretable feedback. This combination will enable the development of robust generative models suitable for applications in forensics, healthcare, and predictive modeling, where both transparency and performance are essential.

1.1 Motivation

While GANs and VAEs have significantly advanced generative modeling, several challenges persist. Many existing GAN-based methods have been developed to tackle age progression and rejuvenation tasks, including models like AW-GAN, which utilizes attention mechanisms and wavelets for enhanced face aging and rejuvenation, DiffAge3D, a diffusion-based model for 3D-aware facial aging, and Triple-GAN, which introduces a progressive aging framework with triple translation loss. However, GAN-based models face multiple challenges, such as ensuring the quality of generated images, managing the large amount of training data required, and overcoming mode collapse, a common issue where the generator produces repetitive outputs and fails to capture the full diversity of the data.

In addition to GANs, autoencoder-based approaches have also been employed for facial transformations. Examples include CAAE (Conditional Adversarial Autoencoders), its improved version CAAE++, and CPAVAE (Conditional Perceptual Adversarial Variational Autoencoder). These methods leverage latent-space encoding to perform conditional transformations such as age progression. However, despite their strengths, autoencoder-based models face critical challenges, such as limited interpretability, high sensitivity to hyperparameter selection, and a tendency to overfit when trained on small datasets. Furthermore, like GAN-based methods, these approaches require large amounts of labeled data, which can be difficult to obtain, especially in forensic and real-world scenarios.

To address some of these issues, researchers have introduced xAI (Explainable AI)-based approaches aimed at reducing data dependency and improving interpretability. For example, Interpretable Face Aging enhances CAAE by integrating LIME-based explanations to clarify the influence of individual features on transformations. Similarly, xAI-GAN uses feature-level feedback mechanisms (e.g., SHAP and Saliency Maps) to make the learning process more transparent and data-efficient. Additionally, models like xAI-FR apply explainable AI methods to face recognition tasks, improving the transparency of decisions made by deep neural networks. However, despite their contributions, xAI-based models face limitations, such as the computational resources required, challenges related to inconsistent explainability metrics, and a risk of overfitting due to the added complexity of interpretability tools.

This research aims to address the shortcomings of these individual approaches by combining the conditional generative capabilities of CPAVAE with the interpretable feedback mechanisms of xAI-GAN. The proposed hybrid model will take advantage of CPAVAE's ability to generate identity-preserving transformations while using xAI-guided feedback to reduce data requirements and improve training transparency. This combination offers an opportunity to overcome common challenges in GANs and autoencoders, such as mode collapse, data dependency, and interpretability issues. By developing a framework that is both efficient and explainable, this research aims to create a system that can generate realistic facial transformations while ensuring trustworthy outputs—a necessity for forensic investigations, missing child identification, and healthcare applications.

1.2 Objective

The objective of this research is to develop a hybrid generative framework that integrates the strengths of Conditional Perceptual Adversarial Variational Autoencoder (CPAVAE) and xAI-GAN. The proposed model aims to deliver precise attribute-conditioned transformations, such as facial age progression and regression, while incorporating explainable training mechanisms to enhance transparency and usability. The framework's dual focus on identity-preserving transformations and interpretability makes it suitable for a range of applications in forensics, security, and predictive analytics.

Conditional Transformations with Facial Age Progression and Regression. A core objective of the proposed model is to facilitate age progression and regression, essential tasks in various fields. Age progression predicts the future appearance of an individual, while age regression estimates how the individual might have looked at a younger age. These tasks require precise control over age-based features to generate realistic images that maintain the subject's identity. The applications of facial age transformations are vast, spanning several domains:

- **Missing child identification**, where age progression helps estimate how a missing child might appear after several years.
- **Age-invariant verification** is used in biometric systems to ensure accurate recognition across different life stages.
- **Facial appearance prediction** for entertainment and media, such as character development in movies and video games.
- **National security operations**, where surveillance systems need to match faces despite age variations.

