AI-KD: Towards Alignment Invariant Face Image Quality Assessment Using Knowledge Distillation

Žiga Babnik[1], Fadi Boutros[2], Naser Damer[2,3], Peter Peer[1], and Vitomir Štruc[1]
[1]University of Ljubljana, Ljubljana, Slovenia
[2]Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany
[3]Department of Computer Science, TU Darmstadt, Darmstadt, Germany
https://github.com/LSIbabnikz/AI-KD

***Abstract*—Face Image Quality Assessment (FIQA) techniques have seen steady improvements over recent years, but their performance still deteriorates if the input face samples are not properly aligned. This alignment sensitivity comes from the fact that most FIQA techniques are trained or designed using a specific face alignment procedure. If the alignment technique changes, the performance of most existing FIQA techniques quickly becomes suboptimal. To address this problem, we present in this paper a novel knowledge distillation approach, termed AI-KD that can extend on any existing FIQA technique, improving its robustness to alignment variations and, in turn, performance with different alignment procedures. To validate the proposed distillation approach, we conduct comprehensive experiments on 6 face datasets with 4 recent face recognition models and in comparison to 7 state-of-the-art FIQA techniques. Our results show that AI-KD consistently improves performance of the initial FIQA techniques not only with misaligned samples, but also with properly aligned facial images. Furthermore, it leads to a new state-of-the-art, when used with a competitive initial FIQA approach. The code for AI-KD is made publicly available from: https://github.com/LSIbabnikz/AI-KD.**

***Index Terms*—Computer Vision, Face Recognition, Face Image Quality Assessment, Face Detection, Face Alignment**

I. Introduction

Face Image Quality Assessment (FIQA) refers to the process of predicting quality scores for facial images with the goal of providing automated face recognition (FR) models with auxiliary information for the recognition process. This is particularly important for unconstrained image acquisition scenarios, where sample quality can vary significantly and can, therefore, have an adverse impact on performance [1], [2].

Modern FIQA approaches typically predict a single numerical value from the input face samples, also referred to as a *unified quality score*, that aims to capture the biometric utility of the given sample for the recognition task [3]. While significant progress has been made in FIQA techniques over the years [2], existing techniques are still sensitive to the alignment of the input samples [4]. The main reason for this sensitivity is that most FIQA techniques are trained on samples aligned using a specific facial landmark detector (also often referred to as a face keypoint detector), and, as such, also often overfit to that particular landmark detector. Even though modern landmark detectors are robust and perform well on challenging benchmarks [5], using an unknown detector that was not seen during training, still leads to a notable decrease in FIQA performance.

To address this issue, we present in this paper an **A**lignment-**I**nvariant **K**nowledge **D**istillation (AI-KD) procedure that improves the performance of existing FIQA approaches, when dealing with samples produced using any (unknown) face landmark detector. AI-KD relies on a novel distillation process that incorporates simple image transformations, which mimic the (minor) variability between samples produced by different alignment approaches. Using multiple FIQA techniques, FR models and performance benchmarks, we show that AI-KD leads to considerable performance gains when confronted with misaligned face images, but also improves performance when the face samples are optimally aligned.

II. Related Work

A. General FIQA Techniques

General FIQA techniques assess the (biometric) quality of the given face samples, by predicting a single (unified) numerical score, where a higher score usually refers to a higher quality. Based on how the methods compute the quality score, they can be further divided into: unsupervised (analytical) methods and supervised (regression-based) methods.

Unsupervised methods estimate the quality of the provided face samples by analyzing their characteristics. Early techniques observed (human) perceptual characteristics, such as noise levels, blurriness, lighting conditions, etc. [6], [7]. More recent techniques, on the other hand, analyze the image characteristics from the viewpoint of a FR model. These methods most often measure the robustness of the sample's representation in the embedding space of the targeted FR model. The earliest such method, named SER-FIQ [8], uses Dropout to produce several representations of a single input sample, whereas FaceQAN [9] uses adversarial methods to generate a number of perturbed samples that are then utilized for quality estimation. The most recent, DifFIQA [10] approach, uses the forward and backward processes of diffusion models to generated perturbations of the face samples.

Supervised methods commonly train a quality-regression model to assess the quality of the studied face samples. The regression models are trained on pseudo-quality labels obtained with various strategies. Early works relied on visual

Supported in parts by the ARIS junior researcher program, ARIS Programmes P2-0250 "Metrology and Biometric Systems" and P2-0214 "Computer Vision", and ARIS Research Project J2-50069 "MIXBAI".

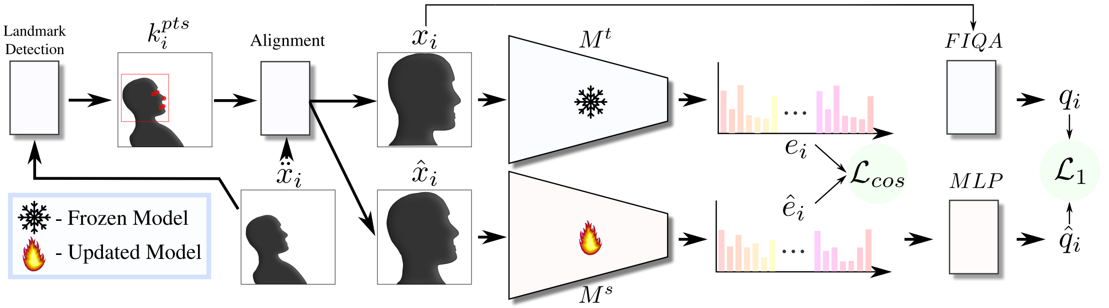

Fig. 1. **Overview of the proposed Alignment Invariant Knowledge Distillation (AI-KD) process.** The proposed approach trains a quality-regression model, consisting of a FR backbone M^s and a quality regression head MLP, on quality labels q_i extracted using any existing FIQA approach. Training samples $\hat{x}_i$ are (mis)aligned on the fly, by perturbing the correct landmark k_i^{pts} of the initial unaligned face samples $\ddot{x}_i$. Additionally, to ensure robustness to alignment variations in the distilled model, we design a distillation objective that ensures consistency between representations of the aligned e_i and (mis)aligned images $\hat{e}_i$, as well as matching the predicted quality scores $\hat{q}_i$ to the quality labels q_i.

image characteristics, similar to unsupervised methods, or human annotations [11] to compile the labels. Conversely, modern approaches consider FR models and the recognition procedure in the label generation process. FaceQnet [12], for example, compares all images of an individual to its highest quality sample. A similar approach, PCNet [13], compares several pairs of images of the same individual to generate labels, while a more advanced approach SDD-FIQA [14] takes into account also the imposter pairs, or the pairs containing two images of distinct individuals when producing pseudo-quality labels for training.

