DRIFTRECT: SELECTIVE TRAJECTORY RECTIFICATION FOR FAITHFUL FLOW-BASED IMAGE EDITING

Anonymous authors
Paper under double-blind review

ABSTRACT

Flow-based image editors can realize strong semantic changes with pretrained generative transport, yet the same editing dynamics may also move content that should remain unchanged. Source-specific inversion provides useful information about the input image, but directly following a reconstruction trajectory can couple source recovery with edit execution and suppress the requested change. We therefore formulate faithful flow editing as *selective trajectory rectification*: retain the native source–target edit transport as the primary actuator and selectively correct collateral drift around it. We introduce DRIFTRECT, a training-free framework that first partitions the latent into an edit core, a transition/release region, and a protected exterior from source–target responses. It then combines late exterior damping with progressive spatial and channel–time release, while an inversion-derived source reference supplies bounded image-specific geometric and channel cues. Together, these components regulate where, when, and how strongly preservation pressure is applied without constructing a replacement editing direction. On a fixed 28-pair controlled diagnostic set, DRIFTRECT improves four complementary preservation metrics by 2.8–6.7% over FlowEdit; on the full 700-image PIE-Bench at seed 0, the corresponding gains are 2.0–3.5%. A complete second-seed replication retains favorable gains on all four measures after averaging seeds within image. In both settings, the paired CLIP-T interval spans zero. Controlled mechanism studies further show that inversion-derived source information provides a complementary source-specific fidelity margin when routed through the bounded controller.

1 INTRODUCTION

Recent advances in large-scale text-to-image generative models have made text-guided editing of real images increasingly practical. Unlike unconstrained image generation, however, real-image editing must not only realize the requested semantic change, but also preserve visual content that is unrelated to the instruction. These two objectives are inherently coupled. Aggressive editing may alter background appearance, geometry, identity, or fine-grained texture, whereas overly conservative updates may preserve the source simply by under-editing. The central challenge is therefore not only to produce a meaningful semantic change, but to concentrate the editing dynamics on the content that should change while suppressing unnecessary motion elsewhere.

Flow matching and rectified flow provide a continuous transport formulation for modern generative models through learned time-dependent velocity fields (Lipman et al., 2023; Liu et al., 2023). Recent flow-based editors exploit these pretrained dynamics directly. In particular, FlowEdit constructs source- and target-conditioned responses and forms an inversion-free editing trajectory without additional model optimization (Kulikov et al., 2025). Its native source–target transport provides strong semantic actuation, but the resulting motion is not inherently restricted to the support of the requested modification. Consequently, content that should remain stable can evolve together with the edited region. Figure 1 illustrates this unintended motion, which we call *collateral trajectory drift*.

Source-specific inversion provides complementary information for controlling such drift. An inverted trajectory captures how a particular source image is represented by the generative dynamics and therefore contains image-dependent structural and appearance cues (Mokady et al., 2023). However, reconstruction and editing serve different objectives: reconstruction favors returning toward

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.

the observed source, whereas editing must depart from it wherever required by the target instruction. Directly injecting source-return velocity or forcing the edit to follow the inverted trajectory can therefore improve source similarity while suppressing the requested change. This suggests that source information is most useful not as a replacement editing direction, but as a signal for regulating the native editing dynamics.

Based on this observation, we formulate faithful flow-based editing as *selective trajectory rectification*. The key idea is to separate two roles that are entangled in the native editing dynamics: *edit actuation*, which realizes the target semantics, and *preservation control*, which suppresses unnecessary deviation from the source. Rather than constructing a new editing direction, selective trajectory rectification retains the native source–target edit transport as the primary driving signal and introduces bounded residual control according to where, when, and how strongly preservation is needed.

We instantiate this principle as DRIFTRECT, a training-free trajectory-rectification framework built on a frozen flow editor. DRIFTRECT first derives a response-guided spatial field that separates an edit core, a transition/release region, and a protected exterior. It then applies late-stage exterior damping to suppress collateral motion in source-stable regions, while progressive spatial and channel–time release restores editing freedom where the requested modification needs to develop. In parallel, an inversion-derived source reference supplies image-specific geometric and channel cues through a bounded correction pathway. Together, these mechanisms form a selective residual controller over the native editing trajectory, concentrating edit-driving motion on the requested change while preserving surrounding source structure.

Experiments validate this design across both controlled and public settings. On a fixed synthetic diagnostic set of 28 source–target pairs generated with FLUX.1-dev (Labs, 2024) and evaluated with two random seeds, DRIFTRECT improves LPIPS-I, DreamSim-I, BG-LPIPS, and Boundary-MSE by 5.11%, 2.79%, 4.51%, and 6.73% over FlowEdit, respectively, while the paired target-text interval spans zero. On the full 700-image PIE-Bench at seed 0 (Ju et al., 2024), DRIFTRECT again improves the same four preservation metrics by 2.48%, 2.01%, 2.80%, and 3.49%, respectively; a complete second seed preserves favorable effects on all four measures. The paired CLIP-T intervals cross zero in both settings. Together, these results show consistent source-fidelity gains across diverse editing tasks at closely matched target-text scores.

Our contributions are summarized as follows:

- We identify *collateral trajectory drift* as a key source of fidelity degradation in flow-based image editing, and formulate faithful editing as *selective trajectory rectification* that separates edit actuation from preservation control.
- We introduce DRIFTRECT, a training-free residual controller for frozen flow editors that combines response-guided spatial partitioning, late exterior damping, progressive spatial/channel–time release, and bounded source-specific anchoring to selectively regulate the native editing trajectory.
- We demonstrate consistent source-fidelity improvements on both controlled and full PIE-Bench evaluations without a statistically detected reduction in target alignment, together with mechanism analyses that disentangle spatial preserve–release control, source-specific anchoring, and edit actuation.

2 RELATED WORK

Inversion and fidelity-preserving editing. Diffusion inversion methods recover source-consistent latents for real-image editing (Song et al., 2021; Mokady et al., 2023); Edit-Friendly DDPM Noise Space, Eta Inversion, and PnP Inversion further show that reconstruction accuracy alone does not ensure editability (Huberman-Spiegelglas et al., 2024; Kang et al., 2024; Ju et al., 2024). Rout et al. control the fidelity–editability balance during rectified stochastic inversion, while Xu et al. combine flow inversion, velocity compensation, and invariance control (Rout et al., 2025; Xu et al., 2025). Attention-control methods such as Prompt-to-Prompt and MasaCtrl preserve layout or appearance through cross- or self-attention manipulation (Hertz et al., 2023; Cao et al., 2023). Our interface keeps the native source–target drive as the actuator and applies bounded preserve–release residuals around it.

Inversion-free flow editing and trajectory regularization. FlowEdit constructs source–target transport from a pretrained flow model without explicit inversion (Kulikov et al., 2025). FlowAlign regularizes the editing trajectory to balance semantic alignment and source consistency (Kim et al., 2026), while SNR-Edit modifies the stochastic source trajectory through structure-aware noise control and segmentation-guided constraints (Jiang et al., 2026). All three treat trajectory behavior as a locus of fidelity control, but at different interfaces: FlowAlign regularizes the trajectory globally, SNR-Edit modifies source-side noise/trajectory construction, and DRIFTRECT applies response-localized residual control to the already formed native edit drive across space, time, and channels.

Source-guided and feature-guided rectified-flow editing. A parallel line improves faithful flow editing through more accurate inversion, noise alignment, or reuse of source features. FireFlow and RF-Edit improve inversion/ODE accuracy, with RF-Edit additionally sharing source features during editing (Deng et al., 2025; Wang et al., 2025); DNAEdit aligns source and target noise paths directly (Xie et al., 2025); and DirectEdit focuses on step-level accurate inversion for flow-based editing (Yang & Ye, 2026). Stable Flow identifies "vital" DiT layers and selectively injects attention features, while ReFlex extracts features from a mid-step inverted latent and adapts source attention before injection (Avrahami et al., 2025; Kim et al., 2025). KV-Edit preserves background content through DiT key–value caching rather than trajectory correction (Zhu et al., 2025). DRIFTRECT differs from these feature- or cache-reuse strategies in that preservation is expressed directly as bounded residual corrections to the editing velocity rather than by reusing source attention/features as the primary preservation channel.

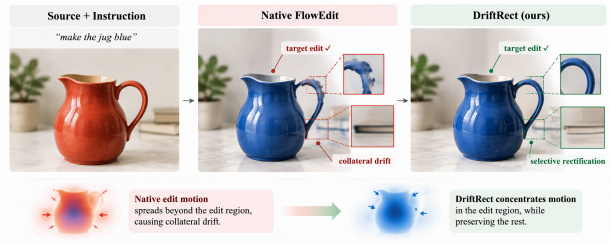

Figure 1: Motivation. FlowEdit completes the requested edit but can induce collateral drift; DRIFTRECT selectively rectifies this motion while preserving source-stable details.

Relation to SteerFlow and the closest source-trajectory interfaces. SteerFlow is the closest inversion-based counterpart: it combines a high-fidelity source inversion with trajectory interpolation between target-editing and source-reconstruction velocities, and further applies adaptive masking to constrain spatial changes (Dao et al., 2026). DRIFTRECT shares the goal of using source-specific information to improve fidelity, but adopts a different interface. Instead of blending the source-reconstruction velocity into the main editing trajectory, DRIFTRECT preserves the native source–target transport as the edit-driving signal and routes source information through bounded spatial, temporal, and channel-wise corrections. This distinction is also the focus of our interface study, which directly compares source-velocity injection, trajectory blending, spatially masked guidance, and selective residual rectification under the same backbone.

