QMagFace: Simple and Accurate Quality-Aware Face Recognition

Philipp Terhörst[124], Malte Ihlefeld[3], Marco Huber[23], Naser Damer[2], Florian Kirchbuchner[23],
Kiran Raja[1], Arjan Kuijper[23]

[1]Norwegian University of Science and Technology, Gjøvik, Norway
[2]Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany
[3]Technical University of Darmstadt, Darmstadt, Germany
[4]Paderborn University, Paderborn, Germany
Email: philipp.terhoerst@uni-paderborn.de

Abstract

In this work, we propose QMagFace, a simple and effective face recognition solution (QMagFace) that combines a quality-aware comparison score with a recognition model based on a magnitude-aware angular margin loss. The proposed approach includes model-specific face image qualities in the comparison process to enhance the recognition performance under unconstrained circumstances. Exploiting the linearity between the qualities and their comparison scores induced by the utilized loss, our quality-aware comparison function is simple and highly generalizable. The experiments conducted on several face recognition databases and benchmarks demonstrate that the introduced quality-awareness leads to consistent improvements in the recognition performance. Moreover, the proposed QMagFace approach performs especially well under challenging circumstances, such as cross-pose, cross-age, or cross-quality. Consequently, it leads to state-of-the-art performances on several face recognition benchmarks, such as 98.50% on AgeDB, 83.95% on XQLFQ, and 98.74% on CFP-FP. The code for QMagFace is publicly available[1].

1. Introduction

Face recognition systems are spreading worldwide and are increasingly involved in unconstrained environments [29]. In these environments, these systems have to deal with large variabilities, such as challenging illuminations, poses, and expressions, that might result in incorrect matching decisions [20][19]. The face image quality of a sample is defined as its utility for recognition [16][34][5][42] and measures the impact of these variabilities on the face recognition performance. Previous works either do not employ face image quality information during comparison

---
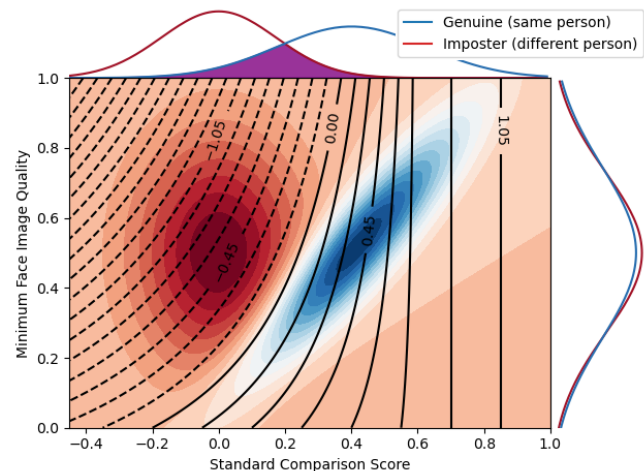
[1]https://github.com/pterhoer/QMagFace

Figure 1: Visualisation of the proposed quality-aware comparison score. The genuine (blue) and imposter (red) distributions are shown with respect to the comparison scores and their face image qualities. While the quality distributions (right) are very similar, the score distributions (top) are strongly overlapping. The proposed quality-aware comparison score is shown via black isolines. Dashed lines indicate negative scores and solid lines positive. The quality-awareness increases the separability of both distributions.

[27][44][9][18] or include quality estimates in comparison process that are not inherently suitable for such a task [40][26]. While the first case results in a loss of valuable information for the comparison, in the second case, the limiting factor lies in the utilized quality estimates.

In this work, we propose QMagFace, a solution that combines a quality-aware comparison function with a face recognition model trained with a magnitude-aware angular margin (MagFace) loss. Incorporating model-specific face image qualities in the comparison process aims at enabling

an improved face recognition performance even under challenging circumstances. Exploiting the linear relationship between the qualities and their comparison scores that is induced by the MagFace loss, our quality-aware comparison function is simple but effective.

In Figure 1, the effect of the proposed quality-aware scoring function is visualized. Using the standard comparison score, the genuine (same person) and imposter (different person) distributions are strongly overlapping (see top plot). Even if their respective quality distributions are very similar (right plot), combining both information with the proposed quality-aware scoring function increases the separability leading to more reliable comparison scores for matching. Especially for lower comparisons and quality scores, the proposed approach adapts the scores more strongly to increase the accuracy. For higher qualities and comparison scores, the proposed solution does not alter the score since high comparison scores imply face pairs of high quality. Consequently, QMagFace is especially effective when dealing with challenging circumstances such as cross-pose or cross-age.

The experiments were conducted on four face recognition databases and six benchmarks. The results demonstrate (a) a constant improvement in the face recognition performance compared to standard comparison scores over a wide range of false match rates, (b) the suitability of the used linear function for quality weighting, and (c) a high generalizability of the proposed approach despite changes in backbone architecture, training databases, and evaluation benchmarks. Moreover, the QMagFace consistently reaches high performances in video-based recognition tasks and achieves state-of-the-art results on three of the four image-based face recognition benchmarks. Especially under challenging circumstances, such as cross-pose, cross-age, or cross-quality, QMagFace achieved high performances.

2. Related Works

2.1. Face Image Quality Assessment

Driven by the international standards, such as ISO/IEC 19794-5 [20] and ICAO 9303 [19], the first generation of face image quality assessment (FIQA) approaches are built on human perceptive image quality factors [14][13][38][34][2][1] [11]. The second generation of FIQA approaches [47][3][7][23][5][16][33][48] consists of supervised learning algorithms based on human or artificially constructed quality labels. However, humans may not know the best characteristics for face recognition systems and artificially labelled quality values, derived from comparison scores, rely on error-prone labelling mechanisms.

The third generation of FIQA approaches [42] completely avoids the use of quality labels by utilizing the face recognition networks themselves. In 2020, Terhörst et al.