B. Quality-Aware FR Techniques

Unlike general FIQA techniques, quality-aware FR techniques combine quality estimation and face recognition into a single task. Here, FR models are trained to discern the identity as well as the quality of the input face samples, by employing standard training procedures combined with a quality regression branch. One of the earliest examples of such methods is PFE [15], which measured the uncertainty of the samples representation in the latent space of the recognition model. Uncertainty, in this case, can be viewed as the inverse measure of quality. The MagFace [16] approach, extends the ArcFace [17] margin-loss with a magnitude-aware term, allowing it to encode quality into the magnitude of the sample representations. The most recent approach, CR-FIQA [18], estimates quality as the ratio between the distance of a sample representation to the positive class-center and the nearest negative-class center.

III. Methodology

Face Image Quality Assessment (FIQA) techniques require properly aligned input samples in order to achieve the best possible performance, consequently limiting the choice of landmark detection algorithms used during inference. Although most landmark detectors achieve high accuracy, their predictions usually differ by several pixels, which is enough to adversely impact the performance of any existing FIQA technique. In this paper, we present a simple, but elegant knowledge distillation approach named AI-KD (**A**lignment-**I**nvariant **K**nowledge **D**istillation), which aims to improve the robustness and overall performance of FIQA techniques when dealing with samples aligned using different landmark detectors. The method, presented in Fig. 1, uses knowledge distillation to fine-tune a preexisting FR model M to predict quality scores $\hat{q}_i$ of the input samples $\hat{x}_i$, through an additional MLP (Multi Layer Perceptron) quality-regression head. During training two copies of M are utilized, i.e., the teacher M^t (with frozen weights) is used to generate ground-truth representations of properly aligned input samples x_i, while the student M^s is trained to be robust to alignment variations by learning from improperly aligned samples $\hat{x}_i$. In the following sections, we present the whole AI-KD methodology in-detail.

A. Method Overview

The proposed AI-KD technique aims to improve the performance of any existing FIQA method, when presented with input samples aligned with an unknown landmark detection algorithm. Formally, given a FIQA technique Q, the goal is to train a quality-regression model, consisting of a pretrained FR backbone M^s and an MLP based quality-regression head, on a large face dataset $\{\ddot{x}_i\}_{i=0}^N$ containing N (in general non-aligned) samples. Towards this end, we first extract facial landmarks k_i^{pts} and pseudo quality labels q_i from all samples in the datasets. During training, we then employ the alignment-invariant knowledge distillation procedure, which dynamically transforms each initial face sample $\ddot{x}_i$ into a properly aligned sample x_i and a misaligned sample $\hat{x}_i$, imitating alignment variations of different landmark detectors. The knowledge distillation process uses both, a quality $\mathcal{L}_1$ and feature loss $\mathcal{L}_{cos}$, to simultaneously ensure optimal FIQA performance and alignment-invariance of the final trained/distilled model.

B. Data Preprocessing

During the preprocessing step, the facial landmarks k_i^{pts} of all N samples $\ddot{x}_i$ are extracted using a chosen landmark detector D, such that $k_i^{pts} = D(\ddot{x}_i)$. The extracted landmarks k_i^{pts}, consist of the coordinates of the left and right eye, the tip of the nose and the corners of the lips, and can be used to properly align $\ddot{x}_i$ by matching the coordinates to a predefined template, resulting in a properly aligned sample x_i [17]–[19]. Additionally, pseudo quality labels q_i are also calculated for all N samples $\ddot{x}_i$ using a chosen FIQA technique with the well-aligned samples $q_i = Q(x_i)$.

TABLE I
SUMMARY OF THE CHARACTERISTICS OF THE EXPERIMENTAL DATASETS.

Dataset	#Images	#IDs	#Comparisons		Use Case
			Mated	Non-mated	
Adience [20]	19,370	2,284	20,000	20,000	General
LFW [21]	13,233	5,749	3,000	3,000	
CPLFW [22]	11,652	3,930	3,000	3,000	Cross-Pose
CFP-FP [23]	7,000	500	3,500	3,500	
CALFW [24]	12,174	4,025	3,000	3,000	Cross-Age
XQLFW [25]	13,233	5,749	3,000	3,000	Cross-Quality

TABLE II
EXPERIMENTS USING PROPERLY ALIGNED SAMPLES.

