AQUAFace: Age-Invariant Quality Adaptive Face Recognition for Unconstrained Selfie vs ID Verification

Shivang Agarwal[*1], **Jyoti Chaudhary**[*1†], **Sadiq Siraj Ebrahim**[1‡], **Mayank Vatsa**[1], **Richa Singh**[1], **Shyam Prasad Adhikari**[2§], **Sangeeth Reddy Battu**[2§]

[1]IIT Jodhpur, India,
[2]Swiggy, India
{shivangagarwal, jyotichaudhary, mvatsa, richa}@iitj.ac.in, ssiraje.etce.ug@jadavpuruniversity.in,
{all.shyam, reddysangeeth93}@gmail.com

Abstract

Face recognition in the presence of age and quality variations poses a formidable challenge. While recent margin-based loss functions have shown promise in addressing these variations individually, real-world scenarios such as selfie versus ID face matching often involve simultaneous variations of both age and quality. In response, we propose a comprehensive framework aimed at mitigating the impact of these variations while preserving vital identity-related information crucial for accurate face recognition. The proposed adaptive margin-based loss function AQUAFace exhibits adaptiveness towards hard samples characterized by significant age and quality variations. This loss function is meticulously designed to prioritize the preservation of identity-related features while simultaneously mitigating the adverse effects of age and quality variations on face recognition accuracy. To validate the effectiveness of our approach, we focus on the specific task of selfie versus ID document matching. Our results demonstrate that AQUAFace effectively handles age and quality differences, leading to enhanced recognition performance. Additionally, we explore the benefits of fine-tuning the recognition model with synthetic data, further boosting performance. As a result, our proposed model, AQUAFace, achieves state-of-the-art performance on six benchmark datasets (CALFW, CPLFW, CFP-FP, AgeDB, IJB-C, and TinyFace), each exhibiting diverse age and quality variations.

Code — https://sadiqebrahim.github.io/AQUAFace/

Introduction

Selfie versus ID face matching, a task centered around verifying a person's identity by comparing their face in a selfie with that in their official identification (ID) document, holds significant importance in various real-life scenarios. This verification process is pivotal in domains such as security, digital identification systems, and authentication processes. The task presents a multifaceted challenge due to the diverse array of factors influencing facial appearance, such as variations in image quality, resolution, illumination, facial expressions, and age-difference between selfies and ID document photos (Shi and Jain 2019; Albiero et al. 2020; Agarwal et al. 2023). Figure 1 depicts the impact of age and quality variations on the recognizability of face pairs in the context of selfie versus ID face matching. The partitioning of face pairs into easy, medium, and hard categories highlights the varying degrees of difficulty encountered in matching faces across different quality and age scenarios. It is evident that recognizability decreases notably when faced with combined age and quality variations, describing the intricate nature of the task.

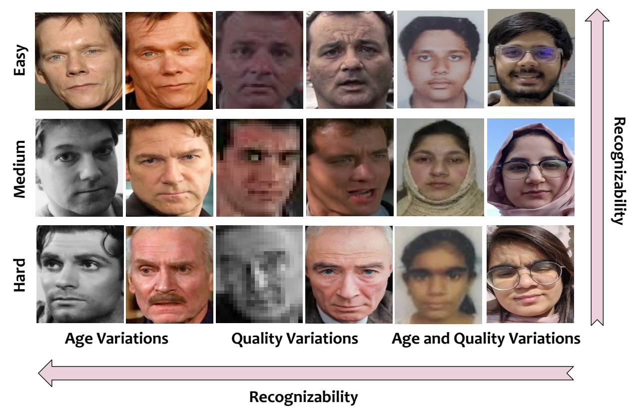

Figure 1: Examples of image pairs depicting variations in age, quality, and combined age + quality factors. The top row illustrates easy pairs with high recognizability, the middle row depicts pairs of medium difficulty in recognition, and the bottom row showcases hard pairs posing challenges in verification. It is evident that the combined effect of age and quality variations notably reduces recognizability.

[*]These authors contributed equally.

[†]Jyoti Chaudhary is a PhD candidate at IIT Patna (1821cs06@iitp.ac.in) and Research Associate at IIT Jodhpur.

[‡]Work done as part of summer internship at IIT Jodhpur.

[§]Work done while at Swiggy.

Margin-based classifiers have emerged as powerful tools in the domain of Face Recognition (FR), particularly in mitigating the impact of quality variations during the recognition process. These classifiers operate by defining a margin, or threshold, between classes in the feature space, thereby facilitating more effective discrimination between samples. ArcFace (Deng et al. 2022), for example, enforces a large margin between intra-class and inter-class variations, leading to improved performance by scaling feature vectors to lie on a hypersphere and encouraging greater class separation. AdaFace (Kim, Jain, and Liu 2022), on the other hand, dynamically adjusts the margin based on input sample quality, effectively pulling easy samples closer to the class center while pushing hard samples farther away, thus enabling more precise classification in scenarios with varying image quality. Recent advancements like MagFace (Meng et al. 2021) and QMagFace (Terhörst et al. 2023) have further extended the capabilities of margin-based classifiers by incorporating quality awareness into the training process. MagFace employs quality-aware loss functions to handle image quality variations, enhancing the system's robustness, while QMagFace integrates quality-aware features into the margin-based classification framework, facilitating more accurate recognition even in the presence of image quality variations.