3 METHOD

3.1 OVERVIEW

Let I_s be the source image, c_s and c_t the source and target prompts, and $\mathbf{x} = \mathcal{E}(I_s)$ its latent. A frozen rectified-flow model predicts $\mathbf{v}_\theta(\mathbf{z}_t, t, c)$ with

$$\frac{d\mathbf{z}_t}{dt} = \mathbf{v}_\theta(\mathbf{z}_t, t, c). \tag{1}$$

At active step i, FlowEdit evaluates the source branch

$$\mathbf{z}_i^s = (1-\sigma_i)\mathbf{x} + \sigma_i \boldsymbol{\epsilon}_i, \qquad \mathbf{z}_i^{\text{share}} = \mathbf{z}_i + \mathbf{z}_i^s - \mathbf{x}, \tag{2}$$

and forms the native edit drive

$$\mathbf{g}_i = \mathbf{v}_\theta(\mathbf{z}_i^{\text{share}}, t_i, c_t) - \mathbf{v}_\theta(\mathbf{z}_i^s, t_i, c_s). \tag{3}$$

DRIFTRECT uses $\mathbf{g}_i$ as the primary editing drive and augments it with bounded residuals that decide *where* to preserve, *when* to apply pressure, and *where* to release it. Figure 2 summarizes the pipeline, and Fig. 4 gives the complementary latent-space view.

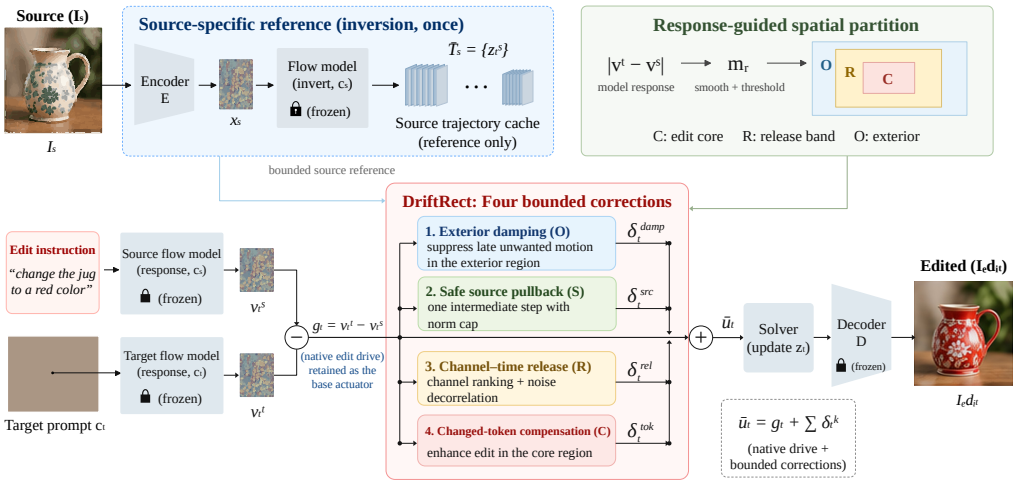

Figure 2: Overall framework of DRIFTRECT. The frozen native source–target transport provides the editing drive, while four bounded corrections concentrate motion on the requested edit and preserve stable source content.

3.2 RECTIFICATION REFERENCES

Response-guided spatial field. At the first active edit step, source and target prompts are evaluated at the same probe state. For latent location p,

$$a(p) = \sqrt{\mathrm{mean}_c \left(v_{p,c}^t - v_{p,c}^s\right)^2}. \tag{4}$$

We threshold a at the 60th percentile, dilate by one token, and Gaussian-smooth the support to obtain $\mathbf{m}_r$. A hard support $\mathbf{M}_h$ and a three-token boundary ring $\mathbf{B}$ define

$$\mathbf{C} = \mathbf{M}_h \odot (1 - \mathbf{B}), \quad \mathbf{R} = \max(\mathbf{M}_h, \mathbf{B}), \quad \mathbf{O} = 1 - \mathbf{m}_r, \tag{5}$$

corresponding to the edit core, release/transition region, and soft exterior preservation field. This partition encodes the preserve–release asymmetry directly: the edit core is allowed to move, the boundary is treated as a transition region, and the exterior receives the strongest preservation pressure. Paired targets and evaluation masks are used exclusively for evaluation and are not provided to the generation pipeline.

Image-specific source reference. We invert the source once under c_s and cache

$$\bar{\mathcal{T}}_s = \{(\bar{\mathbf{z}}_i^s, \bar{\mathbf{v}}_i^s)\}_{i=0}^K. \tag{6}$$

The cache serves as a source-specific reference: its endpoint provides source-adaptive channel information and one aligned intermediate state provides the bounded geometric anchor. These are the only online cache reads; the remaining cached states and velocities are not injected into the edit update, leaving the target-conditioned transport as the edit actuator.

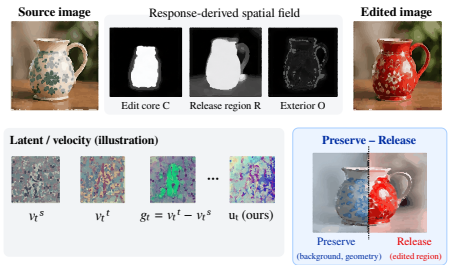

Figure 3: Response-guided spatial control: edit core, release region, and protected exterior.

3.3 FOUR BOUNDED CORRECTIONS

Exterior damping. Late in the trajectory, unnecessary motion is attenuated outside the edit support:

$$\boldsymbol{\delta}_i^{\mathrm{damp}} = -\kappa_i \, \mathbf{O} \odot \mathbf{g}_i, \tag{7}$$

where $\kappa_i = 0$ during the first half of the trajectory and increases smoothly during the second half. Before the remaining residual corrections are applied, the resulting damped drive can equivalently be written as

$$\mathbf{g}_i^{\text{damp}} = \mathbf{g}_i + \boldsymbol{\delta}_i^{\text{damp}} = (1 - \kappa_i \mathbf{O}) \odot \mathbf{g}_i. \tag{8}$$

This form makes the role of exterior damping explicit: it acts as a spatially and temporally varying attenuation of the native edit motion, leaving early trajectory formation unconstrained while increasingly suppressing late motion outside the predicted edit support.

Progressive release. Preservation pressure can over-constrain the requested edit, especially near the edit boundary. We therefore release only selected channel–time degrees of freedom. From the inversion endpoint $\mathbf{z}_{\text{inv}}$, we compute a per-channel discrepancy

$$q_c = \frac{|\mu(\mathbf{z}_{\text{inv},c}) - \mu(\mathbf{x}_c)|}{\text{std}(\mathbf{z}_{\text{inv},c}) + \text{std}(\mathbf{x}_c) + 10^{-6}}, \qquad G_c = \text{sigmoid}\left(\frac{r_c - 0.25}{0.10}\right), \tag{9}$$

where r_c is the normalized rank of q_c. Together with a decreasing time gate s_i, this modulates source-noise correlation. The implementation mixes the new noise with the original noise *inside the spatial release region* before reevaluating the target branch:

$$\rho_{i,c}^{\text{eff}} = 1 - G_c s_i (1 - \rho_i), \qquad \widetilde{\epsilon}_{i,c} = (1 - \mathbf{R}) \odot \epsilon_{i,c} + \mathbf{R} \odot \left[\rho_{i,c}^{\text{eff}} \epsilon_{i,c} + \sqrt{1 - (\rho_{i,c}^{\text{eff}})^2}\, \xi_{i,c}\right]. \tag{10}$$

The target branch is re-evaluated with this spatially mixed $\widetilde{\epsilon}$, yielding $\widetilde{\mathbf{g}}_i$, and only the residual inside the release region is added:

$$\boldsymbol{\delta}_i^{\text{rel}} = \frac{1}{2}\, \mathbf{R} \odot (\widetilde{\mathbf{g}}_i - \mathbf{g}_i). \tag{11}$$

Thus damping and release are complementary: damping removes collateral exterior motion, whereas release prevents the preservation controller from freezing the requested edit.

Bounded source anchor and semantic safeguard. At a single intermediate step $i^\star$, we advance the current state once under the source condition,

$$\widehat{\mathbf{z}}_{i^\star+1}^{s} = \mathbf{z}_{i^\star} + \Delta\sigma_{i^\star} \mathbf{v}_\theta(\mathbf{z}_{i^\star}, t_{i^\star}, c_s), \tag{12}$$

and compare it with the cached next source state. The resulting exterior pullback

$$\boldsymbol{\delta}_{i^\star}^{\text{src,raw}} = \mathbf{O} \odot \frac{\bar{\mathbf{z}}_{i^\star+1}^{s} - \widehat{\mathbf{z}}_{i^\star+1}^{s}}{\Delta\sigma_{i^\star}} \tag{13}$$

is stripped of components that conflict with the native editing direction and capped at 2% of $\|\mathbf{g}_{i^\star}\|_2$; it is zero elsewhere. Separately, target tokens changed relative to the source prompt receive a mild embedding boost, and the resulting edit-core correction $\boldsymbol{\delta}_i^{\text{tok}}$ is capped at 10% of $\|\mathbf{g}_i\|_2$. This token term is a semantic safeguard rather than a central source of preservation gain.

The final update is

$$\boxed{\mathbf{u}_i^{\text{rect}} = \mathbf{g}_i + \boldsymbol{\delta}_i^{\text{damp}} + \boldsymbol{\delta}_i^{\text{rel}} + \boldsymbol{\delta}_i^{\text{src}} + \boldsymbol{\delta}_i^{\text{tok}}} \tag{14}$$

followed by the original solver step $\mathbf{z}_{i+1} = \mathbf{z}_i + \Delta\sigma_i \mathbf{u}_i^{\text{rect}}$. Setting all residuals to zero exactly recovers the native update. This residual form is central to the design: the target-conditioned transport determines the edit direction, while rectification changes only the amount of motion allowed to survive across space, channels, and trajectory phases. DRIFTRECT is training-free and introduces no additional learnable parameters.

