REFINE WHAT REMAINS: BASE-RELATIVE SELECTIVE REFINEMENT FOR REFERENCE-GUIDED FACE RESTORATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Reference-guided restoration can recover most facial content from a degraded image while leaving localized errors. A second full-image pass may correct these errors but can also change regions that the first pass restored well. We study *Base-relative selective refinement*: improving a fixed first-pass result (the Base) while limiting regressions in reliable regions. Our refiner conditions on the Base latent, degraded image, and same-identity reference. First, Base-conditioned adaptation projects repair-oriented updates against selected preservation gradients. These constraints provide a local, first-order safeguard rather than a guarantee for every output. Second, stochastic Flow-Map optimization combines global candidate credit with repair and preservation preferences. Differentiable trajectory replay provides action-sensitivity weights that transfer these preferences to position-level policy updates. Target-derived evidence guides training; at inference, an evidence predictor uses only available inputs. On CelebA-Test-Ref, the deployable Pred-H model improves paired reconstruction and identity measures over the Base, with a remaining distributional trade-off. Under oracle evidence, spatial credit improves Base-relative repair and preservation diagnostics.

1 INTRODUCTION

Reference-guided face restoration uses an image of the same identity to reduce ambiguity in a degraded observation. Recent methods improve direct restoration through diffusion, identity conditioning, and reference interaction (Hsiao et al., 2024; Liu et al., 2025; Zhou et al., 2026; Yin et al., 2026). Their common endpoint is a restored image; they do not specifically address how to improve that image when most of its content is already satisfactory.

We consider this post-restoration setting. A first pass gives a fixed restoration, which we call the *Base*. The next pass should correct its remaining identity, texture, or structural errors while reducing changes to regions already reconstructed well. This is a repair-and-preserve objective, not a claim that every reliable pixel will be unchanged. A simple additional restoration pass offers more computation, but does not explicitly distinguish repair from preservation. We therefore ask how to adapt a Base-conditioned model without immediately sacrificing reliable content, and how to assign repair and preservation preferences to its stochastic actions. Figure 1 illustrates this goal with one example.

During training and evaluation, a fixed Base makes both sides of this objective measurable against the same starting image and target: improvement in a residual error is useful, while deterioration in an already reliable feature is a cost. An image-level gain can hide that trade-off when a local facial cue becomes worse as another region improves. The refiner therefore needs a way to preserve successful first-pass content while directing additional capacity toward what remains wrong.

Our second pass refines a one-step MeanFlow Base (Geng et al., 2025) using its latent, the original degraded image, a reference, and evidence H. The evidence is target-derived in training and predicted from available inputs at inference. A projected adaptation stage is followed by Flow-Map policy optimization (Li et al., 2026) with replay-based spatial credit (Figure 2).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

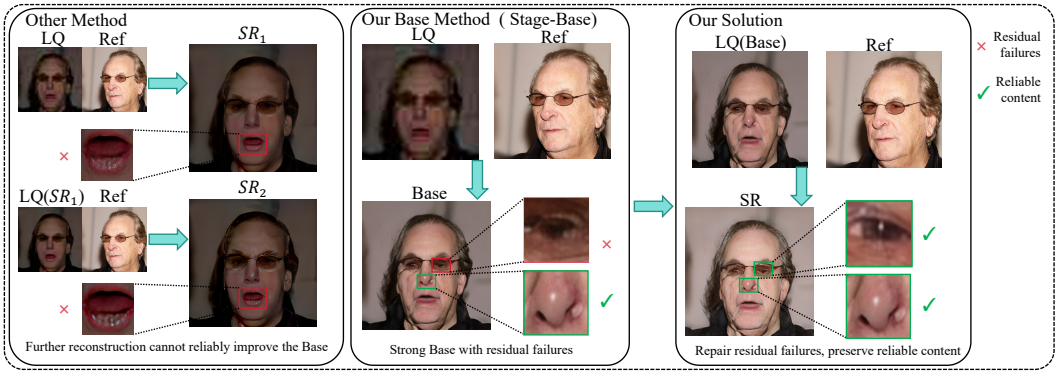

Figure 1: Illustrative Base-relative refinement. A repeated restoration retains a mouth error (left). The Base has an eye error and reliable nose detail (middle); the second pass aims to correct the former and retain the latter (right). The original LQ and reference remain available; marks denote local outcomes only.

Our contributions are:

- We formulate Base-relative selective refinement as a repair-and-preserve objective relative to a fixed first-pass result.
- We adapt a Base-conditioned refiner with a projected supervised update that limits first-order conflict with selected preservation boundaries.
- We augment Flow-Map policy optimization with Base-relative repair and preservation preferences weighted by replay-based action sensitivity.

2 Related Work

2.1 Reference-Guided Face Restoration

Reference-guided methods differ in how they represent and transfer same-identity information. ReF-LDM uses cached reference features (Hsiao et al., 2024); FaceMe introduces personalized identity conditioning (Liu et al., 2025); Zhou et al. (2026) combine high- and low-level reference cues; and RefSTAR studies reference selection and transfer (Yin et al., 2026). IConFace further studies fine-grained identity conditioning (Niu et al., 2026). These works primarily restore the degraded input. Our focus is narrower: improving an already restored image relative to its own fixed Base. This setting requires measuring both correction and regression against that Base.