[42] proposed stochastic embedding robustness for FIQA (SER-FIQ). This concept measures the robustness of a face representation against dropout variations and uses this measure to determine the quality of a face. It avoids the need for training and takes into account the decision patterns of the deployed face recognition model. In 2021, Meng et al. [31] proposed a class of loss functions that include magnitude-aware angular margins encoding the quality into the face representation. Training with this loss results in a face recognition model that produces embeddings whose magnitudes can measure the FIQ of their faces.

While the first two generations of FIQA methods aim to assess the utility of an image for face recognition in general, the third generation aims at determining model-specific quality values. Therefore, we assert that these methods, SER-FIQ [42] and MagFace [31], have the highest potential for improving the recognition performance. SER-FIQ can be applied to arbitrary face recognition networks but produce a complex relation between FIQ and comparison scores. On the other hand, MagFace produces a linear relationship between the qualities and the comparison scores, as we will show in Section 5.4, and thus, it is more suitable for a generalisable enhancement of the recognition performance.

2.2. Face Recognition

In recent years, face recognition is strongly driven by advances in deep representation learning. Early works rely on metric-learning based losses [31], such as contrastive loss [8], triplet loss [37], or penalty angular-margin losses [45][9]. However, due to the combinatorial explosion in the number of face triplets needed for training, the research focus shifted to classification-based approaches. These approaches are typically based on softmax and aim at classifying on a closed-set of identities during training and utilize the previous layer as a feature extractor for unseen faces. Combining the softmax activation with cross-entropy loss, most face recognition losses $\mathcal{L}$ are of the form

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} \log\left(\mathcal{L}_i\right) \quad \text{with} \tag{1}$$

$$\mathcal{L}_i = \frac{e^{r \cos(m_1 \theta_{y_i} + m_2) - m_3}}{e^{r \cos(m_1 \theta_{y_i} + m_2) - m_3} + \sum_{j=1, j \neq y_i}^{N} e^{r \cos(\theta_{y_j})}}. \tag{2}$$

Here, the training set contains N samples and Θ_{y_i} refers to the angle between between last layer weight-vector and the normalized feature vector x_i (with $\|x_i\|_2 = r$). For training without margins $m_1 = 1, m_2 = m_3 = 0$, this refers to a simplified softmax loss. In SphereFace [27], a multiplicative angular margin is deployed with $m_1 = \alpha$ and $m_2 = m_3 = 0$. For keeping the cosine margin penalty $m_3 = \alpha$ ($m_1 = 1$ and $m_2 = 0$), this refers to CosFace [44] and for penalizing an angular margin $m_2 = \alpha$ ($m_1 = 1$ and

$m_3 = 0$), this refers to the ArcFace [9] loss. However, these losses select a fixed margin α assuming that the samples are equally distributed in the embedding space around the class centers, which is not true when dealing with largely intra-class variations. To solve this problem, solutions based on variable margins are proposed. In [6], Boutros et al. proposed ElasticFace in which random margins are drawn from a Gaussian distribution $\mathcal{N}$ in each training iteration. This allows ElasticFace to extract and retract the margins individually for each class (e.g. $m_1 = 1$ and $m_2 \sim N$). Similarly, CurricularFace [18] addresses easy samples in the early training stage and hard ones in the later stage adapting m_2 while keeping $m_1 = 1$ and $m_3 = 0$. In MagFace [31], a magnitude-aware angular margin $m(r)$ (with $m_1 = 1$ and $m_3 = 0$) with a regularization $g(r)$ is proposed that aims at including the utility (face image quality) of a sample in the margin. While the regularization $g(r)$ rewards samples with large magnitudes r, $m(r)$ is a simple linear function that aims at concentrating high-quality samples in a small region around the class centers. This results in more suitable margins that are based on the utility of the samples and are encoded in the magnitude of the embeddings.

2.3. Quality-Aware Face Recognition

The qualities of face images are often used in video-to-video recognition tasks where a set of images from one person is matched to a set of images from another [28][33][48][36][49]. There, the quality of each image is used for a weighted aggregation of information. For recognition tasks based on single images, only a few works included the face image quality to enhance the face recognition performance. In EQFace [26], Liu et al. attached a quality-prediction network on a face recognition model to include the qualities in the training process. However, this method is limited by computationally-expensive training that is not end-to-end.

In [40], Shi and Jain proposed probabilistic face embeddings (PFE). Representing face images as Gaussian distributions in the embedding space, the variance of each feature is interpreted as its uncertainty and thus, as its quality. For comparison, they make use of a mutual-likelihood score to include the quality in the comparison score. However, performance is determined by the quality estimate that is limited by (a) the used uncertainty estimation module which is trained separately from the face recognition network and (b) the assumption that each feature can be independently represented as a Gaussian Process.

In contrast to previous works, we make use of model-specific quality estimates that were linearly included in an end-to-end fashion. Consequently, this allows our proposed solution to work simplistically and more efficiently.

3. Methodology

The main contribution of this work, QMagFace, combines a quality-aware comparison function with a face recognition model trained with MagFace loss [31]. Including the model-specific face image quality in the comparison process aims to consistently improve the face recognition performance, especially under challenging conditions such as cross-pose, cross-age, or cross-quality. Moreover, the proposed quality-aware scoring function can be robustly trained on any face recognition network based on the MagFace loss as it will be shown in Section 5.4.

3.1. Quality-Aware Comparison Scoring

In face biometrics, a comparison score reflects the identity similarity of two face images. This reflection of similarity is more accurate when the compared samples are of high quality [16]. The biometric quality of a face image is defined as its utility for recognition [16][34][5][42]. In [41], it was shown that model-specific quality assessment reflects challenging situations for the face recognition model, such as comparisons with strong variations in pose or age. Similar to SER-FIQ [42], the qualities of MagFace networks [31] utilize the deployed face recognition system and thus, strongly reflect the decision patterns and model biases. Consequently, we propose a simple, but effective, comparison function that includes these model-specific quality values to enhance the accuracy and robustness of the face recognition system.