- **Information management systems**, such as population tracking databases, where ageing patterns must be modelled for administrative use.

Developing an effective hybrid model requires addressing the complexities involved in these transformations, such as ensuring that subtle changes (e.g., wrinkles, facial structure) are incorporated naturally while avoiding visual artefacts that compromise identity preservation.

Interpretability to Enable Transparent Decision-Making. Another objective of the framework is to introduce explainable feedback mechanisms that improve the transparency of the transformation process. Traditional generative models often produce black-box outputs, leaving users with little understanding of how or why a particular transformation was generated. In contrast, the integration of xAI feedback tools—such as LIME [18], SHAP [19], and Saliency Maps—will allow practitioners to trace the impact of individual features on the generated output. This feature-level insight ensures that users can monitor and refine the model's behavior, especially in applications where explanation and trustworthiness are critical, such as forensic investigations and biometric systems.

This transparency will also make it possible to justify predictions in high-stakes scenarios, such as matching the appearance of a missing child with an age-progressed version. Beyond forensics, these insights will empower users to optimize the model for specific scenarios by understanding which features influence transformations the most.

Achieving Data Efficiency through Targeted Learning. The hybrid model's efficiency will also improve its adaptability to real-world conditions, including varying facial poses, lighting conditions, and expressions. By aligning the strengths of xAI-GAN's interpretability with CPAVAE's identity-preserving transformations, the proposed system aims to generate outputs that are accurate, realistic, and reliable across different environments and constraints.

Generalizability and Broader Applications. Another key objective is to ensure that the hybrid model is flexible and generalizable across multiple domains. While the primary focus is on age-conditioned transformations, the framework will also be designed to support other attribute-based transformations, such as gender changes or style variations. The model will be evaluated using Fréchet Inception Distance (FID) and perceptual loss metrics to ensure that it maintains high visual fidelity across different use cases. Beyond forensics, the hybrid system’s potential applications include:

- Healthcare, where facial predictions could help track ageing-related health markers or anticipate changes over time.
- Data augmentation, where the model can generate realistic variations of data to enhance machine learning pipelines.
- Style transfer applications, such as transforming facial features for artistic purposes or cosmetic simulations.

By ensuring that the framework is robust, versatile, and explainable, this research aims to set a new standard for generative models, prioritizing both performance and transparency.

2. Literature Review

In recent years, the task of face age progression and rejuvenation has garnered significant attention due to its applications in areas such as identity verification, missing persons identification, and entertainment. Various deep learning models, particularly Generative Adversarial Networks (GANs) and Autoencoders (AEs), have been extensively explored for this purpose. They are explained as follows:

GAN-based Approaches. Generative Adversarial Networks (GANs) [8] have been widely adopted for face ageing and rejuvenation tasks. Notable models include:

- **AW-GAN (Attention with Wavelet GAN) [9]:** Utilizes wavelet transforms to preserve high-frequency information during image ageing and rejuvenation. However, GAN-based methods like AW-GAN struggle with the quality of generated images, often introducing artefacts due to insufficient training data and susceptibility to *mode collapse*, where the generator outputs limited variations in images.
- **DiffAge3D (Diffusion-based 3D-aware Face Aging) [10]**: Incorporates diffusion processes to generate 3D facial aging effects, leading to improved realism. Yet, similar to AW-GAN, the need for large amounts of high-quality training data and the potential for *mode collapse* remain significant challenges.
- **Triple-GAN [11]**: Introduces triple translation loss, facilitating progressive face aging. Despite advancements in image quality, GAN-based models generally compromise interpretability for accuracy. This lack of transparency in decision-making creates a barrier for sensitive applications like missing person identification.