4 EXPERIMENTS

4.1 PROTOCOL

Controlled benchmark. For controlled mechanism analysis, we use a fixed synthetic diagnostic set of 28 source–target pairs generated with FLUX.1-dev (Labs, 2024). A deterministic object/color/scene vocabulary defines the prompts, and each source–target pair is rendered from the

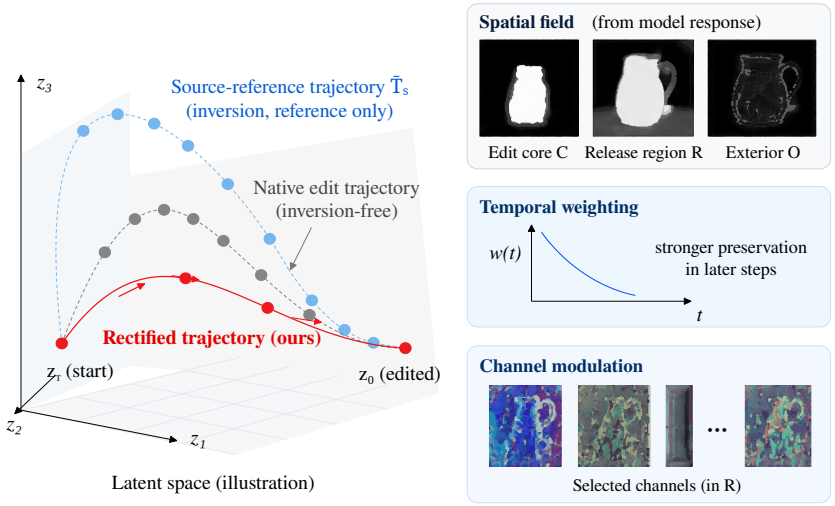

Figure 4: Latent-space view: the native edit trajectory remains the actuator, while source-derived cues provide bounded rectification.

same sampled noise under the corresponding source and target prompts. The set contains 14 local-attribute edits and 14 background edits, together with fixed geometric masks used only for region-wise evaluation. We use this internally constructed set to isolate preservation–editing trade-offs rather than as a substitute for a public benchmark; generalization is evaluated separately on PIE-Bench. We compare with FlowEdit, RF-Edit (Wang et al., 2025), FireFlow (Deng et al., 2025), DNAEdit (Xie et al., 2025), and DirectEdit (Yang & Ye, 2026).

Public benchmark. For public generalization, we evaluate the full 700-image PIE-Bench (Ju et al., 2024). The five-method comparison uses seed 0; we additionally evaluate FlowEdit and full DRIFTRECT on all 700 images at seed 1, then average the two seeds within each image for the paired robustness analysis. Public methods are FlowEdit, ReFlex (Kim et al., 2025), DirectEdit (Yang & Ye, 2026), full DRIFTRECT, and a strict DRIFTRECT-NoInv variant that removes all inversion trajectory/endpoint reads. All public outputs are evaluated through the same RGB256 preprocessing path.

Metrics and statistics. CLIP-T measures target alignment. The controlled benchmark uses DINO-I semantic similarity together with LPIPS-I, DreamSim-I, BG-LPIPS, and Boundary-MSE. For PIE-Bench we additionally report the PnP-style DINO-B/8 structure distance (lower is better). We average seeds within controlled pairs and use 50,000 paired bootstrap resamples; PIE uses the image sample as the paired unit with the same bootstrap count. Additional protocol details are in Appendix B.

4.2 CONTROLLED COMPARISON

Among methods above the edit-adherence floor, DRIFTRECT gives the best DINO-I, LPIPS-I, and DreamSim-I and matches the best BG-LPIPS at the reported precision. Relative to FlowEdit, LPIPS-I, DreamSim-I, BG-LPIPS, and Boundary-MSE improve by 5.11%, 2.79%, 4.51%, and 6.73%, respectively; all four paired intervals favor DRIFTRECT, while the CLIP-T interval crosses zero. The controlled result therefore locates a stronger preservation operating point among methods meeting the fixed edit-adherence criterion. Appendix E.3 adds a paired-target edit-region completion check, while Appendix E.4 reports the fixed 0.90–$1.00\times$ guardrail sweep; the controlled ranking is unchanged from 0.90 through 0.975.

4.3 PUBLIC GENERALIZATION

Relative to FlowEdit, full DRIFTRECT improves LPIPS-I, DreamSim-I, BG-LPIPS, and Boundary-MSE by 2.48%, 2.01%, 2.80%, and 3.49%, respectively. The paired 95% intervals for all four

Table 1: Controlled fidelity–adherence diagnostic on the fixed 28-pair synthetic set. Bold and underline denote the best and second-best results, respectively, among methods satisfying the fixed edit-adherence floor CLIP-T ≥ 27.702; ties at the reported precision are both bolded. $^{\times}$ marks methods below the floor, † denotes a favorable paired 95% interval relative to FlowEdit, and * marks the audited DirectEdit 8-bit transformer adaptation detailed in Appendix C.

Method	CLIP-T ↑	DINO-I ↑	LPIPS-I ↓	DreamSim-I ↓	BG-LPIPS ↓	Bnd. ↓
DNAEdit	**29.411**	0.612	<u>0.243</u>	<u>0.313</u>	**0.095**	**0.0137**
FlowEdit	<u>29.160</u>	<u>0.652</u>	0.252	0.315	0.100	0.0225
DirectEdit*	29.097	0.603	0.259	0.323	0.101	<u>0.0144</u>
DRIFTRECT	29.035	**0.664**	**0.239**†	**0.306**†	**0.095**†	0.0210†
RF-Edit$^{\times}$	*23.104*	*0.630*	*0.212*	*0.291*	*0.086*	*0.0107*
FireFlow$^{\times}$	*24.323*	*0.765*	*0.137*	*0.181*	*0.058*	*0.0065*

Table 2: Full PIE-Bench results on 700 images at seed 0 under a unified RGB256 evaluation. Bold and underline denote the best and second-best results, respectively, among complete methods; DRIFTRECT-NoInv is an ablation and is excluded from this ranking. DINO-B/8 Struct. is the PnP-style structure distance (lower is better). *DirectEdit uses its audited 8-bit transformer/CPU-offload implementation.

Method	CLIP-T ↑	DINO-Str. ↓	LPIPS-I ↓	DreamSim-I ↓	BG-LPIPS ↓	Bnd. ↓
FlowEdit	<u>26.0588</u>	0.02524	0.15827	0.18846	0.07247	0.01746
ReFlex	26.0445	0.03153	0.18243	<u>0.18257</u>	**0.06015**	0.02550
DirectEdit*	25.7605	**0.02172**	<u>0.15561</u>	**0.17395**	0.07238	**0.01111**
DRIFTRECT-NoInv (ablation)	26.0338	0.02454	0.15409	0.18402	0.07016	0.01684
DRIFTRECT	**26.0631**	<u>0.02452</u>	**0.15434**	0.18466	<u>0.07044</u>	<u>0.01685</u>

preservation metrics favor DRIFTRECT. The DINO-B/8 structure distance also decreases, while the CLIP-T difference is $+0.0043$ with a 95% interval of $[-0.0582, 0.0676]$.

The same trend is broad across edit types: LPIPS-I, BG-LPIPS, and Boundary-MSE improve in mean value for all 10 PIE categories, and DreamSim-I improves in 9/10. Compared with ReFlex, DRIFTRECT achieves lower LPIPS-I, Boundary-MSE, and DINO-B/8 structure distance, while ReFlex obtains lower BG-LPIPS; their CLIP-T and DreamSim-I intervals span zero. DirectEdit achieves lower structure distance, DreamSim-I, and Boundary-MSE, whereas DRIFTRECT has higher CLIP-T; their LPIPS-I and BG-LPIPS intervals also span zero. The methods therefore occupy different fidelity–adherence operating points.

The complete second-seed run contains 700 valid FlowEdit and 700 valid DRIFTRECT outputs. After averaging the two seeds within each image, DRIFTRECT improves LPIPS-I, DreamSim-I, BG-LPIPS, and Boundary-MSE by 2.46%, 1.63%, 3.02%, and 2.94%, respectively; all four image-paired 95% intervals favor DRIFTRECT and pass Holm correction for the four comparisons fixed before seed 1 analysis. The CLIP-T difference is $+0.0042$ with interval $[-0.0410, 0.0489]$. Appendix E.2 reports the full two-seed table and figure.

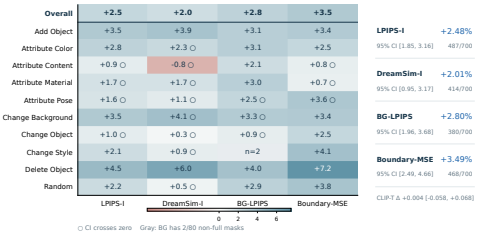

Figure 5: PIE700 category robustness at seed 0. The right rail reports overall paired effects and 95% intervals.

4.4 MECHANISM ANALYSIS

Interface design. A unified interface study compares direct source-velocity injection, source-trajectory blending, spatially masked source guidance, and DRIFTRECT under the same backbone and source reference. The study supports retaining the native source–target transport as the primary carrier of edit semantics, while routing source information through bounded residual control. A

separate actuator stress test provides complementary evidence that same-state conditional-response signals are better used as controller cues than as replacement editing directions. Detailed results are in Appendix F.

What drives the gain? Local drop-one tests identify release as the largest contributor, followed by the mask/damping scaffold; changed-token compensation and the explicit pullback have much smaller effects.

Detailed drop-one values and confidence intervals are reported in Appendix F.

Removing only the explicit one-step source anchor is nearly neutral because endpoint-conditioned channel allocation remains active in this row. A separate *strict NoInv* intervention removes both trajectory and endpoint reads. On Controlled28, full DRIFTRECT shows an additional source-reference margin on DINO-I, DreamSim-I, and BG-LPIPS under the primary control, while the fixed-rank sensitivity gives favorable intervals across all five preservation metrics. Appendix F reports the complete strict-NoInv statistics, the public Full–NoInv comparison, and cache-specificity controls, including a fixed two-seed PIE-Bench Gaussian-replacement study.

