PreFIQs: Face Image Quality Is What Survives Pruning

Anonymous CVPR submission

Paper ID 9

Abstract

Face Image Quality Assessment (FIQA) evaluates the utility of a face image for automated face recognition (FR) systems. In this work, we propose PreFIQA, an unsupervised and training-free FIQA framework grounded in the Pruning Identified Exemplar (PIE) hypothesis. We hypothesize that low-utility face images rely disproportionately on fragile network parameters, resulting in larger geometric displacement of their embeddings under model sparsification. Accordingly, PreFIQA quantifies image utility as the Euclidean distance between L2-normalized embeddings extracted from a pre-trained FR model and its pruned counterpart. We provide a first-order theoretical justification via a Jacobian-vector product analysis, demonstrating that this empirical drift serves as a computationally efficient approximation of the exact geometric sensitivity of the latent embedding manifold. Extensive experiments across eight benchmarks and four FR models demonstrate that PreFIQA achieves competitive or superior performance compared to state-of-the-art supervised and unsupervised FIQA methods, including establishing new state-of-the-art results on several benchmarks, without any training or supervision. These results validate parameter sparsification as a principled and practically efficient signal for face image utility, and demonstrate that quality is, in essence, what survives pruning.

1. Introduction

Face Recognition (FR) systems have achieved remarkable accuracy [11, 29, 49], yet their performance can degrade in unconstrained and challenging real-world settings. Images captured in the wild often exhibit extreme variations in pose, illumination, occlusion, and blur, posing significant challenges for recognition [52]. To address these issues, Face Image Quality Assessment (FIQA) has emerged as a critical preprocessing step, measuring the utility of a face image for automated recognition [26]. High-utility images produce stable and discriminative embeddings leading to reliable recognition, while low-utility images generate uncertain embeddings, undermining the robustness of FR systems [15, 46, 48].

Based on the type of supervision, current state-of-the-art (SOTA) FIQA methods can be broadly categorized as supervised, unsupervised, or self-supervised. Supervised approaches [9, 21, 37] rely on explicit or proxy labels to learn quality scores. Self-supervised approaches [2, 8, 35, 40, 46] jointly optimize FR and FIQA. Unsupervised approaches [3, 6, 31, 42, 48], including our proposed PreFIQA, infer quality by evaluating the robustness of embeddings from pre-trained FR models under stochastic perturbations. A central hypothesis in unsupervised FIQA is that high-utility images produce representations that are resilient to perturbations. For example, SER-FIQ [48] measures embedding variance across multiple forward passes with random dropout, DifFIQA [4] leverages diffusion processes to quantify robustness against noise, and ViTNT-FIQA [42] tracks the stability of feature evolution across transformer blocks. While effective, stochastic methods like SER-FIQ incur substantial computational overhead due to repeated inference, and gradient-based methods such as GraFIQs [31] require backpropagation, which can be prohibitively expensive for real-time deployment.

In this work, we propose **PreFIQA** training-free FIQA framework that measures image utility through model sparsity sensitivity. Our method is grounded in the observation that high-utility images produce feature representations that remain stable under moderate network pruning, whereas low-utility images rely on fragile, easily disrupted parameters. Concretely, we compute the Euclidean distance between L2-normalized embeddings generated by a pre-trained FR model and its pruned counterpart. This embedding drift serves as a proxy for image utility: smaller drift indicates stable identity encoding and high quality, while larger drift signals sensitivity to pruning and lower utility. By interpreting embedding stability under model sparsity as a quality measure, PreFIQA offers a new principled perspective on image utility. The method is fully training-free, requires no labels, and directly captures the functional contribution of each image to the recognition model's robustness. We validate PreFIQA across seven standard bench-

CVPR #9

marks and four FR models, demonstrating competitive or superior performance compared to SOTA supervised and unsupervised FIQA methods.

2. Related Work

Face Image Quality Assessment (FIQA) methods have evolved along several complementary directions, which can be broadly categorized into three paradigms: supervised, unsupervised, and self-supervised approaches.

Supervised approaches typically train quality regressors using explicit or proxy supervision. For example, FaceQnet [21] relies on ICAO compliance labels, SDD-FIQA [37] derives pseudo-labels from similarity-distribution distances, and RankIQ [9] formulates FIQA as a learning-to-rank problem. Subsequent works improve the reliability of these labels: CLIB-FIQA [39] calibrates the confidence of quality anchors, while MR-FIQA [38] leverages multi-reference representations generated from synthetic data to reduce label noise.

Unsupervised approaches can be further divided into non-FR model approaches and FR-specific approaches. Non-FR model approaches estimate face quality without relying on conventional FIQA regressors or pretrained FR. DifFIQA [4] measures sample robustness through diffusion-based modeling, and eDifFIQA [5] distills this into a lightweight predictor. DSL-FIQA [10] combines degradation-aware representation learning with landmark-guided transformers.

FR-specific approaches probe frozen FR backbones without retraining to estimate FIQ. SER-FIQ [48] measures embedding stability under dropout perturbations, GraFIQs [31] exploits gradient-based signals, and FaceQAN [3] links quality to adversarial robustness. Recent training-free methods extend these ideas to transformer architectures and intermediate layers: ViTNT-FIQA [41] tracks embedding-trajectory stability across ViT layers, while FROQ [6] identifies informative intermediate layers via a lightweight calibration step to predict quality in a single forward pass. These methods represent a shift toward efficient, probe-based FIQA without additional supervision.

Self-supervised approaches, often implemented as FR-integrated methods, jointly optimize FR and FIQA. MagFace [35] links quality to embedding magnitude, PFE [46] models uncertainty in embeddings as a quality proxy, and ViT-FIQA [2] introduces a learnable quality token that directly predicts FIQ scores, while CR-FIQA [8] explicitly learns *relative classifiability* across identities, providing a task-relevant measure of utility rather than relying on surrogate labels or embedding magnitude.

Building on these insights, **PreFIQA** introduces a complementary perspective: it quantifies image utility through *sparsity-induced representational drift*. By measuring how sparsifying model parameters affects the embeddings of each image, PreFIQA captures the functional importance of samples for the recognition model itself. This training-free, data-free metric with minimal computational overhead provides a deterministic proxy for utility, emphasizing model robustness and discriminative stability, distinguishing it from prior FIQA approaches.

3. Methodology

In this section, we introduce Pruning-based Face Image Quality Assessment (PreFIQA). PreFIQA quantifies the utility of face images by measuring the representational drift induced by controlled model sparsification. Our method builds on the Pruning Identified Exemplar (PIE) hypothesis [23], which proves that the performance of compressed Deep Neural Networks (DNNs) disproportionately degrades on difficult or low-quality samples. We extend this principle to FR and hypothesize that low-utility face images exhibit higher sensitivity to parameter pruning, resulting in larger geometric displacement in the embedding space. Conversely, high-utility samples produce identity representations that remain stable under moderate structural compression.

Leveraging this asymmetry, we define face image quality as the stability of L2-normalized embeddings under pruning, measured as the Euclidean distance between embeddings extracted from the original and sparsified models. This drift-based formulation provides a deterministic and architecture-aligned proxy for face image utility, requiring neither additional training nor auxiliary supervision.

3.1. Preliminary on Model Pruning

Model pruning is a form of model compression that reduces the effective capacity of a DNN by removing redundant parameters [22, 32]. Recent SOTA FR models are typically over-parameterized [1, 32], allowing substantial parameter removal while maintaining strong verification performance. Pruning strategies can be broadly categorized along two principal dimensions: (i) the *pruning criterion* used to identify removable parameters, and (ii) the *pruning granularity*, i.e., whether parameters are removed individually (unstructured) or in groups (structured).

Let an FR model be denoted by $M_\theta : x \to \mathbb{R}^d$, that map input x to $\mathbb{R}^d$ embedding and parameterized by $\theta \in \mathbb{R}^N$, where N denotes the total number of learnable parameters. Given a target sparsity ratio $\rho \in (0, 1)$, pruning aims to construct a sparsified parameter vector θ_ρ satisfying:

$$\|\theta_\rho\|_0 = (1 - \rho)N, \tag{1}$$

where $\|\cdot\|_0$ denotes the ℓ_0 pseudo-norm counting non-zero entries.

