IConFace: Identity-Structure Asymmetric Conditioning for Unified Reference-Aware Face Restoration

Axi Niu*, **Jinyang Zhang***, **Senyan Qing**
Northwestern Polytechnical University
nax@nwpu.edu.cn zhangjinyang@mail.nwpu.edu.cn qingsenyan@nwpu.edu.cn
Homepage: https://cosmicrealm.github.io/IConFace/

Abstract

Blind face restoration is highly ill-posed under severe degradation, where identity-critical details may be missing from the degraded input. Same-identity references reduce this ambiguity, but mismatched pose, expression, illumination, age, makeup, or local facial states can lead to overuse of reference appearance. We propose **IConFace**, a unified reference-aware and no-reference framework with identity–structure asymmetric conditioning. References are distilled into a norm-weighted global AdaFace identity anchor for image-only modulation, while the degraded image is reinforced as the spatial structure anchor through low-rank residuals and block-wise degraded cross-attention with two-route memory. The resulting single checkpoint exploits references when available and falls back to no-reference restoration when absent, improving identity consistency, fine-detail recovery, and degraded-only restoration quality in a unified model.

Introduction

Blind face restoration (BFR) recovers high-quality faces from low-quality observations with unknown degradations. Modern generative, codebook, transformer, and diffusion priors improve perceptual realism (Menon et al. 2020; Yang et al. 2021; Wang et al. 2021; Gu et al. 2022; Zhou et al. 2022; Wang et al. 2022, 2023; Zhao et al. 2023; Qiu et al. 2023; Yang et al. 2023; Lin et al. 2024; Miao et al. 2025; Wang et al. 2025), but no-reference restoration remains highly ill-posed: severe degradation can remove identity-critical evidence and local facial cues, so a plausible restoration may still drift from the target identity.

Same-identity reference images reduce this ambiguity by providing identity evidence unavailable in the degraded input (Li et al. 2018, 2020b, 2022; Hsiao et al. 2024; Liu et al. 2025; Chong et al. 2025; Yin et al. 2026). This makes reference-aware restoration useful for personal photo enhancement and legacy portrait restoration. However, references are not pure identity carriers: they also encode pose, illumination, expression, age, makeup, and local facial states. Direct transfer or over-attention to reference features may therefore improve identity similarity while overusing reference-specific appearance and failing to preserve the structure implied by the degraded input. Fig. 1 focuses on the complementary detail-recovery challenge: competing methods may produce plausible faces while suppressing or distorting identity-related local details.

Figure 1: **Teaser comparison.** IConFace preserves reference-consistent facial details better than strong blind and reference-aware baselines while remaining anchored to the degraded input.

Practical systems must also handle missing references. Same-identity references may be abundant, limited, noisy, or entirely absent. Existing blind methods lack explicit identity evidence, while many reference-aware methods rely on matching, alignment, personalization, or memory transfer that is less natural in the no-reference case. A useful model should exploit references when available and fall back to robust degraded-only restoration when they are absent.

We propose **IConFace**, a unified framework based on asymmetric identity–structure conditioning. References are compressed into a global identity controller, while the degraded image remains the spatial structure anchor. A hybrid concat backbone keeps degraded and reference tokens as visual evidence, and two lightweight side pathways ex-

*These authors contributed equally.

