FACE RECOGNITION: HOW EASY IS IT TO MATCH NEWBORNS OR TODDLERS TO THEIR ADULTSELVES

Ashish Goli

A Capstone Project (or Thesis) Submitted to the
University of North Carolina Wilmington in Partial Fulfillment
of the Requirements for the Degree of
Master of Science

Department of Computer Science
Department of Information Systems and Operations Management

University of North Carolina Wilmington

2024

Approved by

Advisory Committee

Minoo Modaresnezhad

Ahmed ElSaid

Karl Ricanek Jr., Chair

TABLE OF CONTENTS
(Insert Automatic Table of Contents)

ABSTRACT

Face Recognition: How easy is it to match newborns or toddlers to their adult selves. Goli Ashish, 2024. Capstone Paper, University of North Carolina Wilmington.

In the field of facial recognition, achieving temporal invariance represents a very significant challenge. Facial recognition technology has demonstrated remarkable abilities to identify people at various life milestones thanks to the recent developments in machine learning. Nevertheless, the task is to synchronize infants and children with their adolescent counterparts, as facial features undergo constant changes at this stage. In this project, newborns and toddlers relate to their future adult facial profiles using the latest open-source face recognition algorithms.

Although the accuracy is very high in adult facial recognition, the morphological changes in the children pose some unique challenges. Remarkable strides in the performance and accuracy of face recognition technology have largely been directed towards adults, with a notable gap in the development of systems specifically designed for children and sub-adults. Most existing face recognition systems are tailored for adult faces, leaving a limited scope for the application of these advances to younger age groups. Children are always in a state of growth [1][2][3], and with respect to face recognition, this is their ever-changing youthful appearance. Physical and textural changes of the face in children baffle current technology [4]. The development of newborns and toddlers is characterized by quick facial changes such as the changes in proportions, the formation of expressions, and the evolution of features. It provides a great challenge in the training of highly accurate models based on the limited datasets for this specific demographic. This research aims to assess the viability of recognizing faces from childhood to adulthood by evaluating verification performance. Verification, one of the two face matching methods, is considered less challenging as it involves comparing only two face images to determine if they match or not. To address these challenges, this study utilizes open-source algorithms and extends the application of facial recognition to encompass the intricate developmental stages from infancy to maturity.

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1: INTRODUCTION

As one of the more significant aspects of biometric technology, face recognition plays a role in terms of innovation and use across such systems as security systems to personal devices. The efficiency of facial recognition in matching an adult with their younger counterpart is the dominant subject of this research, where the landscape becomes much more complex by extending the analysis to newborns and toddlers. This research delves into the intriguing question: “How difficult is it to match newborns or toddlers to their adult selves?” Exploring the complexities of facial development from infancy to old age not only advances facial recognition systems but also unveils the complex interplay of genetic and environmental factors in shaping an individual's distinctive facial identity.

From the infant stage to fully mature, humanity is mirrored through a completely different form of face-line representation. Though researchers have critiqued facial recognition algorithms in the process of matching adults from different age groups [5], the case is much more complicated when dealing with the craniofacial complex that is still developing features, i.e., newborns and toddlers as shown in Figure 1. These are the first interactions with the world which lay the groundwork for changes that continue to develop and become final. This research aims at navigating the challenges involved in detecting face patterns at the initial stages of life, to see if it is possible to link such recognizable traits with a person’s appearance in adult years.

Figure 1: Craniofacial changes from infants to adults [13]

This research has more importance than just the technological world; it goes into the fields of mind, genetics, and child development. Conceptualization of the stability and development of morphological features within the early years could reveal valuable information regarding the inheritance of facial morphology and the role of environmental factors in facial development. In addition, the results of this study provide implications for the fields of forensic sciences, where accurate face recognition goes beyond its modern applications by offering the possibility to identify missing persons whose appearance has changed dramatically from childhood.

So, to understand the multifarious issues around the matching of babies or toddlers to their adult selves, we need first to go through the current literature on the development and nature of facial recognition technology, its challenges, and its uses. Following this, an exploration of the world of facial development will be made to

distinguish the genetic and environmental roles that contributed to the constant changes in the face of man. The nexus between technology and developmental science will be discussed alongside the current state of face recognition algorithms and their accuracy in detecting subtle changes in facial transformations.

In the unexplored land of linking the dots in childhood to the adulthood, this research aims to fill the voids in the literature, presenting a complete picture of the complexities of matching with a high degree of accuracy of newborns or toddlers to their adult selves. By doing so, we set out on a voyage that does not conform to the demarcations of disciplines, as technology, genetics, and developmental psychology intersect to unravel the deep-seated secrets etched on the human face.

1.1 Business Use-cases

Missing Persons Identification:

Accurate face recognition technology for infants and toddlers could significantly aid in identifying missing persons who have undergone substantial facial changes since childhood. This has crucial implications for law enforcement and humanitarian efforts, potentially reuniting families separated over time.

Child Protection and Security:

Implementing reliable facial recognition for young children can enhance child protection measures. It could be utilized in childcare facilities, schools, and public spaces to ensure the safety of children and prevent unauthorized access or potential security threats.

Social Media:

Extending facial recognition technology to match newborns and toddlers with their adult selves holds promise for enhanced family-oriented social media experiences. Parents can effortlessly create digital timelines for their children, capturing and sharing precious

moments while seamlessly transitioning from infancy to adulthood, fostering a more comprehensive and personalized social media engagement.

1.2 Technological Challenges

Age progression poses a considerable challenge in accurately predicting the physical transformation from newborn to adult. This complexity arises from the intricate changes in facial features driven by bone growth, muscle development, and fat distribution [1]. Compounding this issue is the limited availability of labeled newborn face data, making it difficult to create a diverse dataset over time due to privacy concerns and logistical challenges. Furthermore, variations in illumination and pose during newborn photography introduce additional hurdles, impacting the precision of facial features matching with adult faces. The minimal expressions exhibited by newborns compared to the wide range of emotions displayed by adults add another layer of intricacy, making the task of matching neutral newborn faces with expressive adult faces highly complex.

1.3 Project Motive

This collaborative project with Virginia Commonwealth University's School of Medicine addresses a critical gap in pediatric research funding by the NIH. Despite the potential benefits, funding for training tools development in pediatrics is contingent upon demonstrating that automatic face recognition cannot match newborns and toddlers to their adult faces (≥ 18 years old). Our thesis project asserts that such matching is not feasible, providing a foundation for further investigation and tool development.

CHAPTER 2: REVIEW OF LITERATURE REVIEW AND ANALYSIS

Face recognition technology has evolved significantly in recent years, with applications ranging from security systems to social media platforms. A fascinating aspect of this technology is its potential to match individuals across different life stages, specifically addressing the challenge of recognizing newborns and toddlers as they mature into adults. This comprehensive, but not exhaustive literature review explores various studies that contribute to our understanding of the complexities and advancements in face recognition, particularly concerning the matching of facial features from infancy to adulthood.

Algorithmic bias is a prominent theme in the literature on face recognition. Notable studies by N. Srinivas et al. [6] and U. Park et al. [7] emphasize the existence of performance differences when processing images of children and adults, highlighting potential biases in face recognition algorithms. This bias poses challenges in accurately recognizing facial features across age groups, with implications for the reliable matching of individuals from infancy to adulthood. The need to mitigate algorithmic bias is further underscored by longitudinal studies conducted by Deb et al. [8] and Ricanek et al. [9], which delve into the evolving nature of facial features during growth, emphasizing the importance of unbiased algorithms in capturing the nuances of age-related changes.

Addressing age-related variations in facial appearance, researchers have proposed age-invariant face recognition approaches. Park et al. [7] and Zhao et al. [10] present methodologies aimed at making face recognition systems more robust to changes in facial features over time. These studies recognize the dynamic nature of facial morphology and seek to enhance the accuracy of matching individuals across different life stages. The pursuit of age-invariant recognition reflects a crucial aspect of face recognition

technology, as it navigates the challenges posed by the natural progression of facial features from infancy to adulthood.

Factors influencing face recognition in children are explored by Hood et al. [11], who investigate the impact of gaze direction on face recognition in both children and adults. Their findings highlight the role of non-static facial cues in the recognition process, emphasizing the need to consider additional factors beyond static facial features. This is particularly relevant in the case of infants and toddlers whose facial features may undergo substantial changes over time. Understanding these factors contributes to the development of more nuanced face recognition systems that can adapt to age-specific characteristics.