Pruning can be formulated as the application of a binary mask $\mathbf{m}_\rho \in \{0, 1\}^N$ to the original parameters:

$$\theta_\rho = \mathbf{m}_\rho \odot \theta, \tag{2}$$

where $\odot$ denotes the element-wise product. The mask $\mathbf{m}_\rho$ is constructed such that a fraction ρ of parameters is set to zero.

Pruning criterion. The pruning criterion determines how parameters are selected for removal. As a baseline, parameters can be removed uniformly at random, independent of their magnitude or functional contribution. However, random pruning does not explicitly target redundant parameters and often leads to lower accuracy compared to methods that remove unimportant ones [14, 47]. Importance can also be estimated using first-order information, e.g., the magnitude of gradients with respect to a loss function $\mathcal{L}$. Parameters with small $|\partial\mathcal{L}/\partial\theta_i|$ are considered less influential and can be pruned. However, this criterion required access to a training dataset to select parameters to be pruned, which is out of scope of this work, where we propose a training- and data-free FIQA approach. A common and effective strategy is magnitude-based pruning, where parameters with the smallest absolute values are removed under the assumption that low-magnitude weights contribute less to the model output. In this case, a threshold τ is determined such that

$$m_{\rho,i} = \mathbb{I}(|\theta_i| > \tau), \tag{3}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. This selection mechanism assumes that parameters with small magnitude contribute less to the network output and can therefore be removed with limited impact on global performance.

Granularity of Pruning. Pruning can be applied either in an unstructured manner, where individual weights are set to zero while preserving the network topology, or in a structured manner, where entire parameter groups (e.g., filters or channels) are removed, requiring corresponding architectural adjustments.

In PreFIQA, pruning is not used for computational acceleration but as a controlled mechanism to systematically reduce model capacity. The sparsified model M_{θ_ρ} therefore provides a principled means of analyzing how embedding representations respond to reductions in network capacity.

3.2. PreFIQs

Recent FR models [11, 33, 49] encode identity information in the angular direction of the embedding space. Consequently, feature representations are L2-normalized and lie on the unit hypersphere. Let $M_\theta(x) \in \mathbb{R}^d$ and $M_{\theta_\rho}(x) \in \mathbb{R}^d$ denote the L2-normalized embeddings of an input sample $x \in \mathcal{X}$ extracted by the original and sparsified FR models, respectively.

Building upon the PIE hypothesis [23, 28], we interpret pruning as a controlled reduction of model capacity that exposes the structural dependence of a sample's representation on specific parameters. If the identity encoding of x relies heavily on parameters removed during pruning, its embedding will undergo a measurable geometric displacement. Conversely, embeddings that are encoded in more redundant or stable parameter subspaces will remain comparatively invariant under moderate sparsification. We therefore quantify the utility of a sample x by measuring the representation drift induced by pruning:

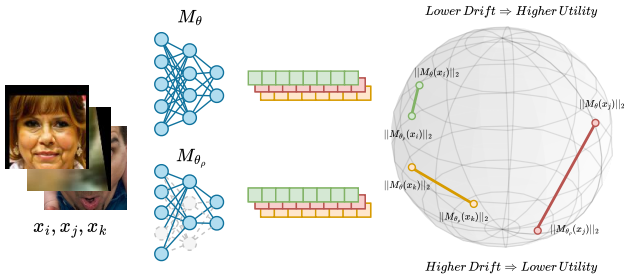


Figure 1. Given face images, e.g, x_i, x_j, and x_k, we extract their L2-normalized embeddings using a pre-trained FR and its sparsified counterpart. FIQ is quantified, for each image, as the Euclidean distance between its corresponding embeddings, measuring the pruning-induced representation drift. Smaller drift indicates stable identity encoding and thus higher image utility, while larger drift reflects structural sensitivity and lower quality.

$$D(x) = \left\| M_\theta(x) - M_{\theta_\rho}(x) \right\|_2 . \tag{4}$$

Since both embeddings are L2-normalized, they lie on the unit hypersphere, and the Euclidean distance is bounded:

$$0 \leq D_\rho(x) \leq 2. \tag{5}$$

Moreover, the Euclidean distance between normalized embeddings is directly related to angular deviation:

$$D^2(x) = 2 - 2\cos\left(\angle\big(M_\theta(x), M_{\theta_\rho}(x)\big)\right), \tag{6}$$

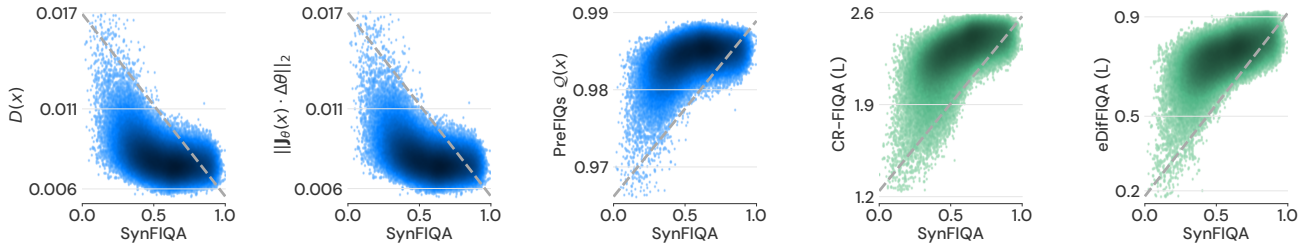
demonstrating that $D(x)$ measures the angular displacement of identity information in latent space.

To obtain a normalized FIQ score $Q(x) \in [0, 1]$, where higher values indicate higher utility, we apply linear rescaling:

$$Q(x) = 1 - \frac{D(x)}{2}. \tag{7}$$

Under this formulation: $Q(x) \approx 1 \Leftrightarrow$ high embedding stability (high utility), and $Q(x) \approx 0 \Leftrightarrow$ large structural sensitivity (low utility).

Importantly, this drift-based formulation is deterministic, requires no auxiliary supervision or stochastic perturbations, and directly aligns the quality estimate with the geometry of the identity embedding manifold.



(a) $D(x)$ (Eq. 4) (b) Jacobian Drift (Eq. 9) (c) PreFIQs (Ours, Eq. 7) (d) CR-FIQA (L) [8] (e) eDifFIQA (L) [5]

Figure 2. Density maps using SynFIQA [38] dataset (550k images), and their proxy labels (x-axis, higher value indicates higher utility) versus various FIQA predictions (y-axis). Figs. 2b and 2a validate our approximation (Eq. 9), showing consistent distributions between Jacobian-based drift (Eq. 9) and empirical pruned-model distance (Eq. 4, lower indicates higher utility). Figs. 2c–2e compare our normalized, unsupervised PreFIQs score with supervised CR-FIQA [8] and eDifFIQA [5] (higher indicates higher utility). Note that SynFIQA labels are algorithmic pseudo-labels, inherently biased by the used synthetic generation, rather than absolute ground truth.

3.2.1. Theoretical Validation via Jacobian-Vector Product

To mathematically validate why the empirical drift $D(x)$ serves as a principled proxy for face image utility, we formalize the model's sensitivity to sparsification using a first-order Taylor expansion. We model the pruning process as an additive structural perturbation $\Delta\theta$ applied to the network weights, where $\Delta\theta_i = -\theta_i$ if weight θ_i is pruned, and 0 otherwise, yielding $\theta_\rho = \theta + \Delta\theta$. Note that this additive formulation is equivalent to the mask-based sparsification in Eq. 2, where $\Delta\theta_i = -\theta_i \cdot (1 - m_{\rho,i})$. Under moderate sparsification, where $\|\Delta\theta\|$ remains sufficiently small, the perturbed face embedding can be approximated as:

$$M_{\theta+\Delta\theta}(x) \approx M_\theta(x) + \mathbf{J}_\theta(x) \cdot \Delta\theta, \tag{8}$$

where $\mathbf{J}_\theta(x) \in \mathbb{R}^{d\times N}$ is the Jacobian of the L2-normalized embedding with respect to the weights θ, evaluated at input x. Rearranging Eq. 8 and taking the ℓ_2-norm of both sides, the magnitude of the theoretical representation drift is governed by the norm of the Jacobian-vector product:

$$||\mathbf{J}_\theta(x) \cdot \Delta\theta||_2 \approx ||M_{\theta+\Delta\theta}(x) - M_\theta(x)||_2. \tag{9}$$

Recent work [31] validates gradient magnitudes as a robust indicator of FIQ, where high-utility images induce low gradient magnitudes, while low-utility samples require parameter updates of higher magnitude to resolve the distribution shift measured by an auxiliary loss. However, rather than depending on backpropagation of an auxiliary distribution shift loss [31], PreFIQs directly probes the geometric sensitivity of the latent face embedding manifold via $\mathbf{J}_\theta(x)$, approximated without any gradient computation through the forward pass of the sparsified model.