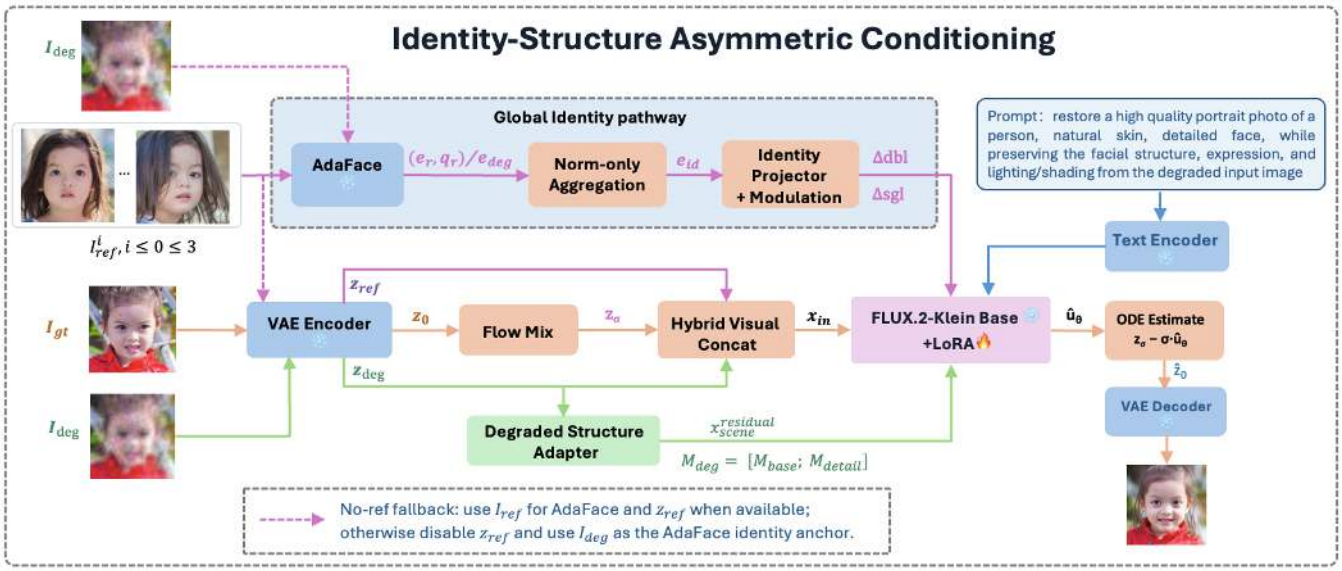

Figure 2: Overview of **IConFace**. The main route keeps the hybrid concat sequence $x_{\text{in}} = [x_{\text{scene}}; x_{\text{deg}}; x_{\text{ref}}]$ in the restoration backbone. A global identity pathway compresses references into a single AdaFace anchor and injects image-only modulation. A degraded structure pathway reinforces the degraded observation with a low-rank input residual and block-wise degraded cross-attention using two-route pooled memory for base structure and local detail. When references are absent, $x_{\text{ref}} = \emptyset$ and the identity pathway falls back to a degraded-image AdaFace anchor as weak self-conditioning.

plicitly control identity and structure. This design improves reference-aligned identity consistency and fine-detail recovery under severe degradation while preserving no-reference restoration capability.

Contributions. Our contributions are:

- A unified optional-reference formulation where one checkpoint supports reference-aware and no-reference restoration through asymmetric identity–structure conditioning.
- Lightweight pathways that use norm-weighted AdaFace modulation for reference identity and low-rank residuals with two-route degraded memory for structure and detail anchoring.
- Extensive results showing improved reference-aligned identity consistency, fine-detail recovery under severe degradation, and strong no-reference perceptual quality.

Related Work

Blind face restoration. Blind and real-world restoration have evolved from geometric priors to degradation modeling, generative priors, dictionaries, codebooks, transformers, and diffusion models (Li et al. 2020a; Yang et al. 2020, 2021; Wang et al. 2021; Gu et al. 2022; Zhou et al. 2022; Wang et al. 2022, 2023; Zhao et al. 2023; Qiu et al. 2023; Yang et al. 2023; Lin et al. 2024; Tsai et al. 2024; Miao et al. 2025; Wang et al. 2025; Niu et al. 2025, 2023, 2024; Li et al. 2025). These methods improve realism, but degraded-only evidence remains ambiguous under strong corruption.

Reference-aware restoration and conditioning. Reference-guided restoration uses exemplars to recover identity cues through warping, adaptive fusion, memory dictionaries, latent diffusion conditioning, or personalized adapters (Li et al. 2018, 2020b, 2022; Varanka et al. 2024; Hsiao et al. 2024; Ying et al. 2024; Zhang et al. 2024; Liu et al. 2025; Chong et al. 2025; Yin et al. 2026). These works show the value of references, but references also carry pose, expression, illumination, and local states that may not match the target. IConFace is also related to broader conditioning and control mechanisms (Vaswani et al. 2017; Rombach et al. 2022; Peebles and Xie 2023; Meng et al. 2022; Zhang, Rao, and Agrawala 2023; Ye et al. 2023; Yan et al. 2019, 2020): auxiliary evidence is most useful when its role is explicit. We specialize this principle by distilling references into a global identity anchor while reinforcing the degraded image as the spatial structure anchor.

Method

Problem Setup

Given a degraded face image I_{deg} and an optional set of same-identity references $\mathcal{R} = \{I^1_{\text{ref}}, \ldots, I^N_{\text{ref}}\}$, our goal is to generate a restored image $\hat{I}$ that preserves the structure and facial state implied by I_{deg}. When references are available, the model should additionally recover reference-consistent identity cues and fine details. When $\mathcal{R} = \emptyset$, it should operate as a no-reference blind restorer.

We train in latent space with a flow-matching objective. Given clean latent z_0, noise $\epsilon \sim \mathcal{N}(0, I)$, and noise level $\sigma \in [0, 1]$, we form

$$z_\sigma = (1-\sigma)z_0 + \sigma\epsilon, \qquad u^\star = \epsilon - z_0. \tag{1}$$

The restoration network predicts $\hat{u}_\theta = f_\theta(z_\sigma, I_{\text{deg}}, \mathcal{R}, \sigma)$ and recovers $\hat{z}_0 = z_\sigma - \sigma\hat{u}_\theta$.

Overview of IConFace

IConFace is built on the FLUX.2-klein-base-4B hybrid concat restoration backbone. The main sequence concatenates noisy scene tokens, degraded-image tokens, and optional reference tokens:

$$x_{\text{in}} = [x_{\text{scene}}; x_{\text{deg}}; x_{\text{ref}}]. \tag{2}$$

In no-reference mode, $x_{\text{ref}} = \emptyset$ and the same degraded-conditioned backbone is retained. IConFace augments this backbone with two side pathways: a global identity pathway that aggregates references into a single AdaFace identity anchor, and a degraded structure pathway that reinforces the spatially aligned degraded observation.

Asymmetric Identity–Structure Conditioning

The central design principle of IConFace is that degraded and reference observations should not be treated symmetrically. The degraded image is the only observation spatially aligned with the target image, so it should remain the structure anchor. References are valuable identity evidence, but they also contain pose, expression, illumination, makeup, age, and local facial states that should not be copied indiscriminately. IConFace therefore assigns references a global identity-control role and assigns degraded tokens a structure-preserving role.

This asymmetry is implemented through three choices. First, the hybrid concat backbone remains the main carrier of degraded and reference evidence. Second, references are aggregated into a single global AdaFace identity anchor for image-only modulation when they are available. Third, degraded tokens are reinforced through an explicit low-rank residual and block-wise degraded memory. In no-reference operation, the reference segment is removed rather than replaced by duplicated placeholders, and the identity pathway uses the degraded-image AdaFace fallback only as weak forward conditioning.

