Improving Low-Resolution Face Recognition under Limited Data: How Synthetic Data Generation Can Close the Domain Gap

Luis S. Luevano Ünsal Öztürk Hatef Otroshi Shahreza Anjith George Sébastien Marcel
Idiap Research Institute, Switzerland
{luis.luevano, unsal.ozturk, hatef.otroshi, anjith.george, sebastien.marcel}@idiap.ch

Abstract

Face Recognition (FR) systems in surveillance settings often encounter Low Resolution *(LR) faces, those whose face region falls below the standard* 112×112 *input size. While labelled High Resolution (HR) training data is abundant, labelled* native-*LR data, and above all* paired *native-LR/HR data, is scarce. One workaround is to* synthesize *LR data from the available HR faces, but how much synthesis effort is repaid in recognition accuracy remains unclear. We present a study of* simple *synthetic generation strategies for a compact, edge device-oriented face recognition system, spanning interpolation-based degradation, knowledge distillation, a Prepended Domain Transformer (PDT), Real-ESRGAN-style degradation, and a learned Super Resolution (SR) front-end with an identity-aware loss. We evaluate these strategies on synthetic cross-resolution face benchmarks (LFW, CFP-FP, AgeDB-30) and on TinyFace, a* real-world *native LR dataset, and expose a* synthetic–real gap*: the degradation setting that is optimal on synthetic benchmarks is not the one that is optimal on real LR. We find that more synthesis effort does not help monotonically: the learned SR front-end does not surpass a direct feed of the aligned LR image into a strong backbone, while simple interpolation augmentation of a compact backbone is the only synthesis that improves over its own baseline. We conclude that generative methods for LR face recognition must be validated on real LR and against a direct-feed baseline, and release our pipeline at* https://idiap.ch/paper/synth-lrfr.

1. Introduction

Modern face recognition (FR) systems are highly accurate on High Resolution (HR) images in uncontrolled conditions, even on edge devices [6, 17, 20, 21]. Real-world deployments, however, frequently encounter Low Resolution (LR) faces as subjects can be captured at long distances and at different heights. After face detection and alignment, the usable facial region may span only a few dozen pixels (Fig. 1), and upsampling naively does not restore identity information [16]. More importantly, this LR setting suffers from *limited data*, and the scarcity is specific: labelled HR training data is plentiful, whereas labelled *native*-LR identities, and above all *paired* native-LR/HR images of the same subject, are costly to collect and largely unavailable. Beyond average accuracy, this setting also raises fairness concerns, since degradation can affect demographic groups unequally. Synthesizing LR data and understanding effective degradation strategies for Low Resolution Face Recognition is therefore highly relevant to improve recognition performance under LR-side data scarcity.

Figure 1. A verification pair at decreasing target resolution, each upsampled back to 112 px. Identity cues vanish progressively.

In critical surveillance applications (e.g. watchlist screening, forensics), it is typical to match a high-quality enrollment (a cooperative capture, an identity document, or a mugshot) against a probe that may be only 16–32 pixels across. This process therefore involves placing an LR probe into the same embedding space as an HR gallery template, which we call the HR→LR setting. We also study the symmetric LR→LR case (both images degraded) as a harder bound. The artifacts present when capturing probes at low resolutions include residual blur, sensor noise, and compression, forming a domain shift to which an HR-trained

recognizer was never exposed. This is also why adapting the probe is often an employed strategy for LR FR.

A common approach synthesizes LR training data by downsampling HR faces and trains (or adapts) the recognizer on it. The synthesis is usually a fixed down-/upsampling chain, and methods are validated on *synthetically* degraded versions of standard verification sets because paired native-LR data is costly to collect. This poses the question: *does a method tuned on synthetic LR actually work on real LR?*

We answer this for a compact, edge device-oriented backbone, EdgeFace [6], and we use the answer to motivate and then evaluate generative alternatives. We compare five adaptation strategies (Sec. 3) under the following protocol: interpolation augmentation, knowledge distillation, a native 32 px stem trained on Real-ESRGAN-style degradation, a Prepended Domain Transformer (PDT) [8], and a learned super-resolution front-end. To find the synthetic–real gap in performance, we evaluate everything on *real* native-LR TinyFace under two alignment pipelines and compare it to synthetic LR FR benchmarks.

Our contributions are the following:

- We present a systematic study of synthetic data generation for LR face recognition on a compact, edge device-oriented backbone under limited LR data, under various generation efforts: deterministic interpolation, simulating Real-ESRGAN-style degradation, and a learned identity-aware super-resolution front-end.
- We provide an analysis on the *synthetic–real domain gap* by evaluating the same models on synthetic cross-resolution benchmarks (LFW, CFP-FP, AgeDB-30) and on the real native-LR set TinyFace, isolating how the choice of degradation and target resolution governs transfer to genuinely captured LR faces.
- We release a pipeline and provide practical guidance for matching generation effort to the deployment setting (compact versus strong backbone), arguing that low-resolution face recognition should be validated on real LR and against a direct-feed strong-backbone baseline rather than on synthetic degradation alone.
- We complement the accuracy study with a demographic fairness assessment on RFW, and find that the average-accuracy gains from LR-aware synthesis do not translate into reduced demographic bias.

2. Related Work

Efficient face recognition. Margin-based losses such as CosFace [27] and ArcFace [4], trained on web-scale data [31] with scalable classification heads [1], define modern face recognition. A related line of work compresses these models for edge computing devices: EdgeFace [6] adapts the hybrid CNN–Transformer EdgeNeXt [18] with a low-rank linear layer, and lightweight designs were compared in the EFaR 2023 competition [13]. We adopt EdgeFace for benchmarking our synthesis approaches, as the LR FR problem is more pronounced when using such compact and deployable models, compared to their performance in HR FR.