			Adience	LFW	CPLFW	CFP-FP	CALFW	XQLFW	$\overline{pAUC}$
Cross-Model	**AdaFace - pAUC@FMR=10^{-3}(↓)**								
	SER-FIQ [8]	B	0.839	0.897∗	0.743∗	0.619∗	0.879∗	0.843	0.803
		T	**0.801**∗	0.915	0.755	0.620	0.915	0.686∗	0.782
	DifFIQA(R) [10]	B	0.816∗	0.848	0.680∗	0.540∗	0.919	**0.610**∗	0.735
		T	0.835	**0.746**∗	0.692	0.551	0.882∗	0.643	0.725
	CR-FIQA [18]	B	0.844	0.851	0.671	0.544	**0.856**∗	0.685	0.742
		T	0.818∗	0.832∗	**0.664**∗	**0.520**∗	0.891	0.641∗	0.728
	SwinFace - pAUC@FMR=10^{-3}(↓)								
			Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	SER-FIQ [8]	B	0.811	0.887∗	**0.698**∗	0.534∗	0.867	0.872	0.778
		T	**0.761**∗	0.939	0.759	0.546	0.840∗	0.636∗	0.747
	DifFIQA(R) [10]	B	0.805∗	0.859	0.736	0.477	0.897	**0.567**∗	0.724
		T	0.811	**0.770**∗	0.734∗	0.459∗	0.851∗	0.633	0.710
	CR-FIQA [18]	B	0.807	0.879	0.724	0.422	**0.813**∗	0.640	0.714
		T	0.788∗	0.860∗	0.718∗	**0.399**∗	0.890	0.621∗	0.713
	TransFace - pAUC@FMR=10^{-3}(↓)								
			Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	SER-FIQ [8]	B	0.837	0.897∗	0.730	0.657	0.910∗	0.820	0.808
		T	**0.771**∗	0.915	0.721∗	0.630∗	0.920	0.611∗	0.761
	DifFIQA(R) [10]	B	0.812∗	0.870	0.640∗	0.528∗	0.920	**0.524**∗	0.716
		T	0.835	**0.784**∗	0.647	0.551	**0.887**∗	0.566	0.712
	CR-FIQA [18]	B	0.829	0.851	0.639	0.512	0.887∗	0.580∗	0.716
		T	0.807∗	0.840∗	**0.625**∗	**0.486**∗	0.935	0.602	0.716
Same-Model	**CosFace - pAUC@FMR=10^{-3}(↓)**								
			Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	SER-FIQ [8]	B	0.825	0.858∗	0.760∗	0.606∗	0.908∗	0.770	0.788
		T	**0.791**∗	0.938	0.789	0.614	0.922	0.563∗	0.769
	DifFIQA(R) [10]	B	0.805∗	0.831	0.707∗	0.522	0.916	**0.557**∗	0.723
		T	0.807	**0.746**∗	0.710	0.517∗	0.893∗	0.566	0.707
	CR-FIQA [18]	B	0.835	0.851	0.696	0.503	**0.889**∗	0.631	0.734
		T	0.803∗	0.847∗	**0.693**∗	**0.477**∗	0.934	0.580∗	0.722

B - *performance of the baseline approach*, **T** - *performance of the extended AI-KD approach*

C. Alignment-Invariant Knowledge Distillation

The knowledge distillation procedure uses the landmarks k_i^{pts} and pseudo-quality labels q_i extracted in the data preprocessing step to train a quality-regression model, consisting of a pretrained FR backbone and a quality regression head, i.e., $M^s \circ MLP$. The training process consists of two main steps: (i) the sample transformation step and (ii) the actual knowledge distillation. The sample transformation step generates samples with varying alignments, whereas the knowledge distillation step transfers the knowledge encoded in the pseudo-quality labels to the student model.

Sample Transformation Step. During this step, the initial face sample $\ddot{x}_i$ is used to generate a properly aligned sample x_i and a misaligned sample $\hat{x}_i$ using the extracted landmarks k_i^{pts}. The properly aligned sample x_i is generated by aligning according to k_i^{pts}, while $\hat{x}_i$ aims to replicate the alignment produced by an unknown facial landmark detector. Since it is not feasible to extract landmarks of all samples x_i using a large number of unique landmark detectors, we make a simple assumption that the predicted landmarks of any well-functioning face landmark detection method will be approximately similar to the baseline landmarks k_i^{pts}. Using this assumption, we then generate new landmarks $\hat{k}_i^{pts}$ corresponding to an unknown method $\hat{D}$, by randomly sampling around the reference coordinates in k_i^{pts}. Formally, this can be written as:

$$\hat{k}_i^{pts} = k_i^{pts} + \mathcal{U}_{[-p,p]}, \tag{1}$$

where $\mathcal{U}_{[-p,p]}$ is a uniform random variable sampled from the interval $[-p,p]$. This means that all coordinates can differ at most by p pixels between the two landmarks (initial and perturbed). The misaligned sample $\hat{x}_i$ is then produced by aligning $\ddot{x}$ using the newly constructed landmarks $\hat{k}_i^{pts}$.

Knowledge Distillation. The N training samples x_i and $\hat{x}_i$ produced by the sample transformation step form the basis for the knowledge distillation procedure. Here, the teacher model M^t is frozen during the entire training process (its weights are not updated), while the parameters of the student model $M^s \circ MLP$ are optimized using dedicated distillation objectives. The properly aligned sample x_i is fed through the frozen teacher model M^t, to produce a feature representation $e_i = M^t(x_i)$ of the input sample. The computed representation e_i and the corresponding pseudo-quality label q_i jointly represent the regression targets for the student model ($M^s \circ MLP$). To be able to define a loss for the distillation procedure, the misaligned sample $\hat{x}_i$ is fed through M^s, producing the feature representation of the misaligned sample $\hat{e}_i$. The computed representation $\hat{e}_i$ is then further processed through an MLP to produce the predicted quality label $\hat{q}_i$. The overall distillation objective used to optimize $M^s \circ MLP$ is then defined as the average of a representation and a quality term. Here, the representation loss, aims to align the representations of $\hat{x}_i$ and x:

$$\mathcal{L}_{cos}(e_i, \hat{e}_i) = 1 - \frac{e_i \cdot \hat{e}_i^T}{\|e_i\| \cdot \|\hat{e}_i\|}, \tag{2}$$

where $\|e_i\|$ represents the norm of the representation, while the quality loss aims to ensure that the original and predicted quality scores are as close as possible, i.e.:

$$\mathcal{L}_1(q_i, \hat{q}_i) = |q_i - \hat{q}_i|. \tag{3}$$

IV. EXPERIMENTS & RESULTS

Experimental setting. We analyze the performance of AI-KD over 3 FIQA methods and in comparison to 7 state-of-the-art competitors: (i) the **unsupervised** FaceQAN [9] and SER-FIQ [8] models, (ii) the **supervised** FaceQnet [26], SDD-FIQA [14] and DifFIQA(R) [10] techniques, and (iii) the **quality-aware** MagFace [16] and CR-FIQA [18] methods. We test all methods on 6 commonly used benchmarks with different quality characteristics, as summarized in Table I, i.e.: Adience [20], Labeled Faces in the Wild (LFW) [21], Cross-Pose Labeled Faces in the Wild (CPLFW) [22], Celebrities in Frontal-Profile in the Wild (CFP-FP) [23], Cross-Age Labeled Faces in the Wild (CALFW) [24] and the Cross-Quality Labeled Faces in the Wild (XQLFW) [25]. Because the performance of FIQA techniques is dependent on the FR model

TABLE III
SAME-MODEL EXPERIMENTS USING MISALIGNED SAMPLES.

