

FunFace: Feature Utility and Norm Estimation for Face Recognition

Žiga Babnik¹, Fadi Boutros², Naser Damer², Deepak Kumar Jain³, Peter Peer¹ and Vitomir Štruc¹

¹University of Ljubljana, Ljubljana, Slovenia

²Fraunhofer Institute for Computer Graphics Research IGD, Germany

³Dalian University of Technology, Dalian, China

Abstract—Face Recognition (FR) is used in a variety of application domains, from entertainment and banking to security, and surveillance. Such applications rely on the FR model to be robust and perform well in a variety of settings. To achieve this, state-of-the-art FR models typically use expressive adaptive margin loss functions, which tie the feature norm to concepts related to sample quality, such as recognizability and perceptual image quality. Recently, through the development of Face Image Quality Assessment (FIQA) techniques, biometric utility has become the preferred measure of face-image quality and has been shown to be a better predictor of the usefulness of samples for face recognition compared to more human-centric aspects, such as resolution, blur, and lighting, tied to general image quality. While image quality expressed through feature norms exhibits a certain level of correlation with biometric utility, it does not fully encapsulate all aspects of utility. To address this point, we propose a new adaptive margin loss, FunFace (Face Recognition Through Utility and Norm Estimation), which incorporates biometric utility, estimated by the Certainty Ratio, into the adaptive margin, taking inspiration from AdaFace. We show that FunFace (when used to train a face recognition model) achieves competitive results to other state-of-the-art FR models on benchmarks containing high-quality samples, while surpassing them on low quality benchmarks. Code will be made publicly available after review.

I. INTRODUCTION

Learning discriminative features is the core focus of Face Recognition (FR) techniques [12]. Through the use of new, larger training datasets [1], [13], [46], and the development of advanced adaptive loss functions [3], [17], [20], [27], models have achieved excellent performance on a variety of benchmark datasets. Most mentioned benchmarks focus on evaluating the model’s performance using images of relatively high quality [16], [30], [44]. Recently, however, more focus has been placed on benchmarks with low-image-quality, replicating difficult security and surveillance scenarios [5], [7], [19]. Domain adaptation (transfer learning) can improve performance in low-quality domains [31], but requires additional processing and can lead to decreased performance in high-quality domains. Adaptive loss functions [3], [20], [27], on the other hand, can facilitate difficult-to-recognize images directly, through adaptive margin functions, enabling excellent performance across all quality-domains.

Modern FR primarily relies on margin-based loss functions, which extend the cross-entropy (softmax) loss, with a simple margin term. The two most commonly used are the angular – ArcFace [9] and the additive – CosFace [41] loss functions, both of which use a static margin value for

This work is supported by the ARIS Young Researcher program. We thank Arnes and the SLING HPC consortium, under the EuroCC 2 project, for providing computational resources.

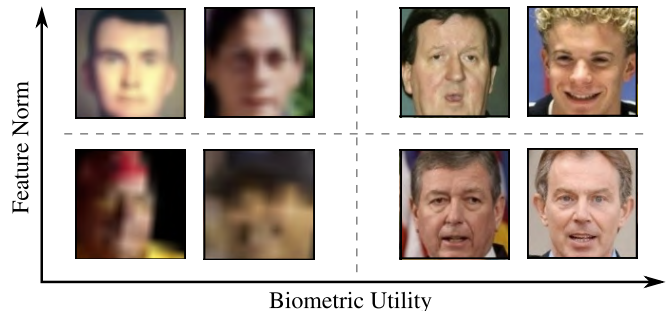


Fig. 1: **Comparison of the feature norms and biometric utility of samples.** Feature norms commonly used by adaptive margin functions [20], [27], have been shown to correlate well with the visual image quality of samples. While visual quality estimates provide a good source of information for the usefulness of samples for the recognition task, recently, biometric utility has been shown to capture the usefulness of samples more accurately. Here, we show visual examples of both high and low feature norm images as estimated by AdaFace [20], and compare their biometric utility, estimated by CR-FIQA [4]. We can observe a disparity between the feature norms and biometric utility, alluding to the fact that the information present in feature norms cannot accurately represent the biometric sample utility.

improved separation of classes. Recently, adaptive margin functions have outperformed their static counterparts by assigning different margin values to individual samples. Adaptive methods generally focus on incorporating information tied to concepts such as the classifiability [3], recognizability [27], and (visual) quality [20] of samples, into the margin function. All of these concepts are in some way related to the usefulness (or utility) of the sample for recognition. The utility, sometimes also referred to as the biometric quality [18], [35], is tied to the likelihood of a sample causing a critical false match error during recognition. Face Image Quality Assessment (FIQA) techniques [2], [4], [15], [32], [39] attempt to estimate the utility of samples, and show excellent performance for the task of sample quality assessment, outperforming standard visual image quality descriptors, which focus on one or more quality components, such as blur, illumination, or sharpness.

Recent adaptive loss functions [20], [27] incorporate additional information, tied to the difficulty of a sample, into the margin function by using feature norms. While feature norms have been shown to correlate well with sample visual

image quality, their relationship to biometric utility remains underexplored. In Fig. 1 we showcase face samples for which the estimated feature norm and biometric utility do not agree. Here, samples with high feature norms can exhibit a large variety of utility, and vice versa. Motivated by the predictive power of biometric utility for sample quality estimation and the disconnect between the feature norms and utility scores, we propose a new adaptive loss function, named FunFace (Feature Utility and Norm Estimation for Face Recognition). The proposed loss function builds on the success of AdaFace [20], by incorporating an additional source of information, alongside the feature norm, into the margin function directly tied to the samples' biometric utility. This established a direct link between the samples' utility and their assigned margin value, enabling the underlying FR model to more easily adapt to samples of varying qualities, resulting in an improvement in performance, specifically in the low-quality domain.

II. RELATED WORK

Face recognition has seen significant performance improvements, especially in low-quality domains, with the development of new deep-learning methods. While aspects such as the use of large datasets [1], [13], [46], and improvements in the architectures [14], [25], [40] can be attributed to these improvements, we focus primarily on the developments of more expressive loss functions [9], [20], [36].

A. Margin Based Loss Functions.

Modern FR models most often employ some form of cross-entropy (softmax) loss, which can be written as:

$$\mathcal{L}_{CE} = -\log \left(\frac{e^{(s \cdot \cos(\theta_i))}}{\sum_{j=1}^C e^{(s \cdot \cos(\theta_j))}} \right), \quad (1)$$

where θ_k , represents the angle between the k -th class center and the given sample's feature, C is the number of classes in the training set, i is the index of the true (positive) class center, and s is a scaling hyperparameter of the model.