The longitudinal studies conducted by Deb et al. [8] and Ricanek et al. [9] offer valuable insights into the dynamic changes in facial features as individuals transition from infancy to adulthood. These studies involve the systematic tracking of facial characteristics over extended periods, allowing for a detailed examination of how facial features evolve with age. The findings from these longitudinal analyses contribute significantly to our understanding of the challenges inherent in recognizing children's faces and provide a solid foundation for the development of face recognition systems capable of accurately matching individuals at various stages of their life trajectories. By analyzing the nuanced transformations in facial features over time, these longitudinal studies enhance the potential for creating more effective and adaptable face recognition algorithms tailored to the unique aspects of each developmental stage.

The integration of deep learning techniques adds another layer to the literature on face recognition. Oo and Oo [12] explore the application of deep learning for child face recognition, demonstrating its potential in addressing the unique challenges posed by the

facial features of young individuals. Deep learning, with its ability to automatically learn hierarchical representations, offers a promising avenue for improving the performance of face recognition systems across diverse age groups. The utilization of deep learning further advances the field by providing more sophisticated tools for recognizing faces with a higher degree of accuracy.

In summary, the literature on face recognition, encompassing algorithmic bias, age-invariant recognition, longitudinal analyses, and deep learning applications, underscores the complexities and advancements in the field. Algorithmic bias poses challenges to the accurate recognition of faces across different age groups, emphasizing the need for unbiased algorithms. The pursuit of age-invariant recognition acknowledges the dynamic nature of facial features over time, seeking to enhance the robustness of face recognition systems. Factors influencing face recognition in children, such as gaze direction, highlight the importance of considering non-static facial cues. Longitudinal studies provide insights into the evolving nature of facial features, contributing to the development of accurate face recognition systems. The integration of deep learning techniques showcases the potential for sophisticated approaches in addressing the unique challenges of recognizing faces, especially in the case of young individuals.

As technology progresses, ongoing research in these areas will contribute to the refinement of face recognition algorithms, making them more adept at accommodating the diverse and dynamic nature of human facial features across the lifespan. The collective findings from these studies provide a foundation for the continued advancement of face recognition technology, with the goal of achieving accurate and reliable matching of newborns and toddlers to their adult selves.

CHAPTER 3: RELATED WORK

This section delves into the existing research on child face recognition, highlighting the limited scope of studies and identifying areas for improvement. The earliest investigation into child face recognition with longitudinal images was conducted by Ricanek et al. [14], while Bharadwaj et al. [15] pioneered research in this field with a dataset of non-longitudinal faces. Ricanek et al. explored the challenge of child face recognition across developmental stages, focusing on childhood, preadolescence, and adolescence. They evaluated the performance of various face recognition systems, including a commercial one and several hand-engineered systems circa 2014, using a longitudinal child corpus called In The Wild Child Celebrate (ITWCC). The ITWCC dataset was aggregated images from public sources of child celebrities, the images had variations in lighting, pose, expression, and age. These systems included PCA, LDA, LRPCA, cohort-LDA, and OpenBR, an open-source face recognition system prevalent at the time. Conversely, Bharadwaj et al. investigated the application of face recognition for newborn identification in birthing hospitals, aiming to prevent baby switching or identify abducted infants. Their approach involved combining SURF and multi-level via Gaussian Pyramids LBP techniques for template generation and matching. The study was conducted on a small dataset comprising 34 newborns, with multiple images captured within 2 to 15 hours post-birth.

Best-Rowden et al. [16] conducted an evaluation of facial recognition (FR) algorithms aimed at identifying newborns and toddlers. They specifically assessed the performance of a Commercial Off-The-Shelf (COTS) algorithm using the Newborns, Infants, and Toddlers (NITL) dataset, which comprises face images of 314 children aged 0 to 4 years. The dataset offers a wide range of variations in lighting, pose, and expression across the

images. Their study revealed that the efficacy of the COTS algorithm declines with increasing age spans between the gallery and probe images. Notably, the algorithm exhibited high verification accuracy for images captured within the same session, but this accuracy degraded when comparing images from different sessions. This research underscores the performance of a single COTS algorithm specifically tailored for newborns and infants.

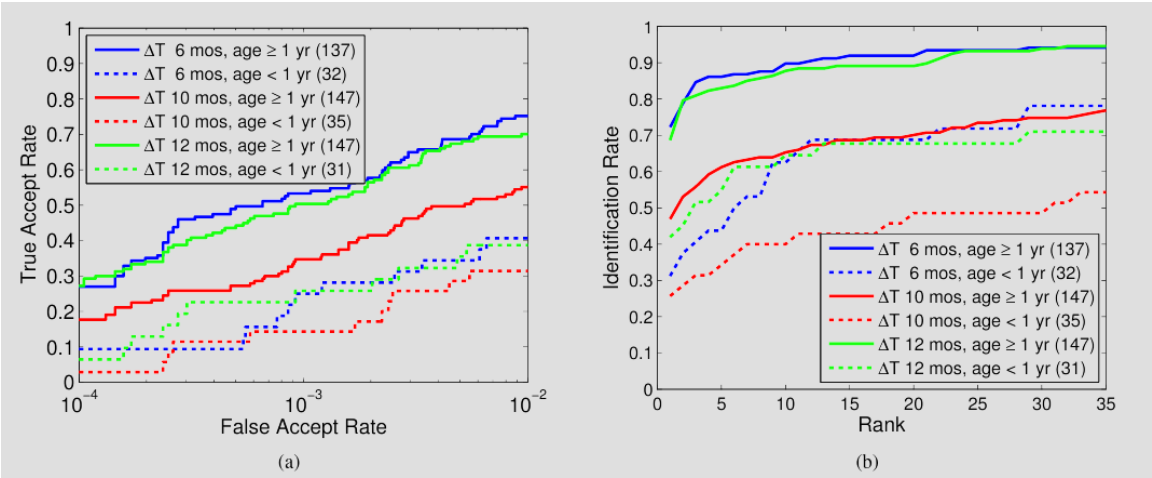

Figure 2: Longitudinal performance of COTS-A for (a) verification and(b) identification experiments of the 206 total subjects in the NIT-S1 subset of the NITL face image database. The gallery for these experiments consists of face images from Session1, and the probe sets are face images fromSessions2,3, and4(with different elapsed times,4T). Performance is further broken down by age at Session 1 enrollment (youngerthan1yearoldvs.at least1yearold). The numbers in parentheses indicate the number of subjects(outof206) that returned for each of the sub sequent sessions. The number of impostors cores in(a) is150,881scores, and the gallery size in(b) is206subjects.

CHAPTER 4: METHODOLOGY

In this section, I will provide details on the face matchers used for this work. The face matchers were selected based on their performance in the open-source space. The models used in this research are: ArcFace and VGG-Face. Further, I provide details on the structure of the matchers, which are implemented in a deep learning framework.

4.1 ArcFace

The ArcFace method for facial recognition offers a sophisticated approach to matching newborns to their adult selves, leveraging deep learning techniques to achieve accurate and reliable results. By utilizing a ResNet34-based architecture, the model can effectively extract and analyze facial features across different stages of development, facilitating the comparison between newborn and adult faces.

At the core of the ArcFace method lies the ability to represent faces in a high-dimensional embedding space, where facial features are encoded into compact and discriminative vectors. This embedding space enables robust facial recognition by ensuring that similar faces are mapped closer together while maintaining distinctiveness between different individuals. Consequently, when matching newborn faces to their adult counterparts, the model can effectively capture the underlying facial similarities despite the significant morphological changes that occur over time.

The ArcFace model architecture incorporates intricate layers, including convolutional, batch normalization, and activation layers, designed to extract hierarchical features from input facial images. By leveraging the ResNet34 backbone, the model can capture both low-level details and high-level facial characteristics, facilitating accurate representation and comparison of newborn and adult faces. Additionally, the inclusion of skip

connections within residual blocks enhances gradient flow during training, enabling efficient learning and optimization of the model parameters.

The ArcFace method excels in handling variations in facial appearance due to age progression, illumination changes, and other environmental factors. Through its robust feature extraction capabilities and metric learning objectives, the model can effectively distinguish between subtle facial differences while mitigating the effects of intra-class variations. This may enable reliable matching of newborn faces to their adult selves, even in challenging real-world scenarios where facial expressions and conditions may vary significantly.

