

Use of (Input) Perturbations/Transformations for Quality Assessment of Samples

Jernej Sabadin

University of Ljubljana, Faculty of Electrical Engineering
Trzaška cesta 25, SI-1000 Ljubljana, Slovenia

js2005@student.uni-lj.si

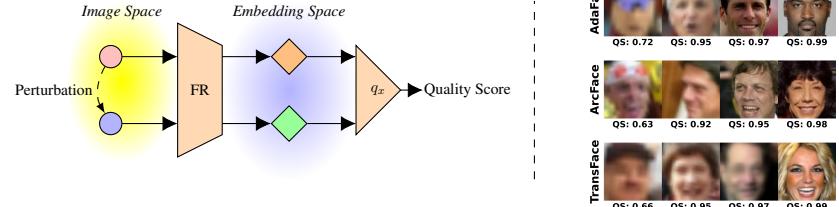
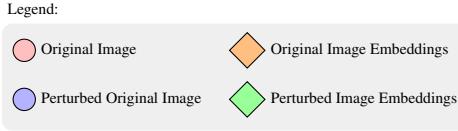


Figure 1: Illustration of our approach to Face Image Quality Assessment (FIQA). This schematic depicts the process of assessing image quality through a novel perturbation technique. Initially, the original image undergoes controlled perturbations to create a variant, symbolizing potential quality degradation. Both the original and perturbed images are then transformed into the embedding space using a Face Recognition (FR) model. In this space, we compute the cosine similarity between the embeddings of the original and perturbed images. The core hypothesis of our technique is that for lower-quality images, perturbations in the input space lead to more significant deviations in the embedding space, thereby resulting in a lower similarity score between the original and perturbed images. This stability of image embedding forms the backbone of our FIQA methodology. The quality scores displayed on the image are obtained using affine perturbation with the ArcFace FR model on the XQFW database.

Abstract

This paper addresses the crucial challenge of evaluating image quality in Face Recognition (FR) systems, especially under non-ideal, real-life circumstances. Assessing the quality of facial images is vital as it directly impacts the accuracy and reliability of FR systems by removing images below a certain predefined quality score threshold. High-quality images enhance the performance and dependability of these systems, while low-quality images can lead to significant errors and reduced functionality. We introduce a novel Face Image Quality Assessment (FIQA) methodology that focuses on the stability of face embeddings in the embedding space, evaluated through controlled perturbations. Our approach is predicated on the hypothesis that the quality of facial images is directly correlated with the stability of their embeddings; lower-quality images tend to exhibit greater variability upon perturbation.

To demonstrate the effectiveness and robustness of our methodology, we conduct comprehensive comparisons with a set of state-of-the-art FIQA techniques. Our evaluation spans various FR models and datasets, ensuring a thorough and diverse examination of performance. The results indicate that our approach, with its emphasis on embedding stability, not only provides an intuitive and direct measure of

image quality but also aligns well with the practical requirements of current FR systems. This contributes to enhancing the overall effectiveness of FR systems in real-world applications.

1. Introduction

In the evolving landscape of biometric recognition, Face Recognition (FR) technologies have significantly advanced, offering robust solutions in diverse applications. However, the efficacy of these systems is intrinsically linked to the quality of the input facial images. The emerging field of Face Image Quality Assessment (FIQA) seeks to enhance the reliability and accuracy of FR systems by rigorously evaluating the quality of these images. The image quality that we are assessing is closely related to utility [18]. The concept of utility in the context of Face Image Quality Assessment (FIQA) encompasses the effectiveness of face recognition comparisons. This effectiveness is influenced by factors such as the topography of the face and the texture of the skin. Additionally, it pertains to the degree of similarity that the facial image holds with the biometric characteristics that the Face Recognition (FR) system is designed to analyze and match. This paper introduces a novel FIQA



methodology that leverages the intrinsic properties of face embeddings, derived from a state-of-the-art FR model, to assess image quality.

Over the years, numerous highly effective Face Image Quality Assessment (FIQA) methods have emerged, broadly classified into three main categories: (i) regression-based, (ii) model-based, and (iii) analytical methods. Regression-based methods are trained using pseudo quality scores, enabling them to infer the quality of images. Model-based methods integrate the extraction of embeddings with the computation of quality scores, harnessing the capabilities of the underlying Face Recognition (FR) models. Analytical methods, on the other hand, deduce the quality score directly from the facial image, circumventing the need for generating pseudo labels. Our methodology, aligning with the analytical category, effectively avoids potential noise or biases that could stem from external sources of information.

Our method begins by applying controlled perturbations to a facial image and analyzing both original and perturbed versions using an FR model in an embedding space. The key hypothesis is that lower-quality images show greater variability in this space, evident from reduced cosine similarity between their original and perturbed embeddings. This stability of embeddings is crucial to our FIQA technique’s assessment of image quality.

We present our contributions in a manner that underscores the practical applicability and theoretical advancement of our methodology:

- We examine embedding stability through controlled perturbations across various parameter spaces as a measure of image quality.
- We extensively compare our method with state-of-the-art FIQA techniques, demonstrating its effectiveness, despite being a straightforward and intuitive method.

