2022 26th International Conference on Pattern Recognition (ICPR)
August 21-25, 2022, Montréal, Québec, Canada

FaceQAN: Face Image Quality Assessment Through Adversarial Noise Exploration

Žiga Babnik*, Peter Peer† and Vitomir Štruc*
*University of Ljubljana, Faculty of Electrical Engineering Tržaška cesta 25, 1000 Ljubljana, Slovenia
†University of Ljubljana, Faculty of Computer and Information Science, Večna Pot 113, 1000 Ljubljana, Slovenia
E-mail: {ziga.babnik, vitomir.struc}@fe.uni-lj.si, peter.peer@fri.uni-lj.si

2022 26th International Conference on Pattern Recognition (ICPR) | 978-1-6654-9062-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICPR56361.2022.9956444

***Abstract*—Recent state-of-the-art face recognition (FR) approaches have achieved impressive performance, yet unconstrained face recognition still represents an open problem. Face image quality assessment (FIQA) approaches aim to estimate the quality of the input samples that can help provide information on the confidence of the recognition decision and eventually lead to improved results in challenging scenarios. While much progress has been made in face image quality assessment in recent years, computing reliable quality scores for diverse facial images and FR models remains challenging. In this paper, we propose a novel approach to face image quality assessment, called FaceQAN, that is based on adversarial examples and relies on the analysis of adversarial noise which can be calculated with any FR model learned by using some form of gradient descent. As such, the proposed approach is the first to link image quality to adversarial attacks. Comprehensive (cross-model as well as model-specific) experiments are conducted with four benchmark datasets, i.e., LFW, CFP–FP, XQLFW and IJB–C, four FR models, i.e., CosFace, ArcFace, CurricularFace and ElasticFace, and in comparison to seven state-of-the-art FIQA methods to demonstrate the performance of FaceQAN. Experimental results show that FaceQAN achieves competitive results, while exhibiting several desirable characteristics. The source code for FaceQAN is available at https://github.com/LSIbabnikz/FaceQAN.**

I. Introduction

In recent years, Face Recognition (FR) techniques have achieved excellent results on datasets containing both high quality frontal images and images of variable quality [1], [2], [3]. Face recognition in completely unconstrained settings, on the other hand, remains challenging, as no quality guarantees can be made for facial images captured in-the-wild [4], [5], [6]. Face Image Quality Assessment (FIQA) methods aim to improve the face recognition performance in such settings, by providing additional information to the FR models corresponding to the quality of the input face images. Based on this information, a FR model can reject low quality images, which often times cause critical False Non-Match (FNM) errors [7].

According to ISO/IEC 29794-1, the quality of a biometric sample can be defined using its character, fidelity or utility [8]. Contemporary FIQA methods typically rely on the latter and most often define quality in terms of the utility of the input face sample for the FR task. Here, the term *utility* is used to describe the fitness of a sample to accomplish or fulfill a biometric function [7]. In this setting, a single numerical score is commonly calculated to capture the quality of the given face image [7]. From a conceptual view point, contemporary state-of-the-art FIQA techniques can be broadly partitioned into two distinct groups: i.e., (i) regression-based methods and (ii) model-based approaches. Techniques from the first group learn a mapping directly from the image space to automatically-generated pseudo ground-truth quality labels. Several labeling approaches have been introduced over the years to facilitate this process, the basis for which are comparison/similarity scores calculated using a selected FR model. Recent techniques, for example, consider either comparison-score distributions generated over mated image pairs from a given dataset as the basis for the reference quality labels [9], [10], or alternatively, the distribution of similarity scores between probe samples and mated reference samples of a known quality [11]. Techniques from the second group, on the other hand, integrate the quality estimation process directly with the considered FR model and commonly predict the quality of the input samples based on the uncertainty of the computed face representations (features/embeddings) [12], [13], [14]. Such techniques are tightly linked to the given FR model and usually designed without the need for supervised learning. While the quality scores produced by existing approaches from either group have been shown to be good predictors of the utility of the input images for face recognition, they are still limited by the validity of the pseudo ground truth labels and poor integration with the targeted FR model.

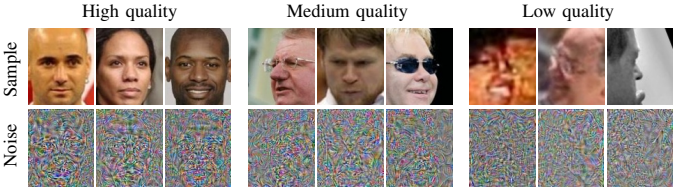


Fig. 1. **Visualization of adversarial noise as a function of image quality.** A clear difference can be observed in the characteristics of the adversarial noise between images of different quality. Higher quality images produce noise masks with a more distinct face-like pattern, while the lower quality masks are closer to random noise. FaceQAN takes advantage of the presented observations to estimate the quality of the input images.

In this paper, we propose a novel, unsupervised FIQA approach called FaceQAN (Face Image Quality Assessment Through Adversarial Noise Exploration) that avoids the pseudo quality label generation process and, consequently, the supervised learning process, by using information available in the input facial image and a given (targeted) FR model only. The basis for FaceQAN are adversarial examples, which can be generated for all modern FR models learned through some form of gradient descent [15]. The proposed method works under the assumption that the difficulty of adversarial example generation is directly correlated with the quality of a given image w.r.t. a certain FR model, as shown on Fig 1. In other words, good quality images are expected to produce

stable and robust representations that are difficult to perturb using adversarial noise. Based on this insight, we make the following main contributions in this paper:

- We propose FaceQAN, a novel approach for generating quality scores for face images that ensures competitive results on several datasets and in comparison to a wide range of recent state-of-the-art FIQA techniques.
- To the best of our knowledge, we are the first to link adversarial attacks to the task of face image quality assessment and show that such linkage leads to highly desirable FIQA characteristics.

II. Related Work

In this section, we position FaceQAN with respect to the two main groups of existing methods, i.e., regression- and model-based approaches. A more comprehensive coverage of the field can be found in the recent survey paper in [7].

