EMBER: EARLY-TRAJECTORY MODULATION AND NO-REFERENCE POLICY OPTIMIZATION FOR OLD PHOTO RESTORATION

Anonymous authors
Paper under double-blind review

ABSTRACT

A faded photograph can be the last visible trace of a person or moment that no longer exists, but restoring it is difficult without clean references or reliable damage masks. Our approach begins with two observations about a pretrained image editor. Its first two denoising steps produce a spatial edit-response signal associated with the regions where the final output later differs from the input. No-reference policy optimization also amplifies restoration and colorization responses that supervised fine-tuning does not consistently express. Based on these observations, we introduce **EMBER**, which combines a Gated 1–2 Step Edit-Response Field (G12) with no-reference group-relative policy optimization (GRPO) on real old photographs. G12 freezes the early response into mask-free rotary-phase modulation, while GRPO adapts the editor using a reward with no explicit color statistic. Across GGT-100K Old Photo and RealOld-74, EMBER performs best or remains competitive on most no-reference metrics not used as training rewards. In a blinded study, 31 participants completed 1,018 valid image-level evaluations across 48 photographs, selecting their two preferred results in each evaluation. Across all seven methods, EMBER achieved the highest Top-2 inclusion rate at 58.3%. EMBER restores these photographs in the form in which they often survive: as a single damaged image, without clean targets during policy optimization or external masks at deployment. Code is available at `https://anonymous.4open.science/r/artifact-af5213a/`.

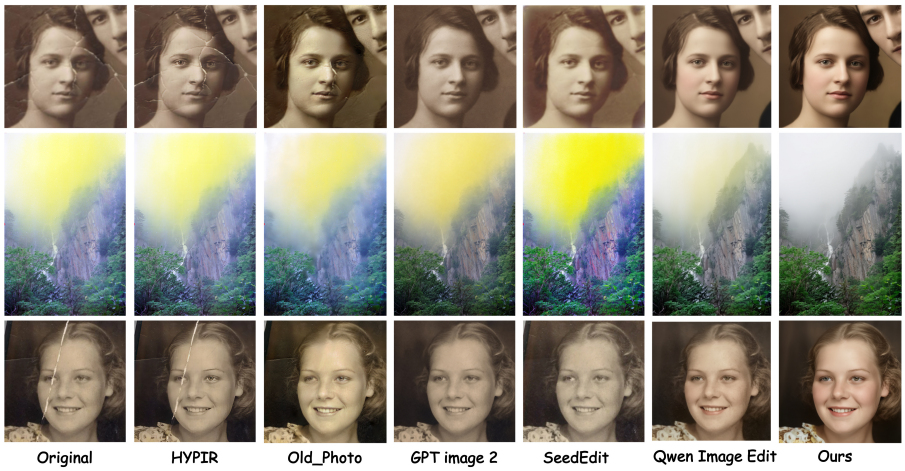


Figure 1: **Real old-photo restoration.** Examples show removal of compound degradations and plausible colorization. Deployment uses neither a clean reference nor a damage mask. All methods receive the same input.

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.

1 INTRODUCTION

Old photographs preserve people and moments long after the scenes they record have disappeared. The photographs themselves, however, accumulate folds, water damage, fungal growth, fading, and scanner artifacts. These degradations interact and rarely follow a known forward process. Unlike a synthetic benchmark sample, a real photograph usually comes alone: there is no clean reference and often no reliable damage mask.

Existing methods, including Bringing Old Photos Back to Life (BOPB), commonly rely on paired or synthetically degraded supervision Wan et al. (2023); Xu et al. (2023); Cai et al. (2024), or draw on large generative priors Lin et al. (2025a); Yang et al. (2026). Many still depend on paired targets, masks, or degradation-specific conditioning. Such assumptions transfer imperfectly to the entangled damage found in real old photographs. This gap motivates evaluation on real photographs in addition to synthetic test pairs Wan et al. (2023); Liu et al. (2025a). Figure 1 places this problem in its real setting: physical damage, fading, and missing color appear together, and every method operates without a user-provided damage mask.

Prior work shows that diffusion representations encode spatial correspondence and semantic structure Tang et al. (2023); Luo et al. (2023). We ask whether an editor's restoration trajectory and no-reference quality feedback can guide restoration. Our experiments yield two bounded observations. During the first two denoising steps, we observe a structured edit-response field associated with later editing, although the field is not an accurate damage map. The policy after supervised fine-tuning (SFT) does not consistently express the quality and color responses observed after brief policy optimization. In particular, no-reference policy optimization can increase a measurable, prompt-supported color response without directly rewarding color.

The first observation motivates G12, the Gated 1–2 Step Edit-Response Field: the field estimated from the first two denoising steps is gated for temporal reliability and frozen, and from step 3 onward it modulates the frame-axis phases of target-token query/key (Q/K) rotary position encoding (RoPE) Bai et al. (2025); Su et al. (2024). This yields an internal, mask-free spatial signal that adds no denoiser pass during rollout or inference (cf. difference-based editing localization Couairon et al. (2023)).

The second observation motivates no-reference group-relative policy optimization (GRPO) on real old photographs Shao et al. (2024); Liu et al. (2025b): sampled candidates are scored by two no-reference assessors and the policy is regularized toward its SFT anchor. The reward contains no explicit color statistic (§3.6).

G12 and GRPO form **EMBER**, a unified system that restores a damaged old photograph without a clean reference or external spatial mask at deployment.

Contributions.

- **Unified restoration from a single damaged photograph.** We introduce EMBER to address compound physical damage, detail loss, and plausible color recovery within one image editor. Deployment requires only the photograph and a fixed instruction, without a clean reference, damage mask, quality assessor, or separate colorization model.

- **Early trajectories as an internal spatial signal.** We show that the first two denoising steps contain a structured field associated with later edits even when downstream injection is disabled. G12 gates and freezes this field, then modulates target-token Q/K rotary phases during later steps without an external mask or an additional denoiser evaluation.

- **Real-photo policy optimization without paired targets.** We adapt the editor on unpaired real photographs using dual-IQA GRPO and an SFT-anchor penalty. Across three seeds, both rewards increase consistently, while the released model amplifies a measurable color response without an explicit color statistic. EMBER leads four of five non-reward metrics on the official GGT-100K OldPhoto test subset. In the 48-photograph blinded study, it is selected on 58.3% of 1,018 valid ballots, compared with 42.9% for RealRestorer.

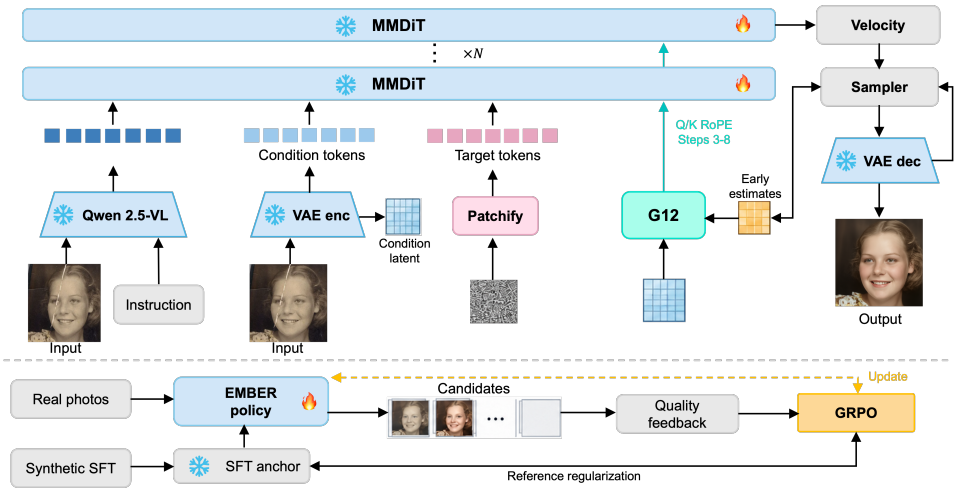


Figure 2: **EMBER architecture and training.** A low-rank-adapted multimodal diffusion transformer uses G12 to convert early trajectory estimates into target-token Q/K RoPE modulation. No-reference GRPO adapts the policy on real photographs while regularizing it toward a frozen SFT anchor.

2 METHOD

2.1 EMBER ARCHITECTURE AND TWO-STAGE OPTIMIZATION

EMBER adapts the pretrained Qwen-Image-Edit-2511 model (Wu et al., 2025; Qwen Team, 2025), whose image editor uses a multimodal diffusion transformer (MMDiT). We train low-rank adaptation (LoRA) weights (Hu et al., 2022) while freezing the remaining parameters. Given a degraded photograph y and fixed instruction $\mathcal{T}$, the model samples $x \sim \pi_\theta(\cdot \mid y, \mathcal{T})$. Training first uses synthetic pairs $\{(y_s, x_s)\}$ for supervised fine-tuning. Appendix E records the deployed instruction and its one-sentence SFT variant. GRPO then adapts the policy on unpaired real photographs $\{y_r\}$, using two no-reference assessors and the frozen SFT model as its reference.

As shown in Fig. 2, frozen Qwen2.5-VL (Bai et al., 2025; Wu et al., 2025) and variational autoencoder (VAE) encoders produce semantic and condition tokens, while noisy latents form target tokens. LoRA adapts the MMDiT Q, K, V, and output projections. G12 estimates an edit-response field from steps 1–2 and modulates only target-token Q/K RoPE during steps 3–8. It applies no attention mask and does not directly modify the semantic or condition streams. At inference, EMBER uses only y and $\mathcal{T}$, without a clean reference, external mask, degradation label, or quality assessor.

2.2 G12: GATED 1–2 STEP EDIT-RESPONSE FIELD

As observed in §1, the model's first two denoising steps produce clean estimates that already exhibit spatially structured responses correlated with subsequent editing locations. G12 converts this observation into a deterministic, mask-free positional modulation through three stages: **field extraction**, **temporal reliability gating**, and **frozen rotary-phase modulation**.

Figure 3 summarizes the data flow. At steps 1–2, G12 compares each predicted clean latent with the same condition latent and extracts D_1, D_2. Their temporal agreement produces a frozen 12×12 field. At steps 3–8, this field is mapped only to target-token Q/K rotary phases. G12 applies no offset to V or to condition-token and semantic-token coordinates. The signal is internal to the sampling trajectory rather than an externally supplied mask. G12 changes relative rotary phases involving target tokens according to the early response field. It neither imposes a binary region mask nor guarantees that any individual attention weight will increase.

Stage 1: Edit-response field extraction. At denoising step $t \in \{1, 2\}$, the flow velocity $v_\theta(\mathbf{z}_t, \mathbf{y}, \mathcal{T}, t)$ yields a clean-latent estimate:

$$\hat{\mathbf{z}}_0^{(t)} = \mathbf{z}_t - \sigma_t \, v_\theta(\mathbf{z}_t, \mathbf{y}, \mathcal{T}, t), \tag{1}$$

where σ_t is the noise-schedule coefficient at step t. The channel-wise median absolute residual between this estimate and the condition latent $\mathbf{z}_y$ provides a spatial discrepancy signal:

$$R_t = \text{median}_c \left| \hat{\mathbf{z}}_{0,c}^{(t)} - \mathbf{z}_{y,c} \right|. \tag{2}$$

We pool R_t to 12×12, subtract a 3×3 average low-pass, and clamp negative responses. We then normalize the high-pass response by its 0.9-quantile and add a condition-normalized global residual term with weight $\alpha = 0.25$. Appendix D gives exact pseudocode. The resulting D_t is an **edit-response field**, not a damage probability map. It is associated with later editing (partial $\rho = 0.307$) but does not accurately segment damage.

Stage 2: Temporal reliability gating. Two observations are the minimum that permits a cross-step agreement statistic before the field is frozen. We combine D_1 and D_2 with an exponential reliability gate:

$$D_{\text{frozen}} = \frac{1}{2}(D_1 + D_2) \odot \exp\left[-\lambda \left| D_1 - D_2 \right|\right], \qquad \lambda = 2. \tag{3}$$

Agreement retains response magnitude, while disagreement is suppressed. The field is computed once and reused at steps 3–8.

Stage 3: Frame-axis rotary-phase modulation. The frozen field is expanded from 12×12 to the full target-token spatial grid via nearest-neighbor interpolation. For each target token at spatial position r, we compute a frame-axis rotary position offset:

$$p_f(r) = -\left\lfloor \kappa \cdot D_{\text{frozen}}(r) \right\rfloor, \tag{4}$$

where κ controls modulation strength. We use $\kappa = 8$ for steps 3–7 and $\kappa = 4$ for the final step 8, tapering the intervention as generation converges. This offset is applied to **both Q and K projections** of target tokens in the frame-axis dimension of the RoPE.

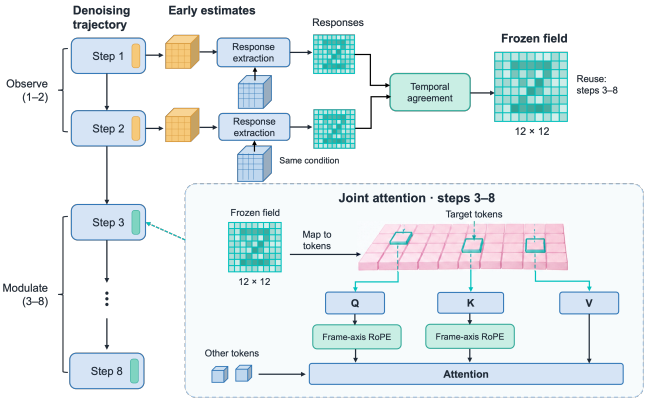


Figure 3: **G12.** The frozen field modulates target-token Q/K RoPE.

The same field construction and κ schedule are used in SFT, GRPO rollout, and inference; during policy optimization, each selected transition replays its rollout field and κ.

The offset changes rotary-phase relations involving target tokens. We interpret this as positional modulation because its effect on individual attention weights need not be monotonic. Section 3.5 tests exact field placement.

Preventing ground-truth leakage during SFT. During supervised fine-tuning, the clean target $\mathbf{x}_s$ is available as the denoising objective. A standard training forward pass mixes $\mathbf{z}_{\mathbf{x}_s}$ with noise. Estimating the field from that pass could leak the clean target into D. To prevent this, we estimate D from a **detached two-step preview trajectory** initialized from pure noise and conditioned on $\mathbf{y}$, run without gradient. The field uses no clean-target values.

2.3 No-Reference Group-Relative Policy Optimization

G12 supplies a field-derived positional modulation. This section defines *what constitutes a good edit* when no ground truth is available. We adapt the model's restoration behavior on real old photographs through group-relative policy optimization with no-reference rewards.

Candidate sampling. For each real photograph $\mathbf{y}_r$, we sample $K = 8$ restoration candidates $\{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$ from the current policy π_θ. Each candidate uses the same eight-step G12 field-estimation and modulation path used at inference. The rollout additionally uses the flow stochastic differential equation (Flow-SDE) exploration process specified in Appendix C.

