Undergraduate Final Project

Improving Kinship Verification using Face Age Transformation

Matheus Levi Rodrigues Aidano

advised by
Prof. Dr. Tiago Figueiredo Vieira

Universidade Federal de Alagoas
Institute of Computing
Maceió, Alagoas
December 13th, 2024

UNIVERSIDADE FEDERAL DE ALAGOAS
Institute of Computing

IMPROVING KINSHIP VERIFICATION USING FACE AGE TRANSFORMATION

Undergraduate Final Project submitted to the Institute of Computing at the Universidade Federal de Alagoas as a partial requirement for obtaining the degree of Computer Science.

Matheus Levi Rodrigues Aidano

Advisor: Prof. Dr. Tiago Figueiredo Vieira

Examining Board:

Bruno Georgevich Ferreira	Prof. Dr., UFAL
Warley Vital Barbosa	Prof. Dr., UFAL

Maceió, Alagoas
December 13th, 2024

Catalogação na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecária: Girlaine da Silva Santos – CRB-4 – 1127

A288I Aidano, Matheus Levi Rodrigues.

Improving kinship verification using face age transformation / Matheus Levi Rodrigues Aidano. – 2025.

48 f.: il. color.

Orientador: Tiago Figueiredo Vieira.

Monografia (Trabalho de Conclusão de Curso em Computação) - Universidade Federal de Alagoas, Instituto de Computação. Maceió, 2025.

Bibliografia: f. 40- 48.

1. Reconhecimento facial (Computação). 2. Visão Computacional. 3. Redes neurais (Computação). 4. Redes generativas adversariais. 5. Parentesco - Percepção facial. I. Título.

CDU: 004.89

,

UNIVERSIDADE FEDERAL DE ALAGOAS/UFAL

Instituto de Computação – IC

Campus A. C. Simões - Av. Lourival de Melo Mota, BL 12
Tabuleiro do Martins, Maceió/AL - CEP: 57.072-970
Telefone: (082) 3214-1401

Trabalho de Conclusão de Curso - TCC

Formulário de Avaliação

Nome do Aluno

M	A	T	H	E	U	S		L	E	V	I							
R	O	D	R	I	G	U	E	S		A	Y	D	A	N	O			

Nº de Matrícula

					1	9	2	1	1	6	0	-	3

Título do TCC (Tema)

Improving Kinship Verification using Face Age Transformation

Banca Examinadora

TIAGO FIGUEIREDO VIEIRA
Nome do Orientador

Documento assinado digitalmente
gov.br
TIAGO FIGUEIREDO VIEIRA
Data: 17/12/2024 15:18:54-0300
Verifique em https://validar.iti.gov.br

BRUNO GEORGEVICH FERREIRA
Nome do Professor

Documento assinado digitalmente
gov.br
BRUNO GEORGEVICH FERREIRA
Data: 19/12/2024 13:54:38-0300
Verifique em https://validar.iti.gov.br

WARLEY VITAL BARBOSA
Nome do Professor

Documento assinado digitalmente
gov.br
WARLEY VITAL BARBOSA
Data: 17/12/2024 15:22:53-0300
Verifique em https://validar.iti.gov.br

Data da Defesa	Nota Obtida
13/12/2024	10,00 (DEZ)

Observações

Coordenador do Curso
De Acordo

Assinatura

Acknowledgments

First, I would like to express my gratitude to my parents for their effort and dedication - they walked so that I could run. I extend my thanks to my friends, who were always by my side, for their companionship and understanding throughout the entire period I devoted to this work. I am also deeply grateful to the professors at the Institute for the opportunities and teachings that enabled a fulfilling journey in my professional career throughout the course, especially Professor Tiago Vieira, who kindly accepted the role of my advisor and provided all the support and guidance necessary for the completion of this work.

My heartfelt thanks go to all the members of the SOIOS research group for their valuable teachings, productive discussions, and the good friendship I have found since joining. Without their contributions, this work would not have been possible. Special thanks to Warley, who personally guided and assisted me in the development of my activities. I am also grateful to all the places where I had the opportunity to work, with a special thanks to the EDGE Innovation Center for providing the necessary infrastructure to train the models used in this research.

Finally, I thank the members of the examination committee for their time and constructive feedback provided.

November 26st, 2024, Maceió - AL

Resumo

Verificação facial de parentesco, a tarefa de determinar relações familiares com base em imagens faciais, tem ganhado atenção significativa nos últimos anos devido às suas aplicações em áreas como mídias sociais, forense e genealogia. No entanto, verificar parentesco com precisão continua sendo um problema desafiador, especialmente ao se considerar as variações de idade entre os membros da família. Esta dissertação explora o potencial de usar técnicas de Transformação de Idade Facial, especificamente através de *Generative Adversarial Networks* (GANs), para melhorar a precisão dos modelos de verificação de parentesco. Este trabalho envolve o desenvolvimento de um modelo de transformação de idade baseado em GAN que pode simular o processo de envelhecimento em imagens faciais. Ao aumentar as bases de dados de parentes com essas imagens transformadas pela idade, buscamos aprimorar a robustez e a confiabilidade dos sistemas de verificação de parentesco. Os resultados experimentais indicam que a incorporação de imagens faciais transformadas pela idade no processo de verificação de parentesco leva a uma representação mais precisa das relações familiares, especialmente em casos onde as diferenças de idade são acentuadas. Este trabalho contribui para o campo emergente da verificação de parentesco ao elaborar uma abordagem inovadora que aproveita o poder das GANs para progressão de idade, oferecendo uma direção promissora para pesquisas futuras e aplicações práticas.

***Keywords*: *Transformação de Idade Facial*, *Reconhecimento Facial de Parentesco*, *Redes generativas adversariais*, *Redes neurais convolucionais*, *Visão Computacional* .**

Abstract

Kinship verification, the task of determining family relationships based on facial images, has gained significant attention in recent years due to its applications in areas such as social networks, forensics, and genealogy. However, accurately verifying kinship remains a challenging problem, particularly when accounting for variations in age between family members. This thesis explores the potential of using Face Age Transformation techniques, specifically through Generative Adversarial Networks (GANs), to improve the accuracy of kinship verification models. This work involves developing a GAN-based age transformation model that can simulate the aging process in facial images. By augmenting kinship datasets with these age-transformed images, we aim to enhance the robustness and reliability of kinship verification systems. Experimental results indicate that incorporating age-transformed facial images into the kinship verification process leads to a more accurate representation of familial relationships, particularly in cases where age differences are pronounced. This work contributes to the growing field of kinship verification by elaborating a novel approach that leverages the power of GANs for age progression, offering a promising direction for future research and practical applications.

***Keywords*: *Face age transformation*, *Kinship Recognition*, *Generative Adversarial Network*, *Convolutional Neural Network*, *Computer Vision* .**

List of Figures

List of Symbols

α_0 Original Image Age

α_1 Target Image Age

$|B|$ Batch size

σ Activation Function

L Loss function.

W Image width.

τ Temperature.

List of Abbreviations

AI Artificial Intelligence

DNA Deoxyribonucleic acid

CNN Convolutional Neural Network

GAN Generative Adversarial Network

FKV Facial Kinship Verification

FS Father-Son

FD Father-Daughter

MS Mother-Son

MD Mother-Daughter

BB Brother-Brother

SS Sister-Sister

GFGS Grandfather-Grandson

GFGD Grandfather-Granddaughter

GMGS Grandmother-Grandson

GMGD Grandmother-Granddaughter

RFIW Recognizing Families in the Wild

FIW Families in the Wild

IoT Internet of Things

LBP Local Binary Patterns

MSE Mean Squared Error

Summary

Chapter 1

Introduction

The face of an individual carries important and unique characteristics of human identification, as shown by the recent success of facial recognition systems. Genetically, these traits are determined based on the DNA of the parent, which carries perceptive similarities, such that humans can tell kinship relationships based on facial similarity [Kaminski et al. (2009)]. Given the importance of these facial cues, various methods have been extensively investigated to verify kinship based on facial images in the fields of computer vision and biometrics [Wu et al. (2022)]. These methods are used to determine the presence of a kinship relationship between two facial images, with applications ranging from the location of missing family members and the analysis of social networks for use in fields such as genealogy.

In the early stages of kinship verification research, feature extractors such as Histogram of Oriented Gradients [Dalal and Triggs (2005)], and Local Binary Patterns [Huang et al. (2011)] were commonly used to extract genetic features (e.g., skin color, eye color) from facial images. However, these approaches often suffered from poor accuracy due to challenges such as varying face angles, lighting conditions, low image resolution, and the presence of facial accessories.

In recent years, Convolutional Neural Networks (CNNs) have demonstrated strong performance in computer vision tasks and have been successfully applied to a variety of face-related tasks, including face recognition. Since then, the main focus of kinship verification models has been deep learning-based techniques, and several high performance kinship verification methods have been proposed [Robinson et al. (2018), Jain et al. (2020), Zhang et al. (2021a)]. However, kinship verification remains a formidable challenge due to critical issues such as the scarcity of labeled data and the bias present in kinship images, such as age and gender differences, which complicate model training. To address these challenges, this thesis proposes the use of a face age transformation model to generate facial images representing various age groups. By augmenting existing kinship datasets with these age-transformed images, a cross-age kinship verification model can be constructed, allowing for the enhancement of insufficient labeled data and enabling more

robust learning of genetic characteristics across different age groups.

The proposed approach involves the construction of a Generative Adversarial Network (GAN)-based age transformation model [Creswell et al. (2018)], which consists of an age encoder and an age classifier. The age encoder is designed to encode target age information into a latent vector, which allows the decoder to generate a facial image of the specified age. Although most GAN-based aging models rely on a conditional discriminator, applying such a condition can negatively impact the discriminator's ability to differentiate between real and fake images, its primary function. To produce more realistic output images, this work employs an age encoder instead of a conditional discriminator. Additionally, to ensure that the synthesized images maintain the identity of the original input image, which is an essential aspect of kinship verification, an identity preservation module is integrated into the face age transformation model. The generated face images, representing a range of ages, are then paired with a kinship dataset and used to train the kinship verification model. Extensive experiments demonstrate that the proposed face age transformation model can generate high-quality facial images across different age groups and that the kinship verification model constructed using these images achieves better performance.

1.1 Motivation

The ability to verify kinship based on facial images has far-reaching implications across various domains, including social media, forensics, genealogy, and law enforcement. In scenarios such as reuniting lost family members or validating family claims, accurate kinship verification can be crucial. However, this task is inherently challenging due to the complexities associated with human faces, particularly the variations that occur due to aging. Traditional kinship verification methods often struggle to account for these age-related changes, leading to diminished accuracy and reliability.

Aging alters facial features in ways that can obscure the genetic similarities shared by family members. As a result, models that fail to account for these transformations may produce inconsistent or inaccurate results. This issue is further compounded by the scarcity of labeled datasets that span a wide range of ages, making it difficult to train models that can generalize effectively across different age groups. Moreover, the inherent variability in age differences within kinship relationships complicates this task; for instance, siblings typically have smaller age gaps than parents and children, yet it is not uncommon for siblings to have significant age differences. This variability makes it difficult to use age differences as a factor in kinship verification, underscoring the need for more sophisticated approaches that can accommodate these complexities.

In recent years, advances in deep learning, particularly Generative Adversarial Networks (GANs), have opened new avenues for addressing these challenges. GANs have shown remarkable capabilities to generate realistic images, including age progression and

facial regression. This presents a unique opportunity to enhance kinship verification models by incorporating age-transformed facial images, thereby bridging the gap between age-disparate kinship pairs.

The motivation for this research stems from the desire to improve the accuracy and robustness of kinship verification systems by leveraging GAN-based age transformation techniques. By generating facial images that represent various stages of aging, we can create more comprehensive datasets and develop models that are better equipped to handle the intricacies of age-related changes. This approach not only addresses a critical gap in current kinship verification methods but also contributes to the broader field of biometrics by introducing novel techniques for facial analysis across age groups.

The ultimate goal of this research is to provide a more reliable and effective tool for kinship verification, with potential applications in real-world scenarios where understanding familial relationships is essential. By developing a path towards age-invariant models, this work aims to pave the way for future research and practical implementations that can benefit society in meaningful ways.

1.2 Objectives

1.2.1 General Objectives

The primary objective of this research is to enhance the accuracy and robustness of kinship verification models by integrating face age transformation techniques, while studying the effects of using synthetic data to compensate for aging effects. Specifically, the research aims to achieve the following:

1.2.2 Specific Objectives

1. Design and implement a Generative Adversarial Network (GAN) capable of generating realistic facial images across different age groups while preserving the identity of the individuals. This model will simulate the aging process to produce age-progressed and age-regressed facial images.

2. Utilize the GAN-based age transformation model to augment existing kinship datasets by generating additional facial images that span a wide range of ages. This augmentation aims to address the scarcity of labeled data, particularly for age-diverse kinship pairs, thereby enhancing the diversity and richness of the training data.

3. Develop a kinship verification model that can effectively utilize the augmented datasets, leveraging the diversity introduced by the age-transformed images. This

model will be fine-tuned to recognize kinship relations across different age groups, ensuring that it can generalize well to various scenarios involving age differences, using state-of-the-art techniques.