4.1.1 ArcFace Implementation

ArcFace is a facial recognition model implemented using deep learning techniques. It begins with the necessary imports, including libraries for handling files, downloading data, and logging information. Additionally, it imports components from the deepface package for model implementation and utility functions.

The ArcFaceClient class serves as the main interface for utilizing the ArcFace model. It is inherited from the Facial Recognition class, implying that it encapsulates functionality related to facial recognition tasks. Within this class, an instance of the ArcFace model is initialized along with relevant specifications such as the input and output shape dimensions.

The load_model function is responsible for constructing the ArcFace model architecture and loading pre-trained weights. It builds the model architecture by assembling layers according to the ResNet34 architecture, incorporating convolutional, batch normalization, dropout, and dense layers. Pre-trained weights are downloaded from a specified URL if they are not already present locally, and then loaded into the model.

The ResNet34 function defines the architecture for the ResNet34 model, which serves as the backbone for the ArcFace model. It constructs a series of convolutional blocks, each containing convolutional, batch normalization, and PReLU activation layers. These blocks are stacked together to form the ResNet34 model.

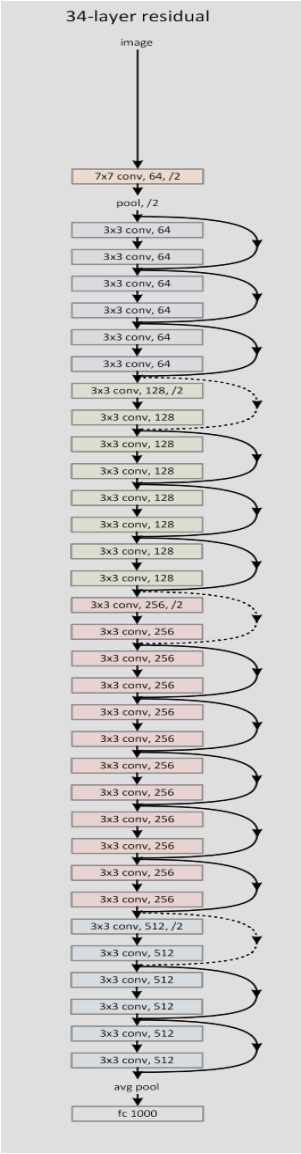

Figure 3[17]: ResNet34 Architecture

The block1 and stack1 functions are helper functions used within the ResNet34 function to define individual residual blocks and stacks of blocks, respectively. Each block

consists of convolutional layers with batch normalization and activation functions, as well as skip connections to facilitate gradient flow during training.

4.2 VGG-Face

The VGG-Face method presents a compelling approach to the challenging task of matching newborns to their adult selves in facial recognition scenarios. This method harnesses the power of deep convolutional neural networks, specifically the VGG-Face model, which has been pre-trained on a vast dataset containing facial images from a diverse range of individuals across various ages. While the model's primary purpose is facial classification, its architecture and learned representations can be effectively repurposed for the task of matching faces across different life stages.

At the core of the VGG-Face method lies the concept of facial embedding. These embeddings are compact numerical representations derived from facial images, capturing essential facial features in a format conducive to computational analysis and comparison. The VGG-Face model, with its deep layers of convolutional and pooling operations followed by fully connected layers, is adept at extracting hierarchical features from input images. These features encode both low-level details, such as texture and color, as well as high-level semantic attributes, such as facial contours and structures.

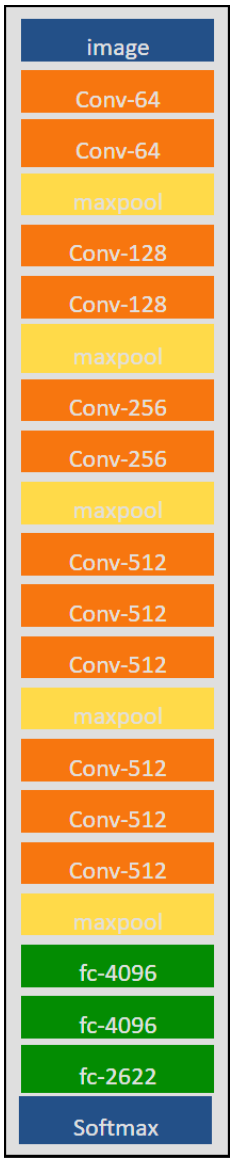

Figure 4[18]: VGG-Face Model

When applied to the task of matching newborns to their adult selves, the VGG-Face method exploits the inherent consistency in facial structures and characteristics that persist throughout an individual's life. Despite the substantial physical changes that occur from infancy to adulthood, certain underlying facial traits remain consistent and recognizable. By extracting and comparing these traits across different images, the

method can effectively identify matches between newborn faces and their corresponding adult counterparts.

An essential aspect of the VGG-Face method's effectiveness lies in its ability to generate consistent embeddings across disparate age groups. This is achieved through normalization techniques applied during the embedding generation process, which help mitigate variations in lighting conditions, poses, and facial expressions. By normalizing the input data, the method ensures that the extracted embeddings encapsulate invariant facial features relevant to the matching task, thereby enhancing the robustness and accuracy of the recognition process.

Practical implementation of the VGG-Face method for matching newborns to their adult selves involves several key steps. First, facial images of newborns and corresponding adults are collected and preprocessed to ensure consistency in terms of resolution, orientation, and quality. Next, the VGG-Face model is utilized to generate facial embeddings from these preprocessed images, capturing the essential facial characteristics of everyone at different life stages. Finally, a matching algorithm is applied to compare the generated embeddings and identify potential matches based on similarity scores or distance metrics.

4.2.1 VGG-Face Implementation

VGG-Face client class and related functions for generating embeddings and loading the VGG-Face model. This class inherits from the FacialRecognition class and implements methods for generating facial embeddings using the VGG-Face model.

The VggFaceClient class initializes by loading the VGG-Face model and setting attributes such as model name, input shape, and output shape. It implements a forward method to generate embeddings from input images. The method utilizes the loaded model

to compute embeddings, incorporating a normalization step to ensure consistent representation.

The base_model function defines the architecture of the VGG-Face model used for classification, which consists of a series of convolutional and pooling layers followed by fully connected layers. This model is designed to classify 2622 identities and is not used for generating embeddings.

The load_model function is responsible for loading the pre-trained weights of the VGG-Face model. It downloads the weights if they are not already available locally and loads them into the defined base model architecture. Additionally, it creates a modified version of the model that outputs 4096-dimensional embeddings instead of classification scores. This modified model is returned for use in generating embeddings from input images.

Throughout the code, dependencies such as TensorFlow/Keras, deepface library modules, and logging utilities are imported. Conditional imports are used to handle different TensorFlow versions, ensuring compatibility with both TensorFlow 1 and 2. The code follows object-oriented principles, encapsulating functionality within classes and functions for modularity and reusability. Additionally, comments are provided to explain key aspects of the implementation and rationale behind certain design choices.

CHAPTER 5: DATASET

Dataset Name	Age Range	# of Subjects	Availability
FG-NET [19]	0 yrs-69 yrs.	82	Public
ITWCC-3 [20]	2 yrs-80 yrs.	304	Public
Large Age-Gap [21]	2 yrs-77 yrs.	1,899	Public
Adience [22]	0 yrs-60+ yrs.	2,284	Public
CLF [23]	2 yrs-18 yrs.	919	Private

Table 1: Available Datasets

The number of child face databases available to the academic community is extremely limited. However, there are some datasets that as a consequence of the problem that they were developed to address, faces of children were included, e.g. FG-Net, Large Age-Gap (LAG), Adience, CLF, and ITWCC. The different datasets composed of child faces used in different studies associated with automatic face recognition are listed in Table 1. Out of the few publicly available datasets, they were filtered based on the requirements. (See Table 2) The requirements used to filter out the existing datasets are:

i. The subject must have at least two images at different ages to qualify.
ii. The subject must have an image of an infant (0yrs-2yrs.) or toddler (2yrs- 4 yrs.), and another image of adult (18+ yrs.).

Dataset Name	Age Range	Number of Subjects	Availability
FG-NET [19]	0 yrs-69 yrs.	74	Public
ITWCC-3 [20]	2 yrs-80 yrs.	42	Public
Large Age-Gap [21]	2 yrs-77 yrs.	140	Public

Table 2: Datasets used for experiments.