No-reference reward. We evaluate each candidate using two no-reference image quality assessors:

$$q_1(\mathbf{x}) = \mathrm{MUSIQ}(\mathbf{x}), \qquad q_2(\mathbf{x}) = \text{CLIP-IQA}(\mathbf{x}). \tag{5}$$

MUSIQ (Ke et al., 2021) is a learned multi-scale image-quality assessor. CLIP-IQA (Wang et al., 2023) supplies a CLIP-based quality estimate. For candidate i in the full rollout batch, let $g(i)$ denote its input group. For reward j, we standardize scores across the $K = 8$ candidates in $g(i)$. We average the two normalized rewards, then standardize over rollout batch $\mathcal{B}$:

$$z_{ij} = \frac{q_j(\mathbf{x}_i) - \bar{q}_{j,g(i)}}{\max(s_{j,g(i)}, \epsilon)}, \quad a_i = \tfrac{1}{2}(z_{i1} + z_{i2}), \quad A_i = \frac{a_i - \bar{a}_{\mathcal{B}}}{\max(s_{\mathcal{B}}, \epsilon)}. \tag{6}$$

Here $\bar{q}_{j,g(i)}$ and $s_{j,g(i)}$ are the within-group mean and standard deviation for assessor j, and $\epsilon = 10^{-6}$ prevents division by zero. Thus z_{ij} is candidate i's standardized relative score under assessor j within the eight candidates generated for the same input. The scalar a_i is their equal-weight composite. It is an intermediate group utility, not a raw IQA score. The terms $\bar{a}_{\mathcal{B}}$ and $s_{\mathcal{B}}$ are the mean and standard deviation over the full rollout batch. The resulting A_i is the group-relative advantage used by GRPO.

Policy update. The LoRA parameters are updated with the clipped group-relative surrogate of GRPO (Shao et al., 2024) and a frozen-anchor reference penalty. No learned critic or value function is used; the policy signal is A_i from Eq. (6). For the selected Flow-SDE transition, the implementation's *x-based reference penalty* is the mean squared shift between policy and anchor next-latent means:

$$\mathcal{L}_{\mathrm{RL}} = \mathcal{L}_{\mathrm{GRPO}}(A_i) + \beta\, \mathbb{E}\left[\|\mu_\theta - \mu_0\|_2^2 / d \right], \qquad \beta = 0.01. \tag{7}$$

Here d is the number of latent elements. Appendix B gives the full per-dimension likelihood-ratio surrogate. The reference term regularizes the policy toward the supervised baseline while proxy quality scores supply the optimization signal.

On reward-amplified color response. The GRPO instruction permits natural-looking colorization (Appendix E), yet the reward in Eq. (5) contains *no* explicit color, saturation, or chrominance term; §3.6 measures the resulting reward-amplified color response, whose mechanism the five-way probes of Appendix H (Table 12) do not isolate.

Eight-step rollout contract. SFT applies the detached field to one random-timestep flow-matching forward with $\kappa = 8$ rather than executing a gradient-bearing rollout; GRPO rollout and deployment use the full eight-step solver, estimating the field at steps 1–2 and reusing it at steps 3–8. Appendix F gives the full pseudocode, Appendix E.1 a cross-implementation agreement check, and Tables 13 and 14 the solver-step and observation-window sweeps.

Table 1: **Results on the official GGT-100K OldPhoto test subset (50 image pairs).** No-reference metrics follow TOPIQ-NR (Chen et al., 2024), MANIQA (Yang et al., 2022), LIQE (Zhang et al., 2023), TReS (Golestaneh et al., 2022), and NIQE (Mittal et al., 2013).

	No-reference quality					Full-reference diagnostics		
Method	TOPIQ-NR ↑	MANIQA ↑	LIQE ↑	TReS ↑	NIQE ↓	PSNR ↑	SSIM ↑	LPIPS ↓
External methods								
BOPB (Wan et al., 2023)	0.384	0.286	1.858	62.704	4.956	<u>23.06</u>	0.7561	0.2722
RealRestorer (Yang et al., 2026)	0.429	0.320	2.543	59.659	4.462	19.93	0.7111	<u>0.2501</u>
HYPIR (Lin et al., 2025b)	0.560	0.394	2.545	<u>73.654</u>	**3.753**	**24.39**	<u>0.7569</u>	0.2844
SeedEdit (Shi et al., 2024)	0.282	0.248	1.185	39.684	5.559	22.81	**0.7837**	0.3056
GPT Image 2 (OpenAI, 2025)	0.399	0.292	2.106	56.977	4.685	22.84	0.7283	**0.2218**
Qwen-Image-Edit-2511 (Wu et al., 2025; Qwen Team, 2025)	0.268	0.254	1.305	36.967	6.145	21.61	0.7065	0.3370
EMBER variants								
EMBER (non-G12)	<u>0.567</u>	<u>0.418</u>	<u>3.233</u>	73.198	5.846	16.10	0.5156	0.4478
EMBER (G12)	**0.615***	**0.429***	**3.909***	**79.354***	<u>4.377</u>	17.94	0.6937	0.2979

Bold and underline mark first and second place. Asterisks on EMBER (G12) mark unadjusted pointwise paired 95% bootstrap intervals against the strongest external baseline for that metric that exclude zero. These marks are descriptive and do not denote multiplicity-adjusted significance. Reward metrics are omitted.

3 EXPERIMENTS

We evaluate restoration quality, early edit prediction, color response, and the released operating point.

3.1 DATASETS, SPLITS, AND EVALUATION PROTOCOL

Synthetic pairs support SFT. The 8,000 pre-degradation targets comprise 7,000 FFHQ images (Karras et al., 2019), 800 DIV2K train-HR images, 100 DIV2K validation-HR images (Agustsson & Timofte, 2017), and 100 additional images from the prepared secondary-image pool. We form each input by randomly composing 2–5 degradations, including fading, blur, scratches, stains, tears, grain, JPEG compression, and paper texture. Neither RealOld-74 nor GGT-100K enters this SFT pool. The official GGT-100K OldPhoto pairs supply generated high-quality references only for secondary full-reference diagnostics (Kong et al., 2026). These references are generative targets, not camera-captured clean ground truth. GRPO uses 500 unpaired real photographs. Each update samples eight inputs and eight candidates per input; a fixed 20-image set only monitors training. The 500-image pool, 20-image monitor, RealOld-74, and GGT-100K test subset are mutually disjoint. RealOld-74 is used for no-reference evaluation, including the checkpoint curves reported in Section 3.3. The official 50-pair GGT-100K OldPhoto test subset (Kong et al., 2026) is never used for training or checkpoint inspection.

The self-assembled real-photo pool and RealOld-74 combine internet-sourced images with images originating from BlueNeg (Liu et al., 2025a). We make no broader dataset claim and do not describe these images as newly captured. We report five *held-out* no-reference metrics (none used as a reward; "held-out NR metrics" below): TOPIQ-NR (Chen et al., 2024), MANIQA (Yang et al., 2022), LIQE (Zhang et al., 2023), TReS (Golestaneh et al., 2022), and NIQE (Mittal et al., 2013). Peak signal-to-noise ratio (PSNR), structural similarity (SSIM) (Wang et al., 2004), and Learned Perceptual Image Patch Similarity (LPIPS) (Zhang et al., 2018) are secondary diagnostics on paired GGT-100K. They measure agreement with the provided generative references, not recovery of an observed historical ground truth. RealOld-74 has no clean targets, so it receives only no-reference evaluation.

Table 2: **No-reference results on the self-assembled RealOld-74 evaluation set.** Metrics follow TOPIQ-NR (Chen et al., 2024), MANIQA (Yang et al., 2022), LIQE (Zhang et al., 2023), TReS (Golestaneh et al., 2022), and NIQE (Mittal et al., 2013). All metrics use 74 images except LIQE, which uses the same 72 images for every method because its implementation requires a short side of at least 224 pixels.

Method	TOPIQ-NR ↑	MANIQA ↑	LIQE ↑	TReS ↑	NIQE ↓
External methods					
BOPB (Wan et al., 2023)	0.435	0.295	2.539	64.949	4.443
RealRestorer (Yang et al., 2026)	0.496	0.356	3.232	67.580	4.261
HYPIR (Lin et al., 2025b)	<u>0.583</u>	0.414	3.088	73.001	**3.789**
SeedEdit (Shi et al., 2024)	0.321	0.239	1.506	43.166	4.926
GPT Image 2 (OpenAI, 2025)	0.466	0.315	2.691	61.300	<u>4.239</u>
Qwen-Image-Edit-2511 (Wu et al., 2025)	0.346	0.270	1.892	45.216	6.150
EMBER variants					
EMBER (non-G12)	0.557	<u>0.420</u>	<u>3.663</u>	<u>77.385</u>	5.131
EMBER (G12)	**0.584**	**0.430**	**3.992***	**80.525***	4.267

RealOld-74 has no clean targets. Formatting follows Table 1.

3.2 BASELINES AND IMPLEMENTATION

We compare six external methods spanning an old-photo specialist, blind restoration systems, and general-purpose generative editors: BOPB (Wan et al., 2023), RealRestorer (Yang et al., 2026), HYPIR (Lin et al., 2025b), SeedEdit (Shi et al., 2024), Qwen-Image-Edit-2511 (Wu et al., 2025; Qwen Team, 2025), and GPT Image 2 (OpenAI, 2025). Methods receive the same input and a matched instruction where applicable. We apply no manual enhancement or per-image selection. The accompanying evaluation artifacts record exact versions, prompts, sizes, and sampling budgets.

The reported EMBER variants follow separate trajectories. EMBER (non-G12) is historical checkpoint-19; EMBER (G12) is formal checkpoint-20. Their difference is an end-to-end comparison, not a causal estimate of field placement. The non-G12 branch is labelled "Vanilla" in the training-dynamics figures. SFT used eight BF16 processes on PPU 810e accelerators. The formal GRPO run used one PPU 810e with DeepSpeed ZeRO-2. Full settings are in Tables 3 and 4.

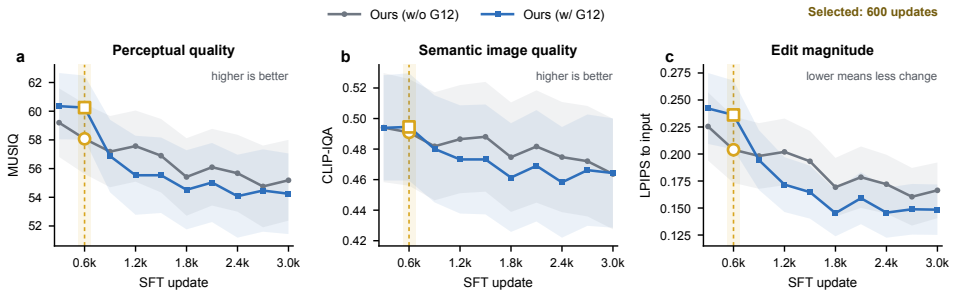


Figure 4: **SFT anchor selection.** Checkpoint curves on RealOld-74 using an aspect-preserving 512^2-pixel target-area budget. Gold markers denote the shared update-600 operating point.

3.3 SFT ANCHOR SELECTION WITH A 512^2-PIXEL TARGET-AREA BUDGET

Figure 4 reports the checkpoint sweep for the two SFT branches under an aspect-preserving 512^2-pixel target-area budget (3,000 updates, evaluated every 300 on RealOld-74). We freeze update 600 as a shared operating point rather than the maximum of every curve: G12 updates 300 and 600 reach similar MUSIQ and CLIP-IQA, while update 600 changes the input less (LPIPS 0.236 versus 0.242) and later checkpoints degrade. The formal G12 release uses this anchor; the main-table non-G12 checkpoint-19 belongs to the separate historical trajectory described above, and the untouched GGT-100K subset provides the separate official test evidence.

3.4 MAIN COMPARISON ON REAL OLD PHOTOGRAPHS

Tables 1 and 2 report the quantitative comparisons on GGT-100K OldPhoto and RealOld-74, respectively. Figure 5 shows representative outputs. On GGT-100K OldPhoto, EMBER has the highest mean TOPIQ-NR, MANIQA, LIQE, and TReS. HYPIR remains better on NIQE. Appendix J gives paired effects and intervals in Table 8.

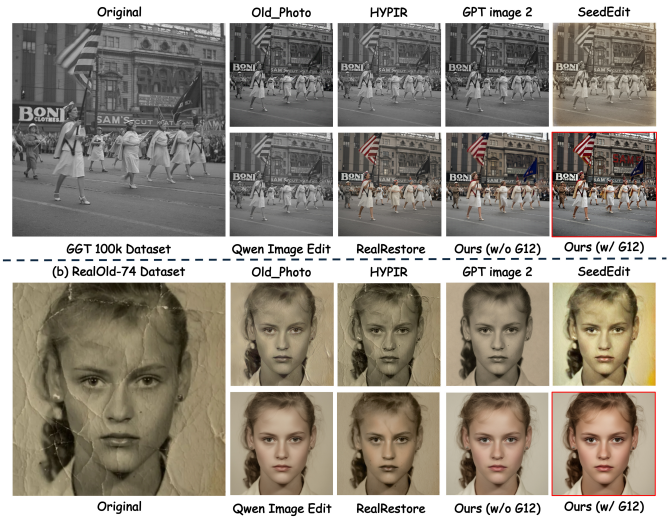


Figure 5: **Qualitative comparison.** Results on GGT-100K OldPhoto and RealOld-74. Red outlines identify EMBER. Zoom in for better view.

The G12 versus non-G12 comparison is end-to-end (§3.2). G12 improves the mean on all five held-out NR metrics within EMBER on GGT-100K, with four pointwise paired intervals excluding zero. It also improves all three full-reference diagnostics. External methods retain stronger absolute full-reference scores in several columns.

On RealOld-74, G12 likewise has the highest mean on four of five metrics across all evaluated methods. The RealOld checkpoint curves are disclosed in Section 3.3; GGT-100K provides a separate official comparison. The paired TOPIQ-NR and MANIQA differences from HYPIR are unresolved, while the within-EMBER gains are more consistent.

3.5 WHAT DOES THE EARLY FIELD ENCODE?

With downstream G12 injection disabled, the early field remained associated with final edit magnitude across 100 synthetic images and three seeds. Spearman ρ was 0.256, with a 95% confidence interval (CI) from 0.206 to 0.305. Partial ρ was 0.307 (95% CI, 0.265–0.348), and the mean within-image permutation correlation was $\rho = 0.002$ (Appendix Fig. 11a). Agreement with binary damage masks was weak on 28 valid cases (area under the precision–recall curve, AUPRC, 0.179 at prevalence 0.169; area under the receiver operating characteristic curve, AUROC, 0.544). We treat the field as an edit-response signal rather than a damage map, since binary masks do not label diffuse fading, tonal imbalance, or regions that the model may recolor.

In a fixed-checkpoint intervention, the correct field did not outperform five exact-histogram permutations or disabled modulation on the mask-defined LPIPS endpoints (Appendix Fig. 11b). Because the permutations preserve the field histogram, that comparison isolates spatial placement at $\kappa = 8$; its scope limits are discussed in §4. The secondary damage-region L1 endpoint also favored the permuted field (scaled effect -3.128, unadjusted $p = 0.0021$; Appendix Table 11). Appendices K and I provide the complete statistics and analysis.

3.6 POLICY OPTIMIZATION ELICITS A MEASURABLE COLOR RESPONSE

Across three 20-update dual-IQA GRPO runs, MUSIQ increased by 3.36 ± 0.50 and CLIP-IQA by 0.034 ± 0.011 (mean $\pm$ seed standard deviation), with all gains positive. Matched non-G12 runs also improved, while G12 produced more late-stage color candidates (38.9% versus 24.6%) without an explicit color reward (Appendix G.2). Relative to its G12 SFT anchor, the released checkpoint improved all five held-out NR metrics on both sets, with all ten pointwise paired intervals excluding zero (Table 9); on GGT-100K, TOPIQ-NR rose by 0.1474 and NIQE fell by 0.5174. Agreement with the provided generative references worsened under all three full-reference diagnostics, consistent with the perception–distortion tradeoff (Blau & Michaeli, 2018). On the fixed monitor set of the released seed-42 run, saturation increased by 0.0390 (95% CI, 0.0280–0.0509; 19/20 images positive) and colorfulness by 2.94 (95% CI, 1.79–4.07; 18/20 positive). Across 1,280 rollout candidates, neither MUSIQ nor the combined advantage showed a clear color association, and CLIP-IQA was weakly associated only with saturation ($\rho = 0.070$, 95% CI 0.002–0.139; Table **??**). These estimates quantify a reward-amplified color response, not historical correctness.