Autoencoder-based Approaches. Autoencoders (AEs) offer an alternative to GANs for age progression and rejuvenation, but they have their limitations:

- **CAAE (Conditional Adversarial Autoencoder) [12]**: Combines autoencoders with adversarial learning to control age transformations. However, its heavy reliance on large datasets and its sensitivity to hyperparameters make it prone to overfitting. Moreover, it lacks the interpretability needed for transparent decision-making.
- **CAAE++ [13]**: An improvement over CAAE, it enhances the ageing process but still shares the limitations of requiring extensive training data.
- **CPAVAE (Conditional Perceptual Adversarial Variational Autoencoder) [14]**: Addresses some issues by introducing perceptual loss through VGG19 for high-quality face ageing. However, despite its effectiveness, the model still suffers from challenges related to interpretability and sensitivity to hyperparameters.

Explainable AI (xAI)-based Approaches. Given the limitations of GAN and autoencoder methods, xAI-based models have emerged as promising solutions:

- **Interpretable Face Aging [15]**: Enhances conditional adversarial autoencoders by integrating LIME explanations, making the decision-making process more transparent. However, inconsistent explainability metrics pose a challenge to the reliability of the interpretations.
- **xAI-GAN [16]**: Introduced in [2], xAI-GAN integrates explainable AI techniques such as saliency maps and SHAP into the GAN training process. By offering richer feedback from the discriminator, xAI-GAN improves both data efficiency and the interpretability of generated faces. Yet, despite these advantages, xAI-based models are resource-intensive and vulnerable to overfitting when dealing with large datasets.

- **xAI-FR [17]**: An xAI-based face recognition model that incorporates explainable AI methods into deep learning frameworks, providing more transparent decision-making. This model demonstrates improvements in interpretability but shares similar challenges with computational cost and inconsistent metrics across different datasets(xAI-CPAVAE).

Current approaches for age progression and rejuvenation, especially GAN and autoencoder models, face challenges such as mode collapse, data efficiency, and interpretability. xAI-based models have been developed to address these limitations, offering enhanced transparency and data efficiency. However, the high computational demands and inconsistent explainability metrics highlight the need for further refinement in xAI methodologies.

The proposed xAI-CPAVAE methodology aims to combine the strengths of xAI and CPAVAE to deliver more interpretable, data-efficient, and realistic face ageing and rejuvenation. By integrating xAI feedback into the CPAVAE framework, this model aspires to overcome the challenges of prior approaches while maintaining high facial feature fidelity and diverse ethnic representation

3. Problem Definition

In today's socio-political landscape, there is a growing need for accurate and objective methods for identifying individuals, especially in cases where age estimation is required but no valid birth records are available. This issue is compounded by the large number of missing persons and human trafficking victims worldwide. Many of these cases involve children, making it imperative to develop systems capable of tracking identity across time through facial changes. The challenges are not only technical but also operational, as law enforcement agencies and social welfare organizations need tools that are efficient, reliable, and transparent to aid in finding missing individuals.

Human trafficking statistics highlight the critical need for better identification systems. In 2024, 50% of trafficking cases in India involved minors, with over 1,300 cases reported in just the first half of the year, meaning around 650 victims were children [4]. The ability to accurately predict age progression and regression from historical photos could significantly improve efforts to find missing children, providing crucial insights into how a child's appearance might change over time. These systems would allow law enforcement agencies to search databases based on predicted facial transformations, narrowing the gap between outdated images and real-world appearances.

Existing facial recognition models used in law enforcement—such as Darpan and Bachpan Bachao Andolan's (BBA) Facial Recognition System (FRS)—have achieved some success in identifying missing children [5]. However, these tools often struggle with long-term appearance changes due to aging, especially when children go missing for extended periods. For instance, matching a photograph of a child taken several years earlier with their current appearance requires accurate age progression models. The lack of robust, interpretable generative systems limits the ability of these models to address complex changes in facial features caused by aging, lighting, pose variations, or expressions.

