Face Aging Synthesis with Identity Drift Heatmap using Generative Adversarial Networks

by

Ayan Pal
24241233
Md.Samiel Islam Sami
21301002

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
June 2025

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Ayan Pal

24241233

Md.Samiel Islam Sami

21301002

Approval

The thesis titled "Face Aging Synthesis with Identity Drift Heatmap using Generative Adversarial Networks" submitted by

1. Ayan Pal (24241233)
2. Md.Samiel Islam Sami (21301002)

Of Spring, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on June 20, 2025.

Examining Committee:

Supervisor:
(Member)

Mr. Dibyo Fabian Dofadar

Lecturer
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

Khondoker Nazia Iqbal

Lecturer
Department of Computer Science and Engineering
Brac University

Thesis Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi

Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

This work aims to generate higher-quality, more accurate age-progressed images using Generative Adversarial Networks (GANs), while introducing Identity Drift Heatmap. In our research, we worked with various GAN models and eventually focused on StarGAN - Hybrid Residual Attention Block to generate the intended face image and identity drift heatmap. While previous research has successfully used GANs for image synthesis, limitations still exist, as most papers mostly focus on generating age-progressed images. In our approach, we can visualize how identity-preservation deviates or changes while aging. Face age synthesis is extensively used in forensics and law enforcement fields, where precise, accurate, and detailed face progression is crucial. To achieve this, we proposed a StarGAN model with Hybrid Residual Attention Block (HRAB) that ensures smoother and more accurate age progression. In addition, our model is also capable of handling diverse facial structures and various aging factors such as hair color, skin texture, wrinkles, while maintaining realistic outputs. With the implementation of Identity Drift Heatmap, our research can be instrumental in various domains, from law enforcement to digital entertainment.

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Introduction

Face age synthesis refers to the procedure of digitally altering or generating a facial image to simulate a person's appearance at different ages while preserving facial identity. This technology can manipulate facial characteristics, including skin tone, wrinkles, and facial structure to predict what a person might look like at a younger or older age. Nowadays, face age synthesis is extensively used by law enforcement agencies to generate age-progressed images of suspects or missing persons to help in investigations [14]. Although face recognition systems are conventional in such cases, they often fail to identify individuals undergoing various changes as the time interval increases. As a result, Face age synthesis can produce better images for identification. Additionally, Face age synthesis is also used in movies and digital media, where actors are required to make appearances of different age groups. Face age synthesis enables convincing image prediction. Moreover, face synthesis can be applied to health and aging-related research, where information about facial changes due to aging is required.

While Generative Adversarial Networks have shown remarkable success in generating age-related facial changes, a major challenge remains in visualizing identity drift between the generated image and the original input image. To address this, we propose integrating an identity drift heatmap, which will show how facial features change from the original image across different age groups. Unlike conventional methods that focus mostly on generating age-progressed images, in our approach, we can specifically highlight regions where identity-preservation deviates or changes. This will visualize identity changes during aging progression and help us understand more about how the human face aging process changes across different age groups. By combining GAN-based aging synthesis with heatmap analysis, our research will generate high-quality age-progressed faces and also visualize and evaluate identity drift.

1.2 What is GAN?

Generative Adversarial Networks, also known as GAN, is an advanced deep learning technique that incorporates generative modeling. With generative modeling, GAN focuses on automatically detecting the patterns and characteristics of input data.

This enables GAN to create new models which is nearly identical to the original data. It is called an adversarial network because the generative process is done by competing two separate neural networks against each other in a min-max game setup [3]. One of these networks, the Generator, is trained to produce newer fake samples from a given input. On the other hand, Discriminator networks try to determine whether the generated input is fake or real. Both models are trained together, and through iteration, the Generator gets better at mimicking the original data and generates outputs that are indistinguishable from the real one. Subsequently, the iterative process concludes when the Generator produces such realistic samples that the Discriminator can no longer accurately distinguish between real and fake. The effectiveness of GAN's image-to-image translation tasks makes it such a valuable resource in face synthesis. The generators can produce face images with older or younger facial attributes while maintaining proper face progression or regression. Furthermore, GAN can be trained on a vast dataset containing diverse age groups to learn natural face aging with essential attributes like wrinkles, hair color, facial sagging, etc. In addition, generators can autonomously learn these attributes from datasets, making GANs more effective in face age synthesis.

1.3 Some GAN Models

Many researchers have shown interest in face synthesis and are actively working with Generative Adversarial Networks (GANs) to explore and expand their functionalities. The ongoing research seeks to discover new ways to use GANs for various applications and challenges. To obtain specific goals related to generative adversarial networks, researchers have been using varied tools, techniques and algorithms. Here, we will introduce some of the GAN tools used for face aging synthesis. First of all, Identity Preserved Generative Adversarial Networks(IPCGANs) are a framework for Conditional GANs (CGANs) that can generate functions to give accurate and realistic results [28]. Additionally, Contextual Generative Adversarial Networks is a method to solve transition pattern-related problems [2]. Contextual Generative Adversarial Networks work better at generating gradual facial changes and show better results. Also, Wavelet-based Multi-pathway Discriminators and Attribute-aware Attentive Generators can solve unnatural changes in semantic details such as face-related attributes [27]. Furthermore, to generate stable and photorealistic outputs, the Deep Convolutional GAN (DCGAN) is a suitable option [8]. DCGAN consists of one Generator and one Discriminator. It uses convolutional layers in both the generator and discriminator to stabilize training and improve the quality of generated images.

Additionally, Attention GAN is another model within the GAN framework that can produce highly realistic results based on the given input [12], [21]. AttentionGAN generates images by incorporating attention mechanisms, allowing the model to focus on specific image characteristics. This GAN model consists of one Discriminator and one Generator, divided into two parts: Attention and Transformation branches. Moreover, SuperResolution GAN (SRGAN) is particularly effective for maintaining background information and improving the quality and resolution of images, ensuring more accurate outputs [18]. S2GAN is capable of learning and extracting aging patterns unique to individuals, offering a personalized approach to face aging

[14]. On the other hand, PFA-GAN stacks of sub-networks instead of using a single generator for large age gaps. This approach reduces ghosting, blur, and aging artifacts. It also introduces an auxiliary age estimator during training and the Pearson correlation coefficient [16].

In addition, CycleGAN uses two Generators and two Discriminators [9]. It does unpaired image-to-image translation by using a cycle consistency loss to ensure that an image converted from one domain can be reverted to its original form. In addition, StarGAN can handle multi-domain image translation within a unified model using one Generator and one discriminator [23]. It uses the target domain and images as initial input. The Discriminator verifies the generated images and the target domain. On the other hand, Wasserstein GAN consists of one Generator and One Critic, which works like a Discriminator. It uses the Wasserstein distance to improve image generation and training stability, producing more diverse and realistic outputs [15]. Furthermore, Diverse Image-to-Image Translation GAN (DRIT) consists of one Generator and one discriminator. The generator is divided into 3 parts, including the Content encoder, Style encoder, and decoder. Discriminator is divided into 2 parts: Global and Local [7].

1.4 Problem Statement

Face aging synthesis has been explored in many research studies; however, only a few models have reached the desired level of quality, leaving ample room for improvement, particularly when addressing diverse cases. A major challenge is generating realistic images of various age groups while accurately capturing facial textures. Preserving key facial features is also crucial for achieving good results. Though some models have generated images with better accuracy, they fail to preserve the facial structures of various age groups and alter the face images too much. Which makes the face synthesis procedure impractical. Our goal is to tackle these challenges and improve the accuracy ratio of the generated outputs using suitable GAN models. More specifically, we will focus on generating accurate and realistic faces by preserving facial structures. For example, the research [8] proposed a model for face synthesis where the accuracy ratio of their work in generated images was 83.6%, compared to 91.8% for real images, showing that there is room for improvement.

Also, many of the research papers used datasets that mostly contain specific classes and age groups. Many papers have worked on celebrity-based datasets (CACD, Celeb A), which always contain high-quality adult face images with heavy makeup and other alterations. Some datasets also didn't provide old or child images, and mostly contained adult face images. Such datasets are not ideal for GAN testing and training, so we tried to find an ideal dataset that contains a large number of facial images for different races and age groups. Additionally, we intend to solve inaccuracies in generated photos by working with multiple GAN models and trying to understand their advantages and weaknesses. We mentioned many GAN models in our research, but we are primarily going to work with Deep Convolutional GAN (DCGAN), CycleGAN, StarGAN, AttentionGAN, Wasserstein GAN, and Diverse Image-to-Image Translation GAN (DRIT). In our research, we came across many GAN-based works that mostly focus on one or two models. Working with five

different models will help us understand different GAN models architecture and how they work. Each GAN has unique strengths in handling image translation, identity preservation, and aging realism. So different GAN model will generate slightly varied aging face images. For example, StarGAN allows smooth transitions across ages, while DCGAN may struggle with gradual changes. By using different models, we can identify which GAN best balances realism, identity preservation, and heatmap interpretability.

The second major part of our research is integrating an Identity Drift Heatmap, which will show how facial features change from the original image across different age groups. In our research, we came across many GAN-based research studies, which mostly showed remarkable results in generating age-related facial changes. But there are lack of research papers about visualizing the changes and deviations while the aging process occurs. To address this, we are planning to integrate an Identity Drift Heatmap, which will show how facial features change from the original image across different age groups. With the identity drift heatmap, we can specifically highlight regions where identity preservation deviates or changes. This will visualize identity changes during aging progression and help us understand more about how the human face aging process changes across different age groups. In forensic or missing person cases, the heatmap will help experts understand specific facial changes in synthesized aged images. For entertainment, it can warn us about regions where identity may appear unnatural.

Machine learning and AI are the talking points right now and for such reason, GAN can also provide many possibilities. The GAN models have potential usability for various real-world applications and various sectors. Our proposed work will provide advantages in the advancement of forensic works, in the commercial industries and ethical AI advancement.

1.5 Research Objectives

In this work, we are going to address key challenges in face aging synthesis, offering a more accurate and realistic age progression solution, while providing Identity Drift Heatmap. Generating such realistic and accurate photos will be instrumental in forensics and security, where realistic facial attributes of a missing child or person are needed. Our key objectives of this research are mentioned below:

1. To implement different GAN models and compare them.
2. To develop more accurate face aging for 5-year age interval.
3. To focus on identity preservation.
4. To integrate an identity drift heatmap and highlight regions where identity-preservation deviates or changes.
5. To test the model with different datasets.
6. To optimize the computational efficiency.

Chapter 2

Literature Review

In the research on face aging, there is a huge concern about the quality of the generated faces. The paper [18] proposes a realistic and detailed face aging with the use of AttentionGAN and SuperResolutionGAN (SRGAN). This work focuses on the security aspect and the importance of a touchless unique identification system. Also, this work aims to generate aged faces without unnatural modifications of facial attributes. Meanwhile, the implementation of SRGAN reduces visual artifacts and provides higher-quality images. Two different subnets are used in AttentionGAN, the first subnet generates numerous attention masks. On the other hand, the second one generates numerous content masks. The corresponding content mask and attention mask are multiplied accompanied by an input face to generate the intended result. Additionally, generated face images are separated from the resultant output with the use of regex filtering. Finally, the SRGAN further improves the quality of the face images with super-resolution. This method used the publicly available UTKFace, CACD databases, and IMDB-WIKI, CelebA, and FGNET are used for cross-age evaluation. The result of generated faces showed that the CACD dataset performed worse than UTKFace because of the different nature of the datasets. UTKFace images are mostly contained with low professional settings, emphasizing more natural settings, which worked better in age progression. Meanwhile, the CACD face images contained extensive lighting and makeup, which the proposed method performs poorly. Cross-age datasets evaluation demonstrates that FGNET has better outputs than IMDB-WIKI and CelebA. Also, the results showed that female faces tend to age faster than the male counterparts. Additionally, the results also demonstrate that the input images with high contrast can produce better information compared to the ones containing low contrast in the face masks. Furthermore, the Face++ tool is used to conduct the age estimation of the generated facial images. The images are divided into three age groups from the CACD and UTKFace datasets and estimated face results are approximately identical to the real images. Additionally, the confusion matrix is also used for evaluation, which shows satisfactory results in the age groups. To conduct the effectiveness and difference in generated face images created for using Super-resolution, evaluation techniques like the SSIM (structural similarity index) and PSNR (peak signal to noise ratio) have been used. Both of the mentioned measuring terms estimate the effectiveness of super-resolution. Finally, a user study was also conducted among many participants to check the pair-wise comparison between other methods and the proposed one. About 60% voted for the proposed work, 30% for the other work better and

the rest voted they are similar.

In this paper [5], the proposed method creates a personality-aware dictionary based on the neighboring age groups. As it is challenging or sometimes not possible to gather data for all age groups, this method mitigates those limitations by having a personalized way to render face aging using specific age group dictionaries. As each individual has personalized attributes that are invariant while aging, it is challenging to collect face images for any specific subject. Thus learning from the neighboring age groups is most relevant. In this procedure, short-term aging pairs of the same persons are collected from MORPH and the CACD (Cross-Age Celebrity Dataset). The aging dictionary is created corresponding to each group's aging attributes. After that, all the dictionaries are trained with the personality-aware attributes learned from the collected database. The proposed model is compared with a state-of-the-art method, Illumination-Aware Age Progression (IAAP) and FT Demo (Transfer Demo) on the database of FGNET. The result shows FT Demo is not consistent and worst in face shape change. Meanwhile, IAAP is slower in progressing age than ground truth. IAAP failed to show some facial textures as it mostly covers smooth faces. Additionally, cross-age face recognition is also conducted between the proposed technique and IAAP. The evaluation results confirm that the mentioned method can generate superior results than IAAP. For an input face, the mentioned technique provides outputs similar to the result of the ground truth. Furthermore, a user study was conducted between volunteers of different ages and walks of life. Among 50 adult participants, 45.35% voted for works that mentioned work better, 18.20% voted they are similar and 36.45% voted that the previous work is better.

In this paper [3], a novel face age progression approach is proposed using the advantage of GAN along with Convolutional Neural Networks (CNN) based generator to produce better face transformation. This work emphasizes its capability to handle both aging accuracy and identity permanence. The CNN-based generators are instrumental in learning age transformation and separately identifying various facial attributes upon the changes over time. Also, the proposed technique doesn't require matching face phases in different age domains to simulate aging faces. As the method is trained with data density of different age pairs it can estimate realistic outputs. Additionally, the mentioned work also tries to simulate the whole face image rather than just the cropped facial areas like many previous works. So, the techniques can generate additional aging features like forehead and hair characteristics to better simulate realistic aging faces. To achieve such a feat, the deep network's intrinsic hierarchy as well as a discriminator of pyramid architecture is utilized which can approximate better aging clues. This approach is also flexible in handling transformation both globally and logically as it doesn't require a single representation. The method also uses a kernel with a stride of 2, which is 3x3 convolution rather than traditional ways in the generator. In this work, the Cross Age Celebrity Dataset (CACD) and the MORPH mugshot dataset are used. Three age groups are created with the pair (30-40), (40-50) and (50-60) years of age. The visual fidelity shows that consistency and smoothness are preserved along with many facial changes like thinner lips, under-eye bags, and deep wrinkles. The result face also managed to make the hair thinner and wispy. Additionally, the Face++ tool is implemented to verify the identification of the original face image. The result of the Morph database

shows that 100%, 98.91%, and 93.09% verification rate for 3 age groups. On the other hand, the values are 99.99%, 99.91%, and 98.28% for the CACD database. Meanwhile, the pyramid neural network structure in the discriminator provided the age progression to be more natural. The experiment clearly shows that the proposed method's use of pyramid architecture generates better results than the traditional use of a one-pathway discriminator. Finally, an evaluation is conducted among many human participants in a pairwise comparison between the proposed work and previous works. The result presented that 69.78% preferred the proposed work, 20.80% favored the previous one, and the rest of them regarded both works as similar.

In the study of face synthesis, there is a lack of works exploring face aging for children under the age of 12. That's where this [22] research paper comes in whose main goal is to generate children's faces. This work emphasizes how important generating children face in finding lost children at their current age. Generally the traditional way to detect blood relation between two people is DNA testing which is costly and time-consuming. For such reason, the proposed work can predict images of the missing children at their current appearance more easily and less expensively. The proposed work does this by combining the effective StyleGAN2 and Facenet methods. The problem with the existing face aging models is that they mostly consider facial attributes that are present in older people. On top of that genetics and head size also affect during aging which is ignored in current models. Current face aging models aren't flexible with the rapid growth of children and adolescents. The proposed system takes genetics into consideration to generate aging faces. The mentioned research uses a pre-trained StyleGAN to mix 2 different face images. Meanwhile, FaceNet identifies similar characteristics of the face images. The model will first take images of the missing children as input along with the children's corresponding parents or a blood-related family member's facial photo. After that, the model will estimate a prediction result. The proposed system also has the function for the family members or relatives of the missing person to submit their information and get a predicted facial image of the present age. To confirm the functionality of the mentioned system an experiment is held in comparison with other methods, CAAE, HRFAE and IPCGAN. In the experiment, 3 children's images about the age of four are set as input along with their parent's image. The output is set to 20 years and the SKEye face compare function is used to identify the similarity between the generated output and the expected output. The proposed model managed to attain similarity of 77%, 76% and 77% for all 3 images, which outperformed the other models tested in the experiment.

In this paper [8], researchers tried to generate stable and photo-realistic images while keeping realistic facial features. They proposed a GAN-based model, Deep Convolutional Generative Adversarial Networks (DCGAN). Implementing a large Neural Network model requires more computational power and a lot of time. To address such problems, they proposed a model that will cover such demand, also cope with the computational power and time. They used FG-NET, MORPH-II, CACD, and IMDB-WIKI datasets. Every dataset has some limitations regarding age and variety in face images. Mainly FG-NET dataset was used to evaluate the work. The FG-NET dataset is a popular dataset that is being widely used for the face synthesis task, as this dataset has a lot to offer for this task. This dataset has

different face images to work with. In the paper, they have discussed face detection, face progression, and regression. However, in their paper, they have proposed a model that generates rich texture and visual reality, which outperforms the related works. In their proposed method, the whole model consists of four parts: Encoder, Connector, Generator and Discriminator. The encoder measures the facial images to a personal latent vector that actually passes through the connector along with the age label and gender label. Then the result goes through the generator, which is involved in generating the images, and the discriminator keeps the generated images separate. When the FG-NET dataset was used as input images and compared it with the model named FT Demo. The images by FT Demo were deformed and had color problems. The proposed model gave far better results.
For age estimation, the accuracy ratio for generated images was 83.6% whereas in real images 91.6%. The MAE error in generated images was 0.2812 and in real images 0.1263. Also, their results were 8.2% less than the real images.

In the paper [9], researchers have proposed a fusion-based GAN to address and solve the issue of identity preservation as well as age estimation accuracy of face aging synthesis. In this work, they proposed a fusion GAN model with two generators. For the face age progression task, they used Cycle-GAN. Meanwhile, ERSGAN was used in face visibility. In the proposed model, there is a Generator which consists of three parts: Encoder, Transformer, and Decoder. The Encoder performs the task of extracting the features of an image. Meanwhile, the Transformers do the job of adding the residual from the input to the output images. On the other hand, the Decoder does the reverse job that the encoder does. The Decoder gets back to the lower-level facial structure of the images through the deconvolution layer. However, one drawback of their proposed method was that it took a lot of time to complete the training for several days, even though they used a computer with higher computational power. They used five popular datasets, which are IMDB-WIKI, UTKFace, CelebA, FG-NET, and CACD. All of those datasets are very common and widely used for the face age synthesis task and other tasks related to this work. This fusion GAN model works better in producing super-resolution face-aging images. The results from their proposed work are satisfactory and can be used in other image datasets. The generated images were compared with the original images and showed an error rate of 0.001%. On the datasets, the confidence score was improved to 95.39, and overcame the previous score of 95.13. In future works, they are hoping to reduce the training time and test the model on different datasets.