Regression-Based FIQA. Techniques from this group utilize pseudo quality (ground truth) labels for the prediction of quality scores. The ground-truth label-generation process is usually automatic [11], [16], [10] but can also involve human assessment [17], [18]. The most recent methods in the literature utilize deep learning models paired with automatic extraction of ground truth labels. One such method, called *FaceQNet*, presented by Hernandez-Ortega *et al.* in [11], is based on a ResNet-50 model trained on pseudo labels generated from the VGGFace2 [19] dataset. The ground truth labels are obtained by first selecting the highest quality image of the individuals in the set, calculated using third party ISO/IEC 19794-5 [20], [21] compliance software. For all other images in the set, the quality is calculated as the normalized Euclidean distance between the embeddings of the given image and of the highest quality image of a particular individual. A pretrained ResNet-50 network is then fine-tuned on the generated ground truth labels. A more advanced labeling process is presented by Ou *et al.* [16] in the *SDD-FIQA* method. Here, the authors suggest using both the inter-class and intra-class distances to produce ground truth labels. For a given image, both mated and non-mated comparison-score distributions are constructed first. The quality score is then computed as the average Wasserstein distance between the distributions over several runs. A face recognition model is finally fine-tuned on the obtained ground truth labels, similar to FaceQNet.

While the discussed approaches perform well on standard datasets such as LFW [22], they heavily rely on the construction of ground truth quality labels. This process often introduces biases associated with either the utilized FR model or the selected dataset. Since the end goal is to improve the face recognition performance of a particular FR model, with its own associated biases, such FIQA methods may not produce optimal quality scores for different datasets and FR models. To avoid such shortcomings, the proposed FaceQAN quality estimation approach relies only on the information contained in the given image and a targeted FR model, avoiding any learning and, consequently, the process of generating pseudo ground truth labels. In cases where the targeted FR model is not available, the FaceQAN quality assessment process can also be performed with an arbitrary surrogate FR model.

Model-Based FIQA. Solutions in this group typically combine the face recognition and FIQA tasks, and learn to produce the embeddings and the quality (or uncertainty) of the input image simultaneously. One such method, *PFE* (Probabilistic Face Embeddings) presented by Shi and Jain [12], learns to predict the uncertainty of a given image, by estimating the mean and variance of the computed feature vectors. Here, the mean vector corresponds to the embedding of the given image, whereas the variance presents the uncertainty of the image in the feature space. A quality score is obtained by calculating the harmonic mean over the variance vector. More recently Meng *et al.* [13] presented *MagFace* an approach that incorporates the prediction of the image uncertainty into the face-recognition learning process. The authors achieved this by introducing a new loss function built upon the ArcFace loss, which enables better separation for embeddings with higher magnitudes. The quality score of a given image is calculated simply as the norm of its embeddings, if the embedding model is trained using the presented loss. While such methods do not need to explicitly extract quality ground truth labels, they enforce a learning regime upon the targeted FR model. The proposed FaceQAN method, on the other hand, is *learning-free* and applicable directly to any (pretrained) targeted model [15] without the need for interventions into the learning procedure.

Our approach is most closely related to *SER-FIQ*, presented by Terhörst *et al.* in [14]. SER-FIQ uses the properties of *dropout*, a regularization technique commonly used in modern convolutional neural networks (CNNs) to avoid overfitting. More precisely, several embeddings are first produced for a single input image using different sub-network layouts, generated using dropout. The quality score is then computed as the sigmoid of the negative normalized sum of distances over all pairs of produced embeddings. Similarly to SER-FIQ, FaceQAN also aims to capture the uncertainty of the embeddings (computed for an input face image) for quality assessment, but does so through the analysis of a set of generated adversarial examples.

III. Methodology

An ideal FIQA method should reflect the biases and performance of the targeted FR model, while being generally applicable to different model topologies trained with arbitrary learning objectives. Thus, only information originating from the targeted FR model and the given input sample should be considered in the quality estimation process. The main contribution of this work, FaceQAN, presented in this section, follows the outlined logic by estimating the sample quality using adversarial noise, which can be generated for all modern FR networks, trained using some variant of gradient descent [15], as well as information available directly in the input image. Different from most adversarial methods, our goal is not to create a specific adversarial example, but rather to measure the difficulty of adversarial example generation.

A. Overview of FaceQAN

Given a targeted FR model M and an input face image $I \in \mathbb{R}^{w \times h \times 3}$, the goal of face image quality assessment is to generate a quality score $Q(I) \in \mathbb{R}$ that reflects the utility of I with respect to M. With FaceQAN, the quality score for

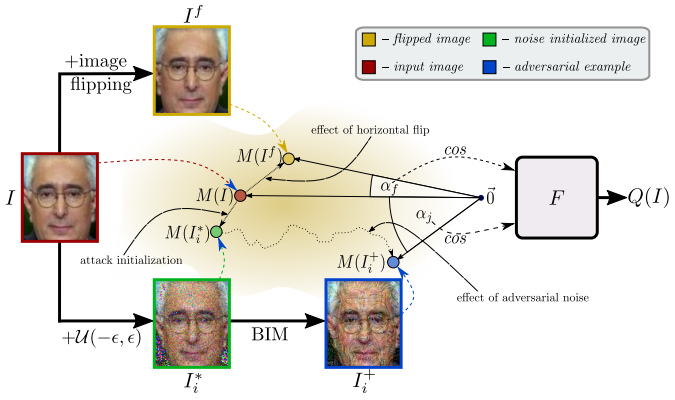


Fig. 2. **High-level overview of FaceQAN.** FaceQAN estimates the quality of an input sample I by exploring the characteristics of adversarial examples I^+ in the embedding space of the targeted FR model M. The final quality is estimated through an aggregation function F that considers the similarity between the embedding of the input sample and k adversarial examples, i.e., $i = 1, \ldots, k$. Moreover, the impact of pose on quality is modeled explicitly within FaceQAN through a face-symmetry estimation procedure based on image flipping. The figure is best viewed electronically and in color.

an input image I is computed by analyzing the characteristics of adversarial examples I^+ in the embedding space of the targeted FR models, i.e., $Q(M(I), M(I^+))$. While the pixel space could also be considered for the exploration of the adversarial noise, using the embedding space ensures tighter integration with the targeted FR models and allows FaceQAN to better capture the utility of the input face samples for FR.