4.5 QUALITATIVE RESULTS AND EFFICIENCY

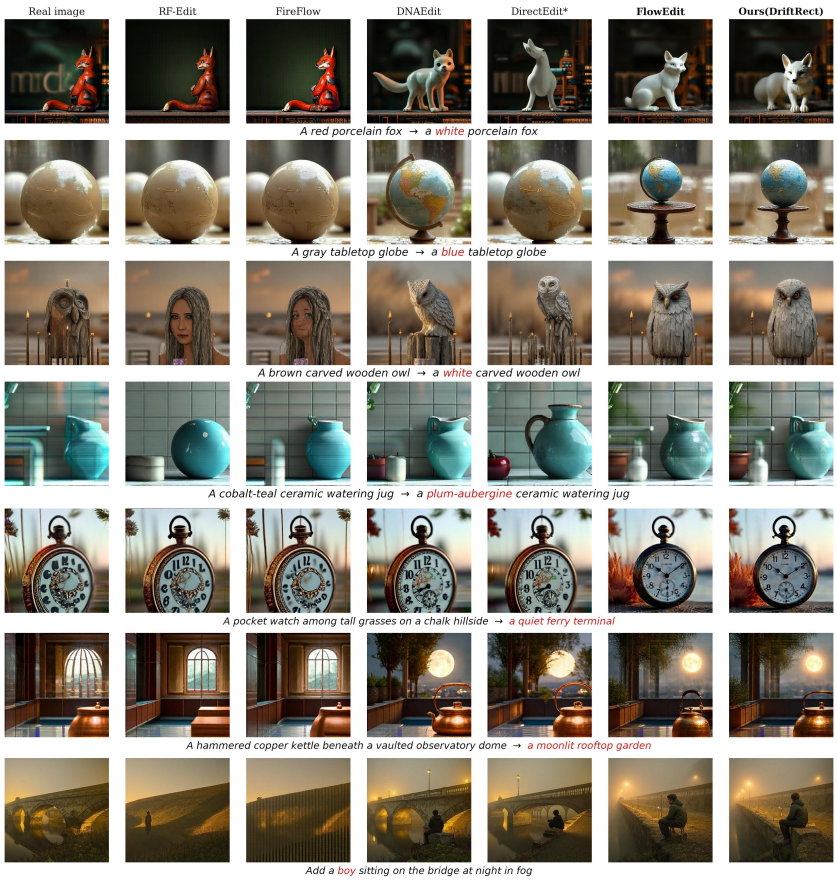

Figure 6: Seven controlled examples from the frozen 28-pair set at seed 0. Columns show source, RF-Edit, FireFlow, DNAEdit, DirectEdit, FlowEdit, and DRIFTRECT under the same source, prompt pair, and edit seed.

The controlled plate gives a wider cross-baseline view, while the public examples focus on source-aligned local details that should remain stable. The paired PIE700 analysis in Fig. 5 establishes the aggregate preservation effect. Appendix E.3 complements the source-fidelity metrics with paired-

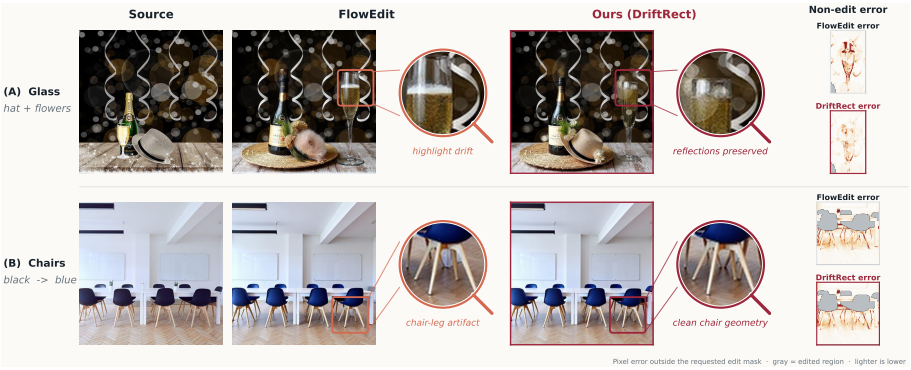

Figure 7: Two public PIE-Bench comparisons with matched detail crops. In the glass case, FlowEdit alters local reflections/highlights while DRIFTRECT better preserves the source detail; in the chair case, FlowEdit introduces a chair-leg/floor artifact while DRIFTRECT retains cleaner geometry.

target edit-region completion, and Appendix G reports cached timing, forward-count, and first-source accounting under separated measurement boundaries.

5 DISCUSSION

Across the controlled set and full PIE700, DRIFTRECT improves perceptual, background, and boundary preservation relative to FlowEdit; the public gains persist over a complete second seed while the target-text interval spans zero. DirectEdit and ReFlex are stronger on selected public metrics, showing complementary fidelity–adherence trade-offs.

Mechanism studies identify response-guided damping and progressive release as the main preserve–release scaffold, with bounded source-reference cues providing a complementary controlled margin. Detailed strict-NoInv, cache-specificity, interface, and paired-target completion diagnostics are reported in Appendices F and E.3.

6 LIMITATIONS

Our evaluation centers on the FLUX.1-dev backbone and combines full PIE-Bench generalization with controlled diagnostics. The primary FlowEdit–DRIFTRECT public comparison is replicated over two complete seeds, while the controlled set supports paired-target and mechanism analyses. A global correction-strength sweep is reported in Appendix F; the remaining controller settings are kept fixed across all reported experiments. Efficiency is reported with explicit cached and first-source accounting in Appendix G. Extending the same protocol to additional backbones, broader human preference studies, and fully instrumented first-use amortization would provide useful complementary evidence.

7 CONCLUSION

We introduced DRIFTRECT, a training-free framework for selective trajectory rectification in flow-based image editing. The method combines response-guided spatial partitioning, late exterior damping, progressive spatial/channel–time release, and a bounded source-specific reference to concentrate editing motion on the requested change while preserving stable source content. Controlled and full PIE700 evaluations show improved perceptual, background, and boundary fidelity, while paired CLIP-T intervals span zero; mechanism studies support complementary roles for preserve–release control and bounded source-derived cues.

AI USE STATEMENT

Generative AI tools were used to assist with language polishing . All methodological decisions, experimental configurations, reported results, and final claims were reviewed and determined by the authors, who take responsibility for the final content.

REPRODUCIBILITY STATEMENT

The method and main experimental protocol are specified in Sections 3 and 4. The appendix documents controlled/public data construction, frozen hyperparameters, statistical procedures, paired-target and threshold checks, mechanism/cache controls, efficiency accounting, additional qualitative evidence, and the verified execution environment. Paired targets and evaluation masks are never generation inputs.

REFERENCES

Omri Avrahami, Or Patashnik, Ohad Fried, Egor Nemchinov, Kfir Aberman, Dani Lischinski, and Daniel Cohen-Or. Stable Flow: Vital layers for training-free image editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7877–7888, 2025.

Mingdeng Cao, Xintao Wang, Zhongang Qi, Ying Shan, Xiaohu Qie, and Yinqiang Zheng. MasaCtrl: Tuning-free mutual self-attention control for consistent image synthesis and editing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 22560–22570, 2023.

Thinh Dao, Zhen Wang, Kien T. Pham, and Long Chen. SteerFlow: Steering rectified flows for faithful inversion-based image editing. *arXiv preprint arXiv:2604.01715*, 2026.

Yingying Deng, Xiangyu He, Changwang Mei, Peisong Wang, and Fan Tang. FireFlow: Fast inversion of rectified flow for image semantic editing. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 13110–13128, 2025.

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.

Inbar Huberman-Spiegelglas, Vladimir Kulikov, and Tomer Michaeli. An edit friendly DDPM noise space: Inversion and manipulations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12469–12478, 2024.

Lifan Jiang, Boxi Wu, Yuhang Pei, Tianrun Wu, Yongyuan Chen, Yan Zhao, Shiyu Yu, and Deng Cai. SNR-Edit: Structure-aware noise rectification for inversion-free flow-based editing. *arXiv preprint arXiv:2601.19180*, 2026.

Xuan Ju, Ailing Zeng, Yuxuan Bian, Shaoteng Liu, and Qiang Xu. PnP Inversion: Boosting diffusion-based editing with 3 lines of code. In *International Conference on Learning Representations*, 2024.

Wonjun Kang, Kevin Galim, and Hyung Il Koo. Eta inversion: Designing an optimal eta function for diffusion-based real image editing. In *European Conference on Computer Vision*, 2024.

Jeongsol Kim, Yeobin Hong, Jonghyun Park, and Jong Chul Ye. FlowAlign: Trajectory-regularized, inversion-free flow-based image editing. In *International Conference on Learning Representations*, 2026.

Jimyeong Kim, Jungwon Park, Yeji Song, Nojun Kwak, and Wonjong Rhee. ReFlex: Text-guided editing of real images in rectified flow via mid-step feature extraction and attention adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 15939–15948, 2025.

Vladimir Kulikov, Matan Kleiner, Inbar Huberman-Spiegelglas, and Tomer Michaeli. FlowEdit: Inversion-free text-based editing using pre-trained flow models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 19721–19730, 2025.

Black Forest Labs. FLUX. https://github.com/black-forest-labs/flux, 2024.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023.

Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Null-text inversion for editing real images using guided diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6038–6047, 2023.

Litu Rout, Yujia Chen, Nataniel Ruiz, Constantine Caramanis, Sanjay Shakkottai, and Wen-Sheng Chu. Semantic image inversion and editing using rectified stochastic differential equations. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/c07bc8b520aa87326b015b0da6a70b48-Abstract-Conference.html.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021.

Jiangshan Wang, Junfu Pu, Zhongang Qi, Jiayi Guo, Yue Ma, Nisha Huang, Yuxin Chen, Xiu Li, and Ying Shan. Taming rectified flow for inversion and editing. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 64044–64058, 2025.

Chenxi Xie, Minghan Li, Shuai Li, Yuhui Wu, Qiaosi Yi, and Lei Zhang. DNAEdit: Direct noise alignment for text-guided rectified flow editing. In *Advances in Neural Information Processing Systems*, volume 38, pp. 137937–137955, 2025.

