# Advancing Contextual Face Image Quality Assessment with the U3FQ: A Unified Tri-Feature Metric

Anonymous CVPR submission

Paper ID 12988

## Abstract

*Facial Image Quality Assessment (FIQA) plays a pivotal role in the enhancement of face matching and recognition systems. Traditional FIQA metrics tend to focus on subjective human visibility, which may not align with the features essential for accurate recognition. To bridge this gap, we propose the Unified Tri-Feature Quality Metric (U3FQ), a novel assessment framework that integrates three critical elements: age variance, facial expression impact, and congruence scores from four state-of-the-art recognition models. Unlike conventional metrics, U3FQ is underpinned by a semi-reference paradigm powered by a multitasking ResNet model. This innovative approach divides its focus between the direct assessment of facial image quality and the prediction of facial age with a mean absolute error (MAE) of ±2 years for both genders. Our method refines congruence scores with quantitative modifiers that account for age discrepancies and expression intensities, ensuring a quality metric that more accurately predicts the likelihood of recognition success. We rigorously evaluated U3FQ against general IQA techniques—BRISQUE, NIQE, and PIQE—as well as specialized FIQA methodologies such as FaceQnet, SER-FIQ, and MagFace. The results confirm that U3FQ marks a substantial advancement in the field of FIQA, offering a holistic and theoretically sound quality assessment tool that is acutely relevant to a wide array of facial recognition scenarios.*

## 1. Introduction

## 2. Related Works

# 3. Methodology

Our work introduces the Unified Tri-Feature Quality Metric (U3FQ) for Contextual Facial Image Quality Assessment (FIQA), a key advancement for the precision of biometric identification systems. This section outlines our integrated methodology for developing and refining the U3FQ metric, which accounts for match scores, age disparities, and facial expressions. We also discuss the machine learning and deep learning frameworks applied in our analysis.

- In Subsection 3.1, we provide a succinct overview of the theoretical foundations of U3FQ and its significance within the realm of FIQA. - Subsection 3.2 details the operational aspects of our model, including how age-related match score adjustments and expression-based calibrations are integrated using the AgeDB [1] dataset. - The design of our computational framework, which utilizes both Random Forest and ResNet models for quality score prediction, is presented in Subsection **??**. - Finally, Subsection 3.3 offers an examination of various computational techniques for match score determination and addresses the handling of non-mated pairs, informed by data from the NIST MEDII dataset to enhance recognition accuracy.

## 3.1. Theoretical Background

**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 illustrated in Figure 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 present a considerable challenge due to the substantial facial feature changes 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.

To empirically underpin this approach, we present Detection Error Tradeoff (DET) plots that demonstrate the variance in performance with different age groups for all four models: VGG-FACE, OpenFace, ArcFace, and FaceNet. Due to page limitations, these plots are included in the supplementary material. Here, we have added the DET plots from VGG-Face in Figure 3, which show the False Non-Match Rate (FNMR) for different age groups. These plots highlight the impact of age difference on the efficacy of face-matching systems, for different anchor age where 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.

**The Influence of Facial Expressions:** The similarity in facial expressions between two images significantly influences recognition performance, as variations in expressions can distort critical facial features used in establishing a match. Consequently, this also affects the overall quality

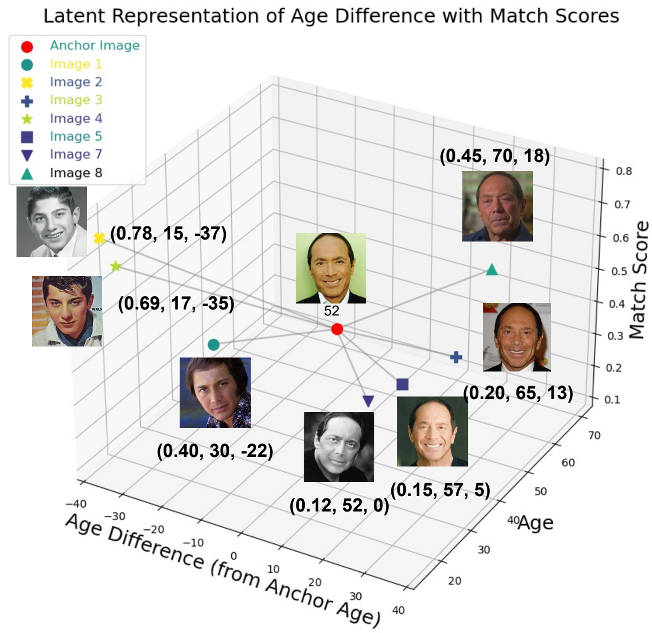


Figure 1. The effectiveness of face matching systems is markedly influenced by the age disparity between compared images. The triplet representation highlights the similarity distance, the age of the compared image, and the age difference relative to the anchor image, with Image 6 serving as the reference point.

