Advancing FIQA with Intrinsic Age Features: Introducing the U3FQ (Unified Tri-Feature Quality) Metric

Anonymous CVPR submission

Paper ID 12988

Abstract

Facial Image Quality Assessment (FIQA) is crucial in enhancing face matching and recognition systems. Traditional FIQA metrics often focus on subjective human visibility, which may not correspond with the features crucial for accurate recognition. To address this issue, 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. U3FQ utilizes an advanced learning paradigm, employing a ResNet model designed for facial image quality assessment. This approach diverges from conventional metrics by focusing on aspects directly influencing recognition accuracy, such as expression intensities and facial features' congruence with recognition models. Our method enhances congruence scores with quantitative modifiers that account for expression intensities, ensuring a more accurate quality metric for predicting recognition success likelihood. U3FQ was rigorously evaluated against general IQA techniques—BRISQUE, NIQE, and PIQE—and specialized FIQA methodologies like FaceQnet, SER-FIQ, and MagFace. The results demonstrate that U3FQ represents a significant advancement in FIQA, offering a holistic and theoretically robust assessment tool that is highly relevant for various facial recognition scenarios.

1. Introduction

Facial Image Quality Assessment (FIQA) is integral to the enhancement of face recognition (FR) systems, especially when dealing with the diversity of image quality often encountered in real-world scenarios. Traditional FIQA approaches primarily assess the standalone biometric utility of images. However, in the context of FR, this method faces a conceptual challenge known as the *"Quality Paradox,"* as discussed by Schlett et al. [22]. This paradox highlights the need to accurately reflect the reliability of comparison scores for image pairs that include the assessed image, thus adding a layer of complexity to FIQA's role in face recognition performance. In recent advancements, FR techniques have shown remarkable results with high-quality frontal images and those of varying quality [13, 17]. However, they still face significant hurdles in completely unconstrained environments [5, 28] where the quality of captured facial images cannot be guaranteed. FIQA methods strive to enhance the performance of FR systems in such settings by offering critical insights into the quality of input images. This input allows FR models to discern and possibly discard images of inferior quality that could lead to erroneous non-matches.

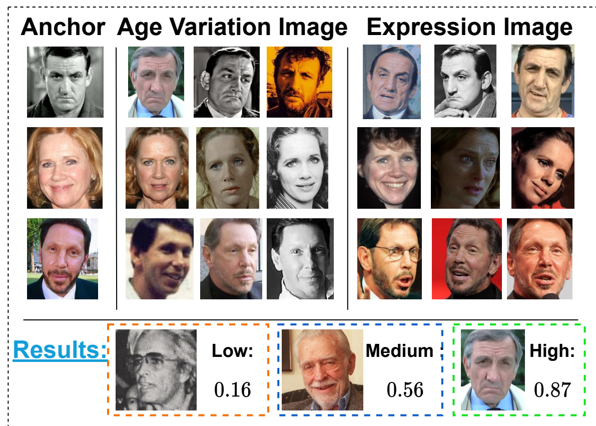

Figure 1. Displaying a range of facial images, each grouped by the set includes an anchor image and its age and expression variants. Results images categorized into low, medium, and high-quality segments, highlighting the model's proficiency in assessing quality variations.

Modern FIQA methods are generally categorized into two distinct styles: regression-based and model-based approaches. Regression-based methods [8, 14, 29] develop a direct mapping from the image space to quality labels generated in a semi-automated manner. These labels often draw on comparison scores across matched image pairs or similarity scores between probe samples and reference images.

On the other hand, model-based [18, 26] FIQA techniques integrate quality assessment directly within the FR model, evaluating the quality based on the certainty or statistics from the generated facial features or embeddings.