Under our hypothesis that sample utility governs reliance on specific parameter subspaces, the following asymmetry is expected. Let x_{high} be a sample with substantially higher utility than x_{low}. Since x_{high} encodes identity in redundant, distributed parameter subspaces, it is comparatively robust to sparsification, and $\Delta\theta$ is expected to isolate near-zero Jacobian entries, yielding $||\mathbf{J}_\theta(x_{\text{high}}) \cdot \Delta\theta||_2 \approx 0$. Conversely, x_{low} relies more heavily on the pruned weights in $\Delta\theta$, yielding Jacobian entries of higher magnitude and consequently a stronger drift:

$$||\mathbf{J}_\theta(x_{\text{low}}) \cdot \Delta\theta||_2 \gg ||\mathbf{J}_\theta(x_{\text{high}}) \cdot \Delta\theta||_2. \tag{10}$$

While the Jacobian-vector product $||\mathbf{J}_\theta(x) \cdot \Delta\theta||_2$ exactly models this structural sensitivity, explicitly computing the full Jacobian $\mathbf{J}_\theta(x)$ is computationally intractable for SOTA architectures with tens of millions of parameters. Even forward-mode automatic differentiation introduces significant overhead. Equation 9 establishes that our proposed empirical distance $D(x)$ (Eq. 4) is a first-order approximation of this exact geometric sensitivity, providing a computationally efficient surrogate that requires neither backpropagation nor an auxiliary distribution shift loss [31].

To validate this theoretical derivation, we analyze the representation drift on SynFIQA [38], a comprehensive synthetic dataset constructed to systematically model diverse intra-class quality degradations. As shown qualitatively in Figures 2b and 2a, the density distributions of the exact Jacobian-vector product ($||\mathbf{J}_\theta(x) \cdot \Delta\theta||_2$) and the empirical Euclidean distance ($D(x)$) align almost perfectly across the quality spectrum, confirming that the static pruned model accurately captures the geometric sensitivity of the latent manifold. Quantitative validation on standard evaluation benchmarks is provided in Section 5.

Furthermore, Figures 2c through 2e contrast our final PreFIQs score ($Q(x)$) against the proxy labels of the SynFIQA. Notably, our training-free PreFIQs yields a quality density distribution closely aligned with SOTA supervised methods, such as CR-FIQA [8] and eDifFIQA [5], which require explicit training phases to learn quality regression.

4. Experimental Setup

Pretrained Model Architecture. We demonstrate the proposed PreFIQs approach using two publicly available pre-

trained FR models released by [11]. Specifically, we utilize a ResNet100 model trained on MS1MV2 [11, 18] and a ResNet50 model trained on CASIA-WebFace [51], both optimized using the ArcFace loss function. Evaluation results based on the ResNet50 architecture are provided in the supplementary material.

Validating the Jacobian Approximation. We verify the accuracy of our proposed distance metric by comparing it directly to the exact mathematical formula (Jacobian-Vector product). To do this, we calculate both the exact gradient-based drift and our simpler Euclidean distance $\mathcal{D}(x)$ across the datasets. The exact Jacobian product is computed using PyTorch's [43] automatic differentiation tools. We perform this comparison by pruning 10% of the model's weights ($\rho = 0.1$). Finally, we measure how closely the two methods align using the pAUC score up to a 30% discard rate.

Model Pruning. We implement a global pruning strategy that includes all parameters within the convolutional and linear layers of the evaluated architectures. To systematically assess the impact of parameter reduction, we perform a comparative analysis across the granularity of pruning unstructured and structured as well as pruning criterion and random pruning (baseline) and magnitude-based pruning. Unstructured pruning is using L_1-norm based magnitude pruning, and the ratio of ρ lowest magnitude parameters are pruned. Structured pruning is performed using the DepGraph framework [13] to manage architectural dependencies, and the final linear layer is not pruned in structural pruning to achieve the same face embedding dimensionality. Random pruning is randomly selecting parameters to prune to match pruning ratio ρ. The models are pruned across a comprehensive spectrum of target sparsity ratios, defined as $\rho \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$.

FR Performance. To assess the impact of model pruning on the FR verification performance, we evaluate the pruned FR models on a set of diverse evaluation benchmarks, including Labeled Faces in the Wild (LFW) [24], AgeDB (using 30 year gap protocol) [36], Celebrities in Frontal-Profile in the Wild (CFP-FP) [45], Cross-Age LFW (CALFW) [54], and Cross-Pose LFW (CPLFW) [53] using their official evaluation protocols..

Evaluation Benchmarks. To ensure alignment with recent SOTA FIQA evaluation protocols [8], we report our results across seven standard benchmarks: LFW[24], AgeDB [36], CFP-FP [45], CALFW [54], Adience [12], CPLFW [53], Cross-Quality LFW (XQLFW) [30] and IARPA Janus Benchmark–C (IJB-C) [34]. These datasets introduce a diverse set of challenging verification scenarios, containing significant variations in age (AgeDB and CALFW), head-pose (CFP-FP and CPLFW) and overall image quality (XQLFW).

Evaluation Metrics. We assess FIQA performance utilizing Error-Versus-Discard Characteristic (EDC) curves [16, 17], a standard evaluation metric in the literature (often referred to interchangeably as Error-Versus-Reject Curves, or ERC [4]). The EDC curve illustrates the impact of sequentially discarding a fraction of the lowest-quality face images on the overall face verification performance. This performance is measured by the False Non-Match Rate (FNMR) [27] evaluated at specific decision thresholds corresponding to fixed False Match Rates (FMR) [27]. In accordance with established SOTA FIQA methodologies [4, 8, 35], we plot the EDC curves across all benchmarks at two fixed FMRs: 10^{-3} and 10^{-4}. Furthermore, we calculate both the Area Under the Curve (AUC) (in supplementary) and the partial AUC (pAUC) for the plotted EDC curves, providing a quantitative metric of verification performance across all rejection thresholds. For better readability, we show $pAUC * 10^3$ and $AUC * 10^3$ values, which we will refer to as pAUC and AUC in the paper. Following standard practices in the literature, the pAUC is evaluated up to a 30% discard rate [4, 5, 44].

FR Models. To evaluate the generalizability of PreFIQs, we report verification performance across different quality discard rates using four distinct FR models: ArcFace [11], ElasticFace (ElasticFace-Arc) [7], MagFace [35], and CurricularFace [25]. For all experiments, we utilize the officially released pre-trained models provided by the respective authors [7, 11, 25, 35]. Each model shares a ResNet100 backbone [19] originally trained on the MS1MV2 dataset [11, 18], and processes 112×112 aligned and cropped input images to generate 512-dimensional feature embeddings.

We evaluate these models under two distinct protocols: *same-model* and *cross-model*. Under the same-model protocol, ArcFace [11] is employed both to compute the image quality scores and to execute the subsequent verification task. Under the cross-model protocol, ArcFace is used exclusively as the quality estimator to establish the discard rankings, while ElasticFace [7], MagFace [35], and CurricularFace [25] act as the independent verification models evaluating the remaining image pairs.

Comparisons with SOTA FIQA. We compare our PreFIQs approach against twelve SOTA FIQA methods: RankIQ [9], PFE [46], SDD-FIQA [37], MagFace [35], CR-FIQA [8], DifFIQA [4], eDifFIQA [5], CLIB-FIQA [40], VIT-FIQA [2] as supervised approaches, SER-FIQ [48], FaceQnet (v1 [21]) [20, 21], GraFIQs [31], ViTNT-FIQA [42] as unsupervised approaches, and FROQ [6] as a semi-supervised approach. A conceptual overview of PreFIQs and SOTA FIQA approaches is given in Table 1.