4. Conduct extensive experiments using benchmark kinship verification datasets to evaluate the effectiveness of the proposed age transformation and verification models. The goal is to demonstrate that the inclusion of age-transformed images leads to improvements in kinship verification accuracy compared to traditional methods.

1.3 Work Organization

This thesis is organized into six chapters, each addressing a critical aspect of the research on improving kinship verification using face age transformation techniques. The structure is designed to guide the reader through the background, methodology, experiments, and findings in a logical and coherent manner.

In Chapter 1, the research topic is introduced, highlighting the motivation behind the study, the problem statement, and the specific objectives of the research. It also provides a brief overview of the challenges in kinship verification and sets the stage for the subsequent chapters.

Chapter 2 presents a literature review, an in-depth analysis of existing research in the fields of kinship verification, facial aging, and generative models. This chapter covers the evolution of kinship verification techniques, the role of facial features in genetic similarity, and the application of generative adversary networks (GANs) in age transformation. It provides the necessary background and highlights the gaps that this work aims to address.

Next, Chapter 3 develops the theoretical bases of the research, including the principles of deep learning, facial aging, the architecture of GANs, and the challenges associated with kinship verification across different age groups. It establishes the conceptual foundation for the proposed methodology.

The methodology is detailed in Chapter 4, where the design and implementation of the GAN-based age transformation model and the kinship verification model are presented. It includes data pre-processing steps, the GAN architecture, the age transformation process, and the integration of transformed images into the kinship verification system.

The results are shown in Chapter 5, where they are presented and analyzed. The chapter provides both quantitative and qualitative evaluations of the models, comparing the performance of the proposed method with existing approaches. The discussion includes an analysis of the improvements achieved through the use of age-transformed images and the limitations of the approach.

Finally, the conclusion summarizes the key findings of the research and highlights the contributions made to the field of kinship verification and biometrics. The work also

discusses the broader implications of the work and suggests directions for future research, including potential improvements to the methodology and applications in related areas.

Chapter 2

Literature Review

This chapter reviews key developments in kinship verification and face age transformation, with a focus on how these techniques have evolved and how they intersect to address the problem of age differences in familial relationships. By examining the existing literature, this review highlights the current state-of-the-art, identifies challenges that persist in the field, and sets the foundation for the proposed research, which seeks to integrate face age transformation into kinship verification systems to improve accuracy and reliability.

2.1 Kinship Verification Overview

Kinship verification, or more specifically, facial kinship verification (FKV), is a task that aims to determine if two individuals have a kin relationship or not, based on their faces, using either images or videos. The most common categories of kinship relationships are: Father-Son (FS), Father-Daughter (FD), Mother-Son (MS), and Mother-Daughter (MD). As the familial relationship becomes more distant, the prediction of kinship becomes increasingly challenging. However, certain databases and methods have been developed to include more distant relations, such as Grandfather-Granddaughter (GFGD) and Grandmother-Grandson (GMGS) pairs, but they cannot reach the same level of accuracy, mostly due to lack of training data. In the last two decades, Kinship Verification has been attracting increasing attention and in 2014 had its first competition [Lu et al. (2014a)], which aimed to evaluate different kinship verification algorithms with three possible experimental protocols: unsupervised, image restricted, and image unrestricted. Another competition worth mentioning is Recognizing Families in the Wild (RFIW) [Robinson et al. (2020)] which has several editions and became the most important competition in the field, using Families in the Wild (FIW) [Robinson et al. (2016)] database as its benchmark.

There are a few reasons that can explain the increase in kinship verification interest Wu et al. (2022). The first is due to its various potential applications: In the anthropology and genetics domain, FKV can help to study the hereditary characteristics of close relatives

in social relationships [A. (2020)]. In the field of public social security, it can be applied to the search for missing children, border control, customs, and criminal investigations [Kohli et al. (2019), Lu et al. (2012b)]. In the social media domain, FKV can be used for the organization of family photo albums and to improve the performance of face recognition systems and social media analysis [Lu et al. (2014a)]. In addition, FKV also has potential applications in smart homes, the Internet of Things (IoT) [Jang et al. (2017)] and personalized software. The second reason is that FKV serves as a fundamental study among visual kinship problems, such as family recognition and family retrieval [Robinson et al. (2018)]. Lastly, the low sensory perception of human eyes to quantify the similarity of two images from different people [Bordallo Lopez et al. (2018)]. Features such as the distance between the eyes and the shape, color, and size of facial parts are not easily judged at a glance, resulting in low recognition accuracy.

One of the directions that the field of Kinship Recognition is following is also adding temporal information, using video-based datasets, which showed promising results [Kohli et al. (2019)], but currently there are only a few good video datasets and, in some cases, they can make the problem harder.

2.1.1 Problem Definition

Given a pair of facial images, the objective of kinship verification is to judge whether two people are biologically related (with a typical kin relationship). It is assumed that these two facial images do not belong to the same individual, since most of the work in the area ignores the self-kinship relationship. Specifically, current kinship verification research only focuses on close family relationships, which can be categorized into three levels of generation, e.g. siblings, parent-child, and grandfather-grandchild.

Therefore, kinship verification can be formulated as a binary classification problem (kin vs. non-kin). FKV deep learning models usually work with a Siamese Neural Network architecture [Bromley et al. (1994)], which can extract useful features from both images using the same process to properly compare them. Formally, as shown in Fig. 2.1, given a pair of faces (X,Y), appropriate feature representations $(\phi(X), \phi(Y))$ are extracted from both images, and then a classifier is used to determine if the two faces have a kin relationship or not, which is normally a form of similarity measurement, such as cosine similarity.

In addition to kinship verification, there is also the classification task to find the exact kin relation of two individuals. In the RFIW challenge [Robinson et al. (2020)], three main tasks are defined: classification, tri-subject verification and search & retrieval. In tri-subject verification, the input consists of three images, one image of a child, and two images of potential parents of that child, and the goal is to verify whether there is a parental kin relation between the parents and the child, which can be applied directly to

find missing children. Finally, search and retrieval attempts to find family members from a large gallery based on a given query image. The system compares the query image with others in the database to identify those that share a familial relationship, like parents, siblings, etc. By ranking images based on similarity to the query, the task helps connect and identify related individuals, which makes it useful in applications such as family identification, missing person searches, and organizing family photos on social media.

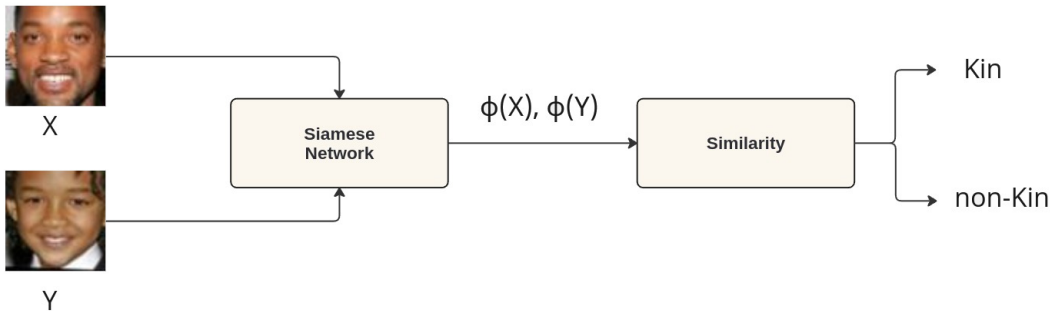

Figure 2.1: General pipeline of Siamese Networks in Kinship Verification

2.1.2 Main Challenges

As defined in the section above, Facial Kinship Verification is a binary classification problem which is harder than face recognition, since kinship pairs do not have the same identity, but only share hidden genetic features, subjected to biological, age and gender variations. In fact, kinship judgment and facial similarity are highly correlated, but not strictly synonymous, which showcases the problem difficulty [DeBruine et al. (2009)]. These challenges are called intrinsic, as they arise from the inherent nature of the task itself. Furthermore, extrinsic challenges involve changes in illumination, camera viewpoint, and face occlusion, for example.

Being a harder problem than face recognition, it is expected that datasets would be larger, but unfortunately that is not the case; they tend to be much smaller in size. In recent years, the number of video-datasets is rising, containing facial expression, head motion and mouth movement, which may increase the accuracy and robustness of kinship verification algorithms, diminishing both the problems of intrinsic and extrinsic difficulties, providing more genetic information in the expressions and more variation on the image conditions.

Small interclass variations are another troublesome problem in FKV, some positive examples may have small facial similarities, whereas negative examples may have high facial similarities. Therefore, small positive and negative variations decrease the separation

between classes and pose significant challenges in learning the real decision boundary. In addition, there is a serious imbalance issue [Li et al. (2021)], evidently the number of negative pairs is significantly more than the number of positive pairs. For this reason, to actually represent the data distribution of families worldwide, a lot of data is required, which is hard to gather because of security and privacy issues, delaying the development of the kinship recognition field.

2.1.3 Datasets

Based on the number of kinship types, existing datasets can be divided into three categories: 4-types, 7-types, and 11-types (nonkin is not considered). The development of public kinship datasets is shown in Fig. 2.2. The first thing to note is the trend of video datasets that started in 2018, with some datasets that carry not only visual face information but also audio, for example. This multisensorial approach might be the key to the next performance breakthrough in the field, but its development is still in early stages. The number of images in most datasets is usually less than 1000, reinforcing the difficulties in data collection mentioned above.

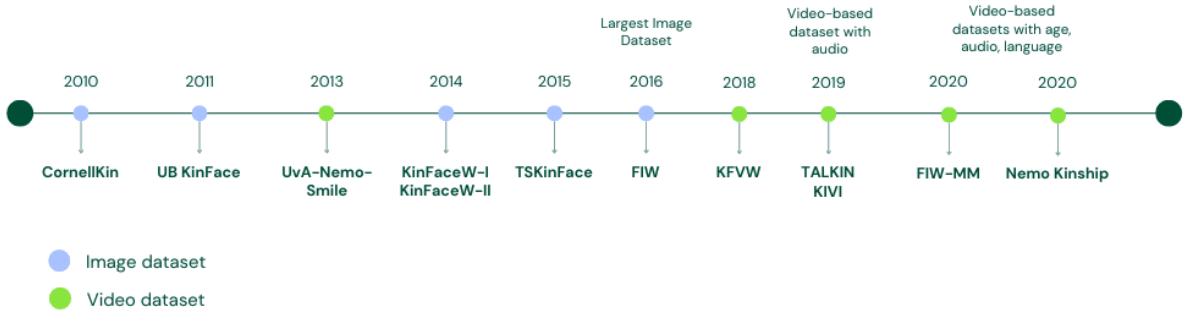

Figure 2.2: Development of representative kinship datasets. Source: [Wang et al. (2023)]

CornellKin [Fang et al. (2010)] is the first kinship dataset that was widely used, where images were collected using a controlled online search, limiting the pose to frontal and neutral facial expressions only. It has 150 pairs of celebrities with family information and four categories: Father-Son (F-S, 40%), Father-Daughter (FD, 22%), Mother-Son (MS, 13%), Mother-Daughter (MD, 26%). In the year following CornellKin, there is UB KinFace [Xia et al. (2011)] (2011), containing three images for each positive set with 270 images collected in total and divided into 90 groups. It is the first database with children, young parents, and old parents collected together. The main issue with this dataset is the high imbalance it carries, with around 80% of the data belonging to father-son relationships.

KinFaceW-I [Lu et al. (2012a)] and KinFaceW-II [Lu et al. (2014b)] are two very important kinship datasets. The main difference between the two is that KinFaceW-

I images are collected from different pictures, i.e., the family members are in different scenarios, illumination, etc. KinFaceW-II consists of pairs obtained from the same photo, usually a big photograph with several family members, with their faces cropped. These photos are unconstrained in terms of pose, lighting, background, expression, age, ethnicity, and partial occlusion and were obtained through an online search. In total, there are 1533 pairs of images in the two datasets combined.

Families in the Wild (FIW) [Robinson et al. (2016), Robinson et al. (2018)] is the largest available image dataset and a reference benchmark in the field. More than 13,000 family photos of 1000 families are labeled and collected from the Internet, making it the largest dataset for a large margin, and also has 11 kinship types. There are 10 images for each family on average. State-of-the-Art methods commonly have this dataset as the main benchmark, besides KinFaceW. It is also very popular for Family Recognition tasks. In 2020, the same authors made FIW with MultiMedia (FIW-MM) [Robinson et al. (2021)], extending FIW with an automated labeling pipeline adding video, audio and text captions.

UvA-NEMO Smile is the first video dataset and it contains 1240 videos of 400 subjects with a resolution of *1920x1080* at 50 fps rate. The dynamics of spontaneous and posed smiles of each subject are recorded. All videos are constrained from an angle and background perspective. It contains seven types of kinship relationships, and it is an important database for studying the impact of facial expressions, such as smile, on kinship feature inheritance. Since then, several other video databases have been collected, such as TALKIN [Wu et al. (2019)], a database collected from YouTube with visual and audio information from celebrities and family TV shows.