3.7 REWARD ABLATIONS AND THEIR LIMITS

Matched 20-update controls isolate the two IQA terms and the reference penalty (Table 5). Each single-assessor arm improved its active monitor score and increased both color statistics. CLIP-IQA-only produced a larger color response than MUSIQ-only, while dual-IQA produced the larger response of the regularized arms. Removing the reference penalty in this run gave slightly larger endpoint proxy gains and color changes. In these single-seed controls, the response was not exclusive to either assessor and persisted without the reference penalty.



Figure 6: **Qualitative component ablation.** Matched outputs and enlargements of the same hull region.

An earlier matched comparison with the DeQA lower confidence bound (DeQA-LCB) (You et al., 2025) appears in Table 6. Both objectives showed transient monitor peaks that did not persist to update 100. These controls use one seed and do not prove that dual-IQA is optimal. We retain the reference penalty to penalize next-latent mean displacement from the SFT anchor. This single run did not establish a better endpoint for the regularized arm.

Figure 6 compares matched SFT and GRPO outputs on one photograph (an end-to-end comparison; §3.2); the G12 crop shows cleaner rail and plate structure.

3.8 BLINDED HUMAN PREFERENCE EVALUATION

Figure 7 summarizes 1,018 valid ballots from 31 participants across 48 photographs. Each ballot selected two of seven anonymous outputs. EMBER was selected on 58.3% of ballots, compared with 42.9% for the runner-up, RealRestorer; the 15.3-point margin has a crossed-bootstrap 95% CI of 2.8 to 28.4 points. A conditional logit model adjusted for display position and gave an odds ratio of 0.59 for RealRestorer relative to EMBER (95% CI, 0.41 to 0.86; Holm-adjusted $p = 0.006$). Standard errors were clustered by participant and photograph.

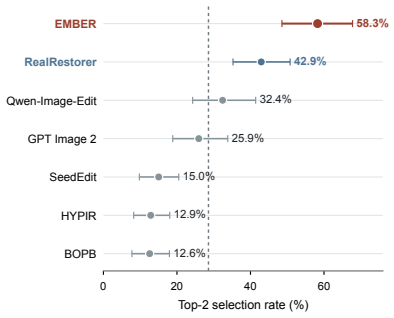


Figure 7: **Blinded Top-2 preference.** The dashed line marks uniform selection.

This blinded study does not reuse either training reward as its outcome. It supports perceived quality but does not isolate color, sharpness, fidelity, identity, or historical accuracy. Appendix N gives the protocol, complete rates, stratified results, sensitivity analyses, and a representative questionnaire page.

4 DISCUSSION AND LIMITATIONS

The five held-out NR metrics do not reuse MUSIQ or CLIP-IQA. However, four are learned no-reference IQA models and may share feature or dataset biases with the reward assessors, so their agreement should not be read as five independent confirmations. The human study provides complementary evidence, but it does not isolate fidelity.

The G12 versus non-G12 comparison remains end-to-end rather than a causal estimate of field placement (§efsec:baselines). The controlled intervention that motivated this reading isolates spatial placement alone: exact placement did not improve the two mask-defined LPIPS endpoints, and neither it nor the unablated design choices behind the field (median residual, 12×12 grid, high-pass operation, quantile normalization, RoPE injection) test alternative field-strength distributions, diffuse color or tonal edits, or G12-assisted training. We therefore interpret the field as an edit-response signal rather than a damage map.



Figure 8: **Failure cases.** inputs (a,c), outputs (b,d).

The SFT data rely on a fixed synthetic degradation library that cannot span the variation of real photographic media and aging, and all experiments use Qwen-Image-Edit-2511, so transfer to other editor architectures is untested. External comparisons pit released baselines against EMBER after adaptation on a disjoint 500-image real-photo pool, evaluating complete systems rather than architecture-matched domain adaptation. Evaluation scale is likewise modest: 74 RealOld photographs (72 for LIQE), 50 GGT-100K OldPhoto pairs, and a 48-photograph human study whose four-photograph strata remain descriptive; no dedicated face-identity evaluation was performed. Under severe damage or complex texture, EMBER may fail to recover fine structure or infer reliable color (Figure effig:failure-cases), so generated colors are plausible reconstructions that should be reviewed before archival use.

5 CONCLUSION

We presented EMBER, a mask-free system for real old-photo restoration. It combines early-trajectory modulation with no-reference policy optimization. EMBER has the highest mean on four of five held-out NR metrics on GGT-100K OldPhoto and achieves a 58.3% blinded Top-2 selection rate across 1,018 valid ballots. These perceptual gains accompany lower agreement with GGT-100K's provided generative references after GRPO. The early field is associated with later editing but is not a damage map. The measured color response does not establish historical color accuracy.

AI USE DISCLOSURE

Generative AI tools assisted literature retrieval, code drafting and debugging, experiment orchestration, figure preparation, and language editing. The authors specified the research questions and comparisons, supervised the experiments, and made all scientific interpretations and manuscript decisions. They reviewed the assisted code and text, reproduced reported statistics from saved artifacts, verified the references, and take responsibility for the final manuscript.

REFERENCES

Eirikur Agustsson and Radu Timofte. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1122–1131, 2017. doi: 10.1109/CVPRW.2017.150.

Omri Avrahami, Dani Lischinski, and Ohad Fried. Blended diffusion for text-driven editing of natural images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18208–18218, 2022.

Omri Avrahami, Ohad Fried, and Dani Lischinski. Blended latent diffusion. *ACM Transactions on Graphics*, 42(4), 2023.

Seung-Yeon Back, Geonho Son, Dahye Jeong, Eunil Park, and Simon S. Woo. Preserving old memories in vivid detail: Human-interactive photo restoration framework. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM)*, pp. 5180–5184, 2024. doi: 10.1145/3627673.3679215.

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, Humen Zhong, Yuanzhi Zhu, Mingkun Yang, Zhaohai Li, Jianqiang Wan, Pengfei Wang, Wei Ding, Zheren Fu, Yiheng Xu, Jiabo Ye, Xi Zhang, Tianbao Xie, Zesen Cheng, Hang Zhang, Zhibo Yang, Haiyang Xu, and Junyang Lin. Qwen2.5-VL technical report, 2025. URL https://arxiv.org/abs/2502.13923.

Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. In *The Twelfth International Conference on Learning Representations*, 2024.

Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6228–6237, 2018. doi: 10.1109/CVPR.2018.00652.

Weiwei Cai, Huaidong Zhang, Xuemiao Xu, Chenshu Xu, Kun Zhang, and Shengfeng He. Delving into important samples of semi-supervised old photo restoration: A new dataset and method. *IEEE Transactions on Multimedia*, 26:9866–9879, 2024. doi: 10.1109/TMM.2024.3400695.

Chaofeng Chen, Jiadi Mo, Jingwen Hou, Haoning Wu, Liang Liao, Wenxiu Sun, Qiong Yan, and Weisi Lin. TOPIQ: A top-down approach from semantics to distortions for image quality assessment. *IEEE Transactions on Image Processing*, 33:2404–2418, 2024. doi: 10.1109/TIP.2024.3378466.

Guillaume Couairon, Jakob Verbeek, Holger Schwenk, and Matthieu Cord. DiffEdit: Diffusion-based semantic image editing with mask guidance. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=3lge0p5o-M-.

Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. DPOK: Reinforcement learning for fine-tuning text-to-image diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, pp. 79858–79885, 2023.

S. Alireza Golestaneh, Saba Dadsetan, and Kris M. Kitani. No-reference image quality assessment via transformers, relative ranking, and self-consistency. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3989–3999, 2022. doi: 10.1109/WACV51458.2022.00404.

David Hasler and Sabine E. Suesstrunk. Measuring colorfulness in natural images. In *Human Vision and Electronic Imaging VIII*, volume 5007 of *SPIE Proceedings*. SPIE, 2003. doi: 10.1117/12.477378.

Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Prompt-to-prompt image editing with cross attention control. In *International Conference on Learning Representations*, 2023.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4396–4405, 2019. doi: 10.1109/CVPR.2019.00453.

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.

Xiangtao Kong, Jixin Zhao, Lingchen Sun, Rongyuan Wu, and Lei Zhang. GGT-100K: Generative ground truth for generalizable real-world image restoration, 2026. URL https://arxiv.org/abs/2605.31039.

Xiaoming Li, Chaofeng Chen, Shangchen Zhou, Xianhui Lin, Wangmeng Zuo, and Lei Zhang. Blind face restoration via deep multi-scale component dictionaries. In *Computer Vision – ECCV 2020*, pp. 399–415, 2020.

Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Yu Qiao, Wanli Ouyang, and Chao Dong. DiffBIR: Toward blind image restoration with generative diffusion prior. In Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol (eds.), *Computer Vision – ECCV 2024*, pp. 430–448, Cham, 2025a. Springer Nature Switzerland. ISBN 978-3-031-73202-7.

Xinqi Lin, Fanghua Yu, Jinfan Hu, Zhiyuan You, Wu Shi, Jimmy S. Ren, Jinjin Gu, and Chao Dong. Harnessing diffusion-yielded score priors for image restoration. *ACM Transactions on Graphics*, 44(6):1–21, 2025b. doi: 10.1145/3763346.

Hanyuan Liu, Chengze Li, Minshan Xie, Zhenni Wang, Jiawen Liang, Chi-Sing Leung, and Tien-Tsin Wong. BlueNeg: A 35mm negative film dataset for restoring channel-heterogeneous deterioration. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 13119–13128, 2025a. doi: 10.1109/ICCV51701.2025.01219.

Jie Liu, Gongye Liu, Jiajun Liang, Yangguang Li, Jiaheng Liu, Xintao Wang, Pengfei Wan, Di Zhang, and Wanli Ouyang. Flow-GRPO: Training flow matching models via online RL. In *Advances in Neural Information Processing Systems*, volume 38, pp. 45650–45685, 2025b. doi: 10.52202/085713-1362. URL https://openreview.net/forum?id=oCBKGw5HNf.

Grace Luo, Lisa Dunlap, Dong Huk Park, Aleksander Holynski, and Trevor Darrell. Diffusion hyperfeatures: searching through time and space for semantic correspondence. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA, 2023. Curran Associates Inc.

Weijia Mao, Hao Chen, Zhenheng Yang, and Mike Zheng Shou. The image as its own reward: Reinforcement learning with adversarial reward for image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5999–6009, 2026. Also available as arXiv:2511.20256.

Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a completely blind image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2013. doi: 10.1109/LSP.2012.2227726.

Maya Okawa, Ekdeep Singh Lubana, Robert P. Dick, and Hidenori Tanaka. Compositional abilities emerge multiplicatively: Exploring diffusion models on a synthetic task. In *Advances in Neural Information Processing Systems*, volume 36, pp. 50173–50195, 2023.

OpenAI. GPT Image 2 model. https://platform.openai.com/docs/models/gpt-image-2, 2025. Accessed 16 September 2026.

Qwen Team. Qwen-Image-Edit-2511. Official model card, 2025. URL https://huggingface.co/Qwen/Qwen-Image-Edit-2511. Accessed 23 September 2026.

Etai Sella, Yoav Baron, Hadar Averbuch-Elor, Daniel Cohen-Or, and Or Patashnik. LooseRoPE: Content-aware attention manipulation for semantic harmonization. In *ACM SIGGRAPH 2026 Conference Papers*, pp. 1–11, 2026. doi: 10.1145/3799902.3811142. Also available as arXiv:2601.05127.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Yichun Shi, Peng Wang, and Weilin Huang. SeedEdit: Align image re-generation to image editing, 2024. URL https://arxiv.org/abs/2411.06686.

Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. RoFormer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, 2024. doi: 10.1016/j.neucom.2023.127063.

Luming Tang, Menglin Jia, Qianqian Wang, Cheng Perng Phoo, and Bharath Hariharan. Emergent correspondence from image diffusion. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=ypOiXjdfnU.

Narek Tumanyan, Michal Geyer, Shai Bagon, and Tali Dekel. Plug-and-play diffusion features for text-driven image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1921–1930, 2023.

Ziyu Wan, Bo Zhang, Dongdong Chen, Pan Zhang, Dong Chen, Fang Wen, and Jing Liao. Old photo restoration via deep latent space translation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):2071–2087, 2023. doi: 10.1109/TPAMI.2022.3163183.

Jianyi Wang, Kelvin C. K. Chan, and Chen Change Loy. Exploring CLIP for assessing the look and feel of images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 2555–2563, 2023.

Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin C. K. Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *International Journal of Computer Vision*, 132(12):5929–5949, 2024.

Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. Towards real-world blind face restoration with generative facial prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9168–9178, 2021.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. doi: 10.1109/TIP.2003.819861.

Chenfei Wu, Jiahao Li, Jingren Zhou, Junyang Lin, Kaiyuan Gao, Kun Yan, Sheng ming Yin, Shuai Bai, Xiao Xu, Yilei Chen, Yuxiang Chen, Zecheng Tang, Zekai Zhang, Zhengyi Wang, An Yang, Bowen Yu, Chen Cheng, Dayiheng Liu, Deqing Li, Hang Zhang, Hao Meng, Hu Wei, Jingyuan Ni, Kai Chen, Kuan Cao, Liang Peng, Lin Qu, Minggang Wu, Peng Wang, Shuting Yu, Tingkun Wen, Wensen Feng, Xiaoxiao Xu, Yi Wang, Yichang Zhang, Yongqiang Zhu, Yujia Wu, Yuxuan Cai, and Zenan Liu. Qwen-Image technical report, 2025. URL https://arxiv.org/abs/2508.02324.

Runsheng Xu, Zhengzhong Tu, Yuanqi Du, Xiaoyu Dong, Jinlong Li, Zibo Meng, Jiaqi Ma, Alan Bovik, and Hongkai Yu. Pik-Fix: Restoring and colorizing old photos. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1724–1734. IEEE, January 2023. doi: 10.1109/wacv56688.2023.00177. URL http://dx.doi.org/10.1109/wacv56688.2023.00177.

Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. MANIQA: Multi-dimension attention network for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1190–1199, 2022. doi: 10.1109/CVPRW56347.2022.00126.

Yufeng Yang, Xianfang Zeng, Zhangqi Jiang, Fukun Yin, Jianzhuang Liu, Wei Cheng, Jinghong Lan, Shiyu Liu, Yuqi Peng, Gang Yu, and Shifeng Chen. RealRestorer: Towards generalizable real-world image restoration with large-scale image editing models, 2026. URL https://arxiv.org/abs/2603.25502.

Zhiyuan You, Xin Cai, Jinjin Gu, Tianfan Xue, and Chao Dong. Teaching large language models to regress accurate image quality scores using score distribution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14483–14494, 2025. doi: 10.1109/CVPR52734.2025.01350.

Fanghua Yu, Jinjin Gu, Zheyuan Li, Jinfan Hu, Xiangtao Kong, Xintao Wang, Jingwen He, Yu Qiao, and Chao Dong. Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25669–25680, 2024.

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. doi: 10.1109/CVPR.2018.00068.