5. Results

This section provides extensive overview of our results. We first provide a quantitatively validation of the Jacobian-Vector product approximation introduced in Section 3.2.1, with results outlined in Table 2. We then provide exten-

Table 1. Comparison overview of the operation and concepts of various FIQA approaches with our PreFIQs. Unsupervised approaches are labeled with BLUE, and supervised and self-supervised methods are labeled with GREEN stripes, respectively.

Method	Quality Labels	Architecture Specific	Additional Training	Custom Loss	Inference				
					Feed-Forward	Backwards	Feature Level	Gradient Level	Representation Level
SDD-FIQA [37]	✓	✗	✓	✗	1	0	✓	✗	✗
PCNet [50]	✓	✗	✓	✗	1	0	✓	✗	✗
eDiFFIQA(L) [5]	✓	✗	✓	✓	1	0	✓	✗	✗
CLIB-FIQA [40]	✓	✓	✓	✓	1	0	✓	✗	✗
MagFace [35]	✗	✗	✓	✓	1	0	✓	✗	✗
CR-FIQA [8]	✗	✗	✓	✓	1	0	✓	✗	✗
ViT-FIQA(T) [2]	✗	✗	✓	✓	1	0	✓	✗	✗
FROQ [6]	✓	✗	✗	✗	1	0	✗	✗	✓
SER-FIQ [48]	✗	✓	✗	✗	100	0	✓	✗	✗
FaceQAN [3]	✗	✗	✗	✗	10	10	✓	✓	✗
GraFIQs [31]	✗	✗	✗	✗	1	1	✗	✓	✗
ViTNT-FIQA [42]	✗	✗	✗	✗	1	0	✓	✗	✗
PreFIQs (Ours)	✗	✗	✗	✗	2	0	✓	✗	✗

sive overview of the FIQA performance when using different granularity of pruning (structured vs. unstructured), and different pruning criteria (L_1 magnitude vs. random pruning) across different pruning ratios ρ. The results of these experiments are shown in Table 3. Additionally, we evaluate the impact of pruning on FR verification performance across a wide set of benchmarks, comparing the granularity of pruning and the pruning criteria across pruning ratios ρ. The results are shown in Table 4. At the end of this Section, we compare our PreFIQs against recent SOTA approaches. The results for four FR models are shown in Table 5.

5.1. Jacobian-Vector Validation

Table 2 presents the quantitative comparison between the exact Jacobian-Vector product (Equation 9) and our proposed discrete representation drift $D(x)$ (Equation 4). The results show that both methods achieve nearly identical pAUC scores across all seven evaluation benchmarks and four FR models. The average pAUC across all datasets and FR models is 10.514 for the theoretical Jacobian drift and 10.558 for our empirical discrete drift. This strong alignment empirically validates our mathematical derivation. It confirms that the computationally efficient PreFIQ, compared to the Jacobian-Vector product, accurately approximates the geometric sensitivity of the latent manifold.

Table 2. Empirical validation of the Jacobian-Vector product approximation. At a sparsity ratio of $\rho = 0.1$, the theoretical Jacobian-Vector product drift (Eq. 9) and the proposed discrete representation drift $\mathcal{D}(x)$ (Eq. 4) achieve nearly identical pAUC scores (discard rate = 0.3, across four used FR models). This quantitative alignment verifies that $\mathcal{D}(x)$ successfully captures the geometric sensitivity of the latent manifold.

Average across FR Models - $pAUC * 10^3$ (FMR= 10^{-3}) [↓]

Methods	Adience	AgeDB-30	CFP-FP	LFW	CALFW	CPLFW	XQLFW	$\overline{pAUC}$
Jacobian Drift (Eq. 9)	10.018	6.863	4.008	0.858	20.727	20.608	139.830	10.514
Discrete Drift $D(x)$ (Eq. 4)	10.068	6.913	3.995	0.858	20.826	20.689	138.775	10.558

5.2. Evaluation of Pruning Approaches

As outlined in our experimental setup, we compare the effects of different pruning strategies across various pruning ratios ρ. The average FIQA results for all FR models and benchmarks are presented in Table 3. This evaluation is divided into two main analytical comparisons.

Granularity of Pruning (unstructured vs. structured). The results clearly show that unstructured pruning consistently achieves the best performance across all tested ratios. It reaches the best average pAUC of 10.516 at a sparsity ratio of $\rho = 0.4$. Furthermore, unstructured pruning maintains highly stable pAUC values across the majority of the tested spectrum. Performance degradation only becomes apparent at extremely high sparsity levels starting at $\rho = 0.8$. In contrast, structured pruning achieves its best average pAUC of 11.137 at the lowest sparsity setting of $\rho = 0.1$ and shows significantly more sensitivity to increases in the pruning ratio. This steep performance decline can be attributed to the aggressive removal of entire architectural structures from the network. More importantly, this is also attributed to the fact that unstructured pruning maintains, to a large extent, FR verification accuracies compared to structured pruning, as shown in Table 4 and discussed in detail in Section 5.3.

Pruning criterion (L_1 magnitude vs. random pruning). Parameter selection based on L_1 magnitude vastly outperforms random parameter selection, as shown in Table 3. Random pruning yields a significantly worse pAUC of 18.298 at $\rho = 0.1$, in comparison to L_1 magnitude at the same pruning ratio. This can be attributed to the lower FR verification accuracies when the model is pruned using random pruning compared to L_1 magnitude criterion, as shown in Table 4 and discussed in detail in Section 5.3. Interestingly, after an initial performance drop, random pruning remains relatively stable across higher sparsity ratios compared to structured pruning. This suggests that pruning random parameters fails to isolate the critical network capacity responsible for encoding Pruning Identified Exemplars (Section 3), which ultimately results in a poor utility score.

5.3. Evaluation of FR Performance

Table 4 compares the underlying FR verification accuracy of the used FR models across different granularities of pruning and pruning criteria across different pruning ratios.

Granularity of Pruning: The results demonstrate that unstructured pruning significantly outperforms structured pruning. Unstructured pruning maintains a highly consistent verification performance across most pruning ratios, experiencing a notable drop only at the extreme ratio of $\rho = 0.9$. In contrast, structured pruning suffers a severe and rapid degradation in accuracy as entire architectural components are removed from the network.

Pruning criterion: Pruning parameters based on L_1

Table 3. FIQ: Average performance of unstructured vs. structured, and structured L_1 magnitude vs. structured random model pruning across different pruning ratios on four evaluated FR models reported using pAUC (discard rate = 0.3, FMR = 10^{-3}). The **best** and *second-best* results per dataset are highlighted. The final column displays the average pAUC across all benchmarks. XQLFW is excluded from this average. Within this column, the best result is shaded per pruning category. Further results provided in the supplementary material. It can be clearly observed that unstructured pruning led to better performance compared to structured pruning. In terms of pruning criterion, L_1 magnitude outperformed, with a clear margin, random selection.

Granularity of Pruning - Comparison between unstructured and structured model pruning - $pAUC * 10^3$ (FMR= 10^{-3}) [↓]

Methods	Adience	AgeDB-30	CFP-FP	LFW	CALFW	CPLFW	XQLFW	$\overline{pAUC}$
Unstructured ρ =0.1	9.924	**6.809**	3.846	0.866	21.435	*20.442*	137.081	10.554
Unstructured ρ =0.2	10.123	7.420	**3.719**	0.895	21.499	**20.099**	**134.395**	10.626
Unstructured ρ =0.3	9.803	6.961	4.040	*0.752*	21.724	20.473	*136.922*	10.626
Unstructured ρ =0.4	9.947	6.982	3.900	0.899	**20.715**	20.651	139.022	10.516
Unstructured ρ =0.5	**9.750**	7.049	*3.759*	0.871	21.081	20.629	139.902	10.523
Unstructured ρ =0.6	*9.797*	*6.867*	4.027	0.789	*20.722*	21.294	144.214	10.583
Unstructured ρ =0.7	10.060	7.066	4.579	0.855	21.336	25.299	150.031	11.533
Unstructured ρ =0.8	11.684	8.449	8.290	0.805	21.841	41.226	161.248	15.382
Unstructured ρ =0.9	14.758	9.464	11.722	0.867	22.995	58.045	177.987	19.642
Structured ρ =0.1	10.695	8.182	4.452	0.791	21.137	21.565	144.276	11.137
Structured ρ =0.2	12.066	9.810	9.218	0.956	21.925	49.520	157.267	17.249
Structured ρ =0.3	16.232	10.433	12.477	1.175	23.157	60.586	169.934	20.677
Structured ρ =0.4	16.651	11.621	12.596	0.918	24.447	60.028	170.537	21.043
Structured ρ =0.5	16.515	11.536	12.023	0.900	24.002	60.155	171.338	20.855
Structured ρ =0.6	16.189	11.070	11.985	0.920	24.049	59.175	171.926	20.565
Structured ρ =0.7	17.402	10.562	12.064	0.996	24.250	59.913	170.943	20.865
Structured ρ =0.8	16.504	10.630	12.155	1.005	24.195	60.143	172.842	20.772
Structured ρ =0.9	16.801	10.195	12.556	0.962	24.011	61.702	176.857	21.038