2.2 Optimization for Generative Restoration and Refinement

Reinforcement learning has also been applied to generative restoration: RealSR-R1 studies real-world super-resolution (Qiao et al., 2025), LRPO studies blind face restoration (Wu et al., 2025), and GDPO-SR studies one-step super-resolution (Yi et al., 2026). IAPO investigates finer-grained preference attribution for diffusion models (Sun et al., 2026). Our optimization target differs in being defined relative to a fixed first-pass output, with distinct repair and preservation measurements.

Our adaptation is related to GEM and PCGrad, which modify gradients to reduce destructive interference (Lopez-Paz & Ranzato, 2017; Yu et al., 2020). We define the active constraints relative to the Base and apply them before policy optimization. Stage B uses the Flow-Map GRPO stochastic interface for deterministic flow-map generators (Li et al., 2026). Its distinguishing component here is a Base-relative, sensitivity-weighted policy surrogate that connects output-space repair and preservation quantities to latent action positions. It remains one globally coupled policy.

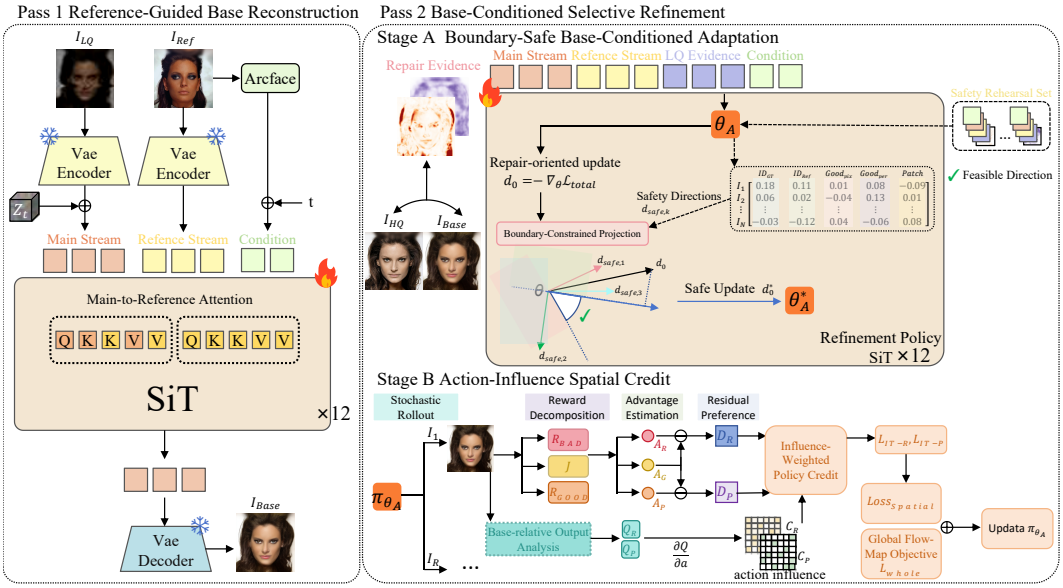

Figure 2: Two-pass framework. Pass 1 creates a frozen reference-guided Base. Stage A adapts a Base-conditioned refiner under preservation boundaries; Stage B adds action-influence-weighted spatial credit to global Flow-Map optimization. Target-derived evidence is used in training and predicted from available inputs at inference. Snowflakes and flames denote frozen and trainable components, respectively.

3 METHOD

The first pass remains frozen throughout both training stages, so all second-pass comparisons share the same Base.

3.1 PROBLEM FORMULATION

Given a degraded face I_{LQ} and a same-identity reference I_{Ref}, we define

$$\begin{aligned}(z_B, B) &= \mathcal{F}_{\theta_{\mathrm{base}}}(I_{\mathrm{LQ}}, I_{\mathrm{Ref}}),\\ \hat{I} &= \mathcal{G}_\theta(z_B, I_{\mathrm{LQ}}, I_{\mathrm{Ref}}, H),\\ H_{\mathrm{oracle}} &= \Phi(B, I_{\mathrm{HQ}}).\end{aligned} \tag{1}$$

Here, B is the decoded first-pass restoration, referred to as the *Base*, while its latent z_B is retained for direct reuse by the second pass. For each input, the Base remains fixed throughout second-pass optimization and provides the common point of comparison for subsequent refinement.

The second pass treats z_B as its primary image condition, I_{LQ} as complementary observation evidence, and I_{Ref} as an identity and appearance prior. The evidence H indicates where further correction should be prioritized; it specifies neither a signed correction direction nor target appearance and does not restrict generation with a hard spatial mask. The full-image process can therefore adjust globally coupled identity and structure while using H as a soft indication of local need. Its output-space coordinates are not assumed to correspond one-to-one with stochastic action coordinates. The target-derived oracle evidence has two bounded channels:

$$\begin{aligned}H_{\mathrm{pix}} &= \mathrm{clip}_{[0,1]}\big(\mathrm{AvgPool}_8\big(\mathrm{mean}_{\mathrm{RGB}}\,|B - I_{\mathrm{HQ}}|\big)\big)\,,\\ d &= \mathrm{AvgPool}_8\big(\max\{\mathrm{LPIPS}_{\mathrm{spatial}}(B, I_{\mathrm{HQ}}), 0\}\big), \qquad H_{\mathrm{perc}} = \frac{d}{1+d}.\end{aligned} \tag{2}$$