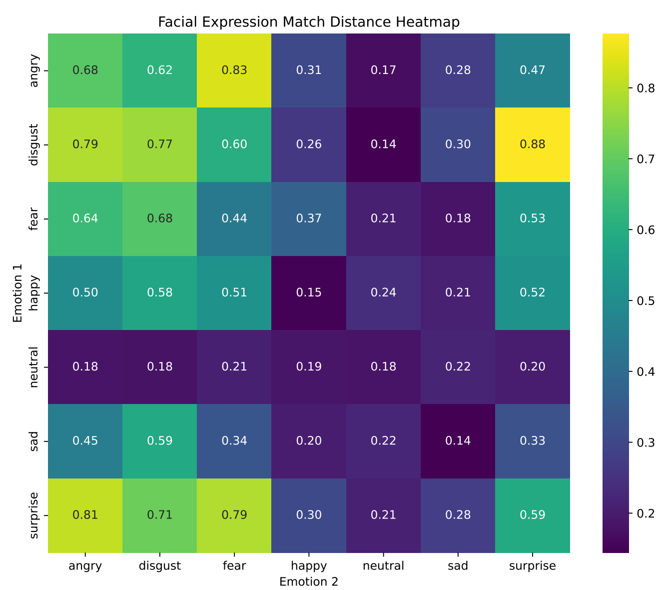


Figure 2. The differential impact of facial expressions on the match score, with weak emotions having a constant effect and strong emotions modifying the score proportionally to their intensity.

CVPR #12988

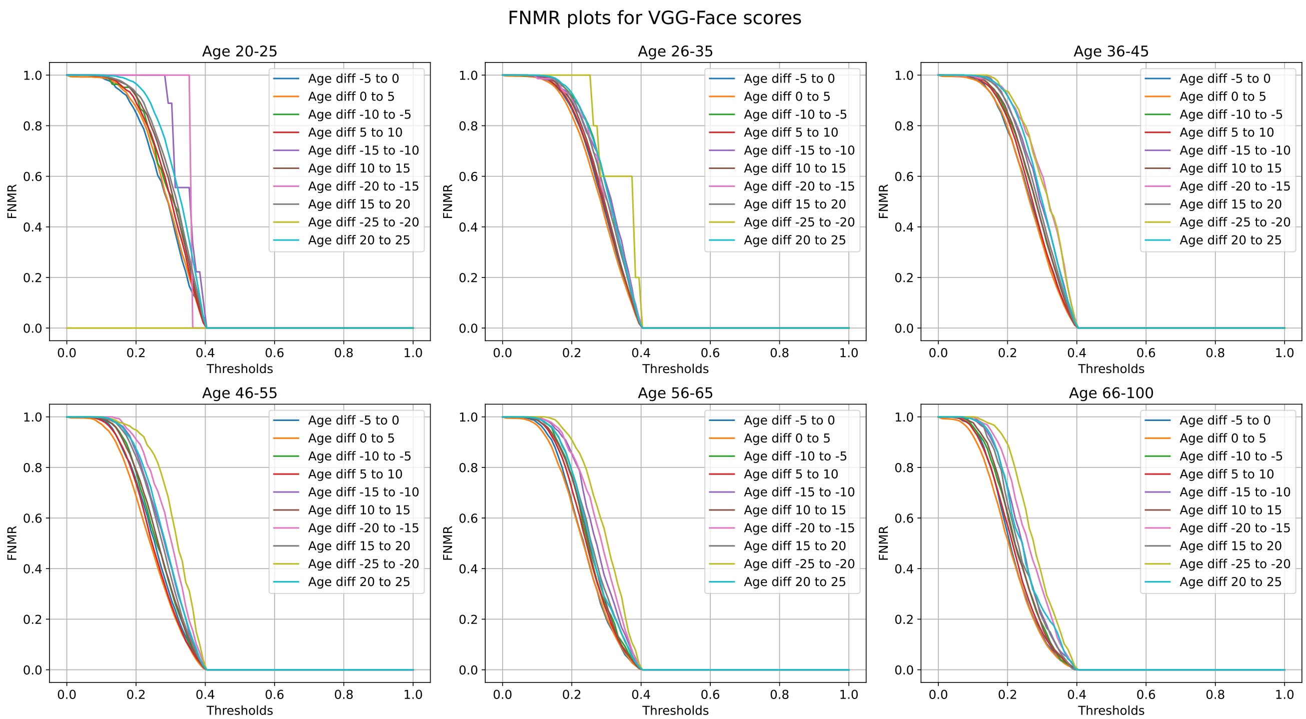


Figure 3. VGG-Face DET plots demonstrating FNMR across different age groups and age difference categories.

of recognition. Figure 2 illustrates the impact that discrepancies in facial expressions have on matching performance, as evidenced by the average match scores across diverse expression pairs.