Pruning Criterion - Comparison between L_1 magnitude and random model pruning - $pAUC * 10^3$ (FMR= 10^{-3}) [↓]

Methods	Adience	AgeDB-30	CFP-FP	LFW	CALFW	CPLFW	XQLFW	$\overline{pAUC}$
L_1 Magnitude ρ =0.1	9.924	**6.809**	3.846	0.866	21.435	*20.442*	137.081	10.554
L_1 Magnitude ρ =0.2	10.123	7.420	**3.719**	0.895	21.499	**20.099**	**134.395**	10.626
L_1 Magnitude ρ =0.3	9.803	6.961	4.040	*0.752*	21.724	20.473	*136.922*	10.626
L_1 Magnitude ρ =0.4	9.947	6.982	3.900	0.899	**20.715**	20.651	139.022	10.516
L_1 Magnitude ρ =0.5	**9.750**	7.049	*3.759*	0.871	21.081	20.629	139.902	10.523
L_1 Magnitude ρ =0.6	*9.797*	*6.867*	4.027	0.789	*20.722*	21.294	144.214	10.583
L_1 Magnitude ρ =0.7	10.060	7.066	4.579	0.855	21.336	25.299	150.031	11.533
L_1 Magnitude ρ =0.8	11.684	8.449	8.290	0.805	21.841	41.226	161.248	15.382
L_1 Magnitude ρ =0.9	14.758	9.464	11.722	0.867	22.995	58.045	177.987	19.642
Random ρ =0.1	15.101	9.258	10.482	**0.729**	22.696	51.522	181.241	18.298
Random ρ =0.2	17.821	10.157	12.858	1.135	24.010	61.720	182.106	21.283
Random ρ =0.3	16.707	10.695	12.351	0.903	23.764	62.029	176.676	21.075
Random ρ =0.4	16.419	10.792	12.560	0.883	23.553	61.756	177.312	20.994
Random ρ =0.5	15.986	10.751	12.271	0.916	23.650	62.384	179.484	20.993
Random ρ =0.6	16.258	11.102	12.256	0.921	24.065	61.378	176.981	20.997
Random ρ =0.7	16.568	10.617	12.211	0.981	23.791	61.519	178.097	20.948
Random ρ =0.8	17.238	10.988	12.203	0.900	24.041	61.293	180.073	21.110
Random ρ =0.9	17.221	10.615	12.025	0.882	24.163	61.529	180.387	21.072

magnitude drastically outperforms random pruning. Unstructured random pruning begins to lose its discriminative power almost immediately and completely collapses to random guessing (accuracy 50.00%) at a relatively low sparsity ratio of $\rho = 0.5$.

This rapid decline in the accuracy of FR verification is directly correlated and explains the corresponding loss in the FIQA performance observed for structured and random pruning strategies discussed in the previous Section 5.2.

5.4. Comparison to State-of-the-Art

Table 5 presents a comparison of our PreFIQs (unstructured L_1 magnitude pruning at $\rho = 0.4$, our best setups Table 3) against recent FIQA approaches across four FR models.

The results demonstrate that PreFIQs achieves highly competitive performance compared to the top-performing SOTA methods. Most notably, PreFIQs establishes the new SOTA performance on the challenging AgeDB-30 benchmark across three evaluated FR models (ArcFace, CurricularFace, and MagFace), while achieving second-best performance on the ElasticFace model. Furthermore, PreFIQs consistently achieves the top or second-best performance on the Adience across all four evaluated FR models.

Table 4. FR verification accuracy (%) of ResNet100 under different pruning strategies at pruning ratios ρ. The **best** and *second-best* results per dataset are highlighted. Further results are provided in the supplementary material. It is evident that unstructured pruning consistently achieves superior performance compared to structured pruning. Regarding the pruning criterion, L_1-magnitude-based pruning clearly outperforms random parameter selection by a substantial margin.

Granularity of Pruning - Comparison between unstructured and structured model pruning [↑]

Methods	LFW	CFP-FP	CFP-FF	AgeDB-30	CALFW	CPLFW	$\overline{\text{Acc}}$ [↑]
ResNet100 (unpruned)	**99.80**	*96.67*	*99.89*	98.35	*96.15*	**93.32**	97.36
Unstructured ρ=0.1	**99.80**	*96.67*	*99.89*	**98.43**	**96.17**	*93.23*	97.37
Unstructured ρ=0.3	**99.80**	96.59	*99.89*	*98.42*	96.08	93.17	97.32
Unstructured ρ=0.5	**99.80**	96.36	**99.90**	98.20	96.00	92.70	97.16
Unstructured ρ=0.7	99.75	94.87	99.79	97.50	95.85	90.38	96.36
Unstructured ρ=0.9	90.45	65.76	91.87	75.27	78.37	58.87	76.76
Structured ρ=0.1	*99.77*	95.30	99.84	97.82	95.88	91.07	96.61
Structured ρ=0.3	82.90	58.71	79.10	71.93	68.22	56.72	69.60
Structured ρ=0.5	72.47	59.39	71.50	58.35	58.92	52.77	62.23
Structured ρ=0.7	51.12	50.19	51.26	50.10	50.17	50.60	50.57
Structured ρ=0.9	50.00	50.00	50.00	50.00	50.00	50.00	50.00

Pruning Criterion - Comparison between L_1 magnitude and random model pruning [↑]

Methods	LFW	CFP-FP	CFP-FF	AgeDB-30	CALFW	CPLFW	$\overline{\text{Acc}}$ [↑]
ResNet100 (unpruned)	**99.80**	*96.67*	*99.89*	98.35	*96.15*	**93.32**	97.36
L_1 Magnitude ρ=0.1	**99.80**	*96.67*	*99.89*	**98.43**	**96.17**	*93.23*	97.37
L_1 Magnitude ρ=0.3	**99.80**	96.59	*99.89*	*98.42*	96.08	93.17	97.32
L_1 Magnitude ρ=0.5	**99.80**	96.36	**99.90**	98.20	96.00	92.70	97.16
L_1 Magnitude ρ=0.7	99.75	94.87	99.79	97.50	95.85	90.38	96.36
L_1 Magnitude ρ=0.9	90.45	65.76	91.87	75.27	78.37	58.87	76.76
Random ρ=0.1	97.43	77.86	98.03	82.82	89.18	73.52	86.47
Random ρ=0.3	71.73	59.83	73.21	56.63	57.88	55.07	62.39
Random ρ=0.5	50.00	50.00	50.00	50.00	50.00	50.00	50.00
Random ρ=0.7	50.00	50.00	50.00	50.00	50.00	50.00	50.00
Random ρ=0.9	50.00	50.00	50.00	50.00	50.00	50.00	50.00

Overall, our entirely training-free PreFIQs approach successfully outperforms several complex supervised approaches across multiple benchmarks. On the large-scale IJB-C dataset, PreFIQs yields highly competitive results, further validating the robustness and generalizability of parameter sparsification as a reliable metric for FIQA.