These fixed transforms bound both channels to $[0, 1]$ without per-image or dataset-level rescaling; equal numerical values need not indicate equal pixel and perceptual error. At inference, an

auxiliary predictor estimates the same two bounded channels from the Base, degraded observation, and reference. This H_{pred} uses no high-quality target and retains the same conditioning interface as H_{oracle}. Both maps have shape $2 \times 64 \times 64$ for 512×512 RGB inputs. We use $W_{\text{BAD}} = \text{upsample}(\max\{H_{\text{pix}}, H_{\text{perc}}\})$ and $W_{\text{GOOD}} = 1 - W_{\text{BAD}}$ for Base-relative region weighting.

We define refinement explicitly relative to the fixed Base: residual failures should be reduced, changes to already reliable content should be limited, and global identity and perceptual consistency should be maintained. We refer to this empirical goal as *Base-relative selective refinement*; it is not guaranteed for every output.

3.2 Strong Reference-Guided MeanFlow Base

The first pass constructs the Base using a reference-guided one-step MeanFlow model (Geng et al., 2025) with a SiT-B/2 backbone (Ma et al., 2024). A frozen VAE (Rombach et al., 2022) maps the degraded observation and same-identity reference to latent space. The main stream combines the current MeanFlow latent with the degraded-image condition, while a separate embedder forms the reference stream. Within the shared-weight dual-stream Transformer, main-stream queries attend jointly to keys and values from both the main and reference streams, whereas the reference stream is updated through self-attention and feed-forward transformations using the same Transformer parameters. A frozen ArcFace encoder (Deng et al., 2019) provides the identity embedding, which enters through AdaLN-Zero modulation (Peebles & Xie, 2023) rather than spatial tokens or attention keys and values.

The first-pass model is trained with the MeanFlow objective together with identity, structural, and perceptual supervision. After training, it is frozen as the Base generator. Its one-step prediction directly produces z_B, which is decoded into B while being retained as the primary second-pass image condition; a copy of the first-pass parameters initializes the trainable refinement model.

3.3 Boundary-Safe Base-Conditioned Adaptation

Before stochastic optimization, we adapt the first-pass model to its new role as a Base-relative selective refiner. The retained Base latent replaces the degraded latent as the primary image condition, and multi-level Base features are supplied through a residual source branch. This makes the current restoration directly available to the refiner, but the Base alone cannot reveal which of its details are wrong. The degraded observation therefore remains available through a residual adapter that operates on the difference between LQ and Base embeddings and uses a zero-initialized output projection. The reference retains its first-pass spatial-token and identity-modulation pathways. The evidence H softly regulates Base-feature retention and learned repair residuals without masking the generation domain or modifying the reference pathway. The LQ residual may supply observation cues absent from the Base. Evidence adjusts Base-feature retention and repair within the feature pathway, not through a final image-space blend; a local error can therefore require nonlocal changes to structure or identity.

Base conditioning alone does not ensure preservation of content that the Base already reconstructs reliably. We therefore form the repair-oriented update proposal

$$d_0 = -\nabla_\theta \left(\mathcal{L}_{\text{anchor}} + \lambda_{\text{general}} \mathcal{L}_{\text{general}} + \lambda_{\text{repair}} \mathcal{L}_{\text{repair}} \right), \tag{3}$$

where $\mathcal{L}_{\text{anchor}}$ retains established restoration behavior, $\mathcal{L}_{\text{general}}$ covers identity and reliable-content preservation, and $\mathcal{L}_{\text{repair}}$ targets residual Base-relative errors.

Base-relative preservation boundaries. To expose preservation risks that may be absent from the current minibatch, we use deterministic safety-domain rehearsal over identity, reliable-content pixel and perceptual fidelity, and local structural quality. For each preservation family, a *worst-unsafe* state captures an existing violation and a *nearest-safe* state captures an admissible state closest to failure. Each active state b_k induces the descent direction $d_{\text{safe},k} = -\nabla_\theta \mathcal{L}_{\text{safe},k}(\theta; b_k)$. Local structural preservation additionally uses a weighted-quantile boundary aligned with the PATCH criterion, so vulnerable local content is not obscured by average preservation losses. The two states cover existing violations and currently safe content close to failure. Considering both matters because an averaged safety loss can hide a local regression behind gains elsewhere.

Boundary-constrained update. Rather than applying d_0 directly, we project it onto directions that remain first-order compatible with all active preservation boundaries and do not reverse the repair direction:

$$\begin{aligned} \mathcal{C} &= \{d : \langle d, d_{\text{safe},k} \rangle \geq 0,\ \forall k, \quad \langle d, d_{\text{repair}} \rangle \geq 0\}, \\ d^\star &= \arg\min_{d \in \mathcal{C}} \frac{1}{2}\|d - d_0\|_2^2, \end{aligned} \tag{4}$$

where $d_{\text{repair}} = -\nabla_\theta \mathcal{L}_{\text{repair}}$. For sufficiently small updates and fixed active states, the selected preservation losses do not increase to first order, while the repair constraint prevents a first-order reversal of repair descent. This is not a guarantee for finite updates or for Stage B. The feasible set contains zero, so projection can reduce a conflicting proposal rather than force an unsafe step. It leaves d_0 unchanged when the proposal satisfies every active preservation and repair constraint. The resulting policy θ_{adapt} provides the preservation-aware Base-conditioned initialization for Stage B.