Pengcheng Xu, Boyuan Jiang, Xiaobin Hu, Donghao Luo, Qingdong He, Jiangning Zhang, Chengjie Wang, Yunsheng Wu, Charles Ling, and Boyu Wang. Unveil inversion and invariance in flow transformer for versatile image editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 28479–28489, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Xu_Unveil_Inversion_and_Invariance_in_Flow_Transformer_for_Versatile_Image_CVPR_2025_paper.html.

Desong Yang and Mang Ye. DirectEdit: Step-level accurate inversion for flow-based image editing. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026.

Tianrui Zhu, Shiyi Zhang, Jiawei Shao, and Yansong Tang. KV-Edit: Training-free image editing for precise background preservation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16607–16617, 2025.

A Supplementary Overview

This appendix contains details moved out of the nine-page-limited main text: evaluation provenance and statistics, frozen implementation settings, complete public two-seed results, paired-target completion and threshold/Pareto checks, response-gate diagnostics, strict no-inversion and cache-specificity controls (including a public Gaussian replacement), interface and correction-strength studies, efficiency accounting, and additional qualitative evidence. The narrative follows the main paper: native actuation first, selective rectification second, with inversion used as a bounded source-specific reference.

Category	Pairs	Preservation focus
Color / local attribute	14	Geometry, identity, texture, background, boundary
Background	14	Foreground identity/geometry and residual non-edited content

Table 3: Composition of the fixed controlled synthetic diagnostic set.

B Evaluation Sets and Protocol Details

B.1 Controlled 28-pair diagnostic

Provenance and construction. The controlled set contains 28 synthetic source–target pairs generated with FLUX.1-dev. A deterministic generator combines a fixed vocabulary of object identities, colors, and scene descriptions; each source and paired target is rendered from the same sampled noise while changing only the relevant prompt attributes. The set contains 14 local-attribute edits and 14 background edits. It was consolidated from earlier fixed internal development manifests and then frozen for the comparisons reported here. We use it for controlled diagnostic evidence, with external generalization assessed separately on the full PIE-Bench.

Every pair records source/target prompts, a source image, a separately generated paired target, and a deterministic edit-region mask. All methods receive only the source image and source/target prompts. Paired targets and evaluation masks are never generation inputs; the latter are used only after generation for region-wise metrics such as BG-LPIPS and Boundary-MSE.

Seeds and resolution. The stochastic methods use edit seeds 0 and 1. For the shared-noise FlowEdit/DriftRect implementation,

$$\mathrm{seed}(s, p, i) = 1{,}000{,}003s + 97p + i, \tag{15}$$

where s is the edit seed, p the pair identifier, and i the scheduler index. RF-Edit and FireFlow are deterministic under the audited official path; DirectEdit's two audited outputs are also identical. We nevertheless retain the common two-record protocol and average within pair before inference. Ordinary outputs are evaluated at 256×256; DirectEdit's 512×512 output is resized to the same RGB256 grid.

B.2 Public PIE-Bench evaluation

We evaluate all 700 PIE-Bench images at seed 0 for five public rows: FlowEdit, ReFlex, DirectEdit, strict DriftRect-NoInv, and full DriftRect. FlowEdit and full DriftRect additionally cover every image at seed 1. The earlier 40-image seed-1 audit is included within this complete second-seed run. No public result is used to alter the sample set or tune method hyperparameters.

Strict NoInv removes all inversion-trajectory and inversion-endpoint reads and replaces endpoint-dependent channel allocation with the frozen uniform-mass control used in the causal study; it is therefore not equivalent to merely dropping the explicit one-step source anchor. All public outputs are evaluated through an identical RGB256 preprocessing path. For two-seed statistics, the two seeds are averaged within each image before 50,000 paired image-level bootstrap resamples.

B.3 Statistics

For the controlled benchmark, if $y_{p,s,m}$ denotes a metric for pair p, seed s, and method m, we first compute

$$\bar{y}_{p,m} = \frac{1}{2} \sum_{s \in \{0,1\}} y_{p,s,m}. \tag{16}$$

Each 95% interval is the 2.5th–97.5th percentile range of 50,000 bootstrap resamples over pair indices; PIE uses the same procedure over image IDs. The controlled CLIP-T floor is $0.95\times$ the FlowEdit mean, or 27.702, and is used only for the controlled operating-point comparison. Section E.4 reports the full 0.90–1.00 sensitivity sweep; PIE700 is analyzed separately.

Algorithm 1 DRIFTRECT: Selective Trajectory Rectification

Require: Source image I_s, prompts (c_s, c_t)
Encode I_s and build the reusable source reference
Estimate response field m_r and regions $(\mathbf{C}, \mathbf{R}, \mathbf{O})$
Compute endpoint-derived channel gate and changed target tokens
Initialize the native inversion-free edit; set $K = 24$ active steps
for $i = 0, \ldots, K-1$ **do**
  Compute native drive $\mathbf{g}_i$
  Compute exterior damping $\boldsymbol{\delta}_i^{\mathrm{damp}}$
  Compute released target and $\boldsymbol{\delta}_i^{\mathrm{rel}}$
  if $i = i^\star = 6$ **then**
    Compute bounded projected source anchor $\boldsymbol{\delta}_i^{\mathrm{src}}$
  end if
  Compute token safeguard $\boldsymbol{\delta}_i^{\mathrm{tok}}$ and update with Eq. 14
end for
return decoded final latent

Component	Frozen value	Role
Source / target guidance	1.5 / 5.5	Native FlowEdit drive
Active editing	final 24 of 28	One solver update per active step
Response field	first active step, 60th percentile	Prompt-responsive support
Postprocess	dilate 1; Gaussian $\sigma = 1$; threshold 0.5	Smooth spatial gate
Boundary ring	3 latent tokens	Transition/release band
Exterior damping	steps 0–11: 0; 12–23: 0.00435→0.10	Late preservation pressure
Source anchor	step 6 only; 2% norm cap	Bounded source correction
Channel gate	$\mathrm{sigmoid}((r_c - 0.25)/0.10)$	Endpoint-derived allocation
Time gate	normalized $\mathrm{sigmoid}(8(0.5 - u))$	Progressive release
Noise correlation	$\rho_{\min} = 0.85$	Controlled stochastic release
Release energy	0.5	Restore edit freedom in $\mathbf{R}$
Token safeguard	LCS difference; $\times 1.10$; 10% cap	Protect changed-token semantics
Precision	bfloat16; CPU offload	Audited 24-GiB execution

Table 4: Frozen DRIFTRECT configuration.

C IMPLEMENTATION DETAILS

All internal methods use frozen FLUX.1-dev with `FluxTransformer2DModel`, CLIP/T5 conditioning, FLUX AutoencoderKL, and `FlowMatchEulerDiscreteScheduler`. The solver has 28 steps and editing acts on the final 24 active steps. DRIFTRECT is training-free and adds no learned parameters.

Release implementation. For channel c,

$$q_c = \frac{|\mu(z_{\mathrm{inv,c}}) - \mu(x_c)|}{\mathrm{std}(z_{\mathrm{inv,c}}) + \mathrm{std}(x_c) + 10^{-6}}, \qquad G_c = \mathrm{sigmoid}\left(\frac{r_c - 0.25}{0.10}\right). \tag{17}$$

Together with the decreasing time gate s_i, this defines $\rho_{i,c}^{\mathrm{eff}} = 1 - G_c s_i (1 - \rho_i)$. New Gaussian noise is mixed with the original noise only inside the release field $\mathbf{R}$, the target branch is re-evaluated, and only the resulting residual inside $\mathbf{R}$ is added. This matches the frozen implementation.

External configurations. RF-Edit uses its pinned FLUX runner with 15 inversion and 15 editing steps, guidance 2.0, and feature injection 3. FireFlow uses 8+8 steps, guidance 2.0, and injection 1. DNAEdit uses the audited FLUX bridge with 48 forwards, guidance 2.5, $\eta = 0.85$, and $T_{\mathrm{start}} = 4$. DirectEdit uses its audited 30-step inversion/edit path, inversion/recovery CFG 1.0/2.0, and attention ratio 0.15. To fit a 24-GiB GPU, its transformer is loaded in 8-bit mode with model CPU offload; this adaptation and its $512 \to 256$ evaluation resize apply only to DirectEdit.

D ADDITIONAL METHOD VISUALIZATION

The latent-space visualization is already shown in the main paper and is therefore not repeated here.

Metric	FlowEdit	DRIFTRECT	Rel. benefit	95% benefit CI	Wins
CLIP-T ↑	29.160	29.035	-0.43%	$[-0.4294, 0.1881]$	13/28
DINO-I ↑	0.652	0.664	+1.87%	$[-0.00189, 0.02558]$	20/28
LPIPS-I ↓	0.252	0.239	+5.11%	$[0.00643, 0.01952]$	22/28
DreamSim-I ↓	0.315	0.306	+2.79%	$[0.00039, 0.01668]$	22/28
BG-LPIPS ↓	0.100	0.095	+4.51%	$[0.00170, 0.00752]$	20/28
Boundary-MSE ↓	0.0225	0.0210	+6.73%	$[0.000726, 0.002358]$	23/28

Table 5: Controlled pair-level uncertainty after averaging two seeds within each pair. Positive benefit favors DRIFTRECT in the metric's favorable direction.

Metric	Local attribute (14)	Background (14)
CLIP-T	-1.23% $[-2.51, 0.25]$	+0.43% $[-1.02, 1.90]$
DINO-I	+3.82% $[1.87, 6.31]$	-0.09% $[-3.47, 3.35]$
LPIPS-I	+6.95% $[3.42, 10.34]$	+3.19% $[-0.13, 6.21]$
DreamSim-I	+4.73% $[2.35, 7.47]$	+0.32% $[-4.81, 4.07]$
BG-LPIPS	+7.99% $[3.98, 12.43]$	+1.38% $[-2.04, 4.16]$
Boundary-MSE	+10.59% $[5.12, 15.92]$	+3.47% $[-0.29, 8.31]$

Table 6: Favorable relative change and pair-bootstrap 95% CI by controlled category.

