# Quality Beyond Perception: Introducing Image Quality Metrics for Enhanced Facial and Fingerprint Recognition

Thesis submitted in partial fulfillment
of the requirements for the degree of

*Master of Science in* ***Computer Science and Engineering*** *by Research*

by

Prateek Jaiswal
2021701009
prateek.jaiswal@research.iiit.ac.in

International Institute of Information Technology
Hyderabad - 500 032, INDIA
July 2024

International Institute of Information Technology
Hyderabad, India

# CERTIFICATE

It is certified that the work contained in this thesis, titled **“Quality Beyond Perception: Introducing Image Quality Metrics for Enhanced Facial and Fingerprint Recognition”** by Prateek Jaiswal, has been carried out under my supervision and is not submitted elsewhere for a degree.

| | |
|---|---|
| Date | Adviser: Dr. Anoop Namboodiri |

To my family and friends

Acknowledgments

I am filled with immense joy as I compile my entire research journey into this thesis. From initially having no understanding of research to eventually writing and publishing papers, my research experience at IIITH has been one of the most memorable phases of my life. I would like to begin by expressing my heartfelt appreciation to my research advisors, Dr. Anoop Namboodiri, for their consistent support and invaluable guidance in shaping my research career. Dr. Anoop's continuous encouragement and guidance at every stage of my research journey have not only helped me develop as an individual but also as a scholar. His practice of recognizing even the smallest achievements in my research work has consistently motivated me to work harder and persevere. Moreover, he has been instrumental in teaching me how to handle pressure effectively. Although my interactions with Dr. Anoop have been relatively limited, his dedication and enthusiasm for his work have left a lasting impact on me. I am genuinely grateful to him for involving me in his projects, which served as crucial milestones in my research career.

I would also like to extend my heartfelt gratitude to all my friends who have made my college experience both memorable and delightful. My journey towards obtaining my Master's degree would never have been complete without the invaluable assistance and support of my friends Aditya Kumar Singh, Dhaval Taunk, Dhruv Srivastava, Bhoomeendra Singh Shisodiya, K.N. Amruth Sagar, Haasith Pasala, Suyash Gupta, Mehul Arora, Nayan Jha. I must express my thanks to my co-author, Sai Amrit Patnaik, and Arka Koner, who played pivotal roles in both of my papers. Their consistent dedication and outstanding efforts throughout the entire research process have been invaluable. Additionally, I want to mention Srivathsa Gotipati, Chirag Jain, Ravi Mishra, Neil Shah, Pawan Baswani, Lokesh Madasu, Gopichand, and Siddhant Bansal for their generous help and support. They have enriched my knowledge by sharing different perspectives and new information in their respective fields.

I reserve a special place of gratitude in my heart for Praguna Manvi, Sourav Karmakar, Shivali Chansoria, Ayan Biswas, Apoorva Srivastav, and Ekta Gavas, who have been by my side, offering unwavering assistance and support throughout my master's journey. My journey would not have been possible without the love and blessings of my dearest family members. Their support and guidance have consistently played a significant role in all my life achievements.

Lastly, I would like to extend my thanks to Saketh Srinivas, Mahen Gandhi, and Kaustubh Kishore, who provided invaluable assistance during my time in Bangalore. Their cooperation during submissions and experiments, especially during my internship, taught me a great deal about the corporate world and how things work beyond the realm of thesis submission. I want to express my gratitude to Neetu Maam, whose support during my internship was exceptional, taking care of every detail during our stay in Bangalore.

# Abstract

Assessing the quality of biometric images is key to making recognition technologies more accurate and reliable. Our research began with fingerprint recognition systems and later expanded to facial recognition systems, underscoring the importance of image quality in both areas.

For fingerprint recognition, image quality is vital for accuracy. We developed the **Fingerprint Recognition-Based Quality (FRBQ)** metric, which improves on the limitations of the NFIQ2 model. FRBQ leverages deep learning algorithms in a weakly supervised setting, using matching scores from DeepPrint, a Fixed-Length Fingerprint Representation Model. Each score is labeled to reflect the robustness of fingerprint image matches, providing a comprehensive metric that captures diverse perspectives on image quality. Comparative analysis with NFIQ2 reveals that FRBQ correlates more strongly with recognition scores and performs better in evaluating challenging fingerprint images. Tested with the FVC 2004 dataset, FRBQ has proven effective in assessing fingerprint image quality.

After our success with fingerprint recognition, we turned to facial recognition systems. In facial recognition, image quality involves more than just perceptual aspects; it includes features that convey identity information. Existing datasets consider factors like illumination and pose, which enhance robustness and performance. However, age variations and emotional expressions can still pose challenges. To tackle these, we introduced the **Unified Tri-Feature Quality Metric (U3FQ)**. This framework combines age variance, facial expression similarity, and congruence scores from advanced recognition models like VGG-Face, ArcFace, FaceNet, and OpenFace. U3FQ uses a Regression Network model specifically designed for facial image quality assessment. We compared U3FQ to general image quality assessment techniques like BRISQUE, BLINDS-II, and RankIQA, as well as specialized facial image quality methodologies like PFE, SER-FIQA, and SDD-FIQA. Our results, supported by analyses such as DET plots, expression match heat maps, and EVRC curves, show U3FQ's effectiveness.

Our study highlights the transformative potential of artificial intelligence in biometrics, capturing critical details that traditional methods might miss. By providing precise quality assessments, we emphasize its role in advancing both fingerprint and facial recognition systems. This work sets the stage for further research and innovation in biometric analysis, underlining the importance of image quality in improving recognition technologies.

# Contents

# List of Figures

# List of Tables

Table Page

*Chapter 1*

# Biometric Image Quality Assessment

## 1.1 Introduction

The advent of deep learning has significantly transformed biometrics, revolutionizing identity verification and authentication methods. At the heart of this technological advancement is image quality assessment (IQA), which is crucial for ensuring the reliability and efficiency of biometric recognition systems. The quality of biometric data, whether from fingerprint scans or facial recognition, directly affects a system's ability to match and verify identities accurately. This thesis is about the complex world of biometric IQA, exploring the challenges and proposing innovative solutions tailored to the unique requirements of fingerprint and facial recognition systems.

Traditional IQA methods [38, 29, 6, 57] often do not meet the specific needs of biometric applications. Hence, this work introduces deep learning-based metrics to enhance IQA for biometric systems, setting a new benchmark in pursuing more secure and effective biometric authentication processes. By comprehensively examining both Fingerprint IQA and Face IQA, we aim to improve the accuracy and reliability of biometric recognition systems. This introduction lays the groundwork for our exploration, setting the stage for a detailed discussion on the evolution of IQA methods, the development of novel quality assessment metrics, and their significant implications for the future of biometric security and identification.

### 1.1.1 Quality in General

Image quality assessment is crucial in image processing, emphasizing human perception of quality, which favors clear images, with minimal noise, distortion, or motion blur. As shown in Figure 1.1, we can see the impact of different levels of compression artifacts on image quality, with the least compression in the image (b) and the most in the image (e). These images are sourced from [57]. Traditionally, image quality is measured against a reference image, serving as a standard for comparing the quality of other images. However, this approach changes when a reference image is not available, such as in images taken in natural or uncontrolled settings. To address this challenge, image quality assessment has branched into two main categories: reference-based and no-reference-based quality assessment.

In reference-based quality assessment, methods aim to measure image quality by comparing a given image to a high-quality reference image. This approach is suitable for scenarios where a clear reference image is

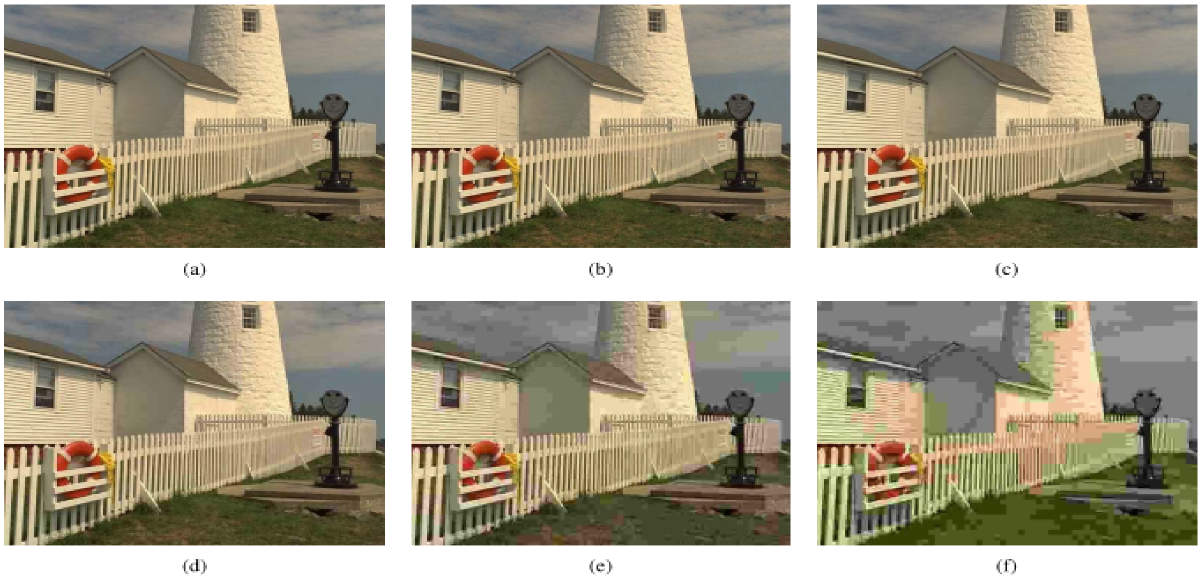



Figure 1.1: This figure demonstrates the variations in image quality after applying different levels of compression artifacts, ranging from minimal (b) to significant (e), based on the study from [57].

available for quality benchmarking. Various techniques have been developed in this category to score the quality of images effectively, drawing from established research in the field [64, 51, 30].

Conversely, no-reference-based quality assessment methods tackle the challenge of assessing image quality when no reference image is present. These methods are designed to evaluate the quality of images independently, without relying on a reference point. Notable works in this category include [39, 47, 57], which have explored the development of quality metrics that do not require a reference image.

## 1.2 Biometric Image Quality Assessment

Image Quality Assessment (IQA) traditionally revolves around human visual perception and aesthetic judgment. IQA is specialized and redefined in biometric recognition for the imperatives of accurate identity verification [5]. This subset of IQA is pivotal for the performance of face and fingerprint recognition systems, focusing on the utility of images for precise biometric identification rather than their visual appeal.

### 1.2.1 Motivation

The motivation for this work stems from the quality paradox in biometric systems. This paradox highlights the discrepancy between general IQA standards and the stringent requirements of biometric image quality, emphasizing the need for a specialized approach to IQA in biometric applications. Figure 1.2 we can see how general image quality and biometric image quality vary.

The deep learning revolution in IQA has brought a paradigm shift in biometric systems, seamlessly integrating IQA into recognition models. This shift allows for proactive image capture enhancement and data quality assessment at the source, which is crucial for the system's recognition capabilities.

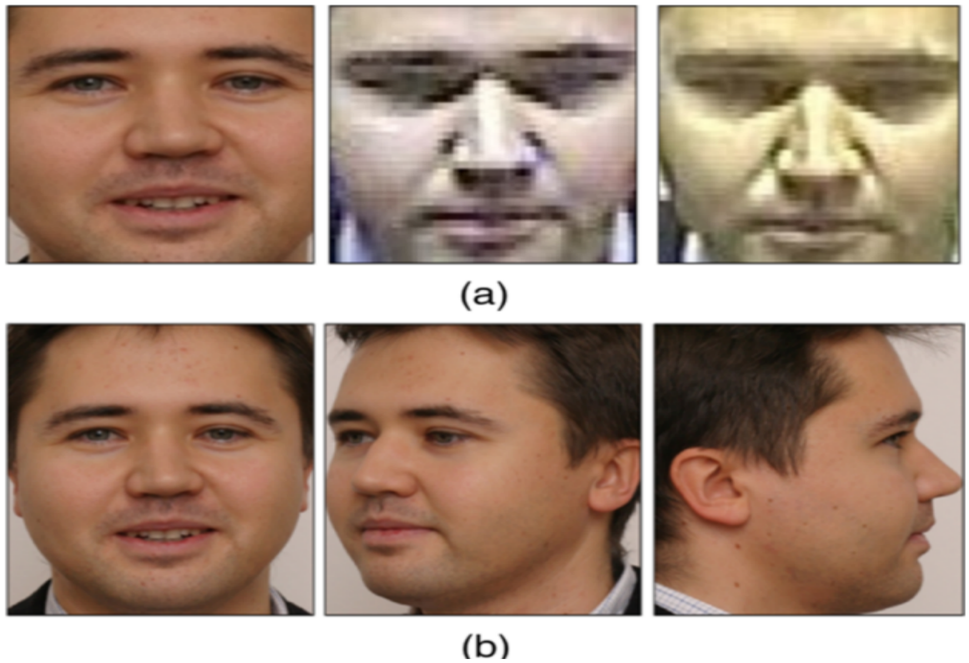



Figure 1.2: Image quality vs biometric quality. While the images (obtained from SCface database) in (a) are of poor image quality, the images in (b) may have lower biometric quality. [5]

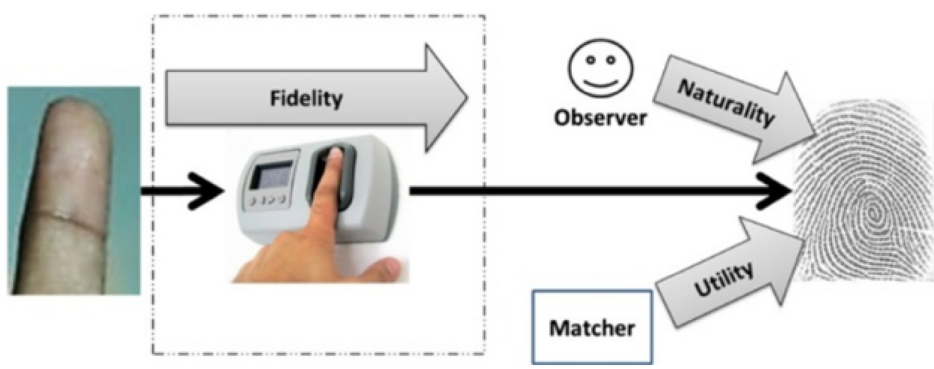



Figure 1.3: Three aspects of quality assessment: naturality, fidelity, and utility, in a typical biometric pipeline [5]

The intricacies of biometric systems demand IQA methodologies that can quantify the impact of various factors on recognition accuracy. This includes analyzing facial feature clarity, occlusions, and lighting conditions for facial recognition. Ridge definition, pattern integrity, and artifact absence are crucial for fingerprint systems. While deep learning-based IQA models automate and refine these assessments, challenges in interpretability and transparency persist, especially where decision-making has significant repercussions.

## 1.3 Biometric Systems and IQA

The complexity of biometric systems necessitates specialized IQA methodologies capable of discerning subtle details that influence recognition accuracy. These methodologies must consider factors such as the clarity of facial features, the presence of occlusions, and lighting conditions for facial recognition systems, as well as ridge definition and pattern integrity for fingerprint systems.

### 1.3.1 Challenges in Deep Learning-Based IQA Models

The advent of deep learning models in IQA presents promising advancements in the automation and refinement of quality assessments. However, the "black box" nature of these models poses significant challenges, particularly regarding interpretability and the need for transparency in sensitive decision-making processes.

### 1.3.2 Fingerprint Image Quality Assessment:

In the specialized field of fingerprint image quality assessment (Fingerprint IQA), the objective is to evaluate the quality of fingerprint images to ensure they adhere to the exacting standards required by fingerprint recognition systems. These systems are crucial for biometric authentication, and their effectiveness largely hinges on the quality of the input fingerprint images. Initially, traditional Fingerprint IQA methods predominantly utilized handcrafted features, relying on the comparison with reference images to gauge quality [52, 54, 56, 55]. However, these methods encounter challenges in effectively addressing various image distortions, such as blurring, noise, and compression artifacts, which can significantly impact the accuracy of biometric authentication. Figure 1.3 shows how the biometric system looks for fingerprints and how we see it.

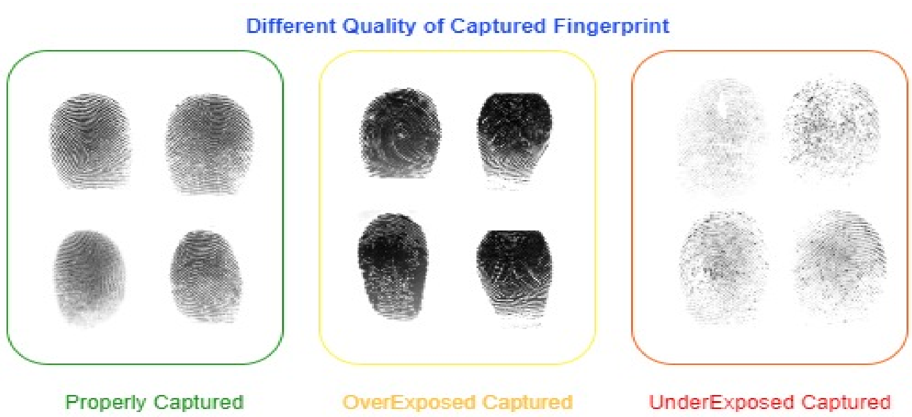



Figure 1.4: (a) High-quality fingerprint with clear ridge patterns.(b) Overexposed fingerprint with excessive contrast results in loss of ridge pattern detail. (c) Degraded quality fingerprint with reduced clarity and disrupted ridge patterns.

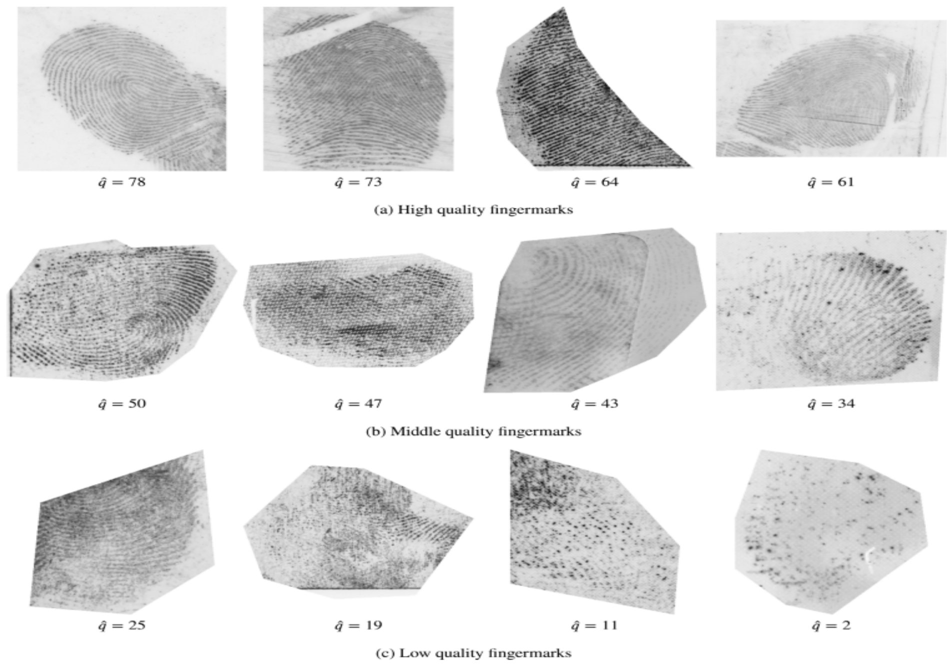



Figure 1.5: Figure showing high quality, middle quality and low quality finger print images.

Figure 1.6: Sample Images of Varying Fingerprint Image Quality

Figure 1.4 shows the scanner output after we capture the fingerprint. It shows how an ideal capture should look and what a poor capture is. Figure 1.5 and 1.6 show a sample of how quality changes in the fingerprint images.

Initially, Fingerprint IQA techniques depended on analyzing a mix of local and global image quality features [27, 32, 31, 52, 12, 13, 22]. These features include variations in Gabor filter responses, frequency and orientation measures, and pixel intensity quality, or a combination thereof, to evaluate the fingerprint image quality. To bring a standard approach to assessing these diverse and intricate features, the National Institute of Standards and Technology (NIST) developed the NIST Fingerprint Image Quality (NFIQ) [56, 55] algorithm. NFIQ synthesizes numerous local and global features intrinsic to the fingerprint image, enabling a more comprehensive and standardized quality assessment. These features are considered handcrafted because they are specifically designed and selected based on predefined image quality criteria.

The introduction of NFIQ marked a significant advancement in Fingerprint IQA by providing a unified metric that encapsulates various aspects of fingerprint image quality. By leveraging such a standardized framework, it became possible to more accurately and consistently determine the suitability of fingerprint images for biometric identification purposes. This development underscores the importance of sophisticated quality assessment methods in enhancing the reliability and efficiency of fingerprint recognition systems.

### 1.3.3 Face Image Quality Assessment:

The core of this thesis is a comprehensive examination of the current landscape of IQA, focusing on face images and the intricate patterns of fingerprints. We investigate the conceptual frameworks, methodologies, and applications, noting the shift towards deep learning paradigms that offer promising improvements in accuracy and interpretability.

In the context of Face IQA, this involves parsing facial features and expressions, environmental conditions, and sensor interoperability. For Fingerprint IQA, it includes analyzing ridge patterns, minutiae clarity, and the impact of presentation attacks. We further discuss the application scenarios of IQA in biometric systems, from pre-enrollment image selection to real-time capture quality enhancement.

We identify the pressing need for standardization in algorithm evaluations to ensure comparability and highlight the challenges that lie ahead, such as developing interpretable deep learning methods that go beyond mere accuracy and utility predictions.

This thesis aims to bridge the gap between Fingerprint IQA and Face IQA by proposing a unified framework that addresses their individual challenges while leveraging their respective strengths. We posit that the future of biometric IQA lies in models that are adept at quality assessment and contribute to the proactive improvement of data capture processes, ensuring that biometric systems are more accurate, reliable, and effective in real-world scenarios.

Within the sequence from fig 1.7, we observe various distortions and their effects on facial images. However, these visual representations alone do not provide clear insights into how such distortions influence the matching process in facial recognition systems. Specifically, the distinction in matching outcomes between poor-quality images paired with different or identical subjects remains unexplored through these images alone. Consequently, while these images may offer a perceptual or subjective assessment of quality, they fall short of predicting the performance during actual recognition tasks.

Recognition quality, in essence, hinges on the ability to gauge the similarity between a facial image and other images within a database, which in turn affects the system's recognizability and, by extension, the overall image quality from a biometric standpoint. This similarity metric is crucial for understanding how distortions impact the recognition capabilities of biometric systems. Therefore, without this comparative analysis, making definitive claims about performance based on image quality alone is challenging.

To address this gap, a more nuanced approach is needed, incorporating the assessment of recognizability alongside traditional quality measures. This involves analyzing how distortions affect the facial recognition system's ability to match a given image against a database, considering both 'different-pair' and 'same-pair' comparisons.

Figure 1.7: A biometric system may process samples that exhibit a wide range of quality variations (Images from MBGC database). Therefore, effective quality assessment metrics that accurately reflect these variations are crucial for the functionality of an automated biometric system. [5]

Furthermore, Face Image Quality Assessment techniques are generally classified into two main categories: regression-based and learning-based approaches. Learning-based strategies align with the operational mechanisms of facial recognition models, adopting similar principles to predict image quality. These methods leverage the recognition model's understanding to estimate the quality of facial images. On the other hand, regression-based methods employ deep neural networks to learn the representation of how a facial image is decomposed, subsequently identifying features crucial for the quality prediction task based on these learned representations. One notable limitation of regression-based approaches is their reliance on labeled data for training to generate quality scores. This requirement for labeled training data can be a constraint, as it necessitates a pre-existing, accurately labeled dataset to train the model for quality assessment tasks effectively.