Age-separated face pairs present a critical challenge in recognition, particularly in scenarios like selfie vs ID face verification. Previous studies have highlighted the detrimental effect of increasing age differences on FR accuracy, underscoring the need for robust solutions in this domain (Zhao, Yan, and Feng 2022). One approach to address this challenge involves developing explainable face verification models to understand the decision-making process of black box models prevalent in FR systems (Mery and Morris 2022; Popescu et al. 2022). However, despite efforts to explore the influencing regions in probe images and improve model transparency, significant gaps remain in understanding how age differences interact with quality-aware FR models (Xie, Pun, and Lam 2022). Thus, there exists a compelling scope for further research to delve into this intersection and develop more comprehensive solutions in age-invariant and quality-aware FR.

In light of the challenges posed by quality variations and age differences in face verification, this paper aims to address the following research questions:

RQ1: How effective are FR models specifically designed for quality-aware FR and age-invariant FR when confronted with these challenges in a combined manner, as observed in the task of selfie versus ID face verification?

RQ2: Is it feasible to develop a unified loss function capable of discerning between easy, medium, and hard samples based on variations in quality and age difference, and subsequently quantify their discriminability based on recognizability?

RQ3: Can synthetic data effectively address privacy concerns associated with real data in the context of selfie vs. ID face verification, and how effective is it for training FR models compared to real data?

Motivated by the observed inverse correlation between increasing age difference and decreasing facial quality, which directly influences face recognition performance (Popescu et al. 2022; Terhörst et al. 2023), our contributions are designed to address critical challenges in this domain. Combining individual losses addressing quality and age variations separately may introduce complexities that cause difficulty in maintaining subject identity. Therefore, a unified loss function is useful to effectively average the impacts of quality and age variations while preserving subject identity. Additionally, to streamline the loss function's complexity, it's crucial to incorporate recognizability metrics for training samples, considering both quality and age differences in a combined manner. Towards this end, the contributions of this paper are:

1. We present AQUAFace, a novel approach that adeptly handles age and quality differences in a unified manner, offering a comprehensive solution for FR tasks.
2. To overcome the privacy concerns imposed by private datasets, we generate Syn_AM, a synthetic dataset comprising $56,612$ images of $14,017$ subjects. The images are sourced from the AgeDB (Moschoglou et al. 2017) and MORPH (Ricanek and Tesafaye 2006) datasets to incorporate significant age and quality variations in a combined manner by leveraging advanced GAN-based techniques.
3. We assess and compare the effectiveness of the proposed method, trained on both real and synthetic data, against state-of-the-art FR models through comprehensive evaluations on 8 benchmark datasets with diverse age and quality variations.

Literature Review

General Face Recognition

In recent years, FR has become increasingly prevalent in various domains, playing a pivotal role in modern technological applications (Zhao et al. 2003; Jain and Li 2011). One prominent application is in the field of security, where FR systems are employed for access control, surveillance, and identity verification. Several notable methodologies have emerged recently within the domain of deep representational learning for FR, each offering distinct approaches to address the challenges inherent in this field (Chopra, Hadsell, and LeCun 2005; Schroff, Kalenichenko, and Philbin 2015; Wang et al. 2017; Deng et al. 2019; Meng et al. 2021). ArcFace (Deng et al. 2022) introduces an Additive Angular Margin Loss (ArcFace loss), enhancing the discriminative power of FR models by incorporating margins into the softmax loss function, effectively maximizing class separability. ElasticFace (Boutros et al. 2022) relaxes the fixed penalty margin constraint seen in prior marginal penalty softmax losses, enabling more flexible class separability learning by utilizing random margin values. EQFace (Liu and Tan 2021) introduces a method that generates a quantitative quality score during feature extraction, eliminating the need for annotated datasets. CurricularFace (Huang et al. 2020) incorporates adaptive curriculum learning into the loss function, enabling the model to handle easy and hard samples at different training stages. CAFace (Kim et al. 2022) presents a two-stage feature fusion paradigm, Cluster and Aggregate, for scalable and order-invariant fusion in unconstrained face recognition. MagFace (Meng et al. 2021) introduces loss functions that learn universal feature embeddings to mitigate performance degradation in diverse environments. QMagFace (Terhörst et al. 2023) enhances recognition by combining quality-aware comparison scores with a magnitude-aware angular margin loss, showing consistent improvements across various datasets. Despite these advancements, there is still a gap in the literature concerning methodologies that concurrently tackle both quality and age variations. Our paper fills this gap, enhancing the performance of face recognition systems, especially in applications like selfie versus ID face verification.

Selfie vs ID Face Matching

Initial research efforts on the selfie vs. ID face verification task (earlier referred to as scanned and digital face matching) have primarily focused on preprocessing frameworks aimed at enhancing the quality of scanned document images to minimize variance between the gallery and probe images (Starovoitov, Samal, and Sankur 2000; Vatsa et al. 2010; Bourlai, Ross, and Jain 2011). These preprocessing steps typically include Discrete Wavelet Transform (DWT) fusion and linear denoising for image quality enhancement, Eigen transforms and rank transformation to reduce variance between digital and scanned images, and mutual information-based functions to decrease age variance. Subsequent research has explored solutions based on two-stage training setups (Shi and Jain 2018; Zhu et al. 2019). More recently, researchers have proposed target domain invariant face recognition, where a model is trained on multiple source domains to adapt to any unconstrained target domain, such as selfies and ID documents (Guo et al. 2020). Addressing age and quality variations present in scanned images at the feature level is paramount for achieving robust and reliable face recognition systems. Our paper underscores the significance of this aspect by proposing a novel loss function that dynamically adapts to the utility of each sample for recognition, considering both quality and age differences.