Low- and cross-resolution face recognition. A large body of work targets recognition when probes are degraded, most commonly by exposing the recognizer to degraded data during training so that high- and low-resolution embeddings become comparable [22]. Approaches fall into three families: resolution-robust *training* (downsampling augmentation, cross-resolution losses, feature adaptation) [22]; *distillation* of an HR teacher into an LR student [5, 10]; and *recovery*, in which a super-resolution or restoration module reconstructs detail before recognition. The synthetic degradation used to build the training and test data is usually treated as an implementation detail; we instead show that this choice governs whether the learned robustness transfers to real LR faces, and we evaluate one method from each family on the same real native-LR protocol.

Super-resolution and blind face restoration for recognition. Sub-pixel convolution [26] and GAN-based super-resolution [30] recover high-frequency detail, and blind face restoration models such as FaceMe [14] reconstruct plausible faces from severely degraded inputs. Restoration can improve downstream recognition on low-quality images [2, 19], but when optimized for perceptual quality alone it can hallucinate identity-inconsistent detail and even reduce accuracy. We therefore tie our super-resolution objective to the recognizer embedding and keep the front-end light enough for an edge model.

Generative degradation synthesis. Real-ESRGAN [29] models real-world degradation with a high-order random pipeline of blur, resizing, noise, and compression. We repurpose this generator to *synthesize* realistic low-resolution training faces, a form of synthetic data generation suited to our setting, where paired native-LR faces are unavailable.

Heterogeneous face recognition and domain transformers. When low resolution is treated as a separate modality, the recognizer can be frozen and only the input adapted. The Prepended Domain Transformer (PDT) [8] and Domain Invariant Units [7] follow this recipe, attractive because the expensive backbone is never retrained. We evaluate how far such input-side adaptation transfers to genuine native-LR data.

3. Synthetic Data Generation and Adaptation Strategies

We frame low-resolution adaptation as two coupled choices: how much effort to spend *synthesizing* the LR domain, and how to *adapt* the recognizer to it. We organize our methods along the following synthesis efforts: *low* is a fixed interpolation chain (downsample, then upsample back to 112 px), with no learned or stochastic component; *medium* is a stochastic, high-order degradation model (Real-ESRGAN-style: blur, resizing, noise, and compression) that mimics real sensor and optics degradation; and *high* adds a learned generative inverse, an identity-aware super-resolution (SR) front-end that reconstructs an HR-like image. These data-side choices are orthogonal to the recognizer-side adaptation (retraining the backbone, distilling from a teacher, or freezing it and adapting only the probe), which we vary independently; all models use the compact EdgeFace-S backbone unless a frozen, stronger backbone is stated.

All synthetic data are derived from HR data, with each setting denoted "$s\downarrow d/\uparrow u$" (downsample to s px with d, upsample to 112 px with u; e.g. "28↓c/↑a" is cubic-down, area-up). For training, interpolation yields one set per setting and the Real-ESRGAN pipeline stores a native 32 px copy paired with its 112 px source. For evaluation, we degrade the verification pairs of standard benchmarks in two modalities, HR→LR (one image degraded) and LR→LR (both), and also test on the real native-LR set.

3.1. Synthesis strategies

Low effort: interpolation augmentation. The simplest adaptation downsamples WebFace4M [31] faces to $s \in \{56, 28, 14, 7\}$ with $\downarrow\in$ {area,cubic}, upsamples back to 112 px with $\uparrow\in$ {area,cubic}, and retrains EdgeFace-S end-to-end with CosFace [27] (PartialFC [1], AdamW [15], 100 epochs). This scheme does not make use of HR data, and considers LR to be purely augmentation. Our comparison set is formed by the combinations of four reliably converging resolution/interpolation settings over {area,cubic}×{area,cubic}.

Medium effort: generative degradation synthesis. To synthesize realistic LR training faces under limited native-LR data, we port the Real-ESRGAN [29] high-order degradation generator. Per image, a first-order block applies a random blur kernel (iso-/anisotropic Gaussian, generalized Gaussian, plateau, or sinc), a random resize (factor $\in [0.15, 1.5]$), Gaussian/Poisson noise, and a JPEG cycle; a second-order block (probability 0.8) repeats blur/resize/noise; a final resize to 32×32 adds an optional sinc low-pass and JPEG. We store the 32 px LR JPEGs alongside the original 112 px HR. We also train a *native* 32 *px variant*: EdgeFace-S with its 4×4 stride-4 stem replaced by a 1×1 convolution, taking the 32 px image directly without upsampling.

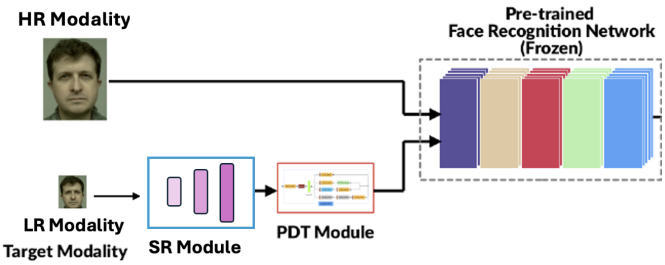


Figure 2. Input-side adaptation: a learned SR upsampler and a PDT translator map a 32 px native-LR face into the input domain of a *frozen* backbone. Only the upsampler is pre-trained (identity-aware in Stage 1), then both upsampler and PDT module are trained with contrastive learning (Stage 2).