Current deep learning models for age progression and regression focus heavily on maximizing accuracy at the cost of interpretability, making it difficult for practitioners to understand and trust the outputs. This trade-off between interpretability and performance hinders real-world usability, particularly in high-stakes applications such as missing child identification, forensic investigations, and national security. There is a growing consensus among researchers and practitioners that future systems must go beyond black-box predictions and offer explainable models that provide transparent insights into the transformation process.

The proposed hybrid framework seeks to address these challenges by combining CPAVAE's conditional generative capabilities with xAI-GAN's explainable feedback mechanisms [1,2]. This new system will offer a more interpretable and efficient approach to facial transformations, meeting the dual need for performance and transparency in real-world scenarios. The model will provide age-conditioned transformations (e.g., predicting future appearances of missing children) while allowing practitioners to visualize and understand which features drive the transformations, making it easier to trust and refine the outputs.

4. Methodology

This section outlines the methodology for integrating the x-AI algorithm into the CPAVAE framework, aimed at enhancing the performance of child face image generation through improved training and latent space manipulation.

Methodology for xAI-CPAVAE. Our framework consists of seven key components: the encoder (E), sampling module (S), generator (G), discriminator for prior distribution (Dz), discriminator for reconstructed images (Dimg), an explainable loss network (xAI), and a perceptual loss network (Ploss). The encoder transforms the input image x into a latent vector z. The generator G then reconstructs the child face image $\bar{x}$, aiming for a close match to the ground truth. The discriminator Dimg engages in adversarial learning, differentiating between the real image x and the generated image $\bar{x}$. Meanwhile, Dz ensures that a uniform distribution is imposed on the latent vector z.

Both x and $\bar{x}$ are processed through the VGG19 network to calculate the perceptual loss. The outputs from Dimg, along with the complete parameter set of Dimg, are fed into the xAI component to generate an explainable matrix M. This matrix is subsequently utilized by the generator G and encoder E to provide insights into the parameters that influenced the discriminator's output.

The explainable matrix M enhances model performance by offering additional context on the significant features. During experimentation, various xAI systems, including Saliency Maps, LIME, DeepSHAP, and DeepLIFT [19], are evaluated to identify the most effective approach for the CPAVAE model.