			CosFace - pAUC@FMR=10^{-3}(↓)						
			Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
RetinaFace(MNet)	**SER-FIQ** [8]	B	0.868	**0.813**∗	0.747∗	0.614	0.896∗	0.744	0.780
		T	**0.832**∗	0.938	0.792	0.586∗	0.920	0.583∗	0.775
	DifFIQA(R) [10]	B	0.861	0.921	0.699∗	0.515	**0.894**∗	**0.551**∗	0.740
		T	0.840∗	0.825∗	0.709	0.499∗	0.901	0.562	0.723
	CR-FIQA [18]	B	0.865	0.851∗	**0.689**∗	0.508	0.895∗	0.638	0.741
		T	0.836∗	0.870	0.699	**0.472**∗	0.931	0.585∗	0.732
MTCNN			**Adience**	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	SER-FIQ [8]	B	0.894∗	0.871∗	0.837	0.734	0.909	0.692	0.823
		T	0.899	0.886	0.773∗	0.695∗	0.899∗	0.590∗	0.790
	DifFIQA(R) [10]	B	0.889∗	0.799	0.727	0.713	0.912	**0.550**∗	0.765
		T	0.906	**0.700**∗	0.715∗	0.692∗	0.905∗	0.564	0.747
	CR-FIQA [18]	B	0.915	0.808	0.718	0.662	**0.889**∗	0.605	0.766
		T	**0.873**∗	0.743∗	**0.709**∗	**0.622**∗	0.934	0.601∗	0.747
DLib			**Adience**	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	SER-FIQ [8]	B	0.910	0.853	0.881∗	0.747∗	0.902∗	0.744	0.839
		T	0.883∗	0.848∗	0.901	0.868	0.907	0.638∗	0.841
	DifFIQA(R) [10]	B	0.880∗	0.809	0.939	0.970	0.892	**0.615**∗	0.851
		T	0.884	**0.687**∗	**0.842**∗	**0.722**∗	**0.879**∗	0.631	0.774
	CR-FIQA [18]	B	0.915	0.789	0.902	0.804	0.882∗	0.667∗	0.827
		T	**0.876**∗	0.735∗	0.864∗	0.755∗	0.910	0.693	0.805

B - *performance of the baseline approach*, **T** - *performance of the extended AI-KD approach*

used, we investigate how well the techniques generalize over 4 state-of-the-art models divided into CNN-based models, i.e., AdaFace[1] [19], and CosFace[1]and Transformer-based models i.e., SwinFace[1] [27], and TransFace[1] [28]. To evaluate the effects of alignment on the performance of FIQA techniques, we employ four different face landmark detection methods i.e., RetinaFace [29] (using ResNet50 and MobileNet backbones), MTCNN [30], and DLib [31].

Evaluation methodology. Using standard evaluation methodology [8], [9], [18], we quantify the performance of the tested methods using the pAUC (partial Area Under the Curve) of the Error-versus-Discard Characteristic (EDC) curves (also referred to as Error-versus-Reject Characteristic (ERC) curves). The EDC curves measure how the performance of a given FR model improves, when rejecting some percentage of the lowest quality images from the dataset, and are calculated using a predefined False Match Rate (FMR) (10^{-3} in our case), while increasing low-quality image discard (reject) rates. In real-world situation, it is not feasible to reject a large percentage of all samples, therefore we report the pAUC at lower values of the discard (reject) rates (30% in our case). Furthermore, for easier interpretation and comparison of scores over different dataset, we normalize the calculated pAUC values using the FNMR at the 0% discard rate, with lower pAUC values indicating better performance.

Implementation Details. We use the VGGFace2 dataset, to train the evaluted models. For the FR backbone of the quality-regression models, we use a ResNet100 model, trained using the CosFace loss. Based on this choice, we split the experiments into Cross- and Same-model scenarios based on whether the evaluated model is also trained using the CosFace loss. For the sample transformation, we use $p = 3$, as preliminary testing showed that different landmark detection methods vary between 3-4 pixels in their predictions. To train the model, we used Stochastic Gradient Descend, with a learning rate of 0.05. Additionally to further improve the model we employed Stochastic Weight Averaging. All experiments were conducted on a desktop PC with an Intel i9-10900KF CPU, 64 GB of RAM and an Nvidia 3090 GPU.

TABLE IV
CROSS-MODEL EXPERIMENTS USING MISALIGNED SAMPLES.