To solve unnatural changes in the face aging synthesis, a GAN-based model called Attribute-Aware Attentive Generator was proposed by the paper [27]. They also addressed the issue of distortion in low-level images due to some age-related modifications. For that reason, they used both Wavelet-based multi-pathway discriminators and attribute-aware attentive generators. The work addresses the problem of not having prior knowledge about the detailed information of images and the insufficient utilization of that information. In this paper, they have introduced a mechanism called the Attention mechanism that improved the visual quality of the images which many works related to this lacked. Age progression was also maintained in this paper. As there is an issue with how many generalizations the proposed model was

tested with the most common and popular datasets, which are FG-NET, CELEB A. They also used MORPH, CACD, FG-NET, CELEB A datasets for certain tasks. In their paper, they mentioned a problem with most of the previous works. The problem is, previous research takes only young faces as the input images to train and learn the model about the different age groups and small parts of facial structures. Later, it gave a ghosting effect in the images. So, their Attention-Aware Attentive Generation-based model worked to overcome those limitations. Also, wavelet-based multi-pathway discriminator helps to solve wrinkles, eye bags, and laugh line-related issues. Results on FG-NET and CELEB A datasets demonstrated the effectiveness of the proposed models. They dealt nicely with the unseen image problem. Also, CACD and MORPH datasets gave satisfactory results. In the datasets used, there was insufficient data on child faces. So, working with large-scale data consisting of different ethnic groups and child faces can be done in the future.

The paper [17], addressed the issue of using images of one age class and trying them into another age class by preserving their own age class identity or characteristics. This work looks different as they tried to bring and train a model with some similar key features of faces and spread it into other images of different age groups. Their method addressed the issue of unsupervised image-to-image translation problems. For that reason, they proposed a model that can solve the issues they have raised. Those were Global Face Aging GAN and Pyramid Face Aging GAN. Here, the global face aging GAN learns about the global transformation, whereas the second one, Pyramid Face Aging GAN works as a pyramid encoding and decoding scheme. To test and implement their proposed model, they experimented using different datasets. Among those, CACD, UTK-face, and FG-NET are the datasets that they used. They tested those models on different datasets based on different key factors and evaluation metrics such as age classification accuracy, MAE(years), FID score, and VGG face score, they got comparatively better results than previous works on PFA-GAN and GFA-GAN. To specify, in age classification one-off accuracy, GFA-GAN gave 78.50 and PFA-GAN gave 70.56. In MAE(years), GFA-GAN gave 10.4667 and PFA-GAN gave 12.5111. For close translations, PFA-GAN succeeded in providing aging effects more accurately than GFA-GAN. By comparing with the ground truth, the proposed model achieved realistic face age progression. Future works can investigate the weaknesses and strengths by using more diverse datasets where facial images of Asian people can be used with the help of the AFAD dataset.

Paper [19] proposed a GAN model called Attention wavelet-transformation-based Generative Adversarial Networks. In the paper [27], we also have seen the same kind of model. It is beneficial that we are gaining knowledge about the same models working for different problems. In the paper [19], they focused on face information and capturing more textures on the face by face aging synthesis, which is different from the problem discussed in the paper [27]. The proposed used CACD, FG-NET, and ICD datasets. In their model, they used an attention generator, wavelet transform, and multi-scale patch generator. They mentioned that most of the GAN-based models used earlier had been implemented based on the convolution layers. In most of the convolution layers, they get the information from just adjacent pixels, and there are some limitations at the receiver ends. In that case, the Attention mechanism works perfectly to solve those kinds of issues with the

combination of Deep Convolutional Neural Networks. The network architecture of their proposed model was divided into several parts. They include: the Generator, Wavelet transform, Multi-scale patch discriminator, AlexNet perceptual network, and Modified-convolution block attention module. They implemented the model and experimented on different datasets. Also, some state-of-the-art face aging models such as CAAE, CPAVAE, SAMSP, and IPCGAN were selected for comparison experiments. The AW-GAN model that they proposed achieved comparative performance to CPAVAE and CAAE as well as SAMSP AND IPCGAN. They find that AW-GAN is competing with IPCGAN when it has been evaluated using some face matches on ICD and CACD. Through COTS matches, IPCGAN and AW-GAN both achieve nearly 100% accuracy on ICD and CACD datasets. They hope to experiment with 3D facial aging modeling, especially for children.

Contextual Generative Adversarial Networks are a useful model to solve the different problems of face aging synthesis, which researchers have used and mentioned in the paper [2]. Contextual Generative Adversarial Networks have been used in several works, but they didn't raise and solved the issue of not capturing the transition patterns of the adjacent age groups. Those patterns can be gradual shapes, texture changes of the faces, etc. So, the proposed method decided to work with this problem to solve it using the Contextual GAN. In the model, there are two Discriminative Networks and one Conditional Transformation Network. First of all, the image the Preprocessing task is done. Then, the face image and input age are given as input. The workflow of this model is shown in the figure 2.1, where we can clearly visualize the process. The Conditional Transformation Network generates synthesized face images and goes through the Age Discriminator Network. To get better results, they specifically model cross-age transition patterns. However, to implement and evaluate those tasks, CACD, FG-NET, LFW, MORPH, and SUP were the datasets they used. In total, 4047 face images were there to work with. But they deleted low-resolution images for better results, they mentioned. If we come to the performance part, when compared with the state-of-the-art models, they achieved better results. For cross-age face verification, the EER on C-GANs Synthetic Pairs and Original Pairs is 8.72% and 17.41% respectively. In the future, they plan to use other GANs to improve their performance, such as LS GAN, Wasserstein GAN, and EB GANs.

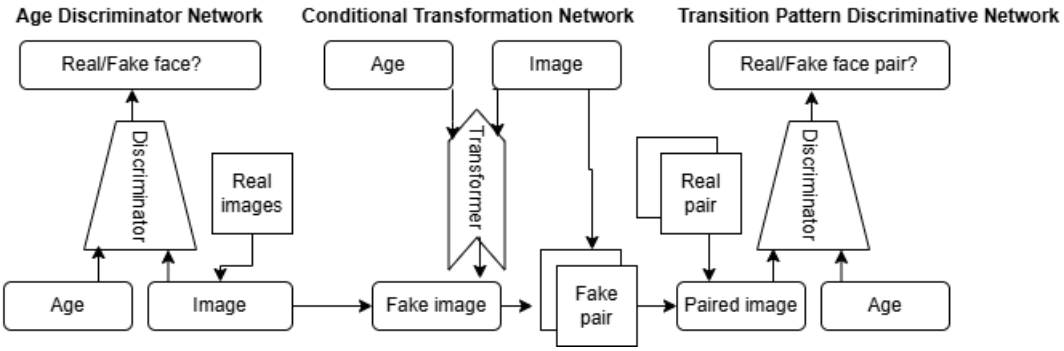


Figure 2.1: Structure of Contextual-GANs in the work of Liu et al. (2017)

The main purposes of the paper [11] is to work with Face aging synthesis on a gender basis and improving the performance and efficiency of existing models using

Cycle-GAN. In this work, gender-based face aging was applied. This work used the IMDB-WIKI dataset. When looking into the implementation part of their work, it was based on a traditional Cycle-GAN model framework which used pairs of some predefined age groups to train their Gender-based GAN. It took almost two weeks to train the model, which suggests more computational power is needed. After comparing with common cycle-GAN methods (CCG), Gender-based GAN (GBG) gave better results. The number of each image and the score of each image are given to compare the female face images aging processss in Cycle-GAN and Gender-based GAN. However, in the proposed paper, some limitations are found, such as not working with a dataset that has a good distribution of faces. Working with a better dataset can be done in the future to make the model work more efficiently.

This paper [26] shows a unique presentation of CGAN (Conditional Generative Adversarial Networks) by generating six highly realistic facial expressions from a single input face image. Six separate Conditional GANs are used, each conditioned to generate one of the facial expressions. In this work, the AR dataset is used for experimentation which contains frontal and face images of different expressions. There are face images of 124 subjects, with 3300 images of various illumination, conditions, and characteristics. Six facial characteristics per person is extracted from this dataset. The expressions are: anger, smile, stare, illumination, wearing a scar, and wearing glasses. 20 percent of images from the datasets are selected for testing and the remaining are used for training the CGAN. The generated images were compared with the original face image and they were indistinguishable. To evaluate the accuracy of the approach several state-of-the-art CNN including VGG-Face, ResNet-50, VGG-Face, FaceNet, DeepFace and FaceNet models are employed. Initially, all these models achieved only about an average of 76 percent accuracy on neutral images from the dataset. However, when the CNN models were fine-tuned with the generated images from the CGAN they achieved around 97 percent accuracy.

This work [20] aims to produce good-quality generated face images with the least computational power and without the need for any external image classifier. This work uses the CelebA dataset for experimentation. The main goal here is to generate fake human faces from random noise and maximize stabilization by removing other noises. Also ensuring no external classifier is used. The generator is trainable and creates fake human faces from random data. Meanwhile, the discriminator tries to detect if the created faces in the generator are real or fake by comparing them with CelebA. Keras's classic sequential modeling is used in both the generator and the discriminator. Five layers are used in the generator. The leaky ReLU activation module is implemented in all the layers except the output layer. To train the generator properly with normalizing and data resizing, the Reshape module and Batch Normalization technique from Keras are used. The main difference in this work is the generators don't update or train after each epoch. The model runs about 100000 epochs with 10 thousand to 203 thousand images from the CelebA dataset. The face images got better as time went on, proving that the model was improving over time. The values from the table of loss function of the discriminator, are mostly between 0 to 1, which is a pretty good result for the discriminator. Also, when calculating high spikes, within 100000 epochs, there are only 771 high spikes

meaning the results are satisfactory for most of the time. The mentioned method works with 3 different image sizes, including 48*48, 100*100, and 150*150, making it dynamic. One downside of the model is that it only works on black and white images.

Additionally, we also took a look at some additional research papers to get a better understanding of GAN models capabilities. In the paper [24], they have done the face aging task by using Explainable Conditional Adversarial Autoencoders. It incorporates explainable components by limiting the model on age-specific attributes. They also provide heatmaps for age classification prediction task, which is a unique research topic. The paper [13] translates thermal face images into visible ones while preserving identity for face recognition. The idea of translating thermal face images into visible ones is novel and practical. Also, the [10] investigates age-related facial analysis using deep learning and proposes methods for age estimation, age-oriented face synthesis, and age-invariant face recognition. They used a Generative Adversarial Network with a Conditional Discriminator Pool and Adversarial Triplet loss to ensure precise age-oriented face synthesis. We also used the researched the popular paper, [4] which helped us learn more about Cycle Consistency. Additionally the paper [1] helped us learn more about Super-Resolution. We have mentioned various models so far and all of them had unique methodologies and features. In paper [17], they bring a model combining Global face aging GAN and Pyramid face aging GAN to take the common features of face images and add those to other images. In the paper [2], the transition of patterns in adjacent age groups was focused on using Contextual GAN. In the paper [19], they focused on capturing more textures on the face by capturing more differences. AttentionGAN and Super Resolution GAN were mentioned in the paper [18] to generate high-quality natural face images with the help of super-resolution. While SRGAN was used to create high-quality images, DCGAN was used to create realistic images which was mentioned in the Paper [8]. The DCGAN architecture was simple and used traditional Discriminator and Generator setup.

Chapter 3

Methodology

3.1 Dataset

We have explored different datasets for the face aging synthesis task. Datasets containing face images of various skin colors, ethnicity and large age groups are challenging to find. During the literature review, we found several datasets in which the researchers have trained and tested their models. Some popular datasets are MORPH, CACD, IMDB-wiki, UTK-Face, Celeb A, etc. Each of these datasets has some advantages and disadvantages. We tried to explore and find limitations in each of the datasets. Manually checking each of the datasets can give a good idea about their characteristics as they are image datasets. We followed some criteria while checking all of the datasets manually. The criteria that we followed are the age ranges, ethnicity, gender, labels of age, or other characteristics. During the age ranges, we checked the starting age of the range and the last age of the range.

In our research, we have found that the UTK-Face dataset [29] contains an age range from 1 to 116, which is crucial when working with a wide range of age groups for face aging synthesis tasks. Also, the UTK-Face dataset contains gender and ethnicity as the file names of each image file of the dataset. Meanwhile, the Morph and IMDB-WIKI datasets also contain good age ranges of face images, but are not as diverse as the UTK-Face dataset. Also, the other datasets do not contain as much diversity as the UTK-Face dataset does. Also in our research, we found out that Celeb A and CACD are quite popular in face aging. While the UTK-Face dataset covers face images from different ethnicities, the Celeb A and CACD datasets are more focused on specific classes and age groups and mostly consist of adult faces. Also, they contain celebrity face images more, which hinders the model's foundation during training. Celebrity faces are mostly high-quality and contain makeup and other alterations, which is not ideal for model testing and training. Diverse face images during training are more essential, which these datasets fail to provide. On the other hand, the IMDB-WIKI dataset fails to contain proper annotations, which leads to larger preprocessing time and makes it less suitable for GAN models to adapt to and learn. Also, the IMDB-Wiki dataset lacks good image quality. On the other hand, we also found satisfactory image files in the MORPH-II dataset [25], which has diverse face images. Though the UTKFace dataset has a good amount of images with varieties in age, gender and ethnicity, the MORPH-II dataset gives a good balance in the dataset after merging. Another reason we have considered

for choosing the MORPH-II dataset is that this dataset contains a good number of black people's images, which UTKFace lacked. So, for face aging-related tasks, the UTK-Face and MORPH-II datasets are more suitable.

Considering the mentioned limitations, we decided to work with the UTK-Face dataset [29] as we found it one of the more preferable datasets to implement. The UTK-Face dataset, contains over 23 thousand face images containing information on age, gender, and ethnicity for each image. Also, this dataset covers different poses of faces, expressions, and variations, which we can observe in the figure 3.1.

Figure 3.1: Face images of various age, gender and ethnicity (UTK-Face)

Figure 3.2: Face images of the same age and gender with different ethnicity (UTK-Face)

The Filenames are also encoded as `age_gender_ethnicity_date&time.jpg`, which gives us information about age (age of the person in the image), gender(0 for male and 1 for female), and ethnicity (ranges from 0 to 4). For ethnicity, 0 means White, 1 means Black, 2 means Asian, 3 means Indian, and 4 means other (Latino, Hispanic etc). This naming convention makes the dataset suitable for working with. The dataset also has many varied face images for any particular age, which is evident in the figure 3.2, where we picked 4 female face images of age 35 with four different

ethnicities. All the images presented in the dataset are set to 200*200 pixels, which is also useful for our work. However, there are still some downsides, such as some blurred and low-quality, or unclear images in the dataset.

In addition to the UTK-Face dataset, we are also working on the MORPH-II dataset, which provides diverse face images. Although the UTKFace dataset has a good amount of images with varieties in age, gender and ethnicity, the MORPH-II dataset gives a good balance in the dataset after merging. Another reason we have considered for choosing the MORPH-II dataset is that this dataset contains a good number of black people images which UTKFace lacked, which is evident in the figure 3.3. We have collected the MORPH-II dataset from Kaggle. The dataset was divided into three parts already on the Kaggle site, such as train, test and validation data. In the training data, the total number of images are 40012, while the total number of male and female images is 34433 and 5579, respectively. The total number of images in the testing part is 5002. Where male and female images are divided into 4266 and 736, respectively. In the validation data, there are a total 5001 number of images, where the total number of female and male images is 760 and 4241, respectively.

Figure 3.3: Face images of various age, gender and ethnicity (MORPH-II)

The image filenames in the MORPH-II dataset are formatted as `00013_00M23` and were in JPG format. Here, 00M23 of the filename, M means male, and 23 means age, 00 represents the number of images. All the images presented in the MORPH-II dataset are set to 180*180 pixels. One of the main reasons we choose to work with MORPH-II is that this dataset contains a good number of diverse race and ethnicity images, which many other datasets don't provide. From the figure 3.3 we can see that the MORPH-II dataset provides good quality images of diverse races.

3.2 Data Preprocessing

Preprocessing is an essential step as it prepares the data for training and testing with different Generative Adversarial Network-based models. At first, we followed the data cleaning process by cleaning unclear and blurred images. To do this task, we downloaded the dataset and checked every image one by one. We did the cleaning process manually. When unclear and blurred images were found, we removed them.

Poorly visible or blurred images can negatively impact model training, preventing the model from learning effectively for future testing. Therefore, we chose to remove such images from our dataset. Dark images are also being removed for the same reason. It took a few hours to complete the image selection process. Working with clean and clear images can improve the training process.

Initially, the number of images in the UTK-Face dataset was 23708. After the cleaning process, the number of images was 22991. So, 717 images were removed in this process. Images of males previously were 12391, which was reduced to 11914. Images of females previously were 11314, which was reduced to 11074. The GAN-based system requires the same sizes of input data, which we ensured by resizing them to the size of 128*128 using torchvision and PIL. First, the PIL library opens the images and manipulates them. From torchvision, we imported the transform module. We used the Resize function given by the transform to keep them the same size. Also, by using this transform from torchvision, we followed the normalization of pixel values. Meanwhile, the ToTensor method converts the pixel values from integer [0, 255] to float [0,1]. Next, using torchvision, the data was standardized from [0, 1] to [-1, 1], which is standard for the GAN-based model. It also ensures stability in the training process of models. Then we ensured the labels of the new file preserve the metadata of the image file. For example, the filename is preserved as: `20_0_1_20170101224524253`. Here, 20 means age is 20, 0 means male, 1 means ethnicity, and 20170110224524253 means date and time. In this example, the date and time are 10th January, 2017, 22 hours (which means 10 PM), 45 minutes, 24 seconds, 253 milliseconds. This naming convention with appropriate labels simplifies the process and does not require any extra file-handling process. We also split the UTK-Face dataset into training and testing parts. The ratio was 80% to 20%. 80% of the data was used for training the model and 20% of the data was used for testing the model. So, from 22991 images, 18393 images were for training and 4598 images were for testing.

Next, we worked on the pre-processing of the MORPH-II dataset. Just like the UTK-Face dataset, we have done manual checking of all the images. We intended to clean the data by finding and removing unclear, blurred, or images that were not clearly visible. By manually removing these images, the total training images was 38989. So, 1023 face images were removed in our manual processing. Then we have followed the next data preprocessing method on the MORPH-II dataset. Similar to the UTKFace dataset, firstly, we have resized the images into 128*128 pixels using the Resize method from the torchvision.transform module. For GAN-based training, resizing the image ensures the uniformity in input dimensions. The images, which were in JPG format, were loaded using the Python Imaging Library (PIL). Using the ToTensor() function, images were converted into tensors, and by using the Normalize transformation, similar to the UTKFace, images were normalized to the range [-1, 1]. After processing the data, the images were saved as PyTorch tensor files, and the filename was maintained as `age_gender_serialNumber.pt`. Where the age and gender were extracted from the previous filename and others were removed. The serial numbers were given from 50000, so the next serial number will be 50001 and onward.

Finally, we merged both UTK-Face and MORPH-II datasets to create one unified dataset ideal for our research. In the merging process, we maintained the same file format for both datasets. We used `age_gender_SerialNumber.pt` as the naming format for all image files. This format is the same as the MORPH-II dataset, so only the UTK-Face dataset needed changes. This is done by going through all the files using python. The serial numbers were maintained properly to prevent clashing with each other. We kept all the files from both datasets in the same folder, which made it suitable for our GAN-based work. Also, we created separate folders for each age bin (age 1-5, age 6-10,. . . , age 116-120), which is needed in some testing cases, such as FID testing. For models like CycleGAN and AttentionGAN, separate young or old age groups are required. For that reason, we organized the data in ascending age order. We considered people with lower than 40 years as young and kept their face images in a separate folder and those over 40 are considered old and kept separately. DCGAN doesn't need this classification. While StarGAN does the multi-domain task, which utilizes the filenames or the labels that we have used. Separate folders were needed for working with WGan and DritGan. After the whole preprocessing task, we acquired high-quality images with a fixed image size of 128*128. Also, the pixel values are standardized to [-1, 1] along with easy label preservation. All these processes ensure that the dataset is ready to work with any GAN models.