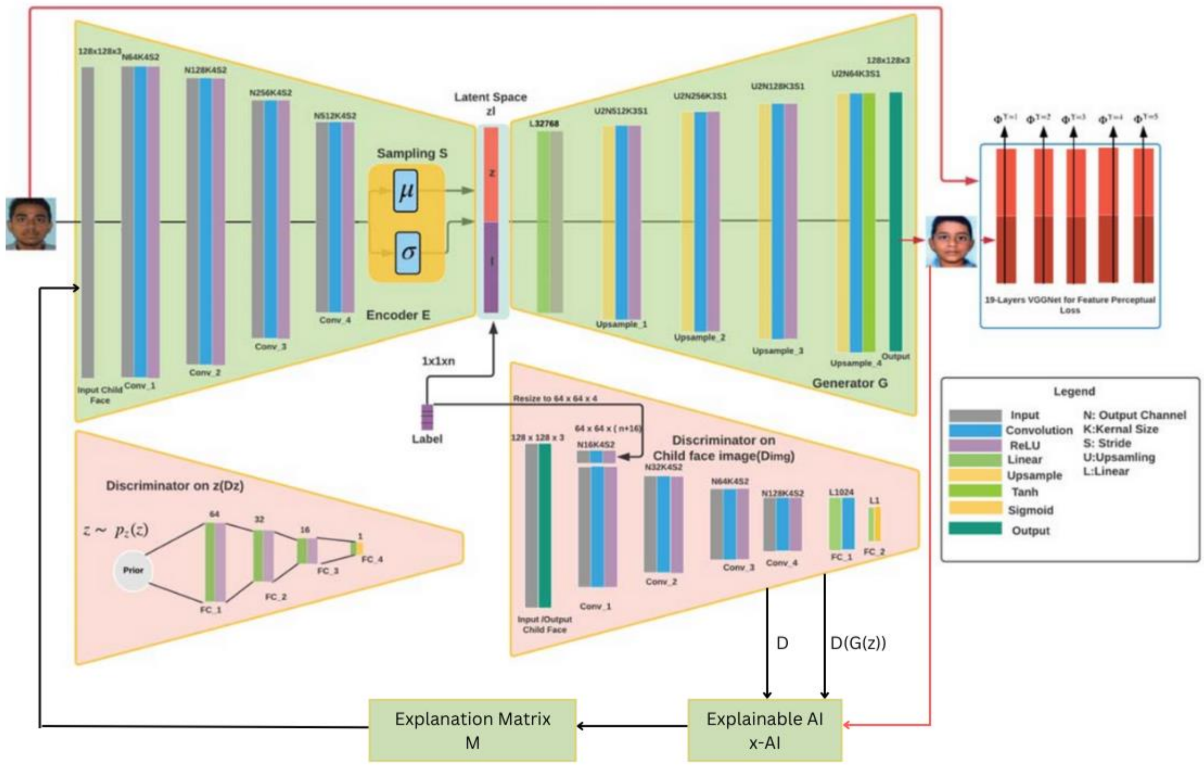

Figure 1. Block diagram of proposed xAI-CPAVAE approach for face rejuvenation and regression.

1. Overview of xAI-CPAVAE Architecture. The xAI-CPAVAE model consists of the following key components:

- **xAI Generator (G)**: Incorporates xAI techniques to improve the generation of child face images by utilizing explanations derived from the input data.
- **CPAVAE Framework**: Composed of an encoder (E), latent space sampling (S), discriminator networks (Dz and Dimg), and a perceptual loss network (Ploss).

2. Data Preprocessing

- **Image Normalization**: Input child face images will undergo normalization using batch normalization parameters (μ and σ) to ensure consistent representation.

- **Data Augmentation**: To increase the robustness of the model, various augmentation techniques (e.g., rotation, flipping, scaling) will be applied to the training images.

3. Generator Training Algorithm. The generator training process will be modified to include the x-AI guidance, as outlined below:

1. **Noise Sample Generation**: Random noise samples zzz are generated.
2. **Initial Loss Calculation**: Compute the initial loss LLL using the discriminator DDD with L=Loss(1−D(G(z)))L = Loss(1 - D(G(z)))L=Loss(1−D(G(z))).
3. **Discriminator Gradient Computation**: Calculate the gradients ΔD\Delta DΔD from the loss LLL.
4. **Generated Example Gradient**: Compute the gradient ΔG(z)\Delta G(z)ΔG(z) based on ΔD\Delta DΔD.
5. **x-AI Integration**: If the x-AI flag is true:
   - Compute the explanation matrix MMM using the x-AI system.
   - Modify the generated example gradient: Δ0G(z)=ΔG(z)+α×ΔG(z)×M\Delta^0 G(z) = \Delta G(z) + \alpha \times \Delta G(z) \times MΔ0G(z)=ΔG(z)+α×ΔG(z)×M.
   - Compute the final generator gradient ΔG\Delta GΔG from the modified gradient Δ0G(z)\Delta^0 G(z)Δ0G(z).
6. **Parameter Update**: Update the generator parameters θG\theta_GθG using the computed gradient ΔG\Delta GΔG.

4. CPAVAE Framework Components

1. **Encoder Network (E)**:
   - Encode input images into latent vectors z.

- Use a stack of convolutional layers with Batch Normalization to maintain feature consistency.

2. **Latent Space Sampling (S)**:
   - Combine age label vectors with latent vectors for controlled age transitions while preserving personality traits.
   - Implement a re-parameterization trick to ensure differentiability during sampling.
3. **Discriminator Networks**:
   - **Dz**: Enforce uniform distribution on latent vectors zzz using adversarial training.
   - **Dimg**: Differentiate between real and generated child face images to ensure realism.
4. **Perceptual Loss Network (Ploss)**:
   - Utilize VGG19 to calculate perceptual loss based on the spatial differences in generated and ground truth images, focusing on maintaining high-frequency details.