			AdaFace - pAUC@FMR=10^{-3}(↓)						
			Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
RetinaFace(MNet)	**SER-FIQ** [8]	B	0.886	0.852∗	0.738∗	0.634	0.865∗	0.829	0.801
		T	**0.841**∗	0.923	0.755	0.594∗	0.916	0.694∗	0.787
	DifFIQA(R) [10]	B	0.880	0.953	0.670∗	0.506∗	0.898	**0.601**∗	0.751
		T	0.866∗	**0.825**∗	0.691	0.507	0.884∗	0.635	0.735
	CR-FIQA [18]	B	0.877	0.844∗	**0.659**∗	0.540	**0.863**∗	0.683	0.744
		T	0.847∗	0.855	0.665	**0.501**∗	0.893	0.649∗	0.735
			SwinFace - pAUC@FMR=10^{-3}(↓)						
			Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	SER-FIQ [8]	B	0.854	**0.840**∗	0.712∗	0.500∗	0.854	0.794	0.759
		T	**0.810**∗	0.947	0.778	0.540	0.837∗	0.641∗	0.759
	DifFIQA(R) [10]	B	0.852	0.945	0.728∗	0.441∗	0.842∗	**0.586**∗	0.732
		T	0.836∗	0.852∗	0.736	0.441	0.870	0.686	0.737
	CR-FIQA [18]	B	0.841	0.872∗	**0.702**∗	0.414	**0.816**∗	0.672	0.720
		T	0.820∗	0.877	0.725	**0.400**∗	0.891	0.620∗	0.722
			TransFace - pAUC@FMR=10^{-3}(↓)						
			Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	SER-FIQ [8]	B	0.886	0.852∗	0.715∗	0.641	0.901∗	0.801	0.799
		T	**0.818**∗	0.923	0.725	0.605∗	0.919	0.609∗	0.766
	DifFIQA(R) [10]	B	0.882	0.953	0.632∗	0.521	0.898	**0.534**∗	0.737
		T	0.859∗	0.847∗	0.647	0.510∗	**0.890**∗	0.551	0.717
	CR-FIQA [18]	B	0.864	**0.844**∗	0.634	0.521	0.896∗	0.588∗	0.724
		T	0.843∗	0.855	**0.628**∗	**0.469**∗	0.933	0.603	0.722
MTCNN			AdaFace - pAUC@FMR=10^{-3}(↓)						
			Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	SER-FIQ [8]	B	0.912	0.910∗	0.802	0.708∗	0.878∗	0.846	0.843
		T	0.909∗	0.911	0.739∗	0.742	0.890	0.697∗	0.815
	DifFIQA(R) [10]	B	0.917∗	0.815	0.695	0.745	0.910	**0.622**∗	0.784
		T	0.933	**0.683**∗	0.686∗	0.729∗	0.895∗	0.655	0.763
	CR-FIQA [18]	B	0.929	0.832	0.685	0.713	**0.857**∗	0.661	0.779
		T	**0.891**∗	0.768∗	**0.678**∗	**0.658**∗	0.895	0.650∗	0.757
			SwinFace - pAUC@FMR=10^{-3}(↓)						
			Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	SER-FIQ [8]	B	0.883	0.901∗	0.833	0.632	0.872	0.832	0.826
		T	0.875∗	0.901	0.777∗	0.601∗	0.823∗	0.681∗	0.776
	DifFIQA(R) [10]	B	0.892∗	0.825	0.745	0.613	0.915	0.564	0.759
		T	0.903	**0.723**∗	0.732∗	0.545∗	0.896∗	**0.563**∗	0.727
	CR-FIQA [18]	B	0.894	0.827	0.732	0.518	**0.813**∗	0.605∗	0.732
		T	**0.862**∗	0.752∗	**0.725**∗	**0.482**∗	0.892	0.621	0.722
			TransFace - pAUC@FMR=10^{-3}(↓)						
			Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	SER-FIQ [8]	B	0.906	0.910∗	0.801	0.748	0.910	0.794	0.845
		T	**0.875**∗	0.911	0.714∗	0.682∗	0.900∗	0.619∗	0.783
	DifFIQA(R) [10]	B	0.904∗	0.837	0.672	0.712	0.922	**0.540**∗	0.764
		T	0.925	**0.722**∗	0.658∗	0.685∗	0.897∗	0.545	0.739
	CR-FIQA [18]	B	0.917	0.840	0.659	0.657	**0.888**∗	0.576∗	0.756
		T	0.878∗	0.768∗	**0.645**∗	**0.589**∗	0.932	0.594	0.734
DLib			AdaFace - pAUC@FMR=10^{-3}(↓)						
			Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	SER-FIQ [8]	B	0.928	0.885	0.858∗	0.736∗	0.884∗	0.868	0.860
		T	0.901∗	0.848∗	0.906	0.769	0.895	0.765∗	0.847
	DifFIQA(R) [10]	B	0.902∗	0.841	0.955	0.871	0.888	**0.718**∗	0.863
		T	0.906	**0.696**∗	**0.832**∗	**0.699**∗	0.875∗	0.724	0.789
	CR-FIQA [18]	B	0.930	0.821	0.881	0.776	**0.854**∗	0.748∗	0.835
		T	**0.897**∗	0.752∗	0.860∗	0.728∗	0.875	0.751	0.810
			SwinFace - pAUC@FMR=10^{-3}(↓)						
			Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	SER-FIQ [8]	B	0.908	0.874	**0.801**∗	0.717∗	0.868	0.727	0.816
		T	**0.861**∗	0.836∗	0.869	0.727	0.844∗	0.652∗	0.798
	DifFIQA(R) [10]	B	0.873∗	0.836	0.847	0.875	0.875	**0.624**∗	0.822
		T	0.873	**0.709**∗	0.807∗	**0.631**∗	0.843∗	0.641	0.751
	CR-FIQA [18]	B	0.891	0.808	0.837	0.699	**0.828**∗	0.707	0.795
		T	0.866∗	0.735∗	0.825∗	0.638∗	0.877	0.678∗	0.770
			TransFace - pAUC@FMR=10^{-3}(↓)						
			Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	SER-FIQ [8]	B	0.919	0.885	0.859∗	0.763∗	0.908	0.806	0.857
		T	**0.870**∗	0.848∗	0.899	0.804	0.906∗	0.688∗	0.836
	DifFIQA(R) [10]	B	0.882∗	0.848	0.919	0.944	0.900	0.620	0.852
		T	0.887	**0.726**∗	**0.801**∗	**0.685**∗	**0.872**∗	**0.616**∗	0.765
	CR-FIQA [18]	B	0.901	0.821	0.872	0.840	0.879∗	0.647∗	0.827
		T	0.881∗	0.752∗	0.834∗	0.745∗	0.913	0.657	0.797

B - *performance of the baseline approach*, **T** - *performance of the extended AI-KD approach*

[1] https://github.com/mk-minchul/AdaFace; https://github.com/deepinsight/insightface; https://github.com/lxq1000/SwinFace; https://github.com/DanJun6737/TransFace

TABLE V
COMPARISON WITH STATE-OF-THE-ART – PROPERLY ALIGNED SAMPLES.