E EXTENDED QUANTITATIVE RESULTS

E.1 CONTROLLED UNCERTAINTY AND CATEGORY ANALYSIS

Holm correction covered CLIP-T, DINO-I, BG-LPIPS, and Boundary-MSE. The adjusted probabilities are .42164, .17448, .0036, and $< 1/50{,}000$, respectively; BG-LPIPS and Boundary-MSE remain significant after correction.

E.2 PUBLIC PIE700 RESULTS

The complete second seed contains 700 valid FlowEdit and 700 valid DRIFTRECT predictions. Nine heuristic artifact flags in the seed-1 records are retained; all six metric columns are finite. Seeds are averaged within image before the paired bootstrap.

Image-paired gains are 2.46% $[2.01, 2.91]$ for LPIPS-I and 1.63% $[0.94, 2.32]$ for DreamSim-I. BG-LPIPS improves by 3.02% $[2.46, 3.59]$, and Boundary-MSE by 2.94% $[2.30, 3.63]$. All four intervals favor DRIFTRECT and pass Holm correction for the fixed family of four comparisons. CLIP-T changes by $+0.0042$ $[-0.0410, +0.0489]$, so target-text non-inferiority is not established. The two-seed category means favor DRIFTRECT on all four preservation measures in 10/10 categories; this is a direction count, not a category-wise significance claim.

E.3 PAIRED-TARGET COMPLETION DIAGNOSTIC

A frozen geometric edit mask selects the requested region. Pixels outside it are set to the same neutral gray in source, paired target, and output before a shared bounding-box crop with 10% context. We compute LPIPS-I and DreamSim-I to the paired target and define progress as source-to-target distance minus output-to-target distance. The paired targets are synthetic references, not human judgments.

Relative to FlowEdit, DRIFTRECT has lower mean target LPIPS by 0.00348 and lower mean target DreamSim by 0.00402, both favorable on 17/28 pairs. The paired intervals are $[-0.00016, 0.00857]$ and $[-0.00238, 0.01145]$, respectively. Both intervals include zero; the diagnostic is therefore used as a completion check alongside, rather than as a replacement for, the source-fidelity metrics. Both FlowEdit and DRIFTRECT have negative mean progress against these synthetic paired targets.

E.4 EDIT-ADHERENCE THRESHOLD AND PARETO SENSITIVITY

The controlled comparison uses a CLIP-T floor of $0.95\times$ the FlowEdit mean. We additionally examine fixed factors $\{0.90, 0.925, 0.95, 0.975, 1.00\}$. Across 0.90–0.975, the qualifying set is unchanged and DRIFTRECT remains the best qualifying row on DINO-I, LPIPS-I, and DreamSim-I;

Method	CLIP-T ↑	DINO-B/8 ↓	LPIPS-I ↓	DreamSim-I ↓	BG-LPIPS ↓	Boundary ↓
FlowEdit	26.0588	0.02524	0.15827	0.18846	0.07247	0.01746
ReFlex	26.0445	0.03153	0.18243	0.18257	0.06015	0.02550
DirectEdit*	25.7605	0.02172	0.15561	0.17395	0.07238	0.01111
DRIFTRECT-NoInv	26.0338	0.02454	0.15409	0.18402	0.07016	0.01684
DRIFTRECT	26.0631	0.02452	0.15434	0.18466	0.07044	0.01685

Table 7: Unified RGB256 means on all 700 PIE-Bench images at seed 0. *DirectEdit uses the audited 8-bit transformer/CPU-offload implementation.

Method	CLIP-T ↑	DINO-B/8 ↓	LPIPS-I ↓	DreamSim-I ↓	BG-LPIPS ↓	Boundary ↓
FlowEdit	26.0578	0.02552	0.15987	0.19002	0.07304	0.01749
DRIFTRECT	26.0620	0.02483	0.15594	0.18693	0.07084	0.01698

Table 8: PIE700 two-seed image-averaged means for the two methods with full seed-1 coverage.

DNAEdit is strongest on BG-LPIPS and Boundary-MSE. At 1.00, only DNAEdit and FlowEdit remain above the floor.

E.5 RESPONSE-GATE LOCALIZATION DIAGNOSTIC

We compare the response-derived edit support with the fixed geometric evaluation mask on a common 16×16 latent-token grid. Gate IoU has Spearman association 0.11 with favorable LPIPS-I gain (10,000-resample interval $[-0.23, 0.43]$) and 0.18 with boundary gain ($[-0.20, 0.52]$). These intervals span zero, so localization overlap alone does not explain the preservation gains. An exploratory negative association between response-support area and LPIPS gain is not used as a confirmed mechanism.

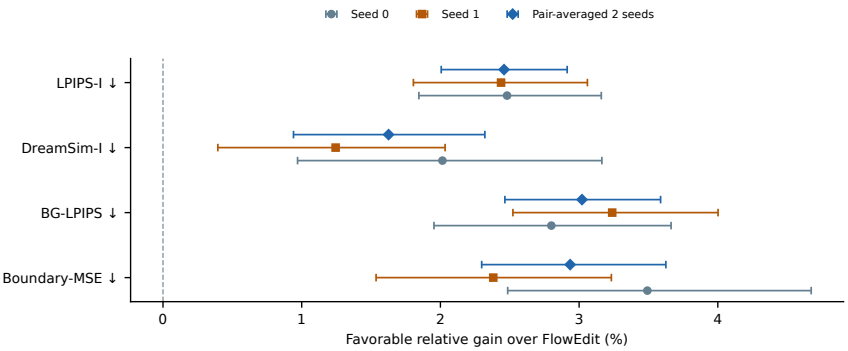

Figure 8: PIE700 preservation gains of DRIFTRECT relative to FlowEdit at seed 0, seed 1, and after averaging two seeds within each image. Whiskers are paired bootstrap 95% intervals.

Method	Target LPIPS ↓	Target DreamSim ↓	LPIPS progress ↑	DreamSim progress ↑
FlowEdit	0.2178	0.2471	-0.0202	-0.0167
DRIFTRECT	0.2143	0.2431	-0.0167	-0.0127
DirectEdit	0.1889	0.2218	+0.0087	+0.0087
DNAEdit	0.1922	0.2205	+0.0054	+0.0100

Table 9: Controlled28 paired-target diagnostic on the fixed edit-region crop, averaged over 28 pairs and two seeds.

Factor	CLIP floor	Qualifying methods	Lowest LPIPS-I
0.900	26.244	DNAEdit, FlowEdit, DirectEdit, DRIFTRECT	DRIFTRECT
0.925	26.973	DNAEdit, FlowEdit, DirectEdit, DRIFTRECT	DRIFTRECT
0.950	27.702	DNAEdit, FlowEdit, DirectEdit, DRIFTRECT	DRIFTRECT
0.975	28.431	DNAEdit, FlowEdit, DirectEdit, DRIFTRECT	DRIFTRECT
1.000	29.160	DNAEdit, FlowEdit	DNAEdit

Table 10: Controlled edit-adherence guardrail sensitivity. The factor multiplies the FlowEdit CLIP-T mean.

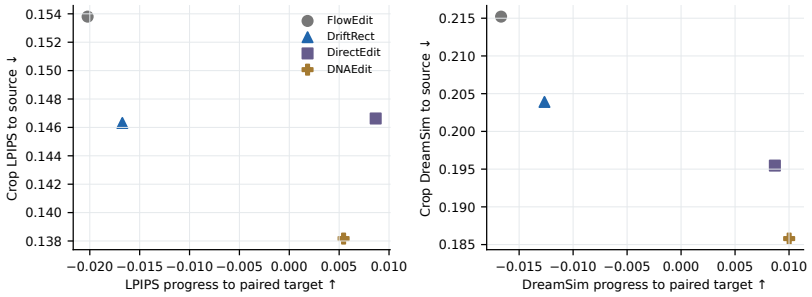

Figure 9: Controlled28 preservation–completion operating points using paired synthetic targets. The plot shows method means, not paired uncertainty.

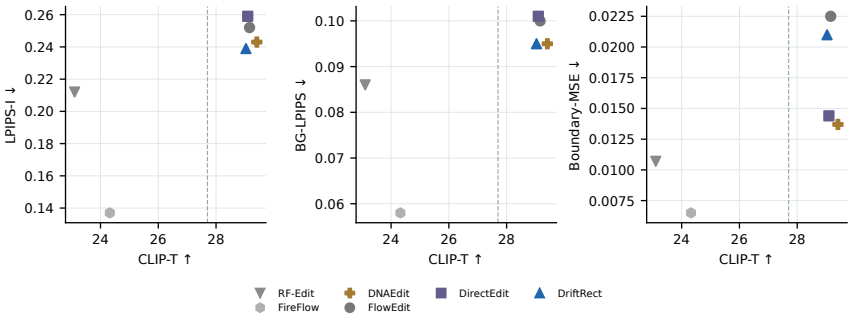

Figure 10: Controlled28 target-text alignment versus source-fidelity operating points. Points are frozen method means and do not imply paired significance between every method pair.

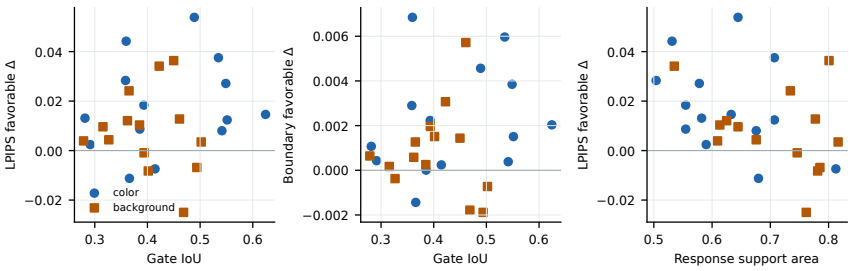

Figure 11: Controlled28 response-gate localization and exploratory association with paired preservation gains. The evaluation mask is used only for this post-generation diagnostic.

F MECHANISM AND CAUSAL CHECKS

F.1 COMPONENT DROP-ONE

Release is the largest local contributor, followed by mask/damping. The explicit pullback is nearly neutral because other inversion-dependent pathways, especially endpoint-conditioned channel allocation, remain active; it must not be interpreted as a strict no-inversion experiment.