2. Related work

In this section, we categorize face image quality assessment (FIQA) methods into three primary branches: analytical techniques, regression-based techniques, and model-based methods. We offer a concise overview of each category and highlighting their unique approaches. A more detailed insight into the field is available in [18].

Analytical FIQA methods. Traditional methods assess image quality through facial symmetry, reflecting illumination and pose variations [7], and extend to capture characteristics including contrast, blur, and sharpness, utilizing random decision forests for utility prediction [8]. A comprehensive approach combines pose reconstruction, frequency analysis for blur, and histogram analysis for brightness, employing a cascaded classification process for efficient quality assessment [11]. Another study leverages CNNs for a

multi-branch assessment focusing on alignment, visibility, deflection, and clarity, and prioritizing score fusion [14], while a different research utilizes pose estimation, GLCM texture analysis, and GMM for quality quantification [17]. The significant advantage of these methods is that they do not rely on artificially generated or human-labeled quality scores. This is beneficial as it reduces potential biases and inconsistencies inherent in subjective labeling. Recently more powerful new methods emerged, taking into account the specificities of the Face Recognition (FR) models being used and the input sample for quality estimation, such as FaceQAN by Babnik et al. [23] and SER-FIQ by Terhörst et al. [21]. FaceQAN is based on the premise that higher quality images possess more stable representations, and the difficulty in generating adversarial samples is directly correlated with image quality. Similarly, SER-FIQ operates on the principle that the quality of a face image can be inferred from the robustness of its embeddings, which are evaluated by observing the variation in representations generated from different subnetworks of a face recognition model. DiFIQA [1], a novel FIQA technique by Žiga Babnik et al., leverages denoising diffusion probabilistic models to assess face image quality by analyzing the stability of image embeddings under perturbations introduced by forward and backward diffusion processes. Another approach called FaceQ-gen [9] is a semi-supervised, data-driven face image quality assessment method using Generative Adversarial Networks (GANs), which transforms a face image of unknown quality into a high-quality version, estimating quality based on the similarity between the original and restored images.

Regression-Based FIQA methods. These methods refer to a category of Face Image Quality Assessment (FIQA) approaches that utilize regression models trained on datasets with pre-assigned (pseudo) quality labels. One of such techniques is FaceQnet [10]. FaceQnet involves a process of selecting a high-quality ICAO-compliant image for each subject in the dataset and then comparing other images of the same subject to this high-quality reference image. The comparison scores obtained from these mated pairs are used to determine the quality of the non-ICAO images. The FaceQnet model, based on the pre-trained ResNet-50 architecture, is then fine-tuned using this labeled data. Another method called PCNet [22] determines image quality by extracting labels from the pairwise verification scores of mated face-image pairs, focusing on the lower predictive confidence of each image in the pair. In a similar vein, SDD-FIQA [16] generates labels using both mated and non-mated pairs, employing the Wasserstein Distance to analyze the difference between intra-class and inter-class similarity distributions. Another method utilizing quality label generation is LightQNet, which is described in [3]. This method adopts a binary classification approach for face quality assessment, focusing on difficult samples near

the classification boundary. LightQNet stands out for its lightweight yet efficient architecture.

Model-based methods These methods combine the face recognition and FIQA tasks. One of such methods is PFE [20] where the output magnitude vector represents the central features of face images, while the output variance vector quantifies the confidence or certainty in those feature. Extending this concept, [4, 13] improve the robustness and speed. Another approach named MagFace [15] by Q. Meng et al. refines ArcFace [6] by adaptively structuring within class distributions and using the feature vector magnitude to determine image quality. Another method called CR-FIQA [2] estimates face image quality by learning to predict the relative classifiability of an image, measured based on its feature representation in relation to its class center and nearest negative class center in angular space.

3. Methodology

In advancing the field of Face Image Quality Assessment (FIQA), our methodology seeks to encapsulate a comprehensive understanding of image stability in the embedding space. By limiting the assessment to the model and the image, the FIQA method ensures that the evaluation is relevant and directly related to the task at hand, while also eliminating the potential noise or bias that might come from external sources of information. While drawing inspiration from principles similar to those in methods like [23, 1, 21], our approach diverges significantly in its application and objectives.

The core of our methodology shown in Figure 2 hinges on the strategic introduction of varied image perturbations. By systematically applying these perturbations to input images and subsequently analyzing their effects on the embeddings within FR models, we gain vital insights into the relationship between image quality and the performance of these models.

3.1. Description of Perturbation Techniques

Our methodology incorporates a wide range of image perturbations to evaluate the resilience of face recognition (FR) models. These include Gaussian noise, salt and pepper noise, Poisson noise, affine transformations, elastic transformations, and both structured and random occlusions.

The process unfolds in two primary phases. The first phase involves the generation of a perturbed image from the original. For an image $\mathbf{I} \in \mathbb{R}^{h \times w \times 3}$, a perturbation function f is applied, resulting in the altered image $\mathbf{I}^f \in \mathbb{R}^{h \times w \times 3}$. This stage transforms \mathbf{I} using one or more of the selected perturbation techniques.