Age-Invariant Quality Adaptive Face Recognition

AQUAFace Loss

Softmax based classifiers focus on sample-to-center comparison that results in lack of uniformity between training and testing scenarios. Therefore, there is a need for a quality-aware training process that is generalizable across all quality levels during testing (Liu and Tan 2021). A quality-aware function should adaptively guide the softmax center to pay attention to low quality samples. Similarly, an age-invariant function should emphasize samples with higher age-difference, and significant aging effect. By leveraging the linear relationship between qualities and comparison scores, as well as the inverse relationship between age-difference and comparison scores enforced by the employed loss function, our age-invariant quality-aware comparison function becomes highly generalizable and efficient particularly in addressing the complex challenges of selfie vs ID face verification. Angular margin-based softmax loss projects the centres in unit so that they capture the intra-class compactness and inter-class separability in a better manner (Deng et al. 2022). In an ideal scenario, margin based classifiers work on increasing the inter-class distances, but in case of unrecognisable samples, the angular disparity is usually very large, but that is because of poor image quality rather than the difference in identity. The target is to weigh samples for training based on their age-difference, quality and recognizability. The overview of the proposed AQUAFace model is presented in Figure 2.

AQUALR: To incorporate age and quality into the likelihood ratio, we utilize a Gaussian Mixture Model (GMM) to estimate the joint density of the comparison score alongside the associated quality and age. Unlike the approach in (Nandakumar et al. 2008), where quality is factored into the likelihood ratio by estimating the joint density of the match score and quality, our method extends this by also considering age, ensuring that the resulting likelihood ratios reflect the combined influence of both factors on the comparison score. In this way, we obtain likelihood ratio taking into account the quality and age labels. During training, at every n-th minibatch, we train GMM to obtain Age and QUAlity-based Likelihood Ratio (AQUALR). For each n-th iterations, we compute pairwise comparison scores of features extracted from ArcFace, and thereafter, at every n-th iteration, GMM is used to estimate the density of positive comparison scores $(\hat{f}_{pos})$ and negative comparison scores $(\hat{f}_{neg})$ along with their quality (q) and age (a) labels. AQUALR(x, q, a) is given by,

$$\text{AQUALR}(x, q, a) = \frac{\hat{f}_{pos}(x, q, a)}{\hat{f}_{neg}(x, q, a)} \tag{1}$$

Adaptive-margin based contrastive loss: We utilize contrastive loss (Hadsell, Chopra, and LeCun 2006) to enhance the distinction between positive and negative pairs within the embedding space. However, in contrast to the fixed margin introduced in (Hadsell, Chopra, and LeCun 2006), we incorporate an adaptive margin in the contrastive loss. We leverage AQUALR to dynamically adjust this margin, as it gauges the subject's likelihood of belonging to either the positive or negative pair, considering all age and quality variations. The adaptive-margin based contrastive loss is given as,

$$\mathcal{L}_{con_{ad}} = \frac{1}{2}(Y * D^2 + (1 - Y) * \max(0, m_{ad} - D)^2) \tag{2}$$

where, $Y = 1$ for positive pair and $Y = 0$ for negative pair. D is the distance metric calculated between the pair embeddings. We calculate cosine similarity as a distance metric and m_{ad} is the adaptive-margin, which is calculated as:

$$m_{ad} = m * \text{AQUALR}(x, q, a) \tag{3}$$

where, m is the fixed margin used in contrastive loss. The adaptive margin imposes a greater penalty on hard pairs because the AQUALR for these samples tends to be closer to

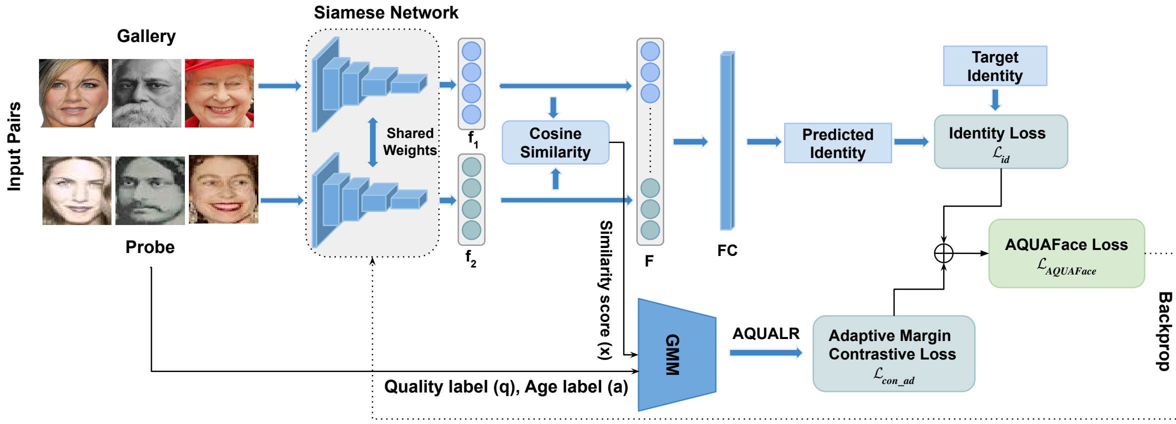

Figure 2: An overview of the proposed AQUAFace model for face recognition.

-1, thereby pushing hard pairs further apart from positive pairs.