Given a face recognition model $\mathcal{M}$ trained with Mag-Face loss and two face images I_1 and I_2, their embeddings are given by $e_1 = \mathcal{M}(I_1)$ and $e_2 = \mathcal{M}(I_2)$ together with the corresponding face image qualities $q_1 = \|e_1\|_2$ and $q_2 = \|e_2\|_2$ encoded through the vector length. The standard comparison score $s = \cos(e_1, e_2)$ is defined over cosine similarity of these templates and thus, represents the angular similarity between both templates. However, comparisons with low-quality images affect the comparison scores and thus, needs to be adjusted. The proposed quality-aware comparison score $\hat{s}$ is defined as

$$\hat{s}(s, q_1, q_2) = \omega(s) * \min\{q_1, q_2\} + s \qquad (3)$$

with a quality-weighting function

$$\omega(s) = \min\left\{0,\ \beta * s - \alpha\right\}. \qquad (4)$$

This comparison function consists of only two trainable parameters (α and β) and thus, can be robustly trained.

Since the biometric sample quality is included linearly in the MagFace loss (through $m(r)$), we assume a linear relationship between the face image qualities and their comparison scores. In Section 5.4, we will demonstrate the suitability of a linear function for the quality weighting. We further assume that the score-adjustment is dependent on the

lowest quality of the comparison and that only comparisons with at least one low-quality sample needs to be adjusted. Please note that the qualities q_1 and q_2 are not easily exchangeable with other FIQ methods since these need to be model-specific with regard to the FR model and require a linear relationship between quality and comparison scores.

A high similarity score s can only be achieved through the comparison of two high-quality samples. In this case, the similarity is well reflected in the comparison score s and thus, no quality-based score adjustments ($\omega = 0$) is needed. A lower comparison score might result from the degradation of a pair with at least one low-quality sample. In this case, the similarity is altered by the sample quality and our proposed function adjusts the score based on the quality. Consequently, if two comparisons result in similar comparison scores $s_1 \approx s_2$ (with $\omega(s) < o$) but have different minimum qualities $q_1^{min} \gg q_2^{min}$, the score with the higher quality undergoes a stronger adjustment and thus, results in a lower quality-aware comparison score ($\hat{s}_1 < \hat{s}_2$).

3.2. Training the Quality-Weighting Function

For training, the comparison scores of the training set $\mathcal{S} = \mathcal{S}_G \cup \mathcal{S}_I$ are separated into genuine and imposter comparisons with the corresponding minimal qualities $\mathcal{Q}^{min}$ of the sample pairs. The training process is divided into three steps and aims at learning the quality weighting function $\omega(s)$ as shown in Algorithm 1.

In the first step, we define the optimal quality weight $w_{opt}(t)$ for a given threshold t. This is given through

$$w_{opt}(t) = \underset{\omega}{\operatorname{argmin}} \frac{1}{|\mathcal{S}_G|} \sum_{s \in \mathcal{S}_G} \Theta(t - \bar{s}(\omega, s)) \qquad (5)$$

where $\Theta(\cdot)$ describes the Heaviside function and

$$\bar{s}(\omega, s) = \sigma(\omega * Q^{min}(s) + s). \qquad (6)$$

is a quality-aware scoring function given by ω and s that is scaled to the range of $[0, 1]$, similar to the range of comparison scores, with a sigmoid function. This optimization aims at minimizing the FNMR at a threshold t through including face image quality information.

In the second step, a relevant threshold range $\mathcal{T}$ needs to be defined that represents the target FMR range. In this work, we choose the range from $FMR^{max} = 10^{-2}$ to $FMR^{min} = 10^{-5}$ to cover a wide variety of potential applications and due to the amount of training data available (in the order of 10^5 images). The relevant threshold range $\mathcal{T} = [t(FMR^{max}), t(FMR^{min})]$ is determined by finding the threshold t that corresponds to the required FMR on the quality-aware scores on the training data. This can be determined by

$$t(FMR) = \underset{t}{\operatorname{argmin}} \left| FMR - \frac{1}{|\mathcal{S}_I|} \sum_{s \in \mathcal{S}_I} \Theta(\bar{s}(\omega_{opt}(t), s) - t) \right|.$$

The third step aims to learn the quality weighting function $\omega(s)$. Since the quality is included linearly in the Mag-Face loss, a linear relationship between the importance of the quality $q(s)$ and its comparison score s is assumed. Therefore, we model the quality weights through a simple linear function $\omega(s) = \beta * s - \alpha$. The parameters α and β can be learned by solving the following optimization

$$\min_{\alpha, \beta} \sum_{t \in \mathcal{T}} \left(\omega_{opt}(t) + \alpha - \beta * t \right)^2, \qquad (7)$$

resulting in the optimal parameters $\hat{\beta}$ and $\hat{\alpha}$

$$\hat{\beta} = \frac{\sum_{t \in \mathcal{T}} (t - \bar{t})(\omega_{opt}(t) - \bar{\omega}_{opt})}{\sum_{t \in \mathcal{T}} (t - \bar{t})^2}, \qquad (8)$$

$$\hat{\alpha} = \hat{\beta} * \bar{t} - \bar{\omega}_{opt} \qquad \text{with} \qquad (9)$$

$$\bar{\omega}_{opt} = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \omega_{opt}(t) \quad \text{and} \quad \bar{t} = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} t. \qquad (10)$$