The complete FaceQAN pipeline, illustrated in Fig. 2, consists of four distinct steps: (i) an *attack initialization* step (§III-B) that generates multiple noisy images from the given input face to facilitate the attack procedure, (ii) an *adversarial-example generation* step (§III-C) that produces the attack samples needed for the analysis, (iii) a *symmetry estimation* step (§III-D) that specifically accounts for the impact of pose on image quality, and (iv) a *quality-score calculation* step (§III-E) that computes the final quality value based on the analysis of the generated adversarial embeddings. All steps are described in-detail in the following sections.

B. Attack Initialization

Standard adversarial attacks typically require an input sample I, a FR model M and the ground truth class/identity label y of the input sample to be able to generate adversarial examples [23], [24], [25]. Because face image quality assessment needs to be applicable to arbitrary facial images irrespective of the identities/labels used to learn the model M, we utilize the similarity-based adversarial attack (SAA) criterion, proposed recently by Wang *et al.* in [26], as the basis for FaceQAN and adapt it to fit our problem. SAA allows us to use the embedding of the input image I as the ground truth label to attack, i.e., $y = M(I)$, and to define an (angular) dissimilarity loss to drive the adversarial noise generation process, i.e.:

$$L(M(I^*), y) = 1 - \frac{M(I^*)^T \cdot y}{\|M(I^*)\|\|y\|}, \tag{1}$$

where $\|\cdot\|$ is the L_2 norm. Since adversarial examples are generated by maximizing L, we ensure that the gradient of L with respect to the input I^* is non-zero, and, therefore, define I^* as a noise perturbed version of the original input image I. As long as the noise infused into I is minute, the embedding of I^* is expected to be close to the embedding of the original input image I, as also illustrated in Fig. 2.

Let $I \in \mathbb{R}^{w \times h \times 3}$ be an input face image with values in the range of $[-1, 1]$. FaceQAN generates the noisy counterpart to I required for the attack initialization as:

$$I^* = \lfloor I + N \rfloor_{[-1,1]}, \quad N \sim \mathcal{U}(-\epsilon, \epsilon), \tag{2}$$

where the noise N is sampled from a uniform distribution $\mathcal{U}$, ϵ is an open hyperparameter, and $\lfloor \cdot \rfloor_{[-1,1]}$ denotes a clipping operator that guarantees that the resulting noisy image I^* contains pixels in the range $[-1, 1]$. Choosing a value of ϵ close to 0 assures that the amount of noise added to the image is minuscule and that the change in the embedding is limited.

Because the added noise initializes the adversarial attack in an arbitrary direction within the embedding space, we minimize this randomness by generating a set of k noisy images, i.e., $\{I_i^*\}_{i=1}^k$, by sampling the noise independently from $\mathcal{U}$ k times. This procedure allows FaceQAN to explore the embedding space around $M(I)$ in various directions and estimate the stability of the embedding for quality estimation.

C. Adversarial Example Generation

In the second step, FaceQAN uses the set of noisy images $\{I_i^*\}_{i=1}^k$ to generate a corresponding set of adversarial examples $\{I_i^+\}_{i=1}^k$. We note again, that the use of adversarial methods in FaceQAN deviates from the usual, as we are not interested in generating adversarial examples that can fool a targeted FR model M, but rather in the characteristics of the generated examples after an attack with *fixed attack hyperparameters*. FaceQAN is in general applicable with any adversarial approach, but we select the Fast Gradient Sign Method (FGSM) [23] together with the Basic Iterative Method (BIM) [27] for the implementation in this paper due to their simplicity and ease of implementation.

Fast Gradient Sign Method. In our setting, FGSM generates an adversarial example I^+ from I such that the difference between the true label $y = M(I)$ and the embedding of the adversarial example $M(I^+)$ is maximized. The procedure starts by first generating the embedding of the input image $M(I)$ and then calculating the loss from Eq. (1) between the generated prediction and the true label, i.e., $L(M(I), y)$. The calculated loss is then back-propagated to the input I, for which the gradient $\nabla_I L(M(I), y)$ is computed. Finally, the adversarial example is constructed by defining the adversarial noise μ as the sign function of the gradient, i.e., $sign(\nabla_I L(M(I), y)$, and in turn: $I^+ = I + \epsilon \cdot \mu$, where ϵ is again an open hyperparameter of the method, that controls the amount of noise added to I. A smaller ϵ corresponds to a lower amount of adversarial noise in the final adversarial example I^+.

Basic Iterative Method. We use BIM to improve the attack capabilities of FGSM and to better explore the embedding space around $M(I)$. BIM extends FGSM by creating adversarial examples over l FGSM runs, where only the input I changes between iterations. Additionally, the parameter ϵ is scaled as $\frac{\epsilon}{l}$, to limit the overall amount of noise added to the image. The initial iteration of BIM is identical to FGSM, whereas for all consequent iterations $m \in [2, l]$ the input I

is defined as the adversarial example I^+_{m-1} produced by the previous iteration. Performing l iterations, we obtain the final adversarial example, i.e.: $I^+ \leftarrow BIM_l(I, y, M)$. The set of adversarial examples $\{I^+_i\}^k_{i=1}$ is generated from the set of noisy images using simple batch processing.

D. Symmetry Estimation

A critical factor known to affect face recognition performance are pose variations [28], [29], [30]. For FaceQAN, we, therefore, design an additional step to explicitly account for the effect of pose in the input images. To avoid unnecessary complexity, we use the following simple logic. Generally faces are vertically symmetrical and as such, if presented with a fully frontal face image, flipping the image horizontally should not have major effects on the produced embedding. This does not hold true for larger pose variations. Thus, we create a horizontally flipped image I^f for each given input I and use its embedding alongside the embeddings of the adversarial set $\{I^+_i\}^k_{i=1}$ when computing the final quality score.