Global Identity Pathway

For each valid reference I^r_{ref}, frozen AdaFace returns a raw embedding $z_r \in \mathbb{R}^{512}$. We separate identity direction and reliability:

$$e_r = z_r/\|z_r\|_2, \qquad q_r = \|z_r\|_2. \tag{3}$$

The norm q_r acts as a quality proxy. With temperature T, multiple references are aggregated by norm-only weights:

$$w_r = \text{softmax}(\log q_r/T), \qquad e_{\text{ref}} = \text{Norm}\left(\sum_r w_r e_r\right). \tag{4}$$

If references are absent, the pathway falls back to the normalized AdaFace embedding of I_{deg}. This fallback is used only as weak forward conditioning, not as a reference supervision target.

The effective identity anchor e_{id} is projected into the backbone hidden space and transformed into modulation deltas for double-stream and single-stream image blocks:

$$h_{\text{id}} = \phi_{\text{global}}(e_{\text{id}}), \qquad \Delta_{\text{dbl}}, \Delta_{\text{sgl}} = \psi_{\text{dbl}}(h_{\text{id}}), \psi_{\text{sgl}}(h_{\text{id}}). \tag{5}$$

The deltas are applied only to image tokens; the text stream is untouched. This gives references a global identity-control role without turning them into a local structure-transfer branch.

Degraded Structure Pathway

The degraded image is the only observation spatially aligned with the target, so IConFace reinforces it as the structure anchor. First, degraded tokens resized to the scene-token resolution are injected through a low-rank input residual:

$$\begin{aligned} x^{\text{residual}}_{\text{scene}} &= s_{\text{deg}} W_{\text{in}}(\tilde{x}_{\text{deg}}), \\ x'_{\text{scene}} &= x_{\text{scene}} + x^{\text{residual}}_{\text{scene}}. \end{aligned} \tag{6}$$

Second, full degraded attention is compressed into a fixed memory budget by two learned-query poolers. A base route pools x_{deg} to preserve global layout and illumination, while a detail route pools $x_{\text{deg}} - \text{smooth}(x_{\text{deg}})$ to emphasize local facial cues:

$$\begin{aligned} M_{\text{base}} &= P_b(x_{\text{deg}}), \\ M_{\text{detail}} &= P_d(x_{\text{deg}} - \text{smooth}(x_{\text{deg}})), \\ M_{\text{deg}} &= [M_{\text{base}}; M_{\text{detail}}], \\ h'_\ell &= h_\ell + \gamma_\ell \text{Attn}(Q_\ell h_\ell, K_\ell M_{\text{deg}}, V_\ell M_{\text{deg}}). \end{aligned} \tag{7}$$

Here P_b and P_d are learned-query poolers, h_ℓ denotes image tokens in block ℓ, and the K/V projectors are low-rank. The adapter therefore outputs both the scene residual and the two-route degraded memory. This additive image-only path lets each block read both coarse degraded structure and high-frequency degraded detail without replacing the main concat route.

Training Objective

The main restoration loss regresses the flow field:

$$\mathcal{L}_{\text{fm}} = \mathbb{E}_{z_0,\epsilon,\sigma}\left[w(\sigma)\|\hat{u}_\theta - (\epsilon - z_0)\|_2^2\right]. \tag{8}$$

In the final model, the flow-matching timestep weight is uniform, i.e., $w(\sigma) = 1$; only the identity loss below uses sigma-dependent weighting. For reference-aware training samples, we decode the current clean-latent estimate $\hat{z}_0$ and compute a frozen-AdaFace identity loss against the reference anchor:

$$\mathcal{L}_{\text{ref-id}} = 1 - \cos(A(\hat{I}), \text{sg}(e_{\text{ref}})), \tag{9}$$

where $A(\cdot)$ is the frozen AdaFace encoder and sg stops gradients through the target. We also use a weak clean-target stabilizer

$$\mathcal{L}_{\text{hard}} = 1 - \cos(A(\hat{I}), \text{sg}(e_{\text{gt}})) \tag{10}$$

for imperfect references. The sigma-weighted identity objective is

$$\mathcal{L}_{\text{id}} = \omega(\sigma)\left[(1 - \lambda^\star_h)\mathcal{L}_{\text{ref-id}} + \lambda^\star_h \mathcal{L}_{\text{hard}}\right], \tag{11}$$

where $\lambda^\star_h = \lambda_h(1 - \cos(e_{\text{ref}}, e_{\text{gt}}))$ increases the stabilizer weight when the selected reference is less consistent with the clean target. We use $\omega(\sigma) = \max(1 - \sigma, \omega_{\min})^2$ with $\omega_{\min} = 0.25$, so high-noise steps are down-weighted rather than hard-skipped. For no-reference samples, $\mathcal{L}_{\text{id}} = 0$ because no reference identity target exists. The final loss is

$$\mathcal{L} = \mathcal{L}_{\text{fm}} + \lambda_{\text{id}}\mathcal{L}_{\text{id}}. \tag{12}$$

Additional implementation details are provided in the supplementary material.

Table 1: Reference-aware restoration on CelebA-Test-Ref, FFHQ-Ref Moderate, FFHQ-Ref Severe, and CelebHQRef100. Ref metrics are computed against the first protocol reference.