In the second phase, we focus on quantifying the similarity between the embeddings of the original and perturbed images. The embeddings are derived using the FR model

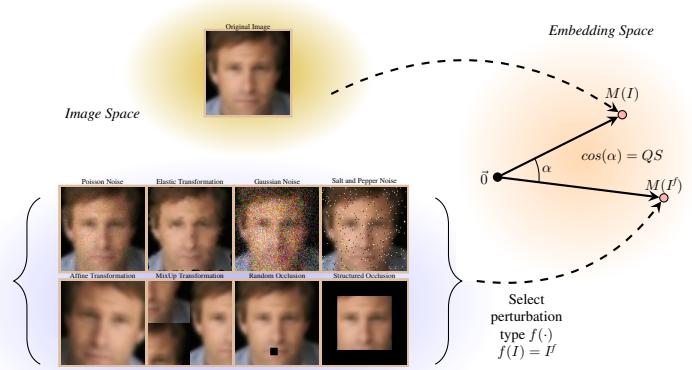


Figure 2: **Overview of our methods.** Illustrating the foundational concept of our approach, which focuses on the stability of embeddings. This stability is quantitatively assessed by measuring the cosine similarity between the embeddings of the original image and its perturbed counterpart. The closer the cosine similarity is to 1, the more stable the embeddings are considered, indicating a higher resilience of the face recognition model to image perturbations.

M , and the quality score $Q(\mathbf{I}) = q_x(M(\mathbf{I}), M(\mathbf{I}^f))$ is determined based on the cosine similarity $q_x(\cdot)$ between vectors $M(\mathbf{I})$ and $M(\mathbf{I}^f)$.

Cosine similarity, crucial for this assessment, is defined for vectors $M(\mathbf{I})$ and $M(\mathbf{I}^f)$ as:

$$q_x(M(\mathbf{I}), M(\mathbf{I}^f)) = \frac{M(\mathbf{I}) \cdot M(\mathbf{I}^f)}{\|M(\mathbf{I})\| \|M(\mathbf{I}^f)\|} \quad (1)$$

where $\|\cdot\|$ represents the L_2 norm. This measure q_x effectively captures the impact of perturbations on the embeddings, thus reflecting the quality of the image in the context of the FR model.

3.2. Perturbations/Transformations Used

Gaussian Noise: For an image matrix \mathbf{I} of dimensions $h \times w \times c$, Gaussian noise perturbation is added:

$$\mathbf{I}^f = \mathbf{I} + \mathbf{G} \quad (2)$$

Here, \mathbf{G} is a matrix with Gaussian-distributed elements of mean $\mu = 0$ and variance σ^2 . In practice, pixel values in \mathbf{I}^f are adjusted to ensure they fall within the standard range of $[0, 255]$.

Salt and Pepper Noise: This perturbation introduces noise into an image \mathbf{I} of dimensions $h \times w \times c$. The perturbed image \mathbf{I}^f is defined as:

$$\mathbf{I}_{ijk}^f = \begin{cases} 255 & \text{with probability } \alpha \times \beta, \\ 0 & \text{with probability } \alpha \times (1 - \beta), \\ \mathbf{I}_{ijk} & \text{otherwise,} \end{cases} \quad (3)$$

where α represents the total amount of noise added to the image, and β dictates the ratio of salt (white) to the total noise. A higher β value increases the proportion of salt noise relative to pepper noise.

Poisson Noise: Poisson noise added to an image \mathbf{I} is modeled using the Poisson distribution for each pixel's intensity:

$$\mathbf{I}^f = \text{Poisson}(\mathbf{I}) \quad (4)$$

This process adds noise based on the original intensity values of \mathbf{I} . Each pixel's intensity directly influences the noise level introduced.

Affine Transformations: Affine transformations applied to an image \mathbf{I} that we use include rotation, scaling, and flipping. The transformed image \mathbf{I}^f is determined as follows:

- *Rotation and Scaling:* Represented by the transformation matrix M_{rot} :

$$M_{rot} = \begin{bmatrix} \alpha \cos(\theta) & -\alpha \sin(\theta) & (1 - \alpha)w/2 \\ \alpha \sin(\theta) & \alpha \cos(\theta) & (1 - \alpha)h/2 \end{bmatrix} \quad (5)$$

where θ is the rotation angle, α is the scaling factor, and w, h are the width and height of the image. The transformation matrix M_{rot} applies rotation and scaling around the top-left corner of an image and then attempts to re-center it by adjusting its position.

- *Flipping:* Horizontal flipping is denoted by M_{hflip} :

$$M_{hflip} = \begin{bmatrix} -1 & 0 & w \\ 0 & 1 & 0 \end{bmatrix}, \quad (6)$$

The cumulative transformation, combining rotation, scaling, and optional flipping, is applied to \mathbf{I} to obtain \mathbf{I}^f .

Structured Occlusion: This perturbation applies an occlusion to the borders of an image \mathbf{I} with dimensions $h \times w \times c$. The occluded image \mathbf{I}^f is defined by:

$$\mathbf{I}_{ijk}^f = \begin{cases} 0 & \text{for } i < t \text{ or } i \geq h - t \\ & \text{or } j < t \text{ or } j \geq w - t, \\ \mathbf{I}_{ijk} & \text{otherwise,} \end{cases} \quad (7)$$

where \mathbf{I}_{ijk} denotes the pixel value at location (i, j, k) , and t represents the thickness of the occlusion border. In this model, pixels within the border defined by t are set to zero, creating an occlusion, while the central part of the image remains unaffected.