In summary, image-based kinship datasets have been well developed for image-based kinship verification. In contrast, there is still a demand for video-based kinship datasets. In addition, most of the datasets are collected in unconstrained settings, which causes many external interference factors and makes it difficult to study kinship verification systematically.

2.1.4 Existing Methods

Kinship verification has evolved significantly over the years, and various methods have been proposed to address the inherent challenges. These methods range from traditional feature extraction techniques to more advanced deep learning approaches. Existing methods can be defined in three main categories: Handcrafted feature descriptors, metric learning based and deep learning methods. This section presents an overview of the key existing methods used for kinship verification.

Early approaches to kinship verification relied heavily on handcrafted features to extract important facial traits. Methods such as Histogram of Oriented Gradients (HOG)

[Dalal and Triggs (2005)], Scale-Invariant Feature Transform (SIFT) [Lowe (2004)], and Local Binary Patterns (LBP) [Ahonen (2004)]were commonly used to represent facial characteristics. These methods aimed to capture low-level features, such as texture, edges, and contours, which could be indicative of genetic similarities between individuals. The first kinship method using this technique was proposed by Fang et al. (2010), using 22 hand-crafted facial features such as color, face geometry, and texture. Zhou et al. (2012) proposed a Gabor [Adini et al. (1997)] wavelet-based gradient orientation pyramid for kinship verification. To represent such complex representations as kinship features, low-level feature descriptors are not enough, better accuracy was achieved by combining different feature detectors [Alirezazadeh et al. (2015)], but the results are still behind current State-of-the-Art methods.

With the limitations of handcrafted features, metric learning methods became popular for kinship verification. These methods aim to learn a similarity function between pairs of images. Cosine similarity is commonly used in these methods to quantify the degree to which two facial images are similar in a learned feature space. The goal of metric learning is basically to decrease the intraclass distance and increase the interclass distance of the facial features. This is achieved by learning a distance metric to measure the similarity between facial images. In addition to cosine similarity, the Mahalanobis distance is often used because it improves upon the traditional Euclidean distance by taking into account correlations between the data points and the variance within the dataset. The Mahalanobis distance can be defined as:

$$d_M(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T M (\mathbf{x} - \mathbf{y})} \tag{2.1}$$

where x and y are feature vectors, in our case, learned facial features of two images, M is a positive semidefinite matrix which defines the space where distances are computed. The objective is to learn an optimal metric matrix M that minimizes the distance between similar pairs and maximizes the distance between dissimilar pairs. Neighborhood repulsed metric learning (NRML) Lu et al. (2014c) uses this concept to ensure that the intraclass samples are close to each other and repulse the interclass samples as far as possible.

Finally, the advent of deep learning, particularly Convolutional Neural Networks (CNNs), brought significant advancements in kinship verification. CNNs automatically learn hierarchical feature representations from raw image data effectively, as shown in Huang et al. (2012), making them well suited for tasks such as kinship verification that require complex feature extraction. The first deep learning model for kinship verification was proposed by Zhang et al. (2015), where the model had three convolutional layers and a fully connected layer, cropping the images with the help of facial landmarks, which showed significant improvement compared to earlier methods which defined the path of kinship research for the next years.

Recent advances in kinship verification have leveraged alongside Siamese Neural Networks the use of contrastive learning, a powerful self-supervised learning framework, to improve model performance by effectively learning discriminative features from facial images. In contrastive learning, models are trained to maximize similarity between positive pairs (e.g., images of family members) and minimize similarity between negative pairs (e.g., non-kin). This approach is particularly well-suited for kinship verification because it focuses on learning fine-grained, relationship-specific features without requiring large amounts of labeled data. The most recent Recognizing Families in the Wild (RFIW) [Robinson et al. (2020)] competition was hosted in 2021, where the best performing model [Zhang et al. (2021b)] utilized a Contrastive Learning approach defined as: Given a set $\mathcal{P} = \{(x_i, y_i)\}_{i=1}^{n}$ where n is the number of positive pairs sampled from different families, contrastive loss L can be defined as:

$$L = \frac{1}{2n} \sum_{i=1}^{n} [L_c(x_i, y_i) + L_c(y_i, x_i)] \tag{2.2}$$

where

$$L_c(x_i, y_i) = -\log \frac{e^{s(x_i, y_i)/\tau}}{\sum_{j=1}^{n} e^{s(x_i, x_j)/\tau} + e^{s(x_i, y_j)/\tau}} \tag{2.3}$$

$s(x, y)$ is defined as the cosine similarity between x and y. τ is used to control the degree of punishment for hard samples, where high values of τ represent a high degree of punishment. With this method, the authors achieved the best result in all three of RFIW's tasks, proving the potential of contrastive learning in kinship recognition. Since then, many work has been done focusing on improve contrastive learning, representing the current state-of-the-art.

2.2 Facial Age Transformation Overview

Facial age transformation is a process that alters the appearance of a face to simulate aging or rejuvenation, while preserving the unique identity of the individual. The primary objective is to generate realistic facial images at different ages, which has numerous applications in fields such as face recognition, movie effects, and social entertainment. Another process closely tied to age transformation is cross-age face recognition. As people age, their facial features change, making it difficult for recognition systems to match images of the same person taken at different ages. Age transformation techniques are often used to help bridge this gap by generating intermediate-aged images [Chen et al. (2019)]. However, these transformations must be highly accurate, as any deviation in identity or unrealistic aging can reduce the performance of the face recognition system. Traditionally, facial age transformation methods were based on physical models and pro-

totypes, but these approaches faced challenges due to their complexity and limited ability to preserve individual facial characteristics.

With the rise of deep learning, especially Generative Adversarial Networks (GANs) [Goodfellow et al. (2014)], facial age transformation has achieved substantial progress. Modern methods are now able to generate more visually realistic results, accurately depicting the aging process while maintaining identity preservation. These techniques include GAN-based models, adversarial encoder-decoder methods, and those that incorporate attention mechanisms to focus on age-related facial regions.

2.2.1 Problem Definition

Facial age transformation can be defined as a problem that takes an image x_0 as input, which is a facial image of an individual of age α_0 and a desired age α_1. The goal is to transform x_0 into the output $x_1 = G(x_0, \alpha_1)$, where G is the age transformation model that represents the individual present in x_0 but looking like someone at age α_1, while maintaining age-unrelated characteristics with x_0, such as identity, emotion, hair, background, and photorealism. A theoretical perfect age transformation model is able to simply convert the age of an individual without zero changes in any other characteristic and would also mean that $x_0 = G(x_0, \alpha_0)$, i.e., the image generated with its original age should output the image itself.

In practical terms, identity preservation is crucial but difficult to maintain, as aging alters many facial features, such as wrinkles, skin texture, and facial structure. A successful age transformation model must retain key identity traits, ensuring that the person remains recognizable at any age. Failing to do so could lead to transformations that appear unrealistic or disjoint from the original individual. In addition, the aging process varies between individuals, making it difficult to generalize. Factors such as genetics, lifestyle, health, and external conditions, such as environmental exposure, contribute to different aging rates [Despois et al. (2020)]. Therefore, a one-size-fits-all aging model is inadequate, as it cannot account for these personal variances. This introduces the need for models that can adapt to individual aging patterns, but a key issue to achieving such a model is the scarcity of datasets that contain images of individuals at multiple age points, especially in controlled settings. Collecting large-scale datasets that span a wide range of ages for the same individual is difficult, limiting the ability of age transformation models to learn effective representations across different age stages.

2.2.2 Main Challenges

Facial age transformation has several potential applications such as biometrics and entertainment [Shu et al. (2016)], but to be used as concrete evidence in the field of forensics, for example, it still has to overcome some challenges to achieve high credibility. These

challenges stem from both the complexity of the aging process itself and the technical limitations of current models. One of the most critical aspect is maintaining the identity of the individual throughout the aging or rejuvenation process. Although facial features change with age, the fundamental traits that define a person's appearance, such as the shape of their eyes, nose, and overall facial structure, must remain consistent. Achieving a balance between altering age-related features and preserving identity is particularly difficult, especially with large age gaps.

The aging process is highly individualized and influenced by various factors, including genetics, lifestyle, and environmental exposure, leading to different aging rates in individuals [Rexbye et al. (2006)]. As a result, two people of the same age can look very different depending on these factors. Guo et al. (2008) showed that different degrees of facial modification by different genders, such as makeup and accessories, can alter the perceptible difference in aging between men and women in images, where these factors are usually not controlled. This variability makes it challenging to build a universal model that can accurately simulate aging across diverse populations. Current models often struggle to generalize the effects of aging between different ethnicities, sex, and other demographic groups, leading to less accurate transformations for underrepresented groups.

Collecting data is also difficult [Liu et al. (2017)], especially if it contains images of the same individual over a wide range of ages, due to the long-term nature of the task and privacy concerns. Most available datasets are limited in size and often lack the diversity needed to train models that can generalize well. With deep learning models, particularly Generative Adversarial Networks (GANs), a lot of training data is required [LeCun et al. (2015)] to achieve good results, while these models have made significant progress in generating visually plausible images, issues like image blurriness, artifacts, and unnatural skin textures still arise, especially when there are large age differences, likely due to the insufficient variety of training data. Ensuring that the generated images are both realistic and free of artifacts, while also maintaining consistency with the original image, is also a difficult balancing act.

2.2.3 Datasets

Face age transformation is a task that is highly dependent on data quality, as it impacts training stability and generated image quality. This section summarizes some relevant and publicly available datasets and compares their characteristics. Table 2.1 shows general information about the datasets, such as age range, number of images, subjects, and average age, which can help selecting a dataset depending on the task.

One of the most popular datasets in terms of age used to be FG-NET [Fu et al. (2016)], published in 2002 and updated in 2014, which contains 1002 facial images from 82 subjects. The images also have face key-point information and vary greatly in age, from

children to elders, which is useful in several scenarios, explaining its popularity. However, the small number of images and low quality of images for current standards make it somewhat obsolete, since there are better options to learn aging patterns. In 2014, the largest available cross-age face recognition and retrieval dataset was published, known as the Cross-Age Celebrity Dataset (CACD) [Chen et al. (2014)]. It contains 163,446 facial images in the wild of 2000 celebrities aged between 16 and 62 years. Each image has 16 facial key points and is widely used for research in cross-age person retrieval.

Flickr-Faces-HQ Dataset (FFHQ) [Karras et al. (2019)] was originally created as a benchmark for generative adversarial networks, consisting of 70,000 high-quality images at 1024x1024, with good coverage of accessories such as eyeglasses, sunglasses, and hats. The images were crawled from Flickr. In addition to being one of the few large high-quality datasets, it also provides face semantic maps that can be used to mask images, segment face regions, and background information to improve age conversion. MORPH [Ricanek and Tesafaye (2006)] was published in 2006 containing around 1,700 facial images, with highly controlled images: frontal pose, neutral expression, moderate lightning and simple background, which makes it a very good benchmark for facial aging. The dataset was later extended and renamed to MORPH2, now having 553,349 images from 13,672 subjects, also providing metadata such as age, gender, and ethnicity, giving it a more balanced distribution of age groups. Another important characteristic of MORPH2 is the presence of images of the same individuals at different points in time, which is valuable for studying age progression, kinship verification, and other tasks.

UTKFace [Zhang et al. (2017)] is the dataset with the longest age range, ranging from 0 to 116 years. It contains 23,709 face images, covering large variations in pose, facial expression, lighting, occlusion, resolution, etc. The images in this dataset provide 68 key points and are labeled by age, gender, and race for tasks such as face detection, age estimation, age progression/regression, and key point localization. Another long-age span dataset is AgeDB [Moschoglou et al. (2017)], which contains 16,488 grayscale images, ranging from 1-101, manually collected, and with manually annotated age and gender. Multi-Racial Child Dataset (MRCD) [Chandaliya and Nain (2022)] is a diversity-focused dataset that provides facial images of children of various racial and ethnic groups, making it valuable for research on issues such as racial bias in facial recognition systems and kinship verification between different racial backgrounds.

Lastly, a recent dataset that has not yet received much attention is the *biometrically filtered famous figure dataset* (B3FD) [Bešenić et al. (2022)]. B3FD is a dataset derived from IMDB-WIKI and CACD, automatically cleaned of faulty web-scraped samples by the unsupervised biometric filtering methods proposed in the paper, which ends up removing 53% of IMDB and 20% of CACD, resulting in 375,592 facial image samples with corresponding age labels. It has 53,759 unique subjects, which amounts to 6.99 samples per subject on average, the age labels are ranging from 0 to 100. As demonstrated in the

paper, the B3FD data outperform all other publicly available data sets evaluated for age estimation, indicating that it is likely also a good dataset for age transformation training.