By applying a margin to the positive class term - $\cos(\theta_i)$ (PCT), we can obtain commonly used FR loss functions. Using a generalized form of PCT , written as:

$$PCT = \cos(m_{Sph} \cdot \theta_i + m_{Arc}) - m_{Cos}, \quad (2)$$

which can represent the three most common static margin-based loss functions currently in use. By applying only one of the included margins, we can derive the SphereFace [23], ArcFace (angular) [9], and CosFace (additive) [41] loss functions, using the m_{Sph} , m_{Arc} , and m_{Cos} margins, respectively. Altering the PCT , we rewrite the CE loss formulation, presented in (1), as:

$$\mathcal{L}_m = -\log \left(\frac{e^{(s \cdot PCT)}}{e^{(s \cdot PCT)} + \sum_{j=1, j \neq i}^C e^{(s \cdot \cos(\theta_j))}} \right). \quad (3)$$

Here, the terms evaluating the angle to the negative classes (contained in the sum) remain the same, while the PCT is altered with the appropriate margin values.

B. Adaptive Loss Functions

The use of adaptive margin functions can yield significant performance gains in unconstrained recognition scenarios, where no quality guarantees can be made regarding the input samples. This can be attributed to the ability of such functions to assign different margin values to each training sample, where the assigned value is often correlated with the sample's difficulty in the recognition process. CurricularFace [17] changes the margin between different stages (curriculum) of training, allowing it to focus on easier samples in the early stages and harder samples in later stages of training. MagFace [27] and AdaFace [20] find that the feature norm aligns well with the recognizability and visual image quality of samples, respectively, and use it for the adaptive margin. ElasticFace [3], on the other hand, focuses on adapting the margin according to the distance to the positive class center.

Our proposed loss function provides an additional source of information to the adaptive function, alongside the feature norm, in the form of the Certainty Ratio (CR) [4]. The CR correlates well with the biometric utility of samples, and is an excellent descriptor of sample quality. Incorporating CR and feature norms allows the model to focus on visual image quality queues and the biometric quality of samples, further improving performance in low-quality domains.

III. PRELIMINARIES

A. AdaFace Loss Function

AdaFace shows that the feature norm correlates well with the image quality of samples, as estimated by BRISQUE [28], a popular blind image quality estimation model. It adapts the commonly used PCT , presented in (2), using:

$$PCT_{Ada} = \cos(\theta_i + g_{angle}) - g_{add}, \quad (4)$$

where g_{angle} and g_{add} are the adaptive angular and additive margin functions. The feature norm is integrated into these margin functions, as shown here:

$$g_{angle} = -m \cdot \|\hat{z}_i\|, \quad g_{add} = m + (m \cdot \|\hat{z}_i\|), \quad (5)$$

where $\|\hat{z}_i\|$ is the normalized feature norm, and m is a static margin value. By defining the angular and additive margin functions in such a way, AdaFace can be seen as a generalized version of both ArcFace and CosFace. When the normalized feature norm $\|\hat{z}_i\|$ equals 0, the angular margin function g_{angle} evaluates to 0, and we are left only with a static additive margin, as in the CosFace loss function. Similarly, when the normalized feature norm $\|\hat{z}_i\|$ equals -1 , the additive margin function g_{add} evaluates to 0, leaving us with only a static angular margin, as in the ArcFace loss function. The normalized feature norm is calculated using:

$$\|\hat{z}_i\| = \left[\frac{\|z_i\| - \mu_z}{\sigma_z/h} \right]_{-1}^1, \quad (6)$$

where μ_z and σ_z are the running mean and standard deviation of the feature norms, and h is a hyperparameter defining the spread of the normalized distribution. The running mean and standard deviation are tracked using a simple EMA (Exponential Moving Average) approach.

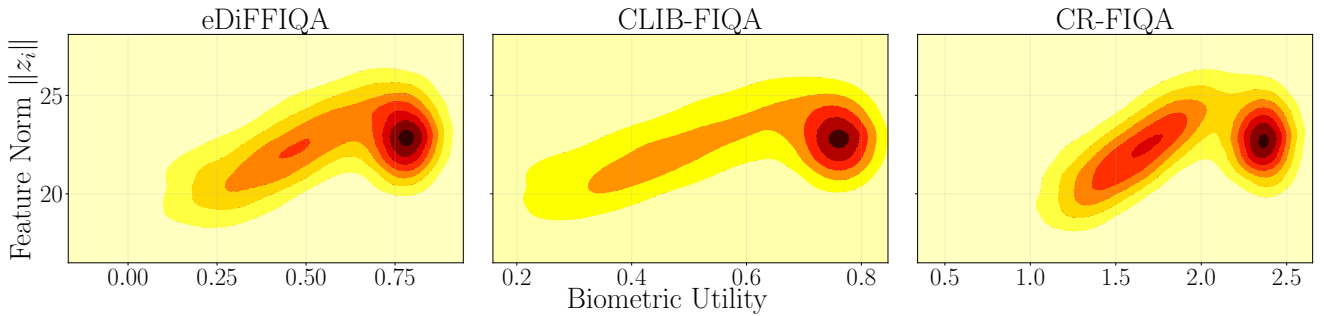


Fig. 2: **Comparison between the feature norm and the utility of samples.** Using samples from the XQLFW [22] benchmark, we generated density maps tying the feature norms and biometric utility of samples. We analyze the feature norms produced by AdaFace, and use three state-of-the-art FIQA techniques: eDiFFIQA, CLIB-FIQA, and CR-FIQA to estimate the biometric utility of samples. We can observe a distinct shape for all included FIQA techniques, where low-scoring samples exhibit linear correlation between the feature norm and biometric utility, while higher-scoring samples exhibit an inverse correlation between the two measures.

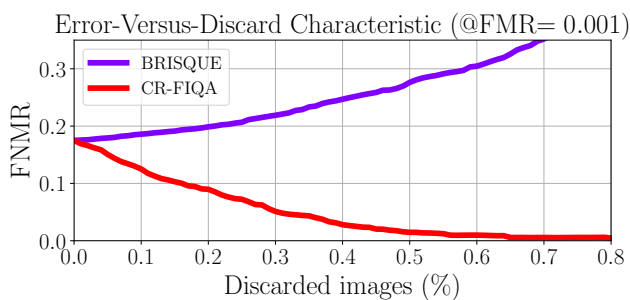


Fig. 3: **Predictive strength of FIQA.** We compare BRISQUE, an IQA technique, against CR-FIQA, a supervised FIQA technique, using the Error-versus-Discard Characteristic (EDC) curves, which measure the False Non-Match Rate (FNMR) at a given False Match Rate (FMR) at various discard rates, on the XQLFW benchmark [22].