3.4 Base-Relative Spatial Credit via Action Influence

Starting from θ_{adapt}, Stage B optimizes refinement behavior through the stochastic policy interface of Flow-Map GRPO (Li et al., 2026). Its global objective is

$$\mathcal{L}_{\text{whole}} = \mathcal{L}_{\text{score}}^{\text{FM}} + \mathcal{L}_{\text{tail}}^{\text{path}} + \mathcal{L}_{\text{deploy}}^{\text{path}}. \tag{5}$$

Here, $\mathcal{L}_{\text{score}}^{\text{FM}}$ is the group-relative score-function term over stochastic Flow-Map transitions. The tail term backpropagates through the deterministic endpoint from a detached stochastic state, while the deployment term follows the standard deterministic second-pass trajectory.

For each restoration condition and stochastic replica, let $\{\hat{I}^{(i)}\}_{i=1}^{G}$ denote its candidate group, with $G = 24$; we suppress the replica index below. The score-function branch assigns whole-candidate credit, but Base-relative refinement also requires distinguishing repair from preservation. The full reward J combines identity, BAD-region repair, GOOD-region preservation, global quality, Pareto preference, and a catastrophic identity penalty. The non-differentiable catastrophic term is excluded from the pathwise branches. The full reward, repair reward R_{BAD}, and preservation reward R_{GOOD} are centered over this condition-local group and normalized by the shared scale of that replica. We denote the resulting advantages by A_G, A_R, and A_P, respectively. We use their residual preferences

$$D_R = A_R - A_G, \qquad D_P = A_P - A_G. \tag{6}$$

Thus, the global branch continues to determine overall candidate preference, while D_R and D_P capture how repair and preservation advantages differ from the full-reward advantage. A candidate favored globally can still perform poorly on one Base-relative objective; the residuals make that disagreement available to the spatial branch. For example, identity and overall quality may improve while a reliable eye worsens. If the candidate's preservation advantage falls below its global advantage, $A_P < A_G$ and $D_P < 0$; this disagreement can guide spatial credit without replacing the global reward.

Continuous Base-relative correspondence. The residual preferences above indicate *what* should receive additional credit, but they do not identify which stochastic actions influence the corresponding output behavior. We therefore construct differentiable output-space proxies. Let p index the spatial evidence lattice, write $H(p) = [H_{\text{pix}}(p), H_{\text{perc}}(p)]$, and let $e_m(I, I_{\text{HQ}}; p)$ denote the local discrepancy for $m \in \{\text{pix}, \text{perc}\}$ with fixed positive calibration scale s_m. We define

$$\begin{aligned} W_R(p) &= \max\{H_{\text{pix}}(p), H_{\text{perc}}(p)\}, \qquad W_P(p) = 1 - W_R(p), \\ g_{i,m}(p) &= \frac{e_m(B, I_{\text{HQ}}; p) - e_m(\hat{I}^{(i)}, I_{\text{HQ}}; p)}{s_m}, \qquad m \in \{\text{pix}, \text{perc}\}. \end{aligned} \tag{7}$$

Here, $g_{i,m}(p) > 0$ means that candidate i improves over the Base at location p under discrepancy m, and $g_{i,m}(p) < 0$ indicates a regression. We then define

$$\begin{aligned} Q_R(\hat{I}^{(i)}) &= \text{mean}_p \left[W_R(p) \frac{g_{i,\text{pix}}(p) + g_{i,\text{perc}}(p)}{2} \right], \\ Q_P(\hat{I}^{(i)}) &= \text{mean}_p \left[W_P(p) \frac{g_{i,\text{pix}}(p) + g_{i,\text{perc}}(p)}{2} \right]. \end{aligned} \tag{8}$$

Q_R and Q_P are not additional rewards or evaluation metrics. They provide continuous signals for locating repair- and preservation-related sensitivity in the stochastic trajectory. For a given condition, regional weights are fixed before candidates are scored, so a candidate cannot redefine where repair or preservation is measured. Positive Q_R indicates aggregate Base-relative repair, while negative Q_P indicates aggregate regression in reliable regions. Both proxies assess the same globally coupled output.