The FG-NET dataset is a collection of facial images depicting subjects across different ages, commonly used in age progression and regression research. It consists of various individuals captured in infancy and adulthood stages. Notably, 74 subjects within the dataset possess images spanning both infancy and adulthood, making it valuable for studying facial aging processes and developing age estimation algorithms.

Figure 5: FG-NET Infants[19]

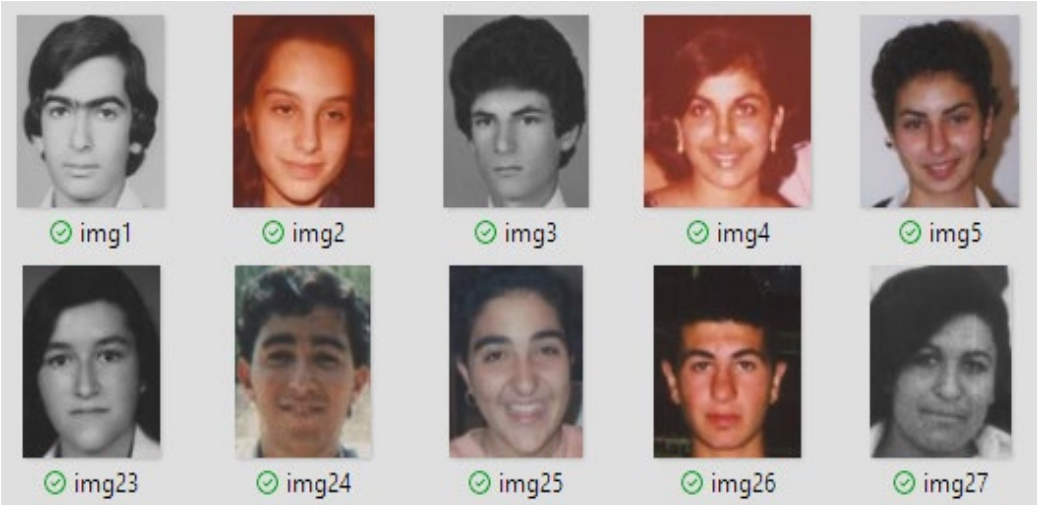

Figure 6: FG-NET Adults[19]

The ITWCC3 (In the Wild Child Celebrity Face Dataset) is a dataset containing images of celebrities' faces, encompassing both childhood and adulthood stages. It features 42 subjects, each with images representing their infancy and adulthood. This dataset serves as a valuable resource for research in age progression, facial recognition, and computer vision applications. Its diverse collection of subjects offers rich data for analysis and algorithm development.

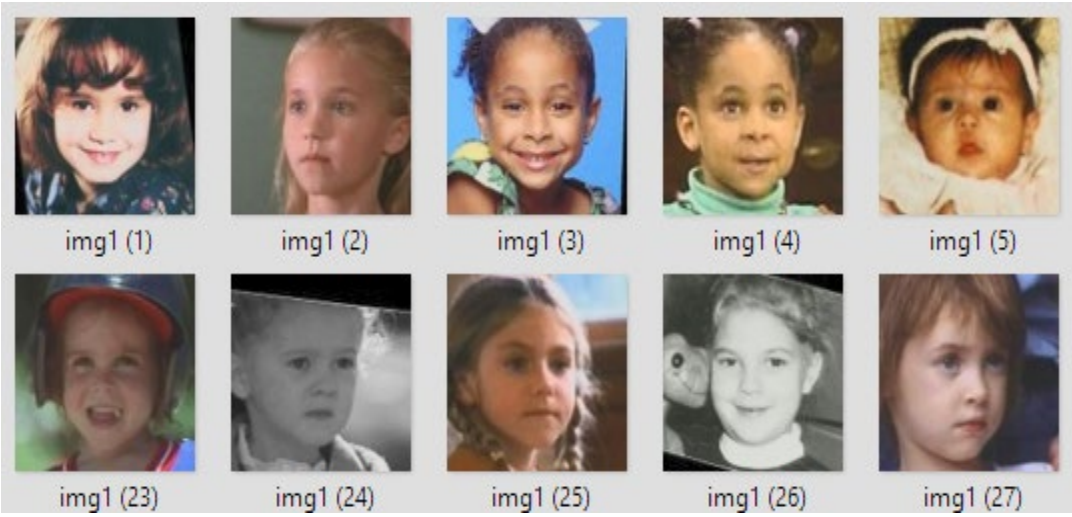

Figure 7: ITWCC-3 Infants[20]

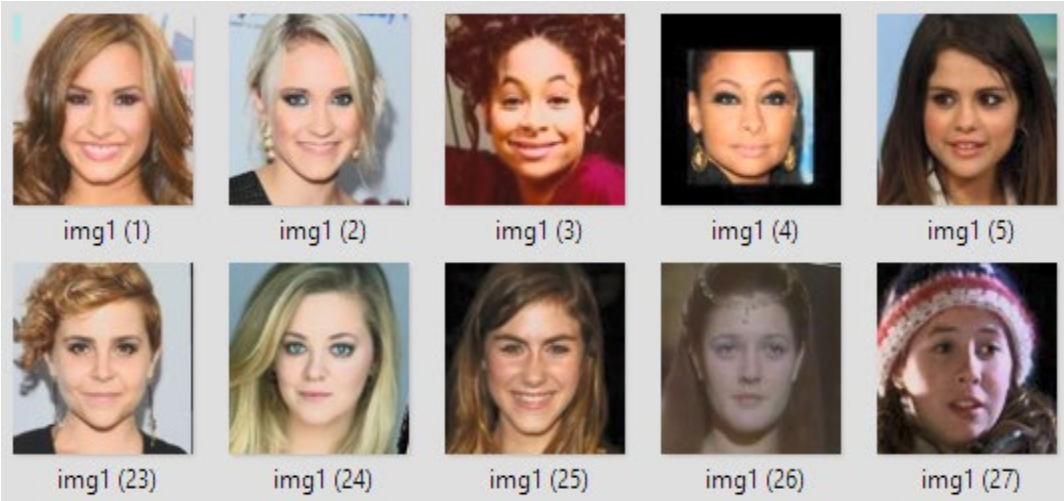

Figure 8: ITWCC-3 Adults[20]

The Large Age-Gap face dataset contains images of individuals at both infancy and adulthood stages, facilitating age progression research and facial recognition algorithm development. It encompasses 140 subjects, showcasing the evolution of facial features over time. This dataset is invaluable for studying age-related changes in facial appearance and enhancing understanding of facial recognition systems' performance across significant age gaps.

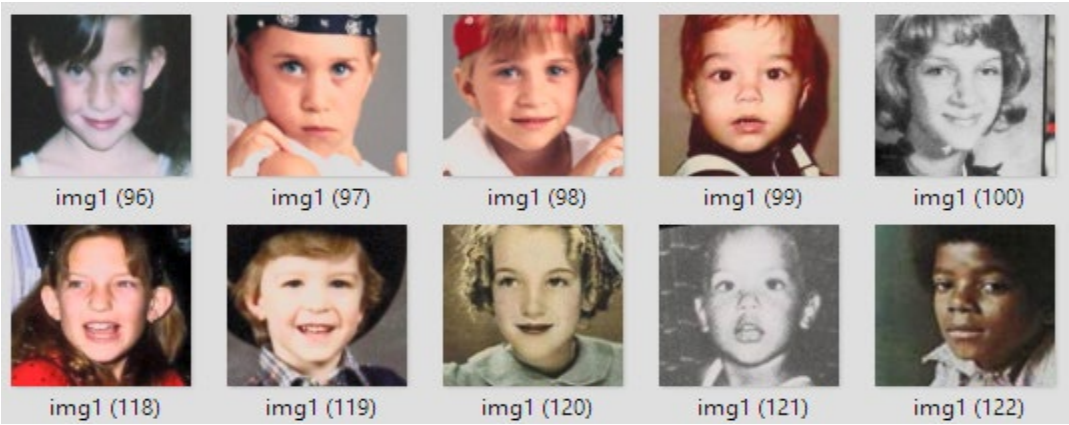

Figure 9: Large Age-Gap Infants[21]