B. CR-FIQA

Estimating the biometric utility of samples has proven to be extremely useful for recognition in low-quality domains, where discarding and quality-weighted averaging of samples can be utilized. CR-FIQA [4] is a supervised FIQA technique that does not require any pseudo-quality labels. During the training of a FR model, the method computes the Certainty Ratio (CR), which is closely tied to the biometric utility, as demonstrated by its excellent performance as an estimator of sample quality. During training of a FR model, CR can be computed as:

$$CR = \frac{CCS}{NNCCS + (1 + \epsilon)}, \quad (7)$$

where CCS is the Class Center Angular Similarity, and $NNCCS$ is the Nearest Negative Class Center Angular Similarity. The ratio is expanded with the $(1 + \epsilon)$ term to ensure stable computation, pushing the value range from $[-1, 1]$ to $[\epsilon, 2 + \epsilon]$. Expressing CR directly with cosine

similarities, we can write:

$$CR = \frac{\cos(\theta_i)}{\max_{j, j \neq i}^C(\theta_j) + (1 + \epsilon)}, \quad (8)$$

where i is the positive class index, and C is the total number of classes in the training dataset.

IV. METHODOLOGY

The performance of FR methods has seen a steady improvement in recent years, thanks to the development of novel adaptive loss functions [3], [17], [20], [27]. These loss functions enable the margin to adapt to individual training samples, usually through information present in the feature norm, which can be associated with the visual image quality of samples [20]. Recently, through the developments of new FIQA techniques, biometric utility has been shown to be an excellent predictor of sample quality for recognition tasks [18], [34], outperforming classical (visual) image quality estimation techniques [24], [28], [29]. In this section, we present our proposed FunFace loss function, which integrates information about the utility of a sample, estimated by a FIQA technique into the adaptive margin. To further motivate the use of FIQA techniques for adaptive functions, we first demonstrate the performance gap between IQA and FIQA techniques for sample quality estimation, and show that the feature norm is an inadequate source of information for accurate estimation of a sample's utility.

A. Comparison of Quality Estimation Techniques

Estimating the quality of samples is an important task for end applications dealing with variable or low-quality images. By using quality estimates, FR systems can discard samples below a specific threshold, which are more likely to cause problematic false match errors, and in turn improve their reliability. In such cases, the predictive strength of the quality estimates is an extremely important factor as it determines the potential improvement in robustness of the FR system.

Recently, FIQA techniques have shown to be excellent predictors of sample quality for recognition, outperforming

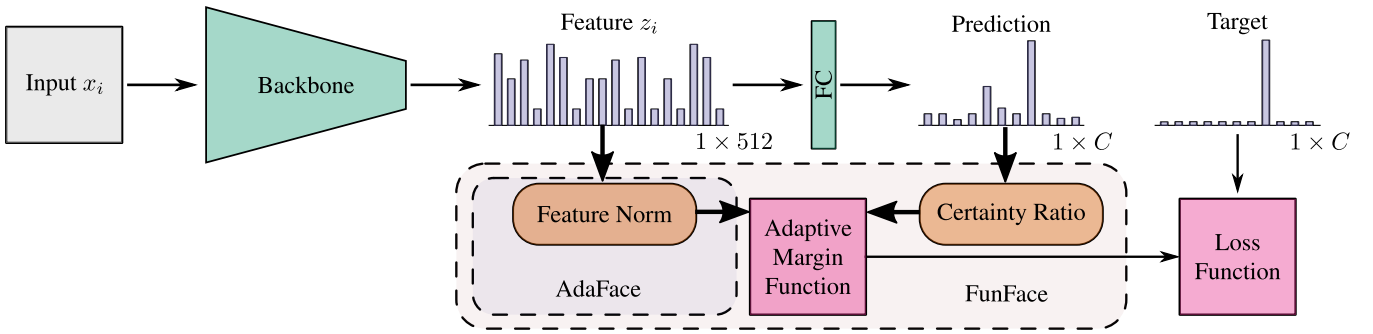


Fig. 4: **Overview of the proposed FunFace loss function.** We extend the AdaFace framework, which adapts the margin according to the measured feature norm, with additional information extracted from the class similarities in the form of the Certainty Ratio. Both the feature norm and certainty ratio are combined to form the final margin value. This allows the final loss to distinguish between samples with varying feature norms as AdaFace, while providing better angular separation, which more closely follows the biometric utility of individual samples.

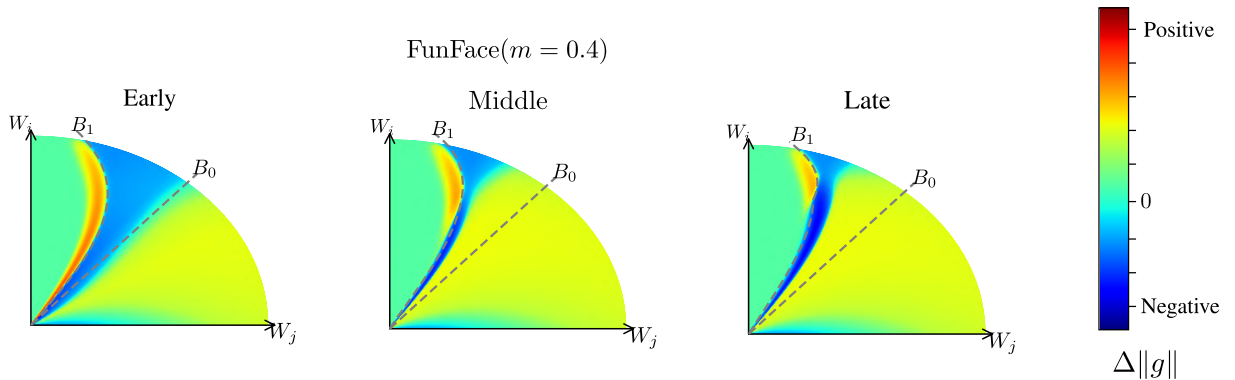


Fig. 5: **Analysis of the gradient scale difference between FunFace and the AdaFace loss function.** We present the difference of the gradient scale between the AdaFace and the proposed FunFace loss function highlighting key differences between the two. Since the mean of the normalized certainty ratio changes drastically during training, we include an analysis from early, middle and late stages of training. Each figure shows the position of the positive class center W_i and the nearest negative class center W_j , as well as the class border before applying the margin function B_0 and after application B_1 .