### 1.3.4 Contributions:

The contributions of this thesis are multifaceted and significant in advancing the field of biometric image quality assessment. They are summarized as follows:

1. **Development of a Deep Learning-Based Fingerprint Quality Assessment Metric (FRBQ):** This thesis introduces the FRBQ (Fingerprint Recognition-Based Quality) metric, a novel approach in the realm of fingerprint image quality assessment. Unlike traditional metrics, FRBQ is designed to operate independently of ground truth label data, relying instead on deep learning techniques to evaluate the quality of fingerprint

images. This innovation represents a significant leap forward, offering a more flexible and potentially more accurate method for assessing fingerprint image quality.

2. **Effective Performance on Low-Resolution Images:** One of the standout features of the FRBQ metric is its ability to assess the quality of low-resolution fingerprint images accurately. This capability is particularly noteworthy as it addresses a common limitation faced by state-of-the-art models, which often struggle to perform effectively on images of lower resolution. By overcoming this challenge, the FRBQ metric enhances the reliability and applicability of fingerprint quality assessment, especially in scenarios where high-resolution images may not be available.

3. **Introduction of U3FQ for Systematic Analysis of Facial Age and Expression Similarity:** The thesis also pioneers the U3FQ framework, marking the first systematic effort to analyze facial age and expression similarity within the context of Facial Image Quality Assessment (FIQA). This analysis sheds light on the age-related dynamics and expression variance in face recognition systems, offering novel insights into how these factors impact recognition performance.

4. **Semantic Insights into Face Age Images:** The research provides a detailed examination of face age images across various age groups, elucidating the differences in how different models represent and interpret these images. This aspect of the study highlights the nuanced ways in which age affects facial recognition, contributing to a deeper understanding of the challenges and opportunities in age-based biometric verification.

5. **Enhanced Accuracy and Robust Generalization:** Through implementing the U3FQ framework, this thesis demonstrates significant improvements in accuracy and robust generalization across a range of benchmark datasets. This achievement not only sets a new standard for FIQA and IQA methodologies but also underscores the potential for further advancements in the field.

Together, these contributions underscore the thesis's role in pushing the boundaries of biometric image quality assessment. By introducing innovative metrics and frameworks, this work aims to improve the accuracy and reliability of biometric recognition systems, paving the way for more secure and efficient authentication processes.

By exploring both Fingerprint IQA and Face IQA domains, this thesis aims to shed light on the sophisticated techniques required to ensure high-quality biometric data, thereby supporting the development of more accurate and reliable biometric recognition systems.

*Chapter 2*

# Related Works

## 2.1 Related Works

The advent of deep learning has ushered in a transformative shift in the field of biometrics, marking a new era characterized by unprecedented levels of accuracy and reliability across various biometric tasks, with a notable emphasis on face recognition. This paradigm shift has facilitated the development of more robust and efficient biometric systems, applicable in diverse domains such as security, authentication, and personal identification.

## 2.2 Fingerprint Image Quality Assessment

In scenarios where explicit label information is unavailable, No Reference Image Quality Assessment (NR-IQA) plays a crucial role [38, 29, 6, 57]. To address this challenge, adopting a weakly supervised learning approach has been proposed [68, 43]. Additionally, Remy et al. have explored fingervein quality assessment [46], and Oblak et al. have conducted a comprehensive survey on deep learning ensemble models for fingerprint image quality assessment [42]. These studies offer valuable insights and alternative methods to enhance recognition accuracy in scenarios with limited label information.

### 2.2.1 Deep Learning in Biometric Applications

DL models have shown significant improvement in accuracy and robustness in various biometric applications, including face recognition, fingerprint recognition, and iris recognition. These models have been shown to outperform traditional feature extraction and classification methods by learning complex representations directly from the raw biometric data. This has led to a paradigm shift in biometric recognition, where deep learning models are becoming the go-to solution for a wide range of biometric recognition tasks. However, challenges remain to be addressed, such as ensuring the privacy and security of biometric data, addressing potential biases and discrimination, and improving the interpretability and transparency of deep learning models.

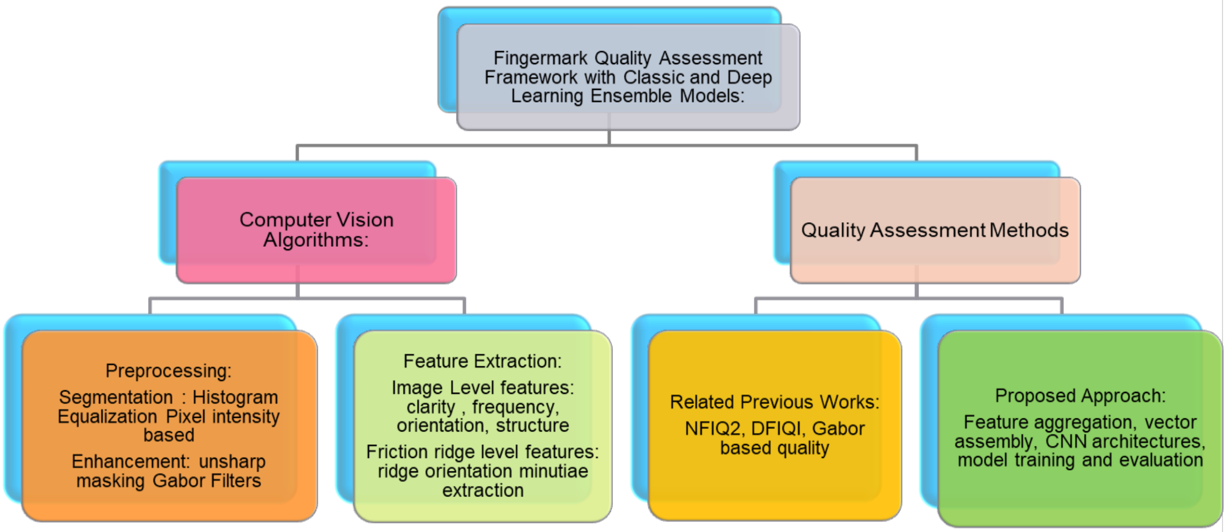



Figure 2.1: Overview of the Fingerprint Image Quality Assessment Process: This figure maps out the multifaceted approach to assessing fingerprint image quality, from preprocessing and feature extraction techniques to the evaluation of existing and proposed quality assessment methods within computer vision algorithms.

Deep learning models such as DeepPrint [18] have been utilized for fingerprint recognition to improve accuracy. DeepPrint extracts important information from fingerprint images without extracting varying feature information through a fixed-length fingerprint representation. It combines deep features extracted by deep networks with minutiae handcrafted features to get the critical feature information. The architecture of DeepPrint uses multitask branches to first extract the fingerprint image representation and subsequently learn how to classify it. The information from both branches is then fused to get a recognition score. Other deep networks have improved specific sub-modules of fingerprint recognition systems such as segmentation [69, 14, 19, 53], orientation field estimation [9, 50, 45], and minutiae extraction [41, 58, 15, 10, 65].

MiDeCon [59] is one of the few works that have utilized an approach incorporating minutiae information for quality score generation in fingerprint recognition. However, this approach has not been widely implemented, and DeepPrint provides a more advanced method for learning quality features. By incorporating minutiae maps during training, DeepPrint's feature-guided deep network has shown superiority over traditional methods in quality-based recognition tasks. As a result, DeepPrint is used in this study to generate matching scores and create proxy ground truth labels for accurate quality score prediction.

Figure 2.1 provides a comprehensive overview of the different components and methodologies employed in Fingerprint Image Quality Assessment (FIQA). It outlines a structured approach that spans from the initial preprocessing of fingerprint images to applying sophisticated computer vision algorithms and quality assessment methods, culminating in the proposed approach for enhancing the assessment process.

In the realm of quality prediction for recognition systems, prior research predominantly relied on handcrafted local and global features. In the figure 2.2, These features were meticulously designed to capture relevant information from the data. However, our novel approach diverges from this tradition. Instead of

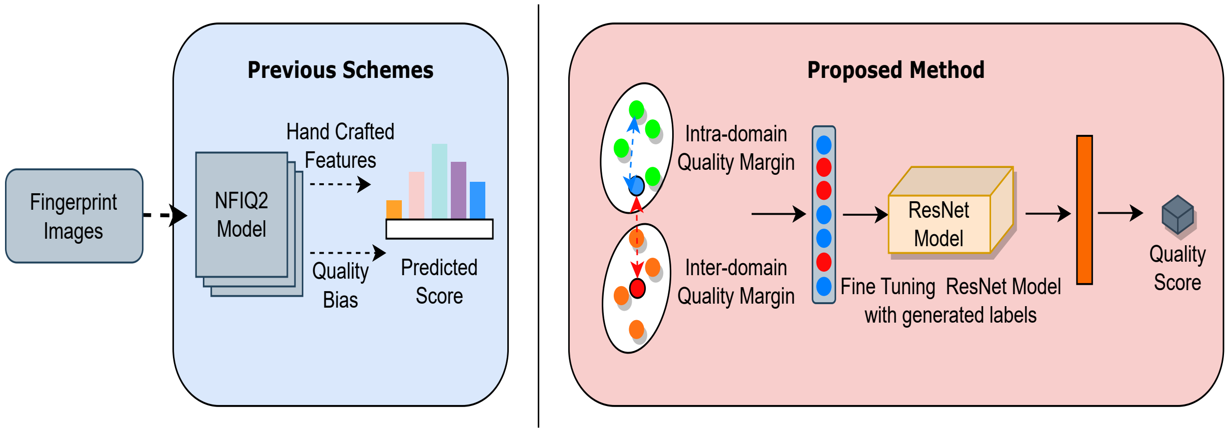



Figure 2.2: Overview of our proposed quality prediction method. Instead of relying solely on handcrafted features, we explore inter and intra-class similarities by labeling them. The matching scores obtained from these labeled similarities serve as the foundation for our quality prediction model.

focusing solely on feature engineering, we delve into the data's intricate inter and intra-class similarities. By meticulously labeling these similarities, we construct a more nuanced representation. Our method leverages matching scores derived from these labeled similarities to predict quality. This shift in perspective promises to enhance recognition system performance significantly.

## 2.3 Face Image Quality Assessment:

The figure 2.3 outlines a structured taxonomy for Face Image Quality Assessment (FIQA), an integral component in ensuring the efficacy of face recognition systems. FIQA methodologies are classified into various categories, each addressing distinct aspects of image quality and their implications for recognition accuracy.

Under the factor-specific category, the taxonomy identifies critical variables such as size, including inter-eye distance and image resolution, which are fundamental for maintaining consistent quality in facial recognition. Illumination is dissected into sub-factors like brightness moments and contrast, significantly affecting facial features' visibility. The pose is analyzed through facial landmarks and appearance templates, among others, to account for variations in face orientation that can impede accurate identification. Additional factors such as noise, skin tone, and other attributes further contribute to the comprehensive nature of this quality assessment framework.

The monolithic section of the taxonomy encompasses methods that treat the image as a whole, employing techniques like edge analysis and symmetry assessment through holistic and whole-image approaches, including opaque machine-learning strategies that do not offer transparency in their assessment process.

The data-driven aspect of FIQA is addressed through various approaches: hand-crafted (Dhc), utility-agnostic (Duat), and several forms of ground truth training, including human-based (Dhgt) and face recognition (FR)-based (Drft, Dfri), as well as FR-integration (Dint). Fusion methods are delineated into explicit (Fe), trained (Ft), and cascade (Fc), with an additional note on instances where fusion is not applied.

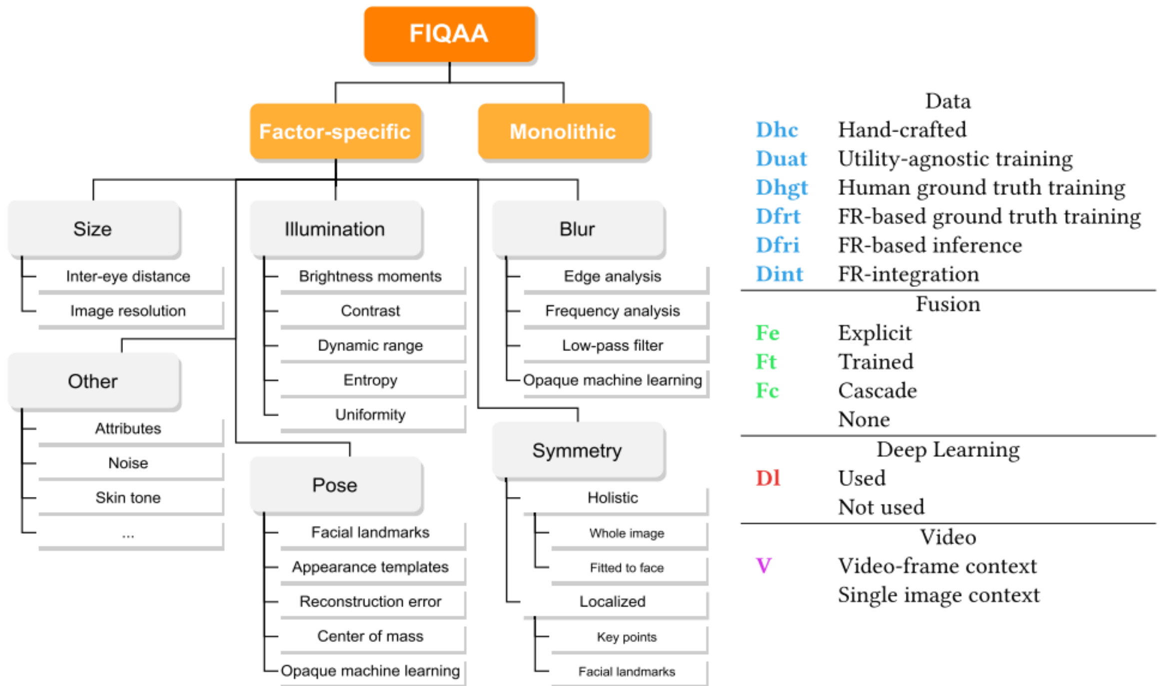



Figure 2.3: Comprehensive Framework for Face Image Quality Assessment (FIQA): This figure categorizes the multifaceted approaches to FIQA, detailing factor-specific metrics, monolithic image evaluation techniques, data-driven methods, fusion strategies, and the application of deep learning and video data analysis in the pursuit of enhancing face recognition systems. Taken from Survey Study [48]

Deep Learning (DI) is recognized as a pivotal component in contemporary FIQA methods, clearly distinguishing between its usage and non-usage. Incorporating video data (V) acknowledges the dynamic nature of image capture, considering both video-frame and single-image contexts.

Facial Image Quality Assessment (FIQA) has become a cornerstone in enhancing the performance of face recognition (FR) systems, especially when faced with the challenges posed by the wide spectrum of image qualities encountered in real-world scenarios. Traditional FIQA methodologies have primarily focused on evaluating the biometric utility of facial images in isolation. However, the integration of FIQA within the broader context of FR systems introduces a conceptual challenge, termed the "Quality Paradox." This paradox, as discussed by Schlett et al. [48], highlights the need for a nuanced representation of the reliability of comparison scores for image pairs that include the assessed image, adding complexity to FIQA's role in optimizing FR performance.

Despite significant advances by contemporary FR technologies in handling high-quality frontal images under varied quality conditions [36, 23], challenges persist in fully unconstrained environments [7, 63], where the consistency of facial image quality cannot be guaranteed. FIQA methodologies are invaluable in these scenarios, providing crucial insights into image quality and enabling FR models to identify and potentially exclude inferior quality images, thus mitigating the risk of erroneous non-matches.

Table 2.1: Comparative Overview of Recent Face Recognition Quality Assessment Methods

| Work | Methodology | Evaluation Criteria | Evaluation Dataset |
|---|---|---|---|
| SER-FIQ [61] | Embedding Variability Analysis: This method utilizes statistical variability in face embeddings to infer image quality. | FNMR, EER: These metrics indicate the frequency of incorrect rejections and the balance point where false acceptances and rejections are equal. | Adience, Color Feret, LFW |
| SDD-FIQA [44] | Similarity Distribution Distancing: This method evaluates the distance in feature space by measuring the distribution of similarity scores. | FMR: The False Match Rate measures the rate of incorrect acceptances, a critical security metric. | Adience, LFW, IJB-C |
| CR-FIQA [8] | Relative Classifiability to assess the ease of correctly classifying a given face image. | FNMR, TAR reflecting the proportion of genuine users correctly verified. | Adience, LFW, AgeDB, CFP-FP, CA-LFW, CP-LFW, XQ-LFW, IJB-C |
| Mag-Face [37] | Enhancing ArcFace with a magnitude-aware loss function for better discriminative learning. | FNMR indicating the frequency of false rejections. | AgeDB, LFW, CFP-FP |
| Face-Qnet [26] | Neural Network Quality Assessment to directly predict the quality score of face images. | Correlation between predicted quality scores and actual biometric utility. | BioSecure Database |
| Auto-matic FIQ [4] | CNN Approaches for feature extraction and quality assessment. | FNMR, FMR indicating the reliability of the authentication system. | LFW |
| PC-Net [66] | Predictive Uncertainty And Confidence Estimation for assessing the quality of face images. | FMR representing the likelihood of incorrect authentications. | IJB-C |
| Opti-miza-tion Based [21] | Employing Quality Label Supervision to guide the quality assessment in a more targeted manner. | FMR, AUC indicating the overall performance of the model. | LFW, CFP-FP, CA-LFW, CP-LFW, XQ-LFW |
| Face-QAN [2] | Adversarial Noise Profiling to measure the robustness of face images against noise. | FMR, AUC to assess the accuracy and reliability across different noise conditions. | LFW, CFP–FP, XQLFW ,IJB–C |
| QMag-Face [60] | Quality Aware Face Recognition | FNMR and EER plots | AgeDB, XQLFQ, CFP-FP |

Contemporary FIQA methodologies are categorized into two main groups: regression-based and model-based approaches. Regression-based strategies [11, 26, 66] establish a direct correlation from the image domain to quality labels, often generated in a semi-automated manner. These labels are typically based on comparison scores from matched image pairs or similarity scores between probe samples and reference images. Conversely, model-based approaches [61, 37] integrate quality assessment directly within the FR model, evaluating quality based on certainty or statistical attributes extracted from generated facial features or embeddings.

Our proposed approach extends the FR model paradigm by incorporating the impact of age variations and emotional expressions on matching scores. Understanding how aging and emotions alter facial features is crucial, as they significantly influence recognition accuracy.

This thesis introduces the Unified Tri-Feature Quality (U3FQ) metric, a novel approach to FIQA. U3FQ redefines FIQA by integrating recognizability and quality estimation through a unique, learning-based methodology. Diverging from traditional paradigms, it leverages match scores in a weakly supervised manner as the primary indicator of quality. Furthermore, U3FQ emphasizes the significant role of facial expressions and age disparity in quality assessment, providing insights into the influence of these factors on matching accuracy.

In addition to facial recognition, this work addresses limitations in fingerprint image quality assessment with the introduction of FRBQ (Fingerprint Recognition-Based Quality), a deep learning-based methodology for no-reference image quality assessment. FRBQ aims to evaluate the recognition performance of fingerprint images across quality scores, utilizing a Convolutional Neural Network (CNN) architecture adapted from the ResNet18 framework. This approach has demonstrated effectiveness even with low-resolution images, surpassing the established NFIQ2 model in accuracy and resilience to various image distortions.

The integration of advanced deep learning models such as U3FQ for facial image quality assessment and FRBQ for fingerprint image quality assessment represents a significant advancement, promising to enhance the accuracy and reliability of biometric identification systems. These developments have far-reaching implications in real-world applications, including security, forensic analysis, authentication, and identification processes.

*Chapter 3*

# Advancing Fingerprint Recognition Quality Assessment: Introducing the FRBQ Metric

## 3.1 Introduction

The assessment of fingerprint image quality is crucial for the efficiency of biometric recognition systems. In this work, we introduce a novel methodology that employs deep learning networks in a weakly supervised setting to learn features specific to fingerprint images. This approach is grounded in the concept of matching fingerprints, where the network is trained to produce a quality score that directly corresponds to the fingerprint's matching capability. By using simulated ground truth data derived from labeled fingerprints, our network can identify and learn the most relevant features for assessing fingerprint quality, thereby improving the precision and reliability of the quality assessment process.

## 3.2 Pipeline

Our pipeline begins with the generation of pseudo ground truth labels, which serve as the foundation for our approach. This process involves using the matching scores between fingerprint images, derived from their cosine similarity, as a basis for label generation. These scores, produced by our DeepPrint model, reflect the degree of similarity between fingerprints, with the threshold set at a predefined value (e.g., 0.4) to distinguish between different quality levels. This step is critical for our weakly supervised learning methodology, as it enables the network to learn from a practical, albeit indirect, representation of fingerprint quality.

### 3.2.1 DeepPrint Feature Extraction and Matching

The DeepPrint [18] model plays a pivotal role in our pipeline, as depicted in Figure 3.1. It processes input fingerprint images to generate a 192-dimensional vector, encapsulating both texture and minutiae details. These vectors are then utilized to compute matching scores through cosine similarity, providing a robust basis for comparing fingerprints. The process of generating these scores, alongside the extraction of minutiae and texture details, is essential for creating a detailed representation of fingerprint features, which is subsequently used for label generation.

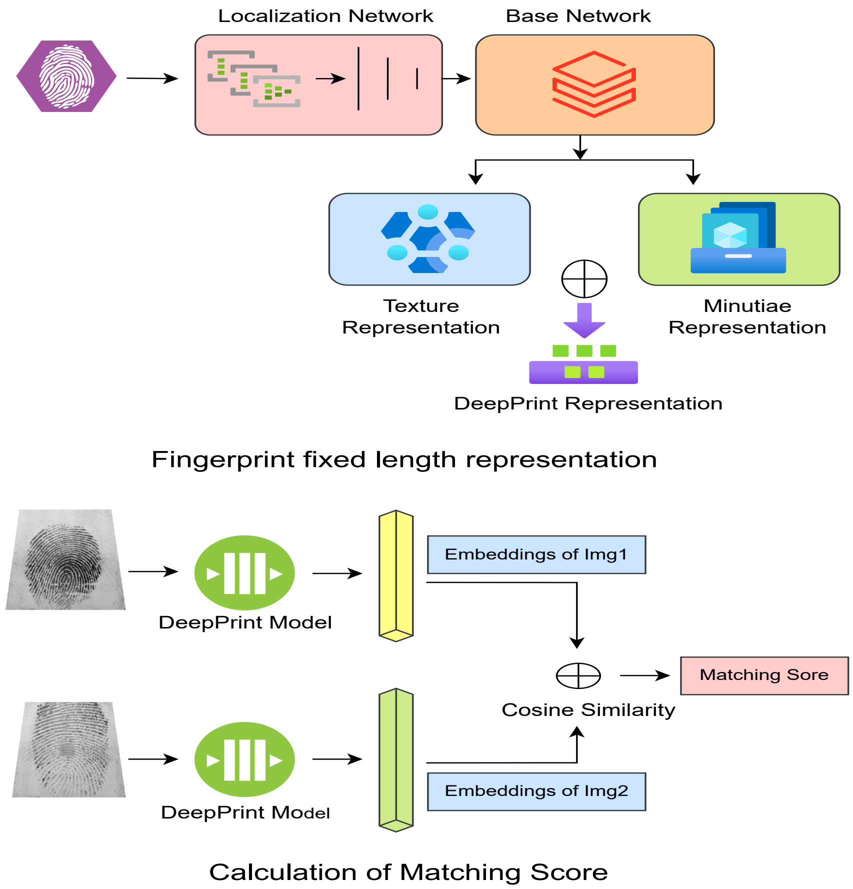



Figure 3.1: **Block 1:** The input image, $\mathcal{I}$, undergoes processing to generate a 192-dimensional vector. Texture and minutiae details are extracted and combined to form a DeepPrint representation $\mathcal{F}_d$. **Block 2:** Embedding vectors from DeepPrint are compared using cosine similarity to produce a matching score for fingerprint comparison.