Action influence transport. Output-space evidence cannot be assigned to the same spatial coordinate of the stochastic action by assumption: the Transformer, subsequent stochastic transitions, deterministic endpoint, and VAE all couple information across space. For each stored old-policy trajectory, we therefore replay the same stochastic path while preserving its sampled innovations. The old-policy parameters remain fixed, while the action under analysis and its downstream dependence are kept differentiable. Fixed innovations remove variation from resampling. Replay through later transitions and the decoder retains cross-location effects that a same-coordinate evidence mask would miss. For $T \in \{R, P\}$, the resulting action influence at transition t and action-lattice position q is

$$\begin{aligned} I_{T,t}(q) &= \frac{1}{G}\sum_{i=1}^{G}\left\| \frac{\partial Q_T(\widetilde{I}_t^{(i)})}{\partial a_{i,t}(q)} \right\|_2, \\ C_{T,t}(q) &= \begin{cases} I_{T,t}(q)/\operatorname{mean}_{q'} I_{T,t}(q'), & \operatorname{mean}_{q'} I_{T,t}(q') > 0, \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{9}$$

Here, $\widetilde{I}_t^{(i)}$ is obtained by differentiably replaying the stored stochastic trajectory rather than generating a separate deterministic candidate. The normalized field $C_{T,t}$ is detached before policy optimization and used only to weight credit. Because it is based on a gradient norm, it measures local sensitivity rather than the signed effect of an action. If its underlying sensitivity vanishes, the corresponding spatial contribution is set to zero. The preference sign comes from D_T; $I_{T,t}$ supplies nonnegative weights. Mean normalization removes the arbitrary overall scale of these influence norms while retaining their relative spatial pattern.

Influence-weighted policy credit. The Gaussian policy still defines a single full-action distribution. For action-lattice position q, let

$$\begin{aligned} \ell_{i,t,q}^{\theta} &= \sum_{c} \log \mathcal{N}\left(a_{i,t,c,q}; \mu_{\theta,i,t,c,q}, \sigma_t^2\right), \\ \log \pi_\theta(a_{i,t} \mid s_{i,t}) &= \sum_{q} \ell_{i,t,q}^{\theta}, \\ \delta_{i,t}(q) &= \ell_{i,t,q}^{\theta} - \operatorname{stopgrad}\left(\ell_{i,t,q}^{\text{old}}\right). \end{aligned} \tag{10}$$

We transport the repair or preservation influence to the policy log-density change by

$$\begin{aligned} \bar{\delta}_{i,t}^{T} &= \operatorname{mean}_q\left[C_{T,t}(q)\,\delta_{i,t}(q)\right], \\ \rho_{i,t}^{T} &= \exp\left(\bar{\delta}_{i,t}^{T}\right), \qquad T \in \{R, P\}. \end{aligned} \tag{11}$$

Unlike the full-action likelihood ratio, ρ^T is a weighted, position-averaged surrogate for spatial credit. Using a clipped group-relative objective, ρ^R is paired with D_R to form $\mathcal{L}_{\mathrm{IT-R}}$, while ρ^P is paired with D_P to form $\mathcal{L}_{\mathrm{IT-P}}$. These terms do not define independent regional policies; they reweight position-level changes of the same globally coupled policy according to their measured action sensitivity.

The spatial contribution and complete Stage-B objective are

$$\begin{aligned} \mathcal{L}_{\text{spatial}} &= \mathcal{L}_{\mathrm{IT-R}} + \mathcal{L}_{\mathrm{IT-P}}, \\ \mathcal{L}_{\text{StageB}} &= \mathcal{L}_{\text{whole}} + \alpha_k \mathcal{L}_{\text{spatial}}. \end{aligned} \tag{12}$$

The two spatial branches are equally weighted; $\alpha_k \geq 0$ controls their joint weight at update k and is absent from the global-only ablation. Only the stochastic score-function branch receives the influence-based spatial credit; the tail and deterministic deployment objectives retain their global definitions.

Candidate comparisons remain condition-local, while each parameter update accumulates gradients over all 256 eligible training conditions. The Base-relative diagnostics below measure whether this policy preserves reliable regions in practice.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Data and protocol. The first-pass model is trained on the Arc2Face corpus (Paraperas Papantoniou et al., 2024) using same-identity reference pairs. Evaluation uses a frozen 2,533-condition protocol based on the public CelebA-Test-Ref benchmark introduced by ReF-LDM (Hsiao et al., 2024), whose source images are from CelebA-HQ (Karras et al., 2018). The 256 refinement-training conditions and 2,533 test conditions have disjoint identities. Each condition contains one high-quality target, one pre-generated degraded observation, and one fixed same-identity canonical reference. The test set is partitioned into four degradation-severity strata (common, hard, severe, and extreme). Every compared method receives the same degraded observation and canonical reference and is evaluated against the same high-quality target. Methods whose official interfaces require multiple reference slots use only duplicated or flipped versions of the same canonical reference, so the unique-reference information budget remains one. The first-pass Base is fixed and shared across all refinement variants.

External baselines. We compare against eight public reference-based restoration methods: ReF-LDM (Hsiao et al., 2024), RefSTAR (Yin et al., 2026), FaceMe (Liu et al., 2025), RestorerID (Ying et al., 2024), InstantRestore (Zhang et al., 2025), DMDNet (Li et al., 2023), PGDiff (Yang et al., 2023), and ASFFNet512 (Li et al., 2020). All methods use their official checkpoints and inference pipelines under the same frozen test protocol. Required preprocessing failures are retained rather than removed, and their handling follows the frozen benchmark protocol.