		Adience	LFW	CPLFW	CFP-FP	CALFW	XQLFW	$\overline{pAUC}$
Cross-Model	**AdaFace**							
	FaceQAN [9]	0.880	0.780	0.679	**0.387**	0.952	0.634	0.719
	SER-FIQ [8]	0.839	0.897	0.743	0.619	0.879	0.843	0.803
	FaceQnet [12]	0.961	0.862	0.883	0.735	0.937	0.977	0.893
	SDD-FIQA [14]	0.839	0.897	0.743	0.619	0.879	0.843	0.803
	DifFIQA(R) [10]	0.816	0.848	0.680	0.540	0.919	**0.610**	0.735
	MagFace [16]	0.862	0.874	0.735	0.692	0.868	0.914	0.824
	CR-FIQA [18]	0.844	0.851	0.671	0.544	**0.856**	0.685	0.742
	AI-KD(SER-FIQ)	**0.801**	0.915	0.755	0.620	0.915	0.686	0.782
	AI-KD(CR-FIQA)	0.818	0.832	**0.664**	0.520	0.891	0.641	0.728
	AI-KD(DifFIQA(R))	0.835	**0.746**	0.692	0.551	0.882	0.643	0.725
	SwinFace							
		Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	FaceQAN [9]	0.860	0.797	0.702	0.409	0.966	0.613	0.725
	SER-FIQ [8]	0.811	0.887	**0.698**	0.534	0.867	0.872	0.778
	FaceQnet [12]	0.918	0.891	0.847	0.652	0.938	0.927	0.862
	SDD-FIQA [14]	0.811	0.887	0.698	0.534	0.867	0.872	0.778
	DifFIQA(R) [10]	0.805	0.859	0.736	0.477	0.897	**0.567**	0.724
	MagFace [16]	0.830	0.888	0.774	0.551	0.855	0.943	0.807
	CR-FIQA [18]	0.807	0.879	0.724	0.422	**0.813**	0.640	0.714
	AI-KD(SER-FIQ)	**0.761**	0.939	0.759	0.546	0.840	0.636	0.747
	AI-KD(CR-FIQA)	0.788	0.860	0.718	**0.399**	0.890	0.621	0.713
	AI-KD(DifFIQA(R))	0.811	**0.770**	0.734	0.459	0.851	0.633	0.710
	TransFace							
		Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	FaceQAN [9]	0.874	0.802	0.633	**0.388**	0.986	0.575	0.710
	SER-FIQ [8]	0.837	0.897	0.730	0.657	0.910	0.820	0.808
	FaceQnet [12]	0.934	0.862	0.885	0.747	0.965	1.007	0.900
	SDD-FIQA [14]	0.837	0.897	0.730	0.657	0.910	0.820	0.808
	DifFIQA(R) [10]	0.812	0.870	0.640	0.528	0.920	**0.524**	0.716
	MagFace [16]	0.869	0.841	0.729	0.652	0.901	0.935	0.821
	CR-FIQA [18]	0.829	0.851	0.639	0.512	0.887	0.580	0.716
	AI-KD(SER-FIQ)	**0.771**	0.915	0.721	0.630	0.920	0.611	0.761
	AI-KD(CR-FIQA)	0.807	0.840	**0.625**	0.486	0.935	0.602	0.716
	AI-KD(DifFIQA(R))	0.835	**0.784**	0.647	0.551	**0.887**	0.566	0.712
Same-Model	**CosFace**							
		Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	FaceQAN [9]	0.866	0.772	0.702	**0.374**	0.987	0.574	0.712
	SER-FIQ [8]	0.825	0.858	0.760	0.606	0.908	0.770	0.788
	FaceQnet [12]	0.948	0.862	0.867	0.691	0.956	0.894	0.870
	SDD-FIQA [14]	0.825	0.858	0.760	0.606	0.908	0.770	0.788
	DifFIQA(R) [10]	0.805	0.831	0.707	0.522	0.916	**0.557**	0.723
	MagFace [16]	0.854	0.866	0.763	0.711	0.902	0.944	0.840
	CR-FIQA [18]	0.835	0.851	0.696	0.503	**0.889**	0.631	0.734
	AI-KD(SER-FIQ)	**0.791**	0.938	0.789	0.614	0.922	0.563	0.769
	AI-KD(CR-FIQA)	0.803	0.847	**0.693**	0.477	0.934	0.580	0.722
	AI-KD(DifFIQA(R))	0.807	**0.746**	0.710	0.517	0.893	0.566	0.707

A. *Analysis of AI-KD*

In this section, we analyze how the presented AI-KD technique improves the performance of state-of-the-art FIQA methods, across a variety of benchmark datasets and FR models. We chose three distinct FIQA methods i.e.: the unsupervised SER-FIQ, the supervised DifFIQA(R), and the quality-aware CR-FIQA methods. We separate the experiments by two criteria: (i) based on the used FR model for evaluation, and (ii) based on the alignment of the evaluation benchmarks. Based on the FR model we consider Cross- and Same-Model experiments, where in the *Cross-Model experiments* the FR models used during the knowledge distillation step M^t differs from the evaluation FR model, while in the *Same-Model experiments* the two are the same. When considering alignment, we separate the experiments into *experiments with properly aligned images*, where the quality scores are predicted from optimally aligned faces samples for the targeted FR model, and *experiments with misaligned images*, where the quality scores are extracted from face samples aligned with an arbitrary (non-optimal) landmark detector.

Experiments with Properly Aligned Images. The results of the experiments with properly aligned images are shown in Table II for both the Cross- and Same-Model scenarios. Here, the average results across all datasets are marked as $\overline{pAUC}$. For each method, we show the baseline (B) results and the extended AI-KD approach (T), the better method of the two is marked with $*$ for individual datasets and with green for $\overline{pAUC}$. The best result of individual datasets is marked with **bold**. From the results, we observe that the proposed AI-KD approach outperforms the baseline FIQA approaches, for all included FIQA techniques and on all tested FR models in terms of overall $\overline{pAUC}$ scores. Interestingly, the results on individual datasets are relatively close between the baseline and extended approaches, with the biggest improvements seen mostly on the hardest of the benchmarks XQLFW.

Experiments with Misaligned Images. From the results in Table III for the Same-Model scenario and in Table IV for the Cross-Model scenario, we again see that for all combinations of benchmarks, FIQA techniques and FR models, the extended methods using AI-KD perform better than the baseline methods. One exception is when using CR-FIQA and DifFIQA(R) in combination with the SwinFace FR model in the Cross-Model scenario, where the baseline approach outperforms the extended approach by a slight margin. When looking at the results per individual benchmark the results appear to vary quite a bit between the different methods. The largest variation can be seen on the most difficult benchmark XQLFW, while for all others the differences between the extended and the baseline approaches appears to be significantly smaller.