3.3 Implementations

In this phase of our work, our data is ready and optimized after the preprocessing part. The next step is to continue training and testing different models for face generation. We decided to implement different models to analyze their architecture, measure their performance, and compare them to learn more about the models that can guide us in our work.

Deep Convolutional Generative Adversarial Networks (DCGAN)

First, we implemented Deep Convolutional Generative Adversarial Networks (DCGAN). DCGAN is one of the popular GAN models, which has a comparatively simple and straightforward architecture. Traditionally, DCGAN consists of one generator and one discriminator. The Generator gets the input images from the dataset that we have provided and generates fake images similar to the real ones. The Generator analyzes the real images to generate fake images. The discriminator tries to distinguish between the real images and fake images generated by the Generator. DCGAN architecture is based on Convolutional Neural Networks (CNNs). So, DCGAN generates images with special features using the convolutional layers. In our experiments with Deep Convolutional Generative Adversarial Networks (DCGAN), we started with a traditional GAN setup, consisting of a single Generator and a single Discriminator. The architecture of DCGAN is shown in the figure 3.4. First, the Generator starts from a random noise vector of size 100. In the initial layer, an input size of 512*4*4 has been passed through, where 512 represents the total channels and 4*4 represents the image size. The generated output size is 256*8*8, which will pass through the next layer as input. Next, we get the output size of 128*16*16, which further converts to 64*32*32 when it passes through the next layer. Again, we

get the output size of 32*64*64 in the next convolution layer, and finally, we get the desired output of 3*128*128 when it passes the next layer. So, each following layer performs upsampling through transposed convolutions to ultimately produce an image of size $3 \times 128 \times 128$. ReLU activation is used in each layer to learn non-linear mappings. For stabilized training Batch Normalization (BN) has been ensured. The Discriminator tries to detect the generated fake images by distinguishing them. The Discriminator layer starts with a 2D vector at the initial stage, working with the downsampling of the images. Which is the opposite of the Generator. The process starts when fake images generated by the Generator size of 128*128 and with RGB 3*128*128, are passed to the discriminator. The Initial layer downsamples images from 3*128*128 to 64*64*64 according to the architecture of the discriminator. Then the next layers generate it to the sizes of 128*32*32, 256*16*16, and finally 512*8*8 respectively. Batch normalization has been maintained for the layers and LeakyReLU activation is used to deal with negative values. Finally, Output passes through the fully connected layers while Sigmoid activation is applied to get the probability of the images being real or fake. In our testing, the learning rate used is 0.0002, the batch size is 64, and the number of epochs is 100.

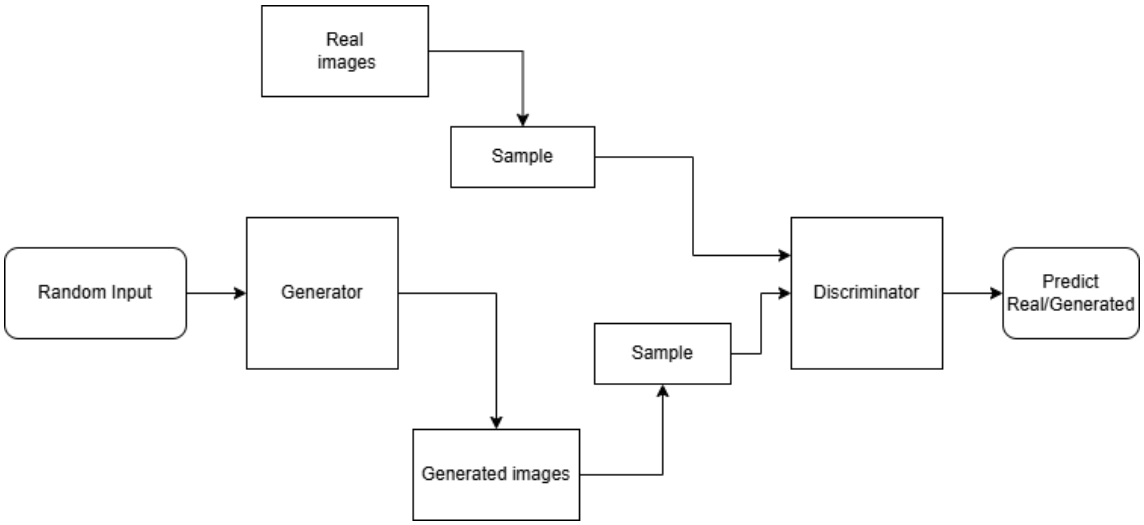


Figure 3.4: Architecture of DCGAN

While the traditional approach produced reasonable results, we tried to improve image quality, training stability, and convergence speed through some changes and modifications. The original Generator design begins with a 100-dimensional noise vector that is reshaped and passed into a higher-dimensional tensor, such as 512*4*4. Each following layer performs upsampling through transposed convolutions to ultimately produce an image of size 3*128*128. To improve the quality and detail of the generated images, we applied the following refinements. The Generator upscales images through successive transposed convolution layers. Each layer is followed by Batch Normalization and a ReLU activation to allow stable gradient flow and to learn complex, non-linear mappings. At the final layer, we apply a tanh activation function instead of ReLU to ensure the output values lie within the normalized image pixel range of [-1, 1], which is more suitable in our case. To keep the architecture fully convolutional, we eliminated fully connected layers at both the input and output ends of the Generator. The Discriminator performs the inverse operation of the

generator, it downsamples the input image using convolutional layers and attempts to classify it as real or fake. Instead of the standard ReLU activation used initially, we use LeakyReLU activation functions with a negative slope ($\alpha = 0.2$). This will allow gradients to flow better for negative input values and improve convergence. Also, this will reduce the risk of dead neurons in ReLUs. Additionally, Spectral Normalization is applied to all convolutional layers in the Discriminator to ensure stable training and better discriminator generalization. This solves issues like mode collapse and discriminator overpower. In the output layer, a final sigmoid activation function is still used, which will indicate the likelihood of the input being a real image. Initially, Binary Cross-Entropy Loss was used in the Generator and Discriminator training. However, to improve gradient flow and avoid vanishing gradients, we used Hinge Loss. This made the Discriminator stricter and the Generator more precise.

CycleGAN

We have also implemented the CycleGAN model on the chosen dataset to generate face images. CycleGAN is designed for working with unpaired images. We prepared the data into two groups: old and young. CycleGAN then works with this data to map the young and old face images. CycleGAN is traditionally designed with two Generators and two Discriminators. The two Generators are Gx and Gy, and the two Discriminators are named as Dx and Dy. Gx converts young face images to old face images and Gy converts old faces into young faces. Each generator has a corresponding discriminator that learns to distinguish between real and generated images from its target domain. Dx discriminator distinguishes between generated young faces and real young faces, whereas Dy distinguishes between generated old faces and real old faces. Thus, fake or real images are detected. Both of the Generators work in the same way. The preprocessed image size of 3*128*128 is passed through, and convolutional layers extract features from it. LeakyReLU is used as the activation function. Initially, the input shape of 3*128*128 generates an output of 64*64*64 with kernel 7*7, stride 1, and padding 3. Then in the layers, it generates the output size of 128*32*32 and 256*16*16 with kernel size 3*3, stride 2, and padding 1. The Residual Blocks (basically the middle layers) have two convolutional layers. Those layers preserved the spatial information (information related to the structures of data). After that, transposed convolutions were used to upsample the data, where ReLU activation is performed. The first transposed convolution gives output 128*32*32. Then the second transposed takes it as input and generates output of the size 64*64*64 and finally generates the size 3*128*128 fake RGB images. In the architecture of Cycle GAN, U-Net or ResNet-based architecture is being used; meanwhile, for the discriminator, the PatchGAN architecture is used. The generated fake image size of 3*128*128 passed through the layers of the discriminator. The LeakyReLU activation is used as it passes through Convolutional layers and images are downsampled again. Convolutional layer 4 generates the output shape of 512*8*8. For the output layer, the final convolutional layer produces a 1*8*8 size feature map, which generates a realistic output. PatchGAN architecture helps in such generations. In this case, adversarial loss ensures that the generated images are indistinguishable from the original images. On the other hand, Cycle consistency loss ensures that image translations are reversible. Here, the learning Rate used is

0.0002, the batch size is 32, and the number of epochs are 50.
To push beyond traditional architecture's performance, we applied some improvements and changes to our CycleGAN model. Initially, the Generator in the CycleGAN uses a ResNet-based model with an encoder, transformer and decoder. The encoder performs downsampling through convolutional layers. Then, the transformer consists of residual blocks that contain structural and spatial features. Finally, the decoder upsamples the feature maps back into the target domain image. To improve the process, instead of using 6 ResNet blocks for 128*128 images, we used 9 ResNet blocks. This will help learning more varied transformations, particularly when mapping complex facial features. Then we applied Instance Normalization over Batch Normalization used previously. Batch Normalization can create artifacts when the batch size is small. Instead, Instance Normalization works independently. Just like before, we used ReLU activation for upsampling the data. In the Discriminator part, previously, a PatchGAN structure was used, which can miss the global facial structure. Instead, we used multi-scale discriminators. This helps acquire both global structures and local textures. Additionally, Spectral Normalization is used in all layers in the Discriminators. This stabilizes GAN training by preventing the Discriminator from becoming too strong. CycleGAN's result depends on the balance of the loss components: Adversarial Loss and Cycle Consistency Loss. To improve translation quality, we used Least Squares GAN loss instead of the standard binary cross-entropy loss. This improves training stability and generates higher-quality images.

AttentionGAN

AttentionGAN is also an advanced Generative Adversarial network-based model that is designed for image-to-image translation while maintaining the attention mechanism. This GAN-based model is specifically focused on the mechanisms that change over time of a person's face, such as wrinkles, eyes, hairstyles, and other facial features, while preserving the unchanged parts of the face. The Generator in AttentionGAN consists of two branches, which are the Attention Branch and the Transformation Branch. Attention Branch works by producing an attention mask that determines which portion of the face images can be transformed or changed. The Transformation branch is responsible for creating images after transforming features from input to output. In the Generator, face images and the target domain were taken as input. Input passes through the convolutional layers first. After that, downsampling occurs. Following, the spatial identity is preserved by the residual blocks. Then, passing through some convolutional layers, it gives the attention output, which is a single-channel mask. The transformation branch generates transformed images by applying changing mechanisms such as facial structures, wrinkles, etc. Following all these steps, we get the final output of the generator. The discriminator detects whether the image is real or fake and the domain label. This Discriminator is a patch-based discriminator that generates adversarial scores and domain classification. For training, AttentionGAN uses several losses such as Attention loss for focused attention masks, Reconstruction loss for Cycle consistency, Adversarial loss to generate realistic images by Generators, and Domain classification loss to ensure the image belongs to the target domain. In our work, the learning Rate used is 0.0001, the batch size is 8, and and number of epochs is 50.

To improve the results of this baseline model, we made some improvements and changes. The Generator in AttentionGAN is dependent on how well the attention branch isolates domain-specific regions for transformation. Thus, we can get better results by improving both the Attention and Transformation branches. Firstly, we tried to generate attention masks at multiple feature levels (low, mid, and high), instead of the previously used single-scale attention mask. These masks are combined using learned weights. Skip Connections are introduced between early layers of the Attention Branch and the Transformation Branch. This ensures better information flow and output details. Traditionally, the Discriminator in AttentionGAN uses a PatchGAN classifier. For better results, we tried multi-scale discriminators, which operate on different image resolutions (e.g., 128×128, 64×64). This helps the model to better understand global and local realism. Also, Spectral Normalization is applied to the convolutional layers of the discriminator. This ensures stable gradients and prevents overpowering the discriminator. We also tried the gradient penalty loss in the adversarial loss, which maintains a smooth loss function in attention-based architectures. Attention GAN uses several losses, such as Reconstruction loss, Domain classification loss, Adversarial loss, and Attention loss. First, we included a perceptual loss (using a pre-trained VGG network) instead of only using L1 or L2 reconstruction loss for Cycle Consistency. This will ensure capturing texture and identity better. We tried changing the weight of the domain classification loss to get better output.

Wasserstein GAN

Wasserstein GAN, commonly called WGAN, is a simple GAN model that uses the Wasserstein distance to improve image generation and training stability, producing more diverse and realistic outputs. It is easy to train and has a reduced likelihood of mode collapse. It operates by transforming input face images to represent various age groups while maintaining the original identity features. WGAN might not be the best choice for face age synthesis tasks, but we wanted to try different models to measure and compare performance with other models. To implement this, a Generator and a Critic are required. Unlike other GAN models, here the discriminator is replaced by the Critic. The Generator starts working with the noise vector z and the target age. So, the input shape for WGAN is (z + target age). After reshaping the dense layer, it goes through some deconvolutional layers and generates output. ReLU activation was being used for the deconvolutional layers. The output layers activated Tanh. Meanwhile, batch normalization was also applied for stable training. The Critic does the work that the Discriminator usually does. The Critic distinguished between fake and real images. The first convolutional layer input shape is 3*128*128 and after passing through three layers, the output is the size of 256*16*16. The Dense layer maps the previous convolutional layer features and integrates features into a single vector. Critic loss helps the Critic to learn to distinguish between real and fake images. While Generator loss helps the Generator produce real-life images. Additionally, the Gradient penalty helps to stabilize the training for the WGAN model. To generate accurate values, the Critic is trained multiple times in each step. Here, the Learning Rate used is 0.0002, the batch size is 32, and the number of epochs is 50.
Like other models, we tried to improve WGAN's performance beyond its base ar-

chitecture. In our original work, the Generator takes a latent vector z concatenated with a target age condition as input. It generates 128×128 RGB face images conditioned on the desired age. In the Generator, for preserving identity across age transformations, we introduced an identity loss using a pre-trained VGGFace architecture. This ensures the identity preservation of input and output images. In WGAN, the Critic is responsible for measuring the Wasserstein distance between real and generated image distributions. First, we applied Spectral Normalization to all convolutional layers in the Critic. This helps in limiting the Lipschitz constant of each layer and stabilizes training. Additionally, residual blocks with skip connections is used while downsampling in the Critic. Although not noticeable, this can help critics learn better features and preserve the gradient flow. For loss functions, Wasserstein GAN is used with Gradient Penalty (WGAN-GP) to improve effectiveness. The gradient penalty was applied to interpolated samples between real and fake images. This encourages the gradient norm to stay close to 1. Finally, we updated the WGAN Critic 5 times for each Generator update. This ensured that the Critic learned a meaningful Wasserstein distance before the Generator was updated.

Diverse Image-to-Image Translation GAN (DRIT)

Diverse Image-to-Image Translation GAN (DRIT) is one of the unique GAN models for generating images. Usually, DRIT GAN can generate multiple images from one input image. After receiving an input image, it keeps the age-invariant features unchanged and creates multiple images by changing the feature that changes with age. So, DRIT GAN separates the content and styles. The Generator consists of the Content Encoder, Style Encoder, and Decoder. The Content Encoder preserves features that are not changeable with age. Subsequently, the Style Encoder extracts the changeable features while the decoder adds the features generated by the Content Encoder and Style Encoder. On the other hand, DRIT uses two discriminators: the Global Discriminator and the Local Discriminator. Global Discriminator ensures the overall realism, whereas Local Discriminator focuses on the fine-grained details. Global Discriminator uses CNN, which reduces the input image into a single scaler value by using convolutional layers, pooling, and dense layers. LeakyReLU has been used in hidden layers. The Local Discriminator focuses on specific features by performing patch-based operations, processing each patch independently. The loss functions include Adversarial Loss to ensure the realism of generated images. Meanwhile, Content Consistency Loss preserves the consistency of content features between real and generated images, and Style Diversity Loss helps generate diverse style-based outputs. Here Learning Rate used is 0.0001, the batch size is 16 and the number of epochs is 50.
To enhance the results of the DRIT model, we applied some modifications and improvements. Firstly, in the Content Encoder of the generator part, we used residual blocks with instance normalization, which preserve the facial structure better. Next, in the Decoder, we used Adaptive Instance Normalization (AdaIN) layers to fuse style into content more smoothly. AdaIN helps in adjusting the mean and variance of content features based on the style input. After that, in the Discriminator part, we improved the local discriminator by using multiple receptive fields to detect both fine lines and subtle changes. This will improve its traditional patch-wise realism. Additionally, we also applied spectral normalization to both Global and Local

Discriminators to prevent overpowering the discriminator. This ensures stable adversarial training and reduces mode collapse. In addition to the content consistency loss, we tried to implement a style reconstruction loss. This is done by modifying the code of the output style and matching it to the original style code.

StarGAN

StarGAN is an advanced Generative Adversarial Networks model that can do multi-domain image synthesis to generate high-quality images. StarGAN can transform face images across different age groups while preserving identity. In the StarGAN architecture, there is one Generator and one Discriminator, as shown in the figure 3.5. First of all, the Generator derived as G(x,c), works with a target domain. Here, x represents the input image, which generates output images depending on the conditions mentioned in the target domain. The conditions in the target domains can be Old or Young. In our work, the input image size of 3*128*128 and a target domain (young or old) is given to the Generator. In this way, the Generator can work in both ways, such as young to old or old to young. Here, the Convolutional layers extract features from the input images. LeakyReLU and Instance Normalization (IN) activation were also used. As 3*128*128 size input images pass through the first convolutional layer, it generates an output shape of 64*64*64 with kernel size 7*7, stride 1, and padding 3. The downsampling nature of the Encoder gives an output shape of 256*16*16. Residual blocks, the middle layer, try to capture complex features. This will help in preserving the image structure. Residual blocks consist of two convolutional layers, which have skip connections. For the Decoder part, the input shape 256*16*16 in one transposed convolutional layer outputs the shape 128*32*32, and the other one takes it as input and outputs the shape of 64*64*64 with kernel size of 4*4, stride of 2, and padding 1. When the final convolutional layer activates, it generates output images in the normalization range. Next, the Discriminator takes an input image size of 3*128*128. In convolutional layers, features are extracted. The number of channels increases and images get downsampled. Here, LeakyReLU activation is being used. In the first convolutional layer, the 3*128*128 input shape passes through, which generates an output shape of 64*64*64. In the next convolutional layer, the output shape is 128*32*32. Similarly, the fourth convolutional layer gives output 512*8*8 with kernel size of 4*4, stride of 2, and padding 1. Discriminator generates the value of whether images are real or fake. Meanwhile, the classification loss is predicting the Domain labels of real images. The adversarial loss ensures that the Generator creates realistic images and the Reconstruction loss keeps the consistency. In this model, the learning Rate used is 0.0001, the batch size is 16, and the number of epochs is 50.