E. Quality Score Calculation

In the last step, the adversarial set $\{I^+_i\}^k_{i=1}$ and the flipped image I^f are passed through the model M to obtain a set of adversarial embeddings $\{y^+_i\}^k_{i=1} = \{\, M(I^+_i) \mid i \in [1,k] \,\}$ and flipped embedding $y_f = M(I^f)$. Both are compared to the embedding $y = M(I)$, using the cosine similarity, which well matches the learning objectives of modern FR models:

$$\cos\alpha_i = S_i = \frac{y^T \cdot y^+_i}{\|y\|\|y^+_i\|}, \quad \cos\alpha_f = s_f = \frac{y^T \cdot y_f}{\|y\|\|y_f\|}, \tag{3}$$

where the set $\{S_i\}^k_{i=1}$ contains similarities for each adversarial example in $\{I^+_i\}^k_{i=1}$, and α_i and α_f are the angles between the embeddings - see Fig. 2. The set of similarities $\{S_i\}^k_{i=1}$ and s_f are then used to calculate the final quality score using the aggregation function $F : \{\{S_i\}^k_{i=1}, s_f\} \mapsto Q$, defined by:

$$q_{adv} = \frac{\mu_S + 1}{2} \cdot \lfloor (1 - \sigma_S) \rfloor_{[0,1]}, \; Q = (\; q_{adv} \cdot s_f)^p \tag{4}$$

where μ_S represents the mean over $\{S_i\}^k_{i=1}$ and σ_S is the corresponding standard deviation. Hence, F considers the mean of the similarity scores as well as their dispersion when computing the quality score, and additionally weights the estimated quality with the computed symmetry prediction. The main intuition behind the adversarial part of this procedure is that lower quality images map to a poorly defined part of the embedding space that is easily perturbed, i.e., the adversarial examples have a low average and high standard deviation.

Because the cosine similarity is bound between $[-1, 1]$, we re-scale both the mean and standard deviation to $[0, 1]$. The final quality is as such defined on $[0, 1]$, where higher values represent images of better quality with respect to M. To ensure the final quality scores $Q(I)$ cover the full range of values on $[0, 1]$, we use a power law as the last computational step in FaceQAN, where p is an open hyperparameter. Note that this final step has no impact on the quality-estimation process and only affects the range of the aggregation function F.

TABLE I
SUMMARY OF THE EXPERIMENTAL SETUP

Dataset	#Images	#IDs	#Comparisons		Main Quality Factors†‡			
			Mated	Non-mated	Pose	O-E	B-R-N	Sc
LFW [22]	13,233	5,749	3,000	3,000	L	L	L	M
CFP-FP [31]	7,000	500	3,500	3,500	H	M	L	M
XQLFW [32]	13,233	5,749	3,000	3,000	L	L	H	M
IJB-C [33]	23,124††	3,531	19,557	15,638,932	H	H	H	Lr

†O-E - Occlusion, Expression; B-R-N - Blur, Resolution, Noise; Sc - Scale.
‡L - Low; M - Medium; H - High; Lr - Large; Values estimated subjectively by the authors.
†† number of templates, each containing several images

IV. Experiments and Results

A. Experimental Setup

Evaluation Setting. We compare FaceQAN with seven state-of-the-art FIQA methods. We chose FaceQNet [11], MagFace [13], PCNet [9], LightQNet [10], and SDD-FIQA [16], as representatives of regression-based techniques, and PFE [12] and SER-FIQ [14] as examples of model-based approaches. To ensure a fair comparison, publicly available official implementations are used or code obtained directly from the authors. As summarized in Table IV, we test FaceQAN on four benchmark datasets with diverse (quality) characteristics, i.e.: Labeled Faces in the Wild (LFW) [22], Celebrities in Frontal-Profile in the Wild (CFP-FP) [31], Cross-Quality LFW (XQLFW) [32] and IJB-C (IARPA Janus Benchmark – C) [33], similarly to [34]. The experiments are conducted with three popular (state-of-the-art) open-source FR models, referred to as ArcFace[1] [35], CurricularFace[2] [36], and ElasticFace[3] [37] hereafter based on their learning objectives. All three FR models are based on the ResNet100 architecture. ArcFace is trained on MS1MV3, ElasticFace on MS1MV2, whereas CurricularFace is trained on CASIA-WebFace and MS1MV2. To enable a fair comparison with the competing state-of-the-art FIQA methods, all results are generated using a cross-model (C) setting, where quality scores are generated using one FR model and performance is measured on another. To compute quality scores for FaceQAN we use the CosFace[1] [38] model trained with the cosface objective with a ResNet100 architecture on the Glint360K dataset [39]. For reference purposes, we also provide model-specific FaceQAN results (S), where access to a target FR model is assumed and, hence, the same FR model is used for quality prediction and performance scoring.

Performance Evaluation. Following standard methodology [13], [14], [16], [7], the Error Versus Reject Characteristic (ERC) is used to score performance. ERC measures the False Non Match Rate (FNMR) at a constant False Match Rate (FMR) usually set to 0.001, when increasing the fraction of unconsidered images. Images are rejected based on their calculated quality label. Additionally, we also measure the Area Under the Curve (AUC) for the ERC plots, where lower scores imply better performance. As suggested in [7], we report AUC values at several drop rates, i.e., 0.1, 0.2, 0.4 and 0.8. Here, smaller drop rates are typically more relevant, since this is when the images of the lowest quality are removed.

Implementation Details. Since we are interested only in exploring the localized space around the computed face

[1] https://github.com/deepinsight/insightface
[2] https://github.com/HuangYG123/CurricularFace
[3] https://github.com/fdbtrs/ElasticFace