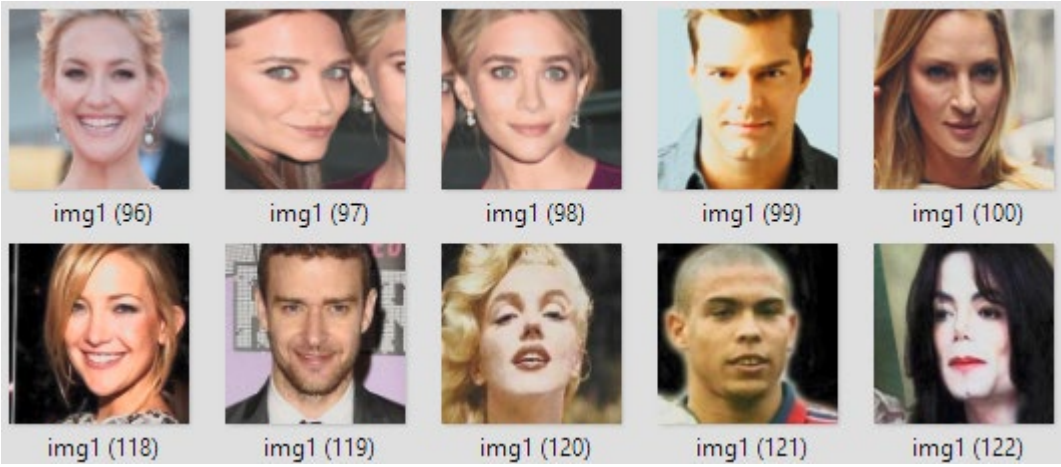

Figure 10: Large Age-Gap Adults[21]

CHAPTER 6: EXPERIMENTS AND RESULTS

6.1 Experiment

We evaluate the performance of the two algorithms under three different datasets, namely, FG-NET, ITWCC3, and Large Age-Gap datasets. In this research, the primary objective was to evaluate the efficacy of facial recognition algorithms (ArcFace and VGG-Face) in distinguishing between infant/toddler and adult selves of individuals. To achieve this, datasets comprising images of subjects at different developmental stages was meticulously curated, with images segregated into two distinct folders: "probe" containing infant/toddler images and "gallery" containing adult self-images. The separation of images facilitated the subsequent analysis, allowing for the comparison of facial features across developmental stages.

To quantify the performance of facial recognition algorithms, the deepface.verify() library was leveraged (as shown in Figure11), enabling the calculation of the distance between facial features for every possible combination of images within the probe and gallery folders (as shown in Figure 12). Each match was categorized as either a "Genuine match," denoting instances where the same individual's facial features matched across developmental stages, or an "Imposter match," indicating matches between different individuals.

Upon analyzing the distribution of matches at varying threshold values, it was observed that the number of genuine matches was notably lower compared to imposter matches. This discrepancy in match instances necessitated a normalization process to ensure equitable comparison between the two categories. Normalization was achieved by dividing the count of genuine and imposter matches by their respective total instances, yielding values within the range of 0 to 1.

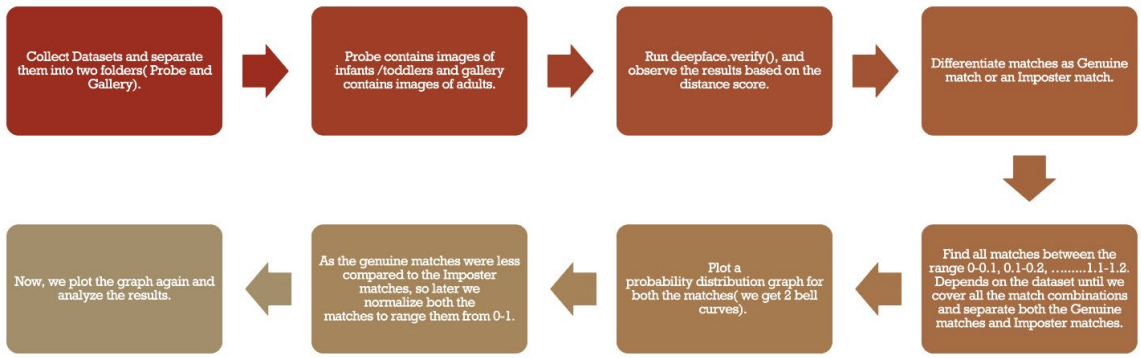

Figure11: Experiment Procedure

```python
from deepface import DeepFace

# Use the DeepFace.verify function
obj = DeepFace.verify(
    img1_path="C:\\Users\\19102\\OneDrive\\Desktop\\Images\\probe\\img1.JPG",
    img2_path="C:\\Users\\19102\\OneDrive\\Desktop\\Images\\gallery\\img1.JPG",
    model_name='VGG-Face',
    detector_backend='retinaface',
    enforce_detection=False
)

# Adjust the threshold value in the result object
threshold = 0.65  # Adjust the threshold value as needed

if obj['distance'] < threshold:
    print(f"The faces are verified with a distance score of {obj['distance']:.4f}.")
else:
    print(f"The faces are not verified. The distance score is {obj['distance']:.4f}.")

print(obj)
```

Figure12: Face matching code snippet using deepface.verify()

```
Out[2]: {'verified': False,
         'distance': 0.8177614859206891,
         'threshold': 0.68,
         'model': 'VGG-Face',
         'detector_backend': 'retinaface',
         'similarity_metric': 'cosine',
         'facial_areas': {'img1': {'x': 132, 'y': 77, 'w': 294, 'h': 390},
          'img2': {'x': 130, 'y': 73, 'w': 274, 'h': 367}},
         'time': 7.36}
```

Figure13: Sample output of the distance value

Following normalization, the data were reevaluated, and a subsequent graph was plotted to visualize the normalized distribution of genuine and imposter matches. This

iterative approach enabled a more nuanced understanding of the algorithm's performance, ensuring that any disparities in match instances were appropriately accounted for and facilitating a fair assessment of the algorithm's efficacy across developmental stages.

6.2 Probability Distribution graph

In the context of identity verification systems, a perfect probability distribution graph for genuine and impostor matches illustrates the ideal separation between the two classes. For example, Figure13 illustrates Genuine match scores typically range from 0.07 to 0.32, representing the likelihood of a match being authentic. Conversely, imposter match scores fall within the range of 0.4 to 0.55, indicating the probability of a match being fraudulent. This clear distinction facilitates effective decision-making in authentication processes.

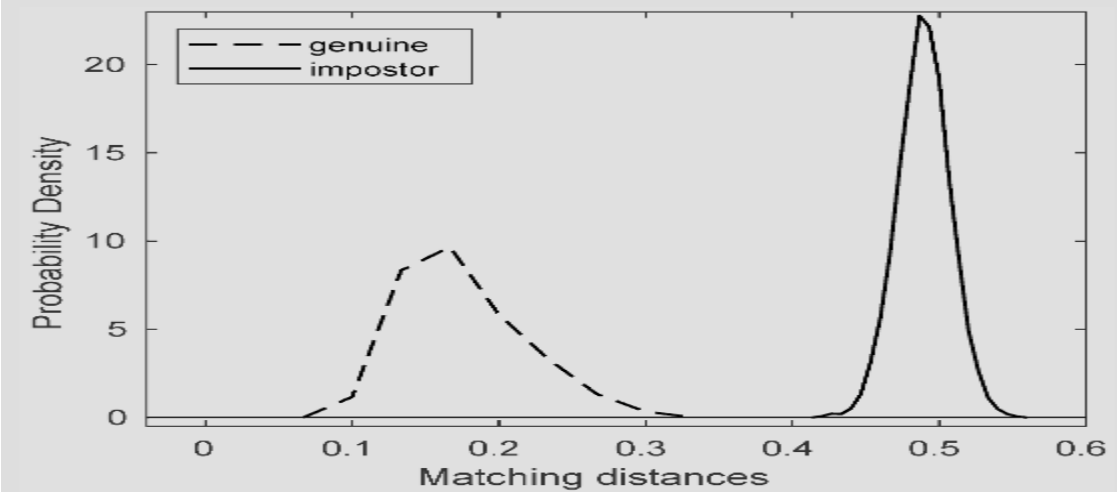

Figure14: Perfect probability distribution graph separating both the Genuine and Imposter match bell curves[24]

6.3 Results

6.3.1 FG-NET

In the investigation conducted on the FG-NET dataset utilizing the Arcface method for facial recognition, the analysis of the plotted graph revealed distinctive characteristics

regarding genuine and imposter match distributions. The observed phenomenon indicated that the imposter match range exceeded that of genuine matches, suggesting a notable discrepancy in the algorithm's ability to differentiate between matches of the same individual across developmental stages and those of different individuals. This disparity in match ranges underscores the importance of discerning genuine matches accurately amidst a larger pool of imposter matches, a crucial factor in assessing the algorithm's effectiveness in real-world applications.