3.4 StarGAN - Hybrid Residual Attention Block

In this research work, we have introduced an advanced version of the StarGAN architecture, which can generate face images across different ages. In this StarGAN architecture, we have used **Hybrid Residual Attention Block (HRAB)** struc-

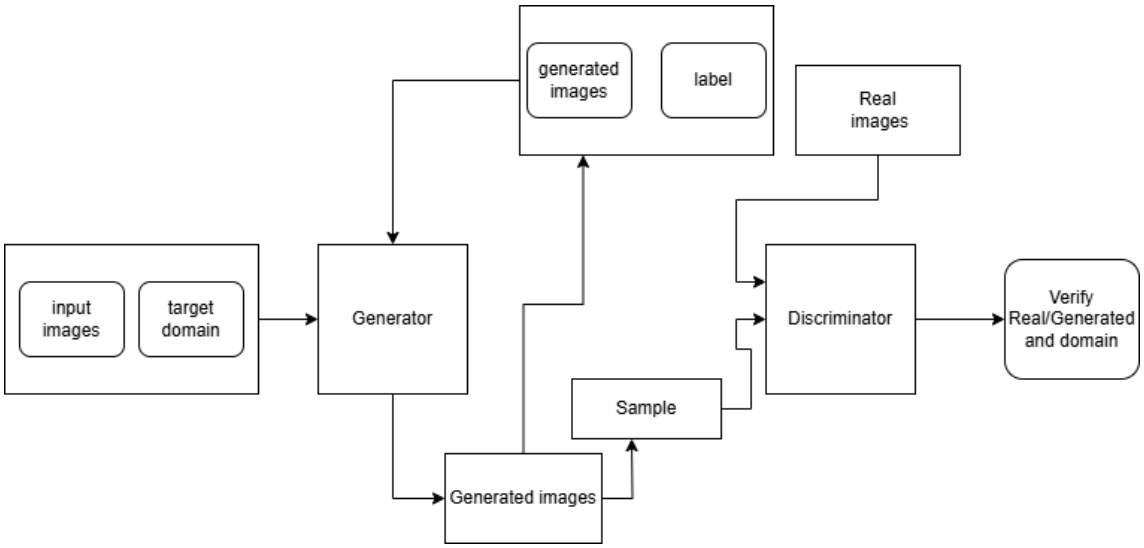


Figure 3.5: Architecture of StarGAN

ture in the generator, which is an innovative StarGAN model architecture for our face aging synthesis task. As accurate face aging depends on the skin feature and changes in the facial regions, such as wrinkles, colors, bone structure shifts and skin sagging, this Hybrid Residual Attention Block structure of the generator improves the accuracy and capability. The generator focuses deeply on the attention features of the faces. Additionally, this StarGAN model is structured to be compatible and operate with the merged UTKFace and MORPH dataset, which works on all the defined age domains. Here we will focus on the architectures of both the Generator and the Discriminator of this model. The figure 3.6 shows the complete diagram of the architecture.

Generator Architecture
Firstly, the Generator consists of Encoder, Transformer and Decoder and it follows their pipeline. But the core innovation in the architecture of the Generator is the Hybrid Residual Attention Block (HRAB) as shown in the figure 3.6. The Generator takes an RGB image of size 128x128x3. Then add it with a domain label tensor. The shape of this is 1x1xN (N = number of age domains, in our work, it is 24). The label is concatenated channel-wise, resulting in an input of shape 128x128x(3+24).

In the Generator, there is an initial convolution called Convolutional layer 2D with 7*7 Kernel, 64 filters or output layers, stride 1 and padding 3. It also consists of InstanceNorm for batch normalization and non-linearity ReLU activation. For the Downsampling layer, one Convolutional layer consists of a 4*4 kernel, 128 filters, stride 2 and padding 1. InstanceNorm -> ReLU. Another Convolutional layer consists of 4*4 kernel, 264 filters, stride 2 and padding 1. InstanceNorm -> ReLU. These layers help to make lower - lower-dimensional latent representation, which is suitable for the transformation.

Then, to focus better on the attention regions of the face, the Hybrid Residual Attention Block (HRAB) was implemented, which has 6 blocks, and each HRAB has two sub-paths. Those two sub-paths are the residual path and the attention

path.

Residual Pathway
Convolutional layer 2D: 3*3 Kernel, stride 1, padding 1. Also InstanceNorm -> ReLU
Convolutional layer 2D: 3*3 Kernel, stride 1, padding 1. Also InstanceNorm
Input to output adds through skip connection.

Attention Pathway
Convolutional layer 2D: 1*1 Kernel, stride 1. ReLU activation
Convolutional layer 2D: 1*1 Kernel, stride 1. Also to generate attention masks there is Sigmoid.
Then in the fusion, residual output multiplies with the attention mask and then through skip connection, it add back to the original input. This actually helps in the special focus on the selective regions.

Upsampling Layers
Deconvolutional layer 2D: 4*4 Kernel, 128 output layers, stride 2 and padding 1.
InstanceNorm -> ReLU
Deconvolutional layer 2D: 4*4 Kernel, 64 output layers, stride 2 and padding 1.
InstanceNorm -> ReLU

Final output layers
Convolutional 2D: 7*7 Kernel, 3 filters, stride 1, padding 3.
Then there is Tanh Activation which helps to map outputs to [-1, 1] range.

Discriminator Architecture
The Discriminator is implemented based on the PatchGAN model which classifies the images into different bins according to their ages and gives the predictions of real and fake images. If we talk about layer configuration, we can see that the first convolutional layer 2D consists of a 4*4 kernel, 64 filters, stride 2 and LeakyReLU activation. The second convolutional layer 2D consists of a 4*4 kernel, 128 filters, stride 2 and InstanceNorm -> LeakyReLU activation. The third convolutional layer 2D consists of a 4*4 kernel, 256 filters, stride 2 and InstanceNorm -> LeakyReLU activation. The fourth convolutional layer 2D consists of a 4*4 kernel, 512 filters, stride 2 and InstanceNorm -> LeakyReLU activation. The 5th convolutional layer 2D consists of a 4 * 4 kernel, 1024 filters, stride 2 and InstanceNorm -> LeakyReLU activation.
The output head consists of an adversarial head and a domain classifier head.

Adversarial Head
Convolutional layer 2D: 3*3 Kernel, 1 filter (real/fake score)

Domain Classifier Head
Convolutional layer 2D: 3*3 Kernel, 24 filters (domain prediction)
About loss functions, hinge loss for adversarial output and for domain prediction, softmax cross-entropy loss. Here, Learning Rate used is 0.0001, the batch size is 16 and the number of epochs is 50. Training time taken about 8.5 hours.

Testing Phase

In this research work, the testing phase is the crucial phase which offers the evaluation of the generated outputs not only in statistical form but also offers visualization to understand it better. Testing phase is designed in such a way that generates face images into different age labels and also quantifies drift heatmaps to evaluate the trained StarGAN model with the enhanced capacity of Hybrid Residual Attention Blocks. So, this pipeline can be considered as a critical component of this whole idea.

Testing pipelines:

Data loading
Input images are loaded from the test datasets and normalized to the range [-1, 1]. Age label is parsed from the filenames.
Domain label encoding
For a given age interval, one hot encoded vector gets created and represents a target age domain. For example, if the target age is 25, it is under the 20-25 domain which represents the 5th index from the total of 24 indexes.
Image generation
The output is the face images across different ages of the input image.
Heatmap generation and computation
After generating images, the real and fake ones got compared pixel-wise and the absolute difference is calculated through per pixel which also generates color coded heatmap using the applyColorMap() function of the OpenCV.
Statistical drift analysis
We computed the mean, max and Standard deviation of drift values for each pixel difference map.
Result storage
For result storage, google drive has been used where designated files were created for face ages, heatmaps and statistical results.
Heatmap Generation pipeline
Heatmap helps to visualize the exact spatial location of the face images which happens due to the aging. Modifications of the faces can also be visualized by this heatmap feature. Firstly, pixel-wise differences were computed. The generated images and the real images got denormalized into [0, 255] and converted to 8-bit. The differences are calculated for each of the RGB channels and we take the average of the channels to create a grayscale difference map. Then applying the color map to visualize it more precisely and clearly. The previously grayscale map was then converted into colored visualization using OpenCV's COLORMAP_JET. This actually generated visually interpreted images where color intensity shows the magnitude of the facial transformation.
Drift Metric Computation
After the drift map generation, we have computed mean drift which shows average pixel difference indicating the overall facial changes. Then, we calculated max drift, which is the highest value of the facial drift map which is localized to key facial features such as wrinkles near mouth or eyes. The. We have computed standard deviation, which measures the variability in the faces due to aging.
We can understand the concept easily by taking examples. If wrinkles happen in the lower part of the mouth, then the lower part the drift intensity would be higher than the upper and the max drift will also be higher. But the mean drift will be

lower in that example. We have saved all these statistics in the CSV log for every testing session.
For architectural justification for testing workflow, we can see one-to-many inference where the training model, StarGAN with the Hybrid Residual Attention Block, enables the generation of multiple age conditions from one input. The drift quantification pipeline allows the system to be used in multiple models. The visualization used here is lightweight and easy to use which can be easily implemented in any field.

Training and Testing Infrastructure

Following components were used for training and testing pipeline:

Training Hardware: Kaggle GPU (P100)
Testing Hardware: Google colab GPU (T4)
Training Model checkpoints storage: Hugging Face
Testing result storage: Google Drive
Frameworks: PyTorch, OpenCV 4.7, NumPy

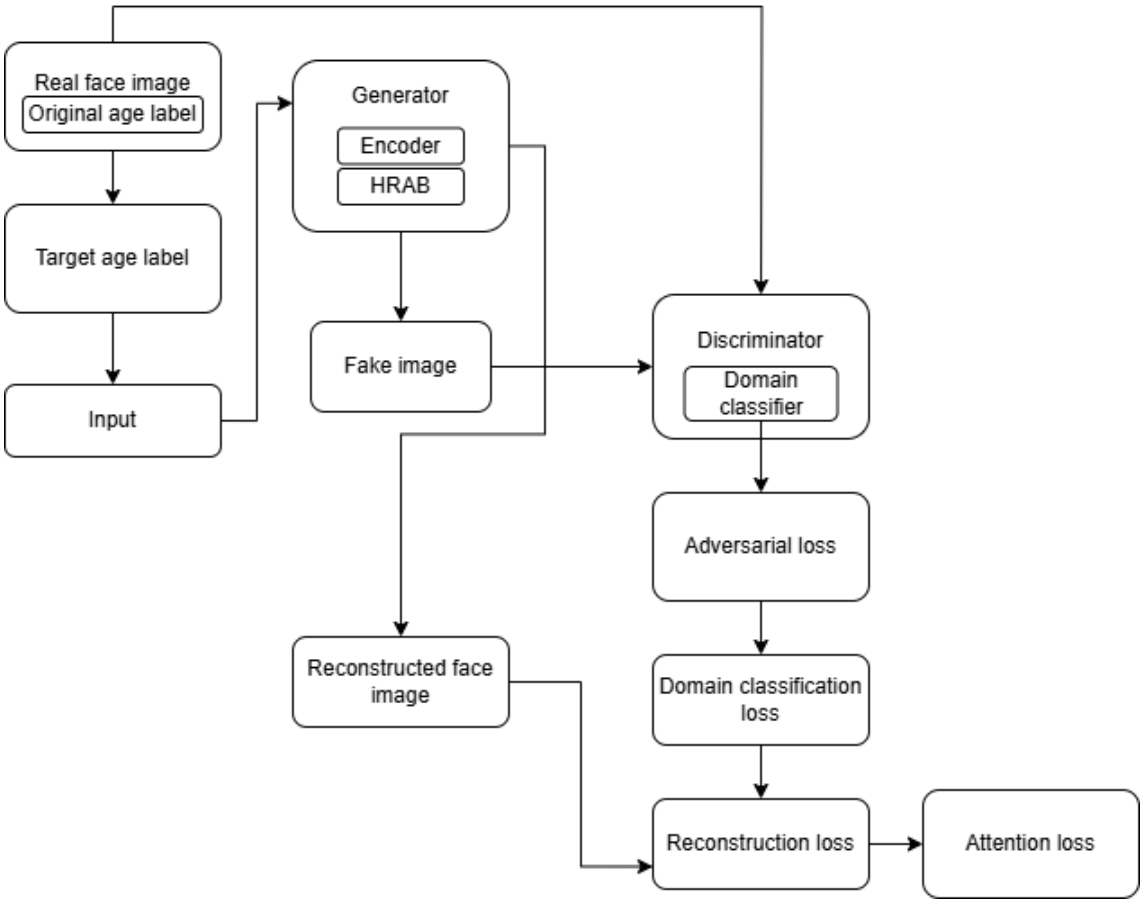


Figure 3.6: Training pipeline of StarGAN - HRAB

Chapter 4

Experimental Results

4.1 Performances

To measure the performances of the models that we implemented for the face aging synthesis task, we have used the Frechet Inception Distance (FID) score and Learned Perceptual Image Patch Similarity (LPIPS) score. We wanted to evaluate the models by using different methods which gave us the results based on different criterias. For our work, FID gives the overall realism of the generated images and LPIPS focuses on the small details such as skin tone, wrinkles, hair, etc. Also FID score and LPIPS score are good for interpret unsupervised face aging pipeline.

FID measures the similarity between the generated image and the real-life counterpart. To find out FID scores, we used the Inception v3 model which is pre-trained. Input images were resized so that the size could be matched with the model. The Pool3 layer of the Inception v3 model extracts the features from the images. To compute the FID score, the average and covariance of the feature vectors were computed for both the real and generated images. The average of the feature vectors is called Mean and the covariance of the feature vectors is called Covariance Matrix. The table 4.1 shows the FID scores of the six selected GAN models.

$$\text{FID} = \|\mu_r - \mu_g\|_2^2 + \text{Tr}\left(\Sigma_r + \Sigma_g - 2\sqrt{\Sigma_r \Sigma_g}\right)$$

Here:
μ_r and μ_g : Means of the feature for real (μ_r) and generated (μ_g) images.
Σ_r and Σ_g : Covariance matrices of the feature for real (Σ_r) and generated (Σ_g) images.
Tr : Trace of sum of diagonal elements of matrix.
$\|\cdot\|_2^2$: Squared Euclidean distance between the means.

Models	DCGAN	StarGAN	Attention GAN	WGAN	CycleGAN	DRIT GAN
FID Scores	73.4	48.1	56.8	65.2	49.4	59.5

Table 4.1: FID scores of implemented GAN models

On the other hand, LPIPS is more focused on finding perceptual similarities between images. To find out the LPIPS score, a feature map had to be extracted. For

this reason, the AlexNet model was used which was already pre-trained. Feature maps were normalized to unit length. Then squared differences were computed between those normalized feature maps. By a learned parameter, the differences were weighted. The table 4.2 shows the LPIPS scores of the six selected GAN models.

$$\text{LPIPS}(x, y) = \sum_{l} \frac{1}{H_l W_l} \sum_{h=1}^{H_l} \sum_{w=1}^{W_l} w_l \cdot \| f_l(x)_{h,w} - f_l(y)_{h,w} \|_2^2$$

Here:
x and y: Two input images (real and generated).
$f_l(x)$ and $f_l(y)$: Feature maps from layer l of a pretrained network for images x and y.
w_l : Learned weight for layer l.
H_l, W_l : Height and width of the feature map at layer l.
$\| \cdot \|_2^2$: The squared Euclidean distance between corresponding feature activations.

Models	DCGAN	StarGAN	Attention GAN	WGAN	CycleGAN	DRIT GAN
LPIPS score	0.36	0.23	0.27	0.41	0.27	0.33

Table 4.2: LPIPS scores of implemented GAN models

In our research, we initially tried to compare the differences in architecture and performance between the 6 mentioned GAN models. For both the FID and LPIPS scores, the lower the score is, the better the performance is. The lower scores for FID and LPIPS demonstrate that realism exists in generated photos. From the performance tables 4.1 and 4.2, we can see that StarGan gives better results in both of the evaluation matrices. Also image generated by the StarGAN model was better than the other models, which we can see in the figure 4.8. So, the satisfactory results make StarGAN ideal for face synthesis tasks.

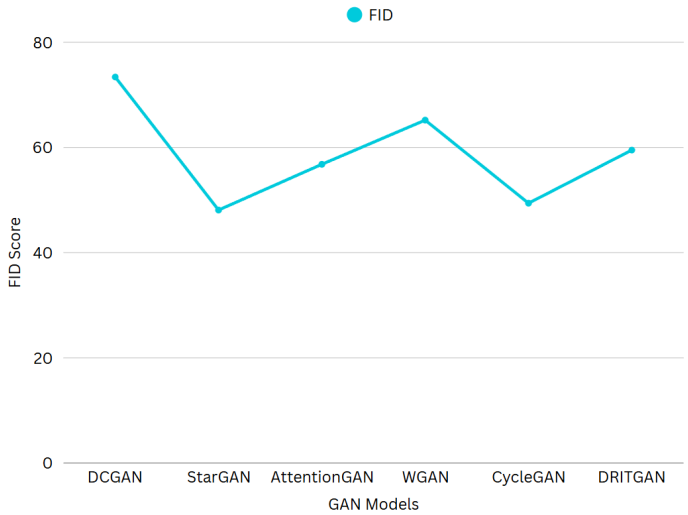


Figure 4.1: Comparison of FID scores

3

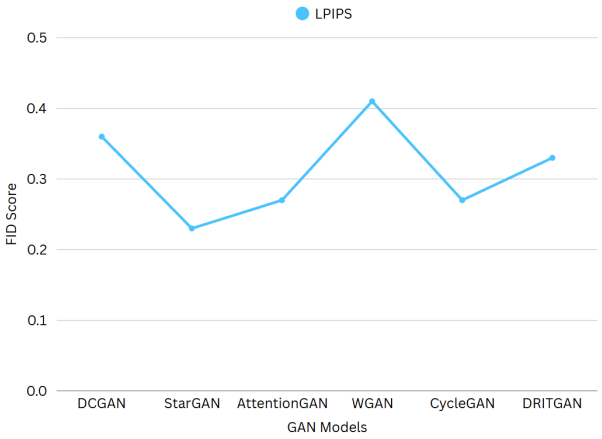


Figure 4.2: Comparison of LPIPS scores

We can get better understanding of the performances of the GAN models from figures 4.1 and 4.2. It is clear that the performances of DCGAN and WGAN were not satisfactory and lots of improvement needs to be done. DRIT GAN and Attention GAN performance on FID is better than DCGAN and WGAN but worse than StarGAN and CycleGAN. However, with some changes in the architecture, those models can give much better output in the face-aging synthesis tasks. The generated images of CycleGAN are also good, however the generated images look red-colored. In the FID evaluation, CycleGAN scores similar to StarGAN and better results than other models. In LPIPS, CycleGAN's score is the same as AttentionGAN. Overall, the satisfactory results in the evaluation matrices prove StarGAN's effectiveness over other models.

While researching, we found that Deep Convolutional Generative Adversarial Networks (DCGAN) have a simple architecture compared to other GAN models discussed in this paper. From the figure 4.8, we can see that the StarGAN architecture outputs better results than DCGAN while both use 1 Generator and 1 Discriminator. Although StarGAN provides better performance, it has obvious changes in the architecture despite having the same number of Generators and Discriminators. As input, StarGAN needs the domain young or old to generate images, whereas DCGAN doesn't need that and lacks realism in the output images. On the other hand, CycleGAN uses two Generators and Discriminators to transition the images from young to old or old to young. So, differences are seen in the architecture of the models. While StarGAN can do both tasks in one unified model. But it also increases the complexity of the architecture. Meanwhile, AttentionGAN is different from the previously discussed three GAN models as its Generator is divided into two parts, where the Attention branch is involved in finding out which portions of the face will be changed from young to old and old to young. While the Transformation branch works on changing these facial features. On the other hand, DRIT GAN has a similar architecture to the Attention GAN. DRIT GAN's Generator contains the Content Encoder and Style Encoder, where the Style Encoder is similar to the Transformation branch of Attention GAN. However, differences are found in the Content Encoder of DRIT GAN, which works differently from the Attention branch of Attention GAN.