High effort: learned super-resolution front-end. We move upsampling *inside* the model: the image stays at 32 px and the upscale operator is learned. Two ×2 sub-pixel (PixelShuffle) stages [26] take $32 \to 64 \to 128$ px, then four valid-padded k=5 convolutions trim 16 px ($128 \to 112$); no interpolation function is used. Two feature-block variants of a plain sub-pixel stack (ESPCN-style, ≈0.86 M parameters) and an RRDB trunk (ESRGAN/Real-ESRGAN-style [29, 30], ≈6.5 M) are studied. The upsampler is prepended to a PDT translator in front of a *frozen* backbone (EdgeFace-base, WebFace12M-pretrained), giving the path `Upsampler`→`PDT`→frozen backbone (Fig. 2). *Stage 1 (SR pretraining)* regresses on (LR-32, HR-112) pairs with an identity-aware perceptual term,

$$\mathcal{L}_{\mathrm{SR}} = \ell_1\big(\mathrm{SR}(x_{\mathrm{LR}}), x_{\mathrm{HR}}\big) + \lambda_{\mathrm{id}}\big(1 - \cos(\phi(\mathrm{SR}(x_{\mathrm{LR}})), \phi(x_{\mathrm{HR}}))\big) \quad (1)$$

where ℓ_1 is the pixel-wise L_1 distance, ϕ is the frozen backbone embedding, and λ_{id} controls the impact of the identity constraint. *Stage 2 (joint contrastive)* warm-starts from Stage 1 and optimizes a contrastive loss [9] over $P \times K$-sampled HR/LR batches, stepping the upsampler and PDT jointly while the backbone stays frozen.

3.2. Baselines

HR and synthetic LR backbones. Our main reference point is the *direct feed* baseline, in which the LR face is aligned and resized to the input layer size and then passed to an HR-trained backbone (EdgeFace-S or base) with *no* particular front-end and no LR-specific training. It is the simplest deployment and, as we show, a hard baseline. We report it for both backbones so that every adaptation can be compared with the direct feed approach of the same model.

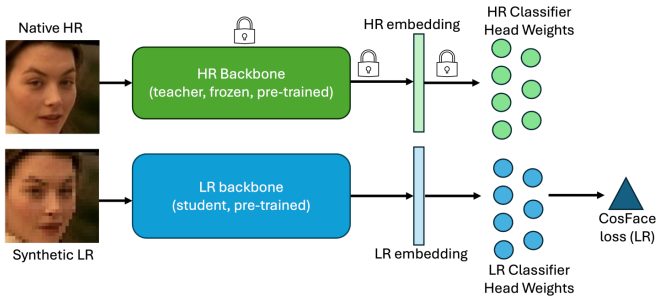


Figure 3. Adopted Knowledge Distillation (KD) paradigm using backbones trained on HR and LR data. The LR backbone is trained to produce embeddings similar to the HR version from LR data.

Knowledge distillation (KD). A *frozen* HR-trained EdgeFace-S teacher guides an LR student of identical architecture (initialized from the LR backbone) via $\mathcal{L} = \mathcal{L}_{\text{CosFace}} + \lambda_e \|\hat{z}_s - \hat{z}_t\|_2^2 + \lambda_w \|\hat{W}_s - \hat{W}_t\|_2^2$, with $\hat{z}$ the L2-normalized embeddings and $\hat{W}$ the L2-normalized sampled PartialFC weights. We include KD as a retraining baseline (illustrated in Fig. 3).

4. Experimental Setup

Training data. WebFace4M [31], in HR, the interpolation LR variants (Sec. 3), and the Real-ESRGAN 32 px synthesis. **Backbones.** EdgeFace-S (3.65 M) for compact models; EdgeFace-base (frozen, WebFace12M) for the SR+PDT front-end. **Synthetic benchmarks.** LFW [11], CFP-FP [25], AgeDB-30 [24] degraded to $56/28/14/7$ px in two face verification modalities: *HR→LR* (one image degraded, one HR) and *LR→LR* (both degraded). We also report standard HR verification and IJB-C [23] TAR@FAR. **Real native-LR.** TinyFace [3] identification (mAP, Rank-k; gallery ∼157k) under two pipelines: aligned crops with padding (`aligned_pad`) and the Differentiable Face Aligner (DFA, ResNet-50) from CVLface [12]. All reported accuracies use horizontal-flip as an augmentation at test time and are given as percentages.

5. The Synthetic–Real Gap

We present our findings comparing baseline results on synthetic cross-resolution matching and synthetic training to real native LR evaluation sets.

5.1. Synthetic cross-resolution

Table 1 reports the macro-mean accuracy over LFW, CFP-FP and AgeDB-30 for the synthetic cross-resolution sets, by modality. Most LR-aware settings outperform the HR baseline, though not all: under HR→LR the 28↓a/↑c setting falls marginally below it for the LR backbone and for KD. The strongest single configuration is 28↓c/↑a, and KD yields the largest gains. LR→LR is consistently harder

Table 1. Synthetic cross-resolution macro-mean accuracy (mean over LFW, CFP-FP, AgeDB-30), by modality. Best LR-trained value per row in **bold**. 28↓c/↑a is best for every method under HR→LR, and for all but PDT under LR→LR.

Method	HR	56↓c/↑a	28↓a/↑c	28↓c/↑a	14↓a/↑c
HR → LR					
LR backbone	78.85	80.25	78.77	**80.74**	79.67
KD	78.85	80.74	78.60	**82.19**	79.82
PDT	78.85	79.86	80.35	**81.22**	79.92
LR → LR					
LR backbone	78.28	78.92	78.74	**79.06**	78.31
KD	78.28	79.14	78.91	**80.59**	78.51
PDT	78.28	78.81	**79.10**	79.08	78.54

Table 2. Per-benchmark synthetic HR→LR mean accuracy. 28↓c/↑a is best for every synthetic benchmark and method (**bold**).

Bench.	Method	HR	56↓c/↑a	28↓a/↑c	28↓c/↑a	14↓a/↑c
LFW	LRb	82.21	83.83	81.70	**84.55**	82.98
	KD	82.21	84.36	81.90	**86.39**	83.43
	PDT	82.21	83.78	84.47	**85.79**	84.02
CFP-FP	LRb	77.57	78.53	77.80	**78.74**	77.59
	KD	77.57	79.09	77.82	**80.87**	77.59
	PDT	77.57	78.10	78.51	**79.16**	78.14
AgeDB	LRb	76.76	78.39	76.82	**78.93**	78.44
	KD	76.76	78.76	76.07	**79.32**	78.44
	PDT	76.76	77.71	78.06	**78.70**	77.61

Table 3. Per-benchmark synthetic *LR→LR* mean accuracy. **Bold**: best LR-trained per row.