Weixia Zhang, Guangtao Zhai, Ying Wei, Xiaokang Yang, and Kede Ma. Blind image quality assessment via vision-language correspondence: A multitask learning perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14071–14081, 2023. doi: 10.1109/CVPR52729.2023.01352.

A Related Work

This appendix positions EMBER against prior restoration, spatial-control, and policy-optimization methods.

Old photo and generative restoration. BOPB and Pik-Fix use paired or synthetically degraded supervision for old-photo restoration Wan et al. (2023); Xu et al. (2023). Face-specialized methods focus on portrait faces rather than complete scenes Li et al. (2020); Wang et al. (2021). Diffusion restorers such as DiffBIR, StableSR, and SUPIR use pretrained image priors with task-specific restoration conditioning Lin et al. (2025a); Wang et al. (2024); Yu et al. (2024). EMBER derives its spatial signal from the sampling trajectory and adds a separate unpaired real-photo policy-optimization stage.

Spatial control in image editing. Some controllable editors use masks or explicit spatial regions Avrahami et al. (2022; 2023). Prompt-to-Prompt manipulates cross-attention maps, whereas Plug-and-Play injects source features and attention during sampling Hertz et al. (2023); Tumanyan et al. (2023). DiffEdit contrasts denoiser predictions under source and target text conditions to estimate an editing mask Couairon et al. (2023). G12 instead compares early clean-latent estimates with the encoded condition under one restoration instruction. It combines two observations as a continuous field and changes target-token RoPE rather than imposing a binary pixel mask. Interactive old-photo restoration can also use user guidance Back et al. (2024). Like LooseRoPE, G12 modifies rotary phases Sella et al. (2026). Its control signal and schedule are specific to restoration: a per-image 12×12 field is computed online, applied only to target-token frame-axis Q/K during later steps, and replayed during GRPO. These operational differences do not by themselves establish superiority over either method.

Reinforcement learning for generative models. Policy-gradient fine-tuning can align diffusion and flow models with human preference or task objectives Black et al. (2024); Fan et al. (2023). GRPO computes advantages within each sample group and avoids a learned value function Shao et al. (2024). MUSIQ Ke et al. (2021) and CLIP-IQA Wang et al. (2023) provide no-reference quality signals that can be used as proxy rewards. Learned image rewards can favor assessor-specific responses Mao et al. (2026). EMBER combines both assessors with a frozen SFT-anchor reference penalty, then evaluates metrics not used as rewards and blinded human preference.

Latent capabilities and reward amplification. Generative models can encode compositional structure and spatial correspondence without explicit supervision for those downstream properties Okawa et al. (2023); Tang et al. (2023). These studies motivate, but do not establish, the mechanism in EMBER. In our setting, no-reference optimization amplifies a prompt-supported color response without a color-specific reward term. We report this as plausible colorization rather than recovery of historical color.

B Implementation-exact group-relative optimization

For each input group g of $K = 8$ candidates, the two reward models are normalized separately using the population mean and standard deviation across that group. Let q_{ij} be reward $j \in \{1, 2\}$ for candidate i, let $g(i)$ denote that candidate's input group, and let $\bar{q}_{j,g(i)}$ and $s_{j,g(i)}$ denote the corresponding group statistics. The implemented per-reward group-relative advantage is

$$z_{ij} = \frac{q_{ij} - \bar{q}_{j,g(i)}}{\max(s_{j,g(i)}, 10^{-6})}, \qquad a_i = \tfrac{1}{2}(z_{i1} + z_{i2}), \qquad A_i = \frac{a_i - \bar{a}_{\mathcal{B}}}{\max(s_{\mathcal{B}}, 10^{-6})}, \tag{8}$$

where $\mathcal{B}$ is the full rollout batch (eight inputs $\times$ eight candidates in the formal run). Thus the two assessors contribute equally after within-group normalization, regardless of their raw numerical scale. Advantages are clipped to $[-5, 5]$ before the policy loss.

For each rollout, Flow-SDE selects one transition index from $\{1, 2, 3\}$. Let $\ell_{\theta,i,t}$ and $\ell_{\mathrm{old},i,t}$ be the mean per-latent-dimension Gaussian transition log-probabilities under the current and rollout policies. The scheduler averages each Gaussian log density over latent elements before the trainer forms the ratio. The implemented per-dimension likelihood-ratio surrogate and clipped loss are

$$\begin{aligned} r_{i,t} &= \exp(\ell_{\theta,i,t} - \ell_{\mathrm{old},i,t}), \\ \mathcal{L}_{\mathrm{GRPO}} &= \mathbb{E}_{i,t}[\max(-A_i r_{i,t}, -A_i \,\mathrm{clip}(r_{i,t}, 0.8, 1.2))]. \end{aligned} \tag{9}$$

Table 3: **Hardware and software used for the formal GRPO run.**

Component	Recorded setting
Accelerator	One PPU 810e
Device memory	96 GiB, measured from the runtime device total
Host allocation	32 CPU cores, 300,000 MB memory, 500,000 MB local disk
Python and PyTorch	Python 3.10.13; PyTorch 2.6.0+ppu1.5.1
PyTorch-reported CUDA compatibility version	12.6
Training stack	DeepSpeed 0.14.5+eb07d41f with ZeRO-2; Accelerate 1.11.0
Precision	BF16 trainable and frozen weights; FP32 optimizer master weights

Table 4: **SFT and formal real-photo GRPO settings.**

Setting	Synthetic SFT	Real-photo GRPO
Base model	Qwen-Image-Edit-2511	Selected G12 SFT update 600
Trainable weights	LoRA rank/alpha 128 on Q/K/V/output	Same LoRA parameterization
Training data	7,000 FFHQ + 900 DIV2K + 100 prepared secondary-pool targets, each with 2–5 sampled degradations	500 unpaired real photographs
Updates	3,000	20
Sampling per update	8-image flow-matching batch	8 inputs, 8 candidates per input
Batching	Per-process batch 1; 8 processes; accumulation 1; effective global batch 8	Per-device batch 1; accumulation 64; effective batch 64 candidate transitions
Optimizer	AdamW; base learning rate 10^{-4}; first five updates at one-third scale	AdamW, learning rate 10^{-4}, betas $(0.9, 0.999)$
Regularization	Weight decay 10^{-2}; frozen non-LoRA weights	Weight decay 10^{-4}, clipping 1.0, reference weight 0.01
Objective	Flow matching	Equal-weight MUSIQ and CLIP-IQA after normalization
Sampling	G12 branch: detached two-step preview; both branches: flow-matching update	8-step rollout, guidance 1.0, Flow-SDE noise 1.0
G12	Non-G12 branch: disabled; G12 branch: $\kappa = 8$ on the training forward	Full rollout with stored-field and κ replay
Checkpointing	Save every 100 updates; evaluate every 300; select update 600	Save and evaluate every update; release update 20

Because ℓ is an average rather than a sum, $r_{i,t}$ is the geometric mean of the elementwise likelihood ratios. It is not the joint transition importance ratio. With the named frozen SFT anchor, let $\mu_{\theta,i,t}$ and $\mu_{0,i,t}$ be the policy and anchor next-latent means, and let d be the number of latent elements. The implementation's "x-based reference penalty" is a next-latent mean displacement:

$$\mathcal{L}_{\text{ref}} = 0.01\ \mathbb{E}_{i,t}\left[\frac{1}{d}\left\|\mu_{\theta,i,t} - \mu_{0,i,t}\right\|_2^2\right], \qquad \mathcal{L} = \mathcal{L}_{\text{GRPO}} + \mathcal{L}_{\text{ref}}. \tag{10}$$

Equation (9) uses a per-dimension ratio-surrogate clipping operator, but the training algorithm is GRPO: A_i is computed from same-input candidate groups and no value network is trained. Ratio clipping is an operator inside the GRPO loss, not the name of a separate training stage. We use one inner optimization epoch and one accumulated optimizer update per rollout batch. Consequently, the recorded ratio was exactly 1.0 and the clipping fraction was zero at all 20 formal updates. This is expected for one on-policy inner epoch. It does not imply a zero policy gradient.

C Hardware, software, and training configuration

Table 3 records the execution environment. Table 4 records the SFT and GRPO hyperparameters.

Base model and adapters. All reported EMBER variants start from Qwen-Image-Edit-2511. We use BF16 LoRA adapters with rank and alpha both 128 on the MMDiT attention projections `to_q`, `to_k`, `to_v`, and `to_out.0`. The text pathway and non-adapter MMDiT weights remain frozen. The formal G12 release contains 480 changed LoRA tensors after checkpoint 0. The release manifest records the adapter SHA-256, the Flow implementation hashes, and the exact base-weight URI.

Synthetic SFT. The two SFT branches are trained for three epochs and 3,000 updates from the same Qwen-Image-Edit-2511 initialization. Each epoch contains 8,000 synthetic pairs. The cache builder requests 10,000 images with an FFHQ ratio of 0.7. It samples 7,000 FFHQ images (Karras et al., 2019) and requests 3,000 images from the prepared secondary-image prefix. Because only 1,000 image files are available there, the sampler retains all of them: 800 DIV2K training-HR images, 100 DIV2K validation-HR images (Agustsson & Timofte, 2017), and 100 additional images from the prepared secondary-image pool. These 8,000 source images are resized to 512×512 and serve as pre-degradation SFT targets. The generator then samples 2–5 distinct operations from scratch, stain, tear, fading, blur, grain, JPEG compression, and paper texture and applies them in that fixed canonical order. Its archived ranges include 2–10 scratches, 1–5 stains, 1–2 tears, contrast $[0.5, 0.9]$, saturation $[0.4, 0.9]$, blur $\sigma \in [0.5, 3.0]$, JPEG quality $[30, 70]$, and paper strength $[0.05, 0.25]$. The launcher uses LoRA rank 128 and eight BF16 processes for the full run. Each process receives one image, with no gradient accumulation, giving an effective global batch of eight and 1,000 updates per epoch. The archived DiffSynth runner constructs AdamW with a base learning rate of 10^{-4} and weight decay 10^{-2}. Its default ConstantLR scheduler applies a one-third factor for the first five optimizer updates and then restores the base rate. The learning rate is not scaled with process count. We save every 100 updates. We evaluate the aspect-preserving 512^2-pixel target-area geometry every 300 updates on RealOld-74.

Formal real-photo policy optimization. The formal release uses one PPU 810e with DeepSpeed ZeRO-2. Each of its 20 updates draws eight unique images from a pool of 500 and samples $K = 8$ candidates per image. Per-device batch size is one, gradient accumulation is 64, and there is one inner epoch and one optimizer update per outer update. AdamW uses learning rate 10^{-4}, betas $(0.9, 0.999)$, weight decay 10^{-4}, $\epsilon = 10^{-8}$, and gradient clipping at 1.0. The two rewards are MUSIQ and CLIP-IQA. Each receives weight 0.5 after per-reward group normalization. The Flow-SDE configuration uses noise level 1.0 and one transition selected from indices $\{1, 2, 3\}$. The named SFT anchor is the reference in Eq. (10).

Geometry and inference. The formal GRPO run uses separate semantic and VAE conditioning scales. The prompt-image semantic path uses an aspect-preserving 384^2-pixel area budget, while the target and VAE-condition paths use an aspect-preserving 512^2-pixel area budget. These are area budgets rather than fixed square resolutions. Native-size evaluation expands the target and VAE-condition paths to the largest source-aligned multiple-of-16 grid while retaining the 384^2-pixel semantic budget. Separate scales are also present in the upstream Qwen image-editing pipeline (Wu et al., 2025; Qwen Team, 2025), so we do not describe this configuration as a registration error or claim that it is generally optimal. Inference uses eight solver steps, guidance scale 1.0, seed 42, and no mask input. The raw decoded grid is retained, followed by at most a 15-pixel-per-axis Lanczos correction to return the exact source dimensions. Appendix E reproduces the fixed instruction.

D G12 IMPLEMENTATION AND PARITY

The complete G12 configuration uses a 12×12 field, temporal reliability coefficient $\lambda = 2$, $Q_{0.9}$ normalization, and a frozen field after the second observation. To make the implementation unambiguous, we detail the calculation. The condition latent is detached, converted to float, bilinearly resized to the clean-estimate spatial shape if necessary, and truncated to the shared channel count. For step t, set

$$R_t = \mathrm{median}_c\, |\hat{z}_{0,c}^{(t)} - z_{y,c}|, \tag{11}$$

$$P_t = \mathrm{AAP}_{12\times12}(R_t), \tag{12}$$

$$L_t = [P_t - \mathrm{AvgPool}_{3\times3}(P_t)]_+, \tag{13}$$

$$\widetilde{L}_t = \mathrm{clip}\big(L_t / \max(Q_{0.9}(L_t), 10^{-6}), 0, 1\big)\,, \tag{14}$$

$$g_t = \mathrm{clip}\big(\mathrm{mean}(P_t) / \max(\mathrm{mean}\,|z_y|, 10^{-6}), 0, 1\big)\,, \tag{15}$$

$$D_t = \mathrm{clip}(\widetilde{L}_t + 0.25 g_t, 0, 1). \tag{16}$$

For the two observations, $D_{\mathrm{frozen}} = \mathrm{clip}(\frac{1}{2}(D_1 + D_2) \exp[-2(\max(D_1, D_2) - \min(D_1, D_2))], 0, 1)$. Here AAP denotes adaptive average pooling. The 12×12 field is nearest-neighbor expanded to

each target image token block of shape $(1, H, W)$. For a target position r, the frame-axis RoPE frequency pairs are replaced using $p_f(r) = -\lfloor \kappa D_{\text{frozen}}(r) \rfloor$; the height/width frequency pairs and all condition-token blocks retain their original values. We use $\kappa = 8$ on steps 3–7 and $\kappa = 4$ on step 8. During policy optimization, the rollout-recorded field, active flag, and κ are selected by the matching stored timestep and replayed during the optimization forward pass. During SFT, field estimation is a detached pure-noise two-step preview that uses no clean-target values. The resulting field is applied with $\kappa = 8$ to the single random-timestep flow-matching forward, preventing clean-target leakage without running a full training rollout.

E RESTORATION INSTRUCTIONS

The formal real-photo GRPO run, its monitoring, and benchmark inference use the following instruction:

> Restore this old damaged photograph conservatively. Remove scratches, stains, tears, and other physical damage. Recover faded colors and tonal balance toward the photograph's natural original appearance when color evidence is present, while preserving identity, composition, geometry, scene content, lighting, and authentic film texture. Introduce natural-looking colors where appropriate. Do not invent colors or objects, oversaturate, over-brighten, or alter undamaged regions.

Formal G12 SFT used the same text except that "Introduce natural-looking colors where appropriate" was replaced by "Keep genuinely monochrome photographs monochrome." We report this one-sentence difference for reproducibility.

E.1 TRAINING–INFERENCE IMPLEMENTATION AGREEMENT CHECK

With the checkpoint, input, prompt, initial latent, solver schedule, and G12 configuration fixed, the DiffSynth and Flow-Factory implementations produced output SSIM 0.995937 on one image. This documents close numerical agreement for that case; it neither establishes equivalence across inputs nor measures restoration quality. The SFT preview uses the first two locations of the deployed eight-step schedule. A four-step sampler is faster but changes the trained schedule; 12 and 20 steps add latency without improving the reported endpoints.

F TRAINING AND INFERENCE PSEUDOCODE

G EXTENDED TRAINING DYNAMICS

G.1 SFT SELECTION DIAGNOSTICS

Figure 9 complements the main anchor-selection curves with input–output SSIM and adjacent-checkpoint LPIPS. All ten measured checkpoints from each branch are retained. The comparatively large G12 drift between updates 600 and 900 is consistent with the simultaneous drop in both monitored quality scores and supports freezing the earlier operating point.