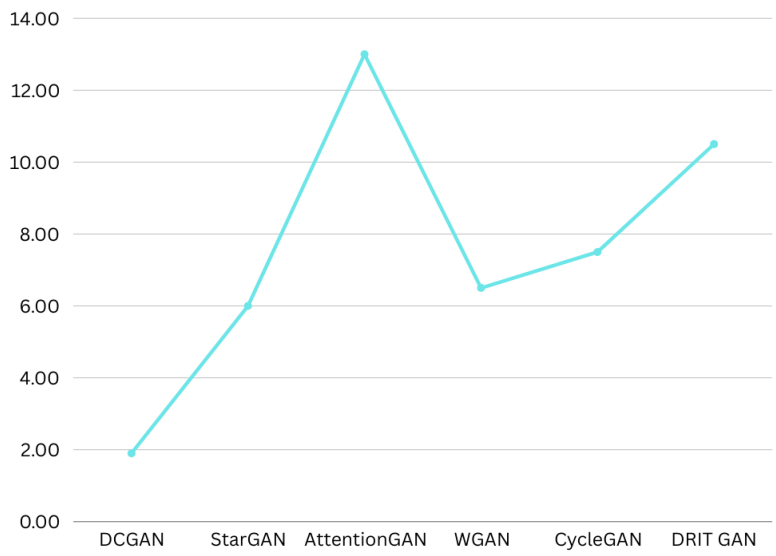


Figure 4.3: Comparisons of computational times taken to train the models

Note: The horizontal line indicates models and the vertical line indicates computational times (in hours) in the above chart.

When looking into the computational times taken by those models in figure 4.3, we find that AttentionGAN and DRIT GAN took more time than other GAN models which are around 13 and 10.5 hours. CycleGAN, StarGAN, and WGAN took around 7.5, 6, and 6.5 hours respectively to be trained. DCGAN took very little time which was less than 2 hours. We have done the work on the P100 GPU offered by Kaggle.

4.1.1 Outputs of the Mentioned GAN models

Figure 4.4: Original image

Figure 4.5: Generated by DCGAN

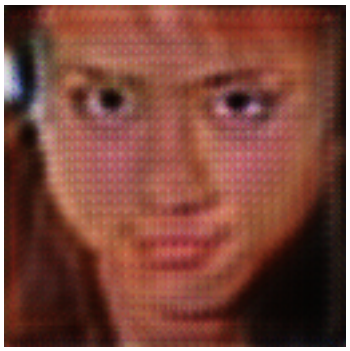


Figure 4.6: Generated by CycleGAN

Figure 4.7: Generated by Attention GAN

Figure 4.8: Generated by StarGAN

Figure 4.9: Generated by DRIT GAN

Figure 4.10: Generated by WGAN

4.1.2 Performance of StarGAN - HRAB

Initially, we worked on the 6 different GAN models and measured their performance in the table 4.1 and 4.2. Later, we specifically focused on StarGAN with HRAB to generate face images alongside drift heatmaps. From the table 4.3 we can see the FID and LPIPS scores for different age intervals. By calculating the average we get the FID score 43.34 and LPIPS score 0.237. Also, we see that both FID and LPIPS scores increase in the older age intervals. The low number of older face images in the dataset can be the reason behind it.

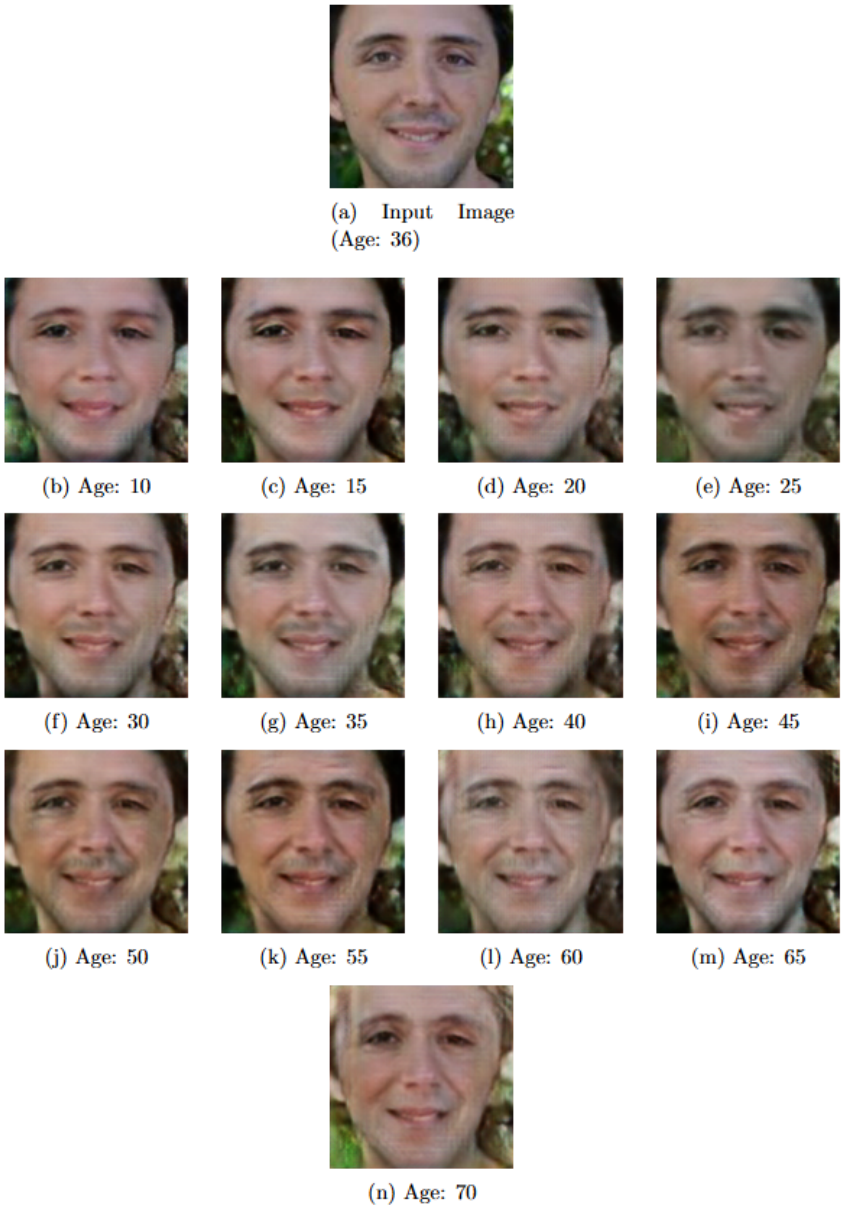
age_bin	FID	LPIPS
5	44.04186239287196	0.19250503840545814
10	42.7589971021879	0.21616310872137547
15	41.62323321930833	0.18529146946966648
20	33.62082430660584	0.1569664487739404
25	29.96339993166248	0.17798145294189453
30	34.5184092064888	0.20330021262168884
35	36.4410981345663	0.21562202525635569
40	33.0268849513414	0.22776743421951938
45	38.48943834092157	0.2423126404484113
50	41.10811540712385	0.23891454982260863
55	38.34985035596733	0.18973742812871933
60	42.67401530340538	0.22487564697861671
65	37.3247232531037	0.22334825600186984
70	37.36796477559955	0.18390294671058655
75	40.3405460025776	0.2422820260127385
80	44.3054347752558	0.2691044519344966
85	49.4448217538827	0.2541332709789276
90	47.8399421589222	0.2918309857447942
95	54.83712869350103	0.2930305834611257
100	55.51852019511415	0.30390563428401947
105	55.18123064885134	0.30725576798121136
110	59.72114912935456	0.3262053060531616
115	58.9353297214995	0.3326810121536255

Table 4.3: FID and LPIPS scores in different age bins using StarGAN - HRAB

4.2 Face Aging Synthesis

In this section, we showed the output of our proposed StarGAN-HRAB model. We generated face-aged images for every five-year age interval. From an input face image, we can generate age-progressed images from the age of 5 to the age of 115. Some example outputs are shown in the figure 4.11 and 4.12.

4.2.1 Sample 1



(a) Input Image (Age: 36)

(b) Age: 10 (c) Age: 15 (d) Age: 20 (e) Age: 25

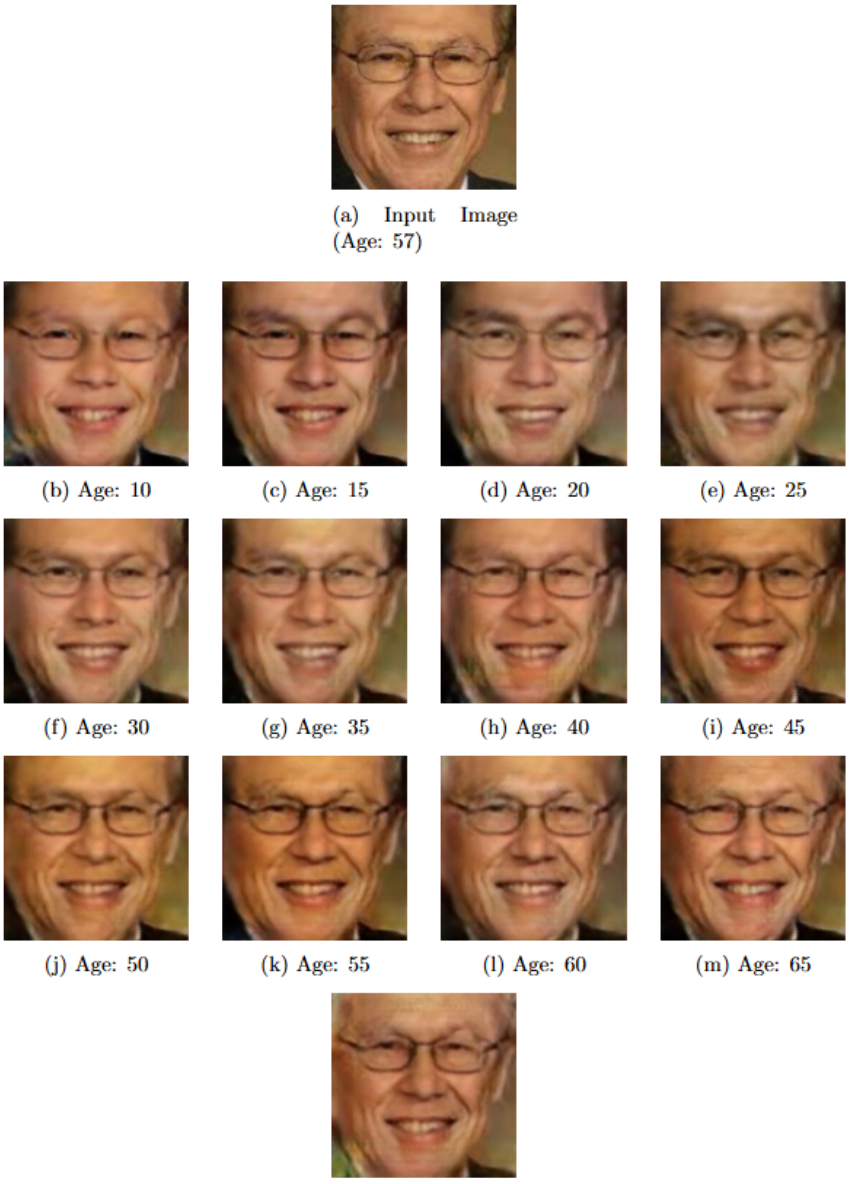
(f) Age: 30 (g) Age: 35 (h) Age: 40 (i) Age: 45

(j) Age: 50 (k) Age: 55 (l) Age: 60 (m) Age: 65

(n) Age: 70

Figure 4.11: Face Aging Synthesis (Example 1)

4.2.2 Sample 2



(a) Input Image (Age: 57)

(b) Age: 10 (c) Age: 15 (d) Age: 20 (e) Age: 25

(f) Age: 30 (g) Age: 35 (h) Age: 40 (i) Age: 45

(j) Age: 50 (k) Age: 55 (l) Age: 60 (m) Age: 65

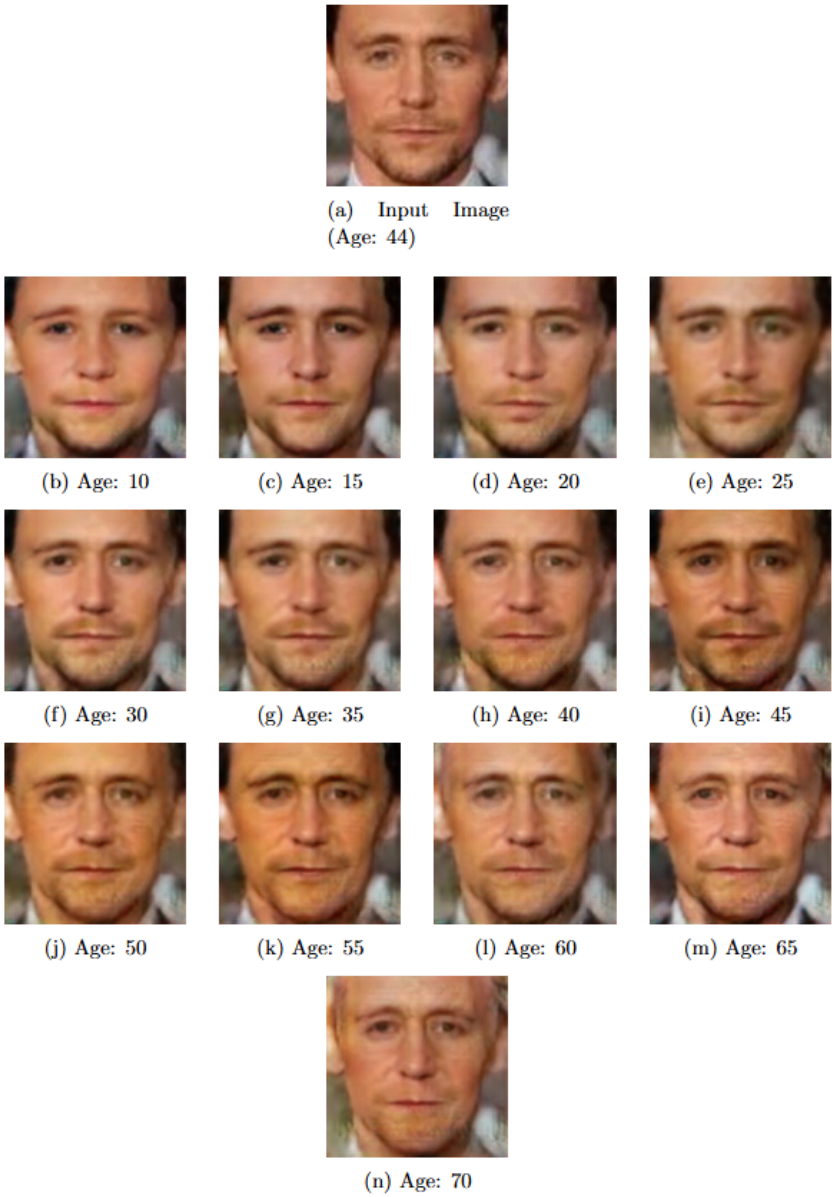
(n) Age: 70

Figure 4.12: Face Aging Synthesis (Example 2)

4.3 Face Aging Synthesis with Heatmap and Drift Heatmap

With the proposed model, we also generated and visualized the heatmaps for identity drift while aging. This identity drift heatmap showed us the gradual changes while aging. The figure 4.13 shows the Aging Synthesis, while 4.14 and 4.15 shows the corresponding Heatmap and Drift Heatmap. Similarly, Heatmap and Drift Heatmap are shown for the outputs 4.16 and 4.19.

4.3.1 Sample 1



(a) Input Image (Age: 44)

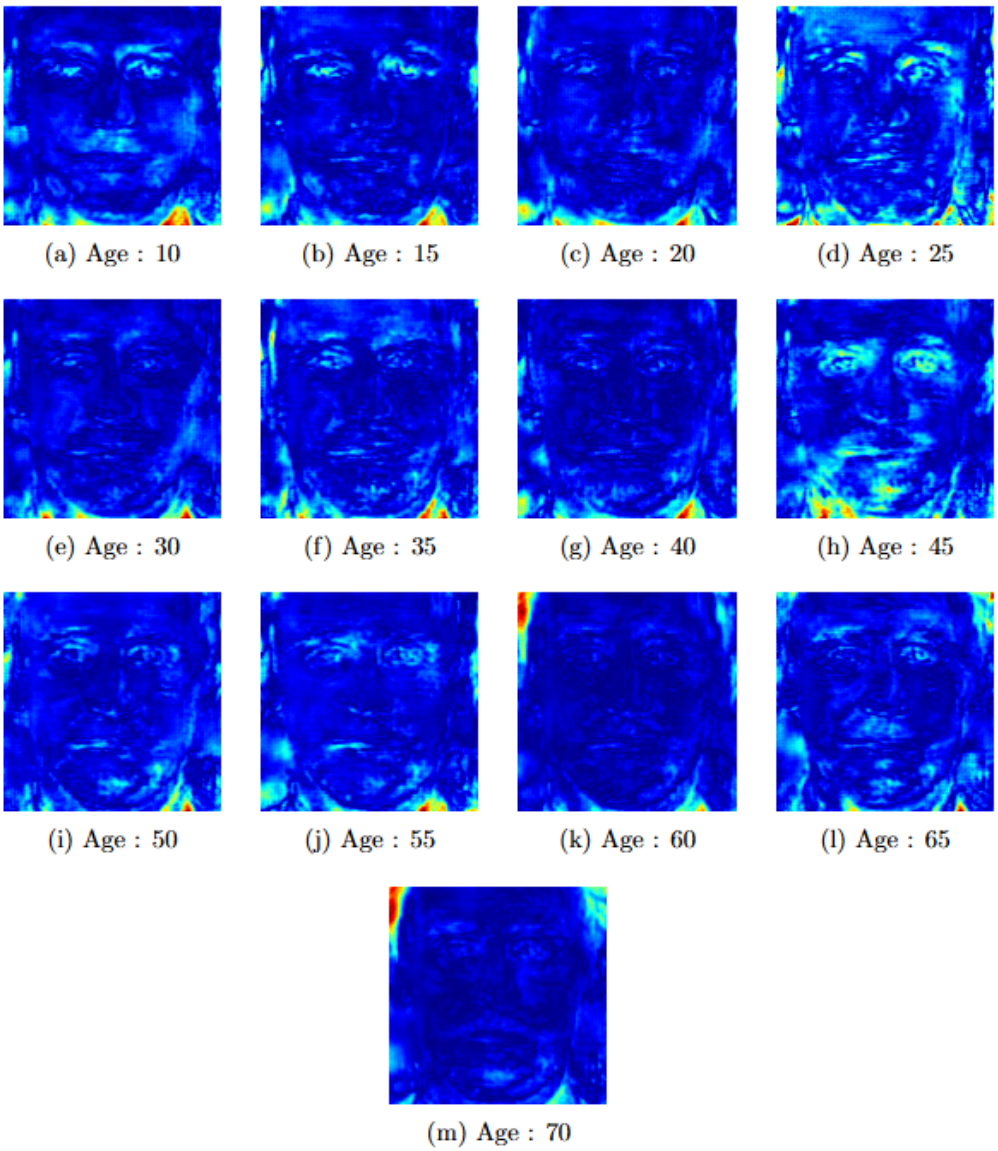
(b) Age: 10 (c) Age: 15 (d) Age: 20 (e) Age: 25

(f) Age: 30 (g) Age: 35 (h) Age: 40 (i) Age: 45

(j) Age: 50 (k) Age: 55 (l) Age: 60 (m) Age: 65

(n) Age: 70

Figure 4.13: Face Aging Synthesis (sample 1)