Algorithm 1: Quality-weighting function

Input: $\mathcal{S}_G, \mathcal{S}_I, Q^{min}, FMR^{min}, FMR^{max}$
Output: $\omega(s)$
/* Determine optimal weights */
1 for $t \in \mathcal{T} \in [t(FMR^{max}), t(FMR^{min})]$ **do**
2 $w_{opt}(t) \leftarrow$
      $\operatorname{argmin}_\omega \frac{1}{|\mathcal{S}_G|} \sum_{s \in \mathcal{S}_G} \Theta(t - (\omega * Q^{min}(s) + s))$
/* Learns quality weighting
   function parameters */
3 $\bar{\omega}_{opt} \leftarrow \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \omega_{opt}(t)$
4 $\bar{t} \leftarrow \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} t$
5 $\hat{\beta} = \frac{\sum_{t \in \mathcal{T}} (t - \bar{t})(\omega_{opt}(t) - \bar{\omega}_{opt})}{\sum_{t \in \mathcal{T}} (t - \bar{t})^2}$
6 $\hat{\alpha} = \hat{\beta} * \bar{t} - \bar{\omega}_{opt}$
/* Define the quality function */
7 $\omega(s) = \min \left\{ 0, \hat{\beta} * s - \hat{\alpha} \right\}$
8 return $\omega(s)$

---

4. Experimental Setup

4.1. Databases and Benchmarks

To compare the performance of the proposed QMagFace approach with ten recent state-of-the-art approaches six face recognition benchmarks are used, LFW [17], AgeDB-30 [32], CFP-FP [39], XQLFW [25], IJB-B [46], and IJB-C[30].

LFW [17] is a face verification benchmark containing 13k images of over 5k identities. In the benchmark experiments, we followed the standard protocol [17] using the 6k

predefined comparison pairs. Moreover, we conducted the experiments on three more challenging benchmarks representing the issues of cross-age (AgeDB [32]), cross-pose (CFP-FP [39]), and cross-quality (XQLFW [25]). AgeDB [32] is unconstrained face recognition benchmark for age-invariant face verification. It contains over 16k images of over 5k identities. In the experiments, we follow the protocol of AgeDB-30 since it is the most reported and challenging one for AgeDB consisting of age gaps of over 30 years. CFP-FP [39] is a face recognition benchmark that addresses the issue of comparing frontal to profile face images. In our experiments, we followed the evaluation protocol of [39] containing 3500 genuine pairs and 3500 imposter pairs. XQLFW [25] is a benchmark that addresses the problem of cross-quality comparisons in face recognition. The protocol defines 6k face image pairs based on the LFW database. However, for each pair, one face image is of much lower quality than the other face. The IARPA Janus Benchmark-B (IJB-B) [46] contains around 21k images and 55k frames from over 7k videos of 1,845 identities. In the experiment, we follow the standard evaluation protocol using around 10k genuine and 8M impostor comparisons. The IARPA Janus Benchmark–C (IJB-C) [30] extends on the IJB-B by adding more identities. In total, it consists of 31k images with over 117k frames of over 11k videos from 3531 identities. The verification protocol considers over 19k genuine and 16M imposter comparisons. In contrast to IJB-B, IJB-C focuses more on occlusion and diversity of subject occupation to improve the representation of the global population.

Besides reporting the verification accuracy based on benchmarks, we make use of four face recognition datasets to cover a much wider range of possible decision thresholds and thus, to cover more potential applications. Morph [22] consists of 55k face images from over 13k subjects. The images are frontal and of high quality. LFW [17] contains 13k images of over 5k identities and the ColorFeret database [35] consists of 14k high-resolution face images from over 1k different individuals. The data possess a variety of poses (from frontal to profile) and facial expressions under well-controlled conditions. The Adience dataset [12] consists of 26k images from over 2k different subjects. The images of the Adience dataset possess a wide range in terms of image quality. In the supplementary material, we included a more detailed discussion, such as on the licenses.

4.2. Evaluation Metrics

Following the international standard for biometric verification evaluation [21], we report the face verification error in terms of false non-match rate (FNMR) at fixed false match rate (FMR). Moreover, we report the equal error rate (EER) and the area under curve (AUC) of the receiver operating characteristic (ROC) curve. The EER equals the FMR

at the threshold where FMR = FNMR and is well known as a single-value indicator of the verification performance. In our experiments, we report the face verification performance over a wide range of FMRs to cover a variety of potential applications. On the benchmarks, we follow the mentioned protocols and report the verification accuracy to be comparable with previous works.

4.3. Face Recognition Models

In the experiments, the proposed QMagFace approach is built on three pre-trained models[2] based on MagFace loss released by the authors [31]. These were trained on the MS1MV2 database [15] and are based on iResNet-18, iResNet-50, and iResNet-100 backbones [10]. In the following, we use the name of the loss function and the model trained with it interchangeably to keep this work easily comprehensible.

Table 1: Learned parameters

Model	Learned parameters	
	α	β
QMagFace-18	0.092861	0.135311
QMagFace-50	0.065984	0.103799
QMagFace-100	0.077428	0.125926

The parameters α and β needed for the proposed QMagFace approach are trained on the Adience dataset [12] due to the large quality variance in its samples and to create a generalizable approach by using the estimated parameters on this dataset. However, the training process is robust and thus, the choice of the training database only affect the performance minimally as we will show in Section 5.4. Due to the simplicity of the proposed approach, the learned parameters are shown in Table 1. To extract a face embedding from a given face image, the image is aligned, scaled, and cropped as described in [31]. Then, the preprocessed image is passed to the face recognition models to extract the feature embedding.

4.4. Investigations

The proposed approach is analysed in three steps. First, we report the performance of QMagFace on six face recognition benchmarks against ten recent state-of-the-art methods in image- and video-based recognition tasks to provide a comprehensive comparison with state-of-the-art. Second, we investigate the face recognition performance of QMagFace over a wide FMR range to show its suitability for a wide variety of applications and to demonstrate that the quality-aware comparison score constantly enhances the

---

[2]https://github.com/IrvingMeng/MagFace (Apache License 2.0)

recognition performance. Third, we analyse the optimal quality weight over a wide threshold range to demonstrate the robustness of the training process and the generalizability of the proposed approach.