G.2 THREE-SEED GRPO OPTIMIZATION TRAJECTORIES

Figure 10 compares matched G12 and non-G12 runs for seeds 42, 43, and 44. Both branches use the same sampled inputs, candidate count, update schedule, and GRPO hyperparameters. Reward changes are measured from each run's own checkpoint 0 on the same 20 monitor images. No checkpoint or seed is excluded.

Both branches improved without a late reward collapse. At update 20, G12 gained 3.36 ± 0.50 MUSIQ and 0.034 ± 0.011 CLIP-IQA. The non-G12 branch gained 3.11 ± 0.99 and 0.030 ± 0.010, respectively. Values are mean $\pm$ seed standard deviation for $n = 3$ independent training seeds. The gain differences were small and did not have the same direction for every seed. These optimized rewards therefore support repeatable reward improvement, not an isolated G12 effect.

Algorithm 1 EMBER training and inference

Stage 1: supervised fine-tuning

Require: Synthetic pairs $\{(\mathbf{y}_s, \mathbf{x}_s)\}$ and LoRA parameters θ

for each $(\mathbf{y}_s, \mathbf{x}_s)$ **do**
  Run a detached two-step pure-noise preview conditioned on $\mathbf{y}_s$
  Estimate D_{frozen} from Eqs. (1)–(3)
  Sample a flow-matching timestep and update θ with G12 modulation
end for
Freeze the selected SFT checkpoint as π_0

Stage 2: no-reference policy optimization

Require: Real photographs $\{\mathbf{y}_r\}$, frozen anchor π_0

for each policy update **do**
  Draw 8 unique $\mathbf{y}_r$ from the 500-image pool
  for each $\mathbf{y}_r$ **do**
    Sample $K = 8$ candidates through the G12 rollout
    Compute MUSIQ and CLIP-IQA; form group-relative advantage A_i
  end for
  Select one Flow-SDE transition and update θ using Eqs. (9)–(10)
end for

Deployment

Require: Degraded photograph $\mathbf{y}$ and fixed instruction $\mathcal{T}$

Steps 1–2: denoise and estimate D_1, D_2
Freeze the temporally gated field D_{frozen}
Steps 3–8: denoise with target-token frame-axis RoPE modulation
Decode the final latent to the restored image $\mathbf{x}$

No external mask or quality-assessor call at deployment.

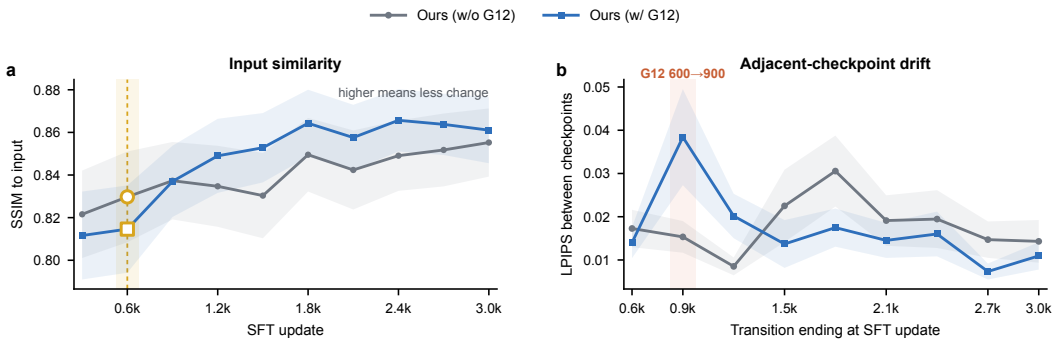


Figure 9: **Additional SFT checkpoint diagnostics. a,** Input–output SSIM for both SFT branches. Points are means over $n = 74$ RealOld-74 images and bands are pointwise 95% confidence intervals. **b,** LPIPS drift between adjacent 300-update checkpoints, computed on the same 74 images; bands again denote 95% confidence intervals. The shaded region marks the G12 600–900 transition.

Color exploration provides a complementary diagnostic. The G12 color-candidate rate increased in every seed, from 24.9% during updates 0–4 to 38.9% during updates 15–19. The non-G12 branch changed from 20.0% to 24.6%. A color candidate meets the fixed saturation and colorfulness thresholds defined in Appendix I. This rate quantifies exploration, not historical color correctness. The branches also start from separate SFT anchors, so their comparison remains end-to-end.

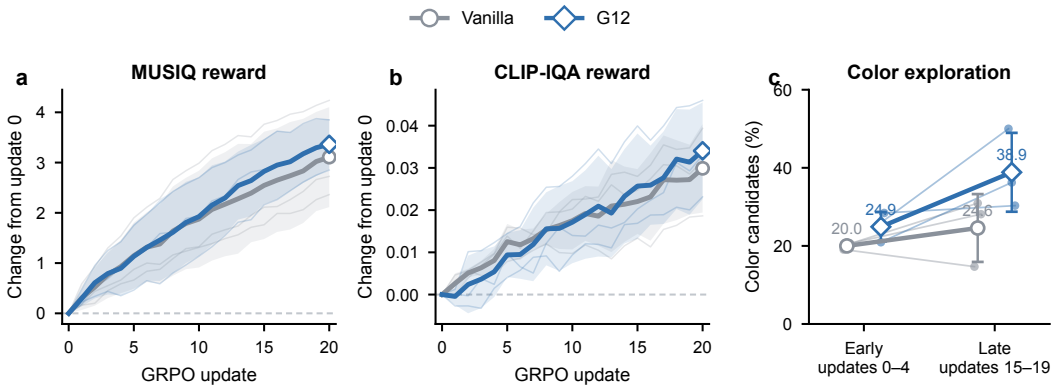


Figure 10: **Matched GRPO dynamics and color exploration. a,b,** Reward change from each run's checkpoint 0. Bold curves and bands show mean ± one seed standard deviation; faint curves show all three seeds. Each reward checkpoint uses the same 20 monitor images within a run. **c,** Rollout candidates with Δ saturation ≥ 0.05 and Δ colorfulness ≥ 5 during early and late updates. Small points are seeds; large markers show mean ± seed standard deviation ($n = 3$). The color measure does not assess historical correctness.

Table 5: **Formal reward-component ablation.** Endpoint changes are measured on the same 20 monitor images.

Reward	β	ΔMUSIQ	ΔCLIP-IQA	ΔSaturation	ΔColorfulness	$\Pr(A > 0 \mid \text{color})$
Dual-IQA	0.01	+3.850	+0.0460	+0.0390	+2.94	0.535
MUSIQ-only	0.01	+3.775	–	+0.0199	+1.73	0.540
CLIP-IQA-only	0.01	–	+0.0364	+0.0333	+2.46	0.511
Dual-IQA, no reference penalty	0	+4.150	+0.0486	+0.0419	+3.41	0.538

H Reward-choice controls and color-sensitivity audit

We first isolate the formal objective components. All four arms start from the same G12 SFT anchor and use the release geometry, seed 42, the fixed 20-image monitor, and 20-update schedule. Table 5 reports changes from the shared update-0 output. A dash indicates an assessor that was inactive and therefore not logged in that arm. Color-candidate support is the fraction of rollout candidates meeting the fixed diagnostic color rule that received positive GRPO advantage.

Both single-assessor arms produced positive color changes. Their color-candidate support rates remained close to one half, so the online rankings do not show a strong direct preference for the fixed diagnostic color rule. The unregularized arm had the largest endpoint proxy gains, but this single 20-update run cannot establish long-horizon stability. We use $\beta = 0.01$ in the released model.

We next compare two matched, single-seed reward choices from an earlier exploratory protocol. Both arms start from the same G12 SFT anchor and use the semantic-512 probe geometry, $K = 4$, 100 updates, and reference weight $\beta = 0.05$. The fixed monitor contains 20 images. Table 6 reports changes from each arm's update-0 output. The dagger marks the arm-specific monitor peak selected on its optimized score or scores. DeQA-LCB is $\mu - 0.5\sigma$, where μ and σ are computed from DeQA's five-level quality-probability output.

Both objectives produced a transient monitor peak, but neither retained a gain at update 100. This result motivated evaluating a shorter release schedule. It does not show that dual-IQA is optimal. The formal controls in Table 5 separately examine the two assessors and the reference term under the release schedule.

We next performed a post hoc color-sensitivity audit on the formal seed-42 rollouts. The archive contains 1,280 candidates in 160 same-input groups, with 140 unique input identifiers. Candidates within a group are subsamples, not independent observations. For each group, we computed Spearman

Table 6: **Matched single-seed reward controls.** Changes are relative to update 0 on the same 20 images. These controls use an earlier protocol and are not the formal release trajectory.

Reward	Update	ΔMUSIQ	ΔCLIP-IQA	ΔDeQA-LCB	ΔSaturation	LPIPS to anchor
Dual-IQA	$1^{\dagger}$	+0.637	+0.0031	+0.0342	−0.0008	0.0050
Dual-IQA	100	−0.732	−0.0085	−0.0320	−0.0130	0.0086
DeQA-LCB	$26^{\dagger}$	+0.511	+0.0043	+0.0290	−0.0080	0.0086
DeQA-LCB	100	−0.207	−0.0026	−0.0238	−0.0032	0.0102

Table 7: **Within-group color sensitivity of the formal reward signals.** Entries are mean Spearman ρ with input-cluster bootstrap 95% intervals.

Candidate attribute	MUSIQ	CLIP-IQA	Combined GRPO advantage
Δ saturation	.030 [−.037,.097]	.070 [.002,.139]	.046 [−.022,.116]
Δ colorfulness	.026 [−.037,.089]	.011 [−.054,.075]	.006 [−.057,.068]

association across its eight candidates. Repeated groups were averaged within input, and 95% intervals were obtained by 10,000 bootstrap resamples of the 140 input identifiers with seed 20260918. Table 7 reports the result. The six intervals are exploratory and are not multiplicity-adjusted.

MUSIQ showed no clear association with either color measure. The only interval that narrowly excluded zero was CLIP-IQA versus saturation. We treat this as weak evidence because the audit contains six exploratory associations. CLIP-IQA showed no clear association with colorfulness, and the combined GRPO advantage showed no clear association with either measure. Using the archived color-candidate rule (Δ saturation ≥ 0.05 and Δ colorfulness ≥ 5, both relative to the degraded input), 272 of 508 candidates received positive advantage (53.5%). These observations do not prove color neutrality. They show that the formal objective did not strongly or consistently rank color change within the sampled candidate groups. Because candidates also differ in structure, contrast, and texture, this is an observational audit rather than a controlled color perturbation.

I Diagnostic metric definitions

Color response. For an RGB image I, mean saturation is the spatial mean of the saturation channel after conversion to hue–saturation–value (HSV), divided by 255. We also use the Hasler–Suesstrunk colorfulness statistic (Hasler & Suesstrunk, 2003):

$$\begin{aligned} r_g &= R - G, & y_b &= \tfrac{1}{2}(R+G) - B, \\ C(I) &= \sqrt{\sigma_{r_g}^2 + \sigma_{y_b}^2} + 0.3\sqrt{\mu_{r_g}^2 + \mu_{y_b}^2}. \end{aligned} \tag{17}$$

RGB channels use the 8-bit $[0, 255]$ scale. Endpoint color changes are paired within image against checkpoint 0. For the rollout color-candidate diagnostic, Δ saturation and Δ colorfulness are instead the candidate output minus its degraded input. These statistics quantify color response, not color correctness.

Early-response association. The confirmatory synthetic diagnostic uses a 12×12 grid on 100 images and generation seeds 42, 43, and 44. The field is estimated during steps 1–2, but downstream G12 injection is disabled during steps 3–8. For region m, final edit magnitude is

$$F_m = \mathrm{Pool}_m\left(\tfrac{1}{3}\sum_c |I_{\mathrm{out},c} - I_{\mathrm{in},c}|\right). \tag{18}$$

Spearman correlation is computed within each image and seed. The three seeds are then averaged within image. For partial Spearman correlation, we rank the variables within each image and seed, then linearly residualize the ranked early response and final edit against pooled input-gradient magnitude and local grayscale variance from a 9×9 window. The null averages five within-image permutations of the early field. Intervals use 10,000 image-level bootstrap resamples with seed 20260918. A separate 32-image closed-loop diagnostic supplies AUPRC and AUROC. A region is marked damaged when its pooled binary-mask coverage is nonzero; 28 images have valid non-degenerate masks.

Table 8: **Paired image-level effects of EMBER (G12).** Positive values favor EMBER. NIQE signs are reversed. Small type gives paired-bootstrap 95% CIs. RealOld LIQE uses $n = 72$; its other metrics use $n = 74$, and GGT-100K uses $n = 50$.

Set	Comparator	TOPIQ-NR	MANIQA	LIQE	TReS	NIQE
RealOld-74	HYPIR	.001 [−.027,.031]	.016 [−.008,.043]	.904 [.704,1.120]	7.524 [4.378,10.987]	−.478 [−.752,−.176]
RealOld-74	EMBER (non-G12)	.026 [.015,.037]	.010 [.0005,.020]	.329 [.231,.430]	3.140 [1.999,4.312]	.864 [.635,1.105]
GGT-100K	HYPIR	.054 [.034,.075]	.034 [.018,.051]	1.364 [1.124,1.604]	5.699 [3.340,8.105]	−.624 [−.775,−.469]
GGT-100K	EMBER (non-G12)	.048 [.031,.066]	.011 [−.003,.025]	.676 [.519,.834]	6.156 [4.430,7.946]	1.469 [1.224,1.728]

Damage repair and intact-region drift. Let Y, X, and $\widehat{X}$ denote the degraded input, clean target, and restored output on the $[0, 1]$ RGB scale. Let M be the binary damage mask. The damage-region L1 gain is

$$G_{\mathrm{L1}} = \mathrm{mean}_{p\in M}\,|Y_p - X_p| - \mathrm{mean}_{p\in M}\,|\widehat{X}_p - X_p|. \tag{19}$$

Damage-region LPIPS gain is defined analogously after replacing all pixels outside M with clean-target pixels. Intact-region L1 and LPIPS retain only M^c by the same compositing rule. Positive damage gains and lower intact errors are better. For each image and generation seed, five exact-histogram permutations match the correct field's values and differ only in spatial position. The five permutations are averaged within seed, and the three generation seeds are averaged within image. The independent unit is therefore an image. Damage-region analyses retain 83 valid masks, while intact-region analyses retain all 100 images. Paired intervals use 10,000 image-level bootstrap resamples with seed 20260918. Two-sided p values use 100,000 paired sign-flip draws. Holm correction is applied across the two co-primary LPIPS endpoints within each comparison. L1 endpoints are secondary and their p values are unadjusted.

J Complete benchmark statistics

The independent unit for every benchmark interval is an image. RealOld-74 uses $n = 74$ for TOPIQ-NR, MANIQA, TReS, and NIQE, and a common $n = 72$ subset for LIQE because images 00054 and 00055 have a short side below the implementation's 224-pixel requirement. The official GGT-100K OldPhoto test subset uses $n = 50$ for every metric. Means and 95% confidence intervals use 10,000 image-level bootstrap resamples with seed 20260916. Paired comparisons resample aligned image IDs and report the mean oriented effect, where positive always favors EMBER. Outputs exist for all benchmark images; only the two common RealOld LIQE inputs are omitted for that metric. The primary no-reference view resizes every output to the corresponding input resolution before scoring. This avoids rewarding or penalizing a method merely because it exports a different number of pixels. Table 8 reports the paired effects against HYPIR and the non-G12 EMBER trajectory on both evaluation sets.