In this paper, we introduce the Unified Tri-Feature Quality (U3FQ) metric, a novel approach in Facial Image Quality Assessment (FIQA). U3FQ redefines FIQA by integrating recognizability and quality estimation through a unique, learning-based methodology. Our method diverges from conventional paradigms by employing match scores in a weakly supervised manner, serving as our primary quality. The salient contributions of our work are:

- U3FQ is the first to systematically analyze facial age and expression similarity within the FIQA framework, offering new insights into age-related dynamics and expression variance in face recognition systems.
- Our method demonstrates enhanced superior accuracy and robust generalization across various benchmark datasets, marking a substantial significant advancement over current FIQA and IQA methodologies.

The notable enhancement in U3FQ is the adoption of a binary classification scheme for facial expressions. This sophisticated approach acknowledges the crucial role of expressions in determining face recognition accuracy. By distinctly categorizing expressions as congruent or disparate, also highlights the substantial impact that facial expressions wield in the overall quality assessment process. Further, as illustrated in Figure 1, U3FQ integrates age disparity as a fundamental component, offering an insightful visualization of how age variations influence matching similarity. This integration is pivotal, as it profoundly acknowledges the impact of age on the authenticity and reliability of face recognition systems. By doing so, U3FQ captures the true essence of biometric image quality, utility and fidelity in face recognition scenarios, transcending the conventional boundaries of FIQA in Face Recognition Systems.

2. Related Works

In this paper, we contextualize our work within the broader FIQA landscape, in which we are introducing the Unified Tri-Feature Quality (U3FQ) metric as a novel perspective in FIQA. Our approach, inspired by the latest trends in unsupervised, semi-supervised, and regression-based learning, uniquely integrates facial biometrics features such as age and facial expressions. This integration enriches the conventional FIQA framework, steering it towards more nuanced and holistic assessments.

2.1. Innovative Approaches in FIQA

Recent innovations in face recognition have been driven by a blend of sophisticated unsupervised and semi-supervised learning methods, fundamentally aimed at enhancing the recognizability of advanced face recognition systems. These methods, exemplified by seminal works such as SER-FIQ [26], SDD-FIQA [20], PCNet [29], and [4], have demonstrated the effectiveness of leveraging intrinsic data characteristics and a robust combination of both annotated and unannotated data. They underscore the potential of using advanced embedding variability analysis and similarity distribution distancing strategies to comprehensively assess facial image quality.

In line with these developments, our U3FQ metric extends the principles of these innovative learning approaches. It diverges from traditional methods by not solely relying on pseudo quality labels or embedding uncertainty. Instead, U3FQ incorporates additional biometric data, refining the quality assessment process and addressing biases inherent in label-dependent methods.

2.2. Integrated Biometric Analysis in FIQA

Drawing on the strengths of both advanced computational techniques and human-perceivable facial attributes, the Unified Tri-Feature Quality (U3FQ) metric represents a sophisticated amalgamation of the finest elements found in FIQA methodologies. U3FQ, while sharing conceptual similarities with notable works like CR-FIQA [6], FaceQnet [14], and FaceQAN [3], distinctively pushes the boundaries of conventional FIQA approaches. It incorporates a deeper, more nuanced integration of biometric analysis, transcending traditional computational assessments.

What distinctly sets the Unified Tri-Feature Quality (U3FQ) metric apart is its meticulous attention to the subtleties of facial biometrics, an aspect often underemphasized in other models. This robust integration ensures that U3FQ not only aligns with but also significantly enhances the practical applications of face recognition systems. Acknowledging the importance of facial expressions, as underscored in seminal studies [7, 16, 24], U3FQ integrates these critical aspects into its framework. Likewise, it draws upon the biometric significance of facial age features, as detailed in pivotal research works [2, 10, 12, 25], demonstrating how age characteristics can profoundly impact recognition tasks.

By factoring in the intricacies of facial expressions and age disparities, U3FQ emerges as an unparalleled tool, resonating profoundly with the real-world demands and complexities of contemporary face recognition technology. It transcends traditional FIQA approaches by offering a more contextually enriched and biometrically informed perspective. This pioneering integration not only elevates U3FQ within the FIQA domain but also paves the way for more contextually aware and accurate face recognition systems, setting a new benchmark for future innovations. U3FQ, therefore, stands as a significant advancement in FIQA, offering a comprehensive solution well-suited to the evolving demands of face recognition systems. Its ability to synthesize various learning approaches and incorporate key biometric features positions it as a groundbreaking tool.