Method	Dataset	Ref-ArcFace ↑	Ref-AdaFace ↑	MUSIQ ↑	CLIP-IQA ↑	MANIQA ↑
DMDNet	CelebA-Test-Ref	0.472	0.471	72.725	0.625	0.495
ReF-LDM		**0.573**	**0.569**	**74.426**	**0.675**	**0.527**
RestorerID		0.539	0.530	**72.820**	**0.695**	**0.536**
InstantRestore		0.533	0.534	71.324	0.555	0.494
FaceMe		**0.544**	**0.542**	70.716	0.586	0.480
Ours		**0.655**	**0.658**	**76.068**	**0.727**	**0.635**
DMDNet	FFHQ-Ref Moderate	0.572	0.564	72.024	0.633	0.477
ReF-LDM		**0.641**	**0.629**	**75.254**	**0.708**	**0.544**
RestorerID		0.605	0.590	72.111	**0.690**	**0.509**
InstantRestore		0.604	0.599	69.358	0.568	0.463
FaceMe		**0.640**	**0.629**	**72.480**	0.610	0.491
Ours		**0.701**	**0.689**	**76.220**	**0.738**	**0.613**
DMDNet	FFHQ-Ref Severe	0.137	0.127	62.177	0.543	0.355
ReF-LDM		**0.595**	**0.588**	**75.744**	**0.717**	**0.552**
RestorerID		0.425	0.407	**73.602**	**0.708**	**0.536**
InstantRestore		**0.473**	**0.477**	69.998	0.581	0.467
FaceMe		0.457	0.455	72.251	0.610	0.493
Ours		**0.701**	**0.692**	**75.679**	**0.726**	**0.600**
DMDNet	CelebHQRef100	0.379	0.370	70.529	0.604	0.465
ReF-LDM		**0.577**	**0.567**	**75.214**	**0.695**	**0.558**
RestorerID		0.459	0.430	**73.228**	**0.712**	**0.549**
InstantRestore		**0.510**	**0.507**	70.671	0.581	0.496
FaceMe		0.451	0.437	72.245	0.618	0.524
Ours		**0.700**	**0.700**	**75.673**	**0.729**	**0.625**

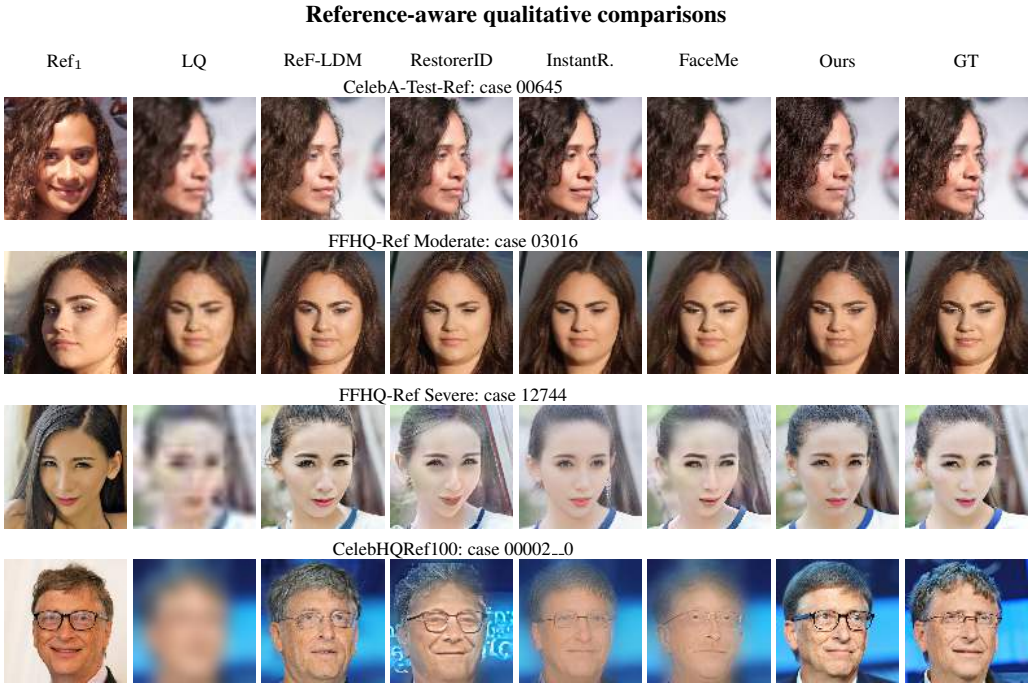

Figure 3: Reference-aware qualitative comparisons across four benchmarks. Each row shows Ref_1, LQ, method outputs, and GT under the same protocol references. IConFace keeps reference-consistent details while preserving the degraded facial structure.

Experiments

Experimental Setup

We train on FFHQ-Ref with online mixed blind degradations combining Real-ESRGAN and BSRGAN style degradation operators, and evaluate a single checkpoint in two modes. FFHQ-Ref train is used for training, while FFHQ-Ref val is used only for preview and qualitative validation. To expose the same model to varying reference availability, train-

ing samples use a mixed protocol: 30% without references, 30% with one reference, 20% with two references, and 20% with three references. When fewer references are available, we use the actual number without duplication.