Bench.	Method	HR	56↓c/↑a	28↓a/↑c	28↓c/↑a	14↓a/↑c
LFW	LRb	84.83	**85.21**	84.88	85.06	84.23
	KD	84.83	85.20	85.47	**86.34**	84.83
	PDT	84.83	85.13	**85.56**	85.45	85.54
CFP-FP	LRb	77.41	77.89	77.41	**78.26**	77.53
	KD	77.41	78.04	77.57	**80.05**	77.53
	PDT	77.41	78.09	**78.41**	78.25	77.84
AgeDB	LRb	72.60	73.66	**73.93**	73.86	73.17
	KD	72.60	74.19	73.70	**75.38**	73.17
	PDT	72.60	73.22	73.34	**73.54**	72.25

than HR→LR at the same resolution: on the 28↓c/↑a column the gap is largest for PDT, followed by the LR backbone and KD, and comes mainly from AgeDB-30 and CFP-FP. A single HR reference in the pair is an advantage that vanishes when both faces are degraded, which is one reason why synthetic HR→LR results look optimistic. Table 2 shows the same pattern per benchmark: under HR→LR, 28↓c/↑a is best on *every* benchmark for all three methods. Under LR→LR (Table 3) it still leads on aggregate, but the best setting per benchmark is less consistent (e.g. 56↓c/↑a edges ahead on LFW).

Table 4. The same EdgeFace-S models, *synthetic* vs *real* LR. Synthetic: HR→LR macro-mean Acc. Real: TinyFace mAP, ×100 (DFA). Best LR-trained per row in **bold**; the resolution differs.

Evaluation	HR-tr.	56↓c/↑a	28↓c/↑a	14↓a/↑c
Synthetic macro (Acc)	78.85	80.25	**80.74**	79.67
Real TinyFace (mAP)	52.55	**54.68**	49.31	—

Table 5. Real native-LR TinyFace under `aligned_pad` crops (mAP / R-1, ×100). Best per row in **bold**.

Method	HR	56↓c/↑a	28↓a/↑c	28↓c/↑a	14↓a/↑c
LR backbone (mAP)	47.24	**49.31**	49.23	48.51	45.67
KD (mAP)	47.24	48.96	47.46	47.94	45.67
PDT (mAP)	47.24	46.69	46.19	45.22	43.66
LR backbone (R-1)	54.48	56.14	**56.41**	55.69	52.28
PDT (R-1)	54.48	53.78	53.49	52.31	51.10

5.2. Distillation gives the largest synthetic gains

Among the three adaptations, knowledge distillation yields the largest gains on HR→synthetic LR (Table 1). The student sees only the degraded image, but its target embedding is computed by the teacher from the pristine one, so the HR-shaped geometry supplies supervision that the LR input alone cannot carry. The effect is largest at the most extreme degradations (14 and 7 px), where the student has least signal of its own. These gains are on *synthetic* low resolution, however, and do not carry over to real native LR.

5.3. Real native-LR

On real LR data the ranking changes. Table 4 compares the two evaluations of the *same* compact models: synthetic HR→LR macro-mean and real native-LR TinyFace mAP (DFA alignment). The setting that is best on synthetic LR, 28↓c/↑a, is the worst of the family on real LR, below even the HR-trained model. On the other hand, the milder 56↓c/↑a setting performs the best. Training at 56 px leaves the embeddings closer to the real native-LR domain than training at a resolution tuned on synthetic data.

The same lesson applies to input-side adaptation under *both* alignments. Table 5 reports TinyFace under the `aligned_pad` crops: every PDT variant sits *below* the HR baseline at every resolution, because a translator fit to synthetic LR does not transfer to genuine degradation and mostly smooths the probe (Fig. 4). Likewise, KD trails the plain LR backbone. Both observations motivate us to test a more realistic degradation model and a learned way to invert it, evaluated on real LR.

6. Generative Pipeline: Results

We show our results on medium to high synthesis efforts tested on real-world LR images.

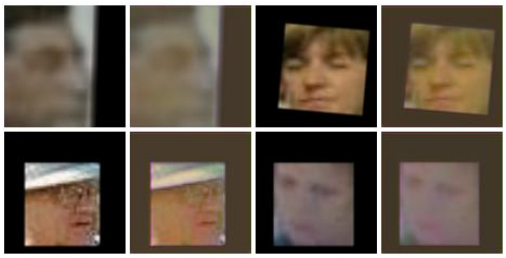


Figure 4. PDT translator outputs on native-LR TinyFace probes. Trained on synthetic LR, the translation appears to have a smoothing effect on the input images.

Table 6. Real native-LR **TinyFace** identification (DFA alignment). "base"=frozen EdgeFace-base (WebFace12M); "S"=EdgeFace-S (3.65 M); *direct* feeds the aligned LR face with no front-end. Generative methods in *italic*. **Bold**=best compact (S) model; underline=best short of the large direct-feed reference. All values ×100.