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 field 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 [19] dataset. - The design of our computational framework, which utilizes both Random Forest and ResNet models for quality score prediction, is presented in Subsection 3.3.

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 2. 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 [27], OpenFace [1], ArcFace [9], and FaceNet [23]. Due to page limitations, these plots are included in the supplementary material. Here, we have added the DET plots from VGG-Face in Figure 4, 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 of recognition. Figure 3 illustrates the impact that discrep-

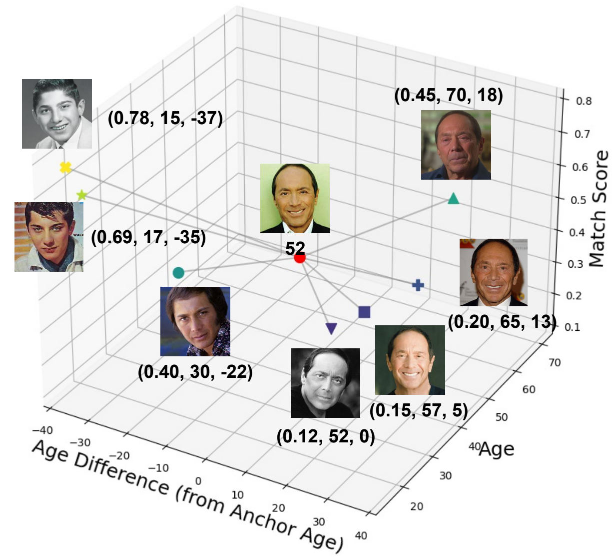

Figure 2. The efficacy of face matching systems is significantly impacted by the age variation between the images being compared. The triplet representation emphasizes the similarity distance, the age of the compared image, and the age difference in relation to the anchor image, with Image 6 as the reference.

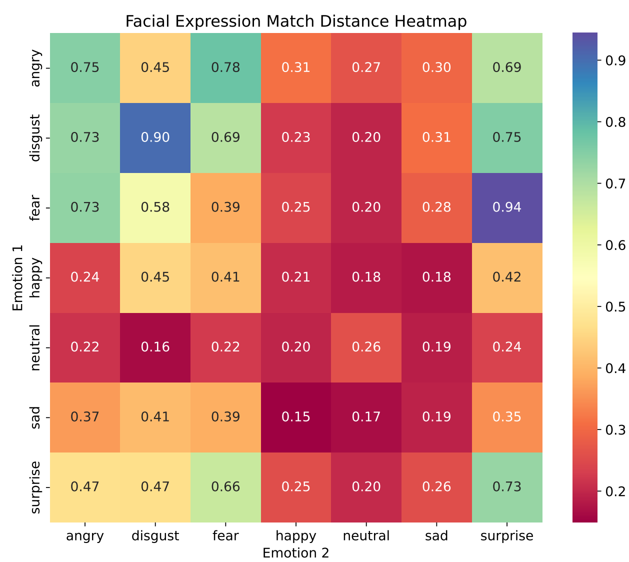

Figure 3. The differential impact of facial expressions on the match score, with weak emotions having a constant effect and strong emotions modifying the score proportionally to their intensity.