Reference-aware benchmarks include CelebA-Test-Ref (2,533 samples), FFHQ-Ref Moderate (857), FFHQ-Ref Severe (857), and CelebHQRef100 (100 identities). No-reference benchmarks include CelebA-Test (3,000), LFW (1,711), CelebChild (360), WebPhoto (407), and Wider-Test (970). Reference-aware baselines include DMDNet, ReF-LDM, RestorerID, InstantRestore, and FaceMe (Li et al. 2022; Hsiao et al. 2024; Ying et al. 2024; Zhang et al. 2024; Liu et al. 2025); blind baselines include CodeFormer, VQFR, GFP-GAN, RestoreFormer++, and DAEFR (Zhou et al. 2022; Gu et al. 2022; Wang et al. 2021, 2023; Tsai et al. 2024).

All reported IConFace results use the same checkpoint with 12 sampling steps and the same restoration prompt in both modes. In reference-aware mode, up to three same-identity references are provided. In no-reference mode, reference tokens are absent and the identity pathway uses the degraded-image AdaFace fallback only as weak self-conditioning. For no-reference benchmarks, including real-world sets without paired ground truth, we report learned no-reference perceptual metrics.

CelebHQRef100 is used as a compact diagnostic split for severe same-identity reference transfer; its construction details are provided in the supplementary material. CelebA-Test provides paired synthetic degraded/GT portraits for degraded-only evaluation, while LFW, CelebChild, WebPhoto, and Wider-Test are real-world no-reference sets without paired targets. We therefore do not mix reference-aware and no-reference baselines in a single aggregate ranking; each method is evaluated in its intended inference mode, and the two result blocks answer separate protocol-specific restoration questions.

Reference-Aware Restoration

Reference-aware identity evaluation should separate reference utilization from paired target-state matching. Protocol references and GT targets are same-identity images but may differ in pose, expression, illumination, age, makeup, or local facial state; on CelebA-Test-Ref, 49.11% and 80.77% of Ref_1–GT pairs fall below AdaFace 0.6 and 0.7, respectively (supplementary Table 2). We therefore use Ref-AdaFace, supported by Ref-ArcFace, as the primary reference-aware identity measure, and report GT-AdaFace and structure checks in the supplement.

Table 1 reports reference-aware restoration results. IConFace achieves the best Ref-ArcFace and Ref-AdaFace on all four benchmarks, with especially large gains on FFHQ-Ref Severe and CelebHQRef100 where degraded identity evidence is weak. IConFace is also consistently strong on MUSIQ, CLIP-IQA, and MANIQA, supporting the intended identity–structure tradeoff rather than a pure identity-score optimization. Supplementary Table 4 further shows a reversal: IConFace is not highest on GT-AdaFace in easier splits, but leads on FFHQ-Ref Severe and CelebHQRef100, indicating that the reference pathway provides genuine identity recovery when degraded evidence is missing rather than merely copying reference state.

Table 2: No-reference restoration on LFW, CelebChild, WebPhoto, Wider-Test, and CelebA-Test.

Method	Dataset	MUSIQ ↑	CLIP ↑	MANIQA ↑
CodeFormer	LFW	75.484	0.689	0.527
GFP-GAN		**75.570**	0.676	**0.551**
VQFR		74.901	**0.725**	**0.543**
RF++		72.251	**0.702**	0.511
DAEFR		**75.840**	0.697	0.542
Ours		**76.712**	**0.758**	**0.645**
CodeFormer	CelebChild	**74.852**	0.686	0.521
GFP-GAN		74.822	0.674	0.530
VQFR		74.459	**0.711**	**0.542**
RF++		71.690	**0.702**	0.506
DAEFR		**74.883**	0.697	**0.537**
Ours		**75.603**	**0.750**	**0.613**
CodeFormer	WebPhoto	**74.004**	0.692	**0.503**
GFP-GAN		**75.213**	**0.702**	**0.543**
VQFR		71.602	0.690	0.502
RF++		71.487	**0.695**	0.490
DAEFR		72.705	0.669	0.494
Ours		**75.795**	**0.719**	**0.593**
CodeFormer	Wider-Test	73.407	0.699	0.496
GFP-GAN		**74.769**	0.700	**0.550**
VQFR		72.011	**0.722**	0.514
RF++		71.518	**0.717**	0.477
DAEFR		**74.143**	0.697	**0.520**
Ours		**75.496**	**0.729**	**0.616**
CodeFormer	CelebA-Test	**75.554**	0.671	0.538
GFP-GAN		**75.466**	0.672	**0.568**
VQFR		74.406	**0.691**	0.552
RF++		73.914	**0.689**	**0.553**
DAEFR		75.251	0.668	0.545
Ours		**75.988**	**0.724**	**0.631**

The largest margins appear in the severe and diagnostic reference-transfer settings. On FFHQ-Ref Severe, IConFace improves Ref-AdaFace by 0.104 over the best baseline (0.692 vs. 0.588), while also giving the best CLIP-IQA and MANIQA. On CelebHQRef100, the Ref-AdaFace margin reaches 0.133 (0.700 vs. 0.567). These gains are not obtained by sacrificing perceptual quality: the method remains first or second on MUSIQ and first on CLIP-IQA/MANIQA in the severe and diagnostic splits. Fig. 3 shows the same tradeoff visually. Under moderate degradation, IConFace preserves the degraded-image structure while restoring hair contours, eyes, skin texture, mouth boundaries, and fine identity marks such as the small mole around the lower chin. Under severe degradation, where competing methods often over-smooth, collapse, or miss identity-specific details, IConFace still produces satisfactory faces with clearer reference-consistent details and a layout anchored to the low-quality input. The joint improvement of Ref-AdaFace and perceptual metrics is important: it indicates that the reference pathway supplies identity evidence without merely optimizing a recognizer score at the expense of visual fidelity.