Random Occlusions: Random occlusions are applied to an

image \mathbf{I} by overlaying a square block of pixels with an occlusion of uniform color. The occluded image \mathbf{I}^f is given by:

$$\mathbf{I}_{ij}^f = \begin{cases} 0 & \text{if } i \in [y, y + s) \text{ and } j \in [x, x + s), \\ \mathbf{I}_{ij} & \text{otherwise,} \end{cases} \quad (8)$$

where $[x, x + s)$ and $[y, y + s)$ define the coordinates of the square occlusion with size s , and (i, j) indexes the pixels of the image. The values of x and y are chosen randomly such that the occlusion fits within the dimensions of \mathbf{I} .

MixUp Transformation: This technique involves partitioning an image \mathbf{I} into a grid of $g_w \times g_h$ segments and randomly shuffling these segments. For an image divided into $N = g_w \cdot g_h$ segments, the MixUp transformed image \mathbf{I}^f is constructed as:

$$\mathbf{I}^f = \text{shuffle}(\{\mathbf{I}_{seg}\}), \quad \text{where } \mathbf{I}_{seg} \subset \mathbf{I} \quad (9)$$

Each segment \mathbf{I}_{seg} is defined by its grid coordinates and shuffled such that no segment remains in its original location. The shuffling function shuffle reassembles the segments into the new image \mathbf{I}^f , ensuring a non-identical arrangement compared to \mathbf{I} .

Elastic Transformation: An elastic transformation perturbs an image \mathbf{I} by applying a displacement field that models natural, non-linear deformations. The transformed image \mathbf{I}^f is given by:

$$\mathbf{I}^f(x, y) = \mathbf{I}(x + \delta_x(x, y), y + \delta_y(x, y)) \quad (10)$$

where δ_x and δ_y are displacement fields along the x and y coordinates, generated from a random field smoothed by a Gaussian filter with standard deviation σ , and scaled by α indicating the deformation intensity. The transformation is applied using bilinear interpolation to preserve continuity.

4. Experimental Results

4.1. Experimental Setup

In our study, we aim to investigate the performance of our FIQA techniques in comparison with various state-of-the-art methods. An essential aspect of this evaluation is the recognition that the effectiveness of FIQA methods is inherently dependent on the face recognition (FR) model employed. To this end, our experimental setup is designed to compare the proposed FIQA techniques against established benchmarks using a range of FR models.

Face Recognition Models: Our analysis includes widely recognized state-of-the-art FR models like ArcFace [6], AdaFace [12], and TransFace [5]. These models have been chosen for their robust performance across diverse facial recognition tasks, providing a comprehensive basis for evaluating FIQA techniques.