CVPR #12988

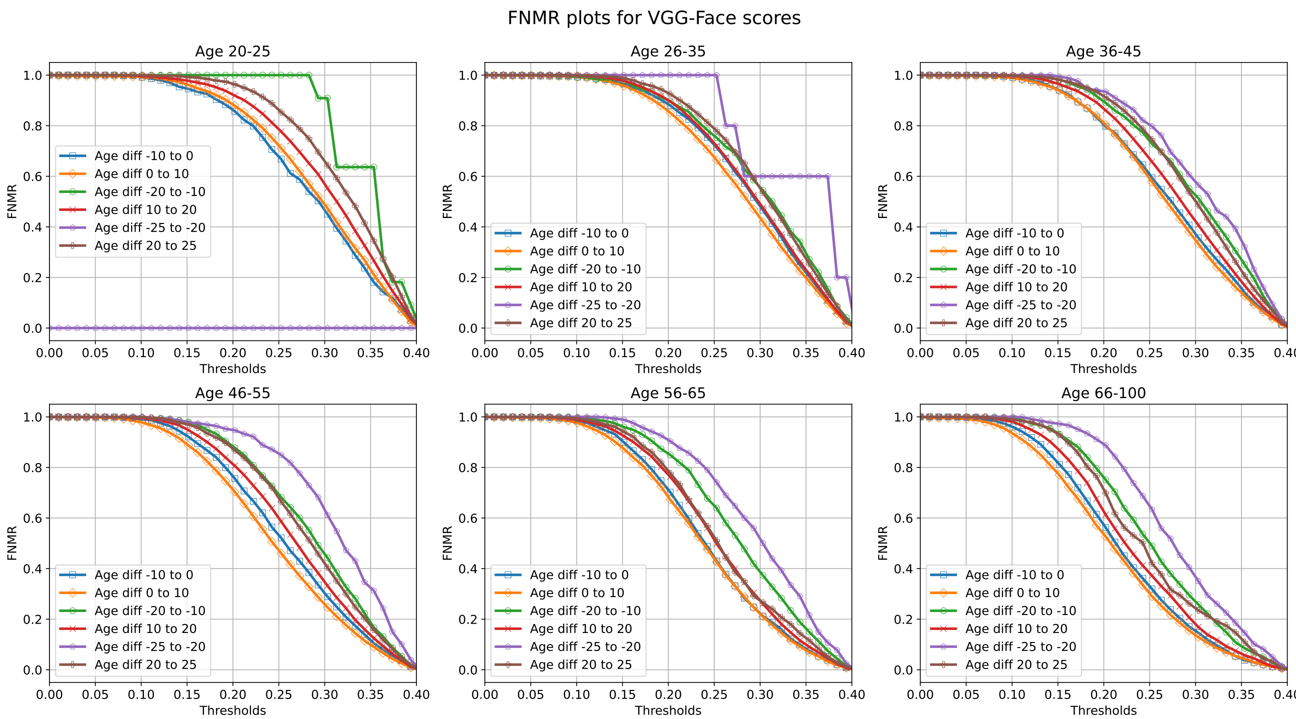

Figure 4. The VGG-Face DET plots, displaying the False Non-Match Rate (FNMR) across various age groups and age difference categories, reveal significant insights about the effects of anchor age on facial recognition accuracy. Notably, the age group of 35-45 years aligns closely with facial images across 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, drawn from the AgeDB dataset [19] using the deepface face verification library.

ancies in facial expressions have on matching performance, as evidenced by the average match scores across diverse expression pairs.

Our methodology ensures a more refined and context-sensitive assessment of facial similarity, taking into account not just the physical resemblance but also the expressive context of each face. This leads to a more accurate and realistic evaluation of facial images, particularly relevant in dynamic real-world scenarios where facial expressions can vary significantly.

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:

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

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:

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

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

In Figure 5 utilizing the equation 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