Model	mAP	R-1	R-5	R-10	R-20
base, direct (HR)	65.12	71.08	74.97	75.83	76.98
base + RRDB-SR + PDT	57.53	64.70	70.52	72.26	74.20
base + ESPCN-SR + PDT	56.64	63.95	69.74	71.62	73.44
S, HR-trained (direct)	52.55	59.66	66.63	69.05	71.49
S, 56↓c/↑a aug	**54.68**	**61.40**	68.16	70.28	72.80
S, Real-ESRGAN 32 stem	53.74	**61.40**	68.62	70.84	73.47
S, 56↓c/↑a PDT	51.94	58.88	66.17	68.51	70.79
S, 28↓c/↑a aug	49.31	57.16	64.14	66.71	69.21

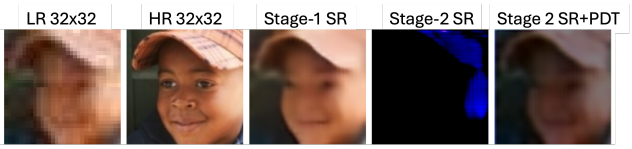
6.1. Native-LR identification

Table 6 is our result on real native-LR TinyFace (DFA alignment), with every model compared against the direct feed of its own backbone. Generative methods are in *italic*.

On the compact backbone, which has no strong direct feed to fall back on, simple LR-aware augmentation is the only synthesis that helps. A mild 56↓c/↑a training gives the best compact model, ahead of both its own direct feed and the Real-ESRGAN 32 px stem. This is the regime our study targets and the one place where synthesis pays off.

In contrast, the learned generative front-end does not help as much. On the frozen EdgeFace-base the full SR+PDT pipeline trails a direct feed of the aligned LR image, and the SR stage on its own is worse still (Table 10). Learned upscaling therefore loses to the interpolation already performed by the aligner resulting in lowered front-end accuracy. The PDT alone shows the same pattern on the compact backbone, again below its direct feed. Fig. 5 shows the upsampler does recover plausible and identity-consistent detail, but this visual quality does not translate into recognition gains.

(a) ESPCN sub-pixel



(b) RRDB (Real-ESRGAN)

Figure 5. Learned $32 \rightarrow 112$ reconstructions on WebFace4M pairs after identity-aware pretraining (input / SR output / HR target).

6.2. Comparison with the literature on TinyFace

Table 7 compares our compact models with the efficient face recognizers of the EFaR 2023 competition [13], which reports TinyFace identification for sub-5 M-parameter models under a common protocol. All our EdgeFace-S models are trained *from scratch for* 100 *epochs*, the budget used for our LR backbones, whereas the competition's EdgeFace-S(γ=0.5) is the fully-trained official checkpoint of the same architecture. Even on this shorter schedule our HR-trained EdgeFace-S lands close to the official checkpoint, which indicates that our protocol and alignment are comparable, unlike cross-paper numbers collected under different alignment, which can differ by tens of points.

At an equal parameter budget and the same schedule, LR-aware augmentation outperforms the official checkpoint of EdgeFace-S on both Rank-1 and Rank-5. Other methods, such as Modified-MobileFaceNet, are also more accurate at a cost of $1.5\times$ the FLOPs. The larger gains come not from the backbone but from training and adaptation: the learned SR front-end on a stronger frozen backbone reaches the level of the best compact competitor (Table 6), while a direct feed of that same backbone is far ahead of everything in Table 7. At comparable complexity, native-LR accuracy is driven by LR-aware synthesis and adaptation rather than by architecture or a longer training schedule.

6.3. LR training as a regularizer (HR benchmarks)

A useful side effect (Table 9): training on LR-augmented data at 56 and 28 px *matches or exceeds* the HR baseline on every standard benchmark and on IJB-C (Fig. 6). The more aggressive 14↓a/↑c setting is the exception, falling below the baseline on five of the six verification sets while still improving model performance on IJB-C. The 28↓c/↑a backbone is best or tied on four of the six verification sets, and 28↓a/↑c gives the best IJB-C TAR@FAR. The native 32 px stem, trained only on very-low-resolution input size performs worse on HR: specializing for native LR trades away HR FR capability, showing to be suitable only when the input is expected to present LR artifacts. We also found that Distillation largely maintains and outperforms different tests against the HR baseline on most sets, although the 28 px KD variant performs worse overall and mostly on CFP-FP and AgeDB-30.

Table 7. TinyFace identification (%) against efficient models from the EFaR 2023 competition [13] (their Table 4). All models are compact ($\leq$4.1 M parameters); the training-schedule difference is discussed in the text. **Bold**: best at the 3.65 M budget.

Method	Params	MFLOPs	R-1	R-5
EFaR 2023 competition [13]				
SAM-MFaceNet	1.10M	236.8	61.31	66.33
EdgeFace-XS(γ=0.6)	1.77M	154.0	58.77	63.89
Mod. MobileFaceNet	2.10M	456.9	64.29	69.60
EdgeFace-XS-Q	2.24M	196.9	59.20	64.75
EdgeFace-S(γ=0.5)	3.65M	306.1	61.02	65.47
GhostFaceNetV1-1	4.09M	215.7	59.54	64.75
Ours (same EdgeFace-S architecture)				
EdgeFace-S, HR-trained	3.65M	306.1	59.66	66.63
EdgeFace-S, 56↓c/↑a aug	3.65M	306.1	**61.40**	**68.16**

Table 8. Standard HR verification (Acc $\times$100) for the KD models. Distillation mostly preserves the LR-training regularization; **bold** marks the best per row.

	HR	56↓c/↑a	28↓a/↑c	28↓c/↑a	14↓a/↑c
LFW	99.45	**99.62**	99.48	99.45	99.58
CALFW	94.53	**95.12**	94.50	94.33	94.77
CPLFW	90.90	**91.37**	90.70	90.85	91.18
CFP-FF	99.29	**99.47**	99.13	99.03	99.39
CFP-FP	96.53	97.06	95.63	95.81	**97.11**
AgeDB-30	94.38	**95.72**	93.88	93.10	94.98

Table 9. Standard HR verification (Acc $\times$100) and IJB-C TAR@FAR (%) for the LR backbones; the last column is the native 32 px stem (*n*32), which collapses on HR.