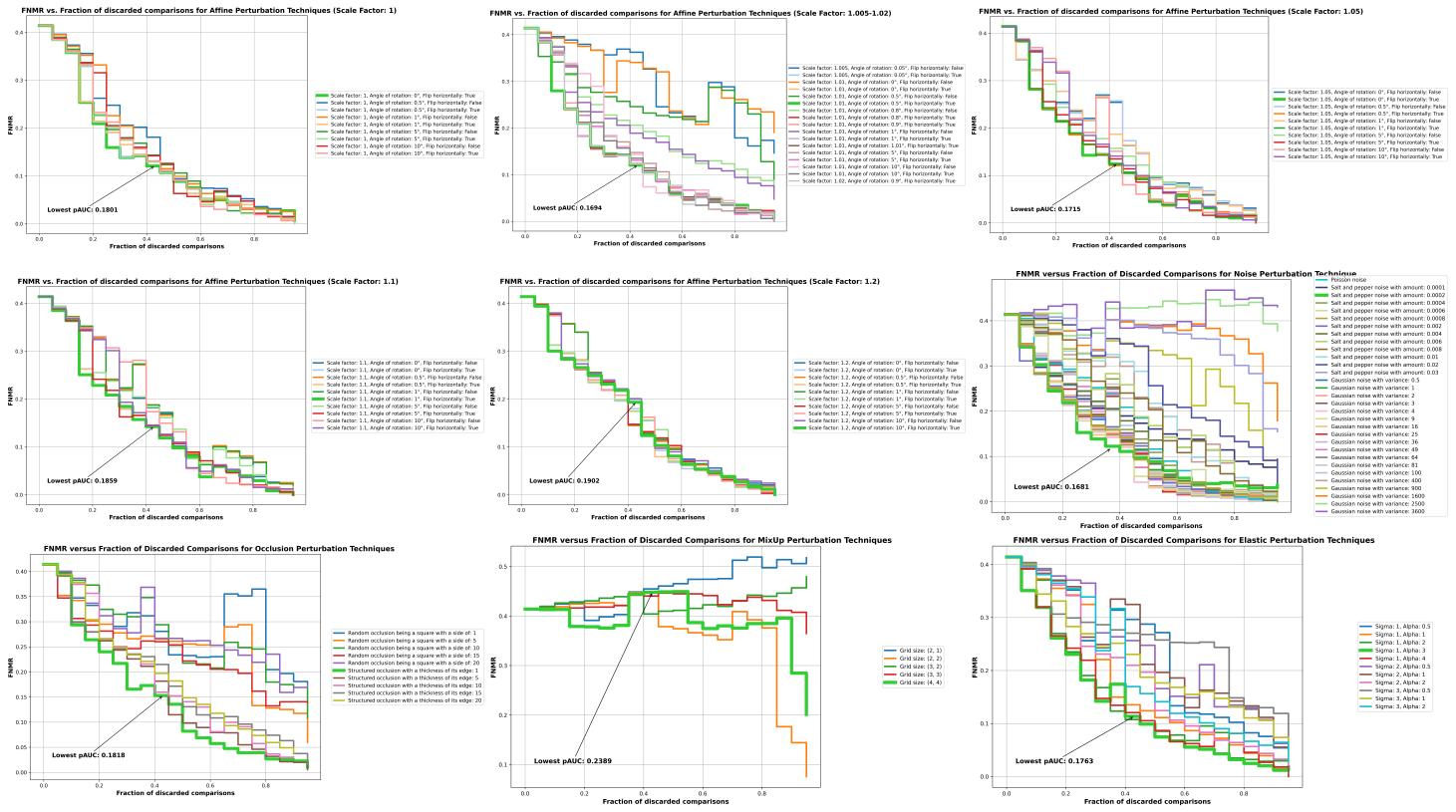


Figure 3: EDC curve analysis for searching the optimal parameters for image Perturbations. This figure demonstrates the influence of various perturbation types on EDC curves, with a special emphasis on Affine transformations. Due to the extensive number of curves, we feature five separate plots for Affine transformations, each representing a distinct scale factor. This analysis is conducted using the ArcFace model on the XQLFW dataset.

Error-versus-Reject Characteristic (EDC) Curves:

EDC curves are a critical component of our evaluation framework. These curves measure the False Non-Match Rate (FNMR) at a constant False Match Rate (FMR), which is typically set to 0.001, as we progressively increase the fraction of images that are excluded based on quality assessments. The EDC curves are crucial for identifying the optimal balance point where the rejection of low-quality images does not result in a significant reduction in FNMR. This balance is fundamental to refining FIQA methodologies: it aims to find a quality threshold that ensures high

recognition accuracy by the FR system without discarding a substantial number of images. By analyzing these curves, we can calibrate the FIQA techniques to maintain an ideal equilibrium, where the system’s accuracy is enhanced without compromising the practical utility of the facial image dataset.

Partial Area Under the Curve (pAUC): The pAUC metric quantifies the performance of Face Image Quality Assessment (FIQA) methods by calculating the area under the Error-versus-Discard Characteristic (EDC) curve for a predefined range of discard fractions. This range is selected to reflect the operational conditions typically encountered by Face Recognition systems, in our case is set to 0.3. The

pAUC thus provides an objective measure of how well the FIQA methods perform in keeping FNMR as low as possible while managing the trade-off with the proportion of discarded verification pairs due to quality considerations. For better comparisons between different datasets, we normalise the FNMR values using the initial one at the discard rate of 0%.

Databases: Our experimental evaluation encompasses several well-known datasets, each with distinct characteristics influencing FIQA performance.

Adience: This dataset includes 19,370 images spanning 2,284 identities with an equal number of 20,000 mated and non-mated comparisons. It features pose and expression varied to medium degree with moderate quality, low in blur and noise and good resolution, suitable for assessing FIQA in everyday image scenarios.

CPLFW: It contains 11,652 images of 3,930 identities. CPLFW is used for 3,000 mated and non-mated comparisons, presenting high pose variation, low expression and medium blur, noise, and resolution, challenging the robustness of FIQA techniques.

XQLFW: Presents low pose and expression variation but high blur and noise, testing FIQA methods ability to handle poor-quality images. XQLFW is used for 3,000 mated and 3,000 non mated pairs.

These datasets have been selected due to their extensive use in the field and the variety of image qualities they represent, encompassing a range of factors like age, pose, expression, and environmental conditions. This diversity is crucial for rigorously testing the robustness and adaptability of FIQA techniques across different real-world scenarios. In summary, our experimental setup is constructed to provide a comprehensive evaluation of FIQA techniques in the context of varying FR models. By analyzing EDC curves and pAUC metrics that are closely examined in the [19] across multiple datasets, we aim to gain deep insights into the efficacy of these techniques in practical facial recognition scenarios.

4.2. Parameter Tuning for Enhanced Perturbation Effectiveness

To attain optimal performance, our study explores a wide array of perturbation parameters. The primary goal is to identify parameter configurations that effectively minimize the Partial Area Under the Curve (pAUC), thereby pinpointing the most efficient perturbation settings.

In exploring the extensive parameter space for optimal perturbation effects, we implemented a methodical grid search approach, complemented by manual examination. This approach enables us to evaluate the impact of different perturbations on the Error-versus-Reject Characteristic (EDC) curves. We focus on adjusting and refining these parameters using the XQLFW dataset, noted for its challenging low-quality images, with the ArcFace model as our primary tool for identifying the best parameters.