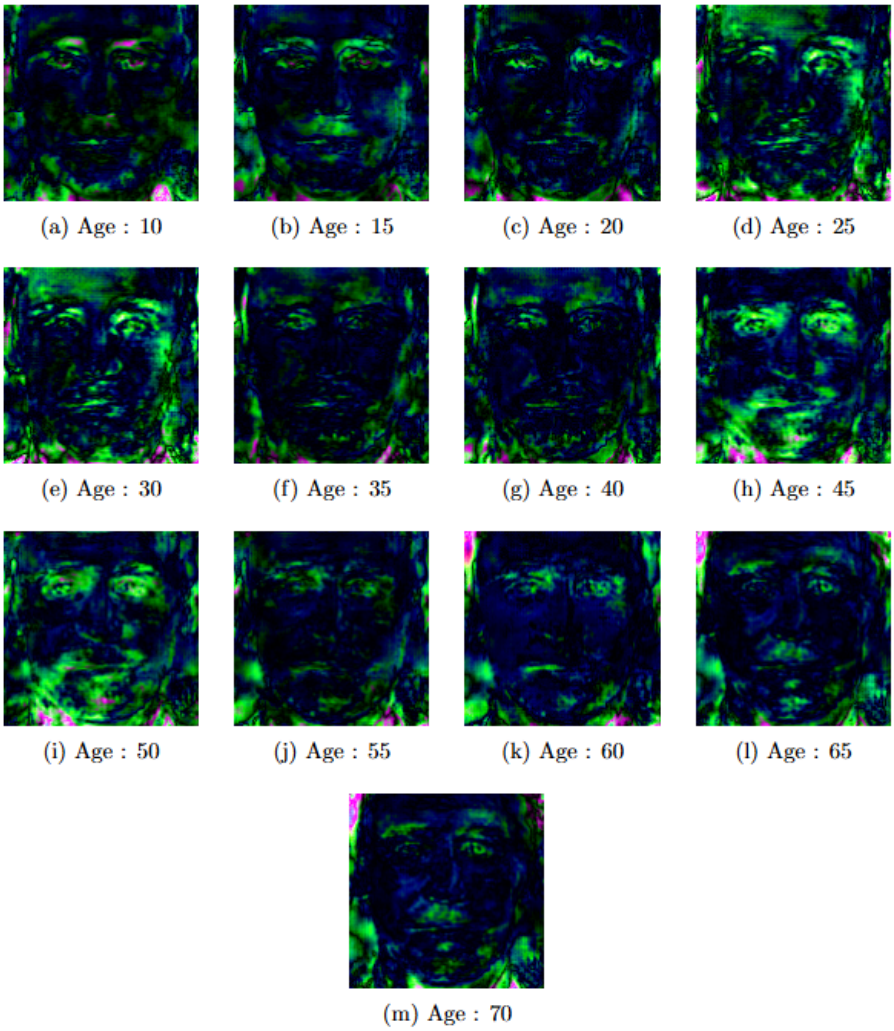
(a) Age : 10 (b) Age : 15 (c) Age : 20 (d) Age : 25

(e) Age : 30 (f) Age : 35 (g) Age : 40 (h) Age : 45

(i) Age : 50 (j) Age : 55 (k) Age : 60 (l) Age : 65

(m) Age : 70

Figure 4.14: Face Aging Synthesis Heatmap (sample 1)



(a) Age : 10 (b) Age : 15 (c) Age : 20 (d) Age : 25

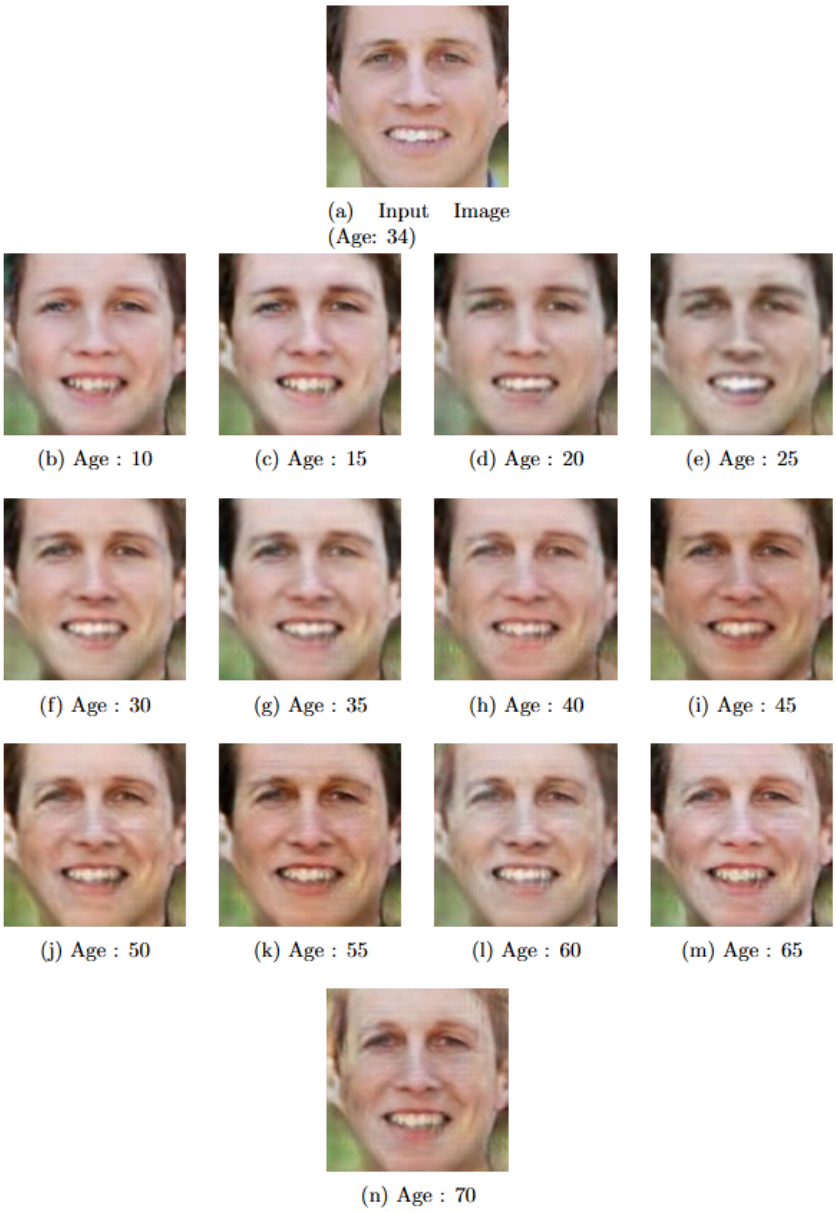
(e) Age : 30 (f) Age : 35 (g) Age : 40 (h) Age : 45

(i) Age : 50 (j) Age : 55 (k) Age : 60 (l) Age : 65

(m) Age : 70

Figure 4.15: Face Aging Synthesis Drift Heatmap (sample 1)

4.3.2 Sample 2



(a) Input Image (Age: 34)

(b) Age : 10 (c) Age : 15 (d) Age : 20 (e) Age : 25

(f) Age : 30 (g) Age : 35 (h) Age : 40 (i) Age : 45

(j) Age : 50 (k) Age : 55 (l) Age : 60 (m) Age : 65

(n) Age : 70

Figure 4.16: Face Aging Synthesis (sample 2)

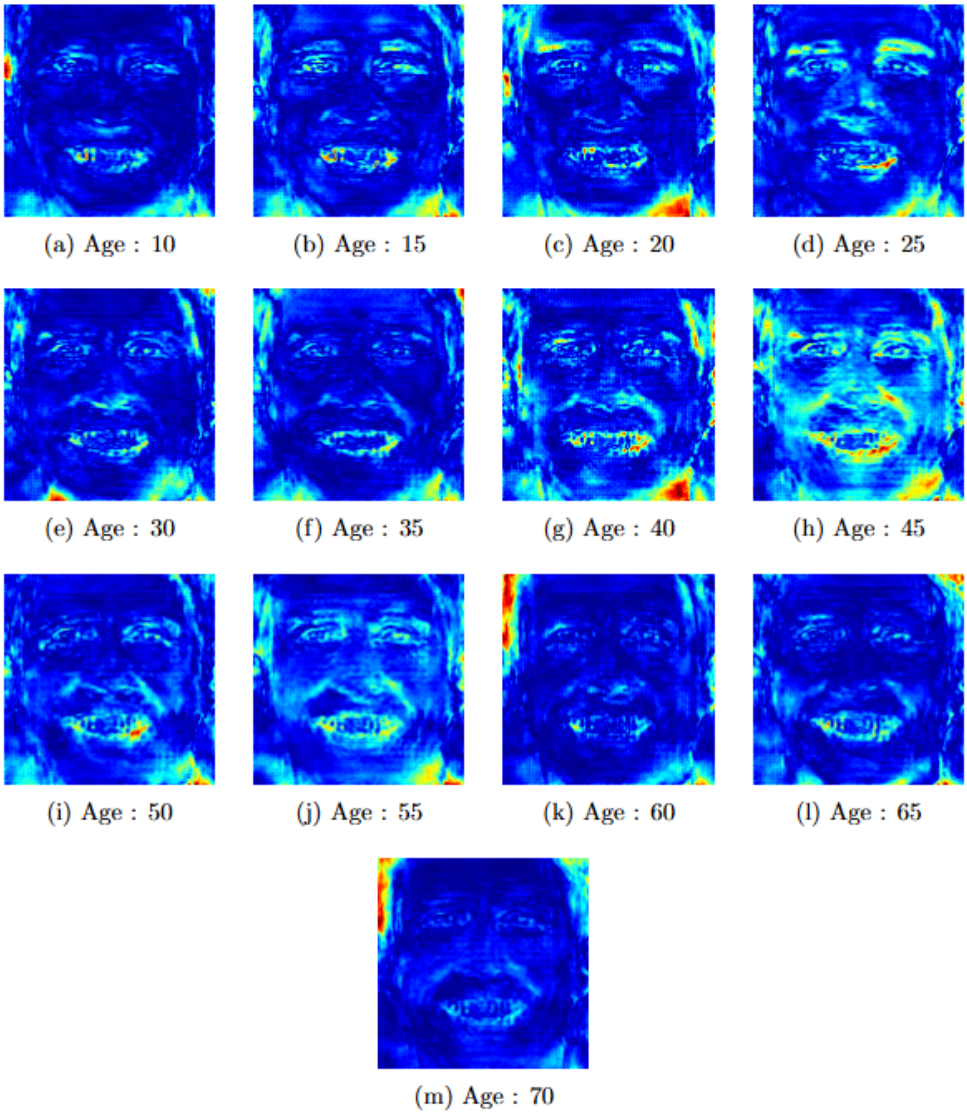
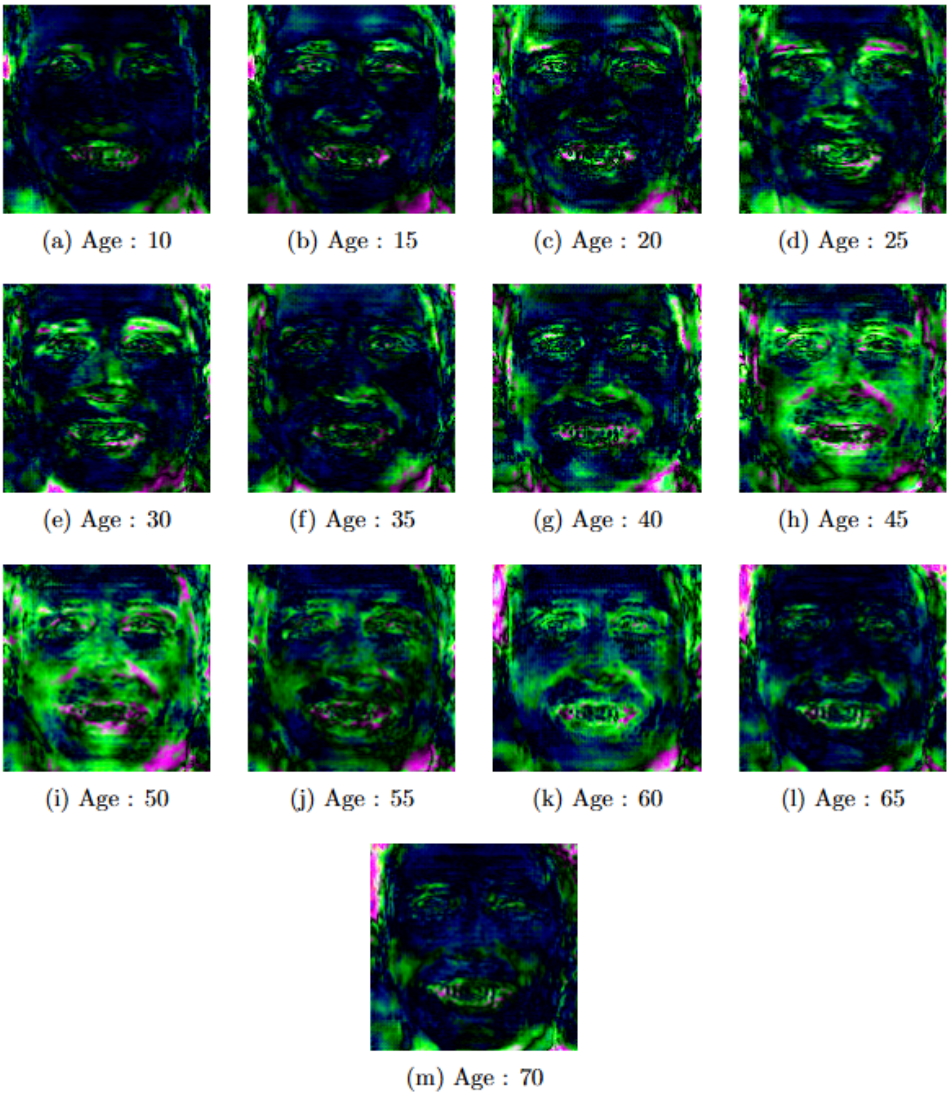


Figure 4.17: Face Aging Synthesis Heatmap (sample 2)



(a) Age : 10 (b) Age : 15 (c) Age : 20 (d) Age : 25

(e) Age : 30 (f) Age : 35 (g) Age : 40 (h) Age : 45

(i) Age : 50 (j) Age : 55 (k) Age : 60 (l) Age : 65

(m) Age : 70

Figure 4.18: Face Aging Synthesis Drift Heatmap (sample 2)

4.3.3 Sample 3

(a) Input Image (Age: 15)

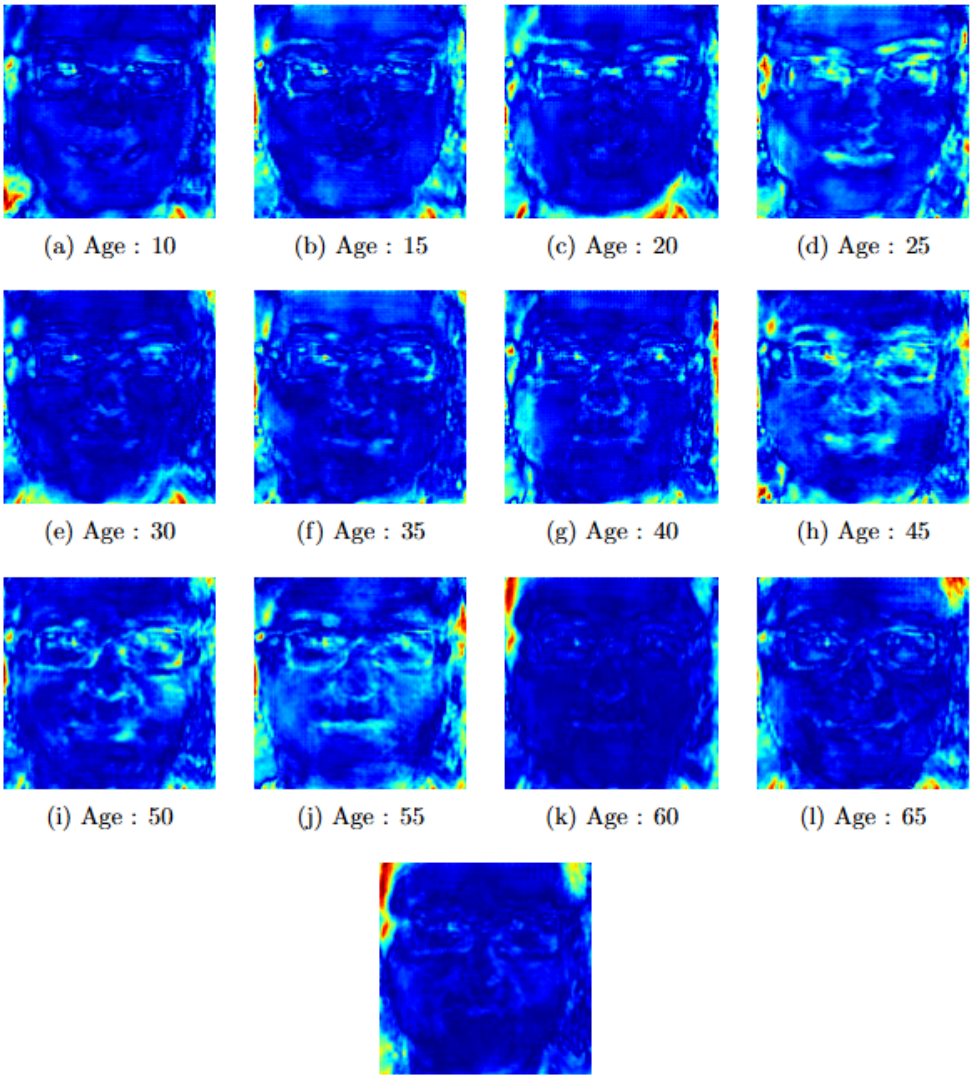
(b) Age : 10 (c) Age : 15 (d) Age : 20 (e) Age : 25

(f) Age : 30 (g) Age : 35 (h) Age : 40 (i) Age : 45

(j) Age : 50 (k) Age : 55 (l) Age : 60 (m) Age : 65

(n) Age : 70

Figure 4.19: Face Aging Synthesis (sample 3)



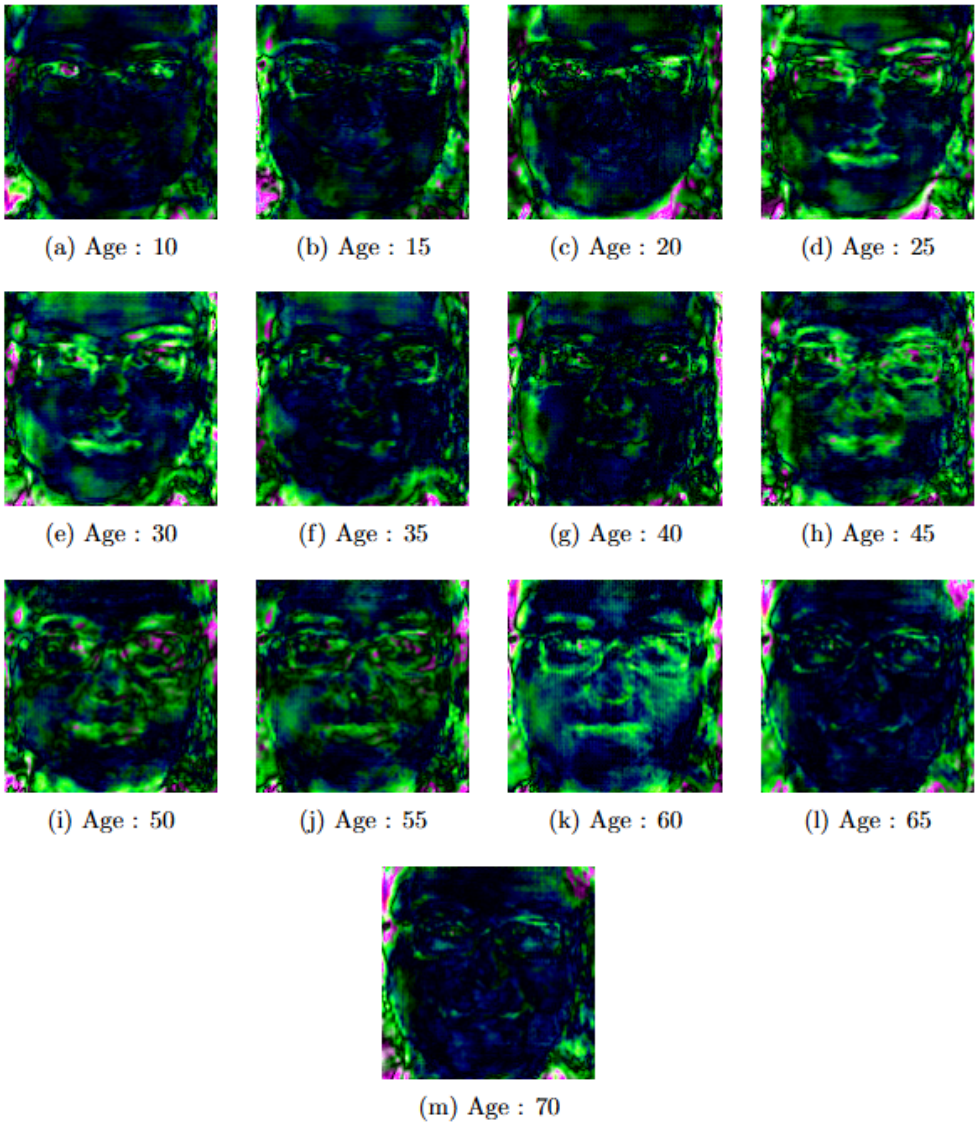
(a) Age : 10 (b) Age : 15 (c) Age : 20 (d) Age : 25