F.2 STRICT NO-INVERSION CONTROLS

The controlled strict-NoInv study removes both source-trajectory and inversion-endpoint reads. Table 12 reports full-minus-NoInv paired effects under the primary uniform-mass replacement and a fixed-rank sensitivity replacement; positive values favor the full method in each metric's favorable direction.

Relative to the full-method gain over FlowEdit, the primary replacement retains 79.8% on LPIPS-I, 82.7% on BG-LPIPS, and 81.6% on Boundary-MSE; the fixed-rank replacement retains 63.6%, 63.1%, and 68.0%. Across five preservation metrics, the median full-versus-NoInv margin is 20.2% under the primary replacement and 36.9% under the sensitivity replacement. On full PIE700, the Full–NoInv paired intervals include zero for all six reported metrics.

F.3 CACHE SPECIFICITY AND STATES ACTUALLY USED

An eight-pair source-reconstruction diagnostic compares correct, shuffled, and distribution-matched Gaussian caches.

For editing, the correct cache is more useful than a category-preserving shuffled cache on DINO-I, LPIPS-I, DreamSim-I, and Boundary-MSE, while CLIP-T is unchanged and BG-LPIPS is inconclusive. However, all six correct-versus-statistics-matched-Gaussian intervals cross zero; we therefore claim correspondence relative to mismatched real trajectories, not unique necessity of exact trajectory identity.

Public PIE-Bench cache replacement. To test cache specificity under the public image distribution, we use a fixed hash-selected, category-balanced subset of 100 PIE-Bench images (10 per official edit category), with seeds 0 and 1. The source image, prompts, edit seed, response mask, editor, and evaluator are fixed; only the correct source cache is replaced by a statistics-matched Gaussian cache. Seeds are averaged within image before 50,000 paired bootstrap resamples over the 100 image IDs.

The DINO-B/8 interval favors the correct cache; the other five intervals include zero. For the prespecified four-metric preservation family (LPIPS-I, DreamSim-I, BG-LPIPS, Boundary-MSE), Holm-adjusted probabilities are 0.7122, 1.0000, 1.0000, and 1.0000. This public subset is a cache-specificity diagnostic and is kept separate from the full 700-image benchmark.

Which cached states are actually used. Although the implementation stores a reconstructed source path, the online controller reads only the inversion endpoint to construct the channel gate and one aligned cached state (state 7 at active step 6) for the bounded geometric pullback. The remaining cached states and cached velocities do not enter the online update. A full-minus-explicit-anchor control keeps the endpoint gate while removing the one-step pullback; this comparison is reported separately from strict NoInv because it isolates only the explicit anchor.

F.4 INTERFACE AND ACTUATOR STUDIES

A unified interface study compares no inversion, direct source-velocity injection, source-trajectory interpolation/blending, spatially masked source guidance, and DRIFTRECT under the same frozen FlowEdit backbone and source reference. Direct injection and trajectory blending move toward the source at the cost of target alignment, whereas residual rectification retains the native actuator and yields a more useful fidelity–editability operating point.

Dropped term	CLIP-T	DINO-I	LPIPS-I	DreamSim-I	BG-LPIPS	Boundary
Mask/Damping	-0.2046	+0.00225	+0.00506	+0.00230	+0.00200	+0.000328
Release	-0.0696	+0.00868	+0.00957	+0.00906	+0.00326	+0.001375
Token safeguard	+0.0621	-0.00258	-0.00093	-0.00059	-0.000335	+0.000183
Explicit pullback	+0.0152	-0.00112	-0.00002	-0.00030	-0.000052	-0.000023

Table 11: Extended controlled drop-one effects. Favorable metric directions are normalized so positive preservation deltas favor the full method.

Metric	Primary favorable Δ [95% CI]	Fixed-rank favorable Δ [95% CI]
CLIP-T	$+0.05882$ $[-0.14566,\ 0.28313]$	-0.08796 $[-0.29621,\ 0.13237]$
DINO-I	$+0.00463$ $[0.00011,\ 0.00963]$	$+0.00721$ $[0.00137,\ 0.01350]$
LPIPS-I	$+0.00260$ $[-0.00009,\ 0.00576]$	$+0.00469$ $[0.00186,\ 0.00778]$
DreamSim-I	$+0.00378$ $[0.00053,\ 0.00702]$	$+0.00493$ $[0.00150,\ 0.00868]$
BG-LPIPS	$+0.00078$ $[0.00002,\ 0.00161]$	$+0.00166$ $[0.00059,\ 0.00282]$
Boundary-MSE	$+0.000279$ $[-0.000064,\ 0.000673]$	$+0.000485$ $[0.000059,\ 0.000955]$

Table 12: Controlled Full–strict-NoInv paired effects.

A separate actuator stress test replaces the native source–target gap with same-state conditional-response constructions. All candidate replacement actuators reduce CLIP-T relative to the corresponding native update, supporting the design choice to retain the native mixed state–condition transport.

Correction-strength sensitivity. The frozen Stage141 data include the one-dimensional correction strengths $s \in \{0.25, 0.50, 1.00, 1.50\}$ over the same 28 pairs and two seeds. Table 17 reports favorable deltas relative to FlowEdit; the nominal paper setting remains $s = 1$, and the $s = 1.5$ row is diagnostic rather than a retuned paper setting.

This sweep directly probes the global correction strength. Response threshold, boundary width, release energy, and anchor cap remain fixed at the values in Table 4 throughout the reported study.

Cache	DINO ↑	LPIPS ↓	MSE ↓	PSNR ↑
Correct	0.640	0.261	0.0366	14.91
Shuffled	0.288	0.657	0.1642	8.25
Gaussian	0.324	0.600	0.0916	10.59

Table 13: Controlled source reconstruction under correct, shuffled, and distribution-matched Gaussian caches.

Configuration	CLIP-T ↑	DINO-I ↑	LPIPS-I ↓	BG-LPIPS ↓	Boundary ↓
Mask only	29.146	0.658	0.248	0.0981	0.02207
Without release	29.104	0.656	0.249	0.0984	0.02238
Correct cache	29.035	0.664	0.239	0.0952	0.02100
Shuffled cache	29.035	0.658	0.242	0.0958	0.02123
Gaussian cache	28.985	0.668	0.239	0.0950	0.02101

Table 14: Focused editing-task checks on Controlled28.

Metric	Correct	Gaussian	Favorable Δ	95% CI	Wins
CLIP-T ↑	25.930395	25.925692	+0.004703	$[-0.046694, 0.060142]$	50/100
DINO-B/8 ↓	0.023313	0.023444	+0.000130	$[0.000039, 0.000226]$	61/100
LPIPS-I ↓	0.149665	0.150101	+0.000436	$[-0.000183, 0.001096]$	58/100
DreamSim-I ↓	0.173585	0.173860	+0.000275	$[-0.000554, 0.001227]$	58/100
BG-LPIPS ↓	0.069379	0.069408	+0.000029	$[-0.000397, 0.000388]$	44/100
Boundary-MSE ↓	0.016923	0.016953	+0.000031	$[-0.000049, 0.000113]$	56/100

Table 15: Correct versus statistics-matched Gaussian cache on the fixed PIE-Bench 100-image, two-seed subset. Positive deltas favor the correct cache.

Actuator	No protection	Mask/Damping	Strict NoInv	Full DRIFTRECT
XCS	-0.9933	-1.0728	-0.3520	-0.4306
XJAC	-0.3589	-0.4306	-0.0703	-0.1189
X-CoSA	-0.3659	-0.4178	-0.0724	-0.1439
X-CAN	-0.1032	-0.1682	-0.0376	-0.0879

Table 16: Editing-gain matrix: CLIP-T change relative to the same-context native actuator. All replacement actuators are negative.

s	CLIP-T	DINO-I	LPIPS-I	DreamSim-I	BG-LPIPS	Boundary
0.25	-0.0283	+0.00418	+0.00380	+0.00306	+0.001382	+0.000695
0.50	-0.1435	+0.00581	+0.00657	+0.00357	+0.001912	+0.001063
1.00	-0.1254	+0.01218	+0.01287	+0.00879	+0.004495	+0.001516
1.50	-0.0894	+0.01204	+0.01571	+0.00884	+0.005540	+0.001854

Table 17: Controlled correction-strength sensitivity. Positive values are favorable in the metric's preferred direction.

G EFFICIENCY AND CORRECTION BUDGET

The earlier controlled timings were collected on RTX 4090 workers. The reported 30.49 s (FlowEdit) and 36.66 s (DRIFTRECT) are cached edit/decode wall times and do not include every first-use reference-building phase. The active DRIFTRECT edit loop uses 97 transformer forwards; the 103-forward per-edit accounting additionally includes four endpoint-inversion evaluations and two response probes. Constructing a new reusable source trajectory adds 28 more forwards, giving 131 first-source model evaluations. These counts must not be read as measured first-use latency.

A separate category-balanced PIE700 pilot times 30 frozen 512×512 images across four GPU lanes; each image's three methods run on the same GPU. Model loading, evaluation, raw-source encoding, and source-reference construction are excluded.

In this cached pilot, full DRIFTRECT is 3.77 s slower than strict NoInv (1.57% paired relative overhead, 95% interval $[1.09, 2.14]\%$) and 123.94 s slower than FlowEdit (103.43% $[102.33, 104.61]\%$). The absolute seconds are not combined with the earlier 256-pixel timing because resolution and timing boundaries differ. First-use wall time remains unmeasured.

Method	Inv./edit	Per-edit fwd.	First-source fwd.	Cached mean	Cached P90	Peak MB
RF-Edit	15/15	30	30	61.69	85.21	24,076
FireFlow	8/8	16	16	43.48	50.84	23,325
DNAEdit	0/48	48	48	35.68	47.35	22,956
FlowEdit	0/28	48	48	30.49	33.95	22,954
DirectEdit*	30/30	60	60	41.34	44.84	22,999
DRIFTRECT-Weak	4/28	55	83	31.54	35.20	22,954
DRIFTRECT	4/28	103	131	36.66	38.61	22,958

Table 18: Controlled efficiency accounting. Measured seconds are cached edit/decode times; first-use source-reference construction is excluded.