Upon determining the optimal parameters, we extend our evaluation to other models and datasets in a cross-model setting. This includes tests within the same model framework and assessments across different models. Such a comprehensive approach not only tests the robustness of our Face Image Quality Assessment (FIQA) methods but also investigates the generalizability of the Quality Score (QS) across various face recognition models. Through this process, we aim to ensure that our FIQA techniques are adaptable and effective in diverse real-world scenarios.

Figure 3 succinctly demonstrates that smaller perturbations are generally more effective than their larger counterparts. Notably, the MixUp method, which introduces the most significant alterations, is identified as the least effective. On the other hand, the most favorable results are achieved through a carefully calibrated affine transformation, characterized by slight rotation, scaling, and flipping. Among the five affine transformation scenarios, the one with the largest scale factor exhibits the poorest performance. It's observed that enhancing outcomes by reducing the scale factor and applying minor rotations yields better results. However, these improvements are not monotonically decreasing the pAUC; they reach a point of diminish-

Table 1: Comparative analysis with leading techniques using partial area under the curve (pAUC) metrics. This table encapsulates a comparison with premier state-of-the-art techniques, utilizing pAUC values computed at a False Match Rate (FMR) of 10^{-3} and a discard threshold of 0.3. We also provide the average of $pAUC$ values calculated across all datasets for a specific model, denoted as \overline{pAUC} . The most exemplary results are highlighted in green, followed by the second-best in yellow, and the third-best in orange, delineating a clear performance hierarchy among the evaluated methods.

ArcFace				
FIQA model	XQLFW	CPLFW	Adience	$pAUC$
CR-FIQA	0.16932	0.17116	0.05225	0.13091
DiffFIQA	0.16241	0.16205	0.06524	0.12990
FaceQAN	0.16411	0.16717	0.10366	0.14498
MagFace	0.22478	0.17861	0.04794	0.15044
SDD-FIQA	0.20938	0.18332	0.04701	0.14657
FIQA - Affine Transformation	0.16936	0.17305	0.05495	0.13245
FIQA - Elastic Transformation	0.18400	0.20173	0.19023	0.19199
FIQA - MixUp Transformation	0.23798	0.25243	0.22022	0.23688
FIQA - Salt and Pepper noise	0.18279	0.20485	0.20725	0.19830
FIQA - Structured Occlusion	0.18180	0.20507	0.12455	0.17047
AdaFace				
FIQA model	XQLFW	CPLFW	Adience	$pAUC$
CR-FIQA	0.17149	0.14046	0.02869	0.11355
DiffFIQA	0.15712	0.15821	0.07070	0.12868
FaceQAN	0.16275	0.16117	0.12151	0.14848
MagFace	0.22943	0.17695	0.02746	0.14461
SDD-FIQA	0.21040	0.17905	0.03592	0.14179
FIQA - Affine Transformation	0.17192	0.17154	0.02500	0.12282
FIQA - Elastic Transformation	0.19184	0.19737	0.20809	0.19910
FIQA - MixUp Transformation	0.25059	0.24997	0.24033	0.24696
FIQA - Salt and Pepper noise	0.18977	0.20070	0.21532	0.20193
FIQA - Structured Occlusion	0.17555	0.19662	0.10954	0.16057
TransFace				
FIQA model	XQLFW	CPLFW	Adience	$pAUC$
CR-FIQA	0.14251	0.14904	0.02500	0.10552
DiffFIQA	0.12846	0.13987	0.06761	0.11198
FaceQAN	0.13629	0.14455	0.12171	0.13418
MagFace	0.23307	0.16786	0.02500	0.14198
SDD-FIQA	0.20530	0.17211	0.03391	0.13711
FIQA - Affine Transformation	0.15318	0.16031	0.02500	0.11283
FIQA - Elastic Transformation	0.18235	0.18419	0.20794	0.19149
FIQA - MixUp Transformation	0.25185	0.25497	0.24278	0.24987
FIQA - Salt and Pepper noise	0.17813	0.19012	0.21248	0.19358
FIQA - Structured Occlusion	0.18027	0.18367	0.11366	0.15920

ing returns at a certain level of minimal perturbation, beyond which further reductions in perturbation cease to yield improvements. This trend is also evident in the application of salt and pepper noise, where a very small amount of noise with an equal salt to pepper ratio performs better. Despite the minimal extent of noise corruption, further decreasing the noise level does not lead to enhanced performance, indicating that the most effective perturbation is not necessarily the smallest, but rather one that is subtly balanced.