Our purpose for using DeepPrint was to achieve better matching performance due to several key advantages:

- Fixed-Length Representation: DeepPrint uses a fixed-length representation, which avoids the need for graph matching and is computationally efficient.
- High Discriminative Power: It maintains high discriminative power even with low-quality fingerprints.
- Performance Metrics: In a large-scale test with 1.1 million fingerprints from the NIST SD4 dataset, DeepPrint achieved a rank-1 search accuracy of 98.80% in just 0.3 seconds. In comparison, a top COTS matcher had a slightly higher accuracy of 98.85% but took 27 seconds to complete the search.

Additionally, we experimented with Verifinger and Morpho, but they did not yield satisfactory results. Thus, DeepPrint's combination of efficiency, speed, and high accuracy made it the optimal choice for our work.

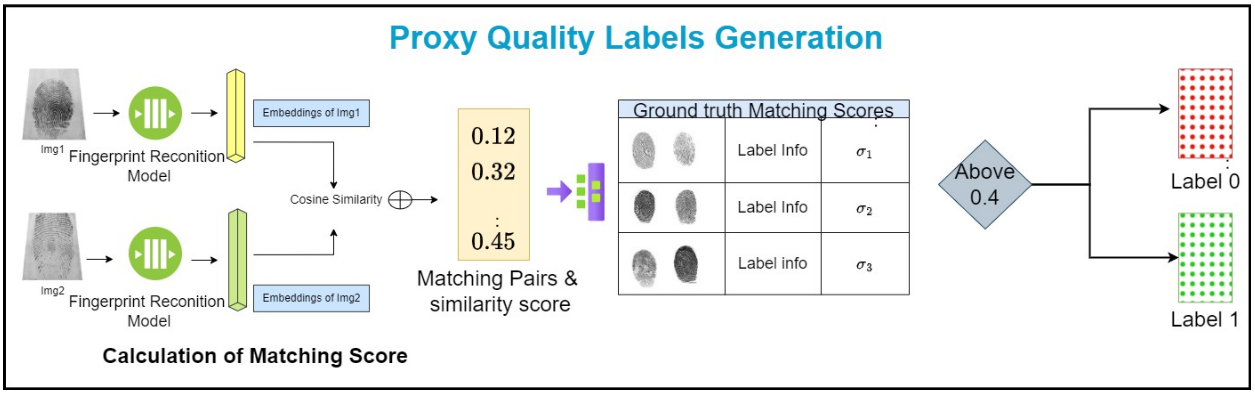



Figure 3.2: The methodology for generating pseudo ground truth labels by evaluating matching scores, a crucial step in our weakly supervised learning approach for assessing fingerprint quality.

### 3.2.2 Proxy Groundtruth label generation

As shown in Figure 3.2, the generation of proxy ground truth labels involves analyzing matching scores to produce quality indicators for fingerprints. This process is crucial for defining the quality of fingerprints in relation to their matching performance. The labels generated through this method inform our network about the expected quality outcomes based on matching scores, facilitating a targeted learning process that focuses on the most relevant features for quality assessment.

### 3.2.3 Quality Score Calculation and Network Training

The core of our methodology lies in the calculation of quality scores and the training of a neural network to accurately predict these scores. We utilize a ResNet architecture, or a similar quality regression network, which is fine-tuned using the labels generated from our initial processing steps. This network learns to differentiate between high and low-quality fingerprints based on their matching scores, as detailed in Figure 3.3. The training process involves a specially designed loss function that accounts for the matching scores and label information, allowing the network to adapt its predictions to closely match the ground truth data.

The network processes two distinct input images, $\mathcal{I}^1$ and $\mathcal{I}^2$, predicting a quality score for each, denoted as $\mathcal{P}^1$ and $\mathcal{P}^2$. Utilizing the matching score $\mathcal{M}$ from DeepPrint, a proxy ground truth for quality is obtained. This process necessitates both images to ensure the precise calculation of the quality score. Alongside the matching score, label information $\mathcal{K}$ augments the accuracy of the quality score calculation. The harmonic mean of the quality score, rather than the average matching score, is utilized for calculating the matching score. This use of harmonic average emphasizes the lower values in the set, capturing the impact of low scores more prominently, and consequently offering a more precise quality score for images with limited matching ability.

**Loss Function Overview:** The employed network, as shown in Figure 3.3, incorporates a specially designed loss function denoted as $\mathcal{L}_{(M,Q_i,K,W_i)}$ or $\mathcal{L}_Q$ to enhance its fine-tuning. In this notation:

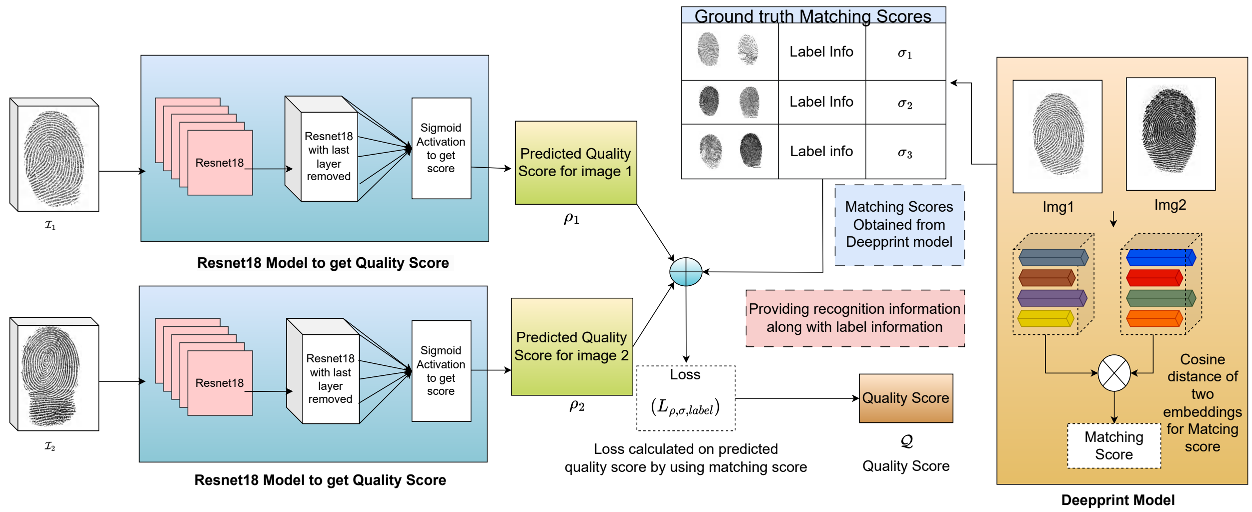



Figure 3.3: In this proposed method, the ResNet neural network takes in the input images, $(\mathcal{I}^1)$ and $(\mathcal{I}^2)$, and produces predicted quality scores, $(\mathcal{P}_I^1)$ and $(\mathcal{P}_I^2)$. These predicted scores are used in a loss function, along with proxy ground truth quality scores that are matching scores $(\mathcal{M})$ with their label info, generated by the DeepPrint model, in order to train the ResNet to produce scores that are as close as possible to the ground truth. The output of this process is a fingerprint quality score, $(\mathcal{Q})$.

- $Q_i$ specifies the two input images.
- $M$ specifies the matching score of these images.
- $K$ indicates whether the two images are from the same class or different classes.
- $W_i$ weights for balancing.

The primary objective of this loss function is to ensure robust performance under imbalanced data conditions. It aims to minimize the discrepancy between the predicted quality scores and the true quality of the input images, particularly in scenarios where:

- The label is 0 (indicating high quality) with a low matching score.
- The label is 1 (indicating low quality) and the matching score is low.

Conversely, the loss should increase when:

- There's a low matching score combined with a label of 1.
- There's a high matching score paired with a label of 0.

This loss function is applied as binary cross-entropy loss for labels 0 and 1.

Expressed as $\mathcal{L}_Q(\mathcal{M}, \mathcal{Q}_1, \mathcal{Q}_2, \mathcal{K}, \mathcal{W}_0, \mathcal{W}_1)$,:

$$\begin{aligned}\mathcal{L}_i(\mathcal{M}, \mathcal{Q}_i, \mathcal{K}, \mathcal{W}_0, \mathcal{W}_1) = \log \mathcal{Q}_i \, (-\mathcal{W}_1\mathcal{M} + \\ (1-\mathcal{K})\mathcal{W}_0(1-\mathcal{M}))\,, \; \forall i = 1, 2\end{aligned} \tag{3.1}$$

$$\mathcal{L}_Q(\mathcal{M}, \mathcal{Q}_1, \mathcal{Q}_2, \mathcal{K}, \mathcal{W}_0, \mathcal{W}_1) = \mathcal{L}_1 + \mathcal{L}_2 \tag{3.2}$$

$\mathcal{L}_Q$ is a combination of the above two losses where $\mathcal{L}_1$ is calculated for $\mathcal{Q}_1$ the predicted quality of *image-1* and the $\mathcal{L}_2$ is calculated on $\mathcal{Q}_2$ the predicted quality of *image-2* where $\mathcal{M}$ represents the Matching score obtained on a pair of images that are used as a proxy ground truth to finetune the network. $\mathcal{Q}_1$ and $\mathcal{Q}_2$ represent the Quality Score of *image-1* and *image-2* respectively and $\mathcal{K}$ represents the Label information that is 0 for different images and 1 for same images. When $\mathcal{K}$ is 1 loss functions try to reduce the loss directly proportional to $\mathcal{M}$ in predicting the $\mathcal{Q}_1$ and $\mathcal{Q}_2$ and when $\mathcal{K}$ is 0 loss function learns the loss in-versely proportional and try to reduce loss incorporates in predicting the $\mathcal{Q}_1$ and $\mathcal{Q}_2$ based on the inverse of $\mathcal{M}$.

**Weight Calculation:** To address the unbalanced nature of the training data, weights $\mathcal{W}_0$ and $\mathcal{W}_1$ are assigned to balance the loss function. The weights are determined based on the relative proportions of the samples in each class. Specifically, they are calculated as:

$$\mathcal{W}_k = \frac{1}{N_k}\left(\frac{N}{2}\right) \tag{3.3}$$

Where:

- $k$ represents the class label (either 0 or 1).
- $N$ is the total number of samples in the dataset.
- $N_k$ denotes the count of samples in class $k$.

The weights $\mathcal{W}_0$ and $\mathcal{W}_1$ effectively balance the contribution of each class by inversely scaling with their prevalence, ensuring that neither class dominates the loss due to its abundance or scarcity.

## 3.3 Label Information effect on Quality Scores

In the context of fingerprint image quality assessment in our research, we employ proxy ground truths obtained from a DeepPrint matcher. The DeepPrint matcher leverages label information to enhance recognition accuracy concerning quality scores.

**Labelling of Pairs:** Table 3.1 presents an overview of the labeling process. Our approach involves training a deep learning model using labeled data that consists of pairs of fingerprint images. Each pair can either be from the same finger or from different fingers. The DeepPrint matcher utilizes this label information to distinguish between genuine pairs and impostors.

**Training with Labels:** During model training, we incorporate these fake ground truth labels, which are derived from matching scores obtained from the DeepPrint matcher. These scores serve as indicators of the likelihood of a pair of fingerprint images matching, with high scores signifying a good match and low scores indicating a poor match. This approach allows us to harness recognition-based information that cannot be obtained through manual image quality annotation, enabling our model to predict the quality score of fingerprint images accurately.

## 3.4 Analytical Study on Match Score:

We have explored alternative approaches that show promise for future applications, especially in the context of large fingerprint image datasets. These methods not only open up new avenues for further research but also provide valuable insights into the potential enhancements of fingerprint image analysis on a larger scale. In this paper, we have introduced an approach that involves utilizing a pair of images during model training. However, for scenarios where only a single image is available, it becomes essential to derive a representative score that indicates its recognizability. This score guides the model in accurately assessing image quality.

To compute a match score for a single image, it is imperative to gather a collection of match samples encompassing a wide range of both good and bad matches.

**Match Score of Single Image** To derive a single representative score from multiple match scores, statistical methods such as mean, harmonic mean, and median percentile prove valuable. The selection between these methods depends on the specific dataset characteristics and requirements. Here's a concise overview of both methods

- **Harmonic Mean:**
  - The harmonic mean is a type of average that gives more weight to lower values in the set. It is particularly effective at handling situations where extreme outliers or very low scores need to be taken into account.
  - The harmonic mean is suitable when the impact of the worst or lowest scores in the dataset. It can help in scenarios where you want to ensure that the overall quality is not solely driven by a few high scores.
  - It is important to note that the harmonic mean is sensitive to extremely low scores, and a single low score can significantly affect the resulting average.
- **Percentile:**
  - The percentile represents the relative position of a particular score within a distribution. It indicates the percentage of scores that are equal to or below a given value.
  - Using percentiles allows you to determine where a specific score lies within the distribution of all scores. It helps capture the overall quality of a matched image compared to others in the dataset.
  - Percentiles are useful when you want to establish a threshold or cutoff point to classify images as poor or good quality based on their relative position in the score distribution.

The choice between the harmonic mean and percentile depends on the specific requirements and characteristics of the dataset, as well as the significance that we want to place on different scores.

| Label Info | Matching Score | Quality Score |
| --- | --- | --- |
| 0 | High | Low |
| 1 | High | High |
| 0 | Low | High |
| 1 | Low | Low |

Table 3.1: The table shows that when the matching score is high and the label information is positive, the quality score is high. Conversely, when the matching score is low and the label information is negative, the quality score is also low.The matching scores are used along with label information to predict the quality score of fingerprint images accurately.

**Dealing with non-mated pairs:** In this paper, we have conducted a comprehensive study encompassing the utilization of both mated and non-mated pairs, which are commonly referred to as genuine and impostor pairs, respectively. This inclusion allows us to thoroughly assess the performance of our proposed method across diverse scenarios, reflecting real-world fingerprint recognition challenges.

- **Exclusive Mated Pair Analysis**
  - Looking ahead, we envision conducting further experiments focused on exclusively using mated pairs. This aligns with real-world scenarios where the primary objective is to verify the matching of images for authorized individuals.
  - On the other hand, non-mated pairs present a distinct challenge, where the emphasis on recognizability may not be as critical.

  For these cases, we consider the possibility of assigning lower or even negative recognizability scores, given that these pairs do not belong to authorized individuals or are not among the designated matches.

- **Non-Mated Image Quality Assessment:**
  - One key aspect is the ability to predict scores for non-mated pairs, potentially indicating zero or negative recognizability. It would greatly impact the quality score, as it would enable us to identify images that do not match any individual.
  - Although, we recognize that solving this complex problem for a single image without using additional references presents significant challenges. Notwithstanding, we maintain a positive outlook on future developments that could potentially empower us to forecast such results.

By conducting these additional experiments, we aim to gain a deeper understanding of the effectiveness of our method across different use cases, and refine its performance to suit specific fingerprint recognition scenarios. This research contributes to the broader field of fingerprint image quality assessment and holds promise for advancing the reliability and accuracy of fingerprint recognition systems in practical applications.

| Database | Genuine Pair | Impostor Pairs | Total Pairs |
|---|---|---|---|
| $DB_A$ | 2800 | 4950 | 7750 |
| $DB_B$ | 280 | 45 | 325 |

Table 3.2: FVC dataset information

## 3.5 Experiments

This section begins with an overview of the datasets utilized for our experiments, outlined in subsection 3.6. It then proceeds to elaborate on the implementation process and the workflow, as detailed in subsection 3.6.1. The experimental setup and its specifics are discussed in subsection 3.6.2, followed by an analysis of the results derived from these experiments in subsection 3.6.3. Lastly, the limitations inherent to the proposed methodology are examined in subsection 3.6.4.

## 3.6 Dataset Overview

Our study leveraged the FVC 2004 dataset[34], a benchmark dataset for fingerprint image evaluation. The dataset's composition, detailed in Table 3.2, includes four distinct databases from different sensor technologies. For our experiments, we focused on two specific databases: $DB_A$ and $DB_B$. $DB_A$ was a primary source, providing 7751 pairs of genuine and impostor images. In contrast, $DB_B$ contributed 326 pairs of images. Each image within this dataset maintains a resolution exceeding 500 DPI, aligning with NFIQ's recommendations for capturing high-quality fingerprint images.

### 3.6.1 Workflow and Implementation details

To enhance the fingerprint recognition capabilities of the ResNet18 model, an alternative ground truth for assessing the quality of fingerprints was derived from matching scores courtesy of the DeepPrint model. This approach entails calculating matching scores through DeepPrint and correlating these scores with labels indicating the authenticity of the image pairs. A label of 1 signifies that the image pair is genuine, originating from the same set, while a label of 0 indicates an impostor pair coming from different sets.

**Employing DeepPrint for Alternate Ground Truth Generation:** This strategy enables the creation of a reliable substitute ground truth for quality assessment, essential for training the ResNet18 model effectively. This technique allows us to overcome the shortcomings of the NFIQ2 standard, which falls short of accurately determining the quality scores necessary for fingerprint verification.

The use of DeepPrint to create an alternative ground truth underpins our evaluation of the FRBQ methodology, showcasing its effectiveness in precise fingerprint image quality assessment. Matching scores and corresponding label information for select image pairs are detailed in Table 3.3. Additionally, this table compares the NFIQ2 and FRBQ quality scores for these pairs, showing the enhanced accuracy of our approach.

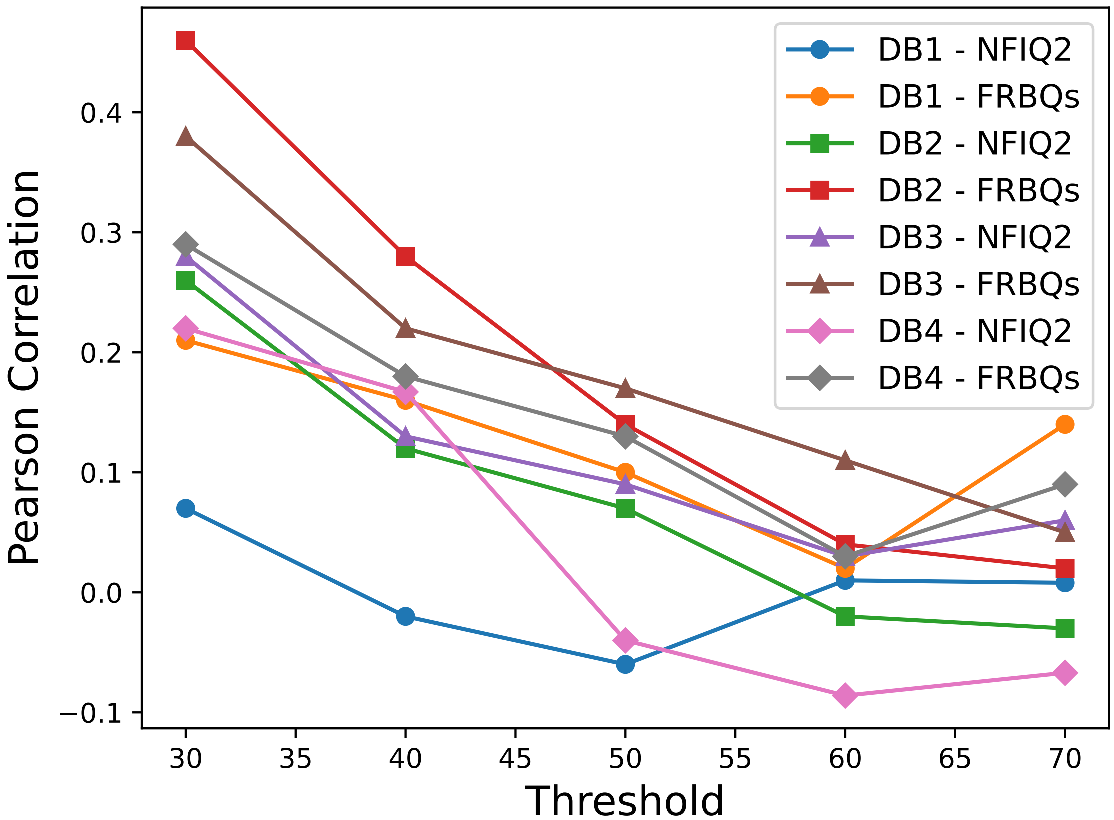



Figure 3.4: The performance of two quality score methods, NFIQ2 and FRBQ, were evaluated using matching scores generated from DeepPrint on four databases in the FVC Dataset A. The correlation between the quality scores and matching scores was analyzed at different quality thresholds. Results showed that FRBQ had higher correlation scores than NFIQ2 at varying quality thresholds for all the datasets.

### 3.6.2 Experiment Setup

In our study, we employed a pre-trained ResNet model, specifically fine-tuned with 7750 image pairs from Database A. For detailed information about the dataset used in our experiments, reference is made to Table 3.2, which outlines the FVC 2004 dataset details, including the number of genuine and impostor pairs, types of sensors used, and the recommended resolution for images.

**Feature Learning with ResNet:** ResNet architectures are renowned for their efficacy in image classification tasks, wherein early layers capture universal features such as textures and edges, while deeper layers are adept at identifying more intricate attributes like parts of objects and their shapes. By locking the early layers, our model capitalizes on these universal features, which are crucial for assessing image quality.

**Model Fine-tuning Process:** During fine-tuning, each image pair is processed through the model individually. A loss function that integrates the predicted scores, matching scores, and label data is employed to refine the model's ability to ascertain a quality score for identification purposes. Initially, the DeepPrint

| | | | | | |
|---|---|---|---|---|---|
| **Img1** |  |  |  |  |  |
| **Img2** |  |  |  |  |  |
| **Matching Score** | 0.25 | 0.66 | 0.42 | 0.05 | 0.27 |
| **FRBQs Img1** | 0.24 | 0.53 | 0.41 | 0.51 | 0.53 |
| **NFIQ2 Img1** | 56 | 39 | 71 | 36 | 26 |
| **FRBQs Img2** | 0.51 | 0.61 | 0.53 | 0.57 | 0.37 |
| **NFIQ2 Img2** | 72 | 19 | 61 | 23 | 5 |
| **Label** | 1 | 1 | 1 | 0 | 0 |

Table 3.3: The table presents information on image pairs, featuring their matching scores, FRBQ, and NFIQ2 scores, along with the label status for each pair. NFIQ2 scores are measured on a scale from 0 to 100, whereas FRBQ scores, ranging from 0 to 1, reflect the capability of the quality score to predict recognition performance. In instances where matching scores are low for distinct images, FRBQ scores are instrumental in pinpointing image pairs of high quality. Furthermore, FRBQ scores are comprehensive, encompassing scenarios where lower matching scores occur for identical images.

model generates a substitute ground truth for the fine-tuning of the ResNet model. After gathering all matching scores, the ResNet model is adapted by substituting its final layer with a one-dimensional fully connected layer, focusing on enhancing the accuracy of predictions by utilizing only this layer's output and freezing the preceding layers.

**Quality Score Estimation:** The determination of the quality score involves processing the image via the ResNet model, which then predicts a probability score indicative of image quality. The FRBQ (Fingerprint Recognition Based Quality) model, once trained, offers precise quality score predictions that mirror the recognition capabilities of the fingerprints.

***Experimental Setup Details:*** *The experiments were conducted using the PyTorch framework on an NVIDIA GPU, with the fine-tuning process optimized through the Adam algorithm. The experimental parameters were configured as follows: the training was conducted over 100 epochs, with a learning rate of* $10^{-5}$ *and a batch size of 8.*

### 3.6.3 Findings

The image pairs were evaluated by averaging their respective scores to maintain fairness in comparison. Scores exceeding a predetermined threshold were classified as high quality, indicating a favorable match. Additionally, matching scores derived from DeepPrint were assessed, and their correlation with the quality scores was analyzed using Pearson correlation coefficients, as depicted in the correlation graphs in Figure 3.4. These graphs incorporate data from Figure 4.1, which shows the quality scores for fingerprint images determined by NFIQ2 and FRBQ (N and F, respectively). The analysis involved adjusting the quality threshold across various datasets to benchmark FRBQ's efficacy against NFIQ2.