Identity-preserving loss: The margin-based softmax loss demonstrates its effectiveness in learning more discriminative features related to subject identity. We use a state-of-the-art pre-trained FR model, ArcFace, and finetune it for the task of selfie versus ID matching. We utilize the ArcFace loss (Deng et al. 2022) to preserve the identity features, which is given by,

$$\mathcal{L}_{id} = -\frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)}+\sum_{j\neq y_i}e^{s\cos(\theta_j)}}\right) \tag{4}$$

where, N is the batch size, θ_{y_i} represents the cosine similarity between the input feature vector and the weight vector of the true class y_i, m is the angular margin, and s is a scaling factor. Finally, we optimize the contrastive loss and identity loss during training, and we term this combined loss as AQUAface loss, given as,

$$\mathcal{L}_{AQUAface} = \mathcal{L}_{con_{ad}} + \mathcal{L}_{id} \tag{5}$$

Results and Analysis

We present the Grad-CAM analysis diagram to highlight the need for addressing facial quality and age variations. Figure 3 illustrates the impact of quality and age differences on model attention. When there is minimal quality and age difference between images, attention remains focused on consistent facial regions, supporting reliable recognition. However, significant age differences and poor image quality lead to dispersed attention across the face. This highlights the need for effective strategies to manage quality and age variations in FR.

Experimental Setup

Quality estimation using an unsupervised assessment model: In training age-invariant quality-aware face recognition models, it is crucial to carefully choose verification pairs that offer the right level of difficulty—neither too easy nor too hard to recognize. This approach suggests training models on challenging samples, avoiding cases where the probe is unrecognizable and the model may learn random features for difficult instances (Kim, Jain, and Liu 2022; Ebrahimi Saadabadi et al. 2023). In light of this, our motivation is to partition the data into three bins—easy, medium, and hard—based on quality or age-difference, to enhance the model's discriminative ability across a range of recognition challenges.

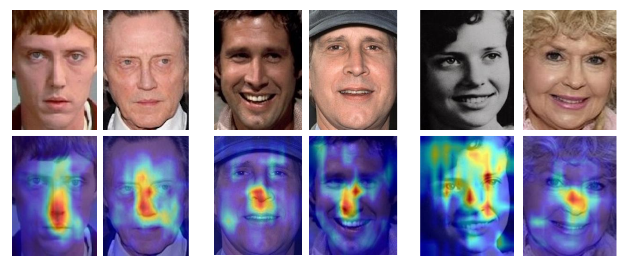

Figure 3: The motivation for addressing age-variations. Grad-CAM visualization demonstrates that attention is given to the key-regions of the face, irrespective of the subject's age. LightCNNv29 (Wu et al. 2018) is used as a backbone for visualization. Samples are taken from the AgeDB database (Moschoglou et al. 2017).

We use a facial quality assessment algorithm to accomplish this task. The unsupervised algorithm, SER-FIQ (Terhörst et al. 2020) takes an image as an input and yields a score in the range of $[0, 1]$, where 0 signifies the lowest quality and 1 denotes the highest quality.

Mean-standard deviation based partitioning: Based on the quality scores acquired by the quality assessment model, we compute the value of mean and standard deviation of the whole set, represented as

$$v_1 = \mu + 0.5\sigma \tag{6}$$

Next, the dataset is divided into two partitions, p_1 and p_2,

Dataset	Subjects	Images	+ Pairs	- Pairs	Age	Quality
LFW	5,749	13,233	3,000	3,000	×	High
CFP-FP	500	7,000	3,500	3,500	×	High
AgeDB	570	16,488	179,490	164,955	✓	High
CALFW	4,025	12,174	3,000	3,000	✓	High
CPLFW	3,930	11,652	3,000	3,000	×	High
IJB-B	1,845	12,115	10,270	8,000,000	×	Mixed
IJB-C	3,531	23,124	19,557	15,638,932	×	Mixed
TinyFace	5,139	169,403	12,308	588,530,780	×	Low
VerifyMe (ours)	12,100	48,400	36,300	363,000	✓	Mixed
Syn_AM (ours)	14,017	56,612	42,459	141,640	✓	Mixed

Table 1: Description of datasets.

based on the value of v_1.

$$p_1 = \{q_i | q_i >= v_1\} \; and \; p_2 = \{q_i | q_i < v_1\} \tag{7}$$

$$v_2 = \mu_{p_1} + 0.5\sigma_{p_1} \; and \; v_3 = \mu_{p_2} + 0.5\sigma_{p_2} \tag{8}$$

where μ_{p_1} and σ_{p_1} are mean and standard deviation of p_1, and μ_{p_2} and σ_{p_2} are mean and standard deviation of p_2. Based on the values of these variables, a query image I_q is classified as follows:

$$\begin{cases} I_q \in Easy(P_1), & \text{if } \{q_i | q_i >= v_2\} \\ I_q \in Medium(P_2), & \text{if } \{q_i | v_2 > q_i >= v_3\} \\ I_q \in Hard(P_3), & \text{if } \{q_i | q_i < v_3\} \end{cases} \tag{9}$$

Implementation details: The experiments are conducted on an Nvidia A40 GPU. We use a pre-trained ArcFace model (trained on CASIA-WebFace) with ResNet18 backbone and finetune it for 25 epochs on our dataset for the selfie vs ID face verification task. Dataset preprocessing involves random horizontal flipping, normalization, and resizing to 112×112 pixels to enhance generalization. The Stochastic Gradient Descent (SGD) optimizer is used, configured with a learning rate of 0.001, momentum of 0.9, and weight decay of $5e - 4$. We leverage the scikit-learn library to train GMM for every 120 iterations.