### 3.2. Formulations and Optimization

Building on the observations from empirical evidence, we proceed to formulate the mathematical model that incorporates age difference penalties into the facial match score. The age difference penalty function is adapted as follows:

$$f(d, a) = \begin{cases} e^{\alpha(d+\beta)} & \text{if } d < 0 \text{ and } a \leq 30, \\ e^{\alpha(d+\beta)/\theta} & \text{if } d < 0 \text{ and } a > 30, \\ \gamma \cdot d & \text{if } d \geq 0, \end{cases} \tag{1}$$

where $d$ represents the age difference between the anchor and the comparison image, $a$ denotes the anchor's age, and $\alpha$, $\beta$, and $\gamma$ are parameters dictating the function's shape. The factor $\theta$ serves as a damping parameter that reduces the penalty for older anchors.

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:

$$g(e) = \begin{cases} c & \text{if } e \text{ is a weak emotion,} \\ d \cdot \text{EXPR\_SCORE}(e) & \text{if } e \text{ is a strong emotion,} \end{cases} \tag{2}$$

where $c$ is a constant factor for weak emotions, and $d$ scales the expression score EXPR_SCORE($e$) for strong emotions.

These formulations, alongside the 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 not only secure but also user-friendly, accommodating the complexities of human features and behaviors.

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.

The U3FQ algorithm initiates with the computation of the match score distance from an ensemble of models $M = \{M_1, M_2, M_3, M_4\}$. The distance metric, denoted as $d$, is derived from the pairwise discrepancies in the features extracted by each model for the given image $I$. The age difference function $f(d, a)$ is applied to adjust $d$ based on the age of the anchor image $a$.

CVPR #12988

**Algorithm 1** U3FQ: Unified Tri-Feature Quality Assessment for Contextual Facial Image Quality

**Input:** Single input image $I$, ResNet model $RN$, age $a$, expression $e$, match score distance models $M = \{M_1, M_2, M_3, M_4\}$

**Output:** U3FQ score $U$, Feature vector $F$, Estimated age, and quality score of input image

$U \leftarrow 0$ ▷ Initialize U3FQ score
$A \leftarrow []$ ▷ Initialize list for age multipliers
$E \leftarrow []$ ▷ Initialize list for expression multipliers
**for all** $model \in M$ **do**
  $d \leftarrow$ ComputeMatchScoreDistance$(I, model)$
  $ageMultiplier \leftarrow f(d, a)$ ▷ Using the age difference function
  $A \leftarrow A + [ageMultiplier]$
  **if** ExpressionsAreSame$(I, model)$ **then**
    $expressionMultiplier \leftarrow c$ ▷ Constant for same expression
  **else**
    $expressionMultiplier \leftarrow d \cdot$ EXPR_SCORE$(e)$ ▷ Scaled for different expressions
  **end if**
  $E \leftarrow E + [expressionMultiplier]$
  $U \leftarrow U + d \cdot ageMultiplier \cdot expressionMultiplier$
**end for**
$F \leftarrow [a, e, U]$ ▷ Feature vector containing age, expression, and U3FQ score
**procedure** U3FQ_ASSESSMENT$(I, F, RN, m = 100)$
  $QualityScores \leftarrow []$ ▷ Initialize list for quality predictions
  $AgeEstimates \leftarrow []$ ▷ Initialize list for age predictions
  **for** $i \leftarrow 1$ to $m$ **do**
    $quality, ageEstimate \leftarrow RN$.Predict$(I, F)$ ▷ Predict quality and age
    $QualityScores \leftarrow QualityScores + [quality]$
    $AgeEstimates \leftarrow AgeEstimates + [ageEstimate]$
  **end for**
  $finalQuality \leftarrow$ Average$(QualityScores)$
  $finalAge \leftarrow$ Average$(AgeEstimates)$
**return** $finalQuality, finalAge$
**end procedure**

Simultaneously, the expression impact function $g(e)$ adjusts the congruence score depending on the facial expression $e$, where $c$ is a constant factor for weak emotions and $d$ is a scaling factor for strong emotions, coupled with the expression score EXPR_SCORE$(e)$.

These functions are crucial as they capture the dynamic nature of facial recognition where age and emotional expression significantly impact the quality of facial features captured in an image.

The algorithm then constructs a feature vector $F$, encompassing the age $a$, expression $e$, and the aggregate congruence score $U$, calculated as a weighted sum of the match score distances adjusted by the age and expression multipliers. This vector is integral in capturing the nuanced elements that contribute to the facial image quality.

$$U \leftarrow \sum_{model \in M} d \cdot f(d, a) \cdot g(e)$$
$$F \leftarrow [a, e, U]$$

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.

### 3.3. Analytical Study

This subsection embarks on a discourse surrounding alternative methodologies that hold potential for future exploration. Considerations include the grouping of facial images by expression, the differential impacts of varying age differences, and other conceivable permutations that could influence the FIQA landscape.

Figure 4. Visualization of alternative approaches and potential experiments for advancing FIQA.

# 4. Experiments

# 5. Results

# 6. Conclusion

# References

[1] Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos 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. 2