CVPR #12988

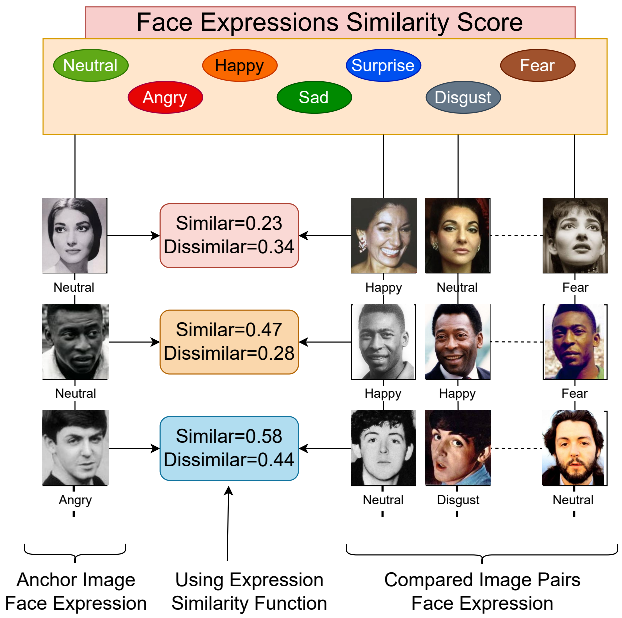

Figure 5. The methodical process of extracting similarity scores for individual images based on their matching counterparts

predictable and consistent for recognition purposes, are assigned the highest scores. In contrast, faces exhibiting strong emotions such as surprise or happiness, despite being similar, receive comparatively lower scores. This adjustment acknowledges the impact of expressive variability on the recognizability of faces.

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.

3.3. Architecture

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.

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

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 or Quality Score

$U \leftarrow 0$
$A \leftarrow []$
$E \leftarrow []$
for all $model \in M$ **do**
  $d \leftarrow$ ComputeMatchScoreDistance$(I, model)$
  $\mathcal{A} \leftarrow f(d, a)$
  $A \leftarrow A + [\mathcal{A}]$
  if ExpressionsAreSame$(I, model)$ **then**
    $\mathcal{E} \leftarrow c$
  else
    $\mathcal{E} \leftarrow d *$ EXPR_SCORE(e)
  end if
  $E \leftarrow E + [\mathcal{E}]$
  $U \leftarrow U + d * \mathcal{A} * \mathcal{E}$
end for
$F \leftarrow [a, e, U]$
procedure U3FQ_ASSESSMENT$(I, F, RN, m = 100)$
  $QualityScores \leftarrow []$
  for $i \leftarrow 1$ to m **do**
    $quality \leftarrow RN$.Predict(I, F)
    $QualityScores \leftarrow QualityScores + [quality]$
  end for
  $finalQuality \leftarrow$ Average$(QualityScores)$
  return $finalQuality$
end procedure