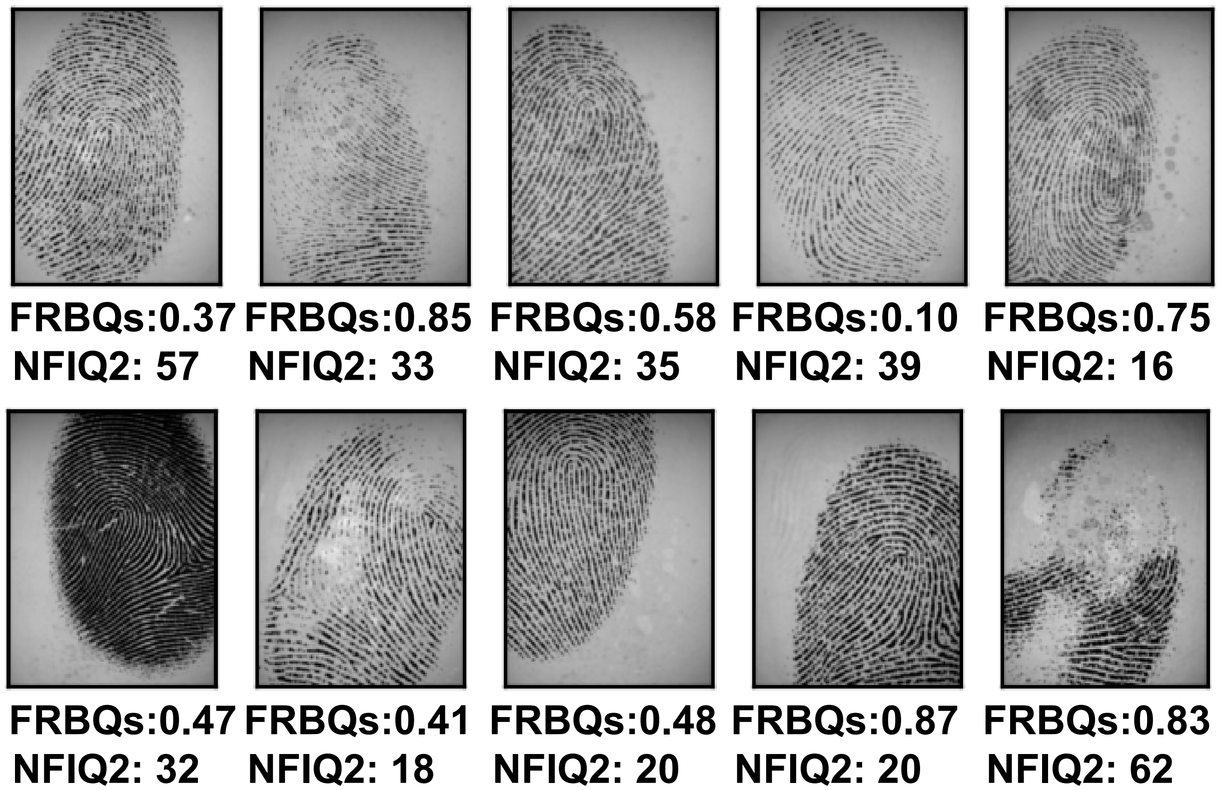



Figure 3.5: The quality scores for fingerprint images are obtained using two different methods, namely NFIQ2 and FRBQ, referenced as N and F, respectively.

Figure 3.5 provides an introductory comparison of the quality scores for fingerprint images obtained using the two different methods, NFIQ2 and FRBQ, which are crucial for understanding the foundation of our analysis. Following this foundational understanding, Figure 3.6 highlights a significant advancement the FRBQ method offers.

**Score Reliability:** NFIQ2 scores provide a dependable measure of recognition accuracy primarily for high-quality fingerprint images, but they do not universally apply. Specifically, for quality scores below 40, NFIQ2's predictive capability for image recognition performance diminishes.

The investigation revealed that the model proficiently identifies images with superior matching scores as high quality. Comparatively, it exhibits a higher correlation with the DeepPrint matcher across all quality thresholds (scaled from 0 to 100) relative to NFIQ2. This suggests that the introduced method is more adept at accurately evaluating the quality of matches and their applicability in fingerprint-matching systems.

Table 3.3 highlights the superiority of FRBQ over NFIQ2 in generating quality scores, with the following insights:

- A threshold of 0.4 was established as optimal for the FRBQ model by analyzing performance across various levels within the FVC 2004 databases.
- FRBQ consistently outperforms NFIQ2 in predicting recognition performance across different quality thresholds, rendering it a more effective tool for assessing the suitability of fingerprints for matching

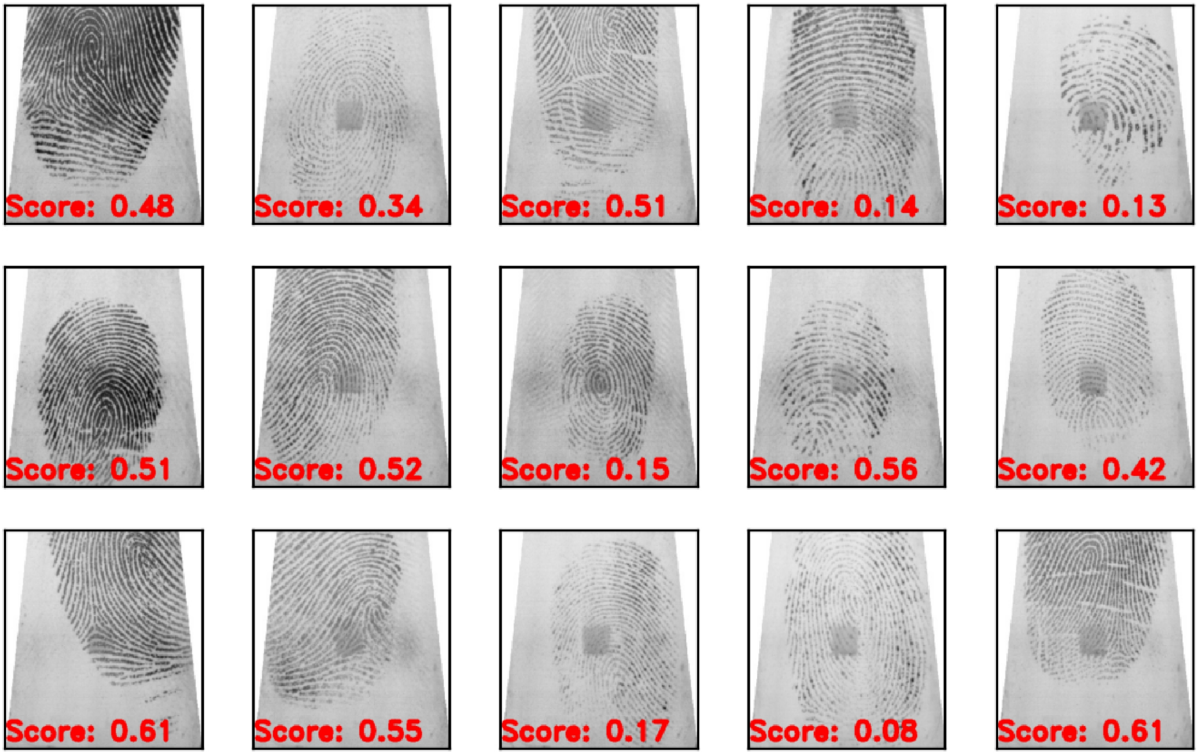



Figure 3.6: FRBQ provided quality scores for images that NFIQ2 could not predict scores on, even though these images were highly distorted and noisy.

applications. Moreover, FRBQ's generated quality score accurately forecasts the matching score, underscoring the advantage of integrating quality and matching data in fingerprint recognition endeavors.

The approach progresses by leveraging label and matching score data to compute precise quality scores, though the assessment process becomes more intricate with lower matching and image quality scores.

**Emerging Scenarios:**

1. Lower quality in one or both images of a pair may lead to an erroneously high matching score due to noise presence. Conversely, in scenarios of high scores, both images may inherently possess similar high quality, or low-quality images might inaccurately reflect high matching scores.

2. Addressing these complexities necessitates the utilization of an extensive and diverse dataset. Such a dataset should cover all imaginable conditions of image quality and matching scores, enriched with a robust collection of image pairs to enable comprehensive evaluation of label and matching score data.

A model's resilience to effectively manage noise is imperative, ensuring a clear demarcation between authentic and noise-generated matches, thereby facilitating precise quality evaluations even in challenging scenarios. The histograms of match scores and quality scores illustrate the correlation between predicted quality scores and average match scores. Match scores range from 0 to 100, while quality scores span from 0 to 1. As demonstrated in Figure 3.7, this distribution reveals a crucial insight: quality scores fill the gaps left by

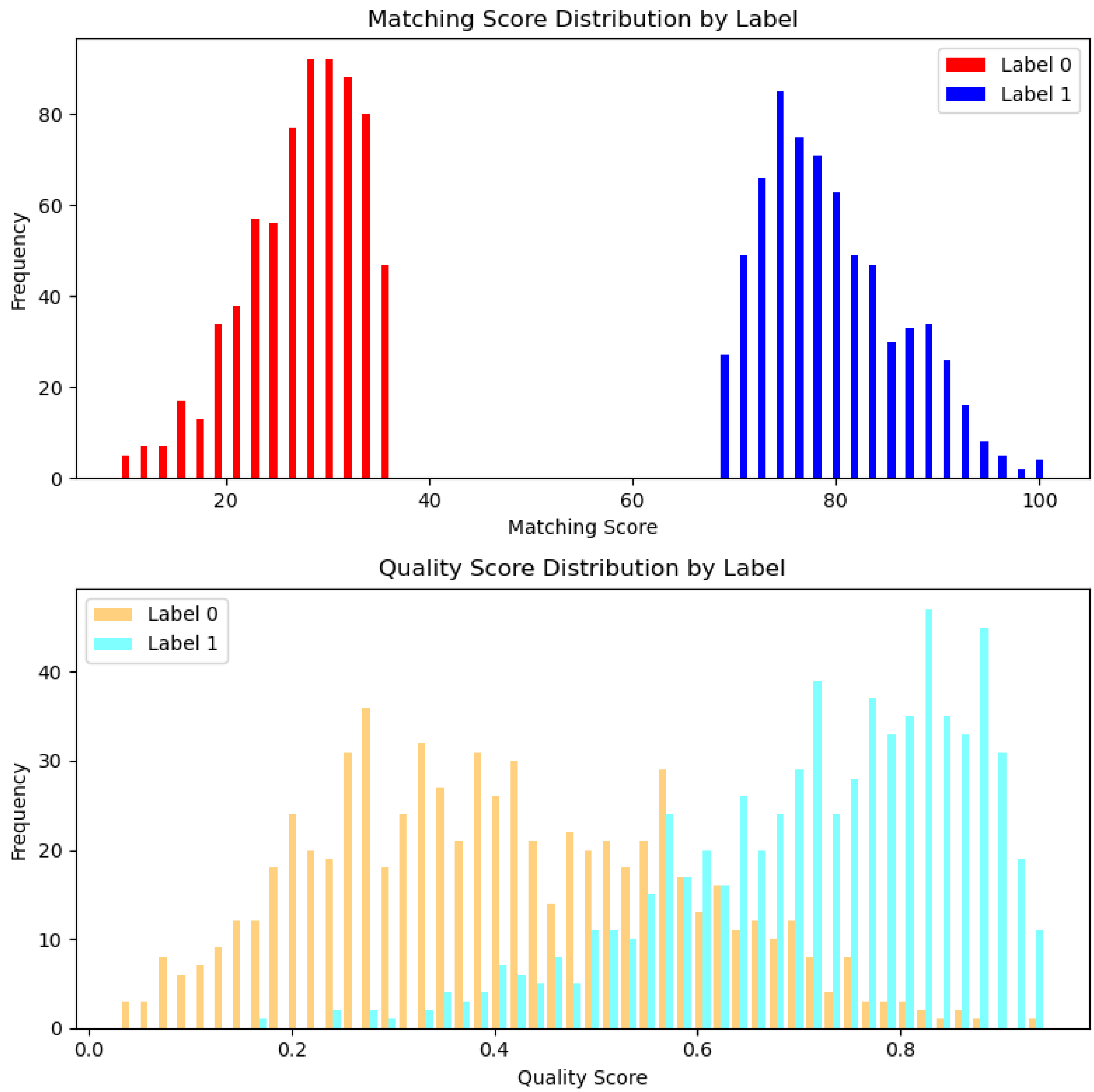



Figure 3.7: The histogram illustrates the match scores derived from the DeepPrint model and the corresponding quality scores obtained from our model, demonstrating improved separation between label 0 scores and label 1 scores.

missing match scores, showcasing their strong predictive capacity even when extensive matching data is unavailable.

### 3.6.4 Discussion and ablation study

This study benchmarks NFIQ2, revealing its shortcomings in accurately determining the quality of fingerprint images.

- Utilizing the specific NFIQ2.2.0 version brought to light its limitations, particularly its failure to provide quality scores for images of low resolution or compromised capture quality, resulting in a default score of zero.

- Conversely, FRBQ demonstrated its robustness by assigning meaningful scores to images that NFIQ2 failed to rate, as shown in Figure 6. This difference highlights FRBQ's enhanced capability to evaluate fingerprint image quality with greater detail, improving the accuracy of fingerprint recognition technologies.

- The consistent and reliable performance of FRBQ, especially with images deemed unassessable by NFIQ2, emphasizes its precision and dependability. Initial results suggest FRBQ's potential dominance in fingerprint recognition technology, particularly for handling images of inferior quality or those captured inadequately.

The findings of this research are pivotal in advancing more accurate and reliable fingerprint recognition systems, with wide-ranging applications in sectors like law enforcement, security, and access management.

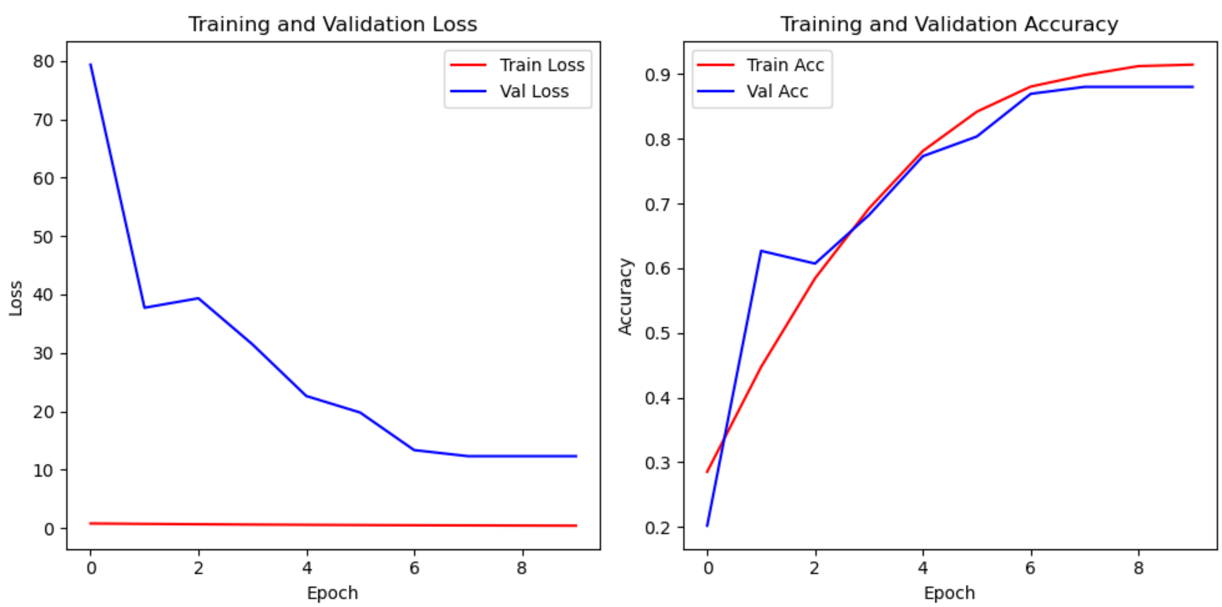



Figure 3.8: Training loss and accuracy results.

**Training of the Model:** Figure 3.8 displays the FRBQ model's loss and training progression, shedding light on its training efficacy. The graph of the loss curve depicts the optimization journey, while the training accuracy graph showcases the model's ability to learn from the data provided. Simultaneously, the graph for validation accuracy showcases the model's ability to generalize to new, unseen data, offering a holistic view of the model's convergence and its success in maintaining high accuracy levels for both training and validation sets.

## 3.7 Conclusion

In summary, the developed FRBQ model, which leverages deep learning to evaluate the quality of fingerprint images, demonstrates significant potential. It excels in producing reliable quality scores for newly scanned fingerprints, outperforming traditional models. The benefits of this method are manifold, including the improvement of fingerprint matching systems, reduction of manual checking and labeling efforts, and its compatibility with existing recognition frameworks. The inclusion of label data and matching scores provides a more comprehensive assessment of image quality. However, accurately evaluating images of poor quality presents a considerable hurdle, necessitating a nuanced approach to understand the relationship between image quality and matching accuracy. The solution to these challenges may be found in advanced algorithms adept at managing the complexities of data and incorporating additional types of information, such as specific features of the fingerprints. Furthermore, a broader training dataset could unveil more about the challenges associated with low-quality fingerprints, helping to surpass the current limitations of using proxy ground truth for training. Ultimately, the FRBQ model has the potential to enhance the precision of fingerprint recognition systems and streamline the process of quality assessment. Overcoming the obstacles related to the evaluation of low-quality images requires further research into optimizing the correlation between image quality and matching scores, aiming to improve the overall effectiveness of the FRBQ model in fingerprint recognition technologies.

*Chapter 4*

# Advancing FIQA with Age and Expressions: Introducing the U3FQ (Unified Tri-Feature Quality) Metric

## 4.1 Introduction

In recent advancements in biometric identification technology, the development of U3FQ marks a significant milestone. U3FQ integrates advanced machine learning and deep learning frameworks to analyze and interpret facial data accurately. This innovative approach allows for a nuanced quantification of the effects of age and expression on match scores, offering a comprehensive assessment tool that transcends traditional metrics by considering the dynamic nature of human faces. Positioned as a holistic, context-aware solution, U3FQ enhances biometric systems' reliability and effectiveness across diverse scenarios, establishing itself as a pivotal tool in the evolution of biometric identification technologies.

## 4.2 Theoretical Background

### 4.2.1 Facial Age Difference

The efficacy of face-matching systems is significantly influenced by the age difference between the anchor image and the comparison image, as shown in Figure 4.1. This influence varies notably with the anchor's age, necessitating a nuanced approach to modeling age difference penalties. For anchors aged between 20 and 30 years, negative age differences typically correlate with child images, which presents a considerable challenge due to the substantial change in facial features that occur during maturation. Conversely, for anchors over 35 years of age, negative age differences represent younger adult images, where changes in facial features are less pronounced.

One of the critical challenges in facial recognition is understanding how age differences between images can affect identity matching. Our algorithm, as detailed in Algorithm 1, specifically addresses this by calculating an age-based match score. This score takes into account the age difference ($d$) and the anchorage ($a$), reflecting the reality that the same age difference can have different implications depending on the age of the individual.

Our algorithm adapts to these variations with a conditional approach based on the anchor's age. In Figures 4.16, 4.17, and 4.15, we present a comprehensive analysis of how age variations impact facial matching accuracy across different age groups, as depicted by the False Non-Match Rate (FNMR) plots. These figures provide a detailed comparison of the recognition performance of various face recognition models in the context of age-related variations. Furthermore, Figures 4.18a, 4.18b, 4.19a, 4.19b, 4.20a, and 4.20b present the probability density plots, showcasing the distribution and variability of matching scores for different age groups across multiple face recognition models. This visual representation offers a nuanced understanding of the models' performance and shows how age factors into the accuracy and reliability of face recognition technologies.

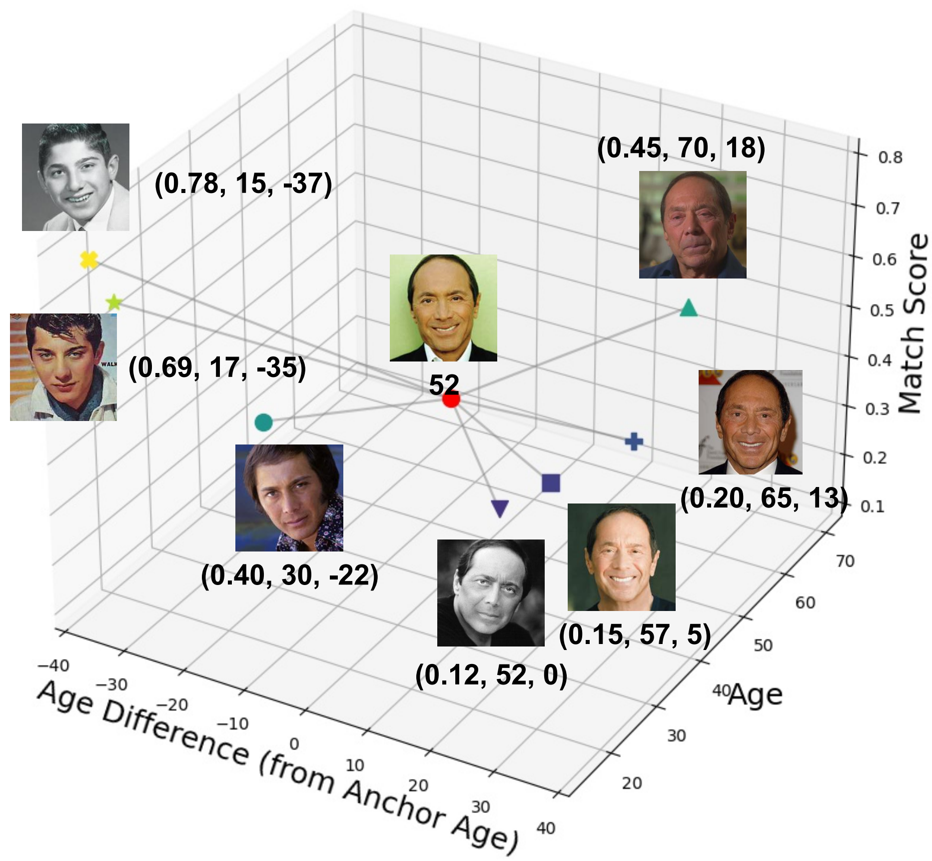



Figure 4.1: The efficacy of face-matching systems is significantly impacted by the noticeable age variation between the images being compared. The comprehensive triplet representation emphasizes the similarity distance, the specific age of the compared image, and the notable age difference relative to the anchor image, with Image 6 serving as the reference. The match score mentioned here is similarity distance, which means more the distance, the lesser the matching.

### 4.2.2 Development of the Nonlinear Function

We explored various mathematical models to formulate the function for calculating the match score. The nonlinear function used in our algorithm emerged as the most effective in correlating the age difference with the match score.

Initially, we tested polynomial and exponential functions. However, these models did not exhibit the same level of correlation with the match scores as our chosen function. The nonlinear function we adopted uniquely captures the nuanced effects of age differences, particularly in how these differences impact the recognition process at different ages.

Adopting this nonlinear function in our algorithm demonstrates its superior ability to accommodate the complex nature of age-related changes in facial features. The mathematical formulation is discussed in detail in Algorithm 1. This approach provides a more accurate and reliable method for facial recognition across different age groups, enhancing the overall effectiveness of identity-matching systems.

The implications of our findings are substantial. By integrating a logarithmic function into our model, we effectively capture the complex influence of age differences on facial recognition. This methodology facilitates a more refined and accurate evaluation of facial matches, particularly when there is a significant age disparity between the images under comparison.