Figure 4: No-reference qualitative examples in the empty-reference mode. Each row shows one benchmark case.

No-Reference Generalization

The same checkpoint also operates without references by removing reference tokens and using the degraded-image AdaFace fallback. Table 2 shows that IConFace ranks first on MUSIQ, CLIP-IQA, and MANIQA across all five no-reference benchmarks, with average margins of 0.667 MUSIQ, 0.026 CLIP-IQA, and 0.069 MANIQA over the strongest baseline in each dataset block. The gains hold on synthetic CelebA-Test and real-world LFW, CelebChild, WebPhoto, and Wider-Test, indicating that reference-aware training does not over-specialize the model to reference-conditioned inputs. Fig. 4 further shows clearer and more realistic eyes, reasonable recovery of nearby hands or context when present, sharper facial boundaries, and stable visual quality under unknown degradations where other methods may fail or produce unstable facial details.

Implementation Details

IConFace is trained at 512 resolution using the FLUX.2-klein-base-4B restoration backbone with rank 16 LoRA adapters. Online degradations mix Real-ESRGAN and BSRGAN style degradation operators so that the model sees both common synthetic corruptions and stronger blind-restoration artifacts. The global identity pathway uses AdaFace IR50 embeddings with norm-only multi-reference aggregation and temperature 1.0. The degraded structure pathway uses low-rank degraded cross-attention with 256 compressed memory tokens, split into base and detail routes. Training mixes samples with zero, one, two, or three references, so the same checkpoint learns both empty-reference restoration and reference-aware restoration. All main results use guidance scale 4.0, seed 42, and 12 sampling steps.

Ablation Study

We ablate IConFace on FFHQ-Ref Moderate and FFHQ-Ref Severe to separate the two design goals: using reference identity evidence and keeping the result anchored to the degraded input. *Concat* keeps only the hybrid degraded-reference token sequence, testing whether concatenation alone can learn the identity–structure balance. *Struct* adds the degraded structure adapter, isolating the value of explicit degraded-image reinforcement. *ID* adds global AdaFace modulation, testing whether a compact identity anchor is more reliable than dense reference transfer. *1-Route* keeps both side pathways but replaces the base/detail split with one degraded-memory route, and *Full* is the complete IConFace design. The cumulative order avoids conflating module effects: it first tests implicit fusion, then explicit degraded-image reinforcement, then global identity anchoring, and finally the separation of coarse layout memory from local detail memory. This makes the comparison strictly modular.

Table 3: Reference-aware ablations on FFHQ-Ref Moderate and FFHQ-Ref Severe. Variants progressively add degraded-structure reinforcement, global identity conditioning, and two-route degraded memory. Ref metrics are computed against the first protocol reference.

Dataset	Variant	Ref-ArcFace ↑	Ref-AdaFace ↑	MUSIQ ↑	CLIP-IQA ↑	MANIQA ↑
FFHQ-Ref Moderate	Concat baseline	0.655	0.633	75.917	0.724	0.599
	+ degraded structure	0.656	0.636	75.980	0.733	0.606
	+ global identity	**0.678**	**0.671**	**76.088**	**0.737**	**0.616**
	+ single-route memory	**0.683**	**0.681**	**76.135**	**0.736**	**0.611**
	IConFace (full)	**0.688**	**0.689**	**76.220**	**0.738**	**0.613**
FFHQ-Ref Severe	Concat baseline	0.602	0.597	75.205	0.717	**0.595**
	+ degraded structure	0.612	0.609	75.316	**0.723**	**0.597**
	+ global identity	**0.654**	**0.660**	**75.430**	0.722	0.593
	+ single-route memory	**0.674**	**0.681**	**75.485**	**0.725**	**0.597**
	IConFace (full)	**0.682**	**0.692**	**75.679**	**0.726**	**0.600**

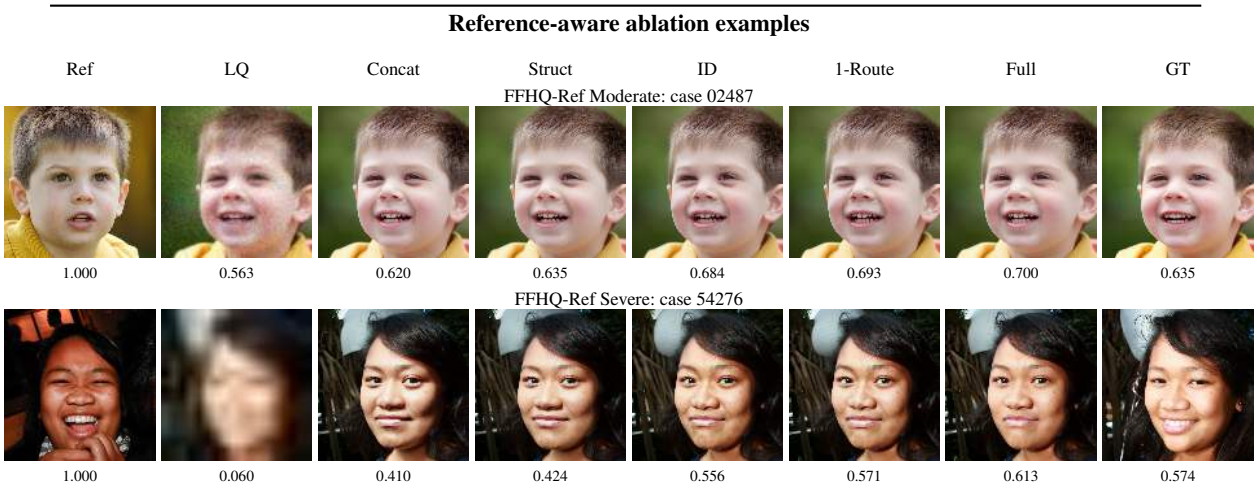

Figure 5: Reference-aware qualitative ablations on FFHQ-Ref Moderate and FFHQ-Ref Severe. Scores under crops report AdaFace similarity to the first protocol reference.