5. Results

5.1. Performance on Single-Image Benchmarks

To demonstrate that the proposed QMagFace approach achieves state-of-the-art performance in image-to-image face recognition tasks, the proposed method is compared against ten recent face recognition models on four benchmarks. For PFE[3] [40], ArcFace[4] [9], and the MagFace variants [31], we used the implementations released by the authors. The remaining benchmark results are taken from [6][26]. In Table 2, the face recognition performances of these are shown. On the LFW benchmark, the proposed QMagFace approach based on the iResNet-100 backbone achieved a performance of 99.83%, which is close to the state-of-the-art performance of 99.85%. On more challenging and less-saturated benchmarks, the proposed approach achieves state-of-the-art performance. This includes 98.50% on cross-age face recognition (AgeDB), 98.74% on cross-pose face recognition (CFP-FP), and 83.95% on cross-quality face recognition (XQLFW). Since the FIQ captures these challenging conditions and the quality values represent the utility of the images for our specific network, the proposed quality-aware comparison score can specifically address the circumstance and their effect on the network. Consequently, it performs highly accurate in the cross-age, cross-pose, and cross-quality scenarios and achieves state-of-the-art performances.

5.2. Performance on Video-Based Benchmarks

In Table 3, the video-based face recognition performance is analysed based on IJB-B and IJB-C. The FNMR is investigated over various FMRs. The performances of most state-of-the-art approaches are taken from the respective works. Since some original works did not investigate the performance on IJB-B/C, the remaining performances are taken from [31] and [6]. For creating an embedding with the corresponding quality-value for a video, the unit-sized embeddings and qualities per frame are aggregated by a quality-weighted sum. Introducing quality-awareness to the MagFace-100 model generally reduces the recognition error. Especially for FMRs up to 10^{-3}, QMagFace achieves state-of-the-art performance. Despite the effectiveness of QMagFace for image-to-image face recognition tasks, the quality of a video-embedding does currently not well rep-

[3] https://github.com/seasonSH/ Probabilistic-Face-Embeddings (MIT License)
[4] https://github.com/deepinsight/insightface (MIT License)

Table 2: Image-to-image face recognition performance on four benchmarks reported in terms of benchmark accuracy (%). The highest performance is marked bold. The proposed approach, QMagFace-100, achieves state-of-the-art face recognition performance, especially in cross-age (AgeDB), cross-pose (CFP-FP), and cross-quality (XQLFW) scenarios.

| Model | Benchmark | | | |
	AgeDB	CFP-FP	LFW	XQLFW
SphereFace [27]	98.17	86.84	99.67	-
CosFace [44]	98.17	98.26	99.78	-
PFE [40]	96.90	97.49	99.80	79.80
ArcFace [9]	98.07	97.31	99.77	79.73
GroupFace [24]	98.28	98.63	**99.85**	-
CurricularFace [18]	98.32	98.37	99.80	-
ElasticFace-Arc [6]	98.35	98.67	99.80	81.87
ElasticFace-Cos [6]	98.27	98.61	99.82	83.78
EQFace [26]	-	98.34	99.80	-
MagFace-18 [31]	93.37	93.11	99.22	69.55
QMagFace-18 (ours)	92.98	94.00	99.30	68.60
MagFace-50 [31]	97.60	97.33	99.72	80.60
QMagFace-50 (ours)	97.88	97.74	99.73	80.63
MagFace-100 [31]	98.18	98.36	99.73	83.90
QMagFace-100 (ours)	**98.50**	**98.74**	99.83	**83.95**

resent its utility for recognition. While the quality corresponding to an embedding for a single frame has a high correlation with the true utility of this frame, the same does not apply for the (weighted-sum) aggregated embedding and thus for the quality of the video. This needs to be addressed by future work.

5.3. Full Performance Analysis

In Table 2, it was already shown that proposed quality-aware face recognition approach leads to stable improvements in the recognition performance. In this section, we will demonstrate these improvements for a wide FMR range. Table 4 shows the recognition performance of MagFace and QMagFace variants. To cover a wide range of potential applications, the performance is analysed over a wide range of decision thresholds (ranging from FMR of 10^{-1} to 10^{-5}) for three databases. The analysis involved over 300k/160k/1.1M comparisons on the Color-Feret/LFW/Morph database. For three backbones (iResNet-18/50/100), the performance of MagFace and the proposed QMagFace approach is compared.

For QMagFace-18, 19 out of 21 scenarios showed an improved recognition performance while for QMagFace-50, 20 out of 21 scenarios showed a performance enhancement. The three cases with a decreased performance took place on ColorFeret, which involves many challenging frontal to pro-

Table 3: Video-based face recognition performance. The performance [%] is reported in terms of FNMR at different FMRs. The best is marked bold. The asterisk (*) denotes a method that is optimized for video-based recognition. In nearly all cases, the quality-awareness increases the performance of MagFace. For FMRs up to 10^{-3}, the proposed quality-aware solution performs best despite that aggregated quality does not reflect well the combined embeddings per video.