Datasets	**Images**	**Subjects**	**Age range**	**Average age**	**Year**
FG-NET	1002	82	0–69	15.84	2002
CACD	163,446	2000	16–62	38.03	2014
FFHQ	70,000	–	–	–	2019
MORPH2	553,349	13,672	16–77	32.69	2006
UTKFace	23,709	–	0–116	33	2017
AgeDB	16,488	568	1–101	50.3	2017
MRCD	64,965	–	0–20	–	2022
B3FD	375,592	53,759	0–100	–	2022

Table 2.1: Dataset summary

2.2.4 Evaluation Metrics

Evaluation metrics can be classified as qualitative and quantitative. Qualitative evaluation relies on subjectivity to judge performance, mainly based on human evaluation. For example, in tasks such as image generation or facial age transformation, humans assess the realism, visual quality, and identity preservation of the generated images. This type of evaluation often involves user studies or expert panels in which subjective opinions of evaluators are gathered to determine the effectiveness of the model. While qualitative methods can provide insights into aspects like naturalness or perceptual quality, they are inherently non-reproducible and can vary between evaluators, leading to inconsistency. Quantitative evaluation, on the other hand, is based on measurable metrics that can be automatically computed. These metrics provide objective and reproducible evaluations of model performance, evaluating the statistical properties of the generated images compared to real images, which is also essential to ensure consistency and reproducibility in model evaluation.

Most studies in face age transformation show output samples from their model, giving the reader the opportunity to make a visual analysis, asserting in an intuitive way the model capabilities, which is usually biased, since the authors tend to choose their best results as demonstration samples. A user study can synthesize this human perception in a more reliable way, where a group of randomly selected people is asked to comment on the generated images or to compare the results of multiple models, thus measuring the model based on human judgment. The problem with user studies is that they require significant time and resources. Coordinating participants, designing effective evaluation protocols, and analyzing subjective responses can be costly and labor intensive. In addition, gathering enough participants to ensure statistical significance is difficult, showing the lack of scalability of user studies.

Qualitative evaluation for face age transformation is based on three main assessments: image quality, age estimation, and identity preservation. Image quality metrics can be calculated with or without a reference image, being classified as Full Reference Image Quality Assessment (FR-IQA) and No Reference Image Quality Assessment (NR-IQA). Peak Signal-to-Noise ratio (PSNR) [Hore and Ziou (2010)] is one of the most common FR-IQA metrics, measuring image quality based on pixel intensity differences. PSNR can be defined as:

$$\text{PSNR} = 10 \log_{10} \left(\frac{(2^n - 1)^2}{\text{MSE}} \right) \tag{2.4}$$

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (I_r(i) - I_g(i))^2 \tag{2.5}$$

where n is the number of bits per pixel, and MSE is the mean squared error (MSE) between the reference image I_r and the generated image I_g. The weakness of PSNR and other metrics such as SSIM [Wang et al. (2004)] and LPIPS [Zhang et al. (2018)] is that they do not take into account human perception of faces, which is processed differently inside the brain [Kanwisher et al. (2002)], suggesting the need for specific metrics for face quality. In this direction, methods for Face Image Quality Assessment (FIQA) have been proposed, such as faceqnet [Hernandez-Ortega et al. (2019)] and SDD-FIQA [Ou et al. (2021)], but they are mostly specific to improve face recognition. Recently, Jo et al. (2023) proposed Interpretable Face Quality Assessment (IFQA) a facial metric based on an adversarial framework where a generator simulates face restoration and a discriminator assesses image quality, which gives an interpretable per-pixel quality measurement, aligned with human judgment. The model is trained to focus on facial regions and ignore background, giving more specificity to the metric and making it a very accurate measure of face quality, which can be used to evaluate face age transformation models.

The age estimation aims to measure the accuracy of the age of the facial images generated with respect to the desired age α_1. There are two ways to evaluate the ability of the models to create age-accurate images, depending on the approach used for age assessment, which could be continuous age or discrete. By choosing to view age as a discrete value, age transformation can be viewed as a classification problem with respect to the age groups, which could be either age single values or age ranges, such as 10-20, 21-30, etc. Being a classification problem, it is possible to calculate the ratio of the predicted ages of the generated images that fall into the correct age group. Usually, an online face recognition API like Face++ and an age estimation model for prediction are used, and then the obtained age estimation distribution is calculated to obtain the aging accuracy. Similarly, a continuous age transformation can be viewed as a regression problem, where the mean absolute error between the estimated age $\hat{a}_i$ and the true age

a_i can be calculated, named Age MAE and defined as:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |a_i - \hat{a}_i| \tag{2.6}$$

where a smaller MAE indicates a smaller error range and a higher accuracy of the algorithm. Finally, age-independent features can be measured by comparing the similarity between the input image and the generated image. This can be evaluated using a pre-trained face recognition model, which should be capable of recognizing both images as the same identity. In addition to this method, identity preservation can also be measured using the Fréchet Inception Distance (FID) [Heusel et al. (2017)] and Kernel Inception Distance [Bińkowski et al. (2018)] scores. Both scores aim to measure how similar the generated images are to the real images in a dataset, which is helpful in both quality assessment and identity preservation, where FID calculates the distance between the feature vector of the real image and the generated image and KID measures the difference between sets of samples by calculating the square of the maximum mean difference between the Inception representations.

2.2.5 Existing Methods

Traditional face age transformation methods can be classified into two general categories: model-based physical methods and prototype methods [Guo et al. (2024)]. The purpose of the physical model-based model is to simulate the time-varying facial appearance, such as the facial muscles [Berg and Justo (2003)] and the skin, through a set of parameters. This is a very mechanical approach and requires a lot of computational power and training data, since its objective is very complex and specific. In contrast, prototype-based methods use the average face as a prototype for each age group and achieve aging or rejuvenation of the face by applying the differences between the prototypes to the input face images. However, given the high variance of faces throughout the world, the average face does not give good results in some cases and causes identity loss in the process.

As for Deep Learning methods, a great variety of methods have been proposed, including Variational Autoencoders (VAE) [Kingma (2013)] and Generative Adversarial Networks (GAN) [Goodfellow et al. (2014)], including its variants such as Conditional GAN (CGAN) [Mirza (2014)] and StyleGAN [Karras et al. (2020)]. To improve the quality of the images and the stability of the training, attention mechanisms were implemented [Xiao et al. (2015)]. In addition, some researchers also explored the fusion of multiple networks to generate high-quality facial images by integrating different components and benefiting from each module. Compared to traditional methods, deep learning-based methods perform better in terms of visual fidelity, aging accuracy, and identity preservation. Figure 2.3 shows a visual representation of this classification.

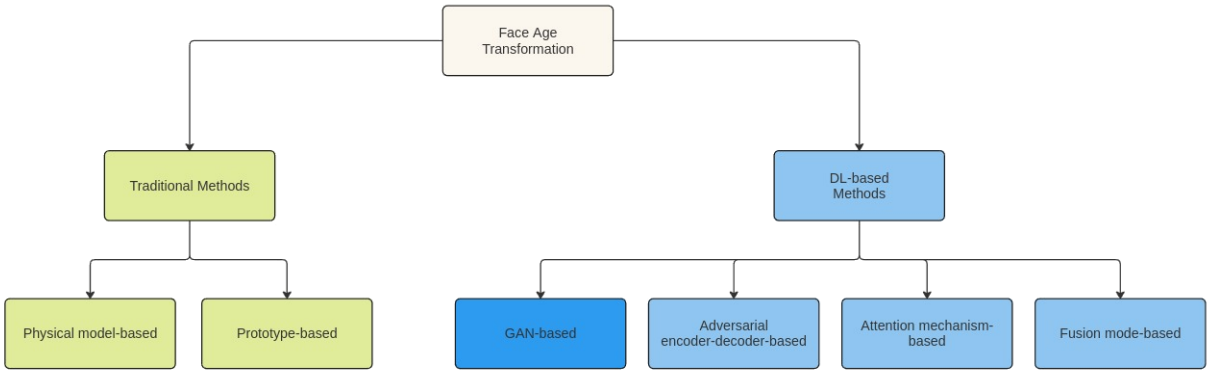

Figure 2.3: Classification of facial age transformation methods. Source Guo et al. (2024)

GAN-based methods are trained using a min-max learning optimization, which mainly consists of a generator G and a discriminator D, where the generator is minimized and the discriminator maximized. The role of the generator is to learn the real image distribution while the goal of the discriminator is to determine whether the current image is a real image or a generated fake image. In the context of face age transformation, the generator needs to create images with the age transformed and the generator needs to determine if the image seems real in the transformed age, without loss in identity, for example. In order to achieve this behavior, several GAN methods have been proposed. Initially, the early GAN-based face age transformation methods were unconditional, i.e., did not use age conditions to guide training, which led to poor aging accuracy and identity persistence. To improve unconditional methods, Yang et al. (2018) proposed the pyramid architecture of GANs (PAGAN) combining face verification with age estimation to capture high-level age information. In PAGAN, a pyramid structure discriminator is designed to ensure that the generated faces exhibit the desired aging effect.

To better enforce identity preservation on the generated images, conditional GANs were created [Mirza (2014)]. Identity-Preserved Conditional GAN (IPCGAN) [Wang et al. (2018)] designed the identity retention module and the age classification module to maintain the identity information while ensuring that the generated faces match the target age. ChildFace [Chandaliya et al. (2020)] added gender and age conditions to the generator to learn gender-aware age distribution, improving face recognition performance. Song et al. (2018) proposed Age-GAN, using an architecture with double CGAN, where an original conditional GAN performs face age transformation while the dual CGANs learn to invert the task, thus improving the general quality of the images. Age-GAN++ [Song et al. (2021)] improves on Age-GAN by sharing the weights of the original and dual parts to simplify the model. Furthermore, a representational disentanglement component was added to enhance the discriminator preservation of age features during generation, thereby improving model performance, but this also lowers the model effectiveness when

the difference between two domains is small. To address this problem, BiTrackGAN [Kuo et al. (2023)] uses a bottom-up approach to train two cascaded CycleGAN blocks, inducing an ideal intermediate state, that is, a constraint mechanism, between the two CycleGAN blocks to achieve more reasonable and accurate facial aging and rejuvenation.

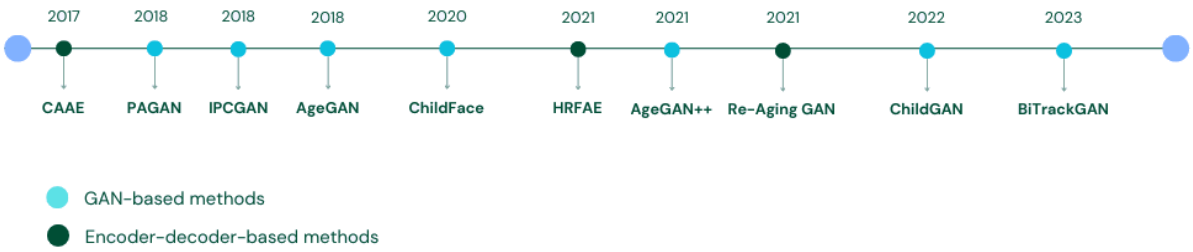

Figure 2.4: Timeline of some Face Age Transformation methods. Source Guo et al. (2024)

The adversarial encoder-decoder-based methods focus on altering the generator to an encoder-decoder architecture. In general, the encoder is in charge of performing feature extraction on the input image, then, the extracted features are used to feed the decoder, which will generate the image. Conditional Adversarial Autoencoder (CAAE) [Zhang et al. (2017)] learns a face representation training the encoder and decoder separately. The main objective of this type of method is to create a latent space containing identity information. High resolution face age editing (HRFAE) [Yao et al. (2021)] combines the encoder-decoder architecture with a feature modulation layer, which inserts an age encoding into the decoder, allowing the use of an unconditional discriminator focused only on image quality, thus significantly improving visual results. Another work that uses the same concept of an age modulation module is Re-Aging GAN [Makhmudkhujaev et al. (2021)], which uses the interaction between a given identity and a target age to learn personalized age features, self-guiding the decoding process and also achieving good results. Most methods are designed to explore changes in adult age, usually not taking into account children's face aging, which is affected by several other factors, such as puberty. ChildGAN [Chandaliya and Nain (2022)] is a Fusion mode-based method which combines the Variational Autoencoder with GAN to improve the continuity and smoothness of the latent space, trying to obtain the good image quality of GANs and the training stability of VAEs. Figure 2.4 shows a timeline of some age transformation methods.

Chapter 3

Theoretical Foundation

In this chapter, the fundamental concepts applied in this work are introduced, including neural networks, convolutional neural networks, and generative adversarial networks. Starting with the history of deep learning, the core features of DL models are explained, and challenges and complications are also addressed during neural network training, to detect these problems and assess how to resolve them. In the following, computer vision is the focus, with convolution and its application on neural networks, convolutional layers, pooling layers, and residual connections, which made this and other computer vision solutions possible. Finally, the concept of Generative Adversarial Networks is explained since it is the basis of the age transformation model.