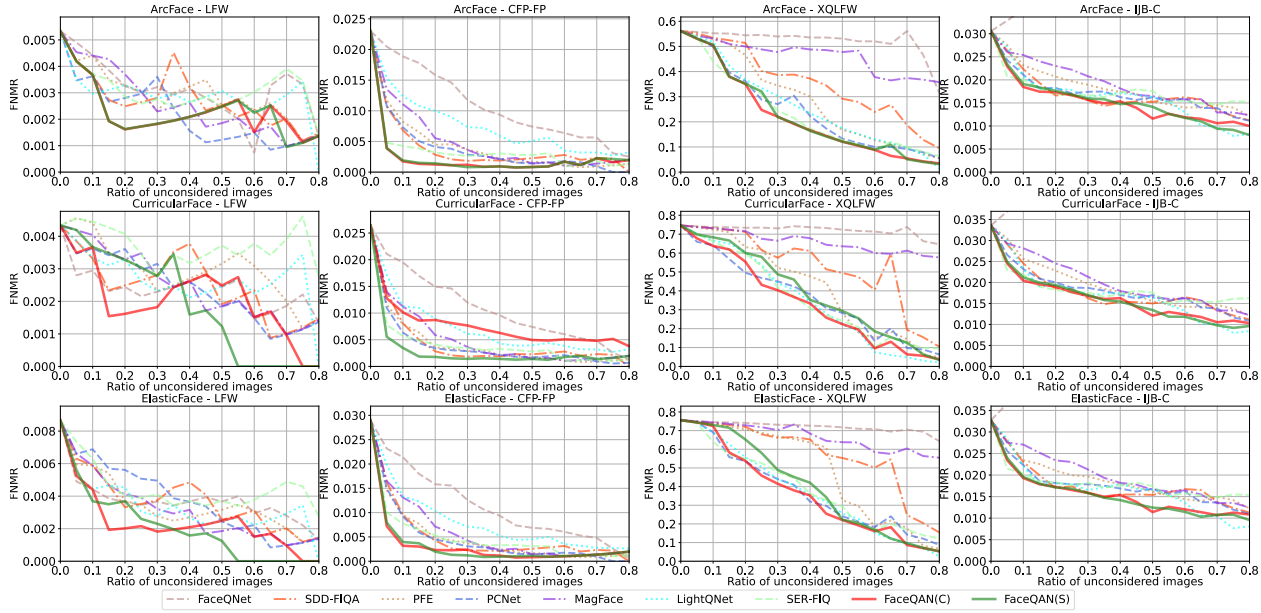


Fig. 3. **ERC results at FMR=**0.001. FaceQAN is evaluated using three state-of-the-art FR models over four datasets and in comparison to seven state-of-the-art baselines and yields highly competitive results. All results correspond to cross-model (C) quality assessment experiments. For reference purposes, model-specific FaceQAN results, marked (S), are also reported. The figure is best viewed electronically and in color.

embedding, $M(I)$, a combination of a small ϵ and low l is chosen for the experiments. Specifically, we set $\epsilon = 0.001$ and $l = 5$ as the values for the BIM parameters. Ideally, the batch size should follow $k \gg d_{M(I)}$, where $d_{M(I)}$ represents the dimensions of the embedding space, so that all directions around $M(I)$ are well explored. However, because the dimensionality of the embedding spaces is usually large (> 512), we set $k = 10$ to ensure a reasonable trade-off between the extent of the embedding space covered and the computational complexity of FaceQAN. We chose $p = 5$ since this enables the full use of values on $[0, 1]$. The benchmark images are cropped and aligned as specified per the chosen face recognition model. The same procedure is also used for the baselines unless a different preprocessing approach was proposed or included in the paper describing the baseline. All experiments are conducted on a desktop PC with an Intel i9-10900KF CPU, 64 GB of RAM and an Nvidia 3090 GPU.

B. Comparison with the State-Of-The-Art

The ERC plots for all combinations of FR models and datasets are shown in Fig. 3 and the calculated AUC scores over the ERC plots (multiplied by 10^3 for readability) in Table II. Below we analyze these results from two aspects: (i) in comparison to all considered (supervised) baselines (COMP; see Table II), and (ii) in comparison to the closely-related SER-FIQ, which does not rely on training (CLOSE).

Baseline Comparisons (COMP). FaceQAN is the most convincing approach when compared to the state-of-the-art methods from the COMP group on all four dataset when considering the ArcFace model despite not relying on supervised training or (pseudo) reference labels. Similar observations can also be made for the CurricularFace and ElasticFace models, where FaceQAN again yields highly competitive results on all four datasets in the cross-model (C) and model-specific setting (S). The only exeption is the cross-model result with CurricularFace on the CFP-FP dataset, where FaceQAN is

TABLE II
COMPARISON WITH STATE-OF-THE-ART - AUC@FMR1E-3 [$\times 10^{-3}$] (↓).
B - BEST OVERALL, - BEST IN COMP, - BEST IN CLOSE

FR Model		ArcFace – AUC@FMR1E-3 [$\times 10^{-3}$]								
DT†	DR†	Comparison Baselines (COMP)‡						Closely Related (CLOSE)‡		
		FQN	SDD	PFE	PCNet	MagFace	LQN	SER	**FQ(C)**†	**FQ(S)**†
LFW	10%	0.49	0.44	0.45	0.4	0.47	**0.38**	0.44	0.43	0.43
	20%	0.87	0.73	0.82	0.69	0.89	0.7	0.77	**0.66**	0.66
	40%	1.49	1.37	1.4	1.25	1.43	1.28	1.3	**1.03**	1.03
	80%	2.52	2.2	2.36	**1.73**	2.08	2.34	2.49	1.84	1.83
CFP-FP	10%	2.07	1.28	1.37	1.3	1.53	1.65	0.92	**0.81**	0.82
	20%	3.83	1.74	2.01	1.84	2.39	2.75	1.3	**0.96**	0.98
	40%	6.29	2.15	2.74	2.46	3.14	4.33	1.87	**1.17**	1.17
	80%	8.75	3.0	3.26	2.95	3.77	5.95	2.67	**1.7**	1.74
XQLFW	10%	55.71	54.99	54.51	53.26	54.44	53.84	**52.06**	53.19	53.24
	20%	110.68	107.41	104.84	94.59	105.72	97.17	**90.9**	93.56	93.7
	40%	219.04	188.38	175.9	152.36	203.53	157.45	142.6	**139.48**	143.3
	80%	422.55	286.99	231.52	199.0	371.23	210.83	187.38	**175.51**	181.18
IJB-C	10%	3.29	2.64	2.72	2.51	2.7	2.64	**2.3**	2.37	2.41
	20%	7.14	4.58	4.91	4.4	5.11	4.51	4.14	**4.14**	4.25
	40%	16.87	7.72	8.7	7.86	9.28	7.94	7.47	**7.32**	7.49
	80%	42.12	13.56	14.6	13.7	15.36	13.01	13.84	**12.11**	12.21