Runtime audit. The environment manifest lists `opencv-python==4.5.5.64`. The imported benchmark module was `cv2==4.11.0` from `opencv-python-headless==4.11.0.86`. Both packages appear in the environment. The metrics used the verified 4.11 runtime.

K Early-response evidence and spatial interventions

Figure 11 summarizes edit prediction and the spatial intervention. Table 10 separates edit prediction from damage localization, and Table 11 reports the complete intervention estimates. An earlier single-seed sweep at $\kappa \in \{2, 4, 8, 12\}$ also produced LPIPS intervals that crossed zero at every strength. The confirmatory study therefore fixes $\kappa = 8$, adds three generation seeds and five permutation replicates, and treats images as the independent units.

The pre-treatment result indicates that the early field carries information about where the model will later edit, even without downstream field injection. The intervention does not detect an LPIPS benefit from exact placement. The negative secondary damage-L1 effect against permutation also argues against a simple pixel-error interpretation. These endpoints cover binary mask-defined regions

Table 9: **Matched G12 SFT-to-GRPO evaluation.** Both checkpoints use identical inference settings. Values in the last column are paired effects oriented so that positive favors GRPO, with pointwise 95% bootstrap CIs. These intervals are not adjusted for multiple comparisons. MUSIQ and CLIP-IQA are omitted because they are the optimized rewards. RealOld LIQE uses $n = 72$; its other metrics use $n = 74$.

Set	Metric	G12-SFT update 600	G12-GRPO update 20	Oriented GRPO effect [95% CI]
GGT-100K	TOPIQ-NR ↑	0.4675	0.6149	+0.1474 [0.1339, 0.1617]
	MANIQA ↑	0.3194	0.4285	+0.1092 [0.0984, 0.1201]
	LIQE ↑	2.8356	3.9086	+1.0730 [0.9073, 1.2455]
	TReS ↑	64.3838	79.3536	+14.9698 [13.3788, 16.6572]
	NIQE ↓	4.8939	4.3765	+0.5174 [0.3964, 0.6492]
RealOld-74	TOPIQ-NR ↑	0.4408	0.5836	+0.1428 [0.1323, 0.1538]
	MANIQA ↑	0.3243	0.4303	+0.1060 [0.0987, 0.1132]
	LIQE ↑	2.9511	3.9923	+1.0412 [0.9110, 1.1712]
	TReS ↑	62.8267	80.5250	+17.6983 [15.7861, 19.7473]
	NIQE ↓	5.0057	4.2666	+0.7390 [0.5991, 0.8797]
GGT-100K	PSNR ↑	19.9766	17.9390	-2.0376 [-2.3620, -1.7448]
	SSIM ↑	0.7662	0.6937	-0.0725 [-0.0820, -0.0637]
	LPIPS ↓	0.2486	0.2979	-0.0493 [-0.0596, -0.0394]

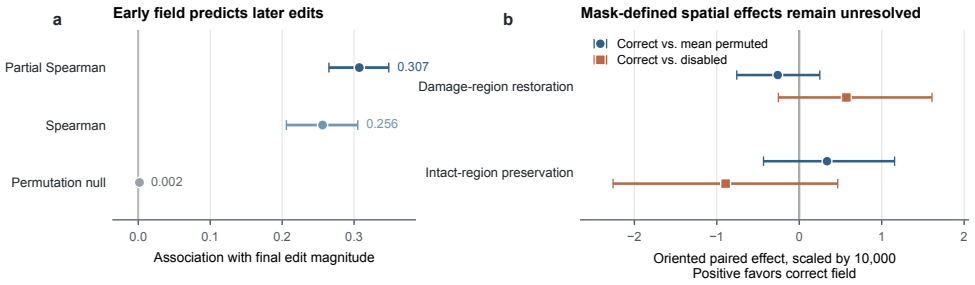


Figure 11: **Early-field prediction and spatial intervention. a**, Step-2 field association with final edits when later G12 injection is disabled ($n = 100$). **b**, Correct-field effects relative to permutation and disabled controls (damage $n = 83$; intact $n = 100$). Bars show 95% image-bootstrap CIs; positive values favor the correct field.

only. They do not measure global fading, tonal imbalance, color recovery, or effects acquired during G12-assisted training. The field is therefore treated as a continuous edit-response signal rather than a damage probability map or a replacement for an external mask.

L Conditioning Geometry and Color-Response Probes

We ran five short dual-IQA policy-optimization probes to characterize the observed color response. Each arm uses 20 updates, one run, and the same eight monitor images. Intervals use 10,000 paired image-bootstrap samples. These are mechanism probes, not independent benchmark replications. "Legacy" denotes the historical non-G12 anchor. Semantic 384 and semantic 512 denote aspect-preserving 384^2- and 512^2-pixel area budgets before Qwen2.5-VL encoding; the target and VAE-condition paths retain an aspect-preserving 512^2-pixel area budget. Archived job names used "mismatch" for the semantic-384 condition, but the two-path scale difference is not itself a registration error. "Condition $\times 0.5$" multiplies a non-mutating copy of the VAE condition latents by 0.5. A color candidate has Δ saturation ≥ 0.05 and Δ colorfulness ≥ 5, where both differences are measured from the degraded input to the rollout candidate. $\Pr(A > 0 \mid \text{color})$ is the fraction of these candidates with positive GRPO advantage. Table 12 reports the checkpoint-0-to-20 saturation change and the rate at which color candidates receive positive advantage.

The legacy semantic-384 arm produced the largest saturation increase. The legacy semantic-512 arm produced a smaller increase, while the G12 semantic-512 interval crossed zero. Visual inspection found line artifacts after halving condition strength, whereas G12 with semantic 384 retained a

Table 10: **Step-2 early-field diagnostics.** The closed-loop localization probe uses 28 valid masks from 32 images. The pre-treatment prediction probe uses 100 images and three generation seeds. Small type gives image-bootstrap 95% CIs.

Quantity	Estimate
Closed-loop mask localization	
Damage prevalence	.1689
AUPRC	.1794 [.1164,.2656]
AUROC	.5444 [.4667,.6181]
Pre-treatment edit prediction; downstream G12 disabled	
Early vs. final-edit Spearman ρ	.2562 [.2059,.3051]
Partial Spearman ρ	.3072 [.2650,.3482]
Mean permuted-field Spearman ρ	.0017 [−.0026,.0057]

Table 11: **Confirmatory intervention on G12 field placement.** Effects are oriented so that positive values favor the correct field and are scaled by 10^4. Damage metrics use 83 valid masks; intact metrics use all 100 images. LPIPS p values are Holm-adjusted within comparison. Secondary L1 p values are unadjusted.

Comparison	Endpoint	Effect	95% CI	p	n
Correct vs. mean permuted	Damage LPIPS gain	-0.261	$[-0.758, 0.250]$	.640	83
	Intact LPIPS preservation	0.337	$[-0.434, 1.156]$	.640	100
	Damage L1 gain	-3.128	$[-5.233, -1.229]$	.0021	83
	Intact L1 preservation	-0.025	$[-0.373, 0.312]$	.887	100
Correct vs. disabled	Damage LPIPS gain	0.572	$[-0.253, 1.610]$	.415	83
	Intact LPIPS preservation	-0.893	$[-2.259, 0.466]$	.415	100
	Damage L1 gain	2.682	$[-1.228, 7.321]$	.227	83
	Intact L1 preservation	-0.146	$[-0.674, 0.378]$	.586	100

positive response without that artifact. Color candidates received positive advantage in only 52–54% of cases across all arms. The geometry variants produced different measured trajectories in this single-run probe. The reward contains no explicit color statistic, but the probe does not establish that its learned assessors are color-neutral. It also cannot verify historical color correctness.

M SOLVER-STEP AND OBSERVATION-WINDOW ABLATIONS

This post-hoc inference sweep fixes one trained checkpoint and changes only the solver length or the field-observation window at inference. The rows therefore represent sampler settings, not separately trained models. Per-image latency was measured on one PPU 810e using 512×512 synthetic inputs and the same inference stack. Table 13 compares four solver lengths. Table 14 separates the number of early observations from frozen versus dynamic field reuse.

Eight steps is a fixed training and deployment contract, not a claim of global solver optimality. The 12-step and 20-step schedules increased latency by 46% and 139%, respectively. Neither improved a reported restoration endpoint. Both also reduced damage-L1 gain and increased intact-region LPIPS. Four steps was faster and competitive. However, it changes the sigma locations of the two field observations and leaves only two modulated transitions. A formal four-step configuration should undergo matched training and evaluation before being treated as a release setting rather than a post-hoc sampler replacement.

One observation cannot measure temporal agreement. Two are the minimum needed for the reliability gate. Three and four frozen observations changed the three quality endpoints by at most 0.00037 from the deployed setting. Dynamic two-step recomputation increased latency from 3.432 to 3.985 seconds without a clear quality gain. We freeze after step 2. This choice preserves the training contract and avoids unsupported claims of numerical optimality.

Table 12: **Conditioning-geometry probe on eight fixed images.** Saturation changes from checkpoint 0 to 20. Small type gives paired-bootstrap 95% CIs.

Anchor / geometry	Δ saturation	$\Pr(A > 0 \mid \text{color})$
Legacy / semantic 384	.01891 [.01026,.02866]	.533
Legacy / semantic 512	.00931 [.00464,.01549]	.519
G12 / semantic 512	−.00456 [−.01967,.00731]	.535
G12 / semantic 512, condition $\times 0.5$	.01275 [−.00489,.02951]	.524
G12 / semantic 384	.01319 [.00442,.02198]	.532

Table 13: **Solver-step sweep on 100 synthetic images.** Damage-region metrics use the 83 images with a non-empty damage region; intact-region metrics and latency use all 100 images. Four steps is competitive and fastest; 12 and 20 steps add cost without improving the reported endpoints.

Steps	Damage L1 gain ↑	Damage LPIPS gain ↑	Intact LPIPS ↓	Latency (s) ↓
4	.12877	.04523	.08651	1.844
8 (deployed)	.12708	.04577	.08845	3.432
12	.12556	.04548	.08979	5.018
20	.12396	.04510	.09144	8.192

N BLINDED HUMAN-EVALUATION PROTOCOL AND COMPLETE RESULTS

The evaluation used 48 photographs in 12 prespecified strata. The strata crossed person-centered content (absent or present), usable input color (absent or present), and degradation severity (light, moderate, or heavy). Each stratum contained four photographs. Each questionnaire page showed one degraded input and seven anonymous outputs. Output positions were randomized independently for each ballot. Figure 13 reproduces a representative deployed page.

Stratum assignments were fixed before survey deployment by visual inspection of the degraded inputs. "Person-centered" indicates that a person is a principal subject. "Usable input color" indicates visible chromatic content that can support a color comparison, rather than a grayscale or near-monochrome input. Severity is an ordinal light, moderate, or heavy assessment of the visible extent and intensity of physical and photometric degradation. These labels are sampling strata, not model inputs or independently validated annotations. The 48 photographs were selected to populate the balanced strata and are not a probability sample of either evaluation set.

Thirty-one participants were recruited through local in-person and online channels. They were informed that their anonymous choices would be used for research evaluation and consented before participating. No demographic variables were collected. Eighteen participants completed all 48 photographs, while 13 submitted partial sets. Each ballot selected exactly two outputs. The study collected 1,018 valid ballots and 2,036 selections, with 19 to 24 ballots per photograph (median 21). No ballot had a missing required value, a repeated method choice, or a duplicate record for the same participant and photograph. No valid ballot was excluded. The seven candidates were SeedEdit, RealRestorer, Qwen-Image-Edit-2511, GPT Image 2, HYPIR, EMBER, and BOPB.

The Top-2 rate is the proportion of valid ballots in which a method was selected. Ballots were not treated as independent observations. Overall 95% confidence intervals used 20,000 crossed bootstrap resamples of participants and photographs with seed 20260921. Pairwise method effects were estimated with a conditional logit model over the 21 possible unordered output pairs. A pair's utility is the sum of its two method effects and their randomized display-position effects. Sandwich standard errors were clustered by both participant and photograph, with finite-cluster corrections and subtraction of the ballot-level intersection term. Holm correction was applied to the six EMBER-versus-baseline contrasts. The global method effect was a two-way cluster-robust Wald test ($\chi^2(6) = 81.87$, $p = 1.47 \times 10^{-15}$). The archived custom Python script implements the likelihood, crossed bootstrap, covariance estimator, and Holm correction directly with NumPy and pandas.

Table 14: **Observation-window sweep at eight solver steps.** The two-observation frozen field is the deployed setting. Differences among one to four frozen observations are small; the dynamic two-step variant adds latency without a consistent restoration gain.

Field construction	Damage L1 gain ↑	Damage LPIPS gain ↑	Intact LPIPS ↓	Latency (s) ↓
1 observation, frozen	.12694	.04565	.08882	3.430
2 observations, frozen (deployed)	.12708	.04577	.08845	3.432
3 observations, frozen	.12695	.04575	.08840	3.433
4 observations, frozen	.12673	.04566	.08839	3.437
2 observations, dynamic	.12689	.04576	.08839	3.985

Table 15: **Overall blinded Top-2 preference.** Rate intervals use crossed participant-by-photograph bootstrap resampling. Odds ratios express each baseline relative to EMBER after adjustment for display position; values below one favor EMBER.

Method	Top-2 count	Rate [95% CI]	OR vs. EMBER [95% CI]	Holm p
EMBER	593	58.3 [48.6, 67.8]	1.00 (reference)	–
RealRestorer	437	42.9 [35.2, 50.8]	0.59 [0.41, 0.86]	0.006
Qwen-Image-Edit-2511	330	32.4 [24.3, 41.4]	0.42 [0.29, 0.61]	1.51×10^{-5}
GPT Image 2	264	25.9 [18.9, 33.8]	0.32 [0.19, 0.54]	4.69×10^{-5}
SeedEdit	153	15.0 [9.8, 20.5]	0.17 [0.10, 0.29]	1.54×10^{-10}
HYPIR	131	12.9 [8.2, 18.0]	0.15 [0.09, 0.25]	2.14×10^{-12}
BOPB	128	12.6 [7.7, 18.0]	0.14 [0.08, 0.25]	3.94×10^{-11}

Table 15 gives the complete overall results. EMBER received 593 selections and a 58.3% Top-2 rate (95% CI, 48.6 to 67.8%). RealRestorer ranked second at 42.9%. Their observed difference was 15.3 percentage points (95% CI, 2.8 to 28.4). The adjusted odds ratio for RealRestorer relative to EMBER was 0.59 (95% CI, 0.41 to 0.86; Holm-adjusted $p = 0.006$).

Figure 12 reports all 12 strata. EMBER ranked first in eight strata and in the top two in 11. It ranked first in all six strata without usable input color. RealRestorer was the strongest competitor when usable input color was present. These strata contain only four photographs each, so they are reported descriptively rather than as separate inferential tests.

Post hoc sensitivity checks did not change the overall ordering. After requiring response times of at least 3, 5, and 10 seconds, EMBER's rates were 59.6%, 59.6%, and 56.5%, respectively; the corresponding RealRestorer rates were 43.4%, 44.2%, and 46.6%. The primary analysis retains every valid ballot because no response-time exclusion was prespecified. Among the 18 participants who completed all 48 photographs, the descriptive rates were 58.8% for EMBER and 43.1% for RealRestorer.