3.1 Neural Networks and Deep Learning

Neural networks were first conceived in the mid-20th century, when Frank Rosenblatt established the concept of a perceptron [Rosenblatt (1958)], which is an operator that takes an input, applies its internal value to it, called weight, and then returns an output. The inspiration comes from human biology, with the perceptron being a representation of a neuron, where dendrites receive input signals from other neurons, process them, and output another signal to nearby neurons. More specifically, perceptron is a linear operator that can be used as a binary classifier, where the output z can be calculated as:

$$z = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + b \tag{3.1}$$

For an input vector x of size n, the vector w represent the perceptron weights. Finally, b represents the bias term, completing the affine function, where a line in the n-dimensional space can be used as a boundary for separating two classes. Because of its linear nature, a perceptron can only classify data that are linearly separable. If the data are not linearly separable, the perceptron will fail to classify with good performance. To solve this issue, activation functions were invented, which are mathematical functions that can be applied

to the output of the perceptron to introduce non-linearity, such as Rectified Linear Unit $\text{ReLU}(x) = \max(0, x)$ [Agarap (2018)], a simple function that clips to zero any negative input. This simple change, combined with the cascade use of several perceptrons, called neurons, created the field of Neural Networks, which were capable of handling complex, non-linear problems.

In a neural network, neurons are organized in layers, arrays of neurons that receive the input from the previous layer and propagate new values to the next layer, generating a final output once it has reached the final layer. For many years, neural networks were primarily regarded as theoretical models due to significant computational limitations that made their practical application challenging. Early neural networks showed promise, but the lack of sufficient computing power, data, and efficient algorithms meant that their potential remained largely unexplored. This situation changed dramatically in 2012 when a neural network model, AlexNet [Krizhevsky et al. (2012)], achieved groundbreaking performance in the ImageNet competition, demonstrating the true power of neural networks and sparking widespread adoption in various fields. This marked the rise of deep learning [Wang et al. (2017)], which is simply a name for neural networks with many layers, which enables it to learn complex patterns in the data, with deep understanding. Since then, advancements in hardware, especially GPUs, the availability of large datasets, and new training techniques have allowed the successful implementation of deep neural networks.

These models learn by an optimization process of a loss function, which measures the difference between the model's predictions and the true labels or desired outputs. Optimization is typically performed using stochastic gradient descent (SGD) [Bottou (2012)] or its variants, where the model iteratively updates the weights by computing the gradient of the loss function with respect to the parameters, since the gradients indicate the direction in which the weights should be adjusted to reduce the loss. Through this process, the model gradually converges toward a set of parameters that minimizes the error and generalizes well to unseen data. This is possible thanks to the backpropagation algorithm that applies the chain rule of calculus, propagating the error backward from the output layer to the earlier layers, allowing the model to update all the weights across the network in a very efficient way.

3.1.1 Convolution Neural Networks

Convolutional Neural Networks [Gu et al. (2015)], or CNNs, are a specialized type of neural network that is mainly used to process grid-like data, such as images. Initially introduced in the 1980s and later gaining widespread popularity in the 2010s, CNNs have become the dominant architecture for image processing and computer vision tasks. This success is largely due to their ability to efficiently and automatically extract spatial hierarchies of features from images. They are particularly well suited for image-related

tasks because they are designed to capture spatial patterns, such as edges, textures, and shapes, across the dimensions of an image. Unlike traditional fully connected neural networks, where every neuron is connected to every other neuron, CNNs make use of convolutions, a mathematical operation that preserves the spatial relationships between pixels by applying filters or kernels across the input image.

There are several key reasons why CNNs are used for image processing tasks. First, CNNs are adept at detecting local features, such as edges and textures, by using small filters or kernels that slide across the input image. This ability allows them to recognize objects regardless of where they appear in the image, making them spatially invariant. In addition, they are robust to translation, meaning that they can detect features such as objects or shapes no matter where they appear in the image. Another advantage of CNNs is their efficient use of parameters through a concept known as parameter sharing. Instead of requiring each pixel to have its own weight, CNNs apply the same filter across the entire image, greatly reducing the number of parameters compared to fully connected networks. Convolutional Neural Networks also excel at learning hierarchical features. In the early layers of a CNN, the network learns simple patterns such as edges and textures, while deeper layers detect more complex patterns such as shapes and objects. This hierarchical learning is essential for recognizing high-level patterns in images, which is critical for tasks like object detection, face recognition, and scene analysis.

Pooling layers are also often used to help reduce the dimensionality of the data by summarizing regions of the image. This process is executed by dividing the images into small grids, similar to convolution, where only one value of that grid is going to be passed along to the network. The most common types of pooling are max pooling and average pooling. In max pooling, only the maximum value within the grid is retained. This operation captures the most prominent feature in that region, enabling the network to focus on the most significant signals. In average pooling, instead of selecting the maximum value, the average of all values within the grid is taken. This approach provides a more generalized view of the features in each region and can be useful in tasks where each pixel contributes equally to the overall pattern. This makes the network more computationally efficient and adds a degree of robustness to small distortions in the image, such as shifts or rotations.

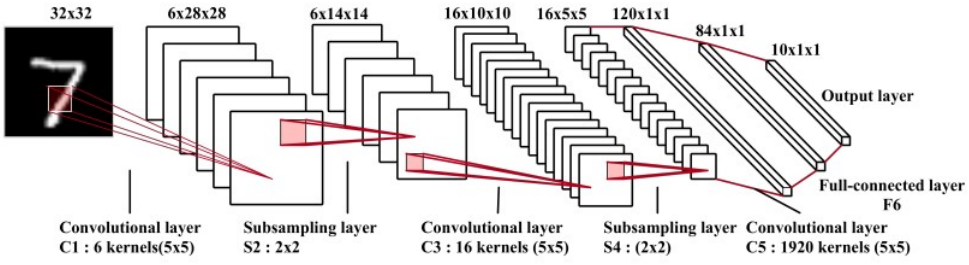

Figure 3.1: Example of a complete CNN architecture, LeNet-5. Source Gu et al. (2015)

Figure 3.1 represents LeNet-5 [LeCun et al. (1989)], where an end-to-end convolutional neural network was built for handwritten digit recognition. First, a 32x32 image of a digit is given as input to the network, and then it passes through a series of alternated convolutional layers and subsampling layers, which are pooling layers with 2x2 grids, causing the image representation to halve its size. After the last convolutional layer, the output is flattened and forwarded to a multilayer perceptron, which receives the representation created by the network and classifies the image.

3.1.2 Residual Connections

Residual connections are an architectural feature in deep neural networks, which address the problem of vanishing gradients and degradation in performance as networks grow deeper [He et al. (2016)]. In a standard deep network, as the number of layers increases, the gradients during backpropagation can become exceedingly small, leading to poor learning in the earlier layers. This phenomenon, known as the vanishing gradient problem, can cause performance to stagnate or even degrade with increasing depth, hindering the network's ability to learn complex patterns. To counteract this, residual connections introduce a shortcut path that skips one or more layers and connects the output of an earlier layer directly to a later layer. Specifically, in a residual block, the input is added directly to the output of a few stacked layers, typically convolutional layers, before being passed to the next block. This addition operation essentially combines the original input with the transformed input, preserving the information from the previous layers. Figure 3.2 shows the flow of a residual block.

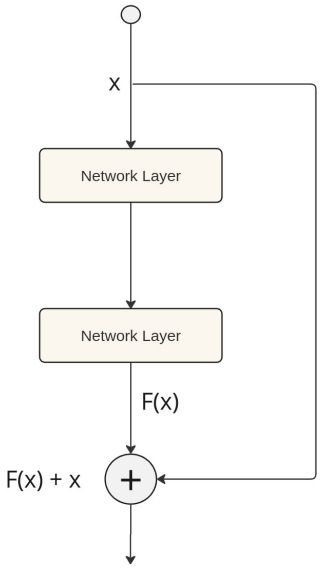

Figure 3.2: Residual block Illustration.

Mathematically, if the original transformation in a network layer is represented as F(x),

where x is the input, then a residual connection reformulates it as F(x) + x. The network now learns a residual mapping that focuses on modeling the differences or adjustments needed instead of the entire transformation. The added input bypasses the main network layers and is summed directly with the output of the transformed data. This simple addition helps ensure that even if certain layers contribute little or no useful information, the original input can still propagate through the network unchanged.

Residual connections provide two main benefits. First, they help maintain gradient flow across layers. Second, they facilitate faster convergence during training by allowing the network to learn smaller incremental changes rather than forcing each layer to learn a complete transformation. Consequently, residual connections have enabled the development of extremely deep architectures, with hundreds or even thousands of layers, that achieve state-of-the-art performance in various tasks like image classification, object detection, and natural language processing.

3.2 Generative Adversarial Networks

Generative Adversarial Networks (GANs) Goodfellow et al. (2014) are a class of deep learning models designed for generative tasks, where the goal is to generate new data instances that resemble a given dataset, thus learning the distribution of the data, as opposed to discriminative models, which map a high-dimensional input to a single label.

GANs consist of two neural networks: a generator and a discriminator, which engage in a competitive process. They are adversaries in a min-max game, where the discriminative model learns to determine whether a sample is from the model distribution or the data distribution, while the generator tries to produce samples that are as close as possible to the real data distribution in order to fool the discriminator. This adversarial setup allows GANs to generate highly realistic data, such as images, by learning the underlying distribution of the training data. Over time, the generator improves in creating realistic data that can fool the discriminator, while the discriminator becomes better at identifying real versus fake data. This dynamic interaction between the two networks drives the learning process.

As neural networks, both generator and discriminator need input in order to provide the output. For the discriminator, it is easy to understand that it's training inputs should be samples both from the training data x and from the generator's output. As for the generator, it receives a random noise vector z, since an input to it does not represent anything. The generator output is a sample $G(z)$ in the same format as the training data, resembling its distribution, while the discriminator outputs a scalar value between 0 and 1, representing the probability of the sample being real, i.e., if the input is a generated image, $D(G(z))$ should be close to zero, while $D(x)$ should be close to 1, since it discriminates a real image.

The training of GANs is framed as a two-player game between the generator and the discriminator that happens simultaneously. This adversarial behavior is represented by this value function $V(G, D)$:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))] \tag{3.2}$$

where p_{data} is the true distribution and p_z is the learned generator distribution, which usually is initialized as a Gaussian distribution. The generator goal is to minimize, because making the term $\log(1 - D(G(z)))$ as low as possible makes the generated data indistinguishable from the real data. The discriminator tries to maximize $\log D(x)$, which is the probability that a real data point is identified as real, and maximize the result of $\log(1 - D(G(z)))$, achieved by low values of $D(G(z))$.

In practice, the training process is quite delicate. Both networks should learn and improve together at roughly the same pace; otherwise, one of the networks would suppress the other. A simple way to think of this is the adversarial nature of the training: if your opponent is much better than you, training will not be beneficial for either of you. To fix this problem, the number of steps that each model is optimized must be balanced, in order to keep their disparity at a limit. Therefore, GAN performance is highly sensitive to hyperparameter choices, such as learning rates, optimization algorithms, and the architecture of both the generator and the discriminator. Fine-tuning these hyperparameters is often necessary to achieve stable and high-quality results.

Generative Adversarial Networks have a wide range of applications across various fields, the most common being image generation, where the generator is trained to produce high-quality images, including faces, landscapes, and objects, that can be used as synthetic training data, art creation, entertainment, etc. This is the application used in this work, where images are generated to be used as training data for a kinship classifier. More specifically, an image-to-image translation is being performed, since the age transformation model is translating images into different ages. Other GAN applications include super-resolution, which is used to increase the resolution of low-quality images and audio generation.

Chapter 4

Methodology

In this chapter, we introduce the methodology applied in this work, including the data processing steps, the design, implementation, and integration of the age transformation model with the kinship verification system, as well as the training breakdown. The methodology begins with a detailed explanation of how the datasets were prepared, addressing any preprocessing techniques necessary to ensure compatibility with the proposed models. This is followed by a comprehensive description of the age transformation model's architecture, highlighting the specific design choices that enable realistic and identity-preserving facial age modifications.

The integration process outlines how the age-transformed images were incorporated into the kinship verification system to enhance its accuracy and robustness. Finally, the training process is broken down into its key components, including the training parameters and optimization strategies used to achieve the desired performance.

4.1 Data Preprocessing

The age transformation model was trained using the *Biometrically Filtered Famous Figure Dataset* (B3FD) [Bešenić et al. (2022)] dataset with some filtering. First, B3FD was filtered only for facial images of ages within the age range, which is between 20 and 70 years old. Then, the age distribution $\mathcal{Q}$ was analyzed, as shown in figure 4.1. Since most of the images are centered around 30 years and there are comparatively few samples of elderly people, training with all the data could cause an imbalance in performance, making the model learn more about transformations on the lower end of the age range spectrum. To prevent this, the dataset was once again filtered, in order to approximate the distribution Q to uniform. By choosing 2000 images per age, except for those that do not have enough, the final dataset used in the training had 93,325 images. After filtering, normalization was applied using the mean and standard deviation of ImageNet, which is common in computer vision pipelines for training.