5. Loss Functions

- **Perceptual Loss**: Calculate perceptual loss Φ(x)\Phi(x)Φ(x) to ensure high-quality image generation.
- **Kullback-Leibler Divergence Loss**: Incorporate KL divergence to retain the texture of generated images.
- **Adversarial Loss**: Compute adversarial loss using both Dimg and Dz to refine the generator's output iteratively.

6. Training Process

1. **Initialize Parameters**: Randomly initialize parameters for the generator, encoder, and discriminator networks.

2. **Iterative Training**: Alternate between updating the generator and discriminator networks using their respective gradients and loss functions.
3. **Performance Evaluation**: Monitor the quality of generated images using metrics such as Inception Score (IS) and Fréchet Inception Distance (FID).
4. **Hyperparameter Tuning**: Experiment with different values of α and learning rates to optimize the training process.

7. Evaluation

- Conduct qualitative and quantitative evaluations of the generated child faces.
- Compare the performance of the xAI-CPAVAE model against traditional CPAVAE and other baseline methods.

5. Experimental Setup

We are using a multitude of datasets combined in order to improve the generalisability of our approach. The datasets being used are – the MRCD dataset (Multi-Racial Child Dataset) and the CLF dataset.

MRCD (Multi-Racial Child Dataset) was developed at DoCSE at Malaviya NIT. It contains 64,965 face images of 4 race groups (Asian-17,221 images, Black-13,354 images, White-19,297 images, Indian-15,103 images). Asian, Black, and White children dataset image (MRCD) to train the ChildGAN model, including web crawl and publicly collected images. The images are labelled in the format "age_genderId_sequenceID", where age is the age of the children, and genderId is the children's id, i.e., 0 or 1. For boys and girls, 0 and 1 are used as gender id, respectively. The MRCD dataset is used in ChildGAN [1]. 0-20 years boys and girls face images are available in this dataset. Multi-Racial Child Dataset (MRCD) containing 64,965 face images of four races (Asian, Black, White, and Indian).[1]

The Children Longitudinal Face (CLF) dataset contains 3,682 face images of 919 children, in the age range of 2 to 18 years. Each subject has an average of 4 images acquired over an average time lapse of 4 years (minimum time lapse of 2 years; maximum time lapse of 7 years). The demographic makeup of the CLF dataset is comprised of 604 (66%) boys and 315 (34%) girls. The face images were captured with a resolution of 354 ×472 pixels. [7]

Fig. 2. Example of longitudinal face data of 4 subjects where images were taken annually, in the CLF dataset. The age of image acquisition is given below each image.

6. Results

7. Future Scope

The proposed hybrid framework that integrates CPAVAE with xAI-GAN offers several exciting avenues for future research. As advancements in generative models and xAI continue to evolve, there are multiple areas where such frameworks can be extended, refined, and applied across various domains. However, to realize the full potential of these systems, several challenges need to be addressed, ranging from improving interpretability metrics and computational efficiency to expanding the applicability of conditional generative models beyond facial aging. Below, we explore the future scope and possible improvements based on insights from [1], [2], and [3].

Improving Interpretability and Explanation Metrics. As emphasized in the ICAAE framework, interpretability plays a key role in ensuring trust in generative models, especially in high-stakes applications such as forensics and law enforcement [3]. However, one of the significant challenges in current xAI-based methods is the inconsistency of explanation metrics. For example, different tools like LIME [18], SHAP [19], and Saliency map feature drive transformations, often making it difficult for practitioners to interpret outputs uniformly. Future research can focus on standardizing explainability metrics, allowing models to generate more consistent and actionable feedback across multiple scenarios.