standard (visual) image quality estimators. Unlike standard approaches, which focus on aspects closely related to human vision, such as the resolution or illumination, these methods estimate the biometric utility (usefulness) of samples, concerned more so with how the FR model perceives the samples. To highlight the effectiveness of FIQA techniques and consequently of biometric utility for FR, we evaluate BRISQUE [28] and CR-FIQA [4] using standard Error-versus-Discard Characteristic (EDC) curves. These curves show the predictive strength of the quality estimates by iteratively discarding a percentage of the lowest quality images, and evaluating the performance of the underlying FR model at each iteration. We chose BRISQUE as the general IQA technique, since it is used in AdaFace as a predictor of image quality. The results of the analysis on the XQLFW benchmark using the AdaFace model are presented in Fig. 3, and show a clear divide between the methods. The performance of the underlying FR model gradually improves when discarding samples through CR-FIQA scores, while for BRISQUE, the performance actually drops, as seen by the increase in the False Non Match Rate (FNMR). This

difference in performance between the methods demonstrates the predictive strength of FIQA techniques for sample quality estimation and further motivates the use of such techniques to guide the model during training.

B. Analysis of the Feature Norm

The feature norm commonly used by adaptive loss functions as a source of information tied to the samples' difficulty is correlated with the visual image quality, as showcased in [20]. However, as shown in IV-A, visual quality is not as important as biometric utility for sample quality estimation, leading us to explore possible correlations between the feature norm and biometric utility of samples.

In Fig. 2 we present a comparison between the feature norms estimated by AdaFace and the biometric utility, estimated by three unique state-of-the-art FIQA techniques eDifFIQA [2], CLIB-FIQA [33], and CR-FIQA [4] using images from the XQLFW benchmark. From the presented density maps, we can observe a similar pattern across all three different FIQA techniques. Because of the unique shape of the observed pattern, we can conclude that the feature norm alone does not contain enough information

to accurately depict the biometric utility of a sample. For example, a sample with a high feature norm could correspond to a wide degree of different biometric utilities. Therefore, a more explicit source of biometric utility information is required in the adaptive margin, to account for the observed discrepancy between the two measures.

C. FUNFace Loss

The proposed FunFace loss function extends on the AdaFace framework, presented in III-A, by incorporating a biometric utility estimator directly into the adaptive margin function, as shown in Fig. 4. Due to its excellent performance and ease of integration, we chose CR-FIQA, described in Section III, as the target biometric utility estimator. The FunFace loss, similarly to AdaFace, defines the adaptive margin functions as:

$$g_{angle} = -m \cdot \kappa, \quad g_{add} = m + (m \cdot \kappa), \quad (9)$$

where κ is a function of the normalized feature norm $\|\hat{z}_i\|$, as in AdaFace but also of the normalized certainty ratio $\|CR_i\|$. The adaptive function κ can be written as:

$$\kappa = \lambda \cdot \|\hat{z}_i\| + (1 - \lambda) \cdot \|\hat{CR}_i\|, \quad (10)$$

where λ is a hyperparameter of the model called the mixing factor, which determines the scaling between the two quality descriptors $\|\hat{z}\|$ and $\|\hat{CR}_i\|$. To compute the CR descriptor of the training samples, we adapt the original formulation, presented in (7), as shown here:

$$CR_i = \frac{\lfloor CSS \rfloor_0^1}{\lfloor NNCCS \rfloor_0^1 + \epsilon} = \frac{\lfloor \cos(\theta_i) \rfloor_0^1}{\lfloor \max_{j,j \neq i}^C(\cos\theta_j) \rfloor_0^1 + \epsilon}, \quad (11)$$

altering the normalization using ϵ , and limiting all cosine values to the interval $[0, 1]$. Through testing, we find that limiting the values impacts training only very early on, and has little effect otherwise, since both the positive and the closest negative class are close enough to the sample to fall into the range of $[0, 1]$. Similar to the feature norm, presented in (6), we normalize the CR values as follows:

$$\|\hat{CR}_i\| = \left[\frac{\|CR_i\| - \mu_{CR}}{\sigma_{CR}/h} \right]_{-1}^1, \quad (12)$$

where the μ_{CR} is the running mean and σ_{CR} the standard deviation of the samples certainty ratio, and h a hyperparameter controlling the spread of the distribution.

D. Analysis of the Gradient Scale

The proposed loss focuses on visual image quality and biometric utility through the use of the feature norm and certainty ratio, respectively. The effects of the feature norm on the gradient scale using the AdaFace loss function have been studied extensively in [20]. Since our proposed loss function directly extends AdaFace, the effects of the feature norm term should correlate well with the original formulation. However, we also include a normalized certainty ratio term in our adaptive margin function.

The certainty ratio is fully dependent on the angular position; the norm of the feature does not affect it, but

TABLE I: Comparison of benchmarks datasets. We present additional information about the benchmark datasets, showcasing the differences in scale and difficulty.

Dataset	Images	IDs	Comparisons		Difficulty
			Mated	Non-mated	
LFW [16]	13,233	5,749	3,000	3,000	Low
CFP-FP [37]	7,000	500	3,500	3,500	Low
CPLFW [44]	11,652	3,930	3,000	3,000	Low
AgeDB [30]	16,488	570	10,000	10,000	Low
CALFW [45]	12,174	4,025	3,000	3,000	Low
IJB-C [26]	138,836	3,531	19,557	15,638,932	Medium
TinyFace [5]	169,403	5,139	3,728	-	High
DroneSURF [19]	411,451	58	712,198	-	High
SurvFace [6]	463,507	15,573	5,320	-	High

rather only the angular distance to the two adjacent class centers. In Fig. 5 we present an analysis of the impact of the certainty ratio on the gradient scale. The figure presents the difference in the gradient scale between the FunFace and AdaFace loss functions, showcasing how the addition of the certainty ratio influences the final decision space. Due to the dynamic nature of the normalized certainty ratio, the differences between the two change during training. Earlier on, the mean of the certainty ratio is close to one, i.e., the features are randomly spaced between the positive and nearest negative class center, however, as the training progresses, the ratio rises, leading training samples away from the negative and closer to the positive class center. From the results, we can see that earlier on FunFace puts less emphasis on more difficult samples, as highlighted by the region with negative gradient scale change, shown in blue, between the two decision boundaries. Gradually, through the training process, more weight is put on difficult samples (far from the boundary B_1), as highlighted by an increase in the gradient scale difference (green), for the middle and late stages of training. An inverse effect can be seen for samples on the positive side of the decision boundary B_1 . Earlier on FunFace emphasizes easier samples (red), while later on this effect dissipates significantly.