	HR	56↓c/↑a	28↓a/↑c	28↓c/↑a	14↓a/↑c	*n*32
LFW	99.45	99.58	99.53	**99.65**	99.58	93.48
CALFW	94.53	95.25	95.28	**95.32**	94.52	—
CPLFW	90.90	91.63	**91.85**	**91.85**	89.97	—
CFP-FF	99.29	**99.57**	**99.57**	99.46	99.03	—
CFP-FP	96.53	**97.24**	97.10	97.09	96.11	81.01
AgeDB-30	94.38	95.63	95.63	**95.67**	93.90	75.21
IJB-C 10^{-3}	96.11	96.68	**96.91**	96.71	96.51	—
IJB-C 10^{-4}	93.07	94.44	**94.56**	94.06	93.43	—

6.4. Impact of the learned SR front-end

To isolate the contribution of the learned super-resolution front-end, we evaluate the identity-aware SR (ESPCN) pretraining checkpoint in isolation before the joint contrastive stage with the PDT (Table 10). On both backbones the SR pretraining alone underperforms a direct feed of the aligned LR image. Adding the joint contrastive

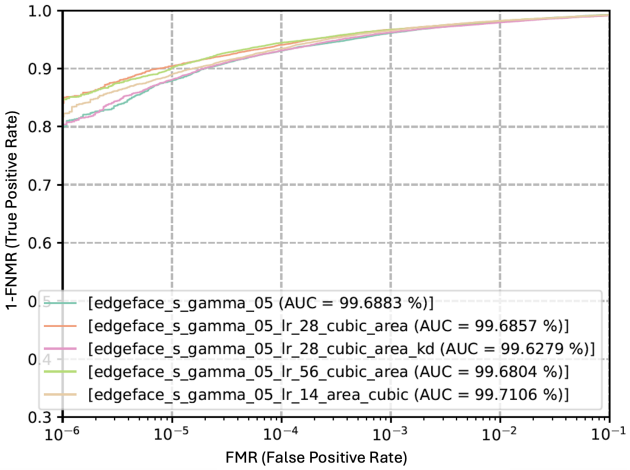


Figure 6. IJB-C False Match Rate (FMR) against 1−False Non-Match Rate (FNMR). The LR-trained EdgeFace-S variants match or exceed the HR baseline across operating points.

Table 10. Effect of the learned SR front-end on native-LR TinyFace (DFA, %). "SR only" is the identity-aware SR (ESPCN) pretraining checkpoint without the joint PDT stage; "+PDT" adds the joint contrastive stage.

Model	mAP	R-1	R-5
EdgeFace-S, direct feed	52.55	59.66	66.63
EdgeFace-S + SR only	35.82	44.02	53.19
EdgeFace-base, direct feed	65.12	71.08	74.97
EdgeFace-base + SR only	43.48	52.23	60.76
EdgeFace-base + SR + PDT	56.64	63.95	69.74

stage recovers much of this loss on EdgeFace-base but still does not reach the direct feed. This reinforces our central finding: a learned SR module is not useful as a standalone front-end, and even with joint training, the added effort does not beat feeding a strong backbone the aligned image.

7. Ablations and Analysis

Interpolation pair. Cubic downsampling preserves more high-frequency structure, which is then smoothed by area upsampling. The cubic-down/area-up pair (28 ↓c/↑a) is the strongest synthetic configuration, while area-down/cubic-up (28 ↓a/↑c) trails it on most HR→LR sets yet leads on IJB-C. When training, only 4 of the 12 down/up combinations converged stably. This shows that the interpolation choice directly affects optimization.

Resolution variation. Accuracy degrades smoothly with training resolution on synthetic LR but the *best* training resolution depends on the target domain: 28 px on synthetic, 56 px on native LR, with 14 px below the HR baseline on TinyFace. At 7 px all methods degrade sharply on CFP-FP and AgeDB-30, an unresolved domain gap.

Does synthesis effort pay off? Not monotonically (Table 11). On the compact backbone the *lowest*-effort option, a single 56 px interpolation, beats both the no-synthesis baseline and the medium-effort degradation stem while 28 px—the setting that was optimal on synthetic data—is the worst among all. The *highest*-effort option is the best *adaptation* in the table, but only on the large frozen backbone, and it is still beaten by feeding that backbone directly. Effort therefore pays off only where it changes the probe embedding most: cheap augmentation for a retrainable compact model, and a learned front-end only when the backbone is fixed and strong.

Table 11. Adaptation strategies by synthesis effort and accuracy on TinyFace (DFA aligned). We show different EdgeFace variants (model), the synthesis degradation type. "Real degr." indicates whether the synthesis reproduces genuine sensor/optics degradation rather than naive interpolation.

Strategy	Effort	Model	Real degr.	mAP (%)
Interp. aug (56 ↓c/↑a)	**low**	**S**	**no**	**54.68**
Interp. aug (28 ↓c/↑a)	low	S	no	49.31
Real-ESRGAN 32 stem	medium	S	yes	53.74
SR+PDT (ESPCN)	high	base†	yes	56.64
SR+PDT (RRDB)	high	base†	yes	57.53
Direct feed (no synthesis)	none	S	–	52.55

† frozen EdgeFace-base; others retrain EdgeFace-S.

Where capacity matters. A strong frozen backbone already absorbs much degradation: direct feed of EdgeFace-base beats every adaptation, including our learned SR front-end on the same backbone, so a trainable front-end is worth most when the backbone is small or cannot be used directly. The two backbone groups in Tables 6 and 11 differ in capacity, training sets (WebFace4M vs. WebFace12M) and whether they are frozen, therefore the accuracy is only comparable within each backbone model variant (S and base). It also explains why KD wins synthetic extremes but not native LR: anchoring to an HR-oriented embedding helps when the test degradation is synthetic and mild, and hurts when the target domain genuinely differs.