### 4.2.3 Analysis of the Correlation Matrix

The correlation matrix derived from our dataset provides insightful observations regarding the relationships between various factors in our facial recognition (FR) model. Notably, the matrix reveals a significant correlation between the age difference-based match score and the overall match score, which underscores the effectiveness of our model in capturing the nuances associated with age differences in FR.

The correlation matrix presented in Figure 4.4 (sub-figure 2) substantiates our approach in accommodating age variations within facial recognition algorithms. The pronounced correlation observed between the age-adjusted match score and the comprehensive match score underscores the effectiveness of our method. Simultaneously, the negligible correlation between mere age differences and the match score validates our decision to employ a logarithmic function, which offers a more sophisticated solution than simpler linear models.

**Key Observations:** A critical finding from the matrix is the strong correlation between the age difference-based match score (Adj_MS) and the normalized combined score (Norm Combined Score). This correlation indicates that our model effectively adjusts match scores based on age differences, a crucial aspect in improving the accuracy of facial recognition across varying age groups.

**Lack of Correlation with Age Difference Alone:** Interestingly, the matrix shows a lack of significant correlation between the age difference and the match score (Adj_MS). This observation reinforces the need for a more sophisticated approach, like our nonlinear logarithmic function, to accurately capture the impact of age differences. Simple linear or direct correlations do not adequately represent the complex dynamics of age-related changes in facial features.

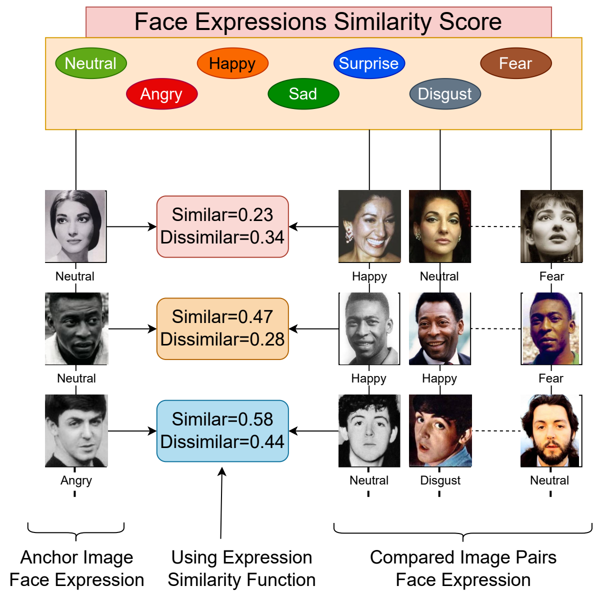



Figure 4.2: Demonstrates the process of calculating similarity and dissimilarity scores for facial images. By comparing emotion similarity within the same image set, each face is assigned a score reflecting its degree of resemblance or deviation. This approach enables a nuanced assessment of facial identity through a concise scoring system.

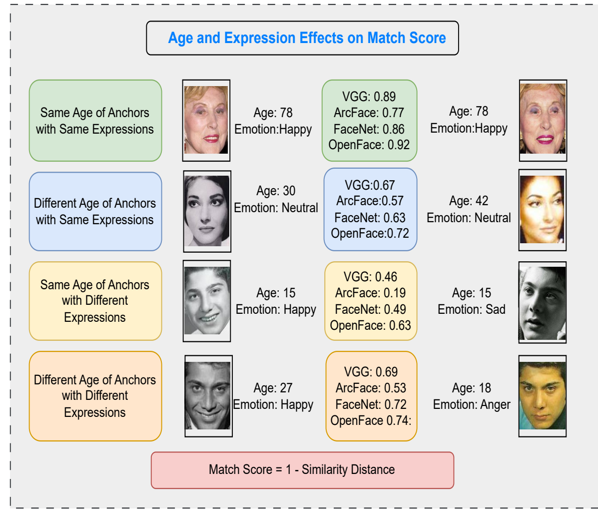



Figure 4.3: Showing the variability in match scores under diverse scenarios, highlighting how age differences and facial expression dissimilarities between anchor and target images influence recognition accuracy. It contrasts the stability in scores when comparing images of the same age and expression against the fluctuating scores observed with varying ages and dissimilar expressions, underscoring the dynamic nature of facial recognition performance.

### 4.2.4 The Influence of Facial Expressions

Similar to age variation in facial images, which can cause identity loss in biometric authentication, facial expressions also impact the matching performance of biometric systems. In unconstrained environments, varying facial expressions are common, and their influence extends beyond mere variations in illumination and pose. These expressions, driven by human emotions such as happiness, sadness, anger, fear, and others, significantly affect the recognition process.

Many face recognition models can predict emotions; however, this ability largely depends on the dataset's diversity. If a particular emotion is present in the dataset, the model can predict it, but its depth of understanding is limited. Consequently, while face recognition models can provide a rough estimation of emotion, they don't significantly impact the matching.

In our study, as shown in Figure 4.2, we develop similarity and dissimilarity metrics to analyze facial expressions and emotions. This approach allows us to assign a quality value associated with these features. We observed that weak emotions such as smiling, neutrality, and fear have a lesser impact than stronger ones. Accordingly, our paper assigns different values to these emotions, reflecting their varying influence on the authentication process. Ultimately, we compute a match score that considers facial expressions, providing a more nuanced understanding of their effect on biometric systems.

The similarity in facial expressions between two images notably influences recognition performance, as variations in expressions can distort critical facial features used in establishing a match. This impacts the

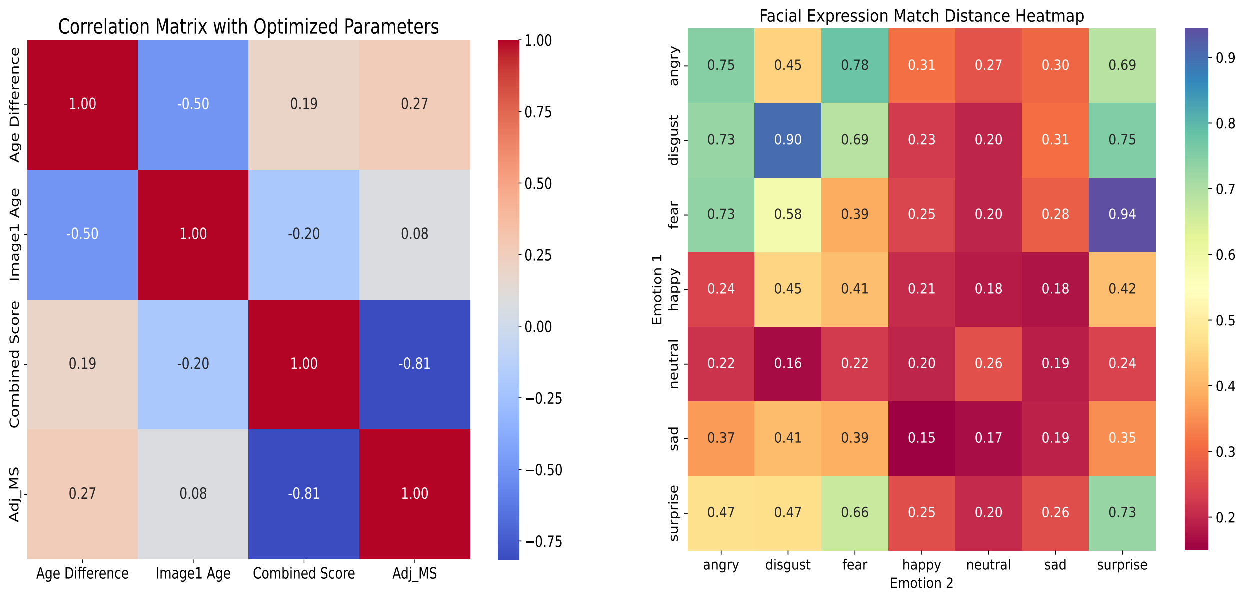



Figure 4.4: **Sub-Figure 1**: Correlation Matrix of Key Variables in Facial Recognition Model. This heatmap displays the correlations among age difference, image age, normalized combined score, and age-based match score (Adj_MS). The matrix highlights the strong correlation between the age-based match score and the overall match score, while indicating a negligible correlation between age difference alone and the match score, underscoring the necessity of our sophisticated model approach. **Sub-Figure 2**: The differential impact of facial expressions on the match score is notable, with weak emotions having a relatively constant effect and strong emotions significantly modifying the score proportionally to their intensity.

overall quality of recognition. Figure 4.4 (sub-figure 2) shows the impact of facial expression discrepancies on matching performance, evidenced by average match scores across expression pairs.

Our methodology ensures a more refined and context-sensitive assessment of facial expression similarity, considering the physical resemblance and the nuanced expressive context of each face. This approach leads to a more accurate and realistic evaluation of facial images, particularly relevant in dynamic real-world scenarios where facial expressions can vary significantly.

To empirically underpin this observation, we present Detection Error Tradeoff (DET) plots that demonstrate the variance in performance with different age groups for all four models: VGG-Face[62], OpenFace[1], ArcFace[16], and FaceNet[49]. Here, we have added the DET plots from VGG-Face in Figure 4.14, which show the False Non-Match Rate (FNMR) for different age groups. These plots highlight that there is a pronounced increase in FNMR as the age difference becomes more negative. The trend gradually inverts with increasing anchor age, reflecting the maturation and stabilization of facial features over time.

## 4.3 Match Score in different Scenario

This analysis serves as the concluding remark, highlighting the implications of match score performance under varying scenarios. It specifically examines how matching accuracy is influenced when the ages and facial expressions in the compared images are either identical or different. These variations have a significant impact on matching accuracy, as demonstrated in Figure 4.3.

Figure 4.3 critically analyzes how match scores fluctuate under different scenarios, focusing on the variations caused by age differences and facial expression dissimilarities between anchor and target images. It visually captures the influence of these two key factors on the effectiveness of face recognition systems. As the age gap between the compared images widens, or as the disparity in their facial expressions increases, the match scores exhibit notable changes.

This analysis is crucial for understanding the dynamic nature of facial recognition accuracy and its sensitivity to age-related and expressive variations. The data presented in the figure is instrumental in developing more adaptive and nuanced face recognition technologies that can effectively handle a wide range of real-world variations.

Our study reveals that these factors profoundly affect the performance of advanced models like VGG-Face and ArcFace. By acknowledging and accounting for age differences and dissimilarities in facial expressions, our model presents a more robust and reliable approach to facial recognition. .

## 4.4 Formulations and Optimization

Building on the observations from empirical evidence, we formulate the mathematical model to incorporate a logistic adjustment based on age difference and anchor age into the facial match score. The adjusted match score function is defined as follows:

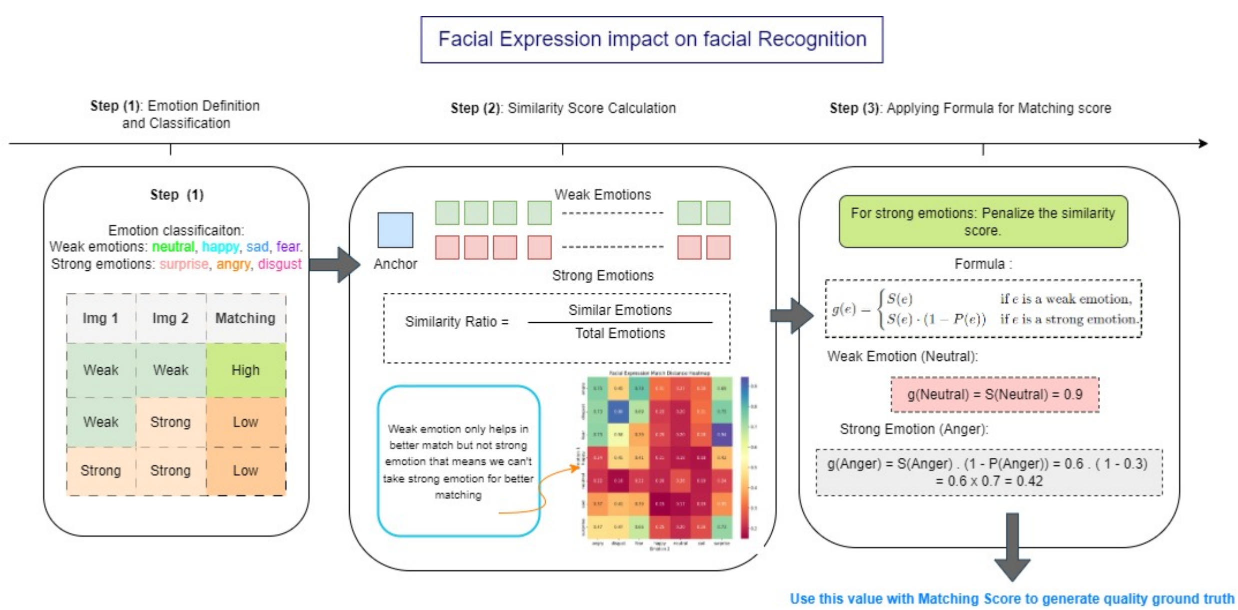



Figure 4.5: Showing Face Expression Impact on Face Recognition.

Figure 4.5 shows Face Expression impacts on matching performance. We can see weak emotions impact less and strong emotions impact more; hence, we penalize the strong emotions while using them for matching.

**Algorithm 1** Calculate Age-Based Match Score Based on Age Difference and Anchor Age

```
procedure CALCULATEAGEBASEDMATCHSCORE(d, a, params)
    Initialize Λ, κ, ξ_0, α, β, γ, δ, ϵ, ζ, η from params
    if a ≤ 30 then
        ξ ← α · d + β · a + γ · d^2
    else
        ξ ← δ · d + ϵ · a + ζ · log(max(a, 1)) + η · d · a
    end if
    score ← Λ / (1 + e^{−κ·(ξ−ξ_0)})
    return score
end procedure
```

$$f(d,a) = \begin{cases} \frac{\Lambda}{1+e^{-\kappa(\xi-\xi_0)}} & \text{if } a \leq 30, \\ \frac{\Lambda}{1+e^{-\kappa(\xi-\xi_0)}} & \text{if } a > 30, \end{cases} \tag{4.1}$$

where:

- $\xi = \alpha d + \beta a + \gamma d^2$ for $a \leq 30$,
- $\xi = \delta d + \epsilon a + \zeta \log(\max(a, 1)) + \eta da$ for $a > 30$,
- $d$ represents the age difference between the anchor and the comparison image,
- $a$ denotes the anchor's age,
- $\Lambda$ is the curve's maximum value,
- $\kappa$ is the logistic growth rate,
- $\xi_0$ is the x-value of the sigmoid's midpoint,
- Parameters $\alpha$, $\beta$, $\gamma$, $\delta$, $\epsilon$, $\zeta$, and $\eta$ control the shape of the function.

The reason for using 30 as a reference point is based on our experiments and observations in the dataset. We found that after the age of 30, facial dynamics do not change significantly, making it a suitable reference. Initially, we considered using 25, but ultimately chose 30 because facial features tend to stabilize after this age.

Our methodology also accounts for the subtle yet significant influence of facial expressions on the match score. This is achieved through the facial expression impact function $g(e)$, which distinguishes between 'weak' and 'strong' emotions, as detailed below. The rationale for using 'weak' and 'strong' emotions as key indicators is evident in the confusion matrix. Weak emotions generally result in higher matching performance, whereas strong emotions lead to lower matching performance.

$$g(e) = \begin{cases} c & \text{if } e \text{ is a weak emotion,} \\ d \cdot \text{EXPR\_SCORE}(e) & \text{if } e \text{ is a strong emotion,} \end{cases} \tag{4.2}$$

where $c$ is a constant factor for weak emotions, and $d$ scales the expression score $\text{EXPR\_SCORE}(e)$ for strong emotions.

Here, utilizing the equation 4.2 designed for face expression similarity function. Our function is calibrated to assign higher scores to faces that are similar, effectively distinguishing them from dissimilar ones.

A key feature of our approach is the nuanced consideration of facial expressions in determining these scores. For instance, neutral expressions, which are generally more predictable and consistent for recognition purposes, are assigned the highest scores. In contrast, despite being similar, faces exhibiting strong emotions such as surprise or happiness receive comparatively lower scores. This adjustment acknowledges the impact of expressive variability on the recognizability of faces.

These formulations and empirical insights collectively enhance the fidelity of the FIQA model's predictions. By incorporating the dynamics of human aging and expressions, we ensure that our facial recognition system is secure and user-friendly, accommodating the complexities of human features and behaviors.

Once the functions $f(d, a)$ and $f(e)$ are computed, they are integrated with the Normalized Average Match Score (NAMS) to derive the Normalized Age Variation Score (NAVS) and Normalized Emotion Similarity Score (NESS). The integration process combines these individual scores to produce comprehensive metrics that reflect both age variations and emotional similarities in facial images. Specifically, NAVS and NESS are formulated as follows:

$$\begin{aligned} \text{NAVS} &= \text{Integration}(f(d, a), \text{NAMS}), \\ \text{NESS} &= \text{Integration}(f(e), \text{NAMS}). \end{aligned}$$

These integrated scores, NAVS and NESS, provide a nuanced understanding of facial image quality, capturing the subtle interplays between age-related features, emotional expressions, and overall image match quality.

The algorithm detailed below outlines the process for computing the contextual quality score and estimating the age for a given input image using a ResNet model. The procedure leverages a feature vector that encompasses age, expression, and congruence score, which are derived from the input image and used to predict the quality score.

## 4.5 Architecture

**Algorithm 2** U3FQ: Unified Tri-Feature Quality Assessment for Contextual Facial Image Quality

**Require:** Single input image $I$, ResNet model $RN$, age $a$, expression $e$, match score distance models $M = \{M_1, M_2, M_3, M_4\}$
**Ensure:** U3FQ Score or Quality Score
$S \leftarrow 0$
MatchScore $\leftarrow 0$
NAMS $\leftarrow 0$ ▷ Normalized Average Matching Score
NAVS $\leftarrow 0$ ▷ Normalized Age Variation Score
NESS $\leftarrow 0$ ▷ Normalized Emotion Similarity Score
**for all** model $\in M$ **do**
    $d \leftarrow$ ComputeMatchScoreDistance$(I, model)$
    MatchScore $\leftarrow$ MatchScore + Normalize$(d)$
**end for**
NAMS $\leftarrow$ Average(MatchScore)$/(0.476)$
**for all** model $\in M$ **do**
    AgeDiffScore $\leftarrow$ AgeDiffScore + $f_{\text{age}}$(NAMS, $a, d$)
    EmotionSimScore $\leftarrow$ EmotionSimScore + $f_{\text{emotion}}$(NAMS, $e$)
**end for**
$S \leftarrow 0.1 \cdot$ NAMS $+ 0.7 \cdot$ NAVS $+ 0.2 \cdot$ NESS
**procedure** U3FQ_ASSESSMENT$(I, RN, m = 100)$
    QualityScores $\leftarrow$ []
    **for** $i = 1$ to $m$ **do**
        quality $\leftarrow RN$.Predict$(I, S)$
        QualityScores $\leftarrow$ QualityScores + [quality]
    **end for**
    finalQuality $\leftarrow$ Average(QualityScores) **return** finalQuality
**end procedure**

The U3FQ algorithm initially commences with the detailed and precise calculation of the match score distance using four distinct and highly sophisticated Face Recognition models, denoted as $M = \{M_1, M_2, M_3, M_4\}$. This crucial distance metric, accurately and effectively represented as $d$, is computed based on the pairwise discrepancies in the features meticulously extracted by each model for a specific given image $I$. It is vital to note that different models have varying embedding spaces, leading to significant and notable differences in distance calculations between images. To effectively address this challenge, we diligently apply thorough and careful normalization to these scores. The normalized scores are then rigorously and systematically evaluated against various thresholds, acknowledging that different models have their own unique and distinct threshold criteria. For instance, models like VGG and FaceNet utilize a threshold of 0.4, whereas ArcFace employs a threshold of 0.6. Through a process of comprehensive maximum voting, we establish an optimal threshold value of 0.476, which is found to be universally effective for most images. Subsequently, this NAMS is used to derive the NAVS and the NESS.

The age difference function, expressed as $f(d, a, NAMS)$, adjusts $d$ in accordance with the age $a$ of the anchor image to calculate NAVS. In contrast, NESS is determined using the expression impact function $g(e)$, which modifies the congruence score based on the facial expression $e$. Here, $c$ represents a constant factor for weak emotions, and $d$ is a scaling factor for strong emotions. This adjustment is complemented

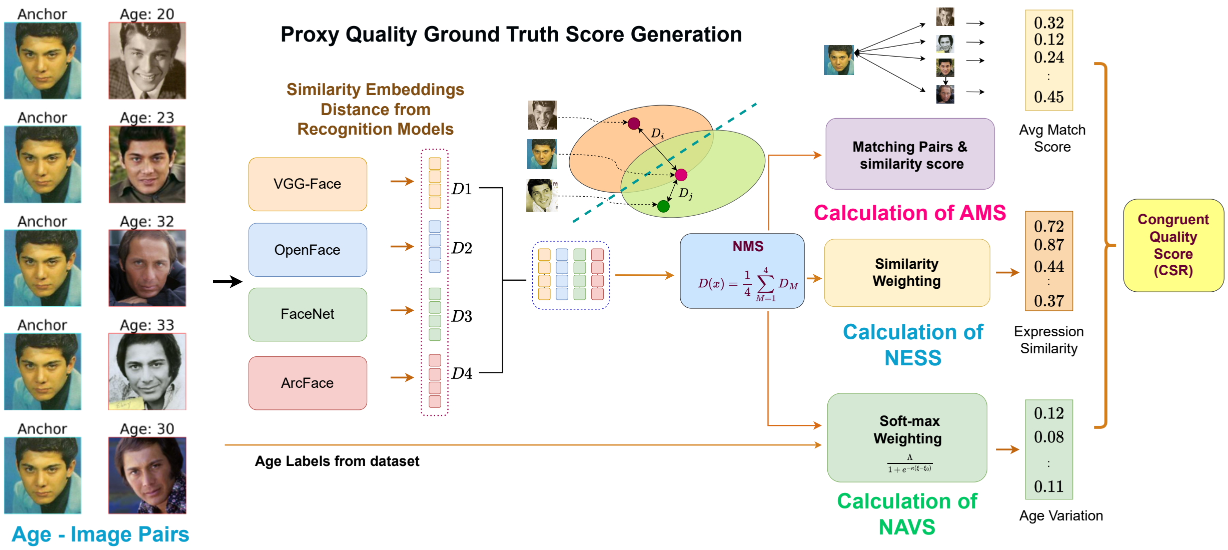



Figure 4.6: The figure presents a method for generating pseudo ground truth labels in face recognition by assessing age-related variations and expression similarity. It starts by calculating similarity distances between images of the same individuals at different ages using face recognition models. These distances are then normalized and combined with age and expression data to get NMS(Normalized Match Score, NESS(Normalized Emotion Similarity Scores) and NAVS (Normalized Age Variation Scores). These Scores are combined based on weighting to provide a combined score that is used are for fine-tuning regression network, leading to a comprehensive quality score that encapsulates recognition accuracy, age differences, and expression similarities.

by the expression score EXPR_SCORE$(e)$. Given that the emotion parameters are also derived from face recognition models, their impact is considered in conjunction with NAMS.

The algorithm calculates a composite score $S$ that integrates three key elements: the matching score, age difference score, and emotion similarity score. These elements are combined in a weighted sum manner, where each feature is assigned a specific weight based on its relative importance. In this revised approach, the weights are 0.1 for the matching score, 0.7 for the age difference score, and 0.2 for the emotion similarity score. This weighting scheme places a higher emphasis on the age difference score, reflecting its greater significance in the evaluation process.

The age difference score and emotion similarity score are derived from the matching score, but are modified by functions that introduce nonlinearity, accounting for variations in age difference and expression. These functions ensure that the scores reflect minute aspects of the facial comparison.