Another area for improvement involves exploring new explainability techniques that offer more granular insights into how specific latent variables influence transformations. For instance, ICAAE introduces perturbations in the latent space to guide facial ageing processes [3]. Building on this, future models could incorporate adaptive perturbations that dynamically adjust explanations during training, ensuring that the feedback aligns more closely with the model's objectives. Additionally, combining multiple xAI methods (e.g., LIME with Saliency Maps) could yield more comprehensive explanations, improving the interpretability of complex transformations, such as those involving simultaneous age and expression changes.

Reducing Computational Overhead and Improving Data Efficiency. Both CPAVAE and xAI-GAN rely on complex adversarial and explainable AI architectures, which demand significant computational resources [1,2]. As the size and complexity of models increase, the need for efficient training techniques becomes more urgent. One promising direction for future research is the development of lightweight architectures that maintain performance without compromising on interpretability. For instance, ICAAE's use of LIME-based perturbations adds a layer of complexity to the adversarial training process [3]. To mitigate this, future research can explore model pruning, quantization, or knowledge distillation techniques to reduce computational overhead while retaining the benefits of xAI-guided feedback.

Another critical improvement involves reducing data dependency. Current models, including CPAVAE, require large labelled datasets for effective training, which may not always be available in real-world applications [1]. Future experiments can explore self-supervised or semi-supervised learning methods to reduce the dependency on labelled data. Additionally, xAI-based data augmentation strategies—where synthetic data generated with interpretable feedback is used to enhance real datasets—can improve both performance and generalizability. This approach would align with xAI-GAN's goal of data-efficient training [2].

Expanding the Scope of Conditional Transformations. While the focus of this research is on facial age progression and regression, the hybrid framework offers the potential to extend beyond age-based transformations. For example, future research could explore how the model can perform multi-attribute transformations—such as changing age, gender, or facial expressions simultaneously—to support more complex real-world applications. Insights from ICAAE and CPAVAE indicate that the use of latent space manipulation and adversarial learning can be extended to include additional conditional labels beyond age [1,3]. These advancements could unlock new use cases in personalized healthcare, identity protection, and entertainment applications.

Further, integrating temporal dynamics into the generative model could enable it to predict how facial features change over time, supporting applications such as chronic illness monitoring or predictive modelling for long-term surveillance. For instance, models could learn how different factors, such as stress or health conditions, affect facial ageing patterns over time, making them valuable tools in biometric systems or long-term healthcare management.

Enhancing Robustness for Real-World Conditions. To ensure real-world usability, the hybrid framework must be robust enough to handle diverse conditions, such as lighting variations, poses, and occlusions. Future research can focus on improving the model's generalization capabilities by integrating domain adaptation techniques, allowing it to perform well across different datasets and environments. Adversarial robustness is another critical area, as generative models are vulnerable to attacks that can manipulate outputs by introducing subtle changes in input data. Developing defensive mechanisms to detect and mitigate such attacks will be essential to ensure that the system performs reliably in critical applications like forensic investigations and child recovery efforts.

Additionally, future research could explore how adaptive learning mechanisms can be incorporated to ensure that the model improves over time as it encounters new data. For example, the ICAAE framework demonstrates that xAI-based corrective feedback can dynamically influence training by highlighting key facial features. A similar approach could be used in the proposed hybrid model to continuously refine its predictions, ensuring long-term adaptability across different use cases [3].

Generalizing the Framework to Other Domains. Although this research focuses on facial transformations, the proposed hybrid framework can be adapted for other generative tasks beyond age progression and regression. For instance, in healthcare, the framework could be applied to predict the progression of disease symptoms based on visual markers (e.g., skin conditions or facial asymmetry). Similarly, style transfer applications could benefit from interpretable generative models that transform visual features in real-time. The insights generated from xAI-based feedback would be particularly useful in artistic applications, where users need precise control over transformations.