6. Conclusion

This paper introduced PreFIQs, a novel data-free and training-free framework for FIQA. Departing from prior approaches, PreFIQs reframes image utility as structural robustness under model sparsification. Grounded in the PIE hypothesis, we demonstrated that the representation drift between an original FR model and its pruned counterpart provides a principled and computationally efficient proxy for image quality. We provided both theoretical and empirical validation of this formulation. A first-order Taylor analysis showed that the proposed discrete embedding drift approximates the Jacobian-vector product governing geometric sensitivity of the latent identity manifold. Extensive experiments across eight benchmarks and four SOTA FR models confirmed this alignment, demonstrating that PreFIQs achieves highly competitive, and in several cases

Table 5. FIQ SOTA comparison using four FR models reported as pAUC scores (discard rate = 0.3, FMR = 10^{-3}). The **best** and *second-best* results per dataset are highlighted. The final column displays the average pAUC across all benchmarks. We exclude XQLFW from this average to prevent evaluation bias, as its quality labels were derived using SER-FIQ. The best average pAUC is highlighted in GREEN for supervised and self-supervised approaches (marked using green stripes) , and BLUE for unsupervised approaches (marked with blue stripes). Our training-free PreFIQ is among the top-performing methods.

ArcFace [11] - $pAUC * 10^3$ (FMR= 10^{-3}) [↓]

Methods	Adience	AgeDB-30	CFP-FP	LFW	CALFW	CPLFW	XQLFW	IJB-C	$\overline{pAUC}$
RankIQ [9]	14.572	10.700	8.323	*0.748*	22.522	34.567	148.080	7.898	14.190
PFE [46]	10.740	8.211	5.893	0.795	22.256	26.604	142.459	7.470	11.710
SDD-FIQA [37]	11.844	8.621	7.810	0.800	22.354	31.146	159.151	7.236	12.830
MagFace [35]	11.154	7.428	4.952	**0.680**	21.066	27.665	160.833	7.154	11.443
CR-FIQA(L) [8]	10.901	7.605	3.660	0.810	*20.937*	20.374	140.016	6.579	10.124
DifFIQA(R) [4]	11.226	9.269	3.931	0.789	21.801	*20.185*	137.822	6.482	10.526
eDifFIQA(L) [5]	10.210	*6.880*	**3.546**	0.785	21.012	**20.086**	142.316	*6.469*	9.856
CLIB-FIQA [40]	10.931	7.387	4.070	0.790	21.064	20.431	*137.399*	6.596	10.181
ViT-FIQA(T) [2]	**9.948**	8.234	*3.568*	0.771	21.771	20.531	140.465	6.563	10.198
FROQ [6]	12.463	8.297	5.550	0.835	**20.914**	22.968	140.843	**6.438**	11.066
SER-FIQ [48]	11.627	7.776	3.797	0.800	22.053	21.570	**132.368**	6.528	10.593
FaceQnet [20, 21]	15.273	8.804	9.009	1.007	23.339	50.881	183.144	8.502	16.688
GraFIQs(L) [31]	10.541	7.717	4.348	0.840	21.425	22.495	144.309	6.863	10.604
ViTNT-FIQA [42]	10.706	9.674	4.568	0.988	22.292	21.802	140.730	6.732	10.966
PreFIQs (Ours)	*10.009*	**6.876**	3.755	0.921	20.979	21.180	141.716	6.770	10.070

CurricularFace [25] - $pAUC * 10^3$ (FMR= 10^{-3}) [↓]

Methods	Adience	AgeDB-30	CFP-FP	LFW	CALFW	CPLFW	XQLFW	IJB-C	$\overline{pAUC}$
RankIQ [9]	12.521	11.441	9.088	*0.748*	21.544	31.151	132.029	7.654	13.449
PFE [46]	9.523	8.478	6.497	0.795	21.723	22.546	120.134	7.090	10.950
SDD-FIQA [37]	10.522	9.492	8.394	0.800	21.750	26.271	142.492	6.904	12.019
MagFace [35]	10.179	7.652	5.352	**0.680**	20.796	24.451	150.727	6.794	10.844
CR-FIQA(L) [8]	10.111	7.556	4.054	0.830	20.701	17.364	*119.232*	6.305	9.560
DifFIQA(R) [4]	9.806	9.949	3.772	0.789	21.065	*17.046*	124.298	6.185	9.802
eDifFIQA(L) [5]	8.996	7.576	**3.520**	0.785	20.597	**16.947**	131.594	**6.164**	9.226
CLIB-FIQA [40]	9.768	8.103	3.841	0.790	*20.489*	17.367	123.186	6.321	9.526
ViT-FIQA(T) [2]	**8.899**	8.606	3.973	0.771	21.439	17.304	124.911	6.337	9.618
FROQ [6]	10.656	8.972	7.177	0.835	**20.212**	19.619	125.059	*6.174*	10.521
SER-FIQ [48]	10.404	7.796	3.811	0.851	21.193	18.447	**117.754**	6.253	9.822
FaceQnet [20, 21]	13.608	9.627	8.580	1.007	22.762	42.412	159.222	8.081	15.154
GraFIQs(L) [31]	9.694	*7.449*	4.081	0.880	20.886	19.500	125.193	6.494	9.855
ViTNT-FIQA [42]	9.625	10.058	5.358	0.988	21.549	18.389	129.235	6.383	10.336
PreFIQs (Ours)	*8.968*	**7.020**	*3.752*	0.921	20.577	18.239	123.709	6.445	9.417

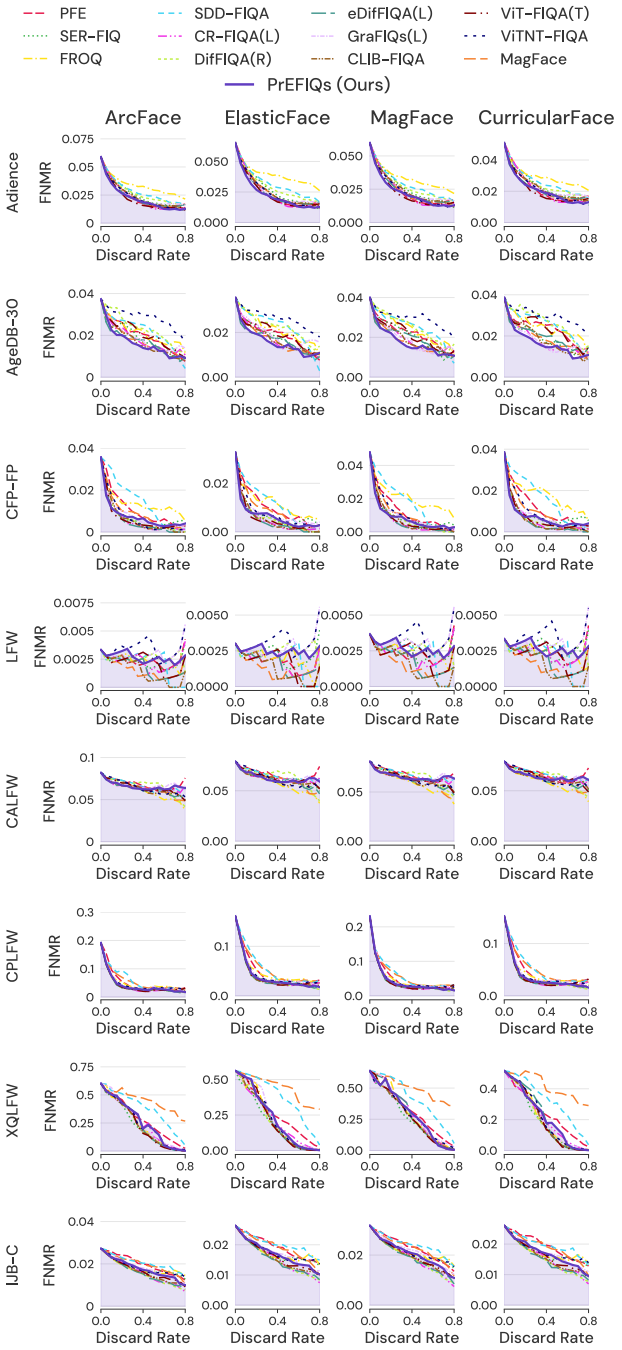
ElasticFace [7] - $pAUC * 10^3$ (FMR= 10^{-3}) [↓]