V. EXPERIMENTS & RESULTS

A. Experimental Setup

Datasets. We divide the benchmarks according to their scale and difficulty into: (i) small-scale, low difficulty LFW [16], CFP-FP [37], CPLFW [44], AgeDB [30], and CALFW [45] benchmarks, (ii) large-scale, medium difficulty IJB-C [26], and (iii) small-scale, high difficulty TinyFace [5], DroneSURF [19], and SurvFace [6]. More information about the benchmarks can be found in Table I. For training, we employed three commonly used face datasets: (i) MS1MV2 [9], [13], (ii) MS1MV3 [11], [13], and (iii) WebFace4M [46], containing 5.8M, 5.1M, and 4.2M images of 85k, 93k and 205k subjects, respectively.

Training Setting. The training datasets have been preprocessed to a final size of 112×112 pixels using a standard five-point landmark alignment [9] using the RetinaFace [8], and MTCNN [43] detectors. For the FR model, we make use of the ResNet-100 [14] backbone and train it using SGD,

TABLE II: **Experimental results on low and medium difficulty benchmarks.** For the low difficulty benchmarks the results are presented using the verification accuracy (%), and for the mixed difficulty using the True Acceptance Rate (TAR) at three False Acceptance Rates (FARs): $1e^{-4}$, $1e^{-5}$, $1e^{-6}$. For each training dataset, we mark the **best**, and *second-best* result.

Method	Dataset	LFW	CFP-FP	CPLFW	AgeDB	CALFW	IJB-C (TAR@FAR=?)		
							$1e^{-4}$	$1e^{-5}$	$1e^{-6}$
CosFace(m=0.35) [41]	MS1MV2	99.81	98.12	92.28	98.11	95.76	96.37	-	-
ArcFace(m=0.50) [9]		99.83	98.27	92.08	98.28	95.45	96.03	-	-
MV-SoftMax [42]		99.80	98.28	92.83	97.95	96.10	95.20	-	-
CurricularFace [17]		99.80	98.37	93.13	98.32	96.20	96.10	-	-
BroadFace [21]		99.85	98.63	93.17	98.38	<i>96.20</i>	96.03	<i>94.11</i>	90.41
MagFace [27]		99.83	98.46	92.87	98.17	96.15	95.97	94.08	<i>90.24</i>
ElasticFace-Arc+ [3]		99.82	98.60	93.28	<i>98.35</i>	96.17	96.40	-	-
ElasticFace-Cos+ [3]		99.80	98.73	93.23	98.28	96.18	96.65	-	-
PartialFC(ArcFace) [1]		<i>99.83</i>	98.45	93.00	98.20	96.18	96.20	-	-
PartialFC(CosFace) [1]		99.83	98.51	93.10	98.03	96.20	96.40	-	-
AdaFace(m=0.4) [20]		99.82	98.49	<i>93.53</i>	98.05	96.08	96.89	-	-
FunFace (ours)		99.82	<i>98.67</i>	93.83	98.13	96.25	<i>96.82</i>	94.47	89.01
VPL-ArcFace [10]	MS1MV3	<i>99.83</i>	<i>99.11</i>	93.45	98.60	<i>96.12</i>	96.76	-	-
AdaFace [20]		99.83	99.03	<i>93.93</i>	98.17	96.02	<i>97.09</i>	-	-
FunFace (ours)		99.83	99.16	94.12	<i>98.27</i>	96.15	97.20	95.54	90.03
ArcFace [9]	WebFace4M	99.83	<i>99.19</i>	94.35	97.95	96.00	97.16	-	-
AdaFace [20]		99.80	99.17	<i>94.63</i>	97.90	96.05	97.39	-	-
FunFace (ours)		99.85	99.14	94.63	97.90	96.15	97.31	95.62	<i>90.81</i>

TABLE III: **Experimental results on the TinyFace benchmark.** We present the Rank-1 and Rank-5 scores for each training dataset; we mark the **best** and *second-best* result.

Method	Dataset	Rank-1	Rank-5
URL [38]	MS1MV2	63.89	68.67
CurricularFace [17]		63.68	67.65
AdaFace(m=0.4) [20]		<i>68.21</i>	<i>71.54</i>
FunFace (ours)		69.07	72.21
AdaFace(m=0.4) [20]	MS1MV3	<i>67.81</i>	<i>70.98</i>
FunFace (ours)		69.88	73.04
ArcFace [9]	WebFace4M	71.11	74.38
AdaFace(m=0.4) [20]		<i>72.02</i>	<i>74.52</i>
FunFace (ours)		73.31	76.05

with an initial learning rate of $1e^{-1}$, weight decay of $5e^{-4}$, and momentum of 0.9. The learning rate is decreased by a factor of 10 at the end of the 12th, 20th, and 24th epoch. We use a mini-batch of 1024 samples and train for 26 epochs. For the implementation of FunFace, we closely follow the findings in [20], setting the hyperparameter h , which controls the spread of the normalized distributions to 0.333, and the scaling hyperparameter s to 64. The models were trained using a compute node equipped with two H100 80GB GPUs, an AMD EPYC 9124 64-core CPU, and 256GB of RAM.

Augmentations. In [20] they show that while static margin loss functions do not see benefits from aggressive augmentations of training samples, the same does not necessarily hold for adaptive margins. Our focus is to excel in both high- and low-quality domains. Therefore, we extend the augmentations proposed in [20], i.e., cropping, rescaling, and photometric jittering, with noising, grayscaling, masking, and affine transformations. The use of augmentations allows the

FR model to better adapt to a larger variety of image qualities, ensuring robust performance in low-quality domains.

Noising produces a weighted average of the original sample and a noise mask, sampled from a uniform or Gaussian distribution. Affine transformations can rotate, translate, and shear the sample, addressing potential issues of the face detection step. Grayscaling blends data from the RGB channels into a single grayscale channel. Masking obscures a rectangular area of the image and can imitate various occlusions due to face masks, glasses, hats, or any other potential objects blocking a part of the face. All operations are applied randomly with a probability of 20%, except for grayscaling, which is applied with a probability of 5%.