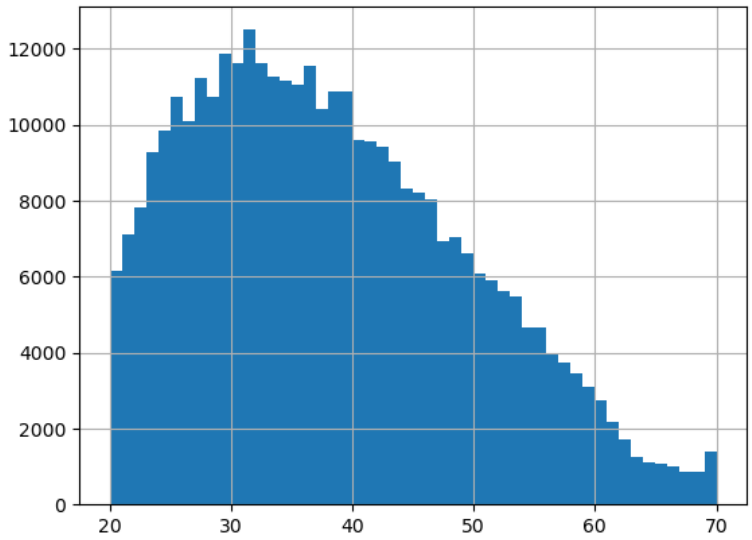

Figure 4.1: Age distribution of B3FD dataset

This model is applied on *Families in the Wild* (FIW) [Robinson et al. (2016)], to generate an augmented dataset, where each image from the original data is used to generate five images, with target ages α_1 20, 30, 40, 50 and 60, respectively. This dataset is used as training data for the classification model, where the data is organized in pairs, so that the pairs in each minimum batch come from different families, as is done to ensure that the restrictions on contrastive learning are respected and the results reproducible, as is done in Zhang et al. (2021b).

4.2 Face age transformation model

The proposed approach for age transformation is an adaptation from *High Resolution Face Age Editing* [Yao et al. (2021)], fit not for high-resolution, but for low resolution, since kinship datasets like KinFace and FIW are made of low resolution images. A custom-designed encoder-decoder architecture is implemented. This model utilizes a combination of latent identity features and age-specific modulation to achieve photorealistic age modifications with minimal artifacts.

4.2.1 Architecture

The main components of the architecture for generating the images are the image encoder, the age encoder, and the decoder. First, the image encoder is constructed using convolutional layers and four residual blocks, where the first convolutional layer has a stride of 1 to capture fine details, and the next two layers use a stride of 2, progressively downsampling the image. The residual blocks process the downsampled feature map, helping the model retain facial identity information by allowing the gradient to flow through multiple layers

effectively, thus preserving facial details that are not related to age. In summary, the image encoder processes the input face image to generate a deep feature representation that captures identity, expression, and other non-age-related characteristics, and outputs a feature map denoted by $C \in \mathbb{R}^{n \times c}$, where $c = 128$ is the number of channels, and n is the product of the two spatial dimensions.

The feature map created by the image encoder does not contain age information inherently; this is the role of the age encoder, which is a component inspired by style transfer techniques. It allows for age transformation by directly modifying the encoded features according to the target age. The target age is encoded as a one-hot vector, and passed as input to the age encoder that comprises a fully connected layer with a sigmoid activation, returning an output a modulation vector w in the range [0, 1], representing how much each channel should change to adapt to the target age. This vector has 128 elements that match the encoder output channel count, to modulate each feature channel individually. The encoded features C are multiplied by the diagonal matrix $diag(w)$, effectively scaling each feature channel according to the desired age.

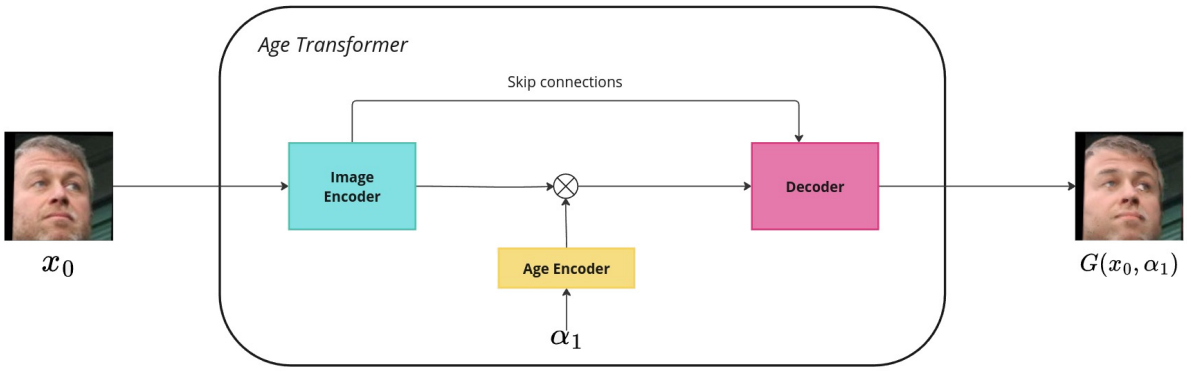

Figure 4.2: General Architecture of the age transformer generator

Finally, these features are forwarded to the decoder, which reconstructs the modulated feature map into a face image, as the final step for the generator. The two upsampling layers progressively increase the spatial resolution of the features. This is followed by convolutional layers to refine the image details, enabling the output at 256x256. Two skip connections link the encoder to the decoder, bypassing parts of the network to help preserve age-irrelevant details like background and finer textures. The skip connections prevent the model from altering these features unnecessarily, improving image consistency and reducing artifacts. : The final output, denoted as $G(x_0, \alpha_1)$, represents the input face at the target age, with the identity, expression, and non-age-related features maintained. Figure 4.2 illustrates the proposed generator.

Opposed to the generator, there is the discriminator, which plays a crucial role in ensuring the output appears photorealistic by distinguishing real images from edited ones. Unlike many traditional face-editing methods, this discriminator does not rely on age

labels; instead, it focuses solely on the realism of the generated images, which helps prevent the introduction of age-conditioned artifacts. By using a PatchGAN [Isola et al. (2017)] framework, where the discriminator divides each image into smaller patches and assesses the photorealism of each patch individually. This patch-based approach enables the model to capture fine-grained details across localized regions of the face, helping to ensure that the generated image appears coherent and realistic on a smaller scale and at the full-image level. This technique is especially effective for maintaining facial texture and minimizing unrealistic artifacts that may otherwise arise from the generative process.

To optimize the performance of both the generator and discriminator, the model employs a loss function tailored for this adversarial training setup. This is the key to guiding the generator toward producing images that not only appear realistic to the discriminator but also maintain identity consistency and accurately depict the target age. The following section delves into the specific loss function used in this model, explaining its components and how it balances realism with identity preservation and age progression in the generated images.

4.2.2 Loss Function

The training process for the age transformer model incorporates three primary losses: reconstruction loss, classification loss, and adversarial loss. Each loss plays a distinct role in guiding the model to produce realistic and accurate age-modified face images. The final loss is a balanced composition of those three.

Reconstruction loss is crucial for preserving non-age-related features, such as identity, expression, and background details, when the input and target ages are the same. This loss encourages the model to generate an output that is identical to the input image when $\alpha_1 = \alpha_0$. By minimizing this loss, the model learns to retain essential details and avoid unnecessary modifications, ensuring that only age-relevant changes are applied. This can be achieved by calculating the L1 norm (mean absolute error) between the original image x_0 and the reconstructed image $G(x_0, \alpha_0)$:

$$\mathcal{L}_{\text{recon}} = \mathbb{E}_{\mathbf{x}_0 \sim p(x)} \left[\| G(\mathbf{x}_0, \alpha_0) - \mathbf{x}_0 \|_1 \right]. \tag{4.1}$$

By minimizing $\mathcal{L}_{\text{recon}}$, the L1 norm penalizes significant differences between input and output while being robust to minor variations. In terms of classification loss, it ensures that the output transformed by age matches the specified target age, guiding the model to produce accurate age transformations. This is done by comparing the generated image's age α_0 with the target age α_1, helping the model accurately render age-relevant changes, like wrinkles or skin texture adjustments, that are characteristic of the desired age group. A pre-trained age classifier, denoted by V, is used to predict the age of the generated image. The classifier produces a probability distribution over a set of possible ages, and

the classification loss is calculated as the cross-entropy between the target age, represented as a one-hot vector z_1, and the predicted distribution:

$$\mathcal{L}_{\text{class}} = \mathbb{E}_{\mathbf{x}_0 \sim p(x)} \mathbb{E}_{\alpha_1 \sim q(\alpha|\alpha_0)} \left[\ell(\mathbf{z}_1, V(G(\mathbf{x}_0, \alpha_1))) \right] \tag{4.2}$$

where $p(x)$ represents the training image distribution over X and ℓ denotes the categorical cross-entropy loss. The target age α_1 is sampled such that there is a sufficient age difference from α_0, which prevents the network from making only minor, indistinguishable adjustments, the minimum difference of 25 years is applied. The last and most important loss is the adversarial loss, which guides the generator versus discriminator dynamic. For this, the LSGAN objective is adopted [Mao et al. (2017)], where the generator loss is:

$$\mathcal{L}_{\text{GAN}}(G) = \mathbb{E}_{\mathbf{x}_0 \sim p(x)} \mathbb{E}_{\alpha_1 \sim q(\alpha|\alpha_0)} \left[\left(D(G(\mathbf{x}_0, \alpha_1)) - 1 \right)^2 \right], \tag{4.3}$$

and the discriminator loss is:

$$\mathcal{L}_{\text{GAN}}(D) = \mathbb{E}_{\mathbf{x}_0 \sim p(x)} \mathbb{E}_{\alpha_1 \sim q(\alpha|\alpha_0)} \left[\left(D(G(\mathbf{x}_0, \alpha_1)) \right)^2 \right] + \mathbb{E}_{\mathbf{y} \sim p(x)} \left[\left(D(\mathbf{y}) - 1 \right)^2 \right] \tag{4.4}$$

Finally, the overall loss function combines all three components:

$$\mathcal{L} = \mathcal{L}_{GAN} + \lambda_{recon} \mathcal{L}_{recon} + \lambda_{class} \mathcal{L}_{class} \tag{4.5}$$

where λ_{recon} and λ_{class} are weights that balance the influence of each loss.

4.3 Kinship Verification Model

The kinship verification model in this work is designed to take advantage of the enhanced dataset created by the age transformation model to improve the accuracy of facial kinship recognition. The integration of age-transformed images aims to address the challenges posed by significant age gaps among family members, which often hinder traditional kinship verification systems. In this section, the core components of the kinship verification model are outlined and it's integration with the transformed images is explained.

4.3.1 Architecture

The kinship verification model is a Siamese Neural Network based on Zhang et al. (2021b) architecture, which is widely used for pairwise similarity tasks. Given a pair of facial images (x, y), intermediate features (h_x, h_y) are extracted using a backbone network. The features are then fed into a multilayer perceptron to obtain a low-dimensional feature pair (f_x, f_y). In the training stage, f_x and f_y are compared using cosine similarity and

optimized using the contrastive loss function, as defined in Section 2.1.4. Negative samples are collected by combining positive samples from different families, since the contrastive loss is going to bring samples in the same family close together in the feature space, while samples from different families should be apart from each other. An important component is the hyperparameter τ, defined in Equation 2.3. It controls the degree of punishment for hard samples, where smaller values correspond to a great punishment.

The validation is split into two, where 90% of validation data is used to evaluate the model every epoch and the rest is used to find the optimal verification threshold after training by optimizing the Area Under the Curve (AUC). For prediction, the projection component is discarded and the similarity is computed directly using h_x and h_y and the calculated threshold. The feature extraction network is a pre-trained ResNet101 called ArcFace [Deng et al. (2019)], which has been successfully used for the extraction of kinship characteristics in the past [Shadrikov (2020)]. The projection network is a simple multilayer perceptron with a fully connected layer, with batch normalization and ReLU activation, followed by another fully connected layer. The training pipeline approach is illustrated in Figure 4.3.

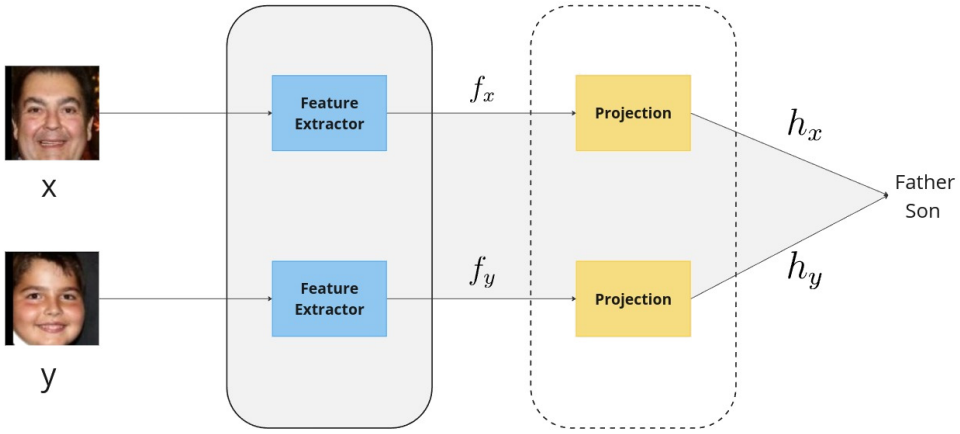

Figure 4.3: Kinship Model Training pipeline

4.3.2 Integration with the Age Transformation Model

To integrate the age transformation model with the kinship verification framework, multiple strategies were employed to effectively leverage the age-transformed images and enhance the feature representation of individuals across varying age groups. The first approach was using images from a specific age group, where, for each individual, the images generated at a specific target age (e.g., 30 years) were incorporated into the dataset. This method assumes that choosing a consistent age group for all individuals can provide a

standardized basis for feature comparison and improve kinship verification accuracy. The second approach utilizes all the images from the 5 age groups, passing them through the network, and calculating the final feature vectors (f_x, f_y) by taking the mean of the five vectors. Finally, the third approach, which yielded the best results is random age sampling for data augmentation. To augment the diversity of the dataset, randomly selected age-transformed images for each individual were added to the dataset. This approach aims to introduce variability, making the model more robust to age-related differences, and reduce over-fitting. These strategies were designed to exploit the strengths of the age transformation model, ensuring that the kinship verification model benefits from the additional age-aligned information while maintaining identity preservation.