Moreover, the graph exhibited intriguing patterns in the distribution of genuine and imposter matches across varying threshold values. Notably, both the genuine and imposter match bell curves exhibited peaks at the threshold value of 0.9.

Threshold	Genuine match		Imposter match	
	Before Normalization	After Normalization	Before Normalization	After Normalization
0.1	0	0	0	0
0.2	0	0	9	0.001
0.3	0	0	12	0.002
0.4	0	0	7	0.001
0.5	2	0.027	5	0.001
0.6	5	0.068	20	0.003
0.7	14	0.191	221	0.04
0.8	19	0.26	992	0.183
0.9	25	0.342	2134	0.395
1	7	0.095	1537	0.284
1.1	1	0.013	293	0.054
1.2	0	0	26	0.004

Table3: Instances of Genuine and Imposter match of FG-NET dataset using ArcFace method.

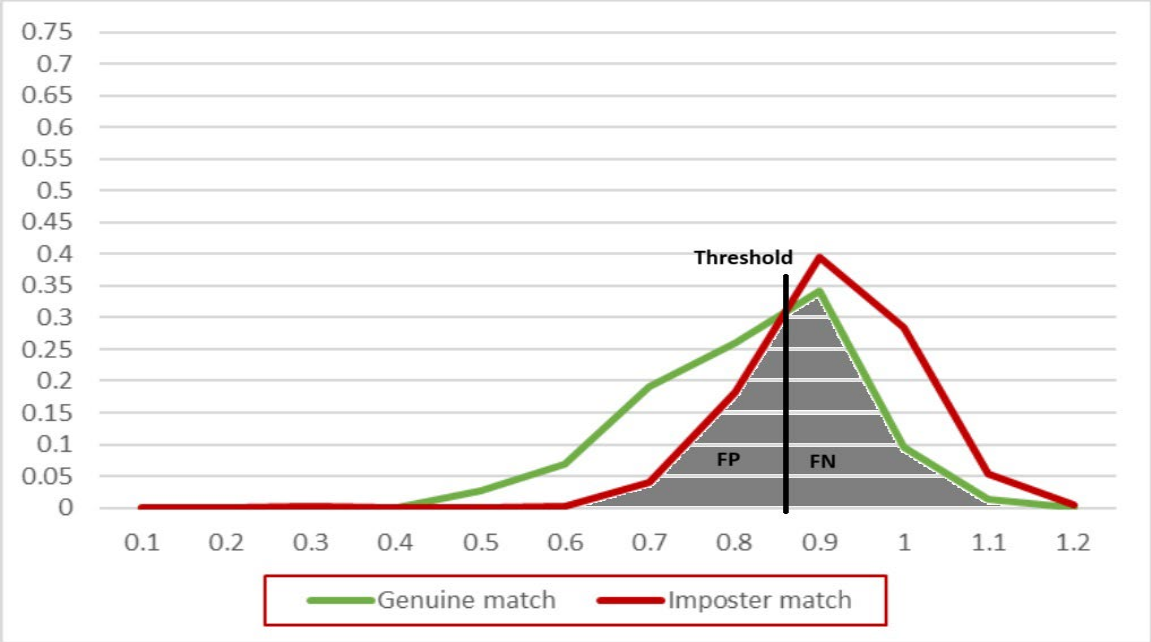

Figure15: Probability distribution of FG-NET dataset after normalization using ArcFace method

Analysis of the FG-NET dataset utilizing the VGG-Face method reveals distinct characteristics in the distribution of genuine and imposter matches. Both curves exhibit a range between 0.6 and 1.0, indicating a considerable overlap in match scores. However, closer inspection reveals notable differences in peak concentrations. The genuine match bell curve prominently peaks at 0.8, suggesting a higher concentration of matches within this score range, indicative of the algorithm's effectiveness in recognizing individuals across developmental stages. Conversely, the imposter match bell curve peaks at 1, indicating a concentration of matches at the highest score range, potentially signaling challenges in accurately distinguishing between different individuals. This nuanced analysis provides valuable insights into the performance nuances of facial recognition algorithms on the FG-NET dataset, contributing to the ongoing discourse on algorithmic accuracy and applicability across diverse datasets and developmental stages.

Threshold	Genuine match		Imposter match	
	Before Normalization	After Normalization	Before Normalization	After Normalization
0.1	0	0	0	0
0.2	0	0	0	0
0.3	0	0	0	0
0.4	0	0	0	0
0.5	0	0	0	0
0.6	2	0.002	9	0.001
0.7	7	0.095	67	0.012
0.8	25	0.342	507	0.093
0.9	21	0.287	2150	0.398
1	18	0.246	2523	0.466
1.1	0	0	0	0
1.2	0	0	0	0

Table4: Instances of Genuine and Imposter match of FG-NET dataset using VGG-Face method.

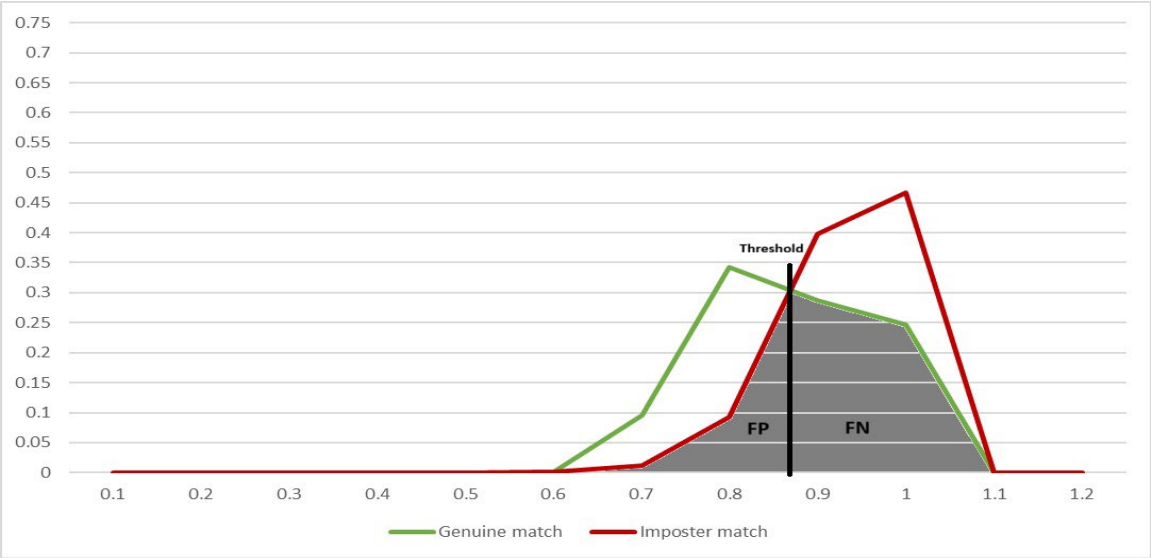

Figure16: Probability distribution of FG-NET dataset after normalization using VGG-Face method

6.3.2 ITWCC-3 and LAG Datasets

The study combines the ITWCC3 and LAG datasets, both sourced from real-world environments, featuring images of celebrities. By integrating these distinct datasets, the

research seeks to leverage the collective insights offered by each, enhancing the comprehensiveness and robustness of the experiment into various aspects of Face recognition and related fields.

The graphical analysis of combined ITWCC3 and LAG datasets employing the ArcFace method reveals distinct trends in imposter and genuine match distributions. Notably, imposter matches demonstrate a broad range spanning from 0.1 to 1.3 threshold values, indicating a considerable number of false positive identifications across the datasets. Conversely, genuine matches exhibit a narrower range, primarily falling between 0.3 to 1.1 thresholds. This observation underscores the challenge in accurately distinguishing between infant/toddler and adult selves, as evidenced by the prevalence of imposter matches at various threshold levels. Further investigation into algorithmic optimization and dataset refinement is warranted to enhance the discriminative capacity and robustness of facial recognition systems across developmental stages.