B. Comparison with State-of-the-art

In this section, we compare the results of the proposed FUNFace loss function with 13 state-of-the-art approaches, including a variety of static and adaptive margin loss functions. We divide the experiments according to the difficulty of the benchmark datasets, as shown in Table I.

Low and Medium Difficulty. Table II presents the experimental results across the low difficulty, i.e., LFW, CFP-FP, CPLFW, AgeDB, and CALFW, as well as the medium difficulty IJB-C benchmark. For the low difficulty benchmarks, we show the verification accuracy, while for the medium difficulty benchmark, we show the True Acceptance Rate (TAR) at three common False Acceptance Rates: $1e^{-4}$, $1e^{-5}$, and $1e^{-6}$. The methods are further divided according to the training dataset. From the results, we can see that the performance of FunFace is similar to current state-of-the-art methods such as AdaFace. While variations in specific scenarios can be seen, such as the increase in performance for

TABLE IV: **Experimental results on the high difficulty DroneSURF and SurvFace benchmarks.** We present the Rank-N(%) score for DroneSURF and the True Acceptance Rate (TAR) at 30%, 20%, 10%, 1% False Acceptance Rates (FARs).

Method	Dataset	DroneSURF-A		DroneSURF-P		SurvFace (TAR@FAR=?)			
		Rank-1	Rank-5	Rank-1	Rank-5	30%	20%	10%	1%
MagFace [27]	MS1MV2	56.92	68.57	25.07	37.27	58.15	47.04	32.53	9.45
ElasticFace [3]		58.23	69.94	26.42	38.70	56.76	46.18	32.38	10.37
AdaFace [20]		59.00	70.37	27.51	39.59	55.97	45.39	29.71	7.55
FunFace (ours)		65.20	75.76	32.65	44.46	60.50	48.83	34.39	11.91
Adaface [20]	MS1MV3	63.03	73.86	31.30	43.64	58.83	47.50	31.74	6.88
FunFace (ours)		68.68	78.25	35.58	47.16	59.69	47.98	32.70	8.85
Adaface [20]	WebFace4M	72.91	81.87	41.62	53.67	67.93	57.04	41.20	10.37
FunFace (ours)		75.06	83.51	44.10	55.93	68.53	58.08	43.12	14.90

TABLE V: **Experimental results of the ablation study.** We present a study of the mixing parameter λ and augmentations on the performance of the FunFace loss, across a wide range of benchmark datasets.

Method	LFW	CFP-FP	CPLFW	AgeDB	CALFW	IJB-C (TAR@FAR=?)			TinyFace Rank-1	
						$1e^{-3}$	$1e^{-4}$	$1e^{-5}$		
AdaFace	99.85	99.23	94.63	97.93	96.08	97.36	95.37	90.85	72.99	
FunFace(λ)	0.1	99.85	99.14	94.63	97.90	96.15	97.31	95.62	90.81	73.31
	0.3	99.85	99.20	94.77	97.92	96.02	97.28	95.55	91.81	73.26
	0.5	99.83	99.24	94.78	97.88	96.10	97.34	95.52	90.33	72.88
	0.7	99.85	99.21	94.63	97.83	96.12	97.34	95.60	91.23	72.83
	0.9	99.82	99.23	94.65	97.82	96.02	97.31	95.48	91.23	72.32
Aug.	Ada.	99.80	99.24	94.62	98.07	96.03	97.08	94.60	87.07	71.03
	Fun.	99.85	99.24	94.52	97.88	96.03	97.23	95.36	89.88	71.14

TABLE VI: **Comparison of the runtime between methods.** We present the results for three techniques, ArcFace, AdaFace, and FunFace, showcasing the differences in runtime.

Method	ArcFace [20]		AdaFace [20]		FunFace	
Time ($\mu \pm \sigma$)	0.0032	\pm 0.0021	0.0075	\pm 0.0021	0.1220	\pm 0.0058

the MS1MV3 dataset on all benchmarks, or a slight decrease on IJB-C, with WebFace4M, overall, the method performs well on both low and medium difficulty benchmarks.

High Difficulty. The experimental results for the low-quality (high difficulty) benchmarks, i.e., TinyFace, DroneSURF, and SurvFace, are shown in Tables III and IV. In Table III, we show the results on the TinyFace benchmark, using the Rank-1 and Rank-5 metrics. The results of competing methods are taken from [20]. From the experimental results, we see that our method outperforms AdaFace across all training datasets, achieving state-of-the-art performance. This result showcases the power of biometric utility as a descriptor of sample quality and its importance in the training process of FR models. In Table IV, we show the results on both the DroneSURF and SurvFace benchmarks. Since results have not been reported on these benchmarks for any method, we evaluate them ourselves. We use models directly provided by the authors and closely follow the provided image preprocessing for each included method. We include results from three excellent adaptive loss functions, i.e., MagFace, ElasticFace, and AdaFace, and compare them to results from

FunFace. For the DroneSURF benchmark, we divide the results into the Active scenario (DroneSURF-A) and the Passive scenario (DroneSURF-P), aligning with the division of data in the dataset. For both scenarios, we evaluate each frame individually, reporting the Rank-1 and Rank-5 metrics. For the SurvFace benchmark, we perform verification tests and report them using the True Acceptance Rate (TAR) at four False Acceptance Rates (FARs), i.e., 30%, 20%, 10%, and 1%, similar to the author suggested values. Unlike the authors of the benchmark, we do not, however, perform fine-tuning on the included training set of images. The results showcase a similar story as before. FunFace outperforms the included adaptive methods across all training datasets, achieving state-of-the-art performance.

C. Ablation Study

In this section, we present an ablation study of the proposed FunFace loss functions, and two important factors surrounding it: (i) the λ parameter and (ii) the image augmentations, as well as a comparison of the time complexity.

Experimental results across a wide-range of benchmarks are presented in Table V. We include results of a FR model

trained using the AdaFace loss function, with our extended augmentations as well as using no augmentations, for better comparison with our proposed method. Focusing first on the effects of the mixing factor λ on the performance of the model. On the low and medium difficulty benchmarks, the results seem to be inconclusive, with performance varying between individual benchmarks. On the high difficulty TinyFace benchmark, however, we can see a clear pattern between the value of λ and the achieved performance of the FR model. We see that the inclusion of the certainty ratio improves performance over the AdaFace model, when using $\lambda = [0.1, 0.3]$. However, we can also see that the performance quickly drops with an increase of λ , showcasing the importance of the balance of information present in the feature norms and biometric utility of samples. Comparing the results of models trained with no augmentations using both AdaFace and FunFace, we see a similar result as when training using augmentations. The performance on low and medium-difficulty benchmarks closely matches between the adaptive functions, with a surprising result on IJB-C, where FunFace achieves a marginal improvement over AdaFace. For high difficulty images, we again see that FunFace outperforms AdaFace, but the difference is not as convincing as with the augmentations, showing the importance of lower quality samples during training of the model.