4.4 Training

For the training of the age transformation model, the hyperparameters are carefully selected to stabilize the adversarial training. There are two phases of training: the first ten epochs are trained in 128x128 and $|B| = 4$, for faster initial learning, then the rest are trained in 256x256 and $|B| = 2$ to refine the results. The parameters used are listed in Table 4.1

Parameter	**Value(s)**
Epochs	20
Batch size ($\|B\|$)	4, 2
Image size	128x128, 256x256
Optimizer	Adam
Weight Decay	0.0005
Learning Rate (α)	0.0001
λ_{recon}	10
λ_{class}	0.1

Table 4.1: Hyperparameters used in the age transformation model

Training files were implemented for training and evaluation of the results. The model architecture and training loop were implemented in PyTorch. The experiments were carried out using an Intel® Xeon(R) Silver 4216 CPU @ 2.10GHz with 512 Gb of RAM, NVIDIA RTX 3090, and a python 3.10 environment. The same environment was used to train the kinship verification model and its training parameters are available in Table 4.2.

Parameter	**Value(s)**
Epochs	80
Steps	100
Batch size ($\|B\|$)	25
Temperature (τ)	0.08
Optimizer	SGD
Learning Rate (α)	0.0001
Momentum	0.9
Threshold	0.108667

Table 4.2: Hyperparameters used in the kinship model

Chapter 5

Results and Discussions

This section presents and analyzes the results of the proposed approach for improving kinship recognition using face age transformation. The results are evaluated to assess the impact of augmenting kinship datasets with age-transformed facial images, generated by the GAN-based model, on the performance of kinship verification systems. Both quantitative metrics and qualitative visual comparisons are used to provide a comprehensive evaluation of the model's effectiveness.

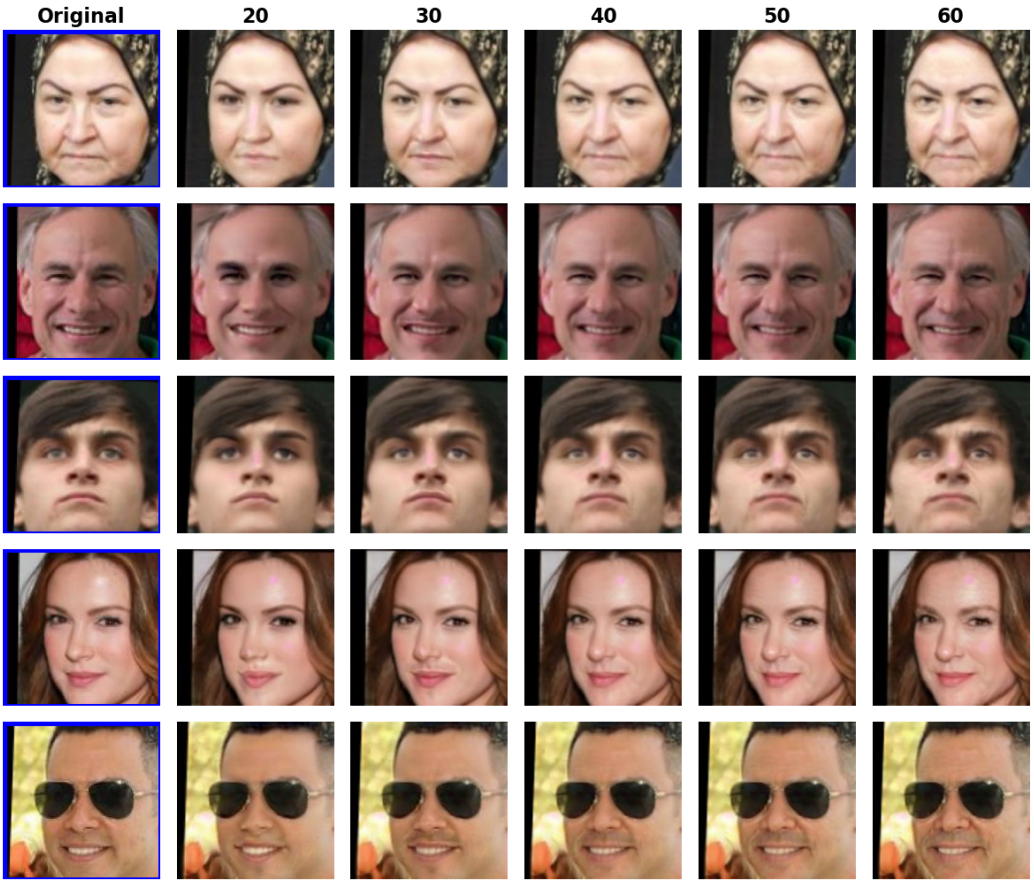

Figure 5.1: Age transformation results of 256x256 FIW images.

The discussion explores how age-transformed images affect the model's ability to bridge age-related gaps in kinship pairs, addressing challenges such as age disparity and limited dataset diversity. In addition, comparisons are made with baseline models and existing state-of-the-art methods to highlight the advantages and limitations of the proposed approach. The insights gained from the results are used to identify key factors that contribute to performance improvements and potential areas for further refinement.

Some examples of transformed image can be seen in Figure 5.1. The differences between the images of the same individual at different ages are subtle, which illustrates the smoothness of the aging process. Compared to the original images, the proposed approach only changes the age-relevant facial features, while the identity, haircut, emotion, and background are well preserved.

Figure 5.2: Comparison with IPCGAN [Wang et al. (2018)] on FIW.

The comparison shown in Figure 5.2 illustrates the performance of the proposed age transformation model against IPCGAN [Wang et al. (2018)] on the FIW dataset. The images show the progression of the age transformation for two individuals in different target age ranges. The proposed model demonstrates more consistent visual fidelity in the age transformation compared to IPCGAN, maintaining a natural and consistent appearance

throughout the aging process. In contrast, IPCGAN introduces noticeable distortions in some transformed images, such as irregular textures and unnatural warping. The proposed method is also more restrictive in terms of identity preservation, not altering features such as skin tone, facial structure, or expressions beyond what is expected for the target age range. This ensures that the transformed images remain easily recognizable as the same individual, an essential factor for downstream tasks like kinship verification.

The accuracy results, calculated using DeepFace [Serengil and Ozpinar (2021)], are presented in Table 5.1 that compare the performance of different methods in various age groups. The proposed model achieves significantly higher accuracy in the age group of 21-30 compared to other methods. PAGAN does not report results for this group, while IPCGAN and AcGAN achieve 46.15% and 25.92%, respectively. This demonstrates the proposed model's ability to handle features of younger faces effectively, which often involve subtler age-related cues. In general, except for older ages, the proposed model achieves better results. This reflects the method's capability to capture and reproduce nuanced age-related transformations while preserving identity. Even with comparable results, aging accuracy in facial age transformation is still a difficult problem, since most methods can barely reach 60% accuracy. Thus, there remains significant potential for improvement in age transformation models, especially for tasks such as kinship recognition, where better aging accuracy tends to yield better results overall.

Method	**21-30**	**31-40**	**41-50**	**50+**
PAGAN [Yang et al. (2018)]	-	42.84	50.78	**59.91**
AcGAN [Zhu et al. (2020)]	25.92	36.49	40.59	47.88
IPCGAN [Wang et al. (2018)]	46.15	55.41	53.86	55.64
Ours	**62.9**	**58.10**	**57.36**	55.64

Table 5.1: Aging accuracy comparison on different age groups.

The results presented in Table 5.2 highlight the face recognition accuracy achieved when comparing the original images with their corresponding transformed images across various target age ranges. These results, calculated using DeepFace [Serengil and Ozpinar (2024)], demonstrate the robustness of the proposed age transformation model in preserving identity in different age transformations. The accuracy progressively increases as the target age range moves from younger to older, starting at 86.65% for the 20-year target and reaching a maximum of 93.10% for the 60-year target.

This trend suggests that the model performs better in preserving identity for transformations to older age groups, possibly because aging features such as wrinkles and skin texture are more distinct and easier to synthesize without distorting unique identity-related traits. The average accuracy indicates that the age transformation model is effective in maintaining key identity features, regardless of the target age group, which is crucial in this application.

Target image (α_1)	**Accuracy**
20 years	86.65
30 years	88.56
40 years	89.57
50 years	91.46
60 years	93.10
Average	**89.87**

Table 5.2: Face recognition accuracy across all target images

The results presented in Table 5.3 demonstrate the performance of various kinship verification methods across 11 kinship relations. The proposed method (Ours) achieves the highest average accuracy compared to state-of-the-art approaches, highlighting the effectiveness of integrating age-transformed images into the kinship verification pipeline. It surpasses only by a narrow but consistent margin. Although this difference in accuracy is not significant, this improvement reflects the contribution of age-transformed images in bridging age gaps and improving feature consistency across different age groups, proving its potential. With better age transformation models, the accuracy gains could be further increased, particularly those involving extended generational differences.

Method	**BB**	**SS**	**SIBS**	**FD**	**MD**	**FS**	**MS**	**GFGD**	**GMGD**	**GFGS**	**GMGS**	**Average**
Vuvko[Shadrikov (2020)]	0.80	0.80	0.77	0.75	0.78	0.81	0.74	0.78	**0.76**	0.69	0.60	0.780
DeepBlueAI[Luo et al. (2020)]	0.77	0.77	0.75	0.74	0.75	0.81	0.74	0.72	0.67	**0.73**	**0.68**	0.760
Ustc-nelslip[Yu et al. (2020)]	0.75	0.74	0.72	**0.76**	0.75	**0.82**	0.75	**0.79**	**0.76**	0.69	0.67	0.760
SupCL [Zhang et al. (2021b)]	0.80	**0.81**	**0.79**	0.75	0.78	0.81	0.76	0.78	0.74	0.65	0.63	0.790
Ours	**0.81**	**0.81**	0.78	**0.76**	**0.79**	**0.82**	**0.77**	0.77	0.70	0.68	0.62	**0.792**

Table 5.3: Comparison of kinship verification accuracy across various methods.

Conclusion

This research explores the potential of Generative Adversarial Networks (GANs) for addressing a critical challenge in facial kinship verification: accounting for age-related variations that obscure familial similarities. Using a GAN-based facial age transformation model, this work successfully simulates aging and rejuvenation processes while preserving individual identity. The age-transformed images were then integrated into kinship verification systems to enhance their robustness and accuracy.

The key contributions of this work include a comprehensive review of both the kinship verification field and the age transformation field, and also the development of a GAN-based Encoder-Decoder Age Transformation Model which was designed to generate realistic age-progressed and age-regressed facial images in low resolution, maintaining identity consistency, addressing a common limitation of earlier approaches. The use of the age transformation model in augmentation on existing kinship datasets with synthetic age-transformed images enriches training data diversity, which may overcome the scarcity of labeled data for age-diverse kinship pairs. The augmented datasets allowed the kinship verification model to generalize better across age differences, slightly improving accuracy.

The experimental results show how using age-transformed images can affect kinship verification, demonstrating improvements that can be further enhanced by the development of both age transformation and classification methods. These results highlight the importance of addressing age as a variable in kinship recognition tasks and underscore the potential of GANs in the advancement of biometric applications. In conclusion, this research provides a promising direction for improving kinship verification systems by bridging age-related gaps. The findings open opportunities for further exploration, including refining the GAN model for more nuanced transformations, applying similar techniques to other biometric tasks, and leveraging additional data modalities like videos and audio for more comprehensive kinship recognition systems.

Bibliography

A., M. (2020). entacular faces: Race and the return of the phenotype in forensic identification. Am Anthropol.

Adini, Y., Moses, Y., and Ullman, S. (1997). Face recognition: The problem of compensating for changes in illumination direction. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):721–732.

Agarap, A. F. (2018). Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*.

Ahonen, T. (2004). Face recognition with local binary patterns. ECCV.

Alirezazadeh, P., Fathi, A., and Abdali-Mohammadi, F. (2015). A genetic algorithm-based feature selection for kinship verification. *IEEE Signal Processing Letters*, 22(12):2459–2463.

Berg, A. C. and Justo, S. C. (2003). Aging of orbicularis muscle in virtual human faces. In *Proceedings on Seventh International Conference on Information Visualization, 2003. IV 2003.*, pages 164–168. IEEE.