B. *Comparison with the State-of-the-Art*

In this section, we compare the extended/distilled techniques to state-of-the-art FIQA methods, for both properly aligned and misaligned samples. The results using proper alignment are shown in Table V, while the results using misaligned samples are reported in Table VI. In both tables pAUC scores at a discard rate of 0.3 and a FMR of 10^{-3} are shown. Average results across all datasets are marked as $\overline{pAUC}$, the best result on individual datasets are marked bold, the overall best result is marked green, the second-best blue and the third-best red.

Experiments with Aligned Images. From the results in Table V, we observe that not only can the proposed knowledge distillation scheme improve the performance of existing FIQA techniques, it can easily achieve state-of-the-art results on all tested scenarios. The top performing method across the Cross- and Same-model scenarios is the extended DifFIQA(R) approach, achieving either the best or second-best result across all cases, closely followed by the extended CR-FIQA approach and FaceQAN. While the extended SER-FIQ outperforms the baseline, its results do not hold up against the top three contending methods. However, compared to all other tested methods, it still leads to competitive results.

Experiments with Misaligned Images. For readability's sake, the results in Table VI with misaligned images are combined into a single score by averaging over all experiments. The results again tell a similar story, the best performing method is the extended DifFIQA(R) method, followed by the extended CR-FIQA and FaceQAN methods. However,

TABLE VI
COMPARISON WITH THE STATE-OF-THE-ART – MISALIGNED SAMPLES.

		Adience	LFW	CPLFW	CFP-FP	CALFW	XQLFW	$\overline{pAUC}$
Cross-Model	**AdaFace**							
	FaceQAN [9]	0.910	0.760	0.767	**0.622**	0.926	0.684	0.778
	SER-FIQ [8]	0.909	0.882	0.800	0.693	0.876	0.848	0.834
	FaceQnet [12]	0.984	0.850	0.941	0.790	0.940	0.983	0.915
	SDD-FIQA [14]	0.909	0.882	0.800	0.693	0.876	0.848	0.834
	DifFIQA(R) [10]	0.900	0.870	0.773	0.707	0.899	**0.647**	0.799
	MagFace [16]	0.925	0.894	0.795	0.682	0.864	0.894	0.842
	CR-FIQA [18]	0.912	0.832	0.742	0.676	**0.858**	0.697	0.786
	T(SER-FIQ)	0.884	0.894	0.800	0.702	0.900	0.719	0.816
	T(CR-FIQA)	**0.878**	0.792	**0.734**	0.629	0.887	0.683	0.767
	T(DifFIQA(R))	0.902	**0.735**	0.736	0.645	0.885	0.671	0.762
	SwinFace							
		Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	FaceQAN [9]	0.881	0.762	0.784	0.569	0.939	0.643	0.763
	SER-FIQ [8]	0.882	0.872	0.782	0.617	0.864	0.784	0.800
	FaceQnet [12]	0.948	0.875	0.876	0.694	0.942	0.944	0.880
	SDD-FIQA [14]	0.882	0.872	0.782	0.617	0.864	0.784	0.800
	DifFIQA(R) [10]	0.873	0.869	0.773	0.643	0.877	**0.591**	0.771
	MagFace [16]	0.895	0.905	0.818	0.654	0.855	0.895	0.837
	CR-FIQA [18]	0.875	0.836	**0.757**	0.544	**0.819**	0.661	0.749
	AI-KD(SER-FIQ)	**0.848**	0.895	0.808	0.623	0.835	0.658	0.778
	AI-KD(CR-FIQA)	0.849	0.788	0.758	**0.507**	0.887	0.640	0.738
	AI-KD(DifFIQA(R))	0.871	**0.761**	0.758	0.539	0.870	0.630	0.738
	TransFace							
		Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	FaceQAN [9]	0.899	**0.761**	0.740	0.622	0.960	0.636	0.769
	SER-FIQ [8]	0.904	0.882	0.792	0.718	0.906	0.800	0.834
	FaceQnet [12]	0.960	0.847	0.937	0.823	0.967	0.998	0.922
	SDD-FIQA [14]	0.904	0.882	0.792	0.718	0.906	0.800	0.834
	DifFIQA(R) [10]	0.889	0.879	0.741	0.726	0.907	**0.565**	0.784
	MagFace [16]	0.927	0.893	0.780	0.690	0.897	0.918	0.851
	CR-FIQA [18]	0.894	0.835	0.722	0.672	0.888	0.603	0.769
	AI-KD(SER-FIQ)	**0.854**	0.894	0.780	0.697	0.908	0.638	0.795
	AI-KD(CR-FIQA)	0.867	0.792	0.703	**0.601**	0.926	0.618	0.751
	AI-KD(DifFIQA(R))	0.890	0.765	**0.702**	0.627	**0.886**	0.571	0.740
Same-Model	**CosFace**							
		Adience	**LFW**	**CPLFW**	**CFP-FP**	**CALFW**	**XQLFW**	$\overline{pAUC}$
	FaceQAN [9]	0.895	0.738	0.785	0.651	0.962	0.594	0.771
	SER-FIQ [8]	0.891	0.846	0.822	0.699	0.903	0.727	0.814
	FaceQnet [12]	0.976	0.847	0.918	0.776	0.959	0.914	0.898
	SDD-FIQA [14]	0.891	0.846	0.822	0.699	0.903	0.727	0.814
	DifFIQA(R) [10]	0.877	0.843	0.789	0.733	0.899	**0.572**	0.785
	MagFace [16]	0.919	0.877	0.788	0.699	0.900	0.865	0.841
	CR-FIQA [18]	0.899	0.816	0.770	0.658	**0.889**	0.637	0.778
	AI-KD(SER-FIQ)	0.872	0.891	0.822	0.716	0.909	0.604	0.802
	AI-KD(CR-FIQA)	**0.862**	0.783	0.757	**0.616**	0.925	0.626	0.761
	AI-KD(DifFIQA(R))	0.877	**0.737**	**0.756**	0.638	0.895	0.586	0.748

here the divide appears to widen, as the extended methods achieve a marginally better result than for instance the third-best approach FaceQAN. With the extended SER-FIQ, we once again observe that it outperforms all remaining methods, except for the three front-runners. Overall, the results suggest that the alignment-invariant knowledge distillation not only improves performance when using misaligned samples, but is also beneficial for the performance of the FIQA techniques with properly aligned samples as well.