Standard restoration evaluation. We report PSNR and SSIM for reconstruction fidelity, LPIPS-Alex (Zhang et al., 2018) for perceptual distance, MUSIQ (Ke et al., 2021) for no-reference perceptual quality, FID-GT (Heusel et al., 2017) for distributional fidelity, and ArcFace (Deng et al., 2019) cosine similarity to the ground-truth target and canonical reference (ID-GT and ID-Ref) for identity fidelity. FID-GT is computed with clean-fid (Parmar et al., 2022) once over the complete set of 2,533 restored images against the corresponding ground-truth set; the remaining metrics are aggregated over individual test samples. Oracle-H uses target-derived H_{oracle} to isolate the refinement mechanism, whereas Pred-H estimates the same evidence from the Base, degraded observation, and reference without access to the high-quality target. For refinement variants, Table 1 reports one deterministic second-pass output per test sample, consistent with the Base-relative refinement analysis.

Base-relative diagnostics. Standard image-level metrics do not reveal whether refinement repairs residual failures while preserving content that is already reliable in the Base. We therefore evaluate refinement relative to the fixed Base using deterministic outputs and sample-level paired comparisons. The Base-relative protocol measures three aspects of selective refinement: repair of residual failures, preservation of reliable content, and local structural stability.

Let G_{pix} and G_{perc} denote Base-relative repair gains on residual regions, where positive values indicate improvement over the Base. Let D_{pix} and D_{perc} denote degradation on reliable regions, where positive values indicate deterioration. For each sample, these quantities compare the candidate and Base errors to the same target under fixed BAD and GOOD weights, respectively. Event rates use the signs of paired differences, while the continuous gains retain their magnitudes. We further measure local structural worsening using F_{patch}: the fraction of worsened valid 32×32 patches at stride 32. A patch is valid when its GOOD weights sum to a positive value, and worsens when its GOOD-weighted pixel or spatial-LPIPS error strictly exceeds the Base; the fraction is undefined if no patch is valid. This evaluation diagnostic is distinct from the 64×64 training PATCH constraint. We do not report an unfrozen binary Patch-Fail rate.

We report event summaries derived from these continuous quantities: *Repair Success* ($G_{\mathrm{pix}} > 0 \wedge G_{\mathrm{perc}} > 0$), *Reliable-Pix Fail* ($D_{\mathrm{pix}} > 0$), and *Reliable-Perc Fail* ($D_{\mathrm{perc}} > 0$). Patch degradation is reported directly as F_{patch}. All metrics are computed from fixed Base-relative comparisons using

the same paired test samples. The scoring maps are computed from fixed Base–target comparisons and do not change with the map supplied as policy conditioning.

Training and comparison protocol. The Base generator remains frozen throughout refinement. The second-pass model is initialized from the first-pass parameters and adapted in Stage A; Stage B starts from the resulting θ_{adapt}. Stage B uses four replicas and 24 candidates per condition. Advantages are centered within each condition and divided by the replica-level reward standard deviation plus 10^{-4}, then clipped to $[-5, 5]$. The global policy ratio is clipped to $[1 - 10^{-4}, 1 + 10^{-4}]$, and combined gradients are clipped at norm 1. Optimizer-facing gradients are accumulated over the full 256-condition training domain before each parameter update. Within each ablation, variants share the same initialization, training data, optimization budget, and evaluation protocol unless stated otherwise.

4.2 Main Results

Table 1 compares eight public reference-based restoration baselines with the frozen Base and successive refinement variants under the same single-reference protocol. Base denotes the first-pass restoration; Stage-A Adaptation is the post-adaptation model evaluated with Oracle-H. Full Refinement denotes the complete Stage-B objective with Base-relative spatial credit via action influence. Full Refinement (Pred-H) is the deployable configuration.

Table 1: Main results on CelebA-Test-Ref under the frozen single-reference protocol. Bold red denotes the best value in each metric across all rows; † marks Oracle-H rows using target-derived evidence.

Method	PSNR↑	SSIM↑	LPIPS↓	MUSIQ↑	FID-GT↓	ID-GT↑	ID-Ref↑
ReF-LDM	16.8067	0.5050	0.3678	68.6248	23.2034	0.6032	0.4878
RefSTAR	16.5080	0.4808	0.3402	67.4507	**17.9327**	0.6991	0.6326
FaceMe	16.5512	0.4988	0.4152	59.1073	37.6183	0.5680	0.4272
RestorerID	16.4703	0.4902	0.4023	**70.3332**	36.7361	0.5419	0.4285
InstantRestore	16.3984	0.4755	0.3509	66.2196	20.5813	0.6547	0.5387
DMDNet	16.4282	0.4900	0.4249	60.8694	44.2424	0.4687	0.3274
PGDiff	16.0382	0.4728	0.4048	62.5187	31.3040	0.6271	**0.8935**
ASFFNet512	16.2400	0.4808	0.4027	63.0992	32.5668	0.4701	0.3545
Ours: Base	17.2043	0.5518	0.3405	63.7321	18.4164	0.7093	0.6519
Ours: Stage-A Adaptation†	18.4278	0.5673	0.3124	67.6744	20.5415	0.7201	0.6720
Ours: Full Refinement (Oracle-H)†	**18.4687**	**0.5710**	**0.3077**	68.2416	18.1493	**0.7287**	0.6813
Ours: Full Refinement (Pred-H)	18.4371	0.5685	0.3112	67.8907	19.7214	0.7253	0.6788

Among the eight external methods, Pred-H leads on PSNR, SSIM, LPIPS, and ID-GT, but not MUSIQ, FID-GT, or ID-Ref (Table 1). Stage A provides most paired-metric gains over the Base: PSNR rises from 17.2043 to 18.4278 dB and LPIPS falls from 0.3405 to 0.3124. Its FID-GT worsens from 18.4164 to 20.5415. Oracle-H later lowers FID-GT to 18.1493, but this staged comparison changes more than spatial credit and cannot explain the recovery; the controlled ablation below isolates spatial credit for Base-relative diagnostics.