	IJB-B			IJB-C		
FNMR at FMR of	10^{-2}	10^{-3}	10^{-4}	10^{-2}	10^{-3}	10^{-4}
CosFace [44]	-	-	5.99	-	-	4.44
PFE [40]	-	-	-	2.83	4.51	6.75
ArcFace [9]	2.47	3.84	5.77	1.82	2.79	4.36
GroupFace [24]	-	-	5.07	-	-	3.74
CurricularFace [18]	-	-	5.20	-	-	3.90
ElasticFace-Arc [6]	-	-	4.78	-	-	3.51
ElasticFace-Cos [6]	-	-	**4.70**	-	-	**3.43**
EQFace* (QWFA) [26]	2.76	3.69	5.12	1.78	2.55	3.62
EQFace (QW) [26])	2.38	3.52	5.49	1.70	2.61	4.16
MagFace-100 [31]	2.57	3.81	5.50	1.74	2.76	4.03
QMagFace-100 (ours)	**2.28**	**3.52**	5.30	**1.49**	**2.38**	3.81

file face comparisons. Since the recognition performance of MagFace-18 and MagFace-50 is much lower than the models based on iResNet-100, the performance in estimating the model-specific face image quality correctly is lower as well. Therefore, the QMagFace approaches that make use of these quality estimates become less accurate when the quality estimate is failed by a large degree.

For QMagFace-100, the performance, and thus the quality estimation, is higher. Consequently, the proposed QMagFace-100 approach leads to strong performance improvements in all investigated cases.

5.4. Robustness Analysis

Lastly, we demonstrate (a) the suitability of choosing a linear quality-weighting function and (b) the generalizability of the QMagFace solution.

In Figure 2, the correlation between the optimal quality weight and different decision thresholds are shown for different MagFace models and databases. For each of the four databases, the optimal quality weight ω_{opt} is computed with Eq. 5 for several thresholds $t \in \mathcal{T}$. These weights show the optimal solutions for a given database that can be achieved by using the quality-aware score function $\hat{s}(s, q_1, q_2)$ from Eq. 3. Moreover, a linear function is fitted through these points and shown in the same color.

For all three models (Figure 2a, 2b, 2c), two observations are made. First, the optimal quality weights ω_{opt} follow a linear function with respect to the decision threshold t. This is observed for all MagFace models and on each database, proving the suitability of our linear quality weight function $\omega(s)$ from Eq. 4. Second, for each model, the optimal quality weight functions are similar. The only exception is the quality weight function for MagFace-100 that is optimized on LFW. In this case, the database turns out to be too easy to train the model effectively since even for an FMR of 10^{-5} the decision threshold is below 0.5. However, in all the other investigated cases, the relation between the optimal quality weights and the decision thresholds is similarly independent of the analysed database. Utilising these databases for training, QMagFace will lead to similar matching decisions demonstrating the robustness of the QMagFace training process. Moreover, it indicates a high generalizability since the learned function on one database is very similar to the optimal functions of the others.

6. Limitations and Ethical Considerations

Despite the high generalizability and the effectiveness of QMagFace for unconstrained face recognition, the approach has to deal with two limiting factors. First, the additional quality information is most beneficial for images of lower quality and thus, the performance improvements of QMagFace decrease for very low FMRs, such as 10^{-7}. Second, for more effective video-based recognition, a more suitable quality aggregation is needed. For single images, the quality of an embedding well reflects its utility for recognition. However, this does not apply when fusing the frames of a video with the corresponding qualities. To more efficiently exploit the quality information in video-based recognition with QMagFace, future works need to focus on more advanced quality-based fusion techniques for video frames.

While the proposed quality-awareness approach might strongly improve unconstrained face recognition for the sake of higher security or convenience, we want to point out the importance of unbiased quality estimates for fair face recognition. The use of biased quality estimates might lead to unfair, and thus discriminatory, matching decisions depending on demographic and non-demographic factors of their users [43].

7. Conclusion

For recognition in unconstrained environments, FR systems have to deal with challenging situations, such as different illuminations, poses, and expressions. Previous works either focused on learning margin-based approaches while not considering FIQ information or included non-inherently fit quality estimates. In this work, we proposed QMagFace, a simple and robust quality-aware FR approach. It integrates model-specific FIQ information in the comparison process to allow a more accurate performance under challenging situations, such as cross-pose or cross-age.

Table 4: Face recognition performance reported in terms of FNMR [%] over a wide range of FMRs. The MagFace and the proposed QMagFace approach are compared for three backbones on three databases. The better values between both approaches are highlighted in bold. In general, the proposed quality-aware solutions constantly improve the performance, often by a large margin. This is especially true for QMagFace based on the iResNet-100 backbone.

| Database | Model | EER | FNMR at FMR | | | | | AUC |
			10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	
ColorFeret	MagFace-18	5.312	4.067	9.193	**22.094**	83.968	**97.517**	98.48
	QMagFace-18	**4.232**	**3.068**	**6.902**	22.951	**82.531**	97.723	**98.65**
	MagFace-50	3.635	2.553	5.056	7.560	12.393	**22.416**	99.09
	QMagFace-50	**2.941**	**1.464**	**4.173**	**6.832**	**12.247**	23.426	**99.55**
	MagFace-100	2.629	1.789	3.297	4.791	7.523	16.909	99.24
	QMagFace-100	**2.060**	**0.950**	**2.616**	4.409	**7.145**	**16.454**	**99.67**
Morph	MagFace-18	0.883	0.813	0.873	1.185	2.189	50.892	99.43
	QMagFace-18	**0.843**	**0.779**	**0.834**	**1.036**	**1.908**	**40.070**	**99.53**
	MagFace-50	**0.788**	0.784	0.825	0.832	0.843	0.894	99.61
	QMagFace-50	0.821	**0.473**	**0.812**	**0.826**	**0.835**	**0.880**	**99.84**
	MagFace-100	0.848	0.777	0.814	0.824	0.834	0.848	99.58
	QMagFace-100	**0.773**	**0.363**	**0.760**	**0.817**	**0.829**	**0.840**	**99.88**
LFW	MagFace-18	1.057	0.324	1.096	3.710	9.613	20.163	99.86
	QMagFace-18	**0.724**	**0.186**	**0.607**	**2.324**	**7.282**	**14.377**	**99.93**
	MagFace-50	0.349	0.110	0.290	0.462	0.586	0.821	99.97
	QMagFace-50	**0.332**	**0.035**	**0.172**	**0.407**	**0.517**	**0.752**	**99.99**
	MagFace-100	0.277	0.159	0.255	0.297	0.441	0.621	99.94
	QMagFace-100	**0.195**	**0.145**	**0.172**	**0.221**	**0.331**	**0.517**	**99.95**