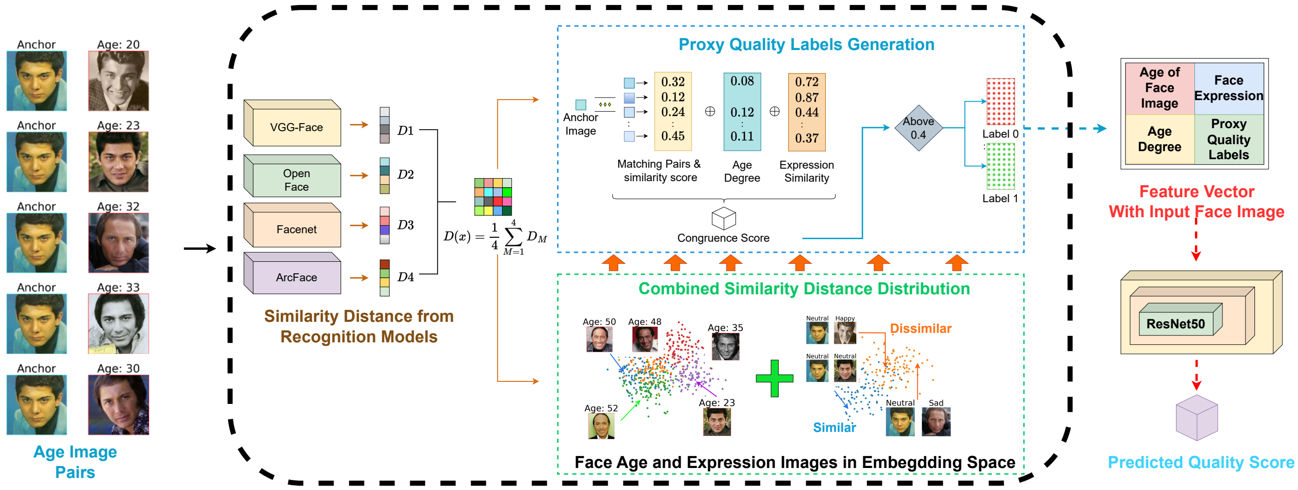

Figure 6. The figure presents a method for generating pseudo ground truth labels in face recognition by assessing age-related variations and expression similarity. It starts by calculating similarity distances between images of the same individuals at different ages using face recognition models. These distances are then normalized and combined with age and expression data to produce labels. These labels are crucial for fine-tuning regression model, leading to a comprehensive quality score that encapsulates recognition accuracy, age differences, and expression similarities.

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.

3.3.1 Regression Network and Quality Estimation

We have advanced and refined an existing Convolutional Neural Network (CNN), originally pre-trained for face recognition tasks, through a process of fine-tuning. This approach of adapting deep learning models to tasks akin to their initial training has been demonstrated effectively in numerous studies. Such networks have been repurposed for detecting facial attributes distinct from identity, including gender, age, and race. In the context of face quality assessment, it is posited that a feature vector containing discriminative facial information should inherently encapsulate 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.

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.

4. Experiments

4.1. Datasets

In our study, we utilize the AgeDB dataset [19], which comprises approximately 16,488 images representing a wide spectrum of age variations across different identities. This dataset serves as the cornerstone for our evaluations. For comprehensive assessment and benchmarking, we employ several esteemed datasets: LFW [15], ColorFeret [21], and Adience [11]. These datasets provide a diverse range of facial imagery that enables us to rigorously compare and analyze the performance of our proposed methodologies. Results are meticulously evaluated, showcasing the robustness of our approach in handling age-variant facial recognition tasks.

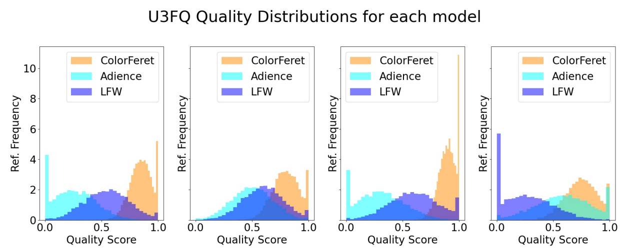

Figure 7. Distribution of scores on different datasets.

5. Results

6. Conclusion

The Unified Tri-Feature Quality Metric (U3FQ) signifies a pivotal advancement in the realm of Facial Image Quality Assessment (FIQA). By integrating critical elements such as age variance and facial expression impact, U3FQ presents a nuanced and comprehensive method for evaluating facial images, which closely aligns with the practical demands of face recognition systems. This research emphasizes the significance of these biometric features in enhancing the accuracy and reliability of recognition models, thereby transcending the confines of traditional FIQA metrics that predominantly rely on subjective human visibility assessments. Through rigorous evaluations on an extensive array 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 by incorporating additional factors such as illumination and pose alongside age variance and expression. This enhancement will aim to further refine the accuracy of reference quality labels, ensuring that U3FQ remains at the forefront of FIQA methodologies. By addressing these additional variables, we intend to broaden the scope and effectiveness of U3FQ, thereby making it an even more robust tool for assessing facial image quality in diverse and challenging recognition scenarios.

