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

Anonymous CVPR submission

Paper ID *****

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 methodology is anchored in the creation of a novel Unified Tri-Feature Quality Metric (U3FQ) for Contextual Facial Image Quality Assessment (FIQA), which is pivotal for enhancing the accuracy of biometric systems. This section elucidates the integrated approach we have employed to devise, enhance, and configure the U3FQ metric, encapsulating the synergetic impact of match scores, age differences, and facial expressions on image quality. It further expounds on the architecture of both conventional machine learning models and advanced deep learning networks used in our study. In Subsection 3.1, we detail the theoretical underpinnings of U3FQ, highlighting its relevance in FIQA and providing a historical context for the selected features. Subsection 3.2 delves into our method's practical application, describing the interplay between age-modulated match scores and expression-weighted adjustments within the model's fine-tuning process, using the AGEDB dataset. The architectural framework, comprising Random Forest and ResNet models, is discussed in Subsection 3.3, which delineates the feature integration for quality score prediction. Finally, Subsection 3.4 presents an analytical study on various computational strategies for match score calculation and the treatment of non-mated pairs, with insights drawn from the NIST MEDII dataset for robust recognition performance

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 gaps 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 2, 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, particularly for anchors aged 20-25 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.

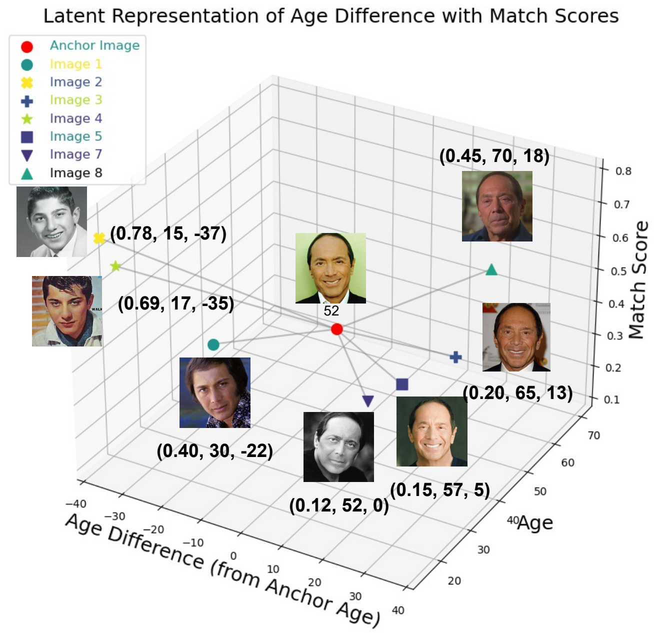

Figure 1. The impact of age difference on face matching system efficacy.

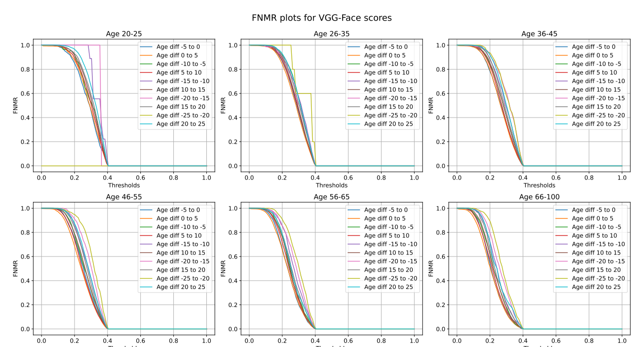

Figure 2. VGG-Face DET plots demonstrating FNMR across different age groups and age difference categories.

These empirical findings underscore the critical role that age plays in the performance of face-matching systems and the importance of integrating such a nuanced understanding into the FIQA model to enhance the system's robustness and reliability.

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 α, β, and γ are parameters dictating the function's shape. The factor θ 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 $\text{EXPR\_SCORE}(e)$ for strong emotions.

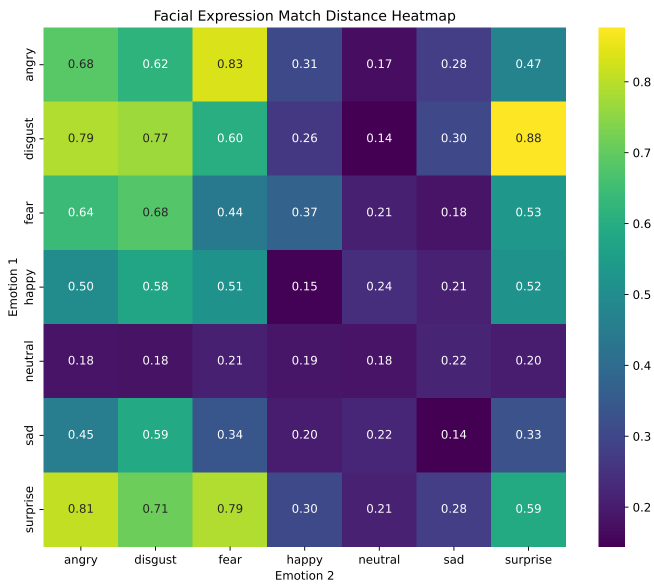

Figure 3. The differential impact of facial expressions on the match score, with weak emotions having a constant effect and strong emotions modifying the score proportionally to their intensity.

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.

Figure 4. The impact of age difference on face match score integration.

Figure 5. Influence of facial expression variations on match score adjustments.

3.3. Architecture

Initially, a Random Forest model was utilized to assess the impact of incorporating match scores, age difference, and facial expression data on image quality predictions. Subsequently, we transitioned to a deep learning framework, employing ResNet50 and ResNet18 architectures to further refine the predictive capability of our system.

Figure 6. The architecture of the proposed U3FQ metric incorporating deep learning models.

The following diagram delineates the workflow pipeline, highlighting the integration of the deep learning network with the Random Forest model to compute the face quality score.

Figure 7. Workflow pipeline utilizing ResNet models and Random Forest for quality score prediction.

3.4. 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 8. Visualization of alternative approaches and potential experiments for advancing FIQA.

4. Experiments

5. Results

6. Conclusion

References