Datasets

Synthesized AgeDB + MORPH: A synthetic dataset containing $14,017$ subjects, made by leveraging single images of each identity from the AgeDB and MORPH datasets. The Lifespan Age Transformation Synthesis (LATS) model (Or-El et al. 2020) was used to perform age transformations on these images, generating synthetic images at three target ages: $15-19$, $30-35$, and $50-65$. To simulate real-world conditions, the generated images were further degraded using an algorithm akin to the one used in GFPGAN (Wang et al. 2021), introducing variations in image quality. There is a significant difference between the original images and the generated images in terms of quality and age. The images were degraded using the following degradation function.

$$I_{LQ} = [(I_{HQ} \otimes k) \downarrow_r + n_\delta]_{JPEG_q} \tag{10}$$

The images were first convolved with a blur kernel k, which includes Gaussian blur with standard deviation $\sigma \in 0:3$ and motion blur with kernel size $s \in 0:7$, followed by a downsampling operation of scale $r \in 0:3$. After which, additive Gaussian noise $\delta \in 0:3$ was added to the images, and then JPEG compression with quality factor $q \in 50:80$

Figure 4: Samples of AgeDB and MORPH datasets along with their synthesized counterparts. The first column of each figure represents the original image, while rest of the images are synthetically generated.

was applied. Finally, we resize the low-quality image back to 256×256. The dataset was organized into pairs, which were subsequently categorized into three bins—easy, medium, and hard. This categorization was based on two criteria: the age difference between the original and transformed images, and the quality of the probe images. Samples belonging to this dataset are represented in Figure 4. A ratio of $80:20$ is maintained while partitioning this dataset into training and testing sets. The pairs information associated with this dataset is provided in the supplementary file.

A private dataset, *VerifyMe* pertaining to $12,100$ subjects is used along with several publicly available datasets. For every subject, there are three scanned document images $(D_1, D_2 \; and \; D_3)$ and a selfie image. While a selfie is a spot face captured using an on-site camera device, the document images are acquired from IDs scanned by employees of an organization using their smartphones. Further details associated with this data are provided in the supplementary file.

Test Datasets: To assess the performance of the proposed FR model in comparison with current state-of-the-art methods, we adhere to a standard experimental protocol using established datasets. For high-quality evaluations, we employ datasets including LFW (Huang et al. 2008), CA-

FR Model	Selfie vs ID (Real)						Selfie vs ID (Synthetic)					
	Easy		Med		Hard		Easy		Med		Hard	
	0.1%	1.0%	0.1%	1.0%	0.1%	1.0%	0.1%	1.0%	0.1%	1.0%	0.1%	1.0%
LightCNN	67.66	79.23	49.89	68.71	17.08	32.21	62.5	78.3	45.8	66.7	15.0	30.5
ElasticFace	77.45	89.02	65.61	78.23	17.87	33.15	75.3	86.7	68.1	79.2	18.3	32.9
MagFace	73.29	85.46	53.93	73.25	14.18	27.66	73.6	85.3	51.3	74.9	12.3	26.5
AdaFace	64.69	79.23	42.54	64.17	10.66	26.80	65.3	80.1	45.6	63.8	11.5	23.3
CurricularFace	83.09	89.91	66.91	81.33	20.92	39.58	86.3	90.6	75.3	80.4	25.3	42.9
EQFace	84.87	92.58	77.14	87.45	23.12	48.12	84.3	91.5	79.7	87.3	26.4	45.6
ArcFace	29.38	49.85	24.01	41.24	6.66	16.14	28.3	46.8	15.3	32.9	5.6	20.2
AQUAFace (Real)	85.46	93.18	78.01	88.10	**51.33**	**66.93**	96.82	98.59	92.09	97.40	82.41	93.86
AQUAFace (Real+Syn)	**88.56**	**94.71**	**80.89**	**89.37**	42.25	59.38	**97.59**	**99.16**	**93.98**	**98.64**	**83.53**	**93.97**

Table 2: Performance evaluation of various face verification models on Real (VerifyMe) and Synthetic (Syn_AM) datasets after bin-wise partitioning. We report GAR@0.1%FAR and GAR@1%FAR in easy, medium and hard categories.

Training Data	LFW	CALFW	CPLFW	CFP-FP	AgeDB	IJB-B	IJB-C	TinyFace	
	Accuracy					TAR@FAR=0.01%		Rank-1	Rank-5
VerifyMe	99.40	96.36	93.66	98.75	98.46	94.83	96.45	69.35	74.94
Syn_AM	88.74	84.37	79.41	89.36	89.74	82.47	87.25	58.59	62.75
VerifyMe + Syn_AM	99.79	97.82	94.57	99.03	98.99	95.23	96.97	70.12	75.87

Table 3: Comparison on benchmark datasets with the AQUAFace model trained on different data types. The first row represents the model trained on real data (VerifyMe). The second row shows the model trained on the synthetic AgeDB and Morph datasets. The last row presents the model's performance when trained on VerifyMe and fine-tuned on the synthetic AgeDB and Morph. 1:1 verification accuracy is reported, following the protocol from (Kim, Jain, and Liu 2022).