Method	Loop forwards	Mean (s)	P90 (s)
FlowEdit	48	119.83	121.98
DRIFTRECT-NoInv	96	240.00	242.94
DRIFTRECT	97	243.77	247.28

Table 19: PIE700 cached edit+decode timing pilot on 30 frozen images. This table excludes first-use source/reference construction.

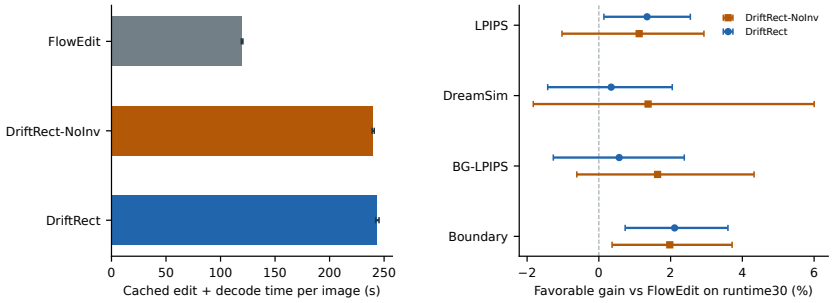

Figure 12: Cached edit/decode seconds and paired preservation gains on the frozen 30-image PIE700 timing subset. Source/reference first-use phases are excluded.

Method	CLIP-T ↑	DINO ↑	LPIPS ↓	BG ↓	Fwd.	Time
FlowEdit	29.160	0.652	0.252	0.100	48	30.49
DRIFTRECT-Weak	29.087	0.654	0.249	0.098	55	31.54
DRIFTRECT	29.035	0.664	0.239	0.095	103	36.66

Table 20: Ordered correction-budget operating points under the controlled protocol. These are categorical configurations rather than a calibrated continuous sweep.

H ADDITIONAL QUALITATIVE EVIDENCE

Outside-mask line-retention F1 improves on 20/28 controlled pairs, with mean difference $+0.009$ and 95% CI $[-0.003, 0.022]$. The interval crosses zero, so this result is treated only as a structural diagnostic.

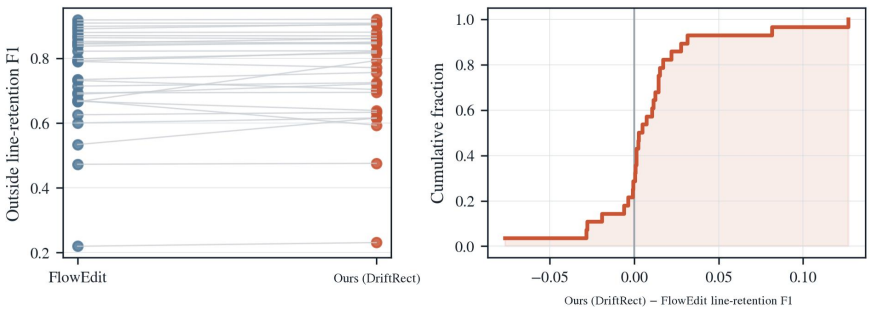

Figure 13: Pairwise outside-mask line-retention diagnostic. Each line connects FlowEdit and DRIFTRECT for one controlled pair; the companion empirical CDF summarizes the pairwise difference. This metric was not used for model selection.

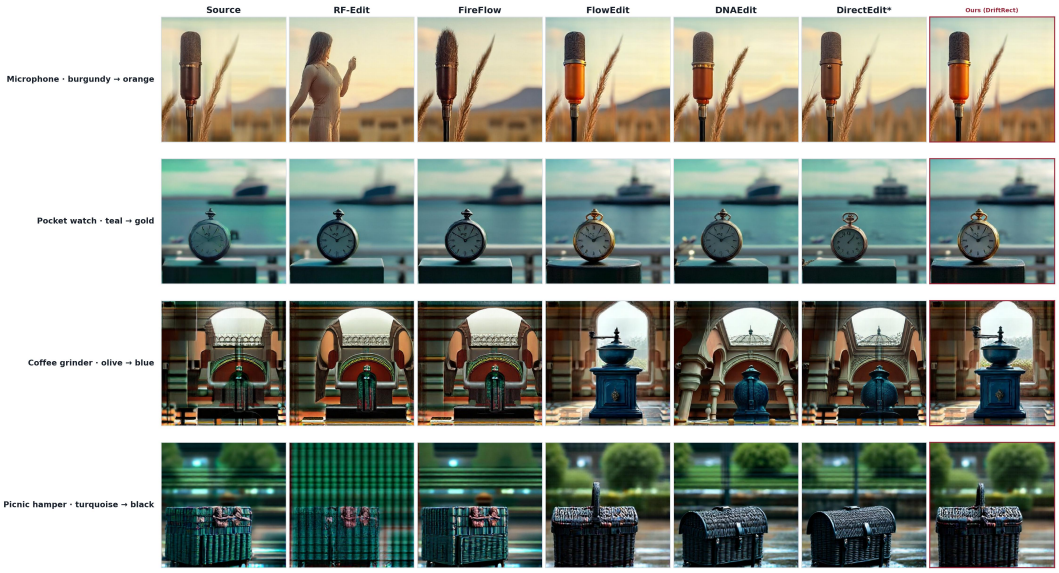

Figure 14: Additional local-attribute stress cases under the prespecified single-run protocol. No color enhancement, sharpening, or generative retouching is applied.

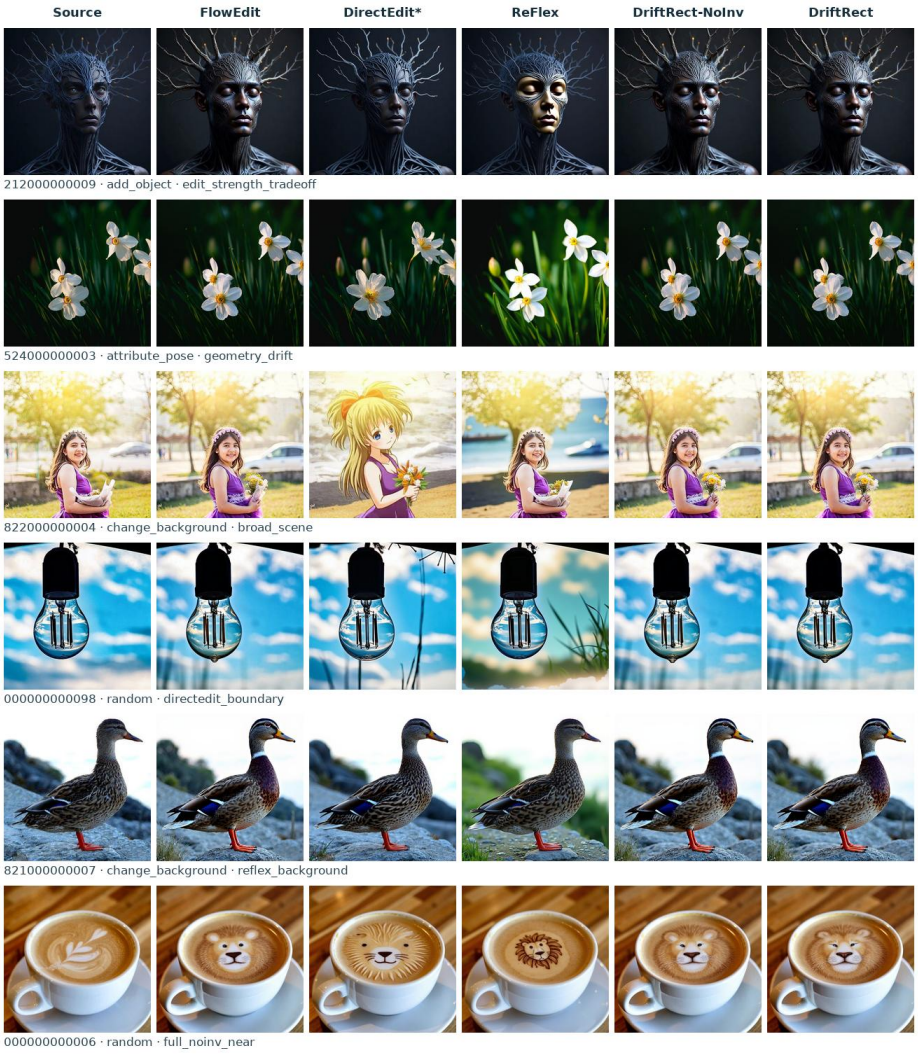

Figure 15: Six PIE-Bench seed-0 mixed cases selected by fixed diagnostic criteria. They illustrate edit-strength, boundary, scene-change, and cross-method trade-offs; the cases are not used to estimate failure prevalence.

I Computing Environment and Reproducibility Notes

The controlled and public evaluations use frozen manifests and paired statistical units. No paired target or evaluation mask is used as a generation input. Controlled mechanism checks reuse the fixed 28-pair protocol and should be interpreted as diagnostics rather than as an independent confirmation set.

Item	Representative inspected worker / execution scope
GPU	NVIDIA GeForce RTX 4090, 24,564 MiB; one GPU per editing job
CPU / memory	dual Intel Xeon Gold 6326; 32 physical cores / 64 threads; 377 GiB RAM
Operating system	Ubuntu 20.04.5 LTS; Linux 5.4
Python	3.10.20
PyTorch / CUDA / cuDNN	2.12.0+cu126 / 12.6 / 9.10.2
diffusers / transformers	0.38.0 / 5.11.0
accelerate	1.14.0
NumPy / SciPy / Pillow	2.2.6 / 1.15.3 / 12.2.0
DriftRect precision / offload	bfloat16 / model CPU offload; no quantization
DirectEdit adaptation	float16 pipeline; 8-bit transformer; model CPU offload
Evaluator grid	RGB256 Lanczos image resize; nearest-neighbor binary-mask resize

Table 21: Representative verified execution environment. Jobs were sharded across hosts; the CPU/memory row does not describe every historical worker.