(a) MagFace-18 (b) MagFace-50 (c) MagFace-100
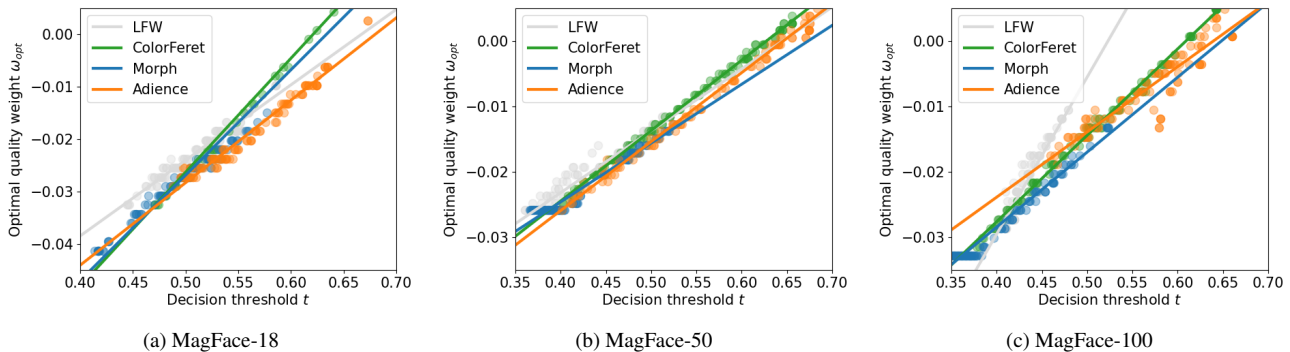
Figure 2: Optimal quality weight for different decision thresholds on four databases. The threshold range reflect FMRs from 10^{-2} to 10^{-5}. Training on different databases lead to similar linear solutions for $\omega(s)$. The results demonstrate that (a) the choice of a linear function is justified and (b) that the learned models have a high generalizability since the weighting function $\omega(s)$ trained on one database is very similar to the optimal functions of the others.

The experiments were conducted on ten FR databases and demonstrated that including the quality-awareness consistently increases the FR performance. Moreover, it was shown that QMagFace reaches competitive recognition results with state-of-the-art solutions. For challenging circumstances, such as cross-pose, cross-age, or cross-quality, QMagFace constantly beat state-of-the-art approaches. Additional experiments indicated a high generalizability of the proposed approach demonstrated the suitability of a linear function for the quality weighting.

References

[1] A. Abaza, M. A. Harrison, and T. Bourlai. Quality metrics for practical face recognition. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 3103–3107, Nov 2012.

[2] A. Abaza, M. A. Harrison, T. Bourlai, and A. Ross. Design and evaluation of photometric image quality measures for effective face recognition. *IET Biometrics*, 3(4):314–324, 2014.

[3] Gaurav Aggarwal, Soma Biswas, Patrick J. Flynn, and Kevin W. Bowyer. Predicting performance of face recognition systems: An image characterization approach. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2011, Colorado Springs, CO, USA, 20-25 June, 2011*, pages 52–59. IEEE Computer Society, 2011.

[4] Tamara L. Berg, Alexander C. Berg, Jaety Edwards, Michael Maire, Ryan White, Yee Whye Teh, Erik G. Learned-Miller, and David A. Forsyth. Names and faces in the news. In *2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004), with CD-ROM, 27 June - 2 July 2004, Washington, DC, USA*, pages 848–854. IEEE Computer Society, 2004.

[5] L. Best-Rowden and A. K. Jain. Learning face image quality from human assessments. *IEEE Transactions on Information Forensics and Security*, 13(12):3064–3077, Dec 2018.

[6] Fadi Boutros, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Elasticface: Elastic margin loss for deep face recognition. *CoRR*, abs/2109.09416, 2021.

[7] J. Chen, Y. Deng, G. Bai, and G. Su. Face image quality assessment based on learning to rank. *IEEE Signal Processing Letters*, 22(1):90–94, Jan 2015.

[8] Sumit Chopra, Raia Hadsell, and Yann LeCun. Learning a similarity metric discriminatively, with application to face verification. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20-26 June 2005, San Diego, CA, USA*, pages 539–546. IEEE Computer Society, 2005.

[9] 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.

[10] Ionut Cosmin Duta, Li Liu, Fan Zhu, and Ling Shao. Improved residual networks for image and video recognition. In *25th International Conference on Pattern Recognition, ICPR 2020, Virtual Event / Milan, Italy, January 10-15, 2021*, pages 9415–9422. IEEE, 2020.

[11] Abhishek Dutta, Raymond N. J. Veldhuis, and Luuk J. Spreeuwers. A bayesian model for predicting face recognition performance using image quality. In *IEEE International Joint Conference on Biometrics, Clearwater, IJCB 2014, FL, USA, September 29 - October 2, 2014*, pages 1–8. IEEE, 2014.

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

[13] M. Ferrara, A. Franco, D. Maio, and D. Maltoni. Face image conformance to iso/icao standards in machine readable travel documents. *IEEE Transactions on Information Forensics and Security*, 7(4):1204–1213, Aug 2012.

[14] Xiufeng Gao, Stan Z. Li, Rong Liu, and Peiren Zhang. Standardization of face image sample quality. In Seong-Whan Lee and Stan Z. Li, editors, *Advances in Biometrics*, pages 242–251, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[15] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III*, volume 9907 of *Lecture Notes in Computer Science*, pages 87–102. Springer, 2016.