LFW (Zheng, Deng, and Hu 2017), CP-LFW (Zheng and Deng 2018), CFP-FP (Sengupta et al. 2016), and Age-DB (Moschoglou et al. 2017). Mixed-quality performance is evaluated using the IJB-B (Whitelam et al. 2017) and IJB-C (Maze et al. 2018) datasets, while the TinyFace (Cheng, Zhu, and Gong 2018) dataset is utilized for low-quality assessments. A brief description of these datasets, including their characteristics and relevance, is provided in Table 1.

Results

Table 2 summarizes the performance evaluation of various state-of-the-art FR models on the *VerifyMe* and Syn_AM datasets, partitioned into easy, medium, and hard categories based on sample complexity related to quality and age differences. The metrics reported include GAR@0.1% FAR and GAR@1% FAR using the ResNet18 backbone.

The evaluation includes state-of-the-art quality-aware FR models alongside the proposed AQUAFace model. Results reveal that AQUAFace consistently outperforms other models across all categories. Specifically, AQUAFace demonstrates superior performance on synthetic data compared to real data, and it excels when trained on real data and fine-tuned on synthetic data. However, in the hard category for real data, the performance of the model fine-tuned on synthetic data is lower than that of the model trained solely on real data, indicating a deficiency in hard pairs within the synthetic data. This highlights the need for future research to focus on generating hard pairs in synthetic datasets to improve model performance in challenging scenarios.

Table 3 provides a comparative analysis of the AQUAFace model's performance when trained on different types of data: real data (VerifyMe), synthetic data (Syn_AM, which includes synthetic counterparts of AgeDB and Morph datasets), and a combination of both. The table is structured into three rows, each representing a different training approach. The first row shows the performance of AQUAFace when trained solely on the real dataset, VerifyMe. The accuracy reported here serves as a baseline for comparison. In the second row, the model is trained exclusively on synthetic datasets (Syn_AM), which includes synthetic aging and quality variations in AgeDB and Morph. The results indicate the challenges of relying solely on synthetic data, as it fails to capture the complexities of age and quality variations found in real-world data, leading to the lowest performance among the three methods. The third row presents the results of a hybrid approach, where the model is initially trained on the real dataset (VerifyMe) and then fine-tuned on the synthetic datasets (Syn_AM). This combination yields the highest performance, demonstrating the effectiveness of synthetic data in enhancing model accuracy when used for fine-tuning. The real data provides a strong foundation by capturing authentic age and quality variations, while the synthetic data adds value by introducing further intra-class variations.

The performance reported across standard high quality, low quality, and mixed quality datasets highlights the potential of synthetic data in improving face recognition models, particularly when used as a supplementary tool for fine-

Method	Train Data	High Quality					Mixed Quality		Low Quality	
		LFW	CALFW	CPLFW	CFP-FP	AgeDB	IJB-B	IJB-C	TinyFace	
		Accuracy					TAR@FAR=0.01%		Rank-1	Rank-5
Backbone-R100										
CosFace	MS1MV2	99.81	95.76	92.28	98.12	98.11	94.80	96.37	61.27	65.14
ArcFace	MS1MV2	99.83	95.45	92.08	98.27	98.28	94.25	96.03	63.81	68.05
CurricularFace	MS1MV2	99.80	96.20	93.13	98.37	98.32	94.80	96.10	63.68	67.65
MagFace	MS1MV2	99.83	96.15	92.87	98.46	98.17	94.51	95.97	64.32	69.97
AdaFace	MS1MV2	99.82	96.08	93.53	98.49	98.05	95.67	96.89	68.21	71.54
Backbone-R18										
ArcFace	CASIA-WebFace	99.14	92.25	90.78	91.63	94.59	91.69	93.52	59.71	63.58
MagFace	CASIA-WebFace	99.35	91.64	89.38	92.85	93.30	91.44	92.93	61.63	66.27
AdaFace	CASIA-WebFace	99.13	92.65	87.00	92.59	92.72	90.61	91.29	61.34	65.31
AQUAFace	VerifyMe+Syn_AM	99.79	**97.82**	**94.57**	**99.03**	**98.99**	95.23	**96.97**	**70.12**	**75.87**

Table 4: Performance comparison of recent methods on benchmark datasets with the AQUAFace model for ResNet100 and ResNe18 backbone. For high quality datasets, 1:1 verification accuracy is reported, following the protocol from (Kim, Jain, and Liu 2022). For mixed quality datasets, TAR@FAR=0.01% is reported. For TinyFace, closed-set rank retrieval (Rank-1 and Rank-5) is reported. Note: The performance of the pretrained models on MS1MV2 is sourced from their respective papers.

tuning rather than as the primary training source.

Table 4 presents a performance comparison of recent FR methods against the AQUAFace model using a ResNet18 and ResNet100 backbones, evaluated across different benchmark datasets. We report $1:1$ verification accuracy, following the protocol from (Kim, Jain, and Liu 2022) in the form of True Accept Rate (TAR) at a False Accept Rate (FAR) of 0.01%, and closed-set rank retrieval, including Rank-1 and Rank-5 accuracy.

These results highlight that the proposed AQUAFace model outperforms current state-of-the-art FR models across all dataset categories, including high-quality, mixed-quality, and low-quality datasets. This performance is achieved by addressing both age and quality variations simultaneously, unlike existing models that focus on either quality-aware or age-invariant FR. For additional details on the performance with a ResNet100 backbone, please refer to the supplementary file.