The composite score $S$ is calculated as follows:

$$S = 0.1 \cdot M + 0.7 \cdot f_{\text{age}}(M, a) + 0.2 \cdot f_{\text{emotion}}(M, e)$$

Here, $M$ represents the basic matching score, $a$ is the age difference, $e$ is the emotional expression, $f_{\text{age}}$ is the function modifying the matching score based on age difference, and $f_{\text{emotion}}$ is the function modifying the matching score based on emotional similarity.

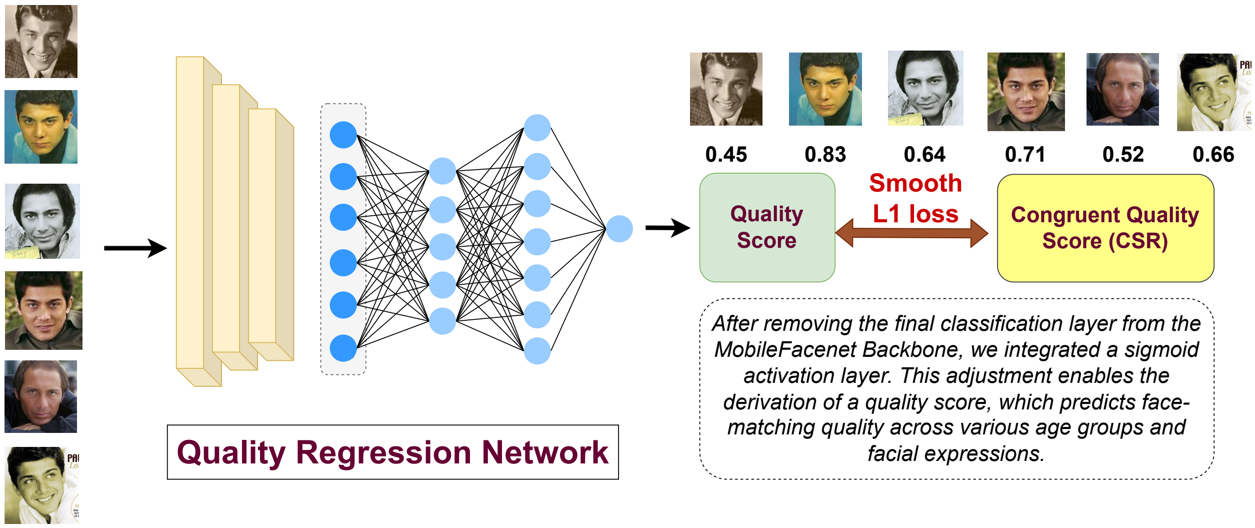



Figure 4.7: Regression Network Utilizing Feed-Forward Mechanism for Precise Quality Score Estimation.

A set of stochastic embeddings are generated through the ResNet model $RN$ across $m$ iterations to provide robust estimates of the image quality $Q$ and the subject's age. The embeddings are processed to yield a final quality score, reflecting the stability and robustness of the features in the presence of inherent variabilities in facial images.

This mathematical and algorithmic formulation of the U3FQ model demonstrates a robust mechanism for assessing facial image quality, providing insights into the complex interplay between age, expression, and recognition robustness. The model's efficacy is further corroborated through empirical evaluations, showcasing its potential to enhance the performance of biometric systems significantly.

## 4.6 Regression Network and Quality Estimation

We have advanced and thoroughly refined an existing Convolutional Neural Network (CNN), originally pre-trained extensively for face recognition tasks, through a meticulous process of fine-tuning. This established approach of expertly adapting deep learning models to tasks closely akin to their initial training has been consistently and effectively demonstrated in numerous influential studies. In figure 4.7 shows overview of Quality Regression Network used in this work. Such versatile networks have been successfully repurposed for detecting a wide range of facial attributes distinct from identity, including gender, age, and race. In the specific context of comprehensive face quality assessment, it is firmly posited that a robust feature vector containing highly discriminative facial information should inherently encapsulate critical aspects of image quality.

For our specific adaptation, we selected the ResNet50 architecture as the foundational network. During the fine-tuning process, we removed the classification layers and augmented the network with fully connected layers, which were then fused with the existing feature vector. This amalgamation was subjected to a sigmoid activation function, designed to yield a quality score.

Table 4.1: AOC at FMR of $1 \times 10^{-2}$ , $1 \times 10^{-3}$ and $1 \times 10^{-4}$. The Blue color text indicates the best overall performance, whereas green represents the second best in comparison, and red signifies the lowest performance.

| LFW | | | | |
|---|---|---|---|---|
| Method | FMR@1e-2 | FMR@1e-3 | FMR@1e-4 | Avg |
| BRISQUE [39] | 0.0467 | 0.0900 | 0.1279 | 0.1127 |
| BLINDS-II [47] | 0.1944 | 0.2354 | 0.2765 | 0.2612 |
| RankIQA [33] | 0.1346 | 0.1120 | 0.1459 | 0.1435 |
| PFE [26] | 0.2035 | 0.2557 | 0.2905 | 0.2499 |
| SDD-FIQA [44] | 0.8101 | 0.7881 | 0.7784 | 0.7979 |
| SER-FIQA [61] | 0.5673 | 0.6534 | 0.7477 | 0.6701 |
| U3FQ (Ours) | 0.8160 | 0.7653 | 0.7880 | 0.8035 |
| Adience | | | | |
| Method | FMR@1e-2 | FMR@1e-3 | FMR@1e-4 | Avg |
| BRISQUE [39] | 0.1845 | 0.2103 | 0.2412 | 0.2235 |
| BLINDS-II [47] | 0.1856 | 0.1546 | 0.1476 | 0.1710 |
| RankIQA [33] | 0.3412 | 0.2978 | 0.2876 | 0.3063 |
| PFE [26] | 0.3526 | 0.2768 | 0.2823 | 0.2870 |
| SDD-FIQA [44] | 0.5970 | 0.6423 | 0.5720 | 0.5996 |
| SER-FIQA [61] | 0.5123 | 0.5687 | 0.4562 | 0.4890 |
| U3FQ (Ours) | 0.7036 | 0.6782 | 0.5610 | 0.6539 |
| AgeDB | | | | |
| Method | FMR@1e-2 | FMR@1e-3 | FMR@1e-4 | Avg |
| BRISQUE [39] | 0.2856 | 0.3235 | 0.3656 | 0.3123 |
| BLINDS-II [47] | 0.3781 | 0.3452 | 0.3708 | 0.3689 |
| RankIQA [33] | 0.3215 | 0.3076 | 0.2765 | 0.2887 |
| PFE [26] | 0.3892 | 0.3187 | 0.2956 | 0.3054 |
| SDD-FIQA [44] | 0.7292 | 0.7238 | 0.7563 | 0.7320 |
| SER-FIQA [61] | 0.6238 | 0.5982 | 0.6286 | 0.6129 |
| U3FQ (Ours) | 0.7630 | 0.7432 | 0.7412 | 0.7520 |

Crucially, we implemented a training strategy where the weights of the pre-existing layers were frozen, ensuring that only the newly integrated layers were subject to training. This training utilized the pseudo ground truth quality labels generated in the preceding step. The outcome of this refined model is a quality score, ranging from 0 to 1, which correlates with the performance of face recognition, offering a robust measure of the quality of facial images in terms of recognition efficacy.

Table 4.2: Summary of the Experimental Setup

| Dataset | #Images | #IDs | Main Quality Factors‡ | | |
|---|---|---|---|---|---|
| | | | P-I | AV-E | N-D |
| AgeDB [40] | 16,487 | 568 | H | H | M |
| Adiance [17] | 5,000 | 1,159 | H | H | L |
| LFW [28] | 5,000 | 1,135 | M | H | H |
| MEDSII [20] | 1,306 | 518 | M | H | L |

† P-I - Pose and Illumination; AV-E - Age-Variation, Expression; N-D - Other Noise & Distortions - Scale.

‡ L - Low; M - Medium; H - High; Lr - Large; Values estimated subjectively by the authors.

## 4.7 Experiments And Results

The AgeDB dataset plays a critical role in our comprehensive study, as cited in Moschoglou et al. [40]. This dataset, comprising 16,487 images, is a foundational resource for examining age variations across different identities. A key visual element in our analysis is presented in Figure 4.8. This figure is composed of two informative pie charts. The first chart offers a detailed description of the age group distribution within the AgeDB dataset, providing a clear overview of the demographic composition. The second chart is particularly insightful, highlighting the age differences between pairs of images. This aspect is fundamental for understanding and improving identity matching in the context of age-related changes.

Figure 4.8 is pivotal in our study, showing the distribution of age gaps between image pairs and the overall age distribution of the dataset.

As detailed in Table 4.2, a key aspect of our analysis involved generating approximately 279,000 pairs from AgeDB. This was achieved by using 568 subjects, each with 15-20 images, resulting in a total of 279K pairs to cover a wider range of identities. An average match score was computed from about 20 images for each identity. This approach allows for in-depth insights into age-related identity matching.

Figure 4.13 shows the intricate distribution of the combined match scores, which are influenced by variations in age and facial expressions. These scores have been utilized as pseudo-ground truth labels to generate quality scores. This depiction showcases the correlation between age and expression factors in facial recognition and highlights their collective impact on the perceived quality of images. The visual representation is a crucial tool in understanding how these variables interact to influence the overall effectiveness of face image quality assessment methodologies.

Additionally, we include the LFW [28] and Adience [17] datasets in the table, while MEDSII [20] is presented in the distribution but not included in the table. These datasets provide diverse facial images, enabling a comprehensive analysis and demonstrating the robustness of our methodologies in age-variant facial recognition.

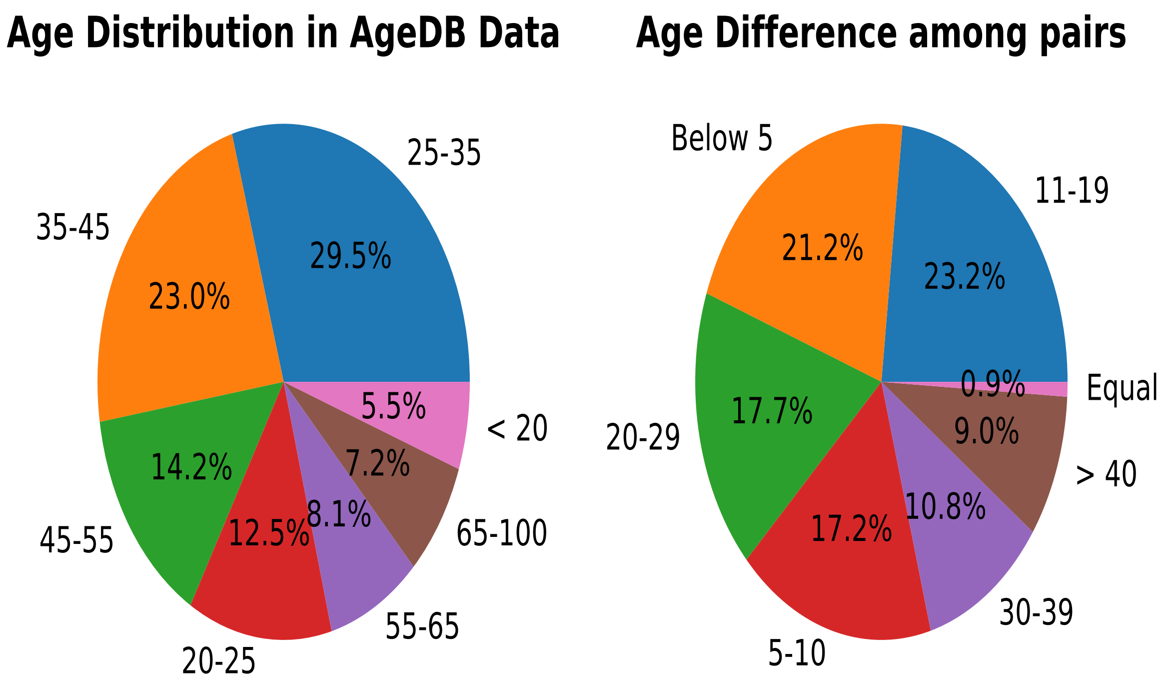



Figure 4.8: Distribution of Age-groups in AgeDB dataset.

### 4.7.1 Implementation Details and Setup

Our computational network is developed using the PyTorch framework following same as [44] and operates on a machine equipped with four NVIDIA GeForce RTX 2080 Ti. For preprocessing, face images are uniformly aligned, scaled, and cropped to a resolution of $112 \times 112$ pixels utilizing the MTCNN algorithm as detailed in [67]. In the training phase, all networks undergo optimization using the Adam optimizer, with a weight decay parameter set to $1 \times 10^{-4}$. The training process starts with an initial learning rate of $1 \times 10^{-3}$, which is subsequently reduced by a factor of $5 \times 10^{-2}$ after every 5 epochs. This systematic adjustment in the learning rate ensures efficient convergence and optimal network performance.

We compared U3FQ with different state-of-the-art Image Quality Assessment methods: BRISQUE [39], BLINDSII [47], RankIQA [33], PFE [26], SDD-FIQA [44], SER-FIQA [61]. Our experiments employed four popular Face Recognition(FR) models: VGG-Face [62], FaceNet [49], ArcFace [16] and OpenFace [3] for computing scores. In our study, we used MobileFaceNet as the backbone for our method, emphasizing its efficiency and suitability in the real world.

### 4.7.2 Evaluation Metrics

In our study, the performance evaluation of the U3FQ was conducted by plotting the Error-Reject Curve (ERC). The ERC is a well-established method for representing Face Image Quality Assessment (FIQA) performance, as documented in the literature [24, 25]. It effectively demonstrates the impact of discarding a proportion of face images—specifically those of the lowest quality—on the face verification performance.

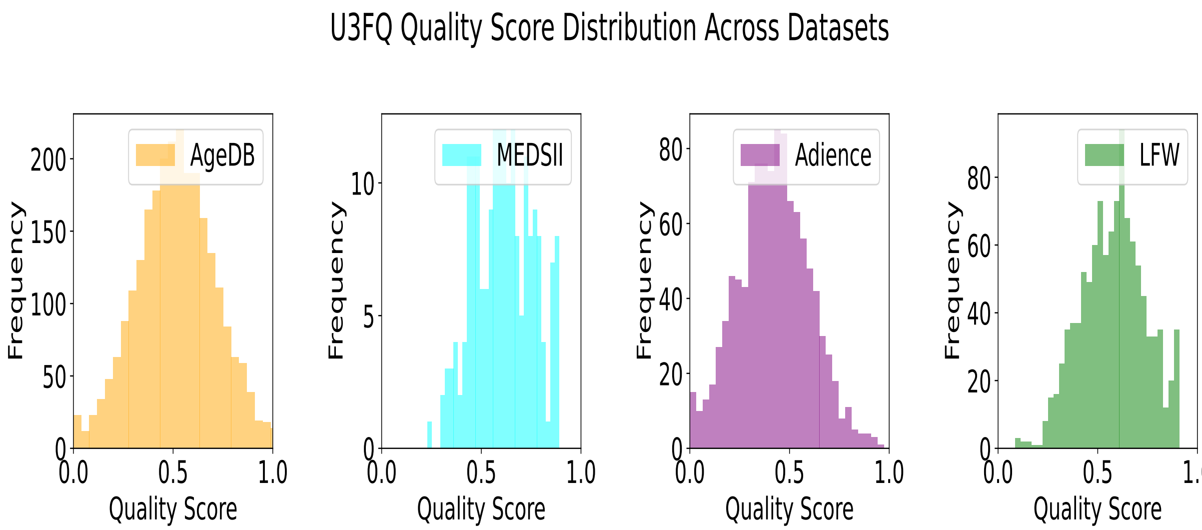



Figure 4.9: Comparative Analysis of Quality Score Distributions for AgeDB, MEDSII, Adience and LFW Datasets.

This impact is measured in terms of the False Non-Match Rate (FNMR) [35] at a predetermined threshold, set at a constant False Match Rate (FMR) [35]. For our analysis, the ERC curves for all benchmarks were plotted at two fixed FMRs: 1e-3, as recommended for border control operations by Frontex, and 1e-4, details of which are included at the end of the thesis. Additionally, we quantified the verification performance using the ERC's Area Over the Curve (AOC). This provides a comprehensive aggregate performance across all rejection ratios.

### 4.7.3 Performance on different recognition models

In the evaluation of U3FQ, as detailed in Table 4.1 and Figures 4.10, 4.11, 4.12, the metric was rigorously compared against both general Image Quality Assessment (IQA) techniques and specialized Face Image Quality Assessment (FIQA) methodologies. General IQA models like BRISQUE, BLINDS-II, and RankIQA, known for their broad application in IQA, were benchmarked alongside U3FQ. Additionally, specialized FIQA techniques such as PFE, SERFIQA, and SDD-FIQA, which are tailored for facial image quality, were also included in the comparison. This comprehensive evaluation using metrics like AUC (Area Under the Curve) or TAR (True Accept Rate) offers a nuanced understanding of U3FQ's performance relative to these established methods. The comparison highlights U3FQ's effectiveness in various contexts and provides valuable insights into its strengths and limitations in the field of FIQA.

### 4.7.4 Quality Score Distribution:

The quality distribution analysis of the U3FQ metric, as shown in Figure 4.9, thoroughly evaluates its performance and versatility across various datasets. This metric exhibits a remarkable ability to adapt, consistently assessing the quality of facial data with diverse characteristics. Such adaptability emphasizes

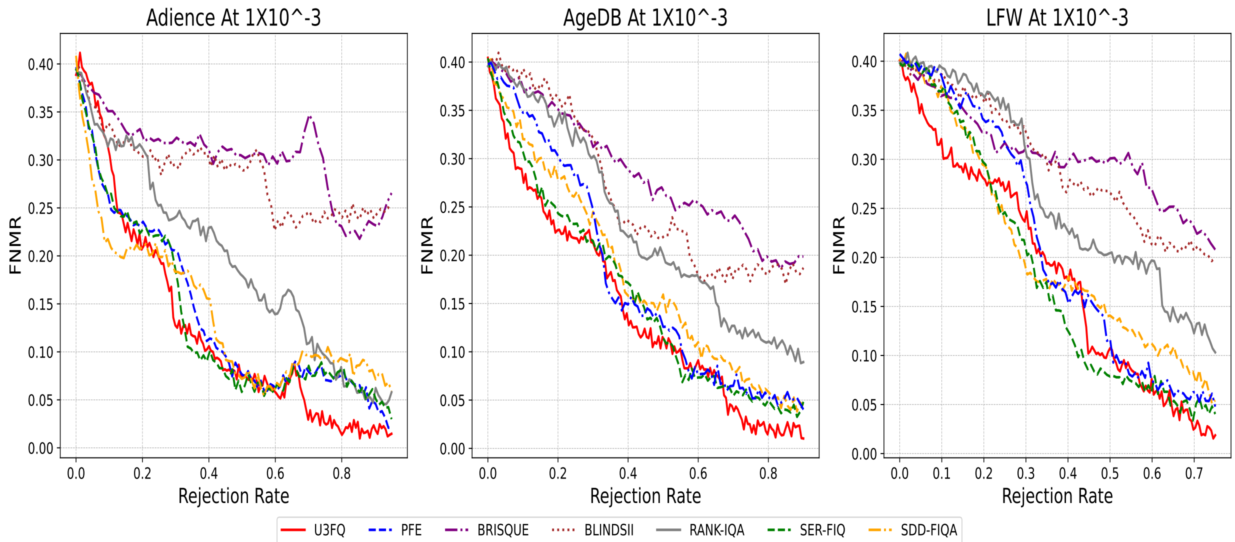



Figure 4.10: Effectiveness of Low-Quality Face Image Rejection in Face Verification: The EVRC (Expected Verification Rate Curve) Graphically Demonstrating FNMR (False Non-Match Rate) at a 1e-3 FMR (False Match Rate) Threshold Based on Predicted Quality Scores

U3FQ's robustness in facial quality assessment and showcases its ease of implementation in multiple contexts. The uniformity of its performance across different datasets reinforces its reliability and efficacy as a facial recognition tool. U3FQ's versatility is invaluable in facial analysis, ensuring high levels of accuracy and efficiency in a broad spectrum of applications.

## 4.8 Conclusion

Through the Unified Tri-Feature Quality Metric (U3FQ), we propose a pivotal advancement in the domain of Facial Image Quality Assessment (FIQA). By integrating age variance and facial expression impact, U3FQ presents a novel and comprehensive method for evaluating facial images. This research emphasizes the significance of these biometric features in enhancing the accuracy and reliability of recognition models, thereby transcending the conventional FIQA metrics that predominantly rely on subjective human visibility assessments. Through rigorous evaluations on an extensive set of face-quality image datasets and benchmark comparisons with state-of-the-art techniques, U3FQ has demonstrated its superiority in delivering relevant and precise quality assessments. Looking ahead, our future work aims to augment the predictive power of U3FQ with additional features such as illumination and pose to refine the accuracy of reference quality labels further, ensuring that U3FQ remains at the forefront of FIQA methodologies. We intend to broaden the scope and effectiveness of U3FQ, making it an even more robust tool for assessing facial image quality in diverse and challenging recognition scenarios under the new version of UXFQ.

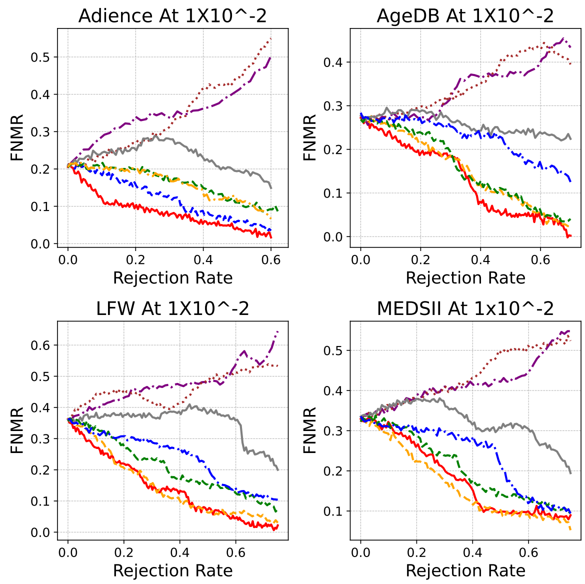



Figure 4.11: Comparative Analysis of U3FQ Performance at an FMR of 0.01 Across Diverse Datasets. This graph shows U3FQ's exceptional performance using four state-of-the-art Face Recognition models on the Adience, LFW, AGEDB, and MEDSII datasets. It demonstrates its competitive edge over contemporary baselines and robustness in varied facial recognition scenarios.

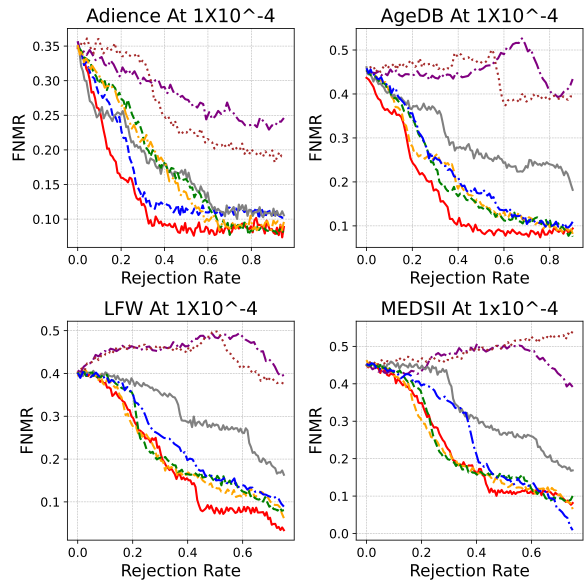



Figure 4.12: Comparative Analysis of U3FQ Performance at an FMR of 0.0001 Across Diverse Datasets. This graph shows U3FQ's exceptional performance using four state-of-the-art Face Recognition models on the Adience, LFW, AGEDB, and MEDSII datasets. It demonstrates its competitive edge over contemporary baselines and robustness in varied facial recognition scenarios.