References

[1] Brandon Amos, Bartosz Ludwiczuk, Mahadev Satyanarayanan, et al. Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science*, 6(2):20, 2016. 3

[2] Raphael Angulu, Jules R Tapamo, and Aderemi O Adewumi. Age estimation via face images: a survey. *EURASIP Journal on Image and Video Processing*, 2018(1):1–35, 2018. 2

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

[4] Lacey Best-Rowden and Anil K Jain. Automatic face image quality prediction. *arXiv preprint arXiv:1706.09887*, 2017. 2

[5] Fadi Boutros, Naser Damer, Jan Niklas Kolf, Kiran Raja, Florian Kirchbuchner, Raghavendra Ramachandra, Arjan Kuijper, Pengcheng Fang, Chao Zhang, Fei Wang, et al. Mfr 2021: Masked face recognition competition. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. IEEE, 2021. 1

[6] Fadi Boutros, Meiling Fang, Marcel Klemt, Biying Fu, and Naser 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. 2

[7] Andrew J Calder and Andrew W Young. Understanding the recognition of facial identity and facial expression. *Facial Expression Recognition*, pages 41–64, 2016. 2

[8] Kai Chen, Taihe Yi, and Qi Lv. Lightqnet: Lightweight deep face quality assessment for risk-controlled face recognition. *IEEE Signal Processing Letters*, 28:1878–1882, 2021. 1

[9] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos 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. 3

[10] Natalie C Ebner. Age of face matters: Age-group differences in ratings of young and old faces. *Behavior research methods*, 40:130–136, 2008. 2

[11] Eran Eidinger, Roee Enbar, and Tal Hassner. Age and gender estimation of unfiltered faces. *IEEE Transactions on information forensics and security*, 9(12):2170–2179, 2014. 6

[12] Yun Fu, Guodong Guo, and Thomas S Huang. Age synthesis and estimation via faces: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 32(11):1955–1976, 2010. 2

[13] Klemen Grm and Vitomir Struc. Deep face recognition for surveillance applications, 2018. 1

[14] Javier Hernandez-Ortega, Javier Galbally, Julian Fierrez, Rudolf Haraksim, and Laurent Beslay. Faceqnet: Quality assessment for face recognition based on deep learning. In *2019 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2019. 1, 2

[15] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric 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. 6

[16] Shan Li and Weihong Deng. Deep facial expression recognition: A survey. *IEEE transactions on affective computing*, 13(3):1195–1215, 2020. 2

[17] Blaž Meden, Peter Rot, Philipp Terhörst, Naser Damer, Arjan Kuijper, Walter J Scheirer, Arun Ross, Peter Peer, and Vitomir Štruc. Privacy–enhancing face biometrics: A comprehensive survey. *IEEE Transactions on Information Forensics and Security*, 16:4147–4183, 2021. 1

[18] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng 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. 2

[19] 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. 3, 4, 6

[20] Fu-Zhao Ou, Xingyu Chen, Ruixin Zhang, Yuge Huang, Shaoxin Li, Jilin Li, Yong Li, Liujuan Cao, and Yuan-Gen 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. 2

[21] P Jonathon Phillips, Hyeonjoon Moon, Syed A Rizvi, and Patrick J Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on pattern analysis and machine intelligence*, 22(10):1090–1104, 2000. 6

[22] Torsten Schlett, Christian Rathgeb, Olaf Henniger, Javier Galbally, Julian Fierrez, and Christoph Busch. Face image quality assessment: A literature survey. *ACM Computing Surveys (CSUR)*, 54(10s):1–49, 2022. 1

[23] Florian Schroff, Dmitry Kalenichenko, and James 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. 3

[24] Caifeng Shan, Shaogang Gong, and Peter W McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and vision Computing*, 27(6):803–816, 2009. 2

[25] Gillian Slessor, Deborah M Riby, and Ailbhe N Finnerty. Age-related differences in processing face configuration: The importance of the eye region. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 68(2):228–231, 2013. 2

[26] Philipp Terhorst, Jan Niklas Kolf, Naser Damer, Florian Kirchbuchner, and Arjan 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. 2

[27] VGG. The visual geometry group (vgg) at the university of oxford. 3

[28] Mei Wang and Weihong Deng. Deep face recognition: A survey. *Neurocomputing*, 429:215–244, 2021. 1

[29] Weidi Xie, Jeffrey Byrne, and Andrew Zisserman. Inducing predictive uncertainty estimation for face recognition. *arXiv preprint arXiv:2009.00603*, 2020. 1, 2