In Table VI we show a comparison of the runtime between three methods: static margin ArcFace, adaptive AdaFace, and our proposed FunFace loss function. The presented times are measured across a single epoch over the WebFace4M dataset, using a batch size of 1024. From the results, we can observe a drastic increase in the runtime of FunFace stemming from the costly computation of the closest negative neighbour of each sample in the batch, as well as the additional time to compute the normalized certainty ratio. Comparing AdaFace and ArcFace, the runtime is much closer, while, expectedly, ArcFace is still faster than AdaFace.

VI. CONCLUSION

In this paper, we presented a new loss function, FunFace, focused on performance improvements in low-quality domains. FunFace builds on the success of AdaFace, which encodes information about the visual quality into the margin using feature norms. We show that the biometric utility is a superior measure of sample difficulty for recognition compared to visual quality estimators, and that it cannot be accurately portrayed using existing information present in the feature norms. This motivates us to include an additional source of information directly related to the biometric utility into the FunFace margin function. We employ the CR-FIQA technique to estimate the utility of samples online during training and combine it with the feature norm information into the final margin value. From the experimental evaluation of the method, we can observe that the performance for high and mixed quality images is comparable to state-of-the-art techniques such as AdaFace. For images of lower quality, as showcased by high difficulty benchmarks, we see a dramatic improvement in quality, stemming from the addition

of the extended augmentations and the predictive power of the certainty ratio for the biometric utility of samples.

Limitations & Future Work. This work focuses on improvements in low-quality domains, as such, its performance in high- and medium-quality domains does not show a clear improvement over state-of-the-art methods. Another clear downside of the approach is its heavier computational runtime, which significantly affects the training time of models. Additionally, the provided analysis shows a benefit of using extended augmentations of input samples, but explores the correlation only briefly. The proposed loss function is derivative work, extending on previous work, namely AdaFace and CR-FIQA, and therefore does not explore all hyperparameters and variables discussed already in the mentioned works.

VII. ETHICAL STATEMENT

A. Ethical Review Boards The present work develops a face-recognition training objective (FunFace) and evaluates it on established, publicly available training and benchmark datasets. We did not recruit participants, perform interventions, or conduct user studies, and we did not collect new biometric data; therefore, formal oversight by an ethical review board was not required for the work reported here.

B. Potential Harms to Human Subjects Although no new data were collected, face datasets inherently contain biometric information, and misuse or leakage could impact individuals' privacy. In line with the conference guidance on informed consent and the protection of privacy and confidentiality, we do not release any raw face imagery or identity annotations beyond what is already provided by the original dataset distributors, and we handle data according to the corresponding dataset terms.

C. Potential Negative Societal Impacts Our primary goal is to improve face recognition robustness in challenging, low-quality scenarios. Such improvements can be beneficial in benign applications (e.g., robust authentication in unconstrained conditions), but they may also enhance surveillance capabilities, which can be used in ways that limit rights, privacy, or exacerbate discrimination. We explicitly acknowledge this dual-use risk and do not endorse deployments that violate applicable laws, human rights, or privacy expectations.

D. Risk-Mitigation Strategies To mitigate risks of misuse and misunderstanding, we (i) clearly scope our contribution as a training objective and report limitations, (ii) commit to releasing code and trained models to support transparency and reproducibility, and (iii) refrain from distributing withdrawn or restricted datasets. In particular, to enable fair comparison with prior work, we report results using MS-Celeb-derived training data, noting that MS-Celeb has been withdrawn; we do not redistribute this data and encourage the use of currently available alternatives when possible. These steps align with the emphasis on risk mitigation via controlled access to research materials and clear communication of intended use and limitations.