Ablation Analysis

In our ablation analysis, we examine the impact of the hyperparameter m on the model's performance. Here, m represents the maximum range of the angular margin as well as the magnitude of the additive margin. From the results presented in Table 5, we observe that the optimal performance is achieved with $m = 0.5$. The motivation behind our ablation study lies in understanding the influence of the margin parameter 'm' on the treatment of quality and age-difference in training samples. By adjusting 'm', we aim to balance the emphasis on low-quality samples, typically associated with higher age-difference, against high-quality samples with minimal age-difference. This adjustment is crucial as it directly impacts the positioning of the softmax center and the delineation of class boundaries, ultimately affecting the discriminative power of the model.

m	VerifyMe			Syn_AM		
	Easy	Med	Hard	Easy	Med	Hard
0.2	73.42	63.83	30.25	89.76	84.36	42.96
0.5	74.49	66.54	39.39	91.46	86.32	46.47
1.0	71.41	65.92	37.01	86.17	85.26	45.39

Table 5: Effect of hyperparameter 'm'.

Conclusions

In conclusion, this paper addresses the challenges posed by age and quality variations in selfie vs. ID face verification. We provide a comprehensive analysis of the effectiveness of state-of-the-art, quality-aware, and age-invariant FR models in tackling these complexities. To address these challenges, we introduce the AQUAFace model, a novel approach that unifies the handling of age and quality variations while preserving the subject's identity. Our proposed model incorporates a unified loss function that balances the impact of these variations, effectively capturing intricate details and adjusting the margin to prioritize samples with poor quality and significant age differences. In addition to these contributions, we evaluated the performance of the proposed model when trained solely on real data, solely on synthetic data, and when fine-tuned on synthetic data after being trained on real data. Our experiments reveal that synthetic data can enhance performance when used for fine-tuning alongside real data. However, the fine-tuned model's performance lags behind its pretrained counterpart when tested on the hard category of real data, highlighting the need for generating hard samples in synthetic datasets. AQUAFace outperforms current state-of-the-art models across high-quality, low-quality, and mixed-quality datasets, validating its effectiveness in unconstrained selfie vs. ID face verification and achieving state-of-the-art performance on CALFW, CPLFW, CFP-FP, AgeDB, IJB-C, and TinyFace.

Acknowledgments

This research was partially supported by Swiggy, India. Ebrahim was supported by the Uplink Internship Program of the India Chapter of ACM SIGKDD.

References

Agarwal, S.; Chaudhary, J.; Savani, H.; Sharma, S.; Vatsa, M.; Singh, R.; Adhikari, S. P.; Reddy, S.; Agrawal, K.; and Misra, H. 2023. Leveraging Synthetic Data and Hard Pair Mining for Selfie vs ID Face Verification. In *IJCB*.

Albiero, V.; Srinivas, N.; Villalobos, E.; Perez-Facuse, J.; Rosenthal, R.; Mery, D.; Ricanek, K.; and Bowyer, K. W. 2020. Identity Document to Selfie Face Matching Across Adolescence. In *IJCB*.

Bourlai, T.; Ross, A.; and Jain, A. K. 2011. Restoring Degraded Face Images: A Case Study in Matching Faxed, Printed, and Scanned Photos. *IEEE TIFS*, 6(2): 371–384.

Boutros, F.; Damer, N.; Kirchbuchner, F.; and Kuijper, A. 2022. ElasticFace: Elastic Margin Loss for Deep Face Recognition. In *IEEE CVPR Workshops*.

Cheng, Z.; Zhu, X.; and Gong, S. 2018. Low-Resolution Face Recognition. In *Computer Vision - ACCV 2018 - 14th Asian Conference on Computer Vision, Perth, Australia, December 2-6, 2018, Revised Selected Papers, Part III*.

Chopra, S.; Hadsell, R.; and LeCun, Y. 2005. Learning a similarity metric discriminatively, with application to face verification. In *CVPR*.

Deng, J.; Guo, J.; Xue, N.; and Zafeiriou, S. 2019. Arcface: Additive angular margin loss for deep face recognition. In *CVPR*.

Deng, J.; Guo, J.; Yang, J.; Xue, N.; Kotsia, I.; and Zafeiriou, S. 2022. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. *IEEE TPAMI*, 44(10): 5962–5979.

Ebrahimi Saadabadi, M. S.; Rahimi Malakshan, S.; Zafari, A.; Mostofa, M.; and Nasrabadi, N. M. 2023. A Quality Aware Sample-to-Sample Comparison for Face Recognition. In *WACV*.

Guo, J.; Zhu, X.; Zhao, C.; Cao, D.; Lei, Z.; and Li, S. Z. 2020. Learning Meta Face Recognition in Unseen Domains. In *CVPR*.

Hadsell, R.; Chopra, S.; and LeCun, Y. 2006. Dimensionality Reduction by Learning an Invariant Mapping. In *CVPR*.

Huang, G. B.; Mattar, M.; Berg, T.; and Learned-Miller, E. 2008. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*.

Huang, Y.; Wang, Y.; Tai, Y.; Liu, X.; Shen, P.; Li, S.; Li, J.; and Huang, F. 2020. CurricularFace: Adaptive Curriculum Learning Loss for Deep Face Recognition. In *CVPR*.

Jain, A. K.; and Li, S. Z. 2011. *Handbook of face recognition*, volume 1. Springer.

Kim, M.; Jain, A. K.; and Liu, X. 2022. AdaFace: Quality Adaptive Margin for Face Recognition. In *CVPR*.