The optimal parameters identified from the analysis presented in Figure 3 are as follows: For noise perturbations, specifically Salt and Pepper noise, the balance between salt and pepper is set with a ratio of $\beta = 0.5$, and the overall noise amount is $\alpha = 0.0002$. In the case of affine transformations, the scaling factor is chosen as $\alpha = 1.01$, the rotation angle is $\theta = 0.5^\circ$, and horizontal flipping is applied. For elastic transformations, the parameters are op-

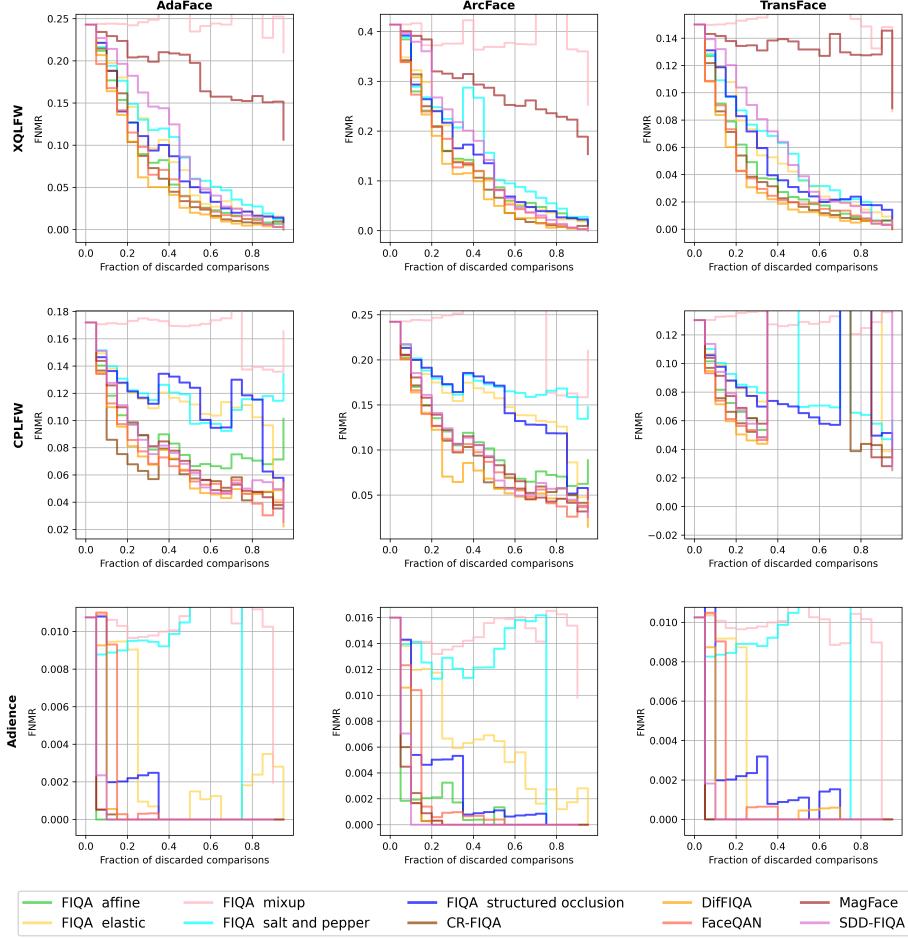


Figure 4: Evaluative Comparison of Perturbation-Based FIQA Techniques Against Established State-of-the-Art Methods Using Non-Interpolated EDC Curves. This comparative analysis encapsulates a diverse range of scenarios, encompassing three distinct facial recognition (FR) models and three different datasets. A critical observation from the resulting curves indicates that the affine transformation technique demonstrates robust competitiveness, notably aligning with the performance of the top two DiffFIQA and FaceQAN methodologies. These insights underline the efficacy of perturbation approaches in the FIQA domain, offering valuable benchmarks for future advancements.

timized with $\alpha = 3$ and $\sigma = 1$. Structured occlusion is implemented with an edge thickness of $t = 1$ pixel, and the MixUp transformation employs a grid size of $g_w = g_h = 4$.

4.3. Comparison with the State-of-the-Art

We compare our perturbation techniques using best parameters with five state-of-the-art methods. The results, in terms of EDC curves, are shown in Figure 4. We use a discard rate of 0.3 and a constant FMR of 0.001 when calculating the pAUC values, as suggested in Section 4.1. The partial Area Under the Curve (pAUC) metrics are detailed in Table 1. It is pertinent to observe that the values in this table may diverge from those optimized in Figure 3 for perturbation types such as Elastic Transformation, Affine Transformation, and Salt and Pepper Noise. This variation can be

attributed to the non-fixed random seed during the generation of these perturbations, leading to stochastic differences in the results.

We can observe that our FIQA method, which uses Affine Transformation, yields very competitive results compared to other techniques. It even outperforms other state-of-the-art methods on the Adience dataset when using the AdaFace FR model. Overall, considering the $p\overline{AUC}$, the Affine Transformation perturbation ranks third twice when using the TransFace and ArcFace models and second once with the AdaFace model. In Figure 5, we present the distribution of quality scores obtained using the ArcFace model coupled with Affine transformations. It is observed that the distribution predominantly skews towards a quality score (QS) near 1. This tendency can be attributed to the minimal

nature of the perturbations applied, which, in turn, induce only slight variations in the embedding space.

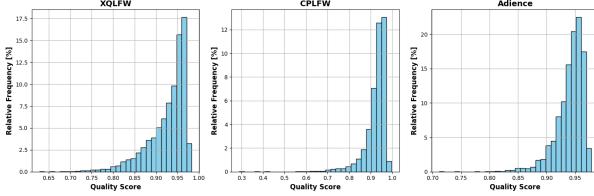


Figure 5: Distribution of quality scores

5. Conclusion

In this study, we have extensively explored the stability of facial image embeddings in response to controlled perturbations. Our methodology focused on assessing how various perturbations impact the embedding space.