Where does the gap come from? Two factors could explain the poor transfer to TinyFace: a *degradation* mismatch (interpolation does not reproduce the blur, noise, and compression of real optics) and an *alignment* mismatch (landmarks are far less reliable on genuinely captured LR faces). Our results implicate both but separate them. The synthetic optimum (28 px) being the worst configuration on real LR (Table 4) points at the degradation model, since all compared models share the same alignment pipeline. Alignment instead confounds absolute accuracy. Moving from `aligned_pad` crops to the DFA aligner raises every model substantially, so any TinyFace comparison must fix the aligner. We therefore attribute the gap primarily to the degradation model, while noting that detection and align-

Table 12. Per-group FMR, its ratio r to the geometric mean, and the Gini index of per-group FMR on RFW at TMR@FMR$= 10^{-2}$.

		African		Asian		Caucasian		Indian		
px	Model	FMR	r	FMR	r	FMR	r	FMR	r	Gini
HR	HR-tr.	0.74	**2.80**	0.27	1.01	0.07	0.25	0.37	1.39	0.37
	56↓c/↑a	1.10	**3.36**	0.53	1.62	0.03	0.10	0.60	1.82	0.36
	28↓c/↑a	1.37	**4.24**	0.60	1.85	0.00	0.05	0.80	2.47	0.39
56	HR-tr.	0.77	**3.41**	0.43	1.92	0.03	0.15	0.23	1.03	0.41
	56↓c/↑a	0.90	**3.25**	0.57	2.04	0.00	0.06	0.70	2.51	0.33
	28↓c/↑a	1.31	**3.85**	0.43	1.28	0.03	0.10	0.70	2.06	0.41
28	HR-tr.	0.33	**3.36**	0.13	1.34	0.00	0.17	0.13	1.33	0.42
	56↓c/↑a	0.60	**3.92**	0.33	2.17	0.00	0.11	0.17	1.08	0.45
	28↓c/↑a	0.64	**2.59**	0.43	1.76	0.03	0.14	0.40	1.62	0.31

ment on native LR remain a major source of absolute error.

7.1. Bias and fairness evaluation

We measure the demographic fairness on RFW [28], which contains African, Asian, Caucasian and Indian subjects. We evaluate the three EdgeFace-S backbones at their respectively calibrated thresholds on IJB-C at TMR@FMR$= 10^{-2}$, using the same IJB-C protocol as in Table 9 on RFW. The probes in RFW are further degraded with bicubic downsampling and upsampling to $56/28/14/7$ px depending on the model. Table 12 reports, for each group d, its FMR as a percentage and the ratio $r_d = \mathrm{FMR}_d/g$, where g is the geometric mean of the FMR over the four groups, so $r_d = 1$ when the group's FMR equals the geometric mean. The last column is the Gini coefficient of the FMR over the four groups, which is 0 when the four groups have equal FMR and grows with the difference between them. In the table the px column is the test resolution and the model name is the training regime. A group with no false matches has its FMR set to $0.5/3000$, half the smallest non-zero rate.

The Caucasian group has better verification accuracy than the other three groups: its r stays far below 1 in every cell, while the African r is the highest throughout, with the Asian and Indian groups in between. This ordering, with the Caucasian group most accurate and the African group least accurate, agrees with bias previously reported on RFW [28].

Training on low-resolution data does not systematically reduce this difference. The $56\downarrow$c/↑a and $28\downarrow$c/↑a backbones, which give the highest average accuracy in Tables 1 and 6, show an FMR Gini that moves in both directions relative to the HR-trained model depending on the test resolution, with no consistent improvement in either backbone. The largest group ratio anywhere in the table also belongs to the $28\downarrow$c/↑a model. Accuracy gains from LR-aware synthesis therefore do not carry over into reduced demographic disparity.

We report resolutions of 28 px and above. At 14 and 7 px the global FNMR is above 0.9 for all models, so the fixed threshold no longer separates the mated and non-mated scores and the per-group comparison is not informative.

8. Limitations and Future Work

Our learned SR front-end is evaluated only on a large frozen backbone, where a direct feed is already a hard baseline. We leave the more informative setup, a *compact* backbone that cannot exploit such a feed, as future work. All our LR training data are synthetic by design as we study the setting in which *no* labelled native-LR identities are available. We therefore do not evaluate the common practice of pre-training on synthetic LR and fine-tuning on limited real LR. We leave for future work using native LR images, such as TinyFace's training split, for training synthesis strategies and assessing LR FR performance. Quantifying how much real LR data each strategy needs before it becomes worthwhile is the clearest next step, alongside learning the *paired* HR-reference/LR-probe conditional directly (e.g. a conditional or diffusion generator, evaluated on real probes against the direct-feed baseline) and analyzing the fairness of LR capture and sampling.

9. Conclusion

We studied five adaptation strategies for low-resolution face recognition under scarce native-LR data and evaluated all of them on real LR. Our four conclusions are as follows: *(i)* The degradation setting that is optimal on synthetic benchmarks is not the one that is optimal on real LR: $28\downarrow$c/↑a is best across the synthetic benchmarks yet is the worst configuration on TinyFace, where the milder 56 px setting wins. *(ii)* Synthesis effort does not pay off monotonically; on a retrainable compact backbone the cheapest interpolation augmentation beats both Real-ESRGAN-style degradation and a learned SR front-end. *(iii)* A learned, identity-aware super-resolution front-end never beats simply feeding a strong frozen backbone the aligned LR image, so a direct-feed baseline should be reported before any restoration or translation pipeline is claimed to help. *(iv)* The average-accuracy gains from LR-aware synthesis do not reduce demographic bias: the FMR disparity on RFW shows no consistent improvement. In practice we therefore recommend matching synthesis effort to the deployment constraint, and validating on real native LR rather than on synthetic degradation alone.

Acknowledgments

We acknowledge the funding provided by the following projects: Frontex under the Frontex Research Grants Programme, Call 2024/CFP/INNOVATE/01, Grant Agreement No. 2025/280; CarMen project, HORIZON-CL3-2023-BM-01, no. 101168325; PopEye project, HORIZON-CL3-2023-BM-01, no. 101168317; and the CERTAIN Project, HORIZON-CL4-2024-DATA-01, no. 101189650.