REFERENCES

- [1] X. An, X. Zhu, Y. Gao, Y. Xiao, Y. Zhao, Z. Feng, L. Wu, B. Qin, M. Zhang, D. Zhang, et al. Partial FC: Training 10 Million Identities on a Single Machine. In *Proceedings of the CVF/IEEE International Conference on Computer Vision (ICCV)*, pages 1445–1449, 2021.
- [2] Z. Babnik, P. Peer, and V. Štruc. eDiffIQA: Towards Efficient Face Image Quality Assessment Based on Denoising Diffusion Probabilistic Models. *IEEE Transactions on Biometrics, Behavior, and Identity Science (TBIOM)*, 6(4):458–474, 2024.
- [3] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper. ElasticFace: Elastic Margin Loss for Deep Face Recognition. In *Proceedings of the CVF/IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022.
- [4] 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 the CVF/IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.
- [5] Z. Cheng, X. Zhu, and S. Gong. Low-Resolution Face Recognition. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pages 605–621. Springer, 2018.
- [6] Z. Cheng, X. Zhu, and S. Gong. Surveillance face recognition challenge. *arXiv preprint arXiv:1804.09691*, 2018.
- [7] D. Cornett, J. Brogan, N. Barber, D. Aykac, S. Baird, N. Burchfield, C. Dukes, A. Duncan, R. Ferrell, J. Goddard, et al. Expanding Accurate Person Recognition to New Altitudes and Ranges: The Briar Dataset. In *Proceedings of the CVF/IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 593–602, 2023.
- [8] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou. RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild. In *Proceedings of the CVF/IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5203–5212, 2020.
- [9] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. Arcface: Additive Angular Margin Loss for Deep Face Recognition. In *Proceedings of the CVF/IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4690–4699, 2019.
- [10] J. Deng, J. Guo, J. Yang, A. Lattas, and S. Zafeiriou. Variational Prototype Learning for Deep Face Recognition. In *Proceedings of the CVF/IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11906–11915, 2021.
- [11] J. Deng, J. Guo, D. Zhang, Y. Deng, X. Lu, and S. Shi. Lightweight Face Recognition Challenge. In *Proceedings of the CVF/IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 0–0, 2019.
- [12] H. Du, H. Shi, D. Zeng, X.-P. Zhang, and T. Mei. The Elements of End-to-End Deep Face Recognition: A Survey of Recent Advances. *ACM computing surveys (CSUR)*, 54(10s):1–42, 2022.
- [13] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao. MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 87–102. Springer, 2016.
- [14] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In *Proceedings of the CVF/IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.
- [15] J. Hernandez-Ortega, J. Galbally, J. Fierrez, R. Haraksim, and L. Beslay. FaceQnet: Quality Assessment for Face Recognition Based on Deep Learning. In *Proceedings of the IAPR International Conference on Biometrics (ICB)*, pages 1–8, 2019.
- [16] 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. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.
- [17] Y. Huang, Y. Wang, Y. Tai, X. Liu, P. Shen, S. Li, J. Li, and F. Huang. CurricularFace: Adaptive Curriculum Learning Loss for Deep Face Recognition. In *Proceedings of the CVF/IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5901–5910, 2020.
- [18] ISO/IEC DIS 29794-1, Biometric Sample Quality. Standard, International Organization for Standardization (ISO), 2022.
- [19] I. Kalra, M. Singh, S. Nagpal, R. Singh, M. Vatsa, and P. Sujit. DroneSURF: Benchmark Dataset for Drone-Based Face Recognition. In *Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pages 1–7. IEEE, 2019.
- [20] M. Kim, A. K. Jain, and X. Liu. AdaFace: Quality Adaptive Margin for Face Recognition. In *Proceedings of the CVF/IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18750–18759, 2022.
- [21] Y. Kim, W. Park, and J. Shin. BroadFace: Looking at Tens of Thousands of People at Once for Face Recognition. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 536–552. Springer, 2020.
- [22] 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 the IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–5, 2021.
- [23] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song. Sphreface: Deep Hypersphere Embedding for Face Recognition. In *Proceedings of the CVF/IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 212–220, 2017.
- [24] X. Liu, J. Van De Weijer, and A. D. Bagdanov. RankIQA: Learning from Rankings for No-Reference Image Quality Assessment. In *Proceedings of the CVF/IEEE International Conference on Computer Vision (ICCV)*, pages 1040–1049, 2017.
- [25] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. SWIN Transformer: Hierarchical Vision Transformer Using Shifted Windows. In *Proceedings of the CVF/IEEE International Conference on Computer Vision*, pages 10012–10022, 2021.
- [26] B. Maze, J. Adams, J. A. Duncan, N. Kalka, T. Miller, C. Otto, A. K. Jain, W. T. Niggel, J. Anderson, J. Cheney, et al. IARPA Janus Benchmark-C: Face Dataset and Protocol. In *Proceedings of the International Conference on Biometrics (ICB)*, pages 158–165, 2018.
- [27] Q. Meng, S. Zhao, Z. Huang, and F. Zhou. MagFace: A Universal Representation for Face Recognition and Quality Assessment. In *Proceedings of the CVF/IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14225–14234, 2021.
- [28] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Transactions on Image Processing (TIP)*, 21(12):4695–4708, 2012.
- [29] A. Mittal, R. Soundararajan, and A. C. Bovik. Making a “Completely Blind” Image Quality Analyzer. *IEEE Signal Processing Letters (SPL)*, 20(3):209–212, 2012.
- [30] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, and S. Zafeiriou. AgeDB: the First Manually Collected, in-the-Wild Age Database. In *Proceedings of the CVF/IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 51–59, 2017.
- [31] K. Narayan, N. G. Nair, J. Xu, R. Chellappa, and V. M. Patel. PetalFace: Parameter Efficient Transfer Learning for Low-Resolution Face Recognition. In *Proceedings of the CVF/IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 804–814. IEEE, 2025.
- [32] 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 the CVF/IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7670–7679, 2021.
- [33] F.-Z. Ou, C. Li, S. Wang, and S. Kwong. CLIB-FIQA: Face Image Quality Assessment with Confidence Calibration. In *Proceedings of the CVF/IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1694–1704, 2024.
- [34] T. Schlett, C. Rathgeb, O. Henniger, J. Galbally, J. Fierrez, and C. Busch. Face Image Quality Assessment: A Literature Survey. *ACM Computing Surveys (CSUR)*, 54(10s):1–49, 2022.
- [35] T. Schlett, C. Rathgeb, J. Tapia, and C. Busch. Considerations on the Evaluation of Biometric Quality Assessment Algorithms. *arXiv preprint arXiv:2303.13294*, 2023.
- [36] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A Unified Embedding for Face Recognition and Clustering. In *Proceedings of the CVF/IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 815–823, 2015.
- [37] 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 the IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016.
- [38] Y. Shi, X. Yu, K. Sohn, M. Chandraker, and A. K. Jain. Towards Universal Representation Learning for Deep Face Recognition. In *Proceedings of the CVF/IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6817–6826, 2020.
- [39] P. Terhorst, J. N. Kolf, N. Damer, F. Kirchbuchner, and A. Kuijper. SER-FIQ: Unsupervised Estimation of Face Image Quality Based on Stochastic Embedding Robustness. In *Proceedings of the CVF/IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5651–5660, 2020.
- [40] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is All You Need.

Advances in Neural Information Processing Systems (NeurIPS), 30, 2017.

- [41] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu. CosFace: Large Margin Cosine Loss for Deep Face Recognition. In *Proceedings of the CVF/IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5265–5274, 2018.
- [42] X. Wang, S. Zhang, S. Wang, T. Fu, H. Shi, and T. Mei. Mis-classified Vector Guided Softmax Loss for Face Recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 12241–12248, 2020.
- [43] J. Xiang and G. Zhu. Joint Face Detection and Facial Expression Recognition with MTCNN. In *2017 4th International Conference on Information Science and Control Engineering (ICISCE)*, pages 424–427. IEEE, 2017.
- [44] T. Zheng and W. Deng. Cross-Pose LFW: A Database for Studying Cross-Pose Face Recognition in Unconstrained Environments. Technical Report 18-01, Beijing University of Posts and Telecommunications, February 2018.
- [45] T. Zheng, W. Deng, and J. Hu. Cross-Age LFW: A Database for Studying Cross-Age Face Recognition in Unconstrained Environments. *CoRR*, abs/1708.08197, 2017.
- [46] Z. Zhu, G. Huang, J. Deng, Y. Ye, J. Huang, X. Chen, J. Zhu, T. Yang, J. Lu, D. Du, et al. WebFace260M: A Benchmark Unveiling the Power of Million-Scale Deep Face Recognition. In *Proceedings of the CVF/IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10492–10502, 2021.