FR Model		CurricularFace – AUC@FMR1E-3 [$\times 10^{-3}$]								
DT†	DR†	Comparison Baselines (COMP)‡						Closely Related (CLOSE)‡		
		FQN	SDD	PFE	PCNet	MagFace	LQN	SER	**FQ(C)**†	**FQ(S)**†
LFW	10%	**0.32**	0.38	0.45	0.37	0.42	0.38	0.45	0.37	0.41
	20%	**0.57**	0.65	0.81	0.72	0.78	0.71	0.87	0.58	0.76
	40%	1.04	1.25	1.35	1.31	1.36	1.21	1.54	**0.99**	1.34
	80%	1.9	2.0	2.38	1.93	2.01	2.19	3.0	1.66	**1.53**
CFP-FP	10%	2.2	1.56	1.44	1.37	1.64	1.8	1.19	1.55	**1.03**
	20%	3.91	2.09	2.04	1.83	2.54	2.85	1.68	2.45	**1.25**
	40%	6.39	2.5	2.65	2.39	3.3	4.17	2.37	3.96	**1.55**
	80%	8.7	3.37	3.19	3.0	3.93	5.61	3.16	5.99	**2.17**
XQLFW	10%	74.15	73.89	73.56	**67.69**	73.85	70.45	70.22	68.48	70.66
	20%	147.75	146.09	142.97	**124.55**	146.23	134.75	132.63	129.23	136.07
	40%	294.9	269.95	252.79	213.54	282.63	227.54	221.3	**211.55**	236.28
	80%	579.14	429.54	338.31	293.0	529.29	281.82	291.27	**272.03**	316.06
IJB-C	10%	3.69	2.91	3.01	2.75	3.01	2.91	**2.53**	2.58	2.62
	20%	7.98	5.04	5.43	4.87	5.65	4.93	**4.49**	4.54	4.63
	40%	18.86	8.37	9.42	8.58	9.92	8.49	7.96	**7.95**	8.05
	80%	47.14	14.21	15.36	14.51	16.1	13.61	14.46	12.86	**12.74**

FR Model		ElasticFace – AUC@FMR1E-3 [$\times 10^{-3}$]								
DT†	DR†	Comparison Baselines (COMP)‡						Closely Related (CLOSE)‡		
		FQN	SDD	PFE	PCNet	MagFace	LQN	SER	**FQ(C)**†	**FQ(S)**†
LFW	10%	**0.57**	0.68	0.66	0.72	0.69	0.59	0.74	0.59	0.59
	20%	0.97	1.14	1.17	1.32	1.18	1.05	1.23	**0.85**	0.95
	40%	1.68	1.94	1.77	2.24	1.86	1.69	2.0	**1.24**	1.42
	80%	2.94	2.97	2.79	3.02	2.53	2.69	3.57	1.88	**1.61**
CFP-FP	10%	2.42	1.78	1.69	1.71	1.88	2.05	1.42	**1.16**	1.22
	20%	4.25	2.44	2.43	2.34	2.95	3.26	2.08	**1.45**	1.55
	40%	6.82	2.95	3.17	2.99	3.84	4.76	2.86	1.84	**1.8**
	80%	9.04	3.87	3.71	3.49	4.49	6.26	3.68	2.3	**2.28**
XQLFW	10%	75.04	74.67	74.54	73.4	75.02	73.23	**71.9**	74.21	74.36
	20%	149.38	147.68	146.64	131.99	148.6	135.74	**130.25**	134.91	144.68
	40%	295.97	282.37	280.81	**219.83**	291.61	226.24	225.59	219.84	247.56
	80%	577.87	460.28	393.29	301.86	534.96	297.76	312.28	**288.4**	320.85
IJB-C	10%	3.65	2.79	2.79	2.66	2.87	2.79	**2.4**	2.47	2.51
	20%	7.96	4.78	5.03	4.6	5.39	4.68	4.3	**4.28**	4.32
	40%	18.91	8.0	8.92	8.14	9.62	8.16	7.83	**7.46**	7.48
	80%	47.41	13.99	14.98	14.04	15.86	13.37	14.26	12.28	**12.13**

†DT – Dataset; DR – Drop Rate (or Ratio of unconsidered images). (C) – cross-model; (S) – model-specific
‡FQN – FaceQNet; SDD – SDD- FIQA; LQN – LightQNet; SER – SER-FIQ. FQ - FaceQAN (ours)

TABLE III
ANALYSIS OF THE TIME COMPLEXITY OVER CFP-FP USING COSFACE

Complexity		FaceQnet	SDD-FIQA	PFE	PCNet	MagFace	LightQnet	SER-FIQ
t [s]	μ	0.0432	0.0006	0.0493	0.0175	0.0011	0.0535	0.1125
	σ	0.0026	0.0004	0.0275	0.0004	0.0004	0.0468	0.0418

Complexity	FaceQAN (ours)				
	$k=2$	$k=5$	†$k=10$	$k=50$	$k=100$
t [s] ($\mu \pm \sigma$)	0.21 ± 0.019	0.23 ± 0.018	0.30 ± 0.006	1.00 ± 0.027	1.87 ± 0.031

† Configuration used in experiments

TABLE IV
AUC [$\times 10^{-3}$] SCORES (↓) OF THE ABLATION STUDY

FR	Dataset	FIQA Model	Image Drop Rate 10%	20%	40%	80%
ArcFace	LFW	FaceQAN	**0.43**	**0.66**	**1.03**	**1.83**
		w/o Symm. Est.	0.53	0.99	1.5	2.5
	CFP-FP	FaceQAN	**0.82**	**0.98**	**1.17**	**1.74**
		w/o Symm. Est.	0.89	1.25	1.56	2.21
Cu. Face†	LFW	FaceQAN	0.41	0.76	1.34	**1.53**
		w/o Symm. Est.	**0.35**	**0.67**	**1.31**	1.63
	CFP-FP	FaceQAN	**1.03**	**1.25**	**1.55**	**2.17**
		w/o Symm. Est.	1.06	1.4	1.75	2.48
ElasticFace	LFW	FaceQAN	**0.59**	**0.95**	**1.42**	**1.61**
		w/o Symm. Est.	0.63	1.05	1.65	2.03
	CFP-FP	FaceQAN	**1.22**	**1.55**	**1.8**	**2.28**
		w/o Symm. Est.	1.38	1.89	2.62	3.32

† Cu.Face – CurricularFace

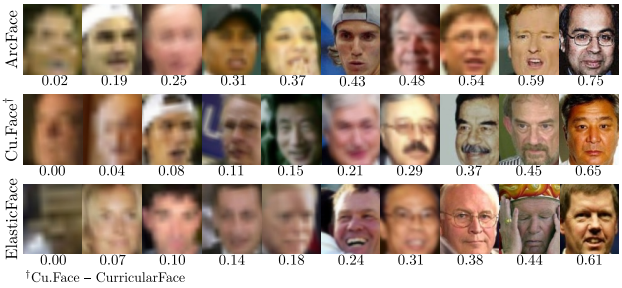
less convincing. It is also interesting to observe that the relative ranking of the competing FIQA techniques changes with different FR models. FaceQAN, on the other hand, is among the top performers with all three considered FR models.