O EVALUATION MATERIALS AND PROVENANCE

The self-assembled real-photo pool and RealOld-74 contain images gathered from internet sources and images originating from BlueNeg, *A 35mm Negative Film Dataset* Liu et al. (2025a). They are not presented as a single homogeneous capture process. We retain source category and split membership in the benchmark ledger. The 500-image GRPO pool, the 20-image monitor, RealOld-74, and the GGT-100K test subset are mutually disjoint. Baseline prompts, exact checkpoints, per-image metric records, and the bootstrap procedure are specified with the evaluation artifacts. GPT Image 2 is identified by the evaluated model label `gpt-image-2`; output hashes are retained in the artifact manifest.

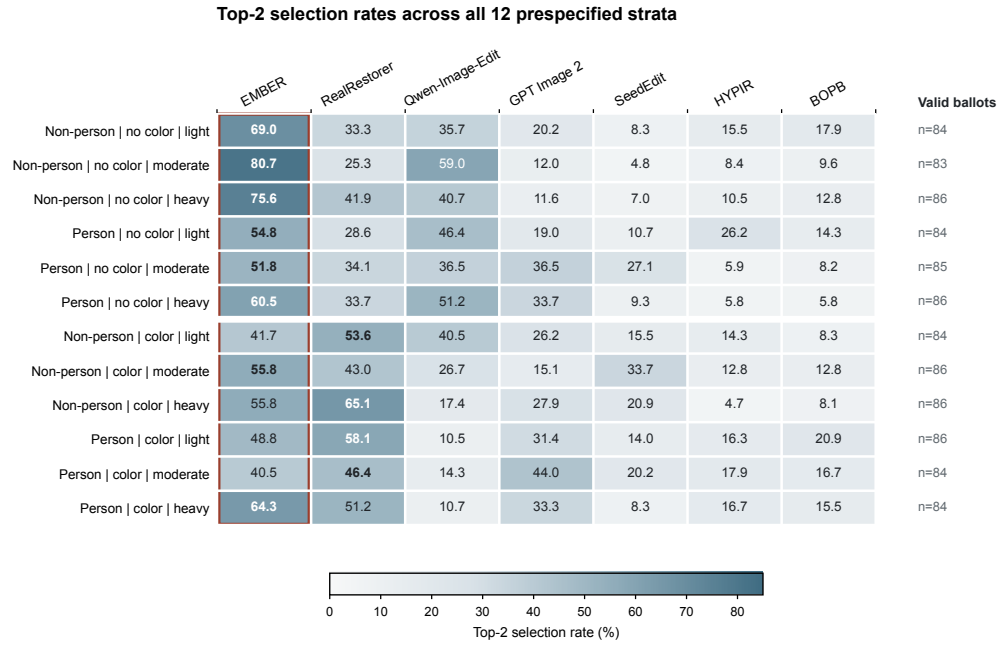


Figure 12: **Stratified blinded Top-2 preference.** Rates for all seven methods across the 12 prespecified strata. Bold values are the highest rate in each row; the red outline marks EMBER. Each ballot contributes two selections, so rates within a row sum to 200%. Valid ballot counts are shown at right.

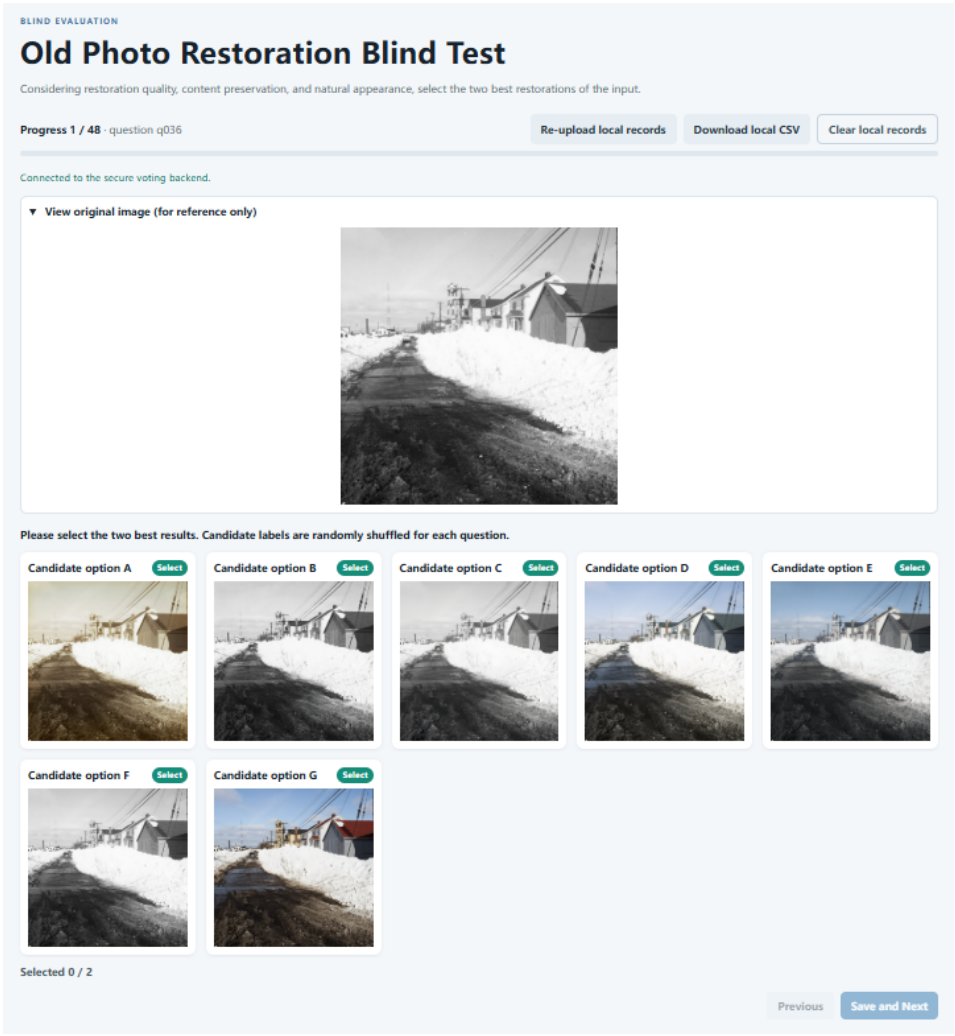
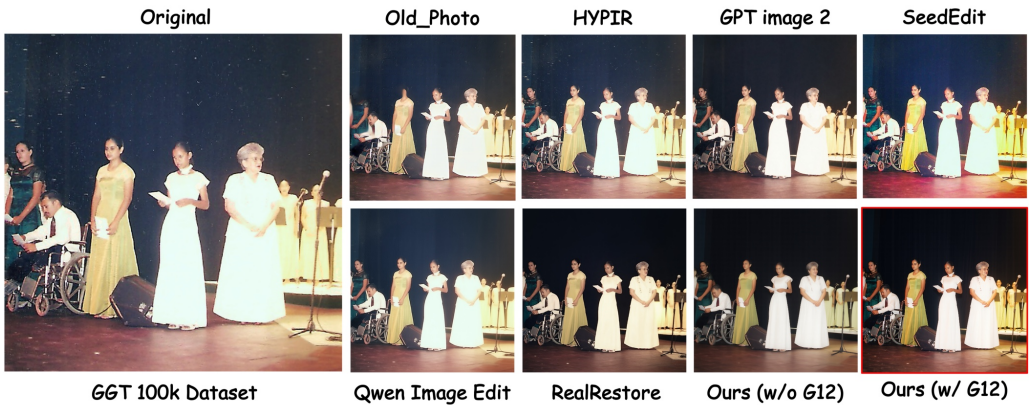
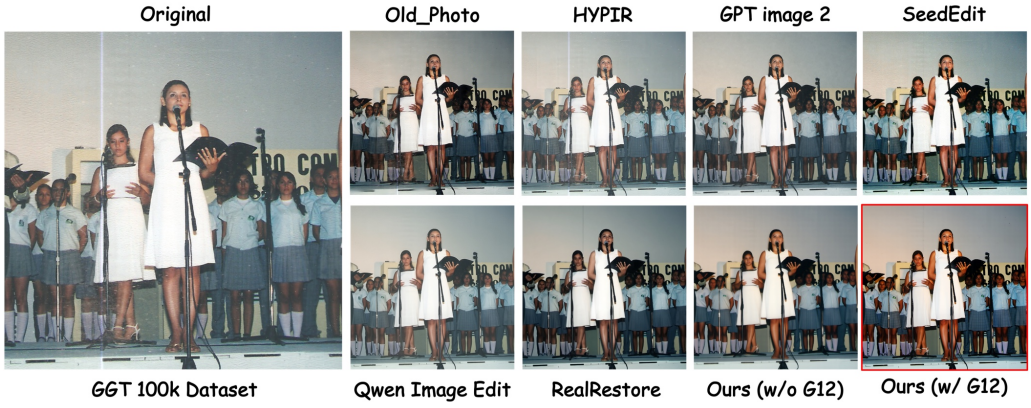


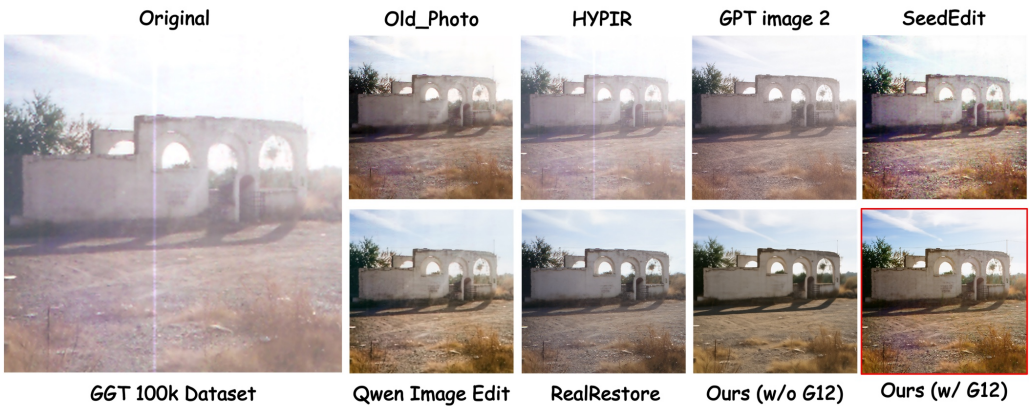
Figure 13: **Representative page of the deployed anonymous questionnaire.** Participants viewed the degraded input for reference and selected the two best results from seven anonymous candidates. Candidate labels were randomized for each ballot, and method identities were not displayed.

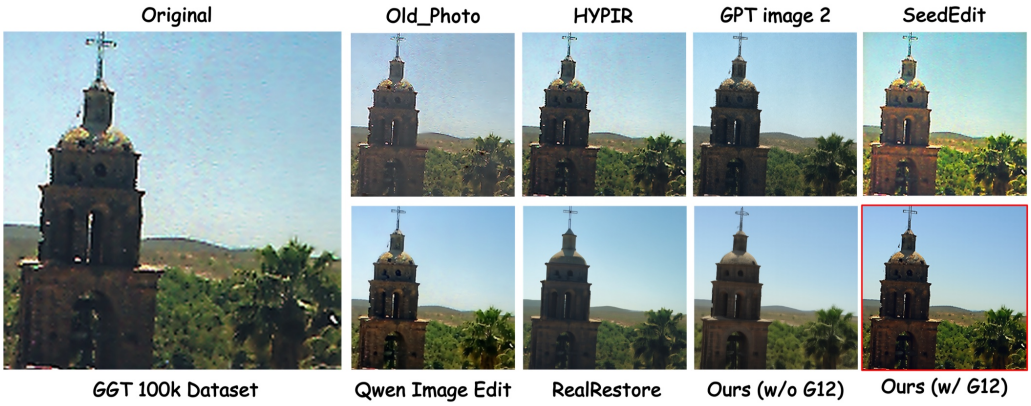
P ADDITIONAL VISUAL RESULTS



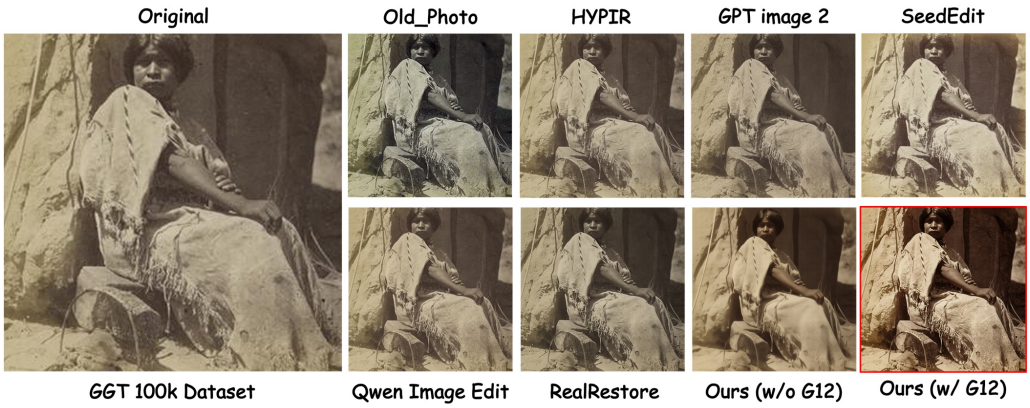


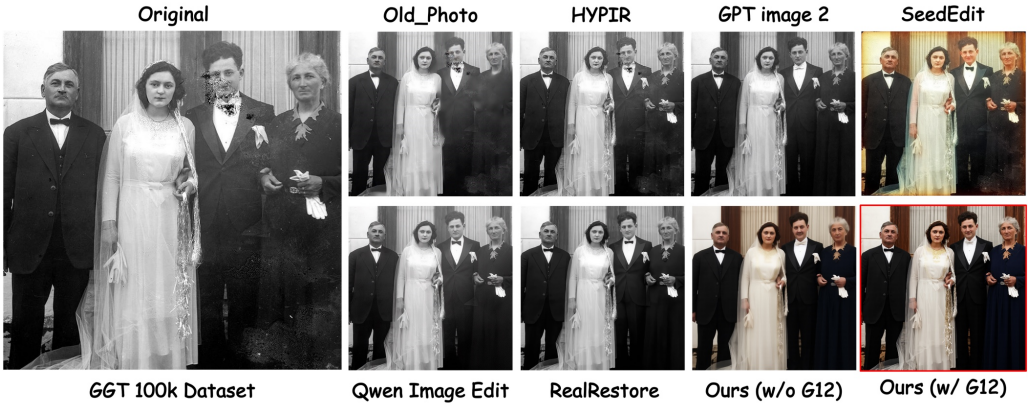


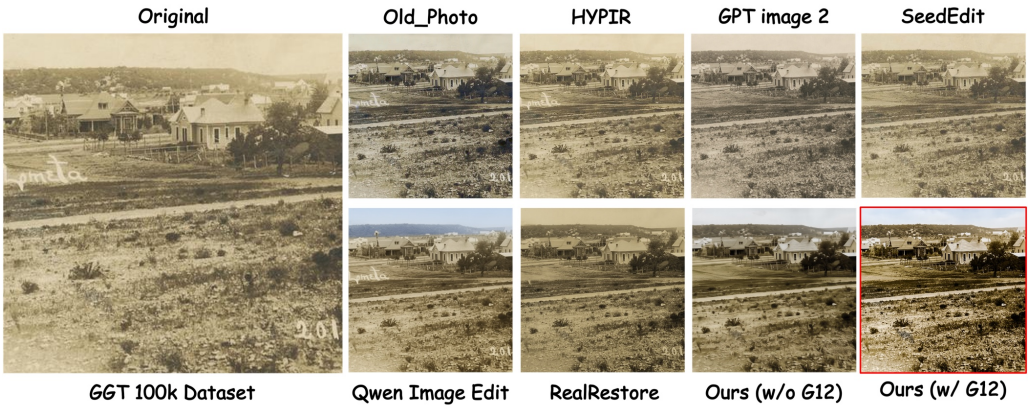

Original
Old_Photo
HYPIR
GPT image 2
SeedEdit
GGT 100k Dataset
Qwen Image Edit
RealRestore
Ours (w/o G12)
Ours (w/ G12)