Threshold	Genuine match		Imposter match	
	Before normalization	After normalization	Before normalization	After normalization
0.1	0	0	8	0.0002
0.2	0	0	13	0.0003
0.3	1	0.005	2	0.00006
0.4	1	0.005	10	0.0003
0.5	2	0.01	14	0.0004
0.6	6	0.032	28	0.0008
0.7	27	0.148	279	0.008
0.8	54	0.296	2318	0.07
0.9	57	0.313	8881	0.269
1	35	0.192	13641	0.414
1.1	9	0.049	6674	0.202
1.2	0	0	1033	0.031
1.3	0	0	41	0.001

Table5: Instances of Genuine and Imposter match of ITWCC3 and LAG datasets using ArcFace method

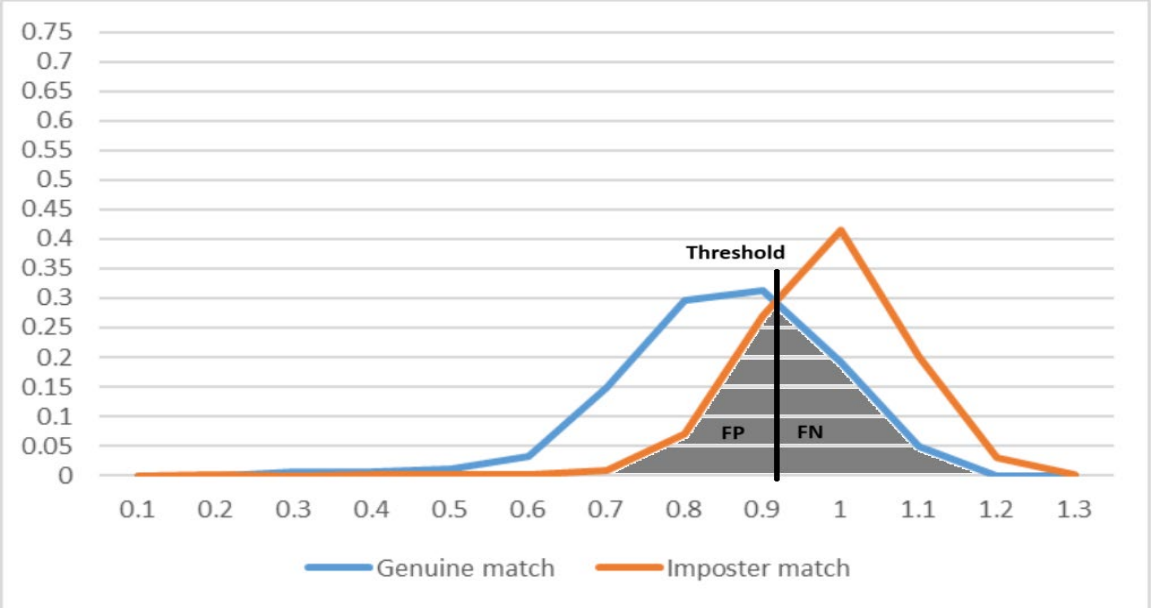

Figure17: Probability distribution of ITWCC3 dataset and LAG Dataset after normalization using Arcface method

The analysis of the combined ITWCC3 and LAG datasets utilizing the VGG-Face method revealed a notable convergence in the range of genuine and imposter match thresholds. Remarkably, both genuine and imposter match ranges extended from 0.4 to 1, suggesting a consistent discriminatory ability of the facial recognition algorithm across the datasets. This observation underscores the robustness and reliability of the VGG-Face method in distinguishing between infant/toddler and adult selves, irrespective of dataset origin. The convergence of match thresholds within this range signifies a balanced discrimination capability, affirming the algorithm's efficacy in accurately identifying matches and discerning between individuals at different developmental stages. This finding substantiates the utility and generalizability of the VGG-Face method in facial recognition applications across diverse datasets.

Threshold	Genuine match		Imposter match	
	Before Normalization	After Normalization	Before Normalization	After Normalization
0.1	0	0	0	0
0.2	0	0	0	0
0.3	0	0	0	0
0.4	1	0.005	1	0.00003
0.5	1	0.005	2	0.00006
0.6	4	0.021	45	0.0013
0.7	14	0.076	302	0.0091
0.8	41	0.225	2287	0.069
0.9	79	0.434	11315	0.3434
1	42	0.23	18993	0.5765
1.1	0	0	0	0

Table6: Instances of Genuine and Imposter match of ITWCC3 and LAG datasets using VGG-Face method.

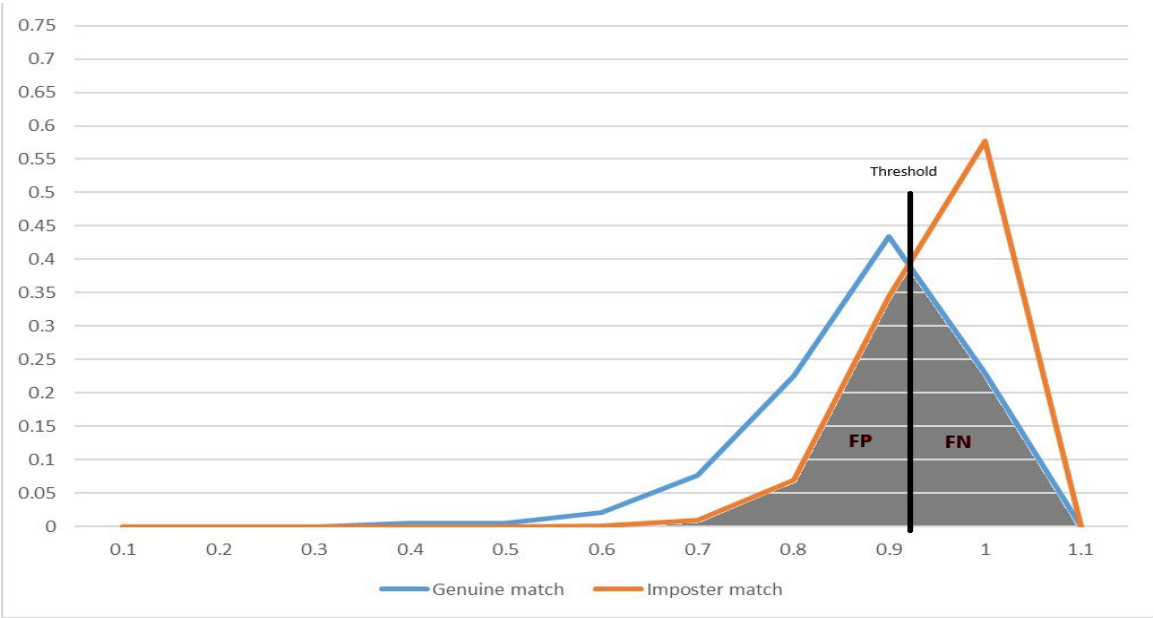

Figure18: Probability distribution of ITWCC3 dataset and LAG dataset(merged) after normalization using VGG-Face method

6.4 Precision

Precision in face matching algorithms refers to the accuracy with which the algorithm identifies or matches a newborn or toddler's face with their future adult self. In the context of facial recognition technology, precision measures how many of the

identified matches are correct out of the total number of matches returned by the algorithm. A high precision indicates that the algorithm is effectively distinguishing between matching and non-matching faces, minimizing false positives. This is particularly crucial in applications like matching newborns or toddlers to their adult selves, where accuracy is paramount for various purposes such as reuniting families or predicting future appearances.

Method	Precision
FG-NET using ArcFace	60.2
FG-NET using VGG-Face	74.1
ITWCC & LAG using ArcFace	65.3
ITWCC & LAG using VGG-Face	47.5

Figure19: Precision values

In the research of face matching algorithms, precision serves as a fundamental metric to assess the accuracy and efficacy of the employed methodologies. The presented precision values provide insights into the performance of two distinct algorithms, namely FG-NET and ITWCC & LAG, utilizing different face recognition models, namely ArcFace and VGG-Face.

In the case of FG-NET, leveraging ArcFace yields a precision value of 60.2%, indicating the algorithm's capability to accurately match newborns or toddlers with their adult selves with a moderate level of precision. Conversely, when employing VGG-Face within the FG-NET framework, a notably higher precision value of 74.1% is achieved,

suggesting enhanced accuracy in the identification process. This discrepancy underscores the significance of the chosen face recognition model in influencing the precision outcomes within the FG-NET context.

Similarly, within the ITWCC & LAG framework, utilizing ArcFace yields a precision value of 65.3%, denoting a moderate level of accuracy in matching newborns or toddlers with their adult selves. However, the application of VGG-Face within the same framework results in a substantially lower precision value of 47.5%, indicative of diminished accuracy compared to its ArcFace counterpart. This discrepancy underscores the nuanced interplay between the choice of algorithm and face recognition model in shaping precision outcomes within the ITWCC & LAG domain.