(e) Age : 30 (f) Age : 35 (g) Age : 40 (h) Age : 45

(i) Age : 50 (j) Age : 55 (k) Age : 60 (l) Age : 65

(m) Age : 70

Figure 4.20: Face Aging Synthesis Heatmap (sample 3)



(a) Age : 10 (b) Age : 15 (c) Age : 20 (d) Age : 25

(e) Age : 30 (f) Age : 35 (g) Age : 40 (h) Age : 45

(i) Age : 50 (j) Age : 55 (k) Age : 60 (l) Age : 65

(m) Age : 70

Figure 4.21: Face Aging Synthesis Drift Heatmap (sample 3)

4.4 Mean, Max and Standard Deviation of Drift

Here, Mean drift refers to the average pixel changes. This also indicates the general aging of the images. While the Max Drift refers to regions with higher changes, such as shifts in structure (Jawline, forehead wrinkle). The Standard Deviation (std) drift measures the consistency and inconsistency of the face aging synthesis. The Mean, Max and Standard Deviations are calculated with their traditional equations. These three metrics are crucial for evaluating the Identity Drift Heatmap. In the table 4.4, 4.5, 4.6 we showed the Mean, Max and Standard Deviation values from the samples we generated in the figure 4.15, 4.18 and 4.21. In table 4.4, it is showing the Mean drift is higher at age 50 and lower at age 35, whereas at age 40, it is showing the maximum drift or peak changes. Standard Deviation (Std) drift is showing consistency across different ages. The table 4.5 shows that the Mean drift is higher at age 50 and lower at age 10. Std drift is consistent at early ages. For table 4.6, it shows that the Mean drift is higher at age 45 and lower at age 10. Whereas at age 40, it is showing the Max drift. Also, Std drift is consistent at early ages.

age_bin	mean_drift	max_drift	std_drift
10	20.9254150390625	139	30.56419231916107
15	21.67413330078125	106	28.21751623576097
20	17.47491455078125	98	26.864423438756322
25	35.24627685546875	146	41.89731307990918
30	35.90313720703125	145	42.93158426680679
35	16.83740234375	144	25.01447437411227
40	20.60467529296875	153	30.38727938070743
45	37.86724853515625	150	48.085409004993274
50	39.27569580078125	142	44.64157364814174
55	19.83575439453125	91	29.916070152370498
60	21.60955810546875	143	35.6832064097073
65	22.68817138671875	147	36.195648798161436
70	21.388916015625	151	34.75073050599049

Table 4.4: Mean, Max and Standard Deviation of sample 1

age_bin	mean_drift	max_drift	std_drift
10	17.25811767578125	143	29.587115870906004
15	23.74249267578125	153	34.019359306926326
20	31.2476806640625	152	40.77463795765343
25	37.6812744140625	104	46.14875400302937
30	31.12628173828125	153	42.278095522717734
35	25.2154541015625	141	34.5649116959055
40	38.62384033203125	150	45.271943547373105
45	41.63623046875	142	49.39982312968093
50	45.35052490234375	146	49.86862864582796
55	33.91680908203125	147	35.43385674440317
60	31.41864013671875	149	36.3811482163769
65	27.86773681640625	151	41.16243249293489
70	24.41912841796875	151	37.442859171693634

Table 4.5: Mean, Max and Standard Deviation of sample 2

age_bin	mean_drift	max_drift	std_drift
10	23.03192138671875	132	35.713377481312605
15	29.25604248046875	132	40.44565347623016
20	32.10443115234375	125	42.196609489806185
25	35.83978271484375	132	38.57821824584624
30	30.306396484375	153	41.021425953716864
35	27.7427978515625	149	36.77130559844475
40	30.0003662109375	157	41.78715278122215
45	45.45843505859375	147	45.61638382782573
50	37.40924072265625	146	42.081080360145386
55	36.6964111328125	129	39.94114929545602
60	31.51531982421875	154	31.632808454098964
65	27.46734619140625	142	43.903413032659365
70	25.7872314453125	149	42.1920297689802

Table 4.6: Mean, Max and Standard Deviation of sample 3

4.5 Comparisons

After working with the 6 mentioned GAN models, we focused on our main goal to generate face images alongside Identity Drift Heatmap using StarGAN - HRAB. We generated face-aged images for every five-year age interval. From an input face image, we can generate age-progressed images from the age of 5 to the age of 115. Moreover, we also generated and visualized the heatmaps for identity drift while aging. In our research, we didn't find any work that exactly matches with our work, and the papers we referenced in our work didn't provide any identity drift heatmap. There are a few works that have researched drift heatmaps on images using GAN models. But none of them specifically implemented it to show the gradual changes in facial aging. For example, in the paper [24], they have done the face aging task

by using Explainable Conditional Adversarial Autoencoders, analyzing the images. They did provide heatmaps, which are for age classification prediction. Also, the paper [6] provides drift heatmap, but not to visualize the face aging changes across different age groups.

Models	FID	LPIPS	Used Dataset
PFA-GAN (2021)	36.2	0.29	FG-NET
CycleGAN (2022)	60.8	-	CelebA-HQ
AttentionGAN (2022)	55.1	-	CelebA-HQ
IPCGAN (2020)	27.4	0.15	CACD
StarGAN - HRAB	35.7	0.26	CACD
StarGAN - HRAB	57.4	0.31	CelebA-HQ

Table 4.7: FID and LPIPS scores comparison with other GAN models

In the table 4.7 we included the FID and LPIPS scores of some recognised models from different research papers to compare them with each other. Here, we can see that our model, StarGAN with Hybrid Residual Attention Blocks (HRAB), performed better than CycleGAN and AttentionGAN on the same dataset. It is also evident that StarGAN - HRAB performs better with the CACD dataset. Overall, StarGAN-HRAB showed promising FID and LPIPS scores among many other models. Also, for better visualization and understanding of comparison between GAN models, figures 4.22 and 4.23 are presented. Which further proves our model's efficiency.

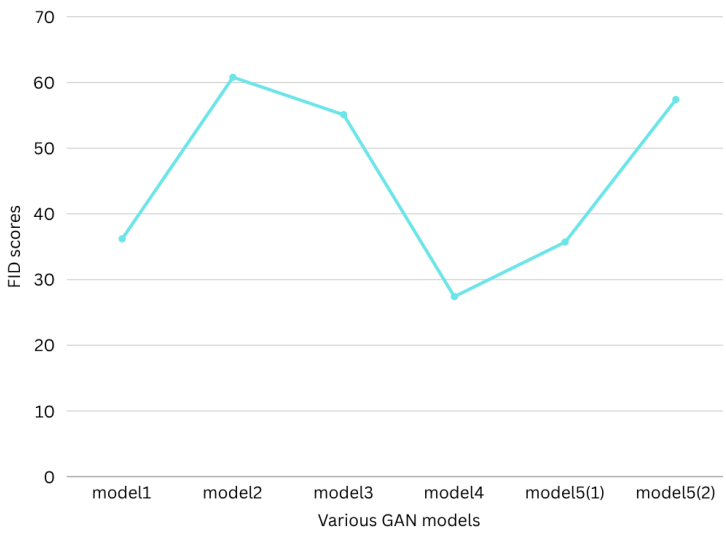


Figure 4.22: FID score comparison of StarGAN - HRAB and other models

Note: Here, model1 -> PFA-GAN
model2 -> CycleGAN
model3 -> AttentionGAN
model4 -> IPCGAN
model5(1) -> StarGAN - HRAB (Tested on CACD)
model5(2) -> StarGAN - HRAB (Tested on CelebA-HQ)

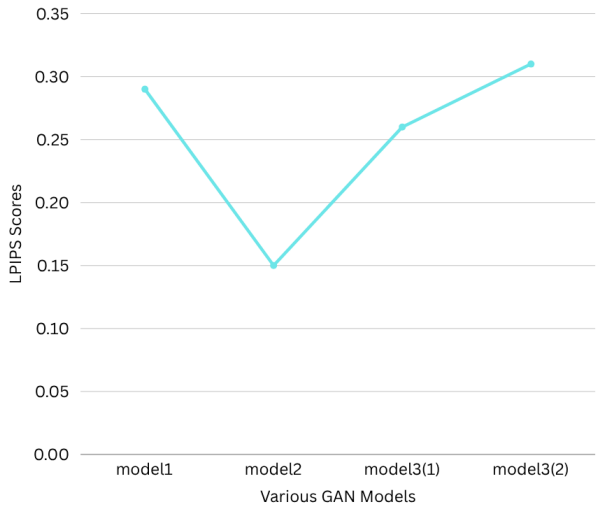


Figure 4.23: LPIPS score comparison of StarGAN - HRAB and other models

Note: Here, model1 -> PFA-GAN
model2 -> IPCGAN
model3(1) -> StarGAN - HRAB (Tested on CACD)
model3(2) -> StarGAN - HRAB (Tested on CelebA-HQ)

4.6 Comparative Analysis of the Outputs

In this section, we generated our own face images to further evaluate the capabilities of the proposed StarGAN model. In the figure 4.24 and 4.27 we showed the aging process. While the figures 4.25, 4.26 and 4.28, 4.29 showed corresponding heatmap and drift heatmap.

4.6.1 Example: 1

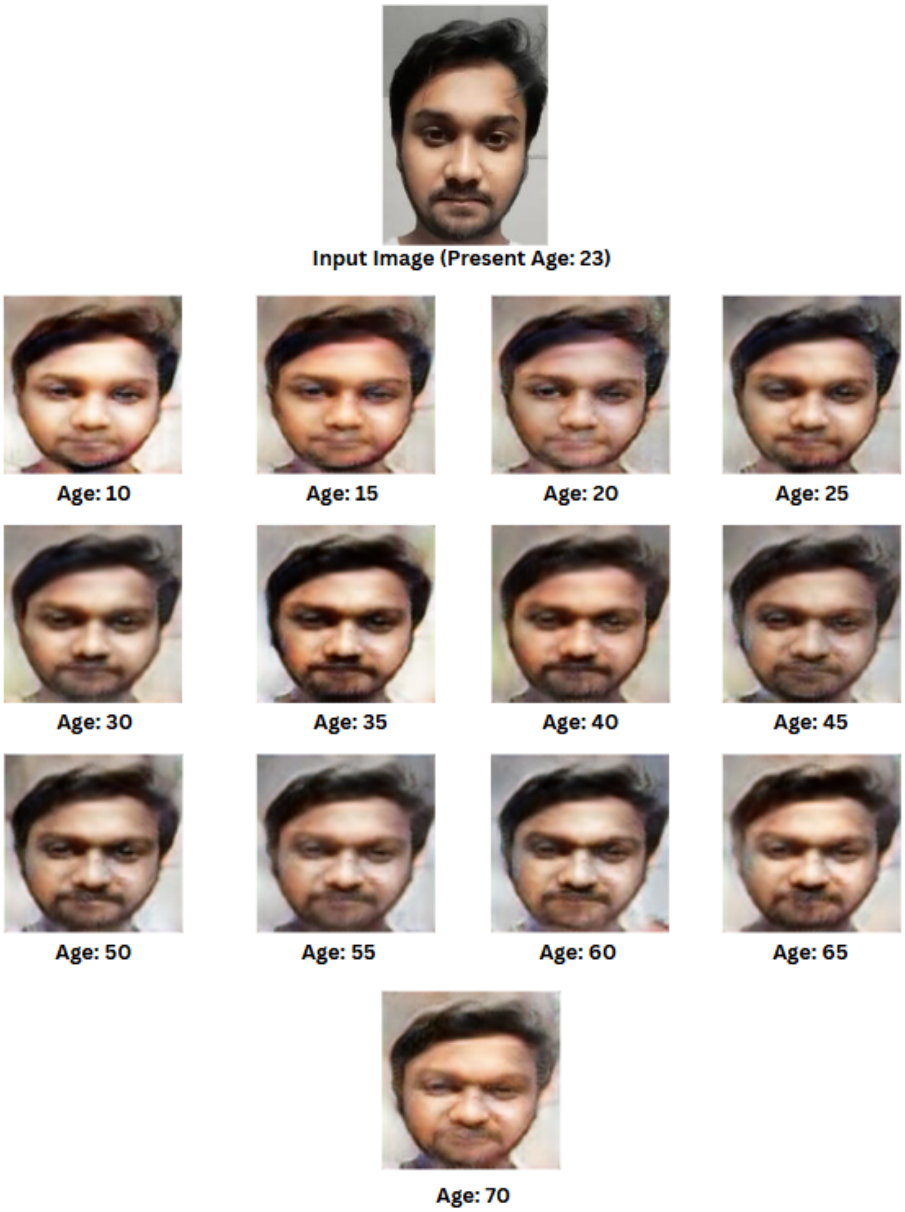


Figure 4.24: Face Aging Synthesis (Example: 1)

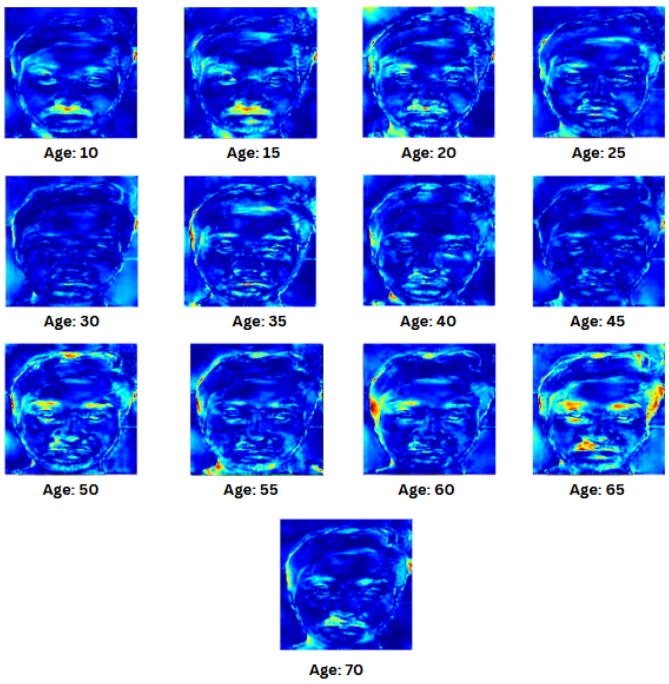


Figure 4.25: Face Aging Synthesis Heatmap (Example: 1)

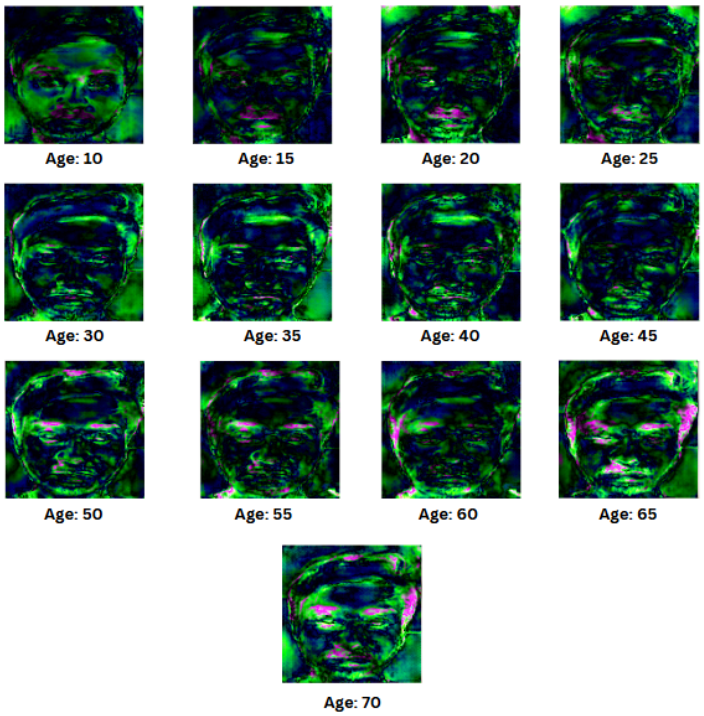


Figure 4.26: Face Aging Synthesis Drift Heatmap (Example: 1)

4.6.2 Example: 2

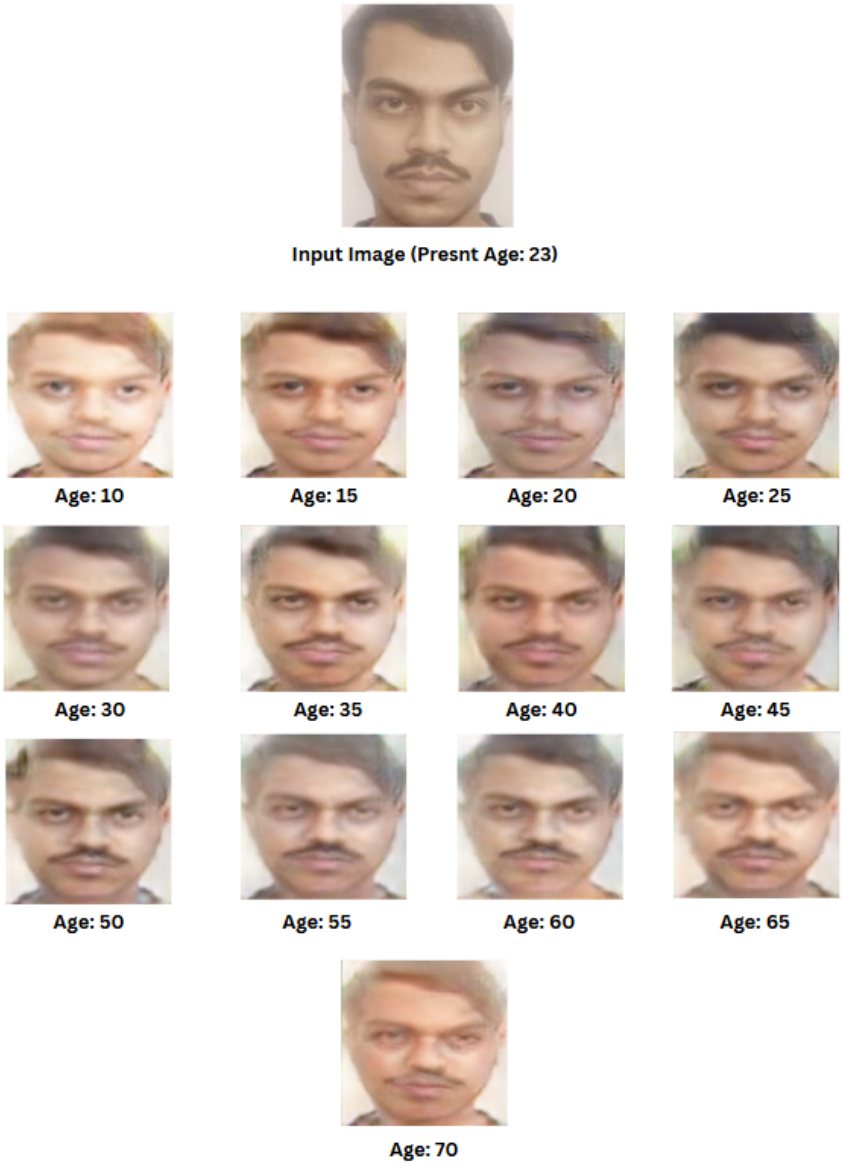


Figure 4.27: Face Aging Synthesis (Example: 2)