Table 3 shows that the modules play different roles. Struct gives modest but consistent gains, matching its purpose as a structure and quality stabilizer rather than an identity controller. ID gives the largest Ref-AdaFace jump, from 0.633 to 0.671 on Moderate and from 0.597 to 0.660 on Severe, showing that the global identity anchor is the main source of reference alignment. 1-Route further improves identity by letting blocks read compressed degraded evidence, but Full performs best because base memory preserves global layout while detail memory emphasizes local facial cues. Relative to Concat, Full improves Ref-AdaFace by 0.056 on Moderate and 0.095 on Severe. Fig. 5 provides a case-level view of the same ablation trend: adding the proposed modules steadily improves AdaFace similarity to the protocol reference, and the full model obtains the highest reference similarity in both examples.

Discussion and limitations. Ref metrics measure alignment to the supplied identity evidence and should be read together with visual comparisons, not as evidence of copying reference pose or expression. The results suggest that IConFace improves reference use while retaining the degraded input as the main spatial observation. This is most useful under severe degradation, where the reference provides identity evidence that the degraded image no longer contains, but the degraded image still offers coarse structure. The method is nevertheless limited by the quality and correctness of the references, the robustness of the frozen identity encoder, and the amount of spatial evidence preserved in the degraded input. Very low-quality, occluded, or identity-inconsistent references may produce unreliable identity anchors, and extremely corrupted inputs can still leave ambiguous local details.

Conclusion. IConFace assigns references a global identity role and the degraded image a spatial-structure role. With this asymmetric design, the same checkpoint supports both reference-aware and no-reference restoration. Experiments show that the global identity pathway improves reference-aligned identity consistency, while degraded-structure reinforcement and two-route memory provide consistent gains in the ablation study. The resulting model improves severe-case identity recovery, maintains perceptual quality, and remains robust in degraded-only mode.

References

Chong, M. J.; Xu, D.; Zhang, Y.; Wang, Z.; and Forsyth, D. 2025. Copy or Not? Reference-Based Face Image Restoration with Fine Details. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*.

Gu, Y.; Wang, X.; Xie, L.; Dong, C.; Li, G.; Shan, Y.; and Cheng, M.-M. 2022. VQFR: Blind Face Restoration with Vector-Quantized Dictionary and Parallel Decoder. In *European Conference on Computer Vision (ECCV)*, 126–143.

Hsiao, C.-W.; Liu, Y.-L.; Yang, C.-K.; Kuo, S.-P.; Jou, Y. K.; and Chen, C.-P. 2024. ReF-LDM: A Latent Diffusion Model for Reference-based Face Image Restoration. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Li, W.; Wang, M.; Zhang, K.; Li, J.; Li, X.; Zhang, Y.; Gao, G.; Deng, W.; and Lin, C.-W. 2025. Survey on Deep Face Restoration: From Non-blind to Blind and Beyond. *ACM Computing Surveys*.

Li, X.; Chen, C.; Zhou, S.; Lin, X.; Zuo, W.; and Zhang, L. 2020a. Blind Face Restoration via Deep Multi-scale Component Dictionaries. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 399–415.

Li, X.; Li, W.; Ren, D.; Zhang, H.; Wang, M.; and Zuo, W. 2020b. Enhanced Blind Face Restoration with Multi-Exemplar Images and Adaptive Spatial Feature Fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2706–2715.

Li, X.; Liu, M.; Ye, Y.; Zuo, W.; Lin, L.; and Yang, R. 2018. Learning Warped Guidance for Blind Face Restoration. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 272–289.

Li, X.; Zhang, S.; Zhou, S.; Zhang, L.; and Zuo, W. 2022. Learning Dual Memory Dictionaries for Blind Face Restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5): 5904–5919.

Lin, X.; He, J.; Chen, Z.; Lyu, Z.; Dai, B.; Yu, F.; Qiao, Y.; Ouyang, W.; and Dong, C. 2024. DiffBIR: Toward Blind Image Restoration with Generative Diffusion Prior. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

Liu, S.; Duan, Z.-P.; OuYang, J.; Fu, J.; Park, H.; Liu, Z.; Guo, C.; and Li, C. 2025. FaceMe: Robust Blind Face Restoration with Personal Identification. In *AAAI Conference on Artificial Intelligence*.

Meng, C.; He, Y.; Song, Y.; Song, J.; Wu, J.; Zhu, J.-Y.; and Ermon, S. 2022. SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations. In *International Conference on Learning Representations*.

Menon, S.; Damian, A.; Hu, S.; Ravi, N.; and Rudin, C. 2020. PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2437–2445.

Miao, Y.; Qu, Z.; Gao, M.; Chen, C.; Song, J.; Han, J.; and Deng, J. 2025. Unlocking the Potential of Diffusion Priors in Blind Face Restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 13471–13480.