Comparison to SER-FIQ (CLOSE). SER-FIQ and FaceQAN are the only models in our experiments that require no supervision and estimate face quality solely based on the input image and a FR model. As can be seen from Table II, the cross-model version of FaceQAN(C) is overall the top performer, suggesting that the adversarial examples generated with the CosFace model are particularly informative for quality estimation and therefore often outperform FaceQAN in the model-specific setting. SER-FIQ is most competitive on the XQLFW dataset. However, it needs to be noted that XQLFW was designed by optimizing SER-FIQ quality scores and is, therefore, highly biased towards this FIQA approach [32]. Convincing results are also achieved with SER-FIQ on the large-scale IJB-C dataset at the lowest drop rate, where the performance is comparable to FaceQAN(C).

Time Complexity. In Table III we analyze the time complexity of FaceQAN in comparison with the considered baselines and as a function of the number of generated adversarial examples k. The analysis is done over the CFP-FP dataset and with the CosFace model. As expected, the time complexity of FaceQAN increases with an increase in k and is overall somewhat higher than that of the competing solutions due to the nature of the BIM approach selected for our implementation.

C. Ablation Study

To demonstrate the impact of the symmetry estimation step on the performance of FaceQAN(S), we perform an ablation study on the LFW and CFP-FP datasets using the three selected FR models (ArcFace, CurricularFace, ElasticFace) with and without (w/o) the symmetry scores s_f from Eq. (3). As can be seen from the results in Table IV, removing the step from the quality assessment process has a detrimental effect on the AUC scores, which consistently increase in comparison to the full FaceQAN model across all tested combinations, except when using CurricularFace on the LFW dataset.



†Cu.Face – CurricularFace

Fig. 4. **Quality-based ranking with FaceQAN.** Note how the ordering of the XQLFW sample images based on the generated quality scores (shown below the images) corresponds to the perceived face quality. Zoom-in for details.

D. Qualitative Evaluation

Image Ranking. In Fig. 4 we show example images from XQLFW ordered according to the computed quality scores. As can be seen, the images follow a reasonable order in terms of perceived quality for all three FR models. Because the ordering is meant to reflect the utility of the samples for face recognition, it is interesting to see how FaceQAN(S) favors blurry frontal images in certain settings over crisp, but less frontal (and w/o neutral expressions) samples.

Quality-Score Distribution. Fig. 5 shows that the quality-score distributions generated for the ArcFace, CurricularFace and ElasticFace models exhibit a relatively similar shape, but have a somewhat different support on some datasets. Overall, however, all models generate reasonably consistent score distributions (also given the results in Fig. 4) that well capture the quality/utility of the input data.

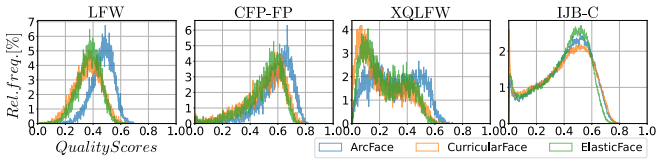


Fig. 5. **Quality-score distribution generated by FaceQAN.** Distributions are shown for (from left to right): LFW, CFP-FP, XQLFW, and IJB-C. The three FR models produce distributions on different scales but of similar shape.

V. CONCLUSION

We presented a novel approach to unsupervised face image quality assessment, called FaceQAN. The proposed approach elegantly avoids the quality-label generation and learning process used by many state-of-the-art FIQA techniques by harnessing adversarial noise, which can be generated for any modern deep learning model [15]. By comparing the embeddings of adversarial examples and the original input sample, FaceQAN is able to calculate a quality score that is an excellent predictor of the sample's utility for face recognition. Extensive experiments with several baselines, datasets and FR models have shown that FaceQAN achieves highly competitive results, while being based on minimal assumptions. As part of our future work, we plan to explore the use of adversarial examples in supervised settings and the predictive power of adversarial noise with (pseudo) reference quality labels.

ACKNOWLEDGMENTS

Supported by the ARRS Research Program P2–0250 (B) as well as the ARRS Junior Researcher Program.

REFERENCES

[1] K. Grm, V. Štruc, A. Artiges, M. Caron, and H. K. Ekenel, "Strengths and weaknesses of deep learning models for face recognition against image degradations," *IET Biometrics*, vol. 7, no. 1, pp. 81–89, 2018.

[2] K. Grm and V. Struc, "Deep face recognition for surveillance applications," *IEEE Intelligent Systems*, vol. 33, no. 3, pp. 46–50, 2018.

[3] B. Meden, P. Rot, P. Terhörst, N. Damer, A. Kuijper, W. J. Scheirer, A. Ross, P. Peer, and V. Štruc, "Privacy–enhancing face biometrics: A comprehensive survey," *IEEE Transactions on Information Forensics and Security*, 2021.

[4] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215–244, 2021.

[5] F. Boutros, N. Damer, J. N. Kolf, K. Raja, F. Kirchbuchner, R. Ramachandra, A. Kuijper, P. Fang, C. Zhang, F. Wang, D. Montero, N. Aginako, B. Sierra, M. Nieto, M. E. Erakin, U. Demir, H. K. Ekenel, A. Kataoka, K. Ichikawa, S. Kubo, J. Zhang, M. He, D. Han, S. Shan, K. Grm, V. Štruc, S. Seneviratne, N. Kasthuriarachchi, S. Rasnayaka, P. C. Neto, A. F. Sequeira, J. R. Pinto, M. Saffari, and J. S. Cardoso, "MFR 2021: Masked Face Recognition Competition," in *Proceedings of the IEEE International Joint Conference on Biometrics (IJCB)*, 2021.