Kim, M.; Liu, F.; Jain, A. K.; and Liu, X. 2022. Cluster and Aggregate: Face Recognition with Large Probe Set. In *NeurIPS*.

Liu, R.; and Tan, W. 2021. EQFace: A Simple Explicit Quality Network for Face Recognition. In *CVPR*.

Maze, B.; Adams, J. C.; Duncan, J. A.; Kalka, N. D.; Miller, T.; Otto, C.; Jain, A. K.; Niggel, W. T.; Anderson, J.; Cheney, J.; and Grother, P. 2018. IARPA Janus Benchmark - C: Face Dataset and Protocol. In *2018 International Conference on Biometrics, ICB 2018, Gold Coast, Australia, February 20-23, 2018*, 158–165.

Meng, Q.; Zhao, S.; Huang, Z.; and Zhou, F. 2021. MagFace: A Universal Representation for Face Recognition and Quality Assessment. In *CVPR*.

Mery, D.; and Morris, B. 2022. On Black-Box Explanation for Face Verification. In *WACV*.

Moschoglou, S.; Papaioannou, A.; Sagonas, C.; Deng, J.; Kotsia, I.; and Zafeiriou, S. 2017. AgeDB: The First Manually Collected, In-the-Wild Age Database. In *CVPR Workshops*.

Nandakumar, K.; Chen, Y.; Dass, S. C.; and Jain, A. K. 2008. Likelihood Ratio-Based Biometric Score Fusion. *IEEE TPAMI*, 30(2): 342–347.

Or-El, R.; Sengupta, S.; Fried, O.; Shechtman, E.; and Kemelmacher-Shlizerman, I. 2020. Lifespan Age Transformation Synthesis. In Vedaldi, A.; Bischof, H.; Brox, T.; and Frahm, J., eds., *ECCV*, volume 12351, 739–755. Springer.

Popescu, A.; Stefan, L.; Deshayes-Chossart, J.; and Ionescu, B. 2022. Face Verification with Challenging Imposters and Diversified Demographics. In *WACV*.

Ricanek, K.; and Tesafaye, T. 2006. Morph: A longitudinal image database of normal adult age-progression. In *FG*.

Schroff, F.; Kalenichenko, D.; and Philbin, J. 2015. Facenet: A unified embedding for face recognition and clustering. In *CVPR*.

Sengupta, S.; Chen, J.-C.; Castillo, C.; Patel, V. M.; Chellappa, R.; and Jacobs, D. W. 2016. Frontal to profile face verification in the wild. In *WACV*.

Shi, Y.; and Jain, A. K. 2018. DocFace: Matching ID Document Photos to Selfies. In *IEEE BTAS*.

Shi, Y.; and Jain, A. K. 2019. DocFace+: ID Document to Selfie Matching. *IEEE TBIOM*, 1(1): 56–67.

Starovoitov, V.; Samal, D.; and Sankur, B. 2000. Matching of faces in camera images and document photographs. In *ICASSP*.

Terhörst, P.; Kolf, J. N.; Damer, N.; Kirchbuchner, F.; and Kuijper, A. 2020. SER-FIQ: Unsupervised Estimation of Face Image Quality Based on Stochastic Embedding Robustness. In *CVPR*.

Terhörst, P.; Ihlefeld, M.; Huber, M.; Damer, N.; Kirchbuchner, F.; Raja, K.; and Kuijper, A. 2023. QMagFace: Simple and Accurate Quality-Aware Face Recognition. In *WACV*.

Vatsa, M.; Singh, R.; Bharadwaj, S.; Bhatt, H. S.; and Noore, A. 2010. Matching digital and scanned face images with age variation. In *IEEE BTAS*.

Wang, J.; Zhou, F.; Wen, S.; Liu, X.; and Lin, Y. 2017. Deep metric learning with angular loss. In *ICCV*.

Wang, X.; Li, Y.; Zhang, H.; and Shan, Y. 2021. Towards Real-World Blind Face Restoration With Generative Facial Prior. In *CVPR*.

Whitelam, C.; Taborsky, E.; Blanton, A.; Maze, B.; Adams, J. C.; Miller, T.; Kalka, N. D.; Jain, A. K.; Duncan, J. A.; Allen, K.; Cheney, J.; and Grother, P. 2017. 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*, 592–600.

Wu, X.; He, R.; Sun, Z.; and Tan, T. 2018. A Light CNN for Deep Face Representation With Noisy Labels. *IEEE TIFS*, 13(11): 2884–2896.

Xie, J.; Pun, C.; and Lam, K. 2022. Implicit and Explicit Feature Purification for Age-Invariant Facial Representation Learning. *IEEE TIFS*, 17: 399–412.

Zhao, J.; Yan, S.; and Feng, J. 2022. Towards Age-Invariant Face Recognition. *IEEE T-PAMI*, 44(1): 474–487.

Zhao, W.; Chellappa, R.; Phillips, P. J.; and Rosenfeld, A. 2003. Face recognition: A literature survey. *ACM computing surveys (CSUR)*, 35(4): 399–458.

Zheng, T.; and Deng, W. 2018. Cross-pose LFW: A database for studying cross-pose face recognition in unconstrained environments. Technical report, Beijing University of Posts and Telecommunications.

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

Zhu, X.; Liu, H.; Lei, Z.; Shi, H.; Yang, F.; Yi, D.; and Li, S. Z. 2019. Large-scale Bisample Learning on ID vs. Spot Face Recognition. *Int J Comput Vis*, 127: 684–700.