V. Conclusion

We presented a novel knowledge distillation technique, named AI-KD, which tries to improve the performance of existing FIQA methods on samples aligned with, from the viewpoint of the FIQA method, an unknown face landmark detector. Through extensive experiments, we showed that the proposed method is able to improve results not only on misaligned but also on properly aligned face images.

References

[1] S. Anwarul and S. Dahiya, "A Comprehensive Review on Face Recognition Methods and Factors Affecting Facial Recognition Accuracy," *Proceedings of ICRIC*, 2020.

[2] T. Schlett, C. Rathgeb, O. Henniger, J. Galbally, J. Fierrez, and C. Busch, "Face Image Quality Assessment: A Literature Survey," *CSUR*, 2022.

[3] "ISO/IEC DIS 29794-1, Biometric Sample Quality," standard, International Organization for Standardization (ISO), 2022.

[4] Y. Peng, L. J. Spreeuwers, and R. N. Veldhuis, "Low-Resolution Face Recognition and the Importance of Proper Alignment," *IET Biometrics*, vol. 8, no. 4, 2019.

[5] A. Kumar, A. Kaur, and M. Kumar, "Face Detection Techniques: a Review," *AIR*, vol. 52, 2019.

[6] X. Gao, S. Z. Li, R. Liu, and P. Zhang, "Standardization of Face Image Sample Quality," in *Proceedings of ICB*, Springer, 2007.

[7] K. Nasrollahi and T. B. Moeslund, "Extracting a Good Quality Frontal Face Image From a Low-Resolution Video Sequence," *TCSVT*, 2011.

[8] P. Terhorst, J. N. Kolf, N. Damer, F. Kirchbuchner, and A. Kuijper, "SER-FIQ: Unsupervised Estimation of Face Image Quality Based on Stochastic Embedding Robustness," in *Proceedings of CVPR*, 2020.

[9] Ž. Babnik, P. Peer, and V. Štruc, "FaceQAN: Face Image Quality Assessment through Adversarial Noise Exploration," in *ICPR*, 2022.

[10] Ž. Babnik, P. Peer, and V. Štruc, "DifFIQA: Face Image Quality Assessment Using Denoising Diffusion Probabilistic Models," in *IJCB*, 2023.

[11] L. Best-Rowden and A. K. Jain, "Learning Face Image Quality from Human Assessments," *TIFS*, vol. 13, no. 12, 2018.

[12] 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 ICB*, 2019.

[13] W. Xie, J. Byrne, and A. Zisserman, "Inducing Predictive Uncertainty Estimation for Face Verification," in *Proceedings of BMVC*, 2020.

[14] F.-Z. Ou, X. Chen, R. Zhang, Y. Huang, S. Li, J. Li, Y. Li, L. Cao, and Y.-G. Wang, "SDD-FIQA: Unsupervised Face Image Quality Assessment with Similarity Distribution Distance," in *Proceedings of CVPR*, 2021.

[15] Y. Shi and A. K. Jain, "Probabilistic Face Embeddings," in *CVPR*, 2019.

[16] Q. Meng, S. Zhao, Z. Huang, and F. Zhou, "MagFace: A Universal Representation for Face Recognition and Quality Assessment," in *Proceedings of CVPR*, 2021.

[17] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive Angular Margin Loss for Deep Face Recognition," in *CVPR*, 2019.

[18] F. Boutros, M. Fang, M. Klemt, B. Fu, and N. Damer, "CR-FIQA: Face Image Quality Assessment by Learning Sample Relative Classifiability," in *Proceedings of CVPR*, 2023.

[19] M. Kim, A. K. Jain, and X. Liu, "AdaFace: Quality Adaptive Margin for Face Recognition," in *Proceedings of CVPR*, 2022.

[20] E. Eidinger, R. Enbar, and T. Hassner, "Age and Gender Estimation of Unfiltered Faces," *IEEE TIFS*, vol. 9, no. 12, 2014.

[21] 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," tech. rep., UMass, Amherst, Oct. 2007.

[22] T. Zheng and W. Deng, "Cross-Pose LFW: A Database for Studying Cross-Pose Face Recognition in Unconstrained Environments," Tech. Rep. 18-01, BUPT, February 2018.

[23] 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 *Proceedings of WACV*, 2016.

[24] T. Zheng, W. Deng, and J. Hu, "Cross-Age LFW: A Database for Studying Cross-Age Face Recognition in Unconstrained Environments," *CoRR*, vol. abs/1708.08197, 2017.

[25] M. Knoche, S. Hormann, and G. Rigoll, "Cross-Quality LFW: A Database for Analyzing Cross-Resolution Image Face Recognition in Unconstrained Environments," in *Proceedings of FG*, 2021.

[26] J. Hernandez-Ortega, J. Galbally, J. Fiérrez, and L. Beslay, "Biometric Quality: Review and Application to Face Recognition with FaceQnet," *arXiv preprint arXiv:2006.03298*, 2020.

[27] L. Qin, M. Wang, C. Deng, K. Wang, X. Chen, J. Hu, and W. Deng, "SwinFace: A Multi-Task Transformer for Face Recognition, Expression Recognition, Age Estimation and Attribute Estimation," *TCSVT*, 2023.

[28] J. Dan, Y. Liu, H. Xie, J. Deng, H. Xie, X. Xie, and B. Sun, "TransFace: Calibrating Transformer Training for Face Recognition from a Data-Centric Perspective," in *Proceedings of ICCV*, 2023.

[29] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild," in *CVPR*, 2020.

[30] J. Xiang and G. Zhu, "Joint Face Detection and Facial Expression Recognition with MTCNN," in *Proceedings of ICISCE*, IEEE, 2017.

[31] V. Kazemi and J. Sullivan, "One Millisecond Face Alignment with an Ensemble of Regression Trees," in *Proceedings of CVPR*, 2014.