Methods	Adience	AgeDB-30	CFP-FP	LFW	CALFW	CPLFW	XQLFW	IJB-C	$\overline{pAUC}$
RankIQ [9]	16.208	10.588	7.716	*0.578*	21.469	32.959	134.681	7.737	13.894
PFE [46]	11.804	7.437	5.414	0.678	21.482	23.735	133.068	7.062	11.088
SDD-FIQA [37]	13.253	8.766	6.139	0.681	21.410	28.248	157.945	6.993	12.213
MagFace [35]	12.355	6.954	4.767	**0.564**	20.546	26.468	158.092	6.907	11.223
CR-FIQA(L) [8]	11.770	7.367	3.288	0.692	*20.193*	19.265	*124.870*	6.355	9.847
DifFIQA(R) [4]	12.572	8.553	3.460	0.685	20.990	*18.780*	127.943	*6.240*	10.183
eDifFIQA(L) [5]	*11.193*	**6.587**	**3.040**	0.681	20.246	**18.774**	135.841	**6.197**	9.531
CLIB-FIQA [40]	11.808	7.144	3.411	0.674	20.196	19.231	129.072	6.397	9.837
ViT-FIQA(T) [2]	11.228	7.607	*3.200*	0.654	20.764	19.469	135.159	6.334	9.894
FROQ [6]	13.919	8.420	5.166	0.718	**20.027**	21.892	136.545	6.242	10.912
SER-FIQ [48]	12.933	7.430	3.428	0.735	20.911	20.168	**118.090**	6.328	10.276
FaceQnet [20, 21]	16.806	8.696	8.261	0.890	22.592	44.257	171.840	8.250	15.679
GraFIQs(L) [31]	11.348	7.678	3.757	0.724	20.796	21.031	146.097	6.536	10.267
ViTNT-FIQA [42]	12.031	9.141	4.136	0.832	21.377	20.535	137.532	6.431	10.640
PreFIQs (Ours)	**10.664**	*6.600*	3.287	0.805	20.324	19.926	135.726	6.461	9.724

MagFace [35] - $pAUC * 10^3$ (FMR= 10^{-3}) [↓]

Methods	Adience	AgeDB-30	CFP-FP	LFW	CALFW	CPLFW	XQLFW	IJB-C	$\overline{pAUC}$
RankIQ [9]	14.574	11.894	11.164	0.812	22.222	37.772	162.060	9.217	15.379
PFE [46]	10.996	8.598	7.291	0.804	22.160	27.130	158.349	8.461	12.206
SDD-FIQA [37]	12.131	9.542	9.532	0.826	22.149	31.445	180.623	8.414	13.434
MagFace [35]	11.258	*7.504*	6.264	**0.706**	21.071	29.659	176.288	8.223	12.098
CR-FIQA(L) [8]	11.321	8.140	4.993	0.818	21.029	22.242	151.180	7.759	10.900
DifFIQA(R) [4]	11.482	9.818	5.447	0.815	21.687	22.351	151.131	7.603	11.315
eDifFIQA(L) [5]	10.614	7.699	**4.756**	0.810	21.085	**22.186**	161.214	**7.554**	10.672
CLIB-FIQA [40]	11.301	8.128	5.317	0.799	*20.967*	22.665	*149.878*	7.715	10.985
ViT-FIQA(T) [2]	*10.184*	8.644	4.926	*0.779*	21.611	*22.228*	151.108	7.686	10.865
FROQ [6]	12.556	9.524	8.486	0.843	**20.786**	24.584	159.701	*7.579*	12.051
SER-FIQ [48]	12.104	8.696	4.918	0.877	21.650	23.579	**144.182**	7.660	11.355
FaceQnet [20, 21]	15.601	9.546	11.081	1.015	22.995	75.482	190.321	9.644	20.766
GraFIQs(L) [31]	10.985	8.041	5.454	0.921	21.267	24.745	160.852	8.024	11.348
ViTNT-FIQA [42]	11.043	9.897	6.528	0.974	22.024	23.590	151.360	7.782	11.691
PreFIQs (Ours)	**10.146**	**7.432**	*4.805*	0.947	20.981	23.261	154.940	7.838	10.773

Figure 3. EDC (FNMR at FMR=$1e^{-3}$) of PreFIQs and recent FIQA approaches. The results are shown for four FR models on eight benchmarks. Unsupervised approaches are visualized using dotted lines. Supervised methods are visualized with dashed lines. PreFIQs is visualized using a continuous line with shaded AUC.



SOTA, performance, particularly on challenging benchmarks such as AgeDB-30 and Adience. Beyond its empirical effectiveness, PreFIQs offers a conceptual shift in FIQA: rather than predicting quality through learned regression or stochastic robustness estimation, it directly measures how well identity information survives controlled capacity reduction. This perspective establishes parameter sparsification as a probe of sample utility. Ultimately, our results support a simple but powerful principle: face image quality is what survives pruning.

References

[1] Fernando Alonso-Fernandez, Kevin Hernandez-Diaz, Jose Maria Buades Rubio, Prayag Tiwari, and Josef Bigun. Deep network pruning: A comparative study on cnns in face recognition. *Pattern Recognition Letters*, 189:221–228, 2025. 2

[2] Andrea Atzori, Fadi Boutros, and Naser Damer. Vit-fiqa: Assessing face image quality using vision transformers. In *2025 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2025. 1, 2, 5, 6, 8

[3] Ziga Babnik, Peter Peer, and Vitomir Struc. Faceqan: Face image quality assessment through adversarial noise exploration. In *2022 26th International Conference on Pattern Recognition (ICPR)*, pages 748–754, 2022. 1, 2, 6

[4] Žiga Babnik, Peter Peer, and Vitomir Štruc. Diffiqa: Face image quality assessment using denoising diffusion probabilistic models. In *2023 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10, 2023. 1, 2, 5, 8

[5] Žiga Babnik, Peter Peer, and Vitomir Štruc. eDifFIQA: Towards Efficient Face Image Quality Assessment based on Denoising Diffusion Probabilistic Models. *IEEE Transactions on Biometrics, Behavior, and Identity Science (TBIOM)*, 2024. 2, 4, 5, 6, 8

[6] Žiga Babnik, Deepak Kumar Jain, Peter Peer, and Vitomir Štruc. FROQ: Observing Face Recognition Models for Efficient Quality Assessment. 2025. 1, 2, 5, 6, 8

[7] Fadi Boutros, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Elasticface: Elastic margin loss for deep face recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2022, New Orleans, LA, USA, June 19-20, 2022*, pages 1577–1586. IEEE, 2022. 5, 8

[8] Fadi Boutros, Meiling Fang, Marcel Klemt, Biying Fu, and Naser Damer. CR-FIQA: face image quality assessment by learning sample relative classifiability. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*, pages 5836–5845. IEEE, 2023. 1, 2, 4, 5, 6, 8

[9] Jiansheng Chen, Yu Deng, Gaocheng Bai, and Guangda Su. Face image quality assessment based on learning to rank. *IEEE Signal Process. Lett.*, 22(1):90–94, 2015. 1, 2, 5, 8

[10] Wei-Ting Chen, Gurunandan Krishnan, Qiang Gao, Sy-Yen Kuo, Sizhuo Ma, and Jian Wang. Dsl-fiqa: Assessing facial image quality via dual-set degradation learning and landmark-guided transformer. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2931–2941, 2024. 2

[11] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 4690–4699. Computer Vision Foundation / IEEE, 2019. 1, 3, 5, 8

[12] Eran Eidinger, Roee Enbar, and Tal Hassner. Age and gender estimation of unfiltered faces. *IEEE Trans. Inf. Forensics Secur.*, 9(12):2170–2179, 2014. 5

[13] Gongfan Fang, Xinyin Ma, Mingli Song, Michael Bi Mi, and Xinchao Wang. Depgraph: Towards any structural pruning. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16091–16101, 2023. 5

[14] Jonathan Frankle, Gintare Karolina Dziugaite, Daniel M. Roy, and Michael Carbin. Pruning neural networks at initialization: Why are we missing the mark? In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. 3

[15] Biying Fu, Cong Chen, Olaf Henniger, and Naser Damer. A deep insight into measuring face image utility with general and face-specific image quality metrics. In *IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2022, Waikoloa, HI, USA, January 3-8, 2022*, pages 1121–1130. IEEE, 2022. 1

[16] P. Grother and E. Tabassi. Performance of biometric quality measures. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29(4):531–543, 2007. 5