Bešenić, K., Ahlberg, J., and Pandžić, I. S. (2022). Picking out the bad apples: unsupervised biometric data filtering for refined age estimation. *The Visual Computer*, pages 1–19.

Bińkowski, M., Sutherland, D. J., Arbel, M., and Gretton, A. (2018). Demystifying mmd gans. *arXiv preprint arXiv:1801.01401*.

Bordallo Lopez, M., Hadid, A., Boutellaa, E., Goncalves, J., Kostakos, V., and Hosio, S. (2018). Kinship verification from facial images and videos: human versus machine. *Machine Vision and Applications*, 29:1–18.

Bottou, L. (2012). Stochastic gradient descent tricks. pages 421–436.

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., and Shah, R. (1994). Signature verification using a "siamese" time delay neural network. In *Advances in Neural Information Processing Systems*, volume 6, pages 737–744.

Chandaliya, P. K. and Nain, N. (2022). Childgan: Face aging and rejuvenation to find missing children. *Pattern Recognition*, 129:108761.

Chandaliya, P. K., Sinha, A., and Nain, N. (2020). Childface: Gender aware child face aging. In *2020 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5. IEEE.

Chen, B.-C., Chen, C.-S., and Hsu, W. H. (2014). Cross-age reference coding for age-invariant face recognition and retrieval. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI 13*, pages 768–783. Springer.

Chen, K., Yao, L., Zhang, D., Wang, X., Chang, X., and Nie, F. (2019). A semisupervised recurrent convolutional attention model for human activity recognition. *IEEE transactions on neural networks and learning systems*, 31(5):1747–1756.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1):53–65.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893 vol. 1.

DeBruine, L. M., Smith, F. G., Jones, B. C., Craig Roberts, S., Petrie, M., and Spector, T. D. (2009). Kin recognition signals in adult faces. *Vision Research*, 49(1):38–43.

Deng, J., Guo, J., Xue, N., and Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699.

Despois, J., Flament, F., and Perrot, M. (2020). Agingmapgan (amgan): High-resolution controllable face aging with spatially-aware conditional gans. In *European Conference on Computer Vision*, pages 613–628. Springer.

Fang, R., Tang, K. D., Snavely, N., and Chen, T. (2010). Towards computational models of kinship verification. In *2010 IEEE International Conference on Image Processing*, pages 1577–1580.

Fu, Y., Hospedales, T. M., Xiang, T., Xiong, J., Gong, S., Wang, Y., and Yao, Y. (2016). Robust subjective visual property prediction from crowdsourced pairwise labels. In *IEEE TPAMI*.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., and Chen, T. (2015). Recent advances in convolutional neural networks. *ArXiv*, abs/1512.07108.

Guo, G., Fu, Y., Huang, T. S., and Dyer, C. R. (2008). Locally adjusted robust regression for human age estimation. In *2008 IEEE Workshop on Applications of Computer Vision*, pages 1–6. IEEE.

Guo, Y., Su, X., Yan, G., Zhu, Y., and Lv, X. (2024). Age transformation based on deep learning: a survey. *Neural Computing and Applications*, 36(9):4537–4561.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Hernandez-Ortega, J., Galbally, J., Fierrez, J., Haraksim, R., and Beslay, L. (2019). Faceqnet: Quality assessment for face recognition based on deep learning. In *2019 International Conference on Biometrics (ICB)*, pages 1–8. IEEE.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.

Hore, A. and Ziou, D. (2010). Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE.

Huang, D., Shan, C., Ardabilian, M., Wang, Y., and Chen, L. (2011). Local binary patterns and its application to facial image analysis: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(6):765–781.

Huang, G. B., Lee, H., and Learned-Miller, E. (2012). Learning hierarchical representations for face verification with convolutional deep belief networks. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2518–2525.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976.

Jain, A., Bhagat, N., Srivastava, V., Tyagi, P., and Jain, P. (2020). A feature-based kinship verification technique using convolutional neural network. *Lecture Notes in Electrical Engineering*, 612:353 – 362.

Jang, W., Chhabra, A., and Prasad, A. (2017). Enabling multi-user controls in smart home devices. In *Proceedings of the 2017 Workshop on Internet of Things Security and Privacy*, page 49–54, New York, NY, USA. Association for Computing Machinery.

Jo, B., Cho, D., Park, I. K., and Hong, S. (2023). Ifqa: interpretable face quality assessment. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 3444–3453.

Kaminski, G., Dridi, S., Graff, C., and Gentaz, E. (2009). Human ability to detect kinship in strangers' faces: effects of the degree of relatedness. *Proc. Biol. Sci.*, 276(1670):3193–3200.

Kanwisher, N., McDermott, J., and Chun, M. M. (2002). The fusiform face area: a module in human extrastriate cortex specialized for face perception.

Karras, T., Laine, S., and Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020). Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119.

Kingma, D. P. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Kohli, N., Yadav, D., Vatsa, M., Singh, R., and Noore, A. (2019). Supervised mixed norm autoencoder for kinship verification in unconstrained videos. *IEEE Transactions on Image Processing*, pages 1329–1341.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, page 1097–1105. Curran Associates Inc.

Kuo, T.-H., Jia, Z., Kuo, T.-W., and Hu, J. (2023). Bitrackgan: Cascaded cyclegans to constraint face aging. *arXiv preprint arXiv:2304.11313*.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., and Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, 2.

Li, W., Wang, S., Lu, J., Feng, J., and Zhou, J. (2021). Meta-mining discriminative samples for kinship verification. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16130–16139. IEEE Computer Society.

Liu, S., Sun, Y., Zhu, D., Bao, R., Wang, W., Shu, X., and Yan, S. (2017). Face aging with contextual generative adversarial nets. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 82–90.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60:91–110.

Lu, J., Hu, J., Zhou, X., Shang, Y., Tan, Y.-P., and Wang, G. (2012a). Neighborhood repulsed metric learning for kinship verification. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2594–2601.

Lu, J., Hu, J., Zhou, X., Zhou, J., Castrillón-Santana, M., Lorenzo-Navarro, J., Kou, L., Shang, Y., Bottino, A., and Vieira, T. F. (2014a). Kinship verification in the wild: The first kinship verification competition. In *IEEE International Joint Conference on Biometrics*, pages 1–6.

Lu, J., Zhou, X., Tan, Y.-P., Shang, Y., and Zhou, J. (2012b). Neighborhood repulsed metric learning for kinship verification. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2594–2601.

Lu, J., Zhou, X., Tan, Y.-P., Shang, Y., and Zhou, J. (2014b). Neighborhood repulsed metric learning for kinship verification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 331–345.

Lu, J., Zhou, X., Tan, Y.-P., Shang, Y., and Zhou, J. (2014c). Neighborhood repulsed metric learning for kinship verification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(2):331–345.

Luo, Z., Zhang, Z., Xu, Z., and Che, L. (2020). Challenge report recognizing families in the wild data challenge. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pages 868–871. IEEE.

Makhmudkhujaev, F., Hong, S., and Park, I. K. (2021). Re-aging gan: Toward personalized face age transformation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3908–3917.

Mao, X., Li, Q., Xie, H., Lau, R. Y., Wang, Z., and Paul Smolley, S. (2017). Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2794–2802.

Mirza, M. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.

Moschoglou, S., Papaioannou, A., Sagonas, C., Deng, J., Kotsia, I., and Zafeiriou, S. (2017). Agedb: the first manually collected, in-the-wild age database. In *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 51–59.

Ou, F.-Z., Chen, X., Zhang, R., Huang, Y., Li, S., Li, J., Li, Y., Cao, L., and Wang, Y.-G. (2021). Sdd-fiqa: Unsupervised face image quality assessment with similarity distribution distance. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7670–7679.

Rexbye, H., Petersen, I., Johansens, M., Klitkou, L., Jeune, B., and Christensen, K. (2006). Influence of environmental factors on facial ageing. *Age and Ageing*, 35(2):110–115.

Ricanek, K. and Tesafaye, T. (2006). Morph: A longitudinal image database of normal adult age-progression. In *7th international conference on automatic face and gesture recognition (FGR06)*, pages 341–345. IEEE.

Robinson, J. P., Khan, Z., Yin, Y., Shao, M., and Fu, Y. (2021). Families in wild multimedia: A multimodal database for recognizing kinship. *IEEE Transactions on Multimedia*, 24:3582–3594.

Robinson, J. P., Shao, M., Wu, Y., and Fu, Y. (2016). Families in the wild (fiw): Large-scale kinship image database and benchmarks. In *Proceedings of the 24th ACM International Conference on Multimedia*, page 242–246, New York, NY, USA. Association for Computing Machinery.

Robinson, J. P., Shao, M., Wu, Y., Liu, H., Gillis, T., and Fu, Y. (2018). Visual kinship recognition of families in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(11):2624 – 2637.

Robinson, J. P., Yin, Y., Khan, Z., Shao, M., Xia, S., Stopa, M., Timoner, S., Turk, M. A., Chellappa, R., and Fu, Y. (2020). Recognizing families in the wild (rfiw): The 4th edition. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pages 857–862.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.

Serengil, S. I. and Ozpinar, A. (2020). Lightface: A hybrid deep face recognition framework. In *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pages 23–27. IEEE.

Serengil, S. I. and Ozpinar, A. (2021). Hyperextended lightface: A facial attribute analysis framework. In *2021 International Conference on Engineering and Emerging Technologies (ICEET)*, pages 1–4. IEEE.

Serengil, S. I. and Ozpinar, A. (2024). A benchmark of facial recognition pipelines and co-usability performances of modules. *Bilisim Teknolojileri Dergisi*, 17(2):95–107.

Shadrikov, A. (2020). Achieving better kinship recognition through better baseline. In *2020 15th IEEE international conference on automatic face and gesture recognition (FG 2020)*, pages 872–876. IEEE.

Shu, X., Xie, G.-S., Li, Z., and Tang, J. (2016). Age progression: Current technologies and applications. *Neurocomputing*, 208:249–261.

Song, J., Zhang, J., Gao, L., Liu, X., and Shen, H. T. (2018). Dual conditional gans for face aging and rejuvenation. In *IJCAI*, pages 899–905.

Song, J., Zhang, J., Gao, L., Zhao, Z., and Shen, H. T. (2021). Agegan++: Face aging and rejuvenation with dual conditional gans. *IEEE Transactions on Multimedia*, 24:791–804.

Wang, H., Raj, B., and Xing, E. (2017). On the origin of deep learning. *ArXiv*, abs/1702.07800.

Wang, W., You, S., Karaoglu, S., and Gevers, T. (2023). A survey on kinship verification. *Neurocomputing*, 525:1–28.

Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.

Wang, Z., Tang, X., Luo, W., and Gao, S. (2018). Face aging with identity-preserved conditional generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7939–7947.

Wu, X., Feng, X., Cao, X., Xu, X., Hu, D., López, M. B., and Liu, L. (2022). Facial kinship verification: A comprehensive review and outlook. *Int. J. Comput. Vis.*, 130(6):1494–1525.

Wu, X., Granger, E., and Feng, X. (2019). Audio-visual kinship verification. *arXiv preprint arXiv:1906.10096*.

Xia, S., Shao, M., and Fu, Y. (2011). Kinship verification through transfer learning. In *Twenty-second international joint conference on artificial intelligence*. Citeseer.

Xiao, T., Xu, Y., Yang, K., Zhang, J., Peng, Y., and Zhang, Z. (2015). The application of two-level attention models in deep convolutional neural network for fine-grained image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 842–850.

Yang, H., Huang, D., Wang, Y., and Jain, A. K. (2018). Learning face age progression: A pyramid architecture of gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 31–39.

Yao, X., Puy, G., Newson, A., Gousseau, Y., and Hellier, P. (2021). High resolution face age editing. In *2020 25th International conference on pattern recognition (ICPR)*, pages 8624–8631. IEEE.

Yu, J., Li, M., Hao, X., and Xie, G. (2020). Deep fusion siamese network for automatic kinship verification. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pages 892–899. IEEE.

Zhang, K., HUANG, Y., SONG, C., et al. (2015). Kinship verification with deep convolutional neural networks. british machine vision conference. *British: BMVA*, pages 148–1.

Zhang, L., Duan, Q., Zhang, D., Jia, W., and Wang, X. (2021a). Advkin: Adversarial convolutional network for kinship verification. *IEEE Transactions on Cybernetics*, 51(12):5883–5896.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595.

Zhang, X., Min, X., Zhou, X., and Guo, G. (2021b). Supervised contrastive learning for facial kinship recognition. In *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pages 01–05. IEEE.

Zhang, Z., Song, Y., and Qi, H. (2017). Age progression/regression by conditional adversarial autoencoder. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5810–5818.

Zhou, X., Lu, J., Hu, J., and Shang, Y. (2012). Gabor-based gradient orientation pyramid for kinship verification under uncontrolled environments. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 725–728.

Zhu, H., Huang, Z., Shan, H., and Zhang, J. (2020). Look globally, age locally: Face aging with an attention mechanism. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1963–1967.