Our findings reveal that Affine perturbations, under a specific set of parameters, yield the most significant insights into the stability of facial embeddings. This approach has proven to be particularly effective in distinguishing between high-quality and lower-quality images.

Moreover, the results are highly competitive when compared with existing state-of-the-art FIQA methods.

This work paves the way for future research in FIQA, suggesting that further exploration of parameter spaces and perturbation techniques could uncover even more profound insights into the quality assessment of facial images. Our approach demonstrates the potential of using embedding stability as a robust and intuitive metric for assessing the quality of facial images in various biometric applications.

References

- [1] Babnik, P. Peer, and V. Štruc. Diffiqqa: Face image quality assessment using denoising diffusion probabilistic models. *arXiv preprint arXiv:2305.05768*, 2023.
- [2] F. Boutros, M. Fang, M. Klemt, B. Fu, and N. Damer. Cr-fiqqa: Face image quality assessment by learning sample relative classifiability. In *Proceedings of the CVF/IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.
- [3] K. Chen, T. Yi, and Q. Lv. Lightqnet: Lightweight deep face quality assessment for risk-controlled face recognition. *Signal Processing Letters*, 28:1878–1882, 2021.
- [4] K. Chen, T. Yi, and Q. Lv. Fast and reliable probabilistic face embeddings based on constrained data uncertainty estimation. In *Image and Vision Computing (IVC)*, page 104429, 2022.
- [5] J. Dan, Y. Liu, H. Xie, J. Deng, H. Xie, X. Xie, and B. Sun. Transface: Calibrating transformer training for face recognition from a data-centric perspective. 2023.
- [6] Deng, Jiankang, G. Jia, Xue, Niannan, Zafeiriou, and Stefanos. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [7] X. Gao, S. Z. Li, R. Liu, and P. Zhang. Standardization of face image sample quality. In *Proceedings of the IAPR International Conference on Biometrics (ICB)*, pages 242–251. Springer, 2007.
- [8] O. Henniger, B. Fu, and C. Chen. On the assessment of face image quality based on handcrafted features. In *Proceedings of the IEEE International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5, 2020.
- [9] J. Hernandez-Ortega, J. Fierrez, I. Serna, and A. Morales. Faceqgen: Semi-supervised deep learning for face image quality assessment. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–8, 2021.
- [10] J. Hernandez-Ortega, J. Galbally, J. Fierrez, and L. Beslay. Biometric quality: Review and application to face recognition with faceqnet. *arXiv preprint arXiv:2006.03298*, 2020.
- [11] H.-I. Kim, S. H. Lee, and Y. M. Ro. Investigating cascaded face quality assessment for practical face recognition system. In *Proceedings of the IEEE International Symposium on Multimedia (ISM)*, pages 399–400, 2014.
- [12] M. Kim, A. K. Jain, and X. Liu. Adaface: Quality adaptive margin for face recognition. 2022.
- [13] S. Li, J. Xu, X. Xu, P. Shen, S. Li, and B. Hooi. Spherical confidence learning for face recognition. In *Proceedings of the CVF/IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, page 15629–15637, 2021.
- [14] Z. Lijun, S. Xiaohu, Y. Fei, D. Pingling, Z. Xiangdong, and S. Yu. Multi-branch face quality assessment for face recognition. In *Proceedings of the IEEE International Conference on Communication Technology (ICCT)*, page 1659–1664, 2019.
- [15] Q. Meng, S. Zhao, Z. Huang, and F. Zhou. A universal representation for face recognition and quality assessment. In *Proceedings of the CVF/IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, page 14225–14234, 2021.
- [16] F.-Z. Ou, X. Chen, R. Zhang, Y. Huang, S. Li, J. Li, Y. Li, L. Cao, and Y.-G. Wang. Sdd-fiqqa: Unsupervised face image quality assessment with similarity distribution distance. In *Proceedings of the CVF/IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7670–7679, 2021.
- [17] R. Raghavendr, K. B. Raja, B. Yang, and C. Busch. Automatic face quality assessment from video using gray level co-occurrence matrix: An empirical study on automatic border control system. In *Proceedings of the IAPR International Conference on Pattern Recognition (ICPR)*, pages 438—443, 2014.
- [18] T. Schlett, C. Rathgeb, O. Henniger, J. Galbally, J. Fierrez, and C. Busch. Face image quality assessment: A literature survey. 2020.
- [19] T. Schlett, C. Rathgeb, J. Tapia, and C. Busch. Considerations on the evaluation of biometric quality assessment algorithms. 2023.

- [20] Y. Shi and A. K. Jain. Probabilistic face embeddings. In *Proceedings of the CVF/IEEE International Conference on Computer Vision (CVPR)*, pages 6902–6911, 2019.
- [21] 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 CVF/IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5651—5660, 2020.
- [22] W. Xie, J. Byrne, and A. Zisserman. Inducing predictive uncertainty estimation for face verification. In *British Machine Vision Conference (BMVC)*, 2020.
- [23] Žiga Babnik, P. Peer, and V. Štruc. Faceqan: Face image quality assessment through adversarial noise exploration. In *Proceedings of the IAPR International Conference on Pattern Recognition (ICPR)*, page 748–754, 2022.