[17] P. Grother, M. Ngan A. Hom, and K. Hanaoka. Ongoing face recognition vendor test (frvt) part 5: Face image quality assessment (4th draft). In *National Institute of Standards and Technology*. Tech. Rep., Sep. 2021. 5

[18] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III*, pages 87–102. Springer, 2016. 5

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778. IEEE Computer Society, 2016. 5

[20] Javier Hernandez-Ortega, Javier Galbally, Julian Fiérrez, Rudolf Haraksim, and Laurent Beslay. Faceqnet: Quality assessment for face recognition based on deep learning. In *2019 International Conference on Biometrics, ICB 2019, Crete, Greece, June 4-7, 2019*, pages 1–8. IEEE, 2019. 5, 8

[21] Javier Hernandez-Ortega, Javier Galbally, Julian Fiérrez, and Laurent Beslay. Biometric quality: Review and application to face recognition with faceqnet. *CoRR*, abs/2006.03298, 2020. 1, 2, 5, 8

[22] Torsten Hoefler, Dan Alistarh, Tal Ben-Nun, Nikoli Dryden, and Alexandra Peste. Sparsity in deep learning: pruning and growth for efficient inference and training in neural networks. *J. Mach. Learn. Res.*, 22(1), 2021. 2

[23] Sara Hooker, Aaron C. Courville, Gregory Clark, Yann Dauphin, and Andrea Frome. What do compressed deep neural networks forget. *arXiv: Learning*, 2019. 2, 3

[24] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, 2007. 5

[25] Yuge Huang, Yuhan Wang, Ying Tai, Xiaoming Liu, Pengcheng Shen, Shaoxin Li, Jilin Li, and Feiyue Huang.

Curricularface: Adaptive curriculum learning loss for deep face recognition. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pages 5900–5909. Computer Vision Foundation / IEEE, 2020. 5, 8

[26] ISO/IEC JTC1 SC37 Biometrics. ISO/IEC TR 29794-5:2010 Information technology - Biometric sample quality - Part 5: Face image data. International Organization for Standardization, 2010. 1

[27] ISO/IEC JTC1 SC37 Biometrics. ISO/IEC 19795-1:2021 Information technology — Biometric performance testing and reporting — Part 1: Principles and framework. International Organization for Standardization, 2021. 5

[28] Zhiyu Jiang, Zhe Liu, Chen Sun, Yantao Shen, Xiaohua Xue, Hongyuan Zha, and Zhiwu Huang. Self-damaging contrastive learning. In *International Conference on Machine Learning (ICML)*, 2021. 3

[29] Minchul Kim, Anil K. Jain, and Xiaoming Liu. Adaface: Quality adaptive margin for face recognition. In *CVPR*, pages 18729–18738. IEEE, 2022. 1

[30] Martin Knoche, Stefan Hörmann, and Gerhard Rigoll. Cross-quality LFW: A database for analyzing cross- resolution image face recognition in unconstrained environments. In *16th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2021, Jodhpur, India, December 15-18, 2021*, pages 1–5. IEEE, 2021. 5

[31] Jan Niklas Kolf, Naser Damer, and Fadi Boutros. Grafiqs: Face image quality assessment using gradient magnitudes. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1490–1499, 2024. 1, 2, 4, 5, 6, 8

[32] Tailin Liang, John Glossner, Lei Wang, Shaobo Shi, and Xiaotong Zhang. Pruning and quantization for deep neural network acceleration: A survey. *Neurocomputing*, 461:370–403, 2021. 2

[33] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. SphereFace: Deep hypersphere embedding for face recognition. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 212–220, 2017. 3

[34] Brianna Maze, Jocelyn C. Adams, James A. Duncan, Nathan D. Kalka, Tim Miller, Charles Otto, Anil K. Jain, W. Tyler Niggel, Janet Anderson, Jordan Cheney, and Patrick Grother. IARPA janus benchmark - C: face dataset and protocol. In *2018 International Conference on Biometrics, ICB 2018, Gold Coast, Australia, February 20-23, 2018*, pages 158–165. IEEE, 2018. 5

[35] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. Magface: A universal representation for face recognition and quality assessment. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, pages 14225–14234. Computer Vision Foundation / IEEE, 2021. 1, 2, 5, 6, 8

[36] Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. Agedb: The first manually collected, in-the-wild age database. In *2017 IEEE CVPRW, CVPR Workshops 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 1997–2005. IEEE Computer Society, 2017. 5

[37] 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 *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, pages 7670–7679. Computer Vision Foundation / IEEE, 2021. 1, 2, 5, 6, 8

[38] Fu-Zhao Ou, Chongyi Li, Shiqi Wang, and Sam Kwong. MR-FIQA: face image quality assessment with multi-reference representations from synthetic data generation. In *IEEE/CVF International Conference on Computer Vision, ICCV 2025, Honolulu, Hawaii, USA, October 19-23, 2025*, pages 12915–12925. Computer Vision Foundation / IEEE, 2025. 2, 4

[39] Fu-Zhao Ou, Chongyi Li, Shiqi Wang, and Sam Kwong. Clib-fiqa: Face image quality assessment with confidence calibration. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1694–1704, 2024. 2

[40] Fu-Zhao Ou, Chongyi Li, Shiqi Wang, and Sam Kwong. Clib-fiqa: Face image quality assessment with confidence calibration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1694–1704, 2024. 1, 5, 6, 8

[41] Guray Ozgur, Eduarda Caldeira, Tahar Chettaoui, Jan Niklas Kolf, Marco Huber, Naser Damer, and Fadi Boutros. Vitnt-fiqa: Training-free face image quality assessment with vision transformers, 2026. 2

[42] Guray Ozgur, Eduarda Caldeira, Tahar Chettaoui, Jan Niklas Kolf, Marco Huber, Naser Damer, and Fadi Boutros. Vitnt-fiqa: Training-free face image quality assessment with vision transformers. *CoRR*, abs/2601.05741, 2026. 1, 5, 6, 8

[43] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. 5

[44] Torsten Schlett, Christian Rathgeb, Juan E. Tapia, and Christoph Busch. Considerations on the evaluation of biometric quality assessment algorithms. *IEEE Trans. Biom. Behav. Identity Sci.*, 6(1):54–67, 2024. 5

[45] Soumyadip Sengupta, Jun-Cheng Chen, Carlos Domingo Castillo, Vishal M. Patel, Rama Chellappa, and David W. Jacobs. Frontal to profile face verification in the wild. In *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016, Lake Placid, NY, USA, March 7-10, 2016*, pages 1–9. IEEE Computer Society, 2016. 5

[46] Yichun Shi and Anil K. Jain. Probabilistic face embeddings. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pages 6901–6910. IEEE, 2019. 1, 2, 5, 8

[47] Hidenori Tanaka, Daniel Kunin, Daniel L. K. Yamins, and Surya Ganguli. Pruning neural networks without any data by iteratively conserving synaptic flow. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2020. Curran Associates Inc. 3

[48] Philipp Terhörst, Jan Niklas Kolf, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. SER-FIQ: unsupervised estimation of face image quality based on stochastic embedding robustness. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pages 5650–5659. Computer Vision Foundation / IEEE, 2020. 1, 2, 5, 6, 8

[49] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *CVPR*, pages 5265–5274. Computer Vision Foundation / IEEE Computer Society, 2018. 1, 3

[50] Weidi Xie, Jeffrey Byrne, and Andrew Zisserman. Inducing predictive uncertainty estimation for face verification. In *31st British Machine Vision Conference 2020, BMVC 2020, Virtual Event, UK, September 7-10, 2020*. BMVA Press, 2020. 6

[51] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z. Li. Learning face representation from scratch. *CoRR*, abs/1411.7923, 2014. 5

[52] Jie Zhao, Yuxiang Xiong, Jian Cheng, Jianshu Li, Yao Zhao, Jian Xing, Shuicheng Yan, and Jiashi Feng. Towards pose invariant face recognition in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1

[53] T. Zheng and W. Deng. Cross-pose lfw: A database for studying cross-pose face recognition in unconstrained environments. Technical Report 18-01, Beijing University of Posts and Telecommunications, 2018. 5

[54] Tianyue Zheng, Weihong Deng, and Jiani Hu. Cross-age LFW: A database for studying cross-age face recognition in unconstrained environments. *CoRR*, abs/1708.08197, 2017. 5