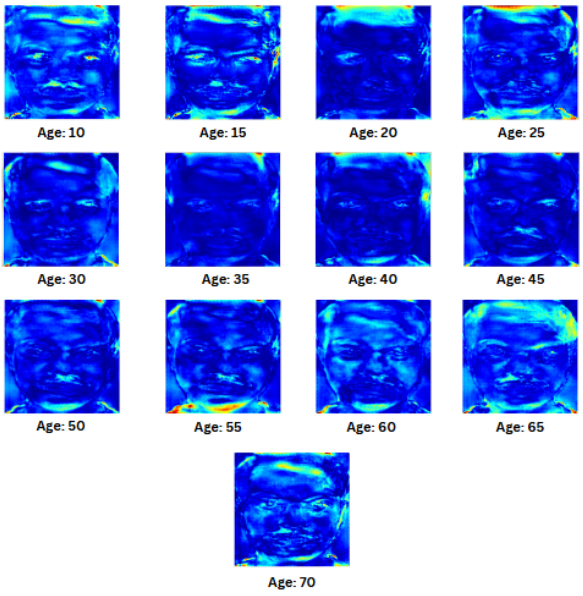


Figure 4.28: Face Aging Synthesis Heatmap (Example: 2)

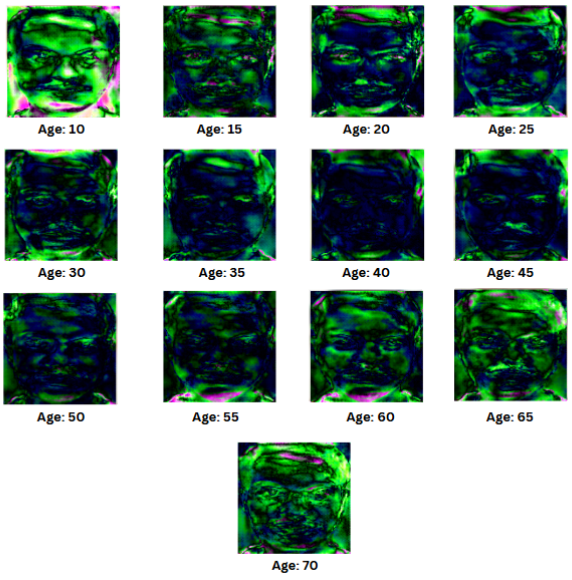


Figure 4.29: Face Aging Synthesis Drift Heatmap (Example: 2)

4.6.3 Comparison with Original Images

Here we showed a side-by-side comparison of the generated face images and our actual face at specific ages.

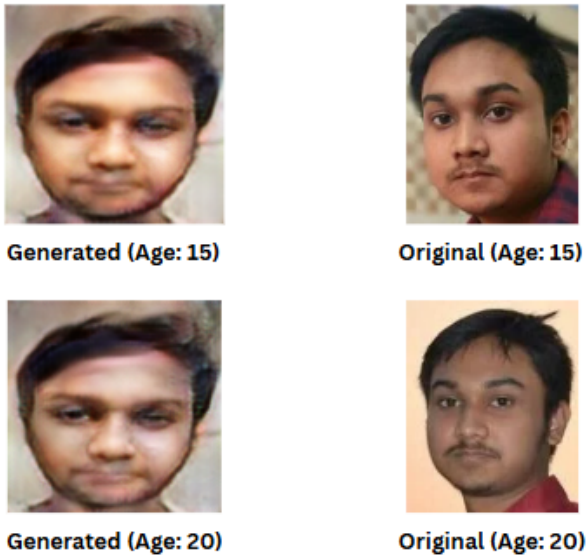


Figure 4.30: Comparison between Generated Image and Original face image

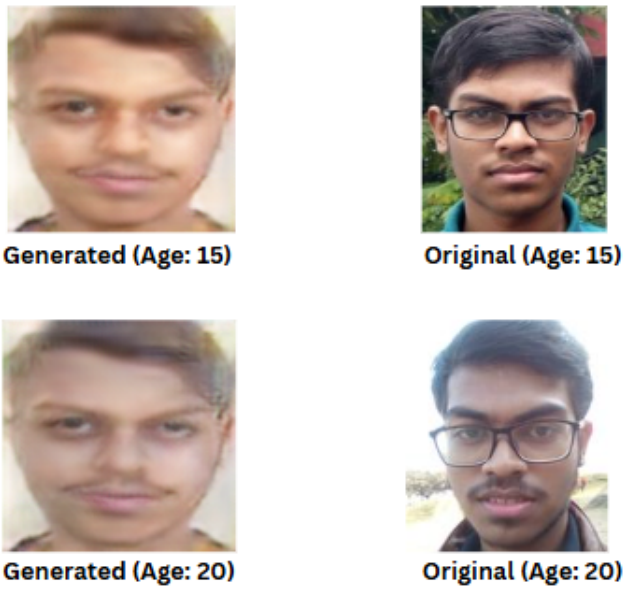


Figure 4.31: Comparison between Generated Image and Original face image

Chapter 5

Novelty and Innovation

5.1 Novelty of the Proposed Work

In our research, we found an innovative way to integrate Identity Drift Heatmap while face aging. Unlike conventional papers that focus mostly on generating age-progressed images, in our approach, we can visualize facial regions where identity-preservation deviates or changes while aging. This visualization helps us understand how the human face aging process changes and deviates across different age groups. At first, we explored and worked with six different GAN models and generated face-aged images. After that, we computed the FID and LPIPS scores of the models and found StarGAN to be a good contender for face age progression. In order to maximize the output of the StarGAN model, we tried to improve the architecture of the model. In both the training and testing phases, we introduced an innovative way to work with the Star GAN model. We introduced Hybrid Residual Attention Blocks in the training phase of the Generator. Traditional residual blocks are used in this case in many research studies, but we found a unique way to use the Residual Attention Blocks in the training phase. The traditional residual blocks are confined to learning identity mapping and feature propagation. But the Hybrid Residual Attention Block is augmented with the channel attention and the spatial attention mechanism. The Channel attention ensures the aging-related changes, such as wrinkles, eye bags, are prominently propagated. While the spatial attention is more focused on the important regions of the faces. So, this attention feature makes the model more prominent in generating more realistic face images, which is particularly helpful in forensic fields, as they require more accurate images to assist in correct identification. Again, in the testing phase, we introduced the Identity Drift Heatmap generation. As we mentioned earlier, this unique architecture offers visual representations with drift-based heatmap generation, which helps us understand the facial changes during the aging process. Also, in our dataset part, we merged both the UTK-Face and the MORPH-II datasets to create a unified dataset that provides face images across various age groups and ethnicities while maintaining the same image file format. This merged dataset is ideal for the face age synthesis task. Furthermore, we tried to improve the performance of the mentioned GAN models by modifying them beyond their traditional architecture. Therefore, our work introduced a novel way to generate aging face images while visualizing the changes with an identity drift heatmap.

5.2 Interpretability and Application Value

As Machine learning and AI studies are advancing at a rapid rate, the GAN models have potential usability for various real-world applications and various sectors. Nowadays, face age synthesis is extensively used in movies and digital platforms, where actors are required to make appearances of different age groups. In such cases, face age synthesis enables convincing prediction of aging faces. With the help of such predictions, actors can put on necessary makeup and other alterations to transform their faces into a realistic aging face. Which reduces the hassle of assuming how an age-progressed face should look. Additionally, Face age synthesis is used by law enforcement agencies to generate age-progressed images of suspects or missing persons to help in investigations [14]. Although face recognition systems are conventional in investigations, they often fail to identify individuals undergoing various changes as the time interval increases. We know that as the time between the original photo increases, the performance of face recognition decreases. As a result, Face age synthesis can produce better predicted images for which can be crucial for identification. Furthermore, the recent pandemic of COVID-19 has shown us the importance of a touchless biometric system to reduce virus spreading. In this case, the face recognition system can be a reliable solution. The implementation of face aging can provide us with a dynamic and intricate touchless biometric recognition system [22]. Moreover, it can be applied to health and aging-related research, where information about facial changes due to aging is required. Also, generated aging face images can make it possible to decrease the visits for updating an individual's photo in various service agencies, reducing the hassles. Moreover, we can specifically highlight regions where identity-preservation deviates or changes in our work, which is not discussed extensively in other research. The Identity Drift Heatmap will visualize identity changes during aging progression and help us understand more about how the human face aging process changes across different age groups. Thus, our proposed work will provide advantages in the advancement of forensic work, in the commercial industries and ethical AI usage.

Chapter 6

Limitations and Future Scopes

GANs make it possible to produce highly realistic age-progressed and age-regressed facial images while maintaining an individual's identical features because of their generative capabilities. In our work, we found a unique and innovative way to take advantage of the GAN capabilities. For the image generation part, an input face image is age-progressed from the age of 5 to the age of 115 with a five-year age interval. In the future, the models can be fine-grained to even lower age intervals and make the synthesis more realistic and user-controllable. We introduced Hybrid Residual Attention Blocks in the training phase of the Generator, and also introduced Identity Drift Heatmap for visualizing the changes. In the future, stronger identity-aware constraints can be implemented to ensure the outputs maintain high similarity with the original image. Additionally, GAN models can be expanded to consider emotion, health, or lifestyle factors and make aging synthesis more personalized and context-aware. Also, training GANs often requires powerful GPUs, large datasets, and many hours of training. With high demand for face synthesis, developing efficient and lightweight models will be crucial in the future. The selected UTK-Face dataset provides us with a vast amount of face images, but most images are really poor quality. Even after our manual preprocessing, many poor-quality images still exist, which can affect the training. Training models on such datasets produced repetitive or overly similar outputs in some cases in our research. Also, fewer images of older people were present in the dataset. In the future, merging more older face images in the dataset can produce better results. Additionally, our work only focused on 2D face aging synthesis, which is a limitation. Also, real-time face aging synthesis with heatmap reports can't be achieved by our work. Although our research didn't achieve the best results, more research on this model could lead to better outcomes. Possibilities for better results and usability are open in our work, and we believe that our work is beneficial in this study.

Chapter 7

Conclusion

Generative Adversarial Network is a groundbreaking and developing field of study that opens endless possibilities for us to research. GANs make it possible to produce highly realistic aging facial images while maintaining an individual's identical features because of their generative capabilities. As GANs can learn from vast amounts of datasets, we can achieve face synthesis of under-researched races and age groups. The technology has been instrumental in different applications, including entertainment industries, forensic investigations and social research. With the proposed method, we can not only produce aging face images from an input image but also visualize the changes during aging by showing the identity drift heatmap. Therefore, the generated results provide us with lots of practicality in face image-related works. Since GAN models are evolving and being studied further, the quality and scope of the technology will also improve substantially. Thus, possibilities for better work are open and our work can contribute to the advancement of this study.

Bibliography

[1] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," *Computer Vision – ECCV 2016*, pp. 694–711, 2016. DOI: 10.1007/978-3-319-46475-6_43. [Online]. Available: https://link.springer.com/chapter/10.1007%2F978-3-319-46475-6_43.

[2] S. Liu, Y. Sun, D. Zhu, *et al.*, "Face aging with contextual generative adversarial nets," Oct. 2017. DOI: 10.1145/3123266.3123431. [Online]. Available: https://doi.org/10.1145/3123266.3123431.

[3] H. Yang, D. Huang, Y. Wang, and A. K. Jain, "Learning face age progression: A pyramid architecture of gans," *arXiv (Cornell University)*, Jan. 2017. DOI: 10.48550/arxiv.1711.10352. [Online]. Available: https://www.researchgate.net/publication/321347784_Learning_Face_Age_Progression_A_Pyramid_Architecture_of_GANs.

[4] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017. DOI: 10.1109/iccv.2017.244. [Online]. Available: https://doi.org/10.1109/iccv.2017.244.

[5] X. Shu, J. Tang, Z. Li, H. Lai, L. Zhang, and S. Yan, "Personalized age progression with bi-level aging dictionary learning," vol. 40, pp. 905–917, Apr. 2018. DOI: 10.1109/tpami.2017.2705122. [Online]. Available: https://doi.org/10.1109/tpami.2017.2705122.

[6] A. Genovese, V. Piuri, and F. Scotti, "Towards explainable face aging with generative adversarial networks," Sep. 2019. DOI: 10.1109/icip.2019.8803616. (visited on 06/11/2023).

[7] H.-Y. Lee, H.-Y. Tseng, Q. Mao, *et al.*, *Drit++: Diverse image-to-image translation via disentangled representations*, arXiv.org, 2019. [Online]. Available: https://arxiv.org/abs/1905.01270 (visited on 06/15/2025).

[8] X. Liu, Y. Zou, C. Xie, H. Kuang, and X. Ma, "Bidirectional face aging synthesis based on improved deep convolutional generative adversarial networks," *Information*, vol. 10, p. 69, Feb. 2019. DOI: 10.3390/info10020069. [Online]. Available: https://doi.org/10.3390/info10020069.

[9] N. Sharma, R. Sharma, and N. Jindal, "An improved technique for face age progression and enhanced super-resolution with generative adversarial networks," *Wireless Personal Communications*, May 2020. DOI: 10.1007/s11277-020-07473-1. [Online]. Available: https://doi.org/10.1007/s11277-020-07473-1.

[10] H. Wang, "Age-related facial analysis with deep learning," 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:247321086.

[11] C. Yang and Z. Lv, "Gender based face aging with cycle-consistent adversarial networks," *Image and Vision Computing*, vol. 100, p. 103 945, Aug. 2020. DOI: 10.1016/j.imavis.2020.103945. [Online]. Available: https://doi.org/10.1016/j.imavis.2020.103945.

[12] H. Zhu, Z. Huang, H. Shan, and J. Zhang, "Look globally, age locally: Face aging with an attention mechanism," *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1963–1967, Apr. 2020. DOI: 10.1109/icassp40776.2020.9054553. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9054553 (visited on 06/15/2025).

[13] D. Anghelone, C. Chen, P. Faure, A. Ross, and A. Dantcheva, "Explainable thermal to visible face recognition using latent-guided generative adversarial network," *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, Dec. 2021. DOI: 10.1109/fg52635.2021.9667018. (visited on 03/10/2024).

[14] M. Grimmer, R. Ramachandra, and C. Busch, "Deep face age progression: A survey," *IEEE Access*, vol. 9, pp. 83 376–83 393, Jan. 2021. DOI: 10.1109/access.2021.3085835. [Online]. Available: https://doi.org/10.1109/access.2021.3085835.

[15] G.-S. Hsu, R.-C. Xie, and Z.-T. Chen, "Wasserstein divergence gan with cross-age identity expert and attribute retainer for facial age transformation," *IEEE Access*, vol. 9, pp. 39 695–39 706, 2021. DOI: 10.1109/access.2021.3062499. (visited on 03/19/2022).

[16] Z. Huang, S. Chen, J. Zhang, and H. Shan, "Pfa-gan: Progressive face aging with generative adversarial network," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 2031–2045, Jan. 2021. DOI: 10.1109/tifs.2020.3047753. (visited on 01/31/2024).

[17] E. Pantraki and C. Kotropoulos, "Face aging using global and pyramid generative adversarial networks," *Machine Vision and Applications*, vol. 32, May 2021. DOI: 10.1007/s00138-021-01207-4. [Online]. Available: https://doi.org/10.1007/s00138-021-01207-4.

[18] N. Sharma, R. Sharma, and N. Jindal, "Prediction of face age progression with generative adversarial networks," *Multimedia Tools and Applications*, Aug. 2021. DOI: 10.1007/s11042-021-11252-w.

[19] P. K. Chandaliya and N. Nain, "Aw-gan: Face aging and rejuvenation using attention with wavelet gan," *Neural Computing and Applications*, vol. 35, pp. 2811–2825, Sep. 2022. DOI: 10.1007/s00521-022-07721-4. [Online]. Available: https://doi.org/10.1007/s00521-022-07721-4.

[20] M. Mahiuddin, M. Khaliluzzaman, S. Azad, and M. N. Arefin, "Fake face generator: Generating fake human faces using gan," *International Journal of Advanced Computer Science and Applications*, vol. 13, Jan. 2022. DOI: 10.14569/ijacsa.2022.0130721.

[21] N. Sharma, R. Sharma, and N. Jindal, "Comparative analysis of cyclegan and attentiongan on face aging application," *Sādhanā*, vol. 47, Feb. 2022. DOI: 10.1007/s12046-022-01807-4.

[22] D.-C. Wang, Z.-J. Tsai, C.-C. Chen, and G.-J. Horng, "Development of a face prediction system for missing children in a smart city safety network," *Electronics*, vol. 11, p. 1440, Apr. 2022. DOI: 10.3390/electronics11091440. [Online]. Available: https://doi.org/10.3390/electronics11091440.

[23] K. Ko, T. Yeom, and M. Lee, "Superstargan: Generative adversarial networks for image-to-image translation in large-scale domains," *Neural Networks*, Mar. 2023. DOI: 10.1016/j.neunet.2023.02.042. (visited on 03/19/2023).

[24] C. Korgialas, E. Pantraki, A. Bolari, M. Sotiroudi, and C. Kotropoulos, "Face aging by explainable conditional adversarial autoencoders," *Journal of Imaging*, vol. 9, p. 96, May 2023. DOI: 10.3390/jimaging9050096. [Online]. Available: https://mdpi-res.com/jimaging/jimaging-09-00096/article_deploy/jimaging-09-00096-v2.pdf?version=1683803834 (visited on 02/13/2024).

[25] C. S. Panuganti, *Morph-2*, Kaggle.com, 2023. [Online]. Available: https://www.kaggle.com/datasets/chiragsaipanuganti/morph (visited on 06/11/2025).

[26] M. A. Iqbal, W. Jadoon, and S. K. Kim, "Synthetic image generation using conditional gan-provided single-sample face image," *Applied Sciences*, vol. 14, pp. 5049–5049, Jun. 2024. DOI: 10.3390/app14125049. [Online]. Available: https://www.mdpi.com/2076-3417/14/12/5049.

[27] Y. Liu, Q. Li, Z. Sun, and T. Tan, *A3gan: An attribute-aware attentive generative adversarial network for face aging*, arXiv.org, 2024. [Online]. Available: https://arxiv.org/abs/1911.06531.

[28] *Face aging with identity-preserved conditional generative adversarial networks — ieee conference publication — ieee xplore*, ieeexplore.ieee.org. [Online]. Available: https://ieeexplore.ieee.org/document/8578926.

[29] *Utkface*, www.kaggle.com. [Online]. Available: https://www.kaggle.com/datasets/jangedoo/utkface-new.