Niu, A.; Pham, T. X.; Zhang, K.; Sun, J.; Zhu, Y.; Yan, Q.; Kweon, I. S.; and Zhang, Y. 2024. ACDMSR: Accelerated Conditional Diffusion Models for Single Image Super-Resolution. *IEEE Transactions on Broadcasting*, 70(2): 492–504.

Niu, A.; Zhang, K.; Pham, T. X.; Sun, J.; Zhu, Y.; Kweon, I. S.; and Zhang, Y. 2023. CDPMSR: Conditional Diffusion Probabilistic Models for Single Image Super-Resolution. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 615–619.

Niu, A.; Zhang, K.; Pham, T. X.; Wang, P.; Sun, J.; Kweon, I. S.; and Zhang, Y. 2025. Learning From Multi-Perception Features for Real-Word Image Super-Resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(7): 6535–6548.

Peebles, W.; and Xie, S. 2023. Scalable Diffusion Models with Transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4195–4205.

Qiu, X.; Han, C.; Zhang, Z.; Li, B.; Guo, T.; and Nie, X. 2023. DiffBFR: Bootstrapping Diffusion Model Towards Blind Face Restoration. In *Proceedings of the 31st ACM International Conference on Multimedia*, 7785–7795.

Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10684–10695.

Tsai, Y.-J.; Liu, Y.-L.; Qi, L.; Chan, K. C.; and Yang, M.-H. 2024. Dual Associated Encoder for Face Restoration. In *International Conference on Learning Representations (ICLR)*.

Varanka, T.; Toivonen, T.; Tripathy, S.; Zhao, G.; and Acar, E. 2024. PFStorer: Personalized Face Restoration and Super-Resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention Is All You Need. In *Advances in Neural Information Processing Systems*, volume 30.

Wang, J.; Gong, J.; Zhang, L.; Chen, Z.; Liu, X.; Gu, H.; Liu, Y.; Zhang, Y.; and Yang, X. 2025. OSDFace: One-Step Diffusion Model for Face Restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Wang, X.; Li, Y.; Zhang, H.; and Shan, Y. 2021. Towards Real-World Blind Face Restoration with Generative Facial Prior. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9168–9178.

Wang, Z.; Zhang, J.; Chen, R.; Wang, W.; and Luo, P. 2022. RestoreFormer: High-Quality Blind Face Restoration from Undegraded Key-Value Pairs. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17512–17521.

Wang, Z.; Zhang, J.; Chen, R.; Wang, W.; and Luo, P. 2023. RestoreFormer++: Towards Real-World Blind Face Restoration from Undegraded Key-Value Pairs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(4): 2555–2566.

Yan, Q.; Gong, D.; Shi, Q.; van den Hengel, A.; Shen, C.; Reid, I.; and Zhang, Y. 2019. Attention-Guided Network for Ghost-Free High Dynamic Range Imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1751–1760.

Yan, Q.; Zhang, L.; Liu, Y.; Zhu, Y.; Sun, J.; Shi, Q.; and Zhang, Y. 2020. Deep HDR Imaging via A Non-Local Network. *IEEE Transactions on Image Processing*, 29: 4308–4322.

Yang, L.; Wang, S.; Ma, S.; Gao, W.; Liu, C.; Wang, P.; and Ren, P. 2020. HiFaceGAN: Face Renovation via Collaborative Suppression and Replenishment. In *Proceedings of the 28th ACM International Conference on Multimedia*, 1551–1560.

Yang, P.; Zhou, S.; Tao, Q.; and Loy, C. C. 2023. PGDiff: Guiding Diffusion Models for Versatile Face Restoration via Partial Guidance. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Yang, T.; Ren, P.; Xie, X.; and Zhang, L. 2021. GAN Prior Embedded Network for Blind Face Restoration in the Wild. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 672–681.

Ye, H.; Zhang, J.; Liu, S.; Han, X.; and Yang, W. 2023. IP-Adapter: Text Compatible Image Prompt Adapter for Text-to-Image Diffusion Models. *arXiv preprint arXiv:2308.06721*.

Yin, Z.; Chen, J.; Liu, M.; Wang, Z.; Li, F.; Pei, R.; Li, X.; Lau, R. W. H.; and Zuo, W. 2026. RefSTAR: Blind Face Image Restoration with Reference Selection, Transfer, and Reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Ying, J.; Liu, M.; Wu, Z.; Zhang, R.; Yu, Z.; Fu, S.; Cao, S.-Y.; Wu, C.; Yu, Y.; and Shen, H.-L. 2024. RestorerID: Towards Tuning-Free Face Restoration with ID Preservation. *arXiv preprint arXiv:2411.14125*.

Zhang, H.; Alaluf, Y.; Ma, S.; Kadambi, A.; Wang, J.; and Aberman, K. 2024. InstantRestore: Single-Step Personalized Face Restoration with Shared-Image Attention. *arXiv preprint arXiv:2412.06753*.

Zhang, L.; Rao, A.; and Agrawala, M. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 3836–3847.

Zhao, Y.; Hou, T.; Su, Y.-C.; Jia, X.; Li, Y.; and Grundmann, M. 2023. Towards Authentic Face Restoration with Iterative Diffusion Models and Beyond. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 7312–7322.

Zhou, S.; Chan, K. C.; Li, C.; and Loy, C. C. 2022. Towards Robust Blind Face Restoration with Codebook Lookup Transformer. In *Advances in Neural Information Processing Systems (NeurIPS)*.