References

[1] X. An, X. Zhu, Y. Gao, Y. Xiao, Y. Zhao, Z. Feng, L. Wu, B. Qin, M. Zhang, D. Zhang, and Y. Fu. Partial FC: Training 10 million identities on a single machine. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 1445–1449, 2021.

[2] M. Bengherabi, D. Laib, F. S. Lasnami, and R. Boussaha. Boosting cross-quality face verification using blind face restoration. In *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2023. arXiv:2308.07967.

[3] Z. Cheng, X. Zhu, and S. Gong. Low-resolution face recognition. In *Asian Conference on Computer Vision (ACCV)*, 2018.

[4] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4690–4699, 2019.

[5] S. Ge, S. Zhao, C. Li, Y. Zhang, and J. Li. Efficient low-resolution face recognition via bridge distillation. *IEEE Transactions on Image Processing*, 29:6898–6908, 2020.

[6] A. George, C. Ecabert, H. Otroshi Shahreza, K. Kotwal, and S. Marcel. EdgeFace: Efficient face recognition model for edge devices. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2024.

[7] A. George and S. Marcel. Heterogeneous face recognition using domain invariant units. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024. arXiv:2404.14343.

[8] A. George, A. Mohammadi, and S. Marcel. Prepended domain transformer: Heterogeneous face recognition without bells and whistles. *IEEE Transactions on Information Forensics and Security*, 18:133–146, 2023.

[9] R. Hadsell, S. Chopra, and Y. LeCun. Dimensionality reduction by learning an invariant mapping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1735–1742, 2006.

[10] G. Hinton, O. Vinyals, and J. Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

[11] 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, 2007.

[12] M. Kim. CVLface: A modular library for face recognition research (differentiable face aligner). https://github.com/mk-minchul/CVLface, 2024.

[13] J. N. Kolf, F. Boutros, J. Elliesen, M. Theuerkauf, N. Damer, et al. EFaR 2023: Efficient face recognition competition. In *IEEE International Joint Conference on Biometrics (IJCB)*, 2023. arXiv:2308.04168.

[14] S. Liu, Z.-P. Duan, J. OuYang, J. Fu, H. Park, Z. Liu, C.-L. Guo, and C. Li. FaceMe: Robust blind face restoration with personal identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025. arXiv:2501.05177.

[15] I. Loshchilov and F. Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*, 2019.

[16] L. S. Luevano, L. Chang, H. Méndez-Vázquez, Y. Martínez-Díaz, and M. González-Mendoza. A study on the performance of unconstrained very low resolution face recognition: Analyzing current trends and new research directions. *IEEE Access*, 9:75470–75493, 2021.

[17] L. S. Luevano, Y. Martínez-Díaz, H. Méndez-Vázquez, M. González-Mendoza, and D. Frey. Swiftfaceformer: An efficient and lightweight hybrid architecture for accurate face recognition applications. In A. Antonacopoulos, S. Chaudhuri, R. Chellappa, C.-L. Liu, S. Bhattacharya, and U. Pal, editors, *Pattern Recognition*, pages 244–258, Cham, 2025. Springer Nature Switzerland.

[18] M. Maaz, A. Shaker, H. Cholakkal, S. Khan, S. W. Zamir, R. M. Anwer, and F. S. Khan. EdgeNeXt: Efficiently amalgamated CNN-transformer architecture for mobile vision applications. In *Computer Vision – ECCV 2022 Workshops*, 2022.

[19] Y. Martínez-Díaz, L. S. Luevano, and H. Méndez-Vázquez. Effectiveness of blind face restoration to boost face recognition performance at low-resolution images. In *International Workshop on Artificial Intelligence and Pattern Recognition (IWAIPR)*, pages 455–467, 2023.

[20] Y. Martínez-Díaz, L. S. Luevano, H. Méndez-Vázquez, M. Nicolás-Díaz, L. Chang, and M. González-Mendoza. Shufflefacenet: A lightweight face architecture for efficient and highly-accurate face recognition. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 2721–2728, 2019.

[21] Y. Martínez-Díaz, M. Nicolás-Díaz, H. Méndez-Vázquez, L. S. Luevano, L. Chang, M. Gonzalez-Mendoza, and L. E. Sucar. Benchmarking lightweight face architectures on specific face recognition scenarios. *Artificial Intelligence Review*, 54(8):6201–6244, Dec. 2021.

[22] F. V. Massoli, G. Amato, and F. Falchi. Cross-resolution learning for face recognition. *Image and Vision Computing*, 99:103927, 2020.

[23] B. Maze, J. Adams, J. A. Duncan, N. Kalka, T. Miller, C. Otto, A. K. Jain, W. T. Niggel, J. Anderson, J. Cheney, and P. Grother. IARPA Janus Benchmark-C: Face dataset and protocol. In *International Conference on Biometrics (ICB)*, pages 158–165, 2018.

[24] 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 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 51–59, 2017.

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

[26] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1874–1883, 2016.

[27] 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 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5265–5274, 2018.

[28] M. Wang, W. Deng, J. Hu, X. Tao, and Y. Huang. Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pages 692–702. IEEE, 2019.

[29] X. Wang, L. Xie, C. Dong, and Y. Shan. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 1905–1914, 2021.

[30] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy. ESRGAN: Enhanced super-resolution generative adversarial networks. In *Computer Vision – ECCV 2018 Workshops*, 2018.

[31] Z. Zhu, G. Huang, J. Deng, Y. Ye, J. Huang, X. Chen, J. Zhu, T. Yang, J. Lu, D. Du, and J. Zhou. WebFace260M: A benchmark unveiling the power of million-scale deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10492–10502, 2021.