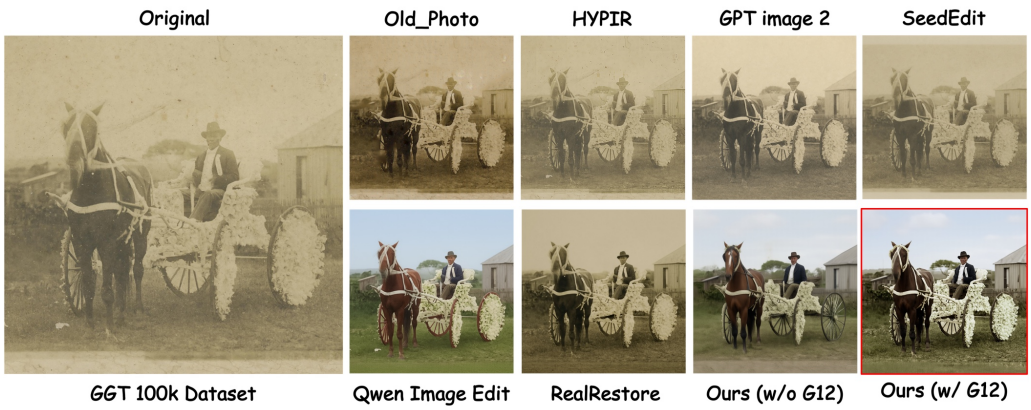

Original
Old_Photo
HYPIR
GPT image 2
SeedEdit
GGT 100k Dataset
Qwen Image Edit
RealRestore
Ours (w/o G12)
Ours (w/ G12)

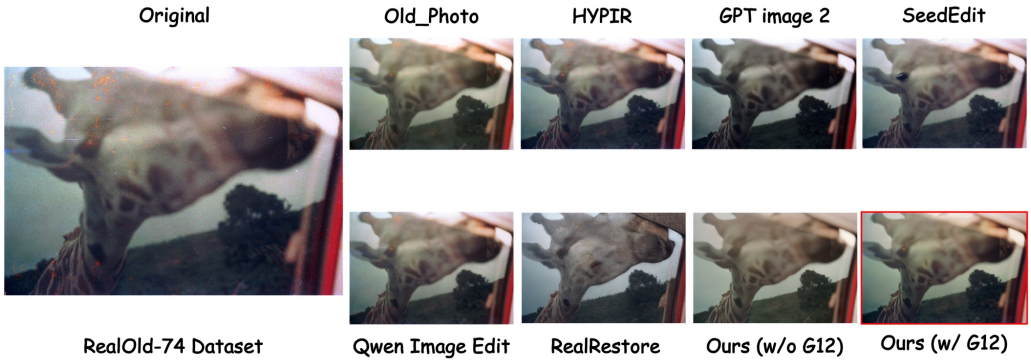

Original
Old_Photo
HYPIR
GPT image 2
SeedEdit
GGT 100k Dataset
Qwen Image Edit
RealRestore
Ours (w/o G12)
Ours (w/ G12)


Original
Old_Photo
HYPIR
GPT image 2
SeedEdit
GGT 100k Dataset
Qwen Image Edit
RealRestore
Ours (w/o G12)
Ours (w/ G12)

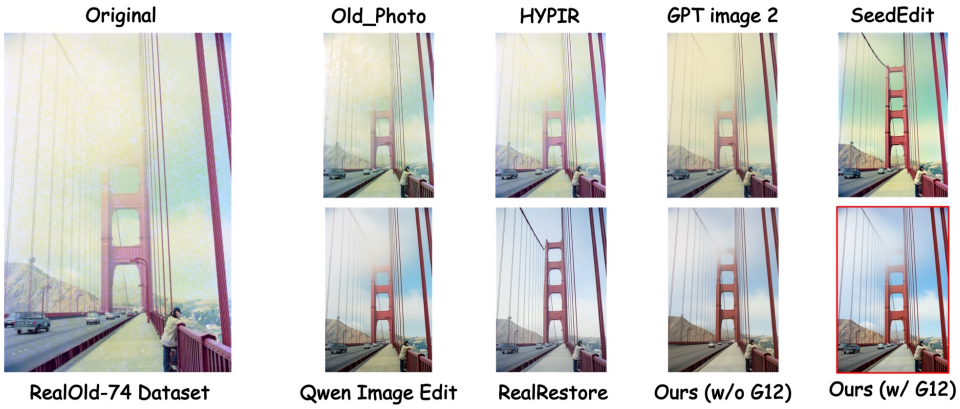

Original
Old_Photo
HYPIR
GPT image 2
SeedEdit
GGT 100k Dataset
Qwen Image Edit
RealRestore
Ours (w/o G12)
Ours (w/ G12)

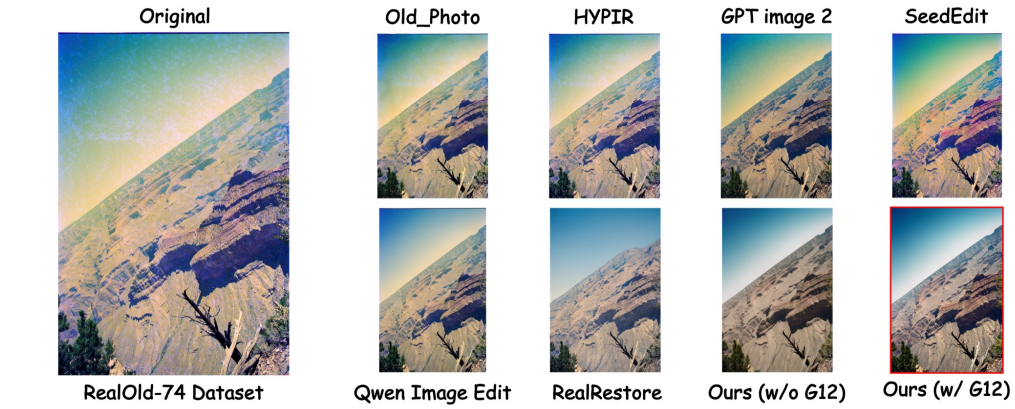

Original
Old_Photo
HYPIR
GPT image 2
SeedEdit
GGT 100k Dataset
Qwen Image Edit
RealRestore
Ours (w/o G12)
Ours (w/ G12)

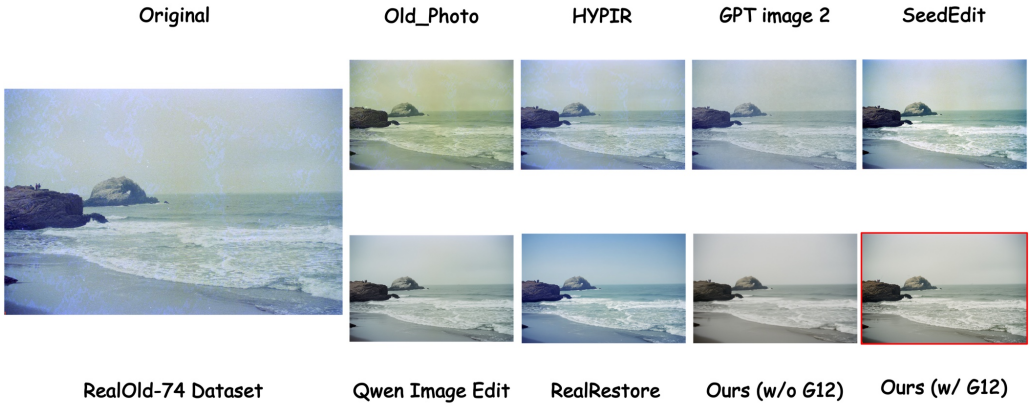

Original
Old_Photo
HYPIR
GPT image 2
SeedEdit
GGT 100k Dataset
Qwen Image Edit
RealRestore
Ours (w/o G12)
Ours (w/ G12)

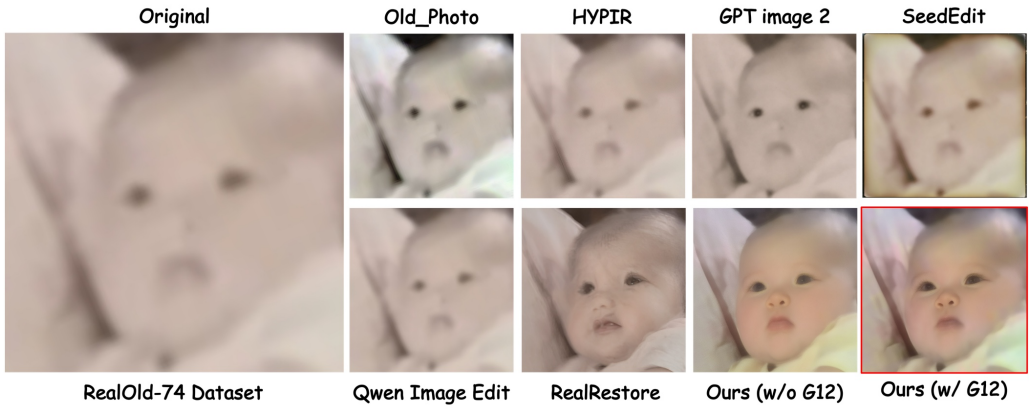

Original
Old_Photo
HYPIR
GPT image 2
SeedEdit
RealOld-74 Dataset
Qwen Image Edit
RealRestore
Ours (w/o G12)
Ours (w/ G12)

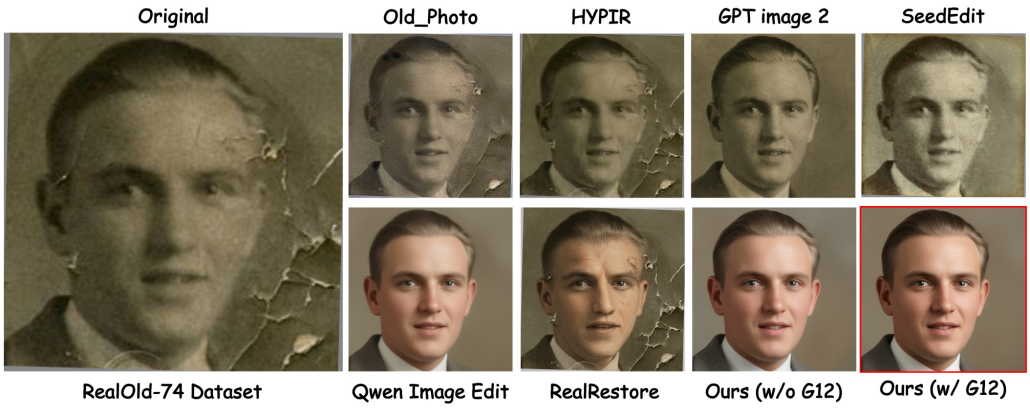

Original
Old_Photo
HYPIR
GPT image 2
SeedEdit
RealOld-74 Dataset
Qwen Image Edit
RealRestore
Ours (w/o G12)
Ours (w/ G12)


Original
Old_Photo
HYPIR
GPT image 2
SeedEdit
RealOld-74 Dataset
Qwen Image Edit
RealRestore
Ours (w/o G12)
Ours (w/ G12)

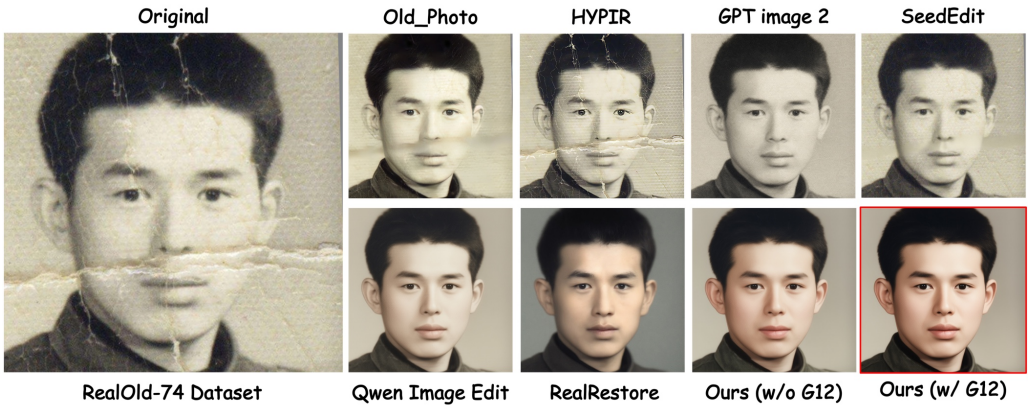

Original
Old_Photo
HYPIR
GPT image 2
SeedEdit
RealOld-74 Dataset
Qwen Image Edit
RealRestore
Ours (w/o G12)
Ours (w/ G12)

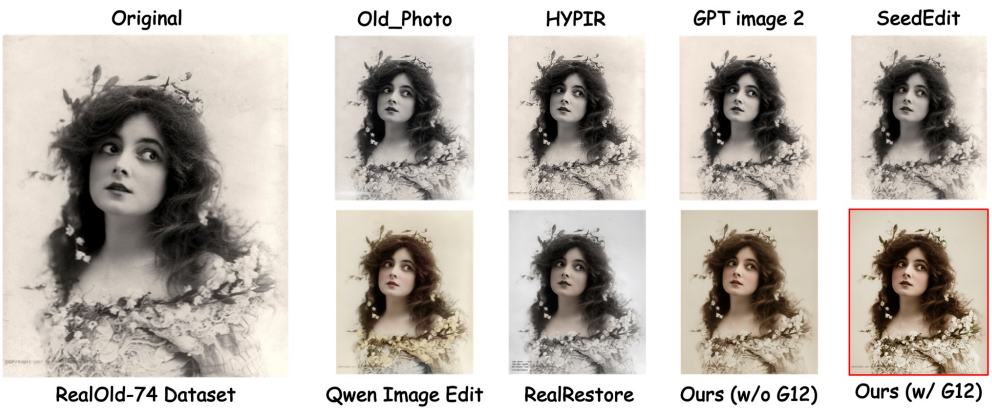

Original
Old_Photo
HYPIR
GPT image 2
SeedEdit
RealOld-74 Dataset
Qwen Image Edit
RealRestore
Ours (w/o G12)
Ours (w/ G12)


Original
Old_Photo
HYPIR
GPT image 2
SeedEdit
RealOld-74 Dataset
Qwen Image Edit
RealRestore
Ours (w/o G12)
Ours (w/ G12)


Original
Old_Photo
HYPIR
GPT image 2
SeedEdit
RealOld-74 Dataset
Qwen Image Edit
RealRestore
Ours (w/o G12)
Ours (w/ G12)


Original
Old_Photo
HYPIR
GPT image 2
SeedEdit
RealOld-74 Dataset
Qwen Image Edit
RealRestore
Ours (w/o G12)
Ours (w/ G12)


Original
Old_Photo
HYPIR
GPT image 2
SeedEdit
RealOld-74 Dataset
Qwen Image Edit
RealRestore
Ours (w/o G12)
Ours (w/ G12)


Original
Old_Photo
HYPIR
GPT image 2
SeedEdit
RealOld-74 Dataset
Qwen Image Edit
RealRestore
Ours (w/o G12)
Ours (w/ G12)