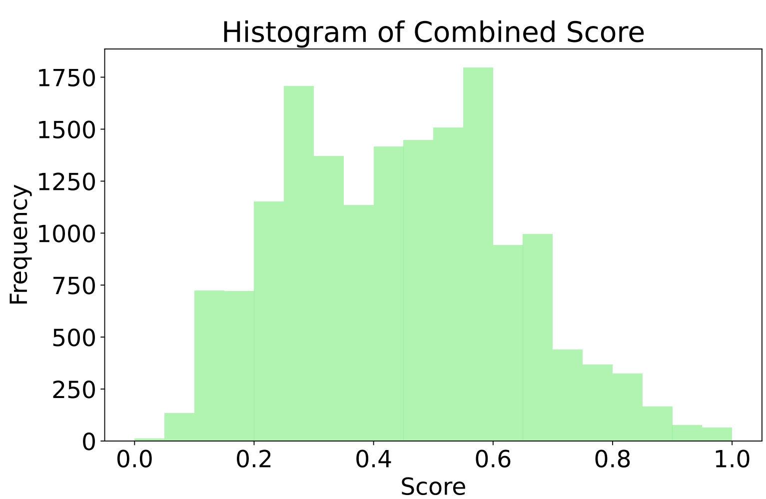



Figure 4.13: Distribution of Combined Match Scores Highlighting the Impact of Age Variation and Facial Expression on Image Quality Assessment.

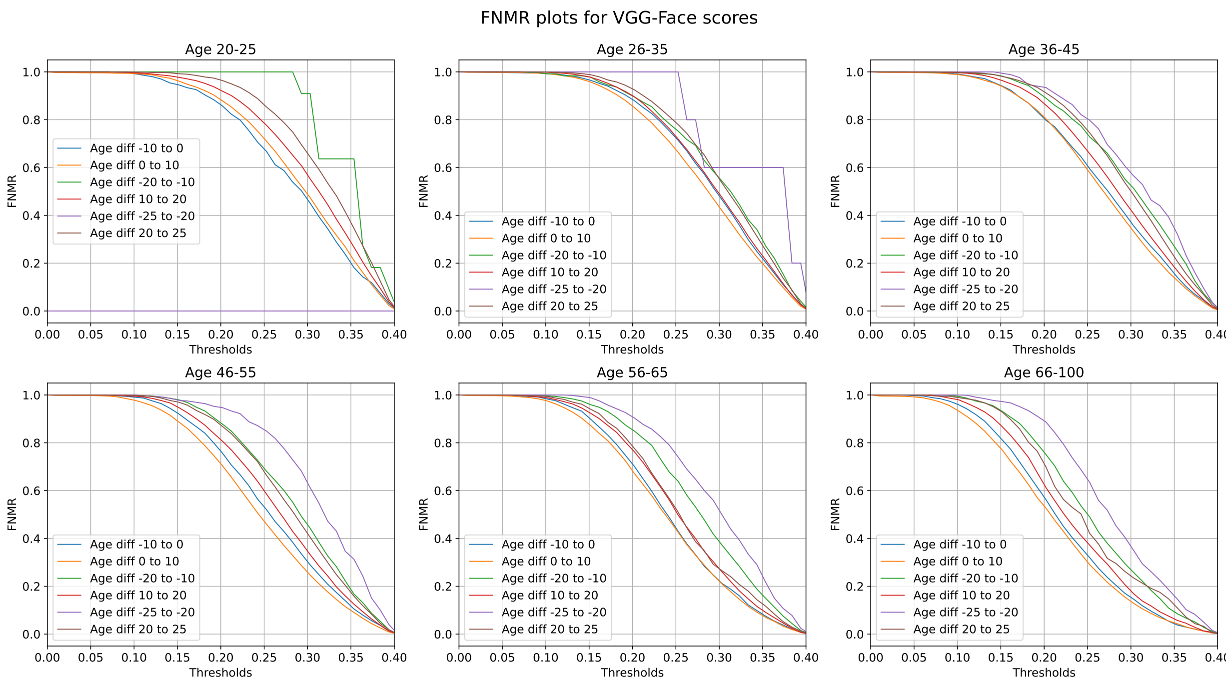



Figure 4.14: The VGG-Face DET plots, displaying the False Non-Match Rate (FNMR) across various age groups and age difference categories, reveal significant insights about the effects of anchor age on facial recognition accuracy. Notably, the age group of 35-45 years aligns closely with facial images across a wide age range, suggesting enhanced feature consistency within this demographic. For individuals over 60 years, a broad age difference (-30 to 30 years) exhibits minimal impact on FNMR, indicating a decreased variation in facial features with age.These observations were drawn from the AgeDB dataset[40] using the deepface face verification library.

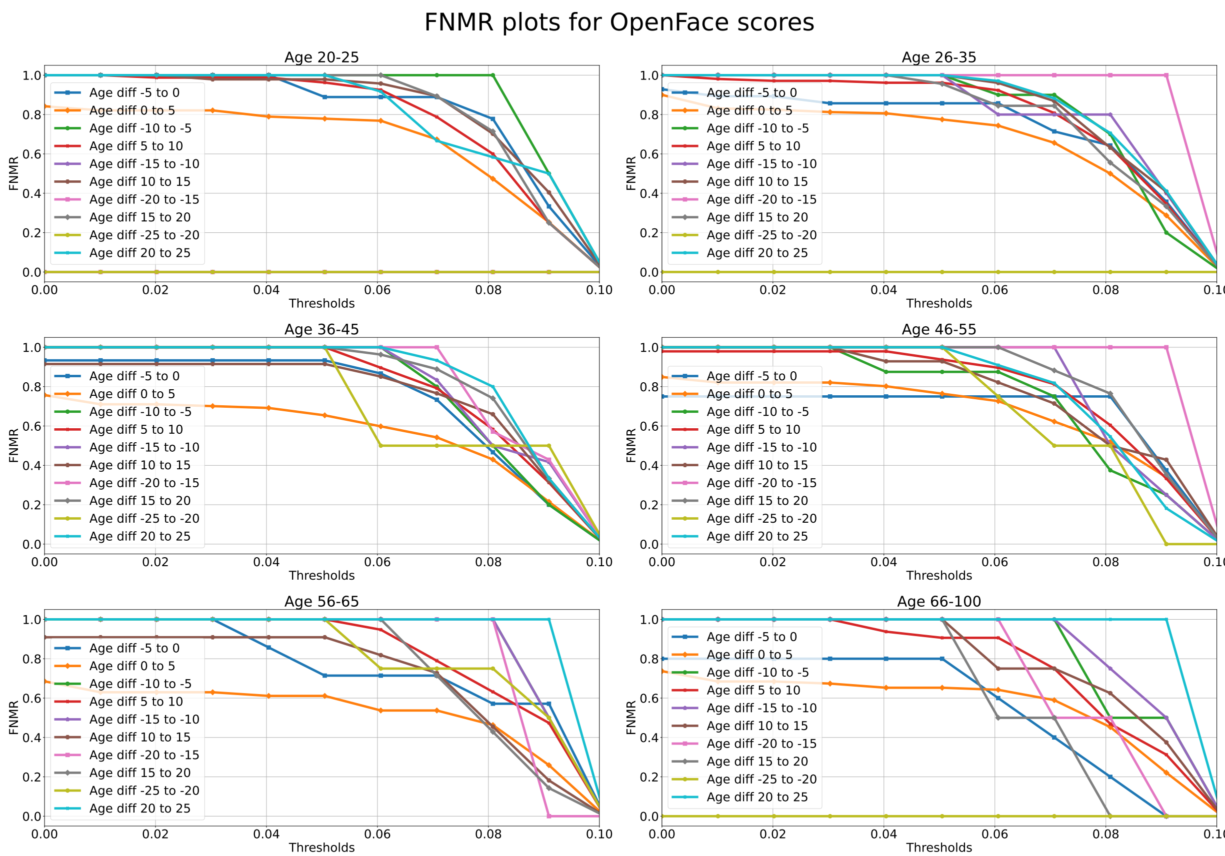



Figure 4.15: OpenFace, renowned for its facial behavior analysis toolkit, excels in facial landmark detection and eye gaze tracking. The Open-Face Detection Error Tradeoff (DET) plots have been meticulously crafted using this advanced technology. These plots display the False Non-Match Rate (FNMR) across a spectrum of age groups and various categories of age differences. The data for these insightful visualizations is sourced from the AgeDB dataset [40] and analyzed using the DeepFace face verification library, showcasing OpenFace's robust capabilities in handling diverse and complex facial recognition scenarios.

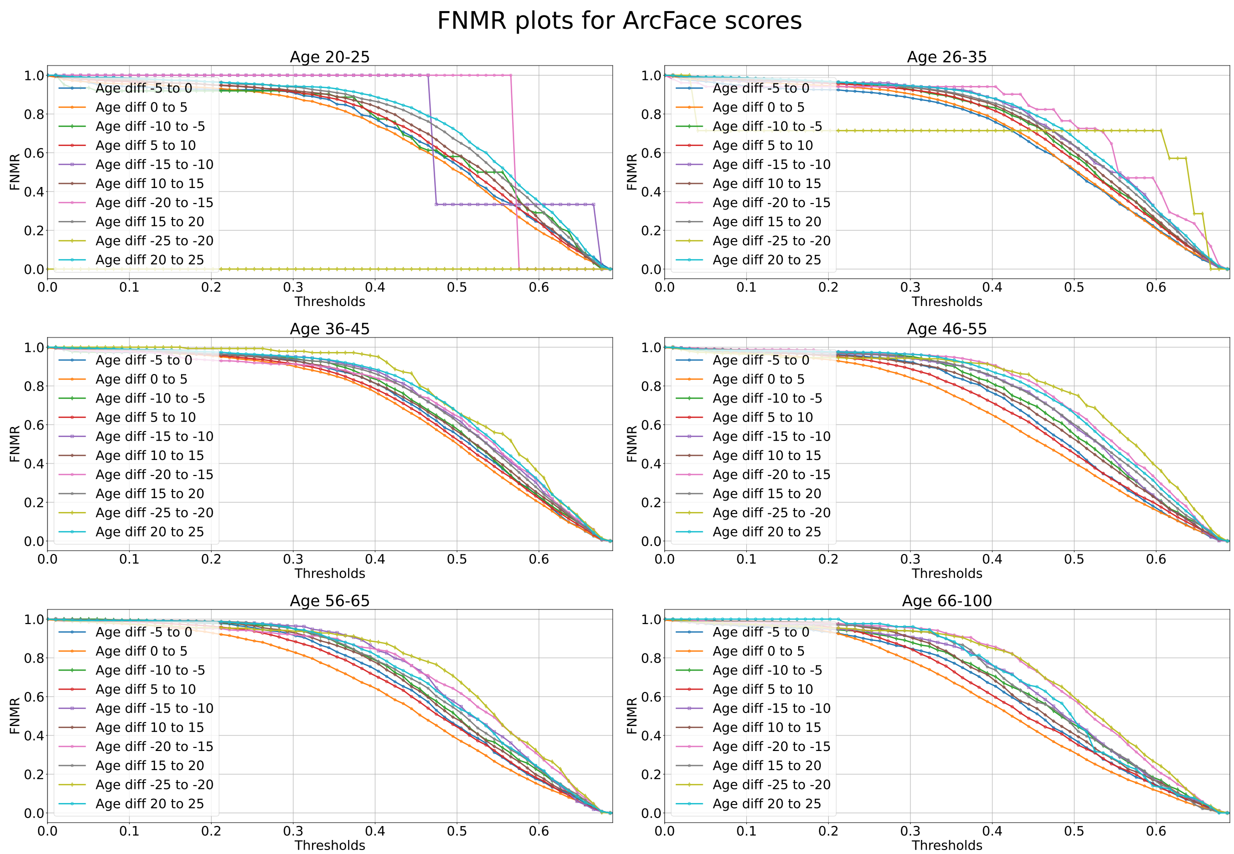



Figure 4.16: ArcFace stands out in the field of facial recognition with its state-of-the-art architecture that focuses on enhancing the discriminative features of faces. Known for its angular margin loss, which significantly boosts the accuracy and robustness of face recognition tasks, ArcFace is particularly adept at capturing subtle facial details. Utilizing this advanced technology, ArcFace has been employed to generate insightful Detection Error Tradeoff (DET) plots. These plots demonstrate the False Non-Match Rate (FNMR) across various age groups and categories of age differences, offering a detailed analysis of recognition performance under different demographic scenarios. The data for these plots is sourced from the AgeDB dataset [40], showing the efficacy of the ArcFace model in managing the complexities inherent in age-diverse facial recognition tasks.

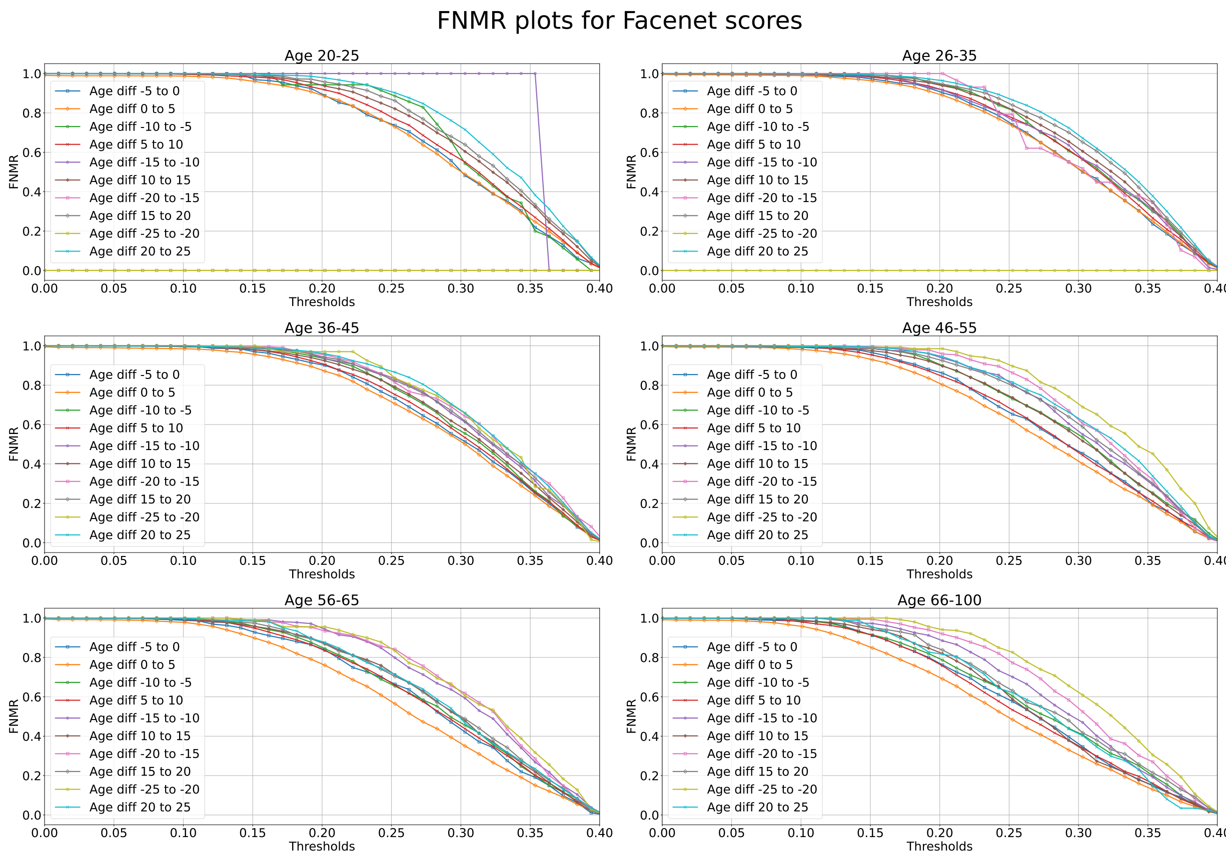
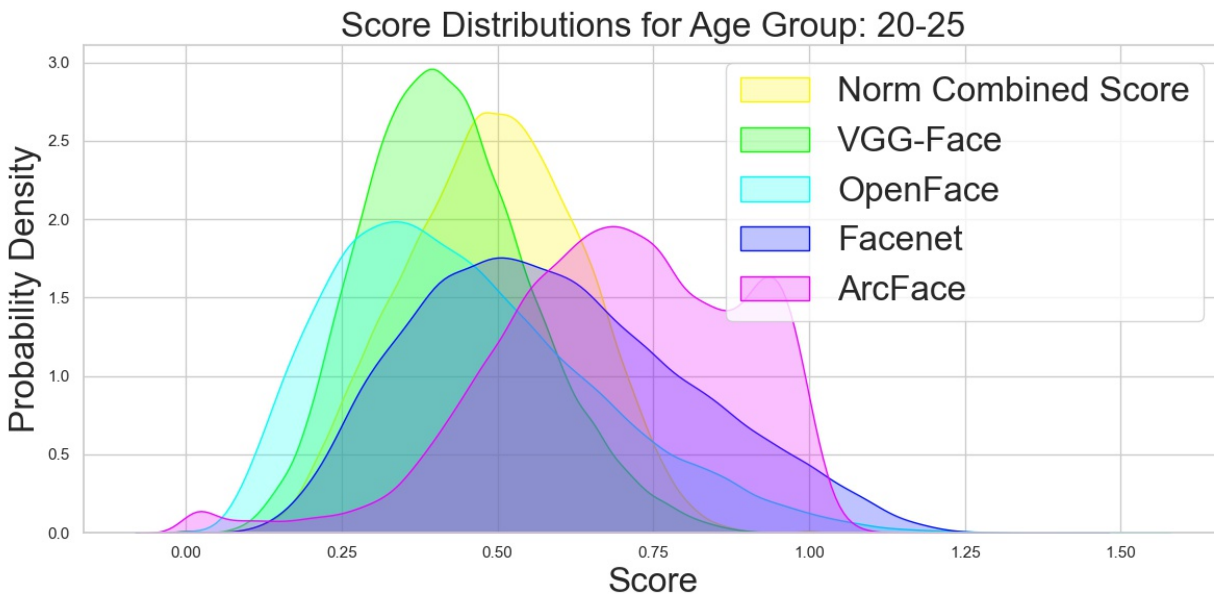
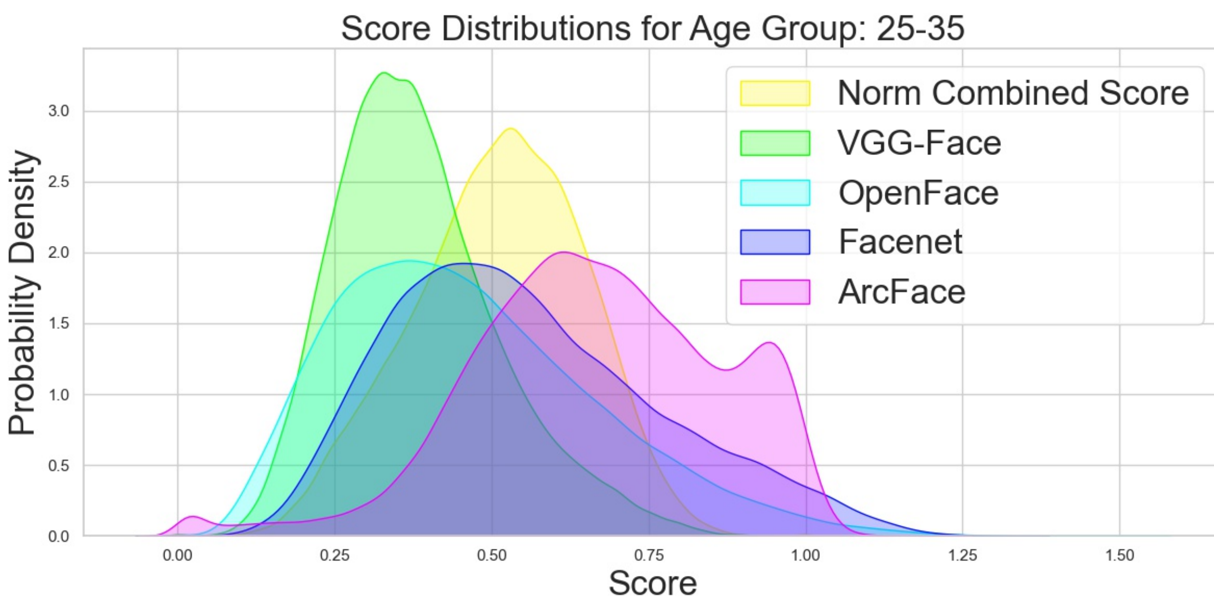



Figure 4.17: FaceNet, distinguished for its innovative approach to facial recognition, leverages deep neural networks to generate a unified embedding for face verification and recognition. Notably proficient in achieving high accuracy with lower computational costs, FaceNet excels in creating compact representations of facial features. The FaceNet-generated Detection Error Tradeoff (DET) plots employ this sophisticated framework and provide insightful analysis. These plots elucidate the False Non-Match Rate (FNMR) across various age groups and different categories of age differences. The underlying data for these comprehensive plots is derived from the AgeDB dataset [40], processed through the high-precision capabilities of the FaceNet model. This approach underscores FaceNet's exceptional performance in diverse facial recognition tasks, particularly in scenarios involving significant age variations.

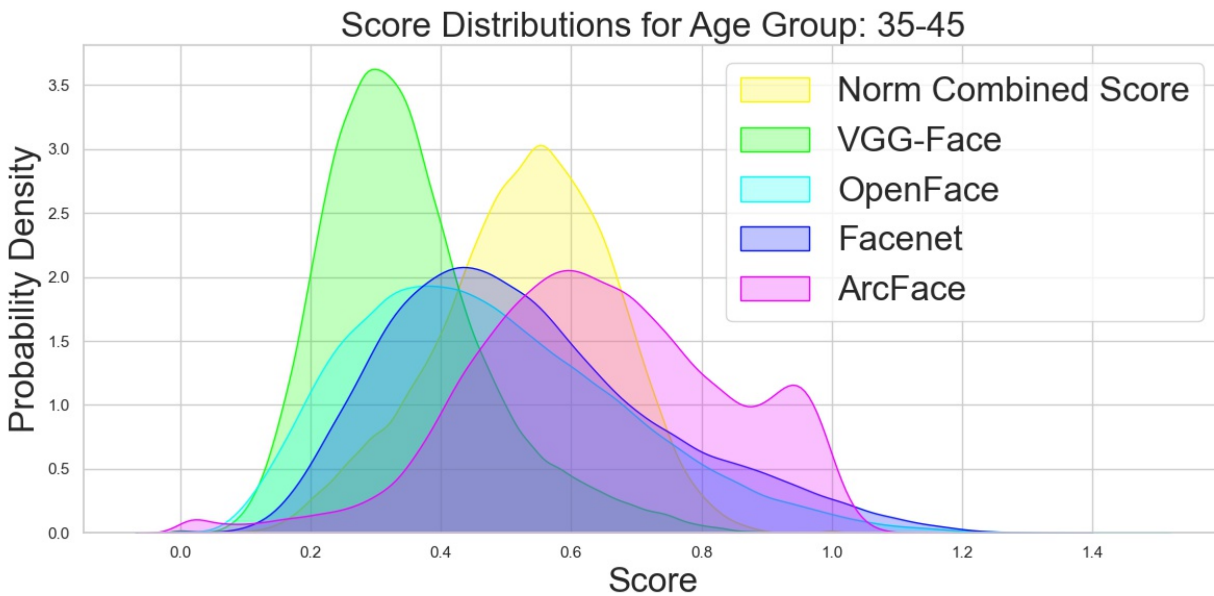



(a)

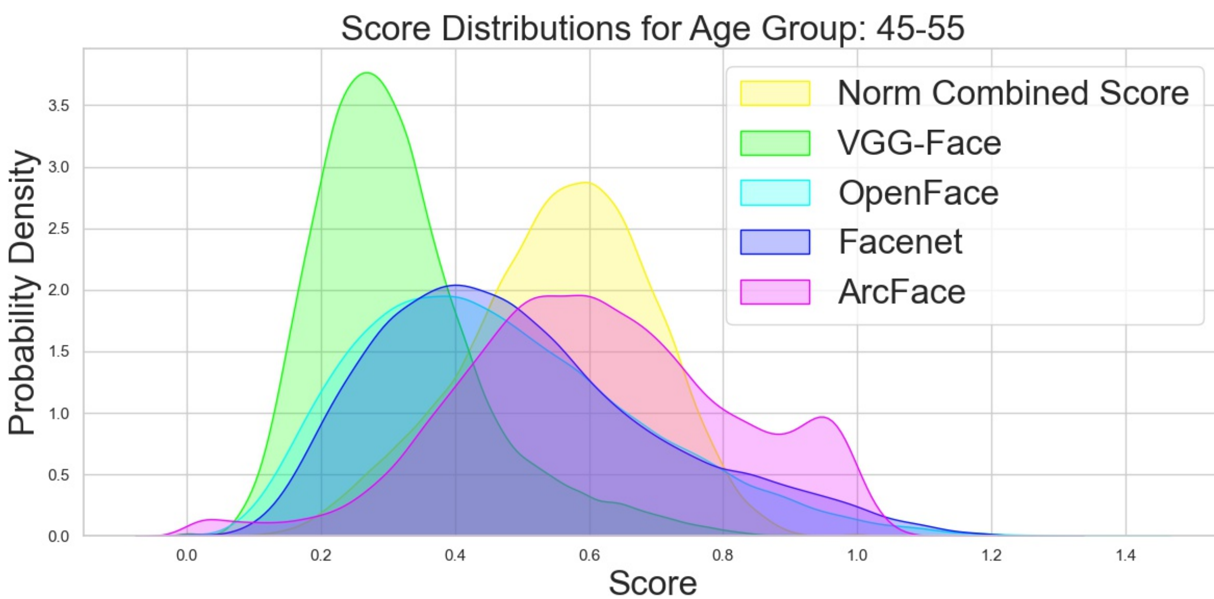



(b)

Figure 4.18: Distribution showing the matching scores of different face recognition models with the Normalized Combined Score for image pairs in the 20-35 age group.