Pred-H uses no target at inference. It reaches 18.4371 dB PSNR, 0.3112 LPIPS, and 0.7253 ID-GT, close to the Oracle-H values of 18.4687 dB, 0.3077, and 0.7287. Its FID-GT of 19.7214 remains worse than the Base, so deployable paired-metric gains carry a distributional trade-off. Table 1 separates oracle mechanism checks from deployable Pred-H evaluation. Highlighted Oracle-H values describe target-informed mechanism checks. Stage A to full Oracle-H introduces Stage B; full Oracle-H to Pred-H changes the evidence source. Neither contrast isolates spatial credit, which Table 2 tests under fixed oracle evidence.

Figure 3 presents two qualitative examples. These illustrations complement the aggregate metrics but do not establish regional repair or preservation rates.

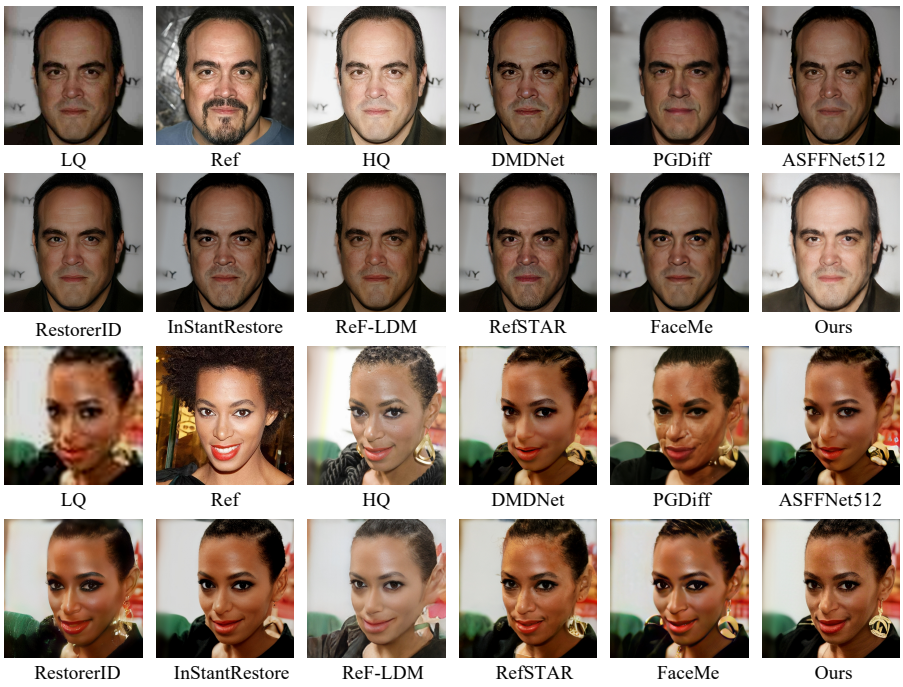

Figure 3: Two CelebA-Test-Ref examples showing the degraded input, fixed same-identity reference, target, external outputs, and our Pred-H output. Table 1 reports aggregate comparisons.

4.3 Ablation and Analysis

Base-relative spatial credit. Under Oracle-H, we isolate spatial credit by comparing the global and full objectives from the same Stage-A model.

Table 2: Controlled Stage-B ablation under oracle evidence on CelebA-Test-Ref. Success requires positive pixel and perceptual repair gains; Pix-Fail and Perc-Fail are degradation events on reliable regions. Larger repair gains and smaller patch degradation are better.

Variant	Repair success↑	Pix-Fail↓	Perc-Fail↓	Pixel gain↑	Perceptual gain↑	Patch degradation↓
Global Flow-Map objective	81.41%	12.79%	12.12%	0.2405	0.6655	0.4377
+ Base-relative spatial credit	83.14%	11.37%	11.25%	0.2478	0.6750	0.4281

Under oracle evidence, adding spatial credit raises Repair Success from 81.41% to 83.14% and lowers both reliable-region failure rates (Table 2). The continuous repair gains also increase, while patch degradation falls. This controlled comparison supports the combined spatial-credit term beyond the global objective. It does not isolate the action-influence weighting from other parts of that term or establish the same regional effect for Pred-H.

5 Conclusion

We study selective refinement of a fixed first-pass Base. Projected adaptation addresses first-order conflicts with preservation boundaries; replay-based spatial credit adds repair and preservation preferences to global Flow-Map optimization. On CelebA-Test-Ref, the deployable Pred-H model improves paired reconstruction and identity measures over the Base, although FID-GT worsens. Under Oracle-H, adding spatial credit improves Base-relative repair and preservation diagnostics.