[6] K. Grm, W. J. Scheirer, and V. Štruc, "Face hallucination using cascaded super-resolution and identity priors," *IEEE Transactions on Image Processing*, vol. 29, pp. 2150–2165, 2020.

[7] T. Schlett, C. Rathgeb, O. Henniger, J. Galbally, J. Fierrez, and C. Busch, "Face image quality assessment: A literature survey," *ACM Computing Surveys*, 2022.

[8] ISO/IEC JTC 1/SC 37 Biometrics, "Information Technology - Biometric Sample Quality - Part 1: Framework," International Organization for Standardization, Standard ISO/IEC 29794-1:2016, 2016.

[9] W. Xie, J. Byrne, and A. Zisserman, "Inducing predictive uncertainty estimation for face verification," in *British Machine Vision Conference (BMVC)*, 2020.

[10] K. Chen, T. Yi, and Q. Lv, "Lightqnet: Lightweight deep face quality assessment for risk-controlled face recognition," *IEEE Signal Processing Letters*, vol. 28, pp. 1878–1882, 2021.

[11] J. Hernandez-Ortega, J. Galbally, J. Fierrez, R. Haraksim, and L. Beslay, "Faceqnet: Quality assessment for face recognition based on deep learning," in *Proceedings of the IEEE International Conference on Biometrics (ICB)*, 2019, pp. 1–8.

[12] Y. Shi and A. K. Jain, "Probabilistic face embeddings," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2019, pp. 6902–6911.

[13] Q. Meng, S. Zhao, Z. Huang, and F. Zhou, "Magface: A universal representation for face recognition and quality assessment," in *CVF/IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 14 225–14 234.

[14] 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 *CVF/IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 5651–5660.

[15] N. Akhtar, A. Mian, N. Kardan, and M. Shah, "Advances in adversarial attacks and defenses in computer vision: A survey," *IEEE Access*, vol. 9, pp. 155 161–155 196, 2021.

[16] O. Fu-Zhao, X. Chen, R. Zhang, Y. Huang, S. Li, J. Li, Y. Li, L. Cao, and W. Yuan-Gen, "SDD-FIQA: Unsupervised Face Image Quality Assessment with Similarity Distribution Distance," in *CVF/IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 7670–7679.

[17] P. Wasnik, R. Ramachandra, K. Raja, and C. Busch, "An empirical evaluation of deep architectures on generalization of smartphone-based face image quality assessment," in *IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2018, pp. 1–7.

[18] L. Best-Rowden and A. K. Jain, "Learning face image quality from human assessments," *IEEE Transactions on Information forensics and security*, vol. 13, no. 12, pp. 3064–3077, 2018.

[19] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2018, pp. 67–74.

[20] D. Maltoni, A. Franco, M. Ferrara, D. Maio, and A. Nardelli, "Biolabicao: A new benchmark to evaluate applications assessing face image compliance to iso/iec 19794-5 standard," in *IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 41–44.

[21] ISO/IEC JTC 1/SC 37 Biometrics, "Information technology - Biometric Data Initerchange Formats - Part 5: Face Image Data," International Organization for Standardization, Standard ISO/IEC 19794-5:2011, 2011.

[22] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.

[23] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

[24] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," in *International Conference on Learning Representations (ICLR)*, 2018.

[25] N. Carlini and D. Wagner, "Towards evaluating the robustness of neural networks," in *IEEE Symposium on Security and Privacy (SSP)*, 2017, pp. 39–57.

[26] H. Wang, S. Wang, Z. Jin, Y. Wang, C. Chen, and T. Massimo, "Similarity-based gray-box adversarial attack against deep face recognition," in *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2021.

[27] A. Kurakin, I. J. Goodfellow, and S. Bengio, "Adversarial examples in the physical world," in *Artificial intelligence safety and security*. Chapman and Hall/CRC, 2018, pp. 99–112.

[28] V. Štruc and N. Pavešić, "The complete gabor-fisher classifier for robust face recognition," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, pp. 1–26, 2010.

[29] J. Zhao, Y. Cheng, Y. Xu, L. Xiong, J. Li, F. Zhao, K. Jayashree, S. Pranata, S. Shen, J. Xing *et al.*, "Towards pose invariant face recognition in the wild," in *CVF/IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 2207–2216.

[30] S. Banerjee, J. Brogan, J. Krizaj, A. Bharati, B. R. Webster, V. Struc, P. J. Flynn, and W. J. Scheirer, "To frontalize or not to frontalize: Do we really need elaborate pre-processing to improve face recognition?" in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2018, pp. 20–29.

[31] S. Sengupta, J. C. Cheng, C. D. Castillo, V. M. Patel, R. Chellappa, and D. W. Jacobs, "Frontal to profile face verification in the wild," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016.

[32] M. Knoche, S. Hormann, and G. Rigoll, "Cross-quality lfw: A database for analyzing cross-resolution image face recognition in unconstrained environments," in *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 2021, pp. 1–5.

[33] B. Maze, J. Adams, J. A. Duncan, N. Kalka, T. Miller, C. Otto, A. K. Jain, W. T. Niggel, J. Anderson, J. Cheney *et al.*, "IARPA Janus Benchmark-C: Face dataset and protocol," in *International Conference on Biometrics (ICB)*, 2018, pp. 158–165.

[34] F. Boutros, M. Fang, M. Klemt, B. Fu, and N. Damer, "CR-FIQA: Face Image Quality Assessment by Learning Sample Relative Classifiability," *arXiv preprint arXiv:2112.06592*, 2021.

[35] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive Angular Margin Loss for Deep Face Recognition," in *CVF/IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4690–4699.

[36] Y. Huang, Y. Wang, Y. Tai, X. Liu, P. Shen, S. Li, J. Li, and F. Huang, "CurricularFace: Adaptive curriculum learning loss for deep face recognition," in *CVF/IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 5901–5910.

[37] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "Elasticface: Elastic margin loss for deep face recognition," 2021.

[38] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, "CosFace: Large margin cosine loss for deep face recognition," in *CVF/IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 5265–5274.

[39] X. An, X. Zhu, Y. Gao, Y. Xiao, Y. Zhao, Z. Feng, L. Wu, B. Qin, M. Zhang, D. Zhang *et al.*, "Partial FC: Training 10 million identities on a single machine," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 1445–1449.