Overall, these precision values offer valuable insights into the comparative performance of FG-NET and ITWCC & LAG algorithms, as well as the differential impact of ArcFace and VGG-Face models on their accuracy in matching newborns or toddlers to their adult selves. Such findings contribute to the ongoing refinement and optimization of face matching algorithms, thereby advancing their applicability in diverse real-world scenarios.

CHAPTER 8: CONCLUSIONS AND FUTURE WORK

In this study, we investigated the precision of face matching algorithms in predicting the adult appearances of newborns or toddlers, utilizing the FG-NET and ITWCC & LAG datasets, along with the ArcFace and VGG-Face models. Our analysis yielded precision values of 60.2% for FG-NET with ArcFace, 74.1% for FG-NET with VGG-Face, 65.3% for ITWCC & LAG with ArcFace, and 47.5% for ITWCC & LAG with VGG-Face.

While these precision values offer valuable insights into algorithm efficacy, it's crucial to contextualize their significance. In the intricate realm of facial recognition, where subtle variations profoundly impact accuracy, achieving precision values above 60% is generally acceptable. However, for applications demanding heightened accuracy, such as forensic investigations or biometric authentication, attaining higher precision remains desirable.

Furthermore, the observed error rates in probability distribution graphs underscore the challenges in accurately discerning genuine matches from imposters. This highlights the necessity for ongoing refinement and optimization of face matching algorithms to mitigate false positives and negatives effectively.

Moreover, the constraint of limited dataset sizes, encompassing only 255 subjects, likely influenced the precision values obtained. Expanding dataset sizes and diversifying sample compositions hold potential to enhance algorithm performance and generalizability.

In conclusion, while the obtained precision values demonstrate promising performance of face matching algorithms, there exists ample room for enhancement.

Future research endeavors should prioritize refining algorithms, augmenting dataset sizes, and addressing identified challenges to bolster the accuracy and reliability of facial recognition technologies for practical applications.

8.1 Future Work:

Determining the matching between newborns or toddlers and their adult selves presents a formidable challenge, particularly with restricted datasets like FG-NET, ITWCC, and LAG Dataset. High error rates observed within the same range for both genuine and imposter matches, using methodologies like ArcFace and VGG-Face, highlight the existing limitations in accurate matching.

To surmount these challenges, future research could benefit from an expansion in data volume, ideally encompassing thousands of instances. A more extensive dataset facilitates the establishment of precise probability distributions for both genuine and imposter matches. Additionally, augmenting the training process to include images of infants or toddlers alongside adult counterparts holds promise for enhancing model performance.

By leveraging larger datasets and refining training methodologies to encompass diverse age groups, future advancements may yield improved accuracy in matching newborns or toddlers to their adult selves. This avenue of exploration is poised to enhance the efficacy of facial recognition techniques, offering valuable insights for applications ranging from age progression modeling to forensic investigations.

REFERENCES

[1] Khoa Luu, Karl Ricanek, Tien D. Bui, and Ching Y. Suen. 2009. Age estimation using active appearance models and support vector machine regression. In Proceedings of the 3rd IEEE international conference on Biometrics: Theory, applications and systems (BTAS'09). IEEE Press, 314–318. https://dl.acm.org/doi/10.5555/1736406.1736456

[2] Karl Ricanek Jr, Eric Patterson, and Amrutha Sethuram, "Diachronic Review of Facial Aging: Mechanisms and Algorithms", in the Proceedings of the First IEEE International Workshop on Robust Biometrics: Understanding the Science and Technology (ROBUST), Waikiki, Hawaii, October 2008.

[3] J. Goldstein and A. Watkins, "She Was Arrested at 14. Then Her Photo Went to a Facial Recognition Database." The New York Times, Aug 1, 2019. [Online] Available: https://www.nytimes.com/2019/08/01/nyregion/nypd-facial-recognition-children-teenagers.html [Accessed Sept 6, 2019]

[4] Craniofacial Changes in Children-Birth to Late Adolescence. (2019). ARC Journal of Forensic Science. https://doi.org/10.20431/2456-0049.0401001

[5] Anggo, M. and Arapu, L. (2018). Face recognition using the fisher face method. Journal of Physics Conference Series, 1028, 012119. https://doi.org/10.1088/1742-6596/1028/1/012119

[6] Srinivas, N., Ricanek, K., Michalski, D., Bolme, D., & King, M. (2019). Face recognition algorithm bias: performance differences on images of children and adults. https://doi.org/10.1109/cvprw.2019.00280

[7] Park, U., Tong, Y., & Jain, A. K. (2010). "Age-Invariant Face Recognition." IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(5), 947-954. DOI: 10.1109/TPAMI.2010.14.

[8] Deb, D., Nain, N., & Jain, A. (2017). Longitudinal study of child face recognition. https://doi.org/10.48550/arxiv.1711.03990

[9] Ricanek, K., Bhardwaj, S., & Sodomsky, M. (2015). "A review of face recognition against longitudinal child faces." BIOSIG 2015.

[10] Zhao, J., Yan, S., & Feng, J. (2022). "Towards Age-Invariant Face Recognition." IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(1), 474-487. DOI: 10.1109/TPAMI.2020.3011426.

[11] Hood, B. M., Macrae, C. N., Cole-Davies, V., & Dias, M. (2003). "Eye remember you: The effects of gaze direction on face recognition in children and adults." Developmental science, 6(1), 67-71. DOI: 10.1111/1469-7610.00142.

[12] Oo, S. L. M., & Oo, A. N. (2019). "Child Face Recognition with Deep Learning." International Conference on Advanced Information Technologies (ICAIT). DOI: 10.1109/AITC.2019.8921152.

[13] A. Midori Albert, et.al, "Craniofacial Changes in Children-Birth to Late Adolescence" *ARC Journal of Forensic Science* 2019:4(1): 1-19.

[14] K. Ricanek, S. Bhardwaj, and M. Sodomsky, "A Review of Face Recognition Against Longitudinal Child Faces," in BIOSIG 2015,
A. Brmme, C. Busch, C. Rathgeb, and A. Uhl, Eds. Bonn: Gesellschaft fr Informatik e.V., 2015, pp. 15–26.

[15] S. Bharadwaj, H. S. Bhatt, R. Singh, M. Vatsa, and S. K. Singh, "Face Recognition for Newborns: A Preliminary Study," in 2010 Fourth
IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), Sept 2010, pp. 1–6.

[16] L. Best-Rowden, Y. Hoole, and A. K. Jain, "Automatic Face Recognition of Newborns, Infants, and Toddlers: A Longitudinal Evaluation,"
in Biosig 2016, A. Brmme, C. Busch, C. Rathgeb, and A. Uhl, Eds. Bonn: Gesellschaft fr Informatik e.V., 2016, pp. 87–98.

[17] Serengil, S. (2023, January 22). Deep Face Recognition with ArcFace in Keras and Python - Sefik Ilkin Serengil. Sefik Ilkin Serengil. https://sefiks.com/2020/12/14/deep-face-recognition-with-arcface-in-keras-and-python/#google_vignette

[18] Serengil, S. (2023b, February 26). Deep Face Recognition with VGG-Face in Keras | sefiks.com. Sefik Ilkin Serengil. https://sefiks.com/2018/08/06/deep-face-recognition-with-keras.

[19] A. Lanitis and T. Cootes. Fg-net aging database. Cyprus College, 2002.

[20] K. Ricanek, S. Bhardwaj, and M. Sodomsky, "A Review of Face Recognition Against Longitudinal Child Faces," in BIOSIG 2015,
A. Brmme, C. Busch, C. Rathgeb, and A. Uhl, Eds. Bonn: Gesellschaft fr Informatik e.V., 2015, pp. 15–26.

[21] *Papers with Code - LAG Dataset*. (n.d.). https://paperswithcode.com/dataset/lag

[22] G. Levi and T. Hassner. Age and gender classification using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR-W), Workshop on Analysis and Modeling of Faces and Gestures, pages 34–42, 2015.

[23] D. Deb, N. Nain, and A. K. Jain, “Longitudinal Study of Child Face Recognition,” in 2018 International Conference on Biometrics (ICB),
Feb 2018, pp. 225–232

[24] Study on the Viability of Canine Nose Pattern as a Unique Biometric Marker
Choi et al. https://www.mdpi.com/2076-2615/11/12/3372