[16] 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.

[17] 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, October 2007.

[18] 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.

[19] Machine Readable Travel Documents. Standard, International Civil Aviation Organization, 2015.

[20] Information technology – Biometric data interchange formats – Part 5: Face image data. Standard, International Organization for Standardization, Nov. 2011.

[21] ISO/IEC 19795-1:2006 Information technology — Biometric performance testing and reporting. Standard, International Organization for Standardization, 2016.

[22] Karl Ricanek Jr. and Tamirat Tesafaye. MORPH: A longitudinal image database of normal adult age-progression. In *Seventh IEEE International Conference on Automatic Face and Gesture Recognition (FGR 2006), 10-12 April 2006, Southampton, UK*, pages 341–345. IEEE Computer Society, 2006.

[23] H. Kim, S. H. Lee, and Y. M. Ro. Face image assessment learned with objective and relative face image qualities for improved face recognition. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 4027–4031, Sep. 2015.

[24] Yonghyun Kim, Wonpyo Park, Myung-Cheol Roh, and Jongju Shin. Groupface: Learning latent groups and constructing group-based representations for face recognition. In *2020 IEEE/CVF Conference on Computer Vision and Pat-*

tern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020, pages 5620–5629. Computer Vision Foundation / IEEE, 2020.

[25] Martin Knoche, Stefan Hörmann, and Gerhard Rigoll. Cross-quality LFW: A database for analyzing cross-resolution image face recognition in unconstrained environments. *CoRR*, abs/2108.10290, 2021.

[26] Rushuai Liu and Weijun Tan. Eqface: A simple explicit quality network for face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2021, virtual, June 19-25, 2021*, pages 1482–1490. Computer Vision Foundation / IEEE, 2021.

[27] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 6738–6746. IEEE Computer Society, 2017.

[28] Yu Liu, Junjie Yan, and Wanli Ouyang. Quality aware network for set to set recognition. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 4694–4703. IEEE Computer Society, 2017.

[29] Iacopo Masi, Yue Wu, Tal Hassner, and Prem Natarajan. Deep face recognition: A survey. In *31st SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI 2018, Paraná, Brazil, October 29 - Nov. 1, 2018*, pages 471–478. IEEE Computer Society, 2018.

[30] 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.

[31] 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.

[32] 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 Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 1997–2005. IEEE Computer Society, 2017.

[33] Fu-Zhao Ou, Xingyu Chen, Ruixin Zhang, Yuge Huang, Shaoxin Li, Jilin Li, Yong Li, Liujuan Cao, and Yuan-Gen Wang. Sdd-fiqa: Unsupervised face image quality assessment with similarity distribution distance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7670–7679, June 2021.

[34] P. J. Phillips, J. R. Beveridge, D. S. Bolme, B. A. Draper, G. H. Givens, Y. M. Lui, S. Cheng, M. N. Teli, and H. Zhang. On the existence of face quality measures. In *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–8, Sep. 2013.

[35] P. Jonathon Phillips, Hyeonjoon Moon, Syed A. Rizvi, and Patrick J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(10):1090–1104, 2000.

[36] Rajeev Ranjan, Ankan Bansal, Hongyu Xu, Swami Sankaranarayanan, Jun-Cheng Chen, Carlos Domingo Castillo, and Rama Chellappa. Crystal loss and quality pooling for unconstrained face verification and recognition. *CoRR*, abs/1804.01159, 2018.

[37] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 815–823. IEEE Computer Society, 2015.

[38] H. Sellahewa and S. A. Jassim. Image-quality-based adaptive face recognition. *IEEE Transactions on Instrumentation and Measurement*, 59(4):805–813, April 2010.

[39] 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.

[40] 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.

[41] Philipp Terhörst, Jan Niklas Kolf, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Face quality estimation and its correlation to demographic and non-demographic bias in face recognition. In *2020 IEEE International Joint Conference on Biometrics, IJCB 2020, Houston, TX, USA, September 28 - October 1, 2020*, pages 1–11. IEEE, 2020.

[42] 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.

[43] Philipp Terhörst, Jan Niklas Kolf, Marco Huber, Florian Kirchbuchner, Naser Damer, Aythami Morales, Julian Fiérrez, and Arjan Kuijper. A comprehensive study on face recognition biases beyond demographics. *CoRR*, abs/2103.01592, 2021.

[44] 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 *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 5265–5274. Computer Vision Foundation / IEEE Computer Society, 2018.

[45] Jian Wang, Feng Zhou, Shilei Wen, Xiao Liu, and Yuanqing Lin. Deep metric learning with angular loss. In *IEEE International Conference on Computer Vision, ICCV 2017,*

Venice, Italy, October 22-29, 2017, pages 2612–2620. IEEE Computer Society, 2017.

[46] Cameron Whitelam, Emma Taborsky, Austin Blanton, Brianna Maze, Jocelyn C. Adams, Tim Miller, Nathan D. Kalka, Anil K. Jain, James A. Duncan, Kristen Allen, Jordan Cheney, and Patrick Grother. IARPA janus benchmark-b face dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 592–600. IEEE Computer Society, 2017.

[47] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell. Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. In *CVPR 2011 WORKSHOPS*, pages 74–81, June 2011.

[48] Weidi Xie, Jeffrey Byrne, and Andrew Zisserman. Inducing predictive uncertainty estimation for face recognition. *CoRR*, abs/2009.00603, 2020.

[49] Jiaolong Yang, Peiran Ren, Dongqing Zhang, Dong Chen, Fang Wen, Hongdong Li, and Gang Hua. Neural aggregation network for video face recognition. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 5216–5225. IEEE Computer Society, 2017.