Moreover, data augmentation with interpretable feedback can enhance machine learning pipelines by generating realistic variations of training data. This would be particularly beneficial for fields such as autonomous driving, where large datasets are required to train models for recognizing various environmental conditions. The use of xAI-guided explanations would ensure that the synthetic data aligns with real-world features, improving model performance across diverse tasks.

7.1. Conclusion

The hybrid framework that integrates CPAVAE and xAI-GAN opens several avenues for future exploration. Key areas include improving interpretability metrics, reducing computational overhead, and expanding conditional transformations beyond age-based attributes. Further research can also focus on enhancing robustness for real-world applications, ensuring the model performs reliably under diverse conditions. Additionally, by generalizing the framework to other domains, this research can contribute to advancements in healthcare, security, entertainment, and machine learning pipelines. With continued experimentation and innovation, the proposed hybrid framework has the potential to set new standards for interpretable and efficient generative modelling.

8. Bibliography

1. Chandaliya, P.K. and Nain, N., 2019, June. Conditional perceptual adversarial variational autoencoder for age progression and regression on child face. In *2019 International Conference on Biometrics (ICB)* (pp. 1-8). IEEE.
2. Nagisetty, V., Graves, L., Scott, J. and Ganesh, V., 2020. xai-gan: Enhancing generative adversarial networks via explainable ai systems. *arXiv preprint arXiv:2002.10438.*
3. Korgialas, C., Pantraki, E. and Kotropoulos, C., 2024, April. Interpretable Face Aging: Enhancing Conditional Adversarial Autoencoders with Lime Explanations. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5260-5264). IEEE.
4. Hindustan Times (2024). *Report*
5. National Crime Records Bureau (2024). *Annual Report on Human Trafficking Statistics in India*.
6. Chandaliya, P.K. and Nain, N., 2022. ChildGAN: Face aging and rejuvenation to find missing children. *Pattern Recognition*, 129, p.108761.
7. Deb, D., Nain, N. and Jain, A.K., 2018, February. Longitudinal study of child face recognition. In *2018 International Conference on Biometrics (ICB)* (pp. 225-232). IEEE.
8. Goodfellow, Ian, et al. "Generative adversarial networks." *Communications of the ACM* 63.11 (2020): 139-144.
9. Chandaliya, Praveen Kumar, and Neeta Nain. "AW-GAN: face aging and rejuvenation using attention with wavelet GAN." *Neural Computing and Applications* 35.3 (2023): 2811-2825.
10. Wahid, Junaid, et al. "DiffAge3D: Diffusion-based 3D-aware Face Aging." *arXiv preprint arXiv:2408.15922* (2024).
11. Fang, Han, et al. "Triple-GAN: Progressive face aging with triple translation loss." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 2020.
12. Zhang, Zhifei, Yang Song, and Hairong Qi. "Age progression/regression by conditional adversarial autoencoder." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
13. Zeng, Jiangfeng, Xiao Ma, and Ke Zhou. "CAAE++: Improved CAAE for age progression/regression." *IEEE Access* 6 (2018): 66715-66722.
14. Chandaliya, Praveen Kumar, and Neeta Nain. "Conditional perceptual adversarial variational autoencoder for age progression and regression on child face." *2019 International Conference on Biometrics (ICB)*. IEEE, 2019.
15. Korgialas, Christos, Evangelia Pantraki, and Constantine Kotropoulos. "Interpretable Face Aging: Enhancing Conditional Adversarial Autoencoders with Lime Explanations." *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024
16. Nagisetty, Vineel, et al. "xai-gan: Enhancing generative adversarial networks via explainable ai systems." *arXiv preprint arXiv:2002.10438* (2020).
17. Rajpal, Ankit, et al. "Xai-fr: explainable ai-based face recognition using deep neural networks." *Wireless Personal Communications* 129.1 (2023): 663-680.
18. Ribeiro, M.T., Singh, S. and Guestrin, C., 2016, August. " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).

19. Lundberg, S., 2017. A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874.*