AI USE STATEMENT

Generative AI tools assisted with drafting and editing prose, checking references and format, providing feedback on the research methodology and experimental presentation, and interpreting the reported results and limitations. The authors verified the AI-assisted changes and take responsibility for the manuscript's scientific claims and conclusions.

REPRODUCIBILITY STATEMENT

Section 3 defines the Base, adaptation, and spatial-credit objectives. Section 4 describes the fixed evaluation conditions, metrics, and paired Base-relative comparisons. All refinement variants are evaluated against the same frozen Base and test conditions.

REFERENCES

Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4690–4699, 2019.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, J. Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Chi-Wei Hsiao, Yu-Lun Liu, Cheng-Kun Yang, Sheng-Po Kuo, Yucheun Kevin Jou, and Chia-Ping Chen. ReF-LDM: A latent diffusion model for reference-based face image restoration. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In *International Conference on Learning Representations*, 2018.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. MUSIQ: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5148–5157, 2021.

Xiaoming Li, Wenyu Li, Dongwei Ren, Hongzhi Zhang, Meng Wang, and Wangmeng Zuo. 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*, pp. 2706–2715, 2020.

Xiaoming Li, Shiguang Zhang, Shangchen Zhou, Lei Zhang, and Wangmeng Zuo. Learning dual memory dictionaries for blind face restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5904–5917, 2023.

Zhiqi Li, Wen Zhang, and Bo Zhu. Flow-Map GRPO: Reinforcement learning for few-step flow-map generators via anchored stochastic composition. *arXiv preprint arXiv:2607.00535*, 2026.

Siyu Liu, Zheng-Peng Duan, Jia OuYang, Jiayi Fu, Hyunhee Park, Zikun Liu, Chun-Le Guo, and Chongyi Li. FaceMe: Robust blind face restoration with personal identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 5567–5575, 2025.

David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Nanye Ma, Mark Goldstein, Michael S. Albergo, Nicholas M. Boffi, Eric Vanden-Eijnden, and Saining Xie. SiT: Exploring flow and diffusion-based generative models with scalable interpolant transformers. In *European Conference on Computer Vision*, pp. 23–40, 2024.

Axi Niu, Jinyang Zhang, and Senyan Qing. IConFace: Fine-grained identity conditioning for reference-aware face restoration. *arXiv preprint arXiv:2605.02814*, 2026.

Foivos Paraperas Papantoniou, Alexandros Lattas, Stylianos Moschoglou, Jiankang Deng, Bernhard Kainz, and Stefanos Zafeiriou. Arc2Face: A foundation model for ID-consistent human faces. In *Proceedings of the European Conference on Computer Vision*, pp. 241–261, 2024.

Gaurav Parmar, Richard Zhang, and Jun-Yan Zhu. On aliased resizing and surprising subtleties in gan evaluation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11410–11420, 2022.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4195–4205, 2023.

Junbo Qiao, Miaomiao Cai, Wei Li, Xudong Huang, Jie Hu, Xinghao Chen, Shaohui Lin, and Hongkai Xiong. RealSR-R1: Reinforcement learning for real-world image super-resolution with vision-language chain-of-thought. *arXiv preprint arXiv:2506.16796*, 2025.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022.

Jiayang Sun, Pin Wang, Hongbo Wang, Xinyue Liu, Huaibo Huang, and Ran He. Towards fine-grained attribution: Instance-aware preference optimization for aligning diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 43155–43164, 2026.

Bin Wu, Yahui Liu, Chi Zhang, Yao Zhao, and Wei Wang. Enhancing blind face restoration through online reinforcement learning. *arXiv preprint arXiv:2509.23339*, 2025.

Peiqing Yang, Shangchen Zhou, Qingyi Tao, and Chen Change Loy. PGDiff: Guiding diffusion models for versatile face restoration via partial guidance. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Qiaosi Yi, Shuai Li, Rongyuan Wu, Lingchen Sun, Zhengqiang Zhang, and Lei Zhang. GDPO-SR: Group direct preference optimization for one-step generative image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2177–2187, 2026.

Zhicun Yin, Junjie Chen, Ming Liu, Zhixin Wang, Fan Li, Renjing Pei, Xiaoming Li, Rynson W. H. Lau, and Wangmeng Zuo. RefSTAR: Blind face image restoration with reference selection, transfer, and reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 12053–12062, 2026.

Jiacheng Ying, Mushui Liu, Zhe Wu, Runming Zhang, Zhu Yu, Siming Fu, Si-Yuan Cao, Chao Wu, Yunlong Yu, and Hui-Liang Shen. RestorerID: Towards tuning-free face restoration with ID preservation. *arXiv preprint arXiv:2411.14125*, 2024.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Howard Zhang, Yuval Alaluf, Sizhuo Ma, Achuta Kadambi, Jian Wang, and Kfir Aberman. InstantRestore: Single-step personalized face restoration with shared-image attention. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, pp. 1–10, 2025.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018.

Mo Zhou, Keren Ye, Viraj Shah, Kangfu Mei, Mauricio Delbracio, Peyman Milanfar, Vishal M. Patel, and Hossein Talebi. Reference-guided identity preserving face restoration. *Transactions on Machine Learning Research*, 2026.