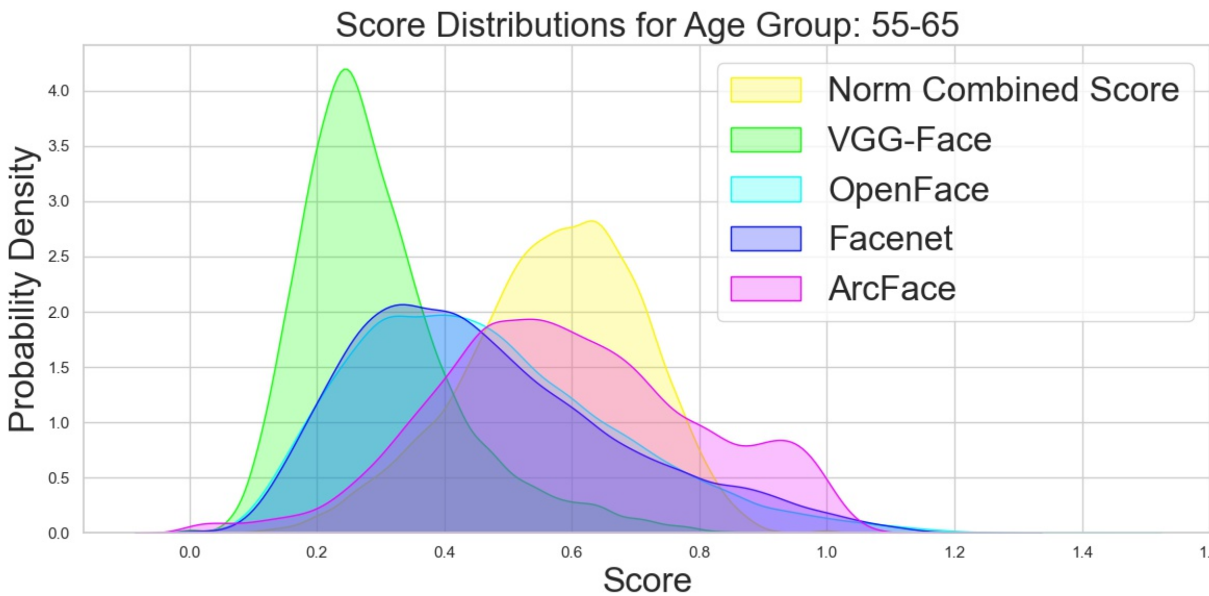



(a)

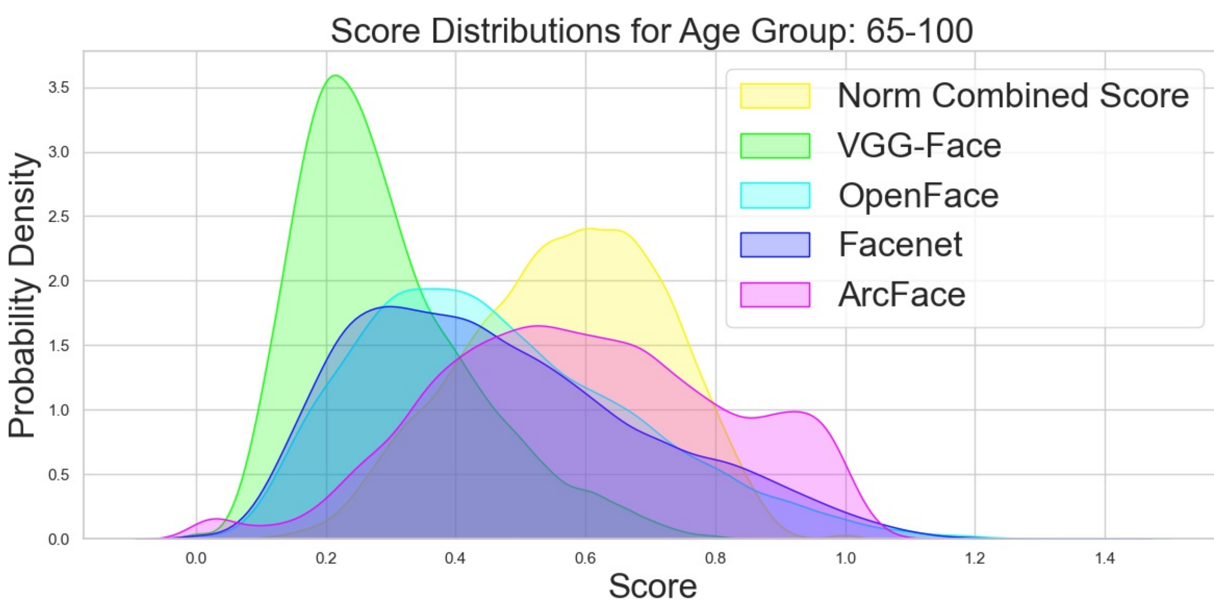



(b)

Figure 4.19: Distribution showing the matching scores of different face recognition models with the Normalized Combined Score for image pairs in the 35-55 age group.




(a)




(b)

Figure 4.20: Distribution showing the matching scores of different face recognition models with the Normalized Combined Score for image pairs in the 55 and above age group.

*Chapter 5*

# Future Work and Conclusion

The FRBQ and U3FQ models represent significant strides in the domain of biometric quality assessment, each targeting crucial but distinct aspects of biometric recognition—fingerprint and facial image quality. Both models leverage deep learning to transcend traditional assessment methods, showcasing a remarkable ability to generate reliable quality scores. The FRBQ model's proficiency in evaluating fingerprint images and the U3FQ model's innovative approach to facial image quality, especially considering age variance and emotional expression, underline a broader shift towards more nuanced and comprehensive quality assessments in biometrics.

## 5.1 Achievements and Challenges

The accomplishments of the FRBQ and U3FQ models are manifold, offering substantial improvements over existing methodologies. By reducing manual inspection efforts and enhancing the compatibility with current recognition frameworks, these models streamline the quality assessment process. Furthermore, the integration of label data and matching scores into the FRBQ model, alongside U3FQ's incorporation of age and expression, provide a richer, more detailed evaluation of biometric quality.

However, challenges remain, notably in accurately assessing images of poor quality. The FRBQ model's struggle with low-quality fingerprints and the ongoing quest to expand U3FQ's capabilities with additional biometric features such as illumination and pose underscore the complexity of biometric quality assessment. These hurdles necessitate a deeper understanding of the intricate relationships between biometric features and recognition accuracy.

## 5.2 Future Directions

In envisioning the future of biometric quality assessment, the development and refinement of the FRBQ and U3FQ models mark a significant leap forward. These models, by leveraging deep learning, have showcased the potential to surpass traditional assessment methods, offering more reliable, comprehensive evaluations of fingerprint and facial image qualities. As we chart the course for future explorations and advancements in this field, several key areas emerge as pivotal to pushing the boundaries of what these technologies can achieve. Among these, the expansion into and development of the Unified eXtensible Feature Quality (UXFQ) framework stands out as a particularly promising direction.

### 5.2.1 Advanced Algorithmic Development:

Future research will undoubtedly benefit from delving into more sophisticated algorithms that can adeptly navigate the complexities inherent in biometric data. This includes enhancing the FRBQ and U3FQ models to integrate additional biometric features—illuminating the path for a more nuanced understanding and assessment of quality across diverse biometric modalities. The goal is to develop algorithms that are not only more inclusive of various biometric characteristics but also capable of distinguishing between subtle variations in quality with greater precision.

### 5.2.2 Expansion of Comprehensive Training Datasets:

To further refine the accuracy and applicability of biometric quality assessment models, there is a pressing need to expand training datasets. These datasets should encompass a wider array of quality variations and biometric features, including those beyond fingerprints and facial images. By integrating a broader spectrum of data, from pristine to significantly degraded quality across multiple biometric modalities, we can significantly enhance the models' generalizability and performance.

### 5.2.3 Integration of Multimodal Biometric Features and UXFQ:

A particularly exciting avenue for future exploration is the integration of multimodal biometric features within a Unified eXtensible Feature Quality (UXFQ) framework. The UXFQ framework aims to provide a holistic, comprehensive platform for assessing the quality of various biometric modalities, including but not limited to fingerprints and facial images. By bringing these different modalities under a single, unified framework, the UXFQ promises to revolutionize biometric quality assessment, offering a more versatile, robust tool for enhancing recognition accuracy across a range of applications.

### 5.2.4 Real-world Application, Testing, and UXFQ Implementation:

The real-world application and testing of these models, particularly within the UXFQ framework, will be crucial. Deploying FRBQ, U3FQ, and UXFQ models in live biometric recognition systems will not only assess their efficacy in practical scenarios but also help bridge the gap between theoretical advancements and their operational utility. This step is essential for understanding how these models perform outside of controlled environments and for identifying areas for further improvement.

### 5.2.5 Ethical and Privacy Considerations:

As we advance in developing more sophisticated biometric quality assessment models, it is imperative to address the ethical and privacy concerns associated with biometric data analysis. Ensuring the ethical use of technologies, especially within the UXFQ framework, requires a commitment to respecting individual privacy and adhering to strict ethical standards. This is particularly important as models become capable of extracting and analyzing increasingly detailed information from biometric data.

## 5.3 Conclusion

The journey ahead for biometric quality assessment is both challenging and exciting. With the FRBQ and U3FQ models laying the groundwork, and the envisioned UXFQ framework set to expand this foundation, the future holds the promise of more accurate, reliable, and comprehensive biometric recognition systems. By focusing on these key areas—advanced algorithmic development, expansive training datasets, multimodal integration, real-world application, and ethical considerations—the field is poised for significant advancements. The ultimate goal is to not only enhance the precision of biometric recognition systems but to also ensure these technologies are developed and utilized in a manner that is ethical, respectful of privacy, and beneficial to society at large.

## Related Publications

- **Advancing Fingerprint Recognition Quality Assessment: Introducing the FRBQ Metric for Enhanced Fingerprint Recognition**, Prateek Jaiswal, Arka Koner, Anoop M Namboodiri. **In ICVGIP 2023:** In Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP ’23), December 15–17, 2023, Rupnagar, India.
- **Advancing FIQA with Age and Expressions: Introducing the U3FQ (Unified Tri-Feature Quality) Metric**, Prateek Jaiswal, Sai Amrit Patnaik, Anoop M Namboodiri. ***Under Review In ICVGIP 2024:*** In Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP ’24), December 13–15, 2024, IT Bangalore, India.
- **Facial De-morphing: Extracting Component Faces from a Single Morph**, Sudipta Banerjee, Prateek Jaiswal, Arun Ross **In IJCB, 2022:** In International Joint Conference on Biometrics, Oct 8-10, 2022, Abu Dhabi, UAE.

# Bibliography


[1] B. Amos, B. Ludwiczuk, M. Satyanarayanan, et al. Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science*, 6(2):20, 2016.

[2] Ž. Babnik, P. Peer, and V. Štruc. Faceqan: Face image quality assessment through adversarial noise exploration. In *2022 26th International Conference on Pattern Recognition (ICPR)*, pages 748–754. IEEE, 2022.

[3] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 59–66. IEEE, 2018.

[4] L. Best-Rowden and A. K. Jain. Automatic face image quality prediction. *arXiv preprint arXiv:1706.09887*, 2017.

[5] S. Bharadwaj, M. Vatsa, and R. Singh. Biometric quality: a review of fingerprint, iris, and face. *EURASIP journal on Image and Video Processing*, 2014:1–28, 2014.

[6] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on image processing*, 27(1):206–219, 2017.

[7] F. Boutros, N. Damer, J. N. Kolf, K. Raja, F. Kirchbuchner, R. Ramachandra, A. Kuijper, P. Fang, C. Zhang, F. Wang, et al. Mfr 2021: Masked face recognition competition. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. IEEE, 2021.

[8] F. Boutros, M. Fang, M. Klemt, B. Fu, and N. Damer. Cr-fiqa: face image quality assessment by learning sample relative classifiability. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5836–5845, 2023.

[9] K. Cao and A. K. Jain. Latent orientation field estimation via convolutional neural network. In *2015 International Conference on Biometrics (ICB)*, pages 349–356. IEEE, 2015.

[10] K. Cao and A. K. Jain. Automated latent fingerprint recognition. *IEEE transactions on pattern analysis and machine intelligence*, 41(4):788–800, 2018.

[11] K. Chen, T. Yi, and Q. Lv. Lightqnet: Lightweight deep face quality assessment for risk-controlled face recognition. *IEEE Signal Processing Letters*, 28:1878–1882, 2021.

[12] T. P. Chen, X. Jiang, and W.-Y. Yau. Fingerprint image quality analysis. In *2004 International Conference on Image Processing, 2004. ICIP'04.*, volume 2, pages 1253–1256. IEEE, 2004.

[13] Y. Chen, S. C. Dass, and A. K. Jain. Fingerprint quality indices for predicting authentication performance. In *International conference on audio-and video-based biometric person authentication*, pages 160–170. Springer, 2005.

[14] X. Dai, J. Liang, Q. Zhao, and F. Liu. Fingerprint segmentation via convolutional neural networks. In *Biometric Recognition: 12th Chinese Conference, CCBR 2017, Shenzhen, China, October 28-29, 2017, Proceedings 12*, pages 324–333. Springer, 2017.

[15] L. N. Darlow and B. Rosman. Fingerprint minutiae extraction using deep learning. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 22–30. IEEE, 2017.

[16] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. 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, 2019.

[17] E. Eidinger, R. Enbar, and T. Hassner. Age and gender estimation of unfiltered faces. *IEEE Transactions on information forensics and security*, 9(12):2170–2179, 2014.

[18] J. J. Engelsma, K. Cao, and A. K. Jain. Learning a fixed-length fingerprint representation. *IEEE transactions on pattern analysis and machine intelligence*, 43(6):1981–1997, 2019.

[19] J. Ezeobiejesi and B. Bhanu. Latent fingerprint image segmentation using deep neural network. *Deep Learning for Biometrics*, pages 83–107, 2017.

[20] A. Founds, N. Orlans, W. Genevieve, and C. Watson. Nist special databse 32 - multiple encounter dataset ii (meds-ii), 2011-07-21 00:07:00 2011.

[21] I. Fraunhofer. Optimization-based improvement of face image quality assessment techniques.

[22] H. Fronthaler, K. Kollreider, and J. Bigun. Automatic image quality assessment with application in biometrics. In *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, pages 30–30. IEEE, 2006.

[23] K. Grm and V. Struc. Deep face recognition for surveillance applications, 2018.

[24] P. Grother and E. Tabassi. Performance of biometric quality measures. *IEEE transactions on pattern analysis and machine intelligence*, 29(4):531–543, 2007.

[25] P. J. Grother, P. J. Grother, M. Ngan, and K. Hanaoka. *Face recognition vendor test (FRVT)*. US Department of Commerce, National Institute of Standards and Technology, 2014.

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

[27] L. Hong, Y. Wan, and A. Jain. Fingerprint image enhancement: algorithm and performance evaluation. *IEEE transactions on pattern analysis and machine intelligence*, 20(8):777–789, 1998.

[28] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.

[29] L. Kang, P. Ye, Y. Li, and D. Doermann. Convolutional neural networks for no-reference image quality assessment. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1733–1740, 2014.

[30] E. C. Larson and D. M. Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of electronic imaging*, 19(1):011006–011006, 2010.

[31] E. Lim, X. Jiang, and W. Yau. Fingerprint quality and validity analysis. In *Proceedings. International conference on image processing*, volume 1, pages I–I. IEEE, 2002.

[32] E. Lim, K.-A. Toh, P. Suganthan, X. Jiang, and W.-Y. Yau. Fingerprint image quality analysis. In *2004 International Conference on Image Processing, 2004. ICIP'04.*, volume 2, pages 1241–1244. IEEE, 2004.

[33] X. Liu, J. Van De Weijer, and A. D. Bagdanov. Rankiqa: Learning from rankings for no-reference image quality assessment. In *Proceedings of the IEEE international conference on computer vision*, pages 1040–1049, 2017.

[34] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain. Fvc2004: Third fingerprint verification competition. In *Biometric Authentication: First International Conference, ICBA 2004, Hong Kong, China, July 15-17, 2004. Proceedings*, pages 1–7. Springer, 2004.

[35] A. Mansfield. Information technology–biometric performance testing and reporting–part 1: Principles and framework. *ISO/IEC*, pages 19795–1, 2006.

[36] B. Meden, P. Rot, P. Terhörst, N. Damer, A. Kuijper, W. J. Scheirer, A. Ross, P. Peer, and V. Štruc. Privacy–enhancing face biometrics: A comprehensive survey. *IEEE Transactions on Information Forensics and Security*, 16:4147–4183, 2021.

[37] Q. Meng, S. Zhao, Z. Huang, and F. Zhou. Magface: A universal representation for face recognition and quality assessment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14225–14234, 2021.

[38] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12):4695–4708, 2012.

[39] A. Mittal, R. Soundararajan, and A. C. Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012.

[40] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, and S. Zafeiriou. 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, 2017.

[41] D.-L. Nguyen, K. Cao, and A. K. Jain. Robust minutiae extractor: Integrating deep networks and fingerprint domain knowledge. In *2018 International Conference on Biometrics (ICB)*, pages 9–16. IEEE, 2018.

[42] T. Oblak, R. Haraksim, P. Peer, and L. Beslay. Fingermark quality assessment framework with classic and deep learning ensemble models. *Knowledge-Based Systems*, 250:109148, 2022.

[43] M. Oquab, L. Bottou, I. Laptev, and J. Sivic. Is object localization for free?-weakly-supervised learning with convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 685–694, 2015.

[44] F.-Z. Ou, X. Chen, R. Zhang, Y. Huang, S. Li, J. Li, Y. Li, L. Cao, and Y.-G. Wang. 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, 2021.

[45] Z. Qu, J. Liu, Y. Liu, Q. Guan, R. Li, and Y. Zhang. A novel system for fingerprint orientation estimation. In *Image and Graphics Technologies and Applications: 13th Conference on Image and Graphics Technologies and Applications, IGTA 2018, Beijing, China, April 8–10, 2018, Revised Selected Papers 13*, pages 281–291. Springer, 2018.

[46] O. Remy, J. Hämmerle-Uhl, and A. Uhl. Fingervein sample image quality assessment using natural scene statistics. In *2022 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–6. IEEE, 2022.

[47] M. A. Saad, A. C. Bovik, and C. Charrier. Blind image quality assessment: A natural scene statistics approach in the dct domain. *IEEE transactions on Image Processing*, 21(8):3339–3352, 2012.

[48] T. Schlett, C. Rathgeb, O. Henniger, J. Galbally, J. Fierrez, and C. Busch. Face image quality assessment: A literature survey. *ACM Computing Surveys (CSUR)*, 54(10s):1–49, 2022.

[49] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[50] P. Schuch, S.-D. Schulz, and C. Busch. Deep expectation for estimation of fingerprint orientation fields. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 185–190. IEEE, 2017.

[51] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11):3440–3451, 2006.

[52] L. Shen, A. Kot, and W. Koo. Quality measures of fingerprint images. In *Audio-and Video-Based Biometric Person Authentication: Third International Conference, AVBPA 2001 Halmstad, Sweden, June 6–8, 2001 Proceedings 3*, pages 266–271. Springer, 2001.

[53] N. Singla, M. Kaur, and S. Sofat. Automated latent fingerprint identification system: A review. *Forensic science international*, 309:110187, 2020.

[54] H. Swofford, C. Champod, A. Koertner, H. Eldridge, and M. Salyards. A method for measuring the quality of friction skin impression evidence: Method development and validation. *Forensic Science International*, 320:110703, 2021.

[55] E. Tabassi. Towards nfiq ii lite: Self-organizing maps for fingerprint image quality assessment. 2014.

[56] E. Tabassi and C. L. Wilson. A novel approach to fingerprint image quality. In *IEEE International Conference on Image Processing 2005*, volume 2, pages II–37. IEEE, 2005.

[57] H. Talebi and P. Milanfar. Nima: Neural image assessment. *IEEE transactions on image processing*, 27(8):3998–4011, 2018.

[58] Y. Tang, F. Gao, J. Feng, and Y. Liu. Fingernet: An unified deep network for fingerprint minutiae extraction. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 108–116. IEEE, 2017.

[59] P. Terhörst, A. Boller, N. Damer, F. Kirchbuchner, and A. Kuijper. Midecon: Unsupervised and accurate fingerprint and minutia quality assessment based on minutia detection confidence. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8. IEEE, 2021.

[60] P. Terhörst, M. Ihlefeld, M. Huber, N. Damer, F. Kirchbuchner, K. Raja, and A. Kuijper. Qmagface: Simple and accurate quality-aware face recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3484–3494, 2023.

[61] P. Terhorst, J. N. Kolf, N. Damer, F. Kirchbuchner, and A. Kuijper. Ser-fiq: Unsupervised estimation of face image quality based on stochastic embedding robustness. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5651–5660, 2020.

[62] VGG. The visual geometry group (vgg) at the university of oxford.

[63] M. Wang and W. Deng. Deep face recognition: A survey. *Neurocomputing*, 429:215–244, 2021.

[64] Z. Wang and A. C. Bovik. *Modern image quality assessment*. PhD thesis, Springer, 2006.

[65] S. Wu, B. Liu, Z. Wang, Z. Jia, and J. Feng. Minutiae-awarely learning fingerprint representation for fingerprint indexing. In *2022 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8. IEEE, 2022.

[66] W. Xie, J. Byrne, and A. Zisserman. Inducing predictive uncertainty estimation for face recognition. *arXiv preprint arXiv:2009.00603*, 2020.

[67] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters*, 23(10):1499–1503, 2016.

[68] Z.-H. Zhou. A brief introduction to weakly supervised learning. *National science review*, 5(1):44–53, 2018.

[69] Y. Zhu, X. Yin, X. Jia, and J. Hu. Latent fingerprint segmentation based on convolutional neural networks. In *2017 IEEE Workshop on Information Forensics and Security (WIFS)*, pages 1–6. IEEE, 2017.