ER-CHORD: SPATIALLY CONSISTENT ONE-STEP IMAGE EDITING VIA EQUAL-ENERGY RELIABILITY REDISTRIBUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

One-step generative models enable fast text-guided image editing, yet they often introduce unintended changes to regions that should remain unchanged. Temporal aggregation stabilizes editing fields across noise levels, but leaves their spatial allocation insensitive to local variations in reliability. Consequently, temporal stability alone does not ensure spatially consistent editing. To address this limitation, we propose ER-Chord, a training-free and inversion-free framework for reliability-guided spatial redistribution. Our key insight is that agreement between editing fields at neighboring noise levels provides a local reliability signal for allocating editing strength. We formulate this allocation as an equal-energy optimization that redistributes control energy according to local reliability while preserving the original editing directions and total control energy. This formulation admits a unique global solution that can be efficiently computed using existing field evaluations, requiring no additional denoiser calls. Experiments on two benchmarks, PIE-Bench and TEdBench, demonstrate improvements over state-of-the-art one-step editing baselines, achieving superior source preservation alongside consistent gains across multiple generative backbones.

1 INTRODUCTION

Distilled text-to-image models, such as SD-Turbo, InstaFlow, and SwiftBrush-v2 (Sauer et al., 2024; Liu et al., 2024; Dao et al., 2024), compress diffusion sampling into a single neural function evaluation. Although this acceleration enables real-time image synthesis, editing real images remains difficult because it requires introducing requested semantic changes while preserving the identity, structure, and unedited regions of the source image. Existing accelerated editors address this trade-off through one-step inversion (Nguyen et al., 2025), few-step reconstruction (Deutch et al., 2024; Gong et al., 2025), or inversion-free transport (Kulikov et al., 2025; Lu et al., 2026). Among these approaches, training-free and inversion-free transport is especially attractive because it eliminates both auxiliary model training and iterative latent inversion. However, executing an edit through a single transport update imposes strict requirements on the spatial organization of the resulting control field.

ChordEdit (Lu et al., 2026) stabilizes one-step editing by aggregating source-to-target proxy fields from nearby noise levels. This formulation primarily addresses temporal variations in the estimated transport field. Because this aggregation rule is applied uniformly across the latent grid, the magnitude at each spatial location is merely a passive result of combining the endpoint fields. However, the two endpoint estimates do not agree equally across all locations. Some regions exhibit consistent local control vectors across noise levels, whereas others show substantial disagreement. Uniform temporal aggregation ignores this spatial disparity and distributes editing strength without considering local reliability. Consequently, it often introduces large updates into unedited backgrounds, causing visual artifacts. Therefore, temporal stability does not ensure spatial consistency, leaving spatial magnitude allocation as a crucial yet unaddressed degree of freedom.

This observation motivates a distinction between *where the estimated control is reliable* and *where a semantic edit should occur*. Existing localized editing methods typically use cross-attention maps or internal feature representations to identify edit regions (Couairon et al., 2023; Simsar et al., 2025; Soni et al., 2025; Nguyen et al., 2025; Hu et al., 2026). In contrast to these localization approaches,

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

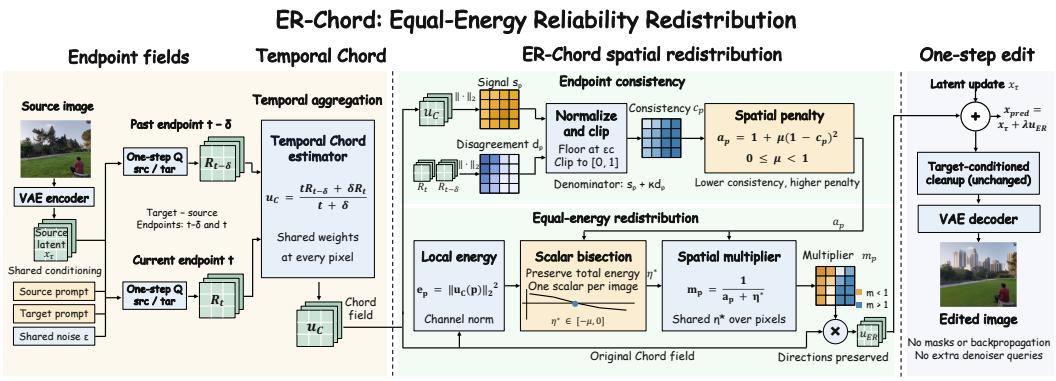

Figure 1: **Method overview.** Starting from a temporally aggregated control field, ER-Chord measures local endpoint consistency and redistributes a fixed control energy budget while preserving local directions. The framework requires no auxiliary guidance, training, inversion, additional denoiser evaluations, or backpropagation.

we retain the temporally aggregated editing directions and use endpoint agreement solely to determine their relative spatial strengths. Thus, endpoint consistency serves as an intrinsic reliability signal for the control field rather than a spatial localization map.

Based on this principle, we propose ER-Chord, a training-free and inversion-free framework for reliability-guided spatial redistribution. ER-Chord preserves the local directions of the Chord field and maintains the total control energy, modifying only how this fixed budget is distributed across spatial locations. We measure local reliability through the agreement between the two Chord endpoint fields and formulate the reallocation as a reliability-weighted quadratic optimization problem subject to an exact equal-energy constraint. This constraint prevents trivial source preservation caused by simply reducing the global editing strength. Although this formulation couples all spatial locations through a shared energy budget, it yields a unique global solution that rescales each local Chord vector by a positive scalar. The entire solution depends on a single scalar variable that can be found efficiently through one-dimensional bisection. ER-Chord therefore requires no additional denoiser evaluations, auxiliary training, or backpropagation.

The equal-energy formulation also provides a direct interpretation of the resulting redistribution. Locations with higher endpoint consistency receive larger scaling factors relative to their original Chord vectors, while a shared scalar coordinates these adjustments to preserve the total control energy exactly. Hence, ER-Chord modifies only the relative spatial allocation of energy without altering local control directions or the overall update scale (Figure 1). As shown in Figures 2 and 3, under identical total control energy, ChordEdit disperses editing strength into irrelevant backgrounds, causing color bleeding and artifacts. In contrast, ER-Chord concentrates energy into semantically consistent regions, producing clean target edits while faithfully preserving source structures.

Empirically, ER-Chord consistently outperforms one-step editing baselines across diverse benchmarks using the same denoiser budget. Without tuning hyperparameters for specific datasets or backbones, our configuration substantially improves background preservation on PIE-Bench, generalizes effectively to other benchmarks such as TEdBench in perceptual quality and source fidelity, and transfers across multiple generative backbones (InstaFlow, SwiftBrush-v2, and SD-Turbo). These results establish reliability-guided spatial redistribution as an effective, orthogonal complement to temporal stabilization.

Our main contributions are summarized as follows:

- We identify spatial magnitude allocation as an unaddressed degree of freedom in temporally aggregated one-step editing, demonstrating that temporal stability alone does not guarantee spatial consistency.
- We propose ER-Chord, a framework that uses endpoint consistency as an intrinsic reliability signal to redistribute total control energy under an exact equal-energy constraint, which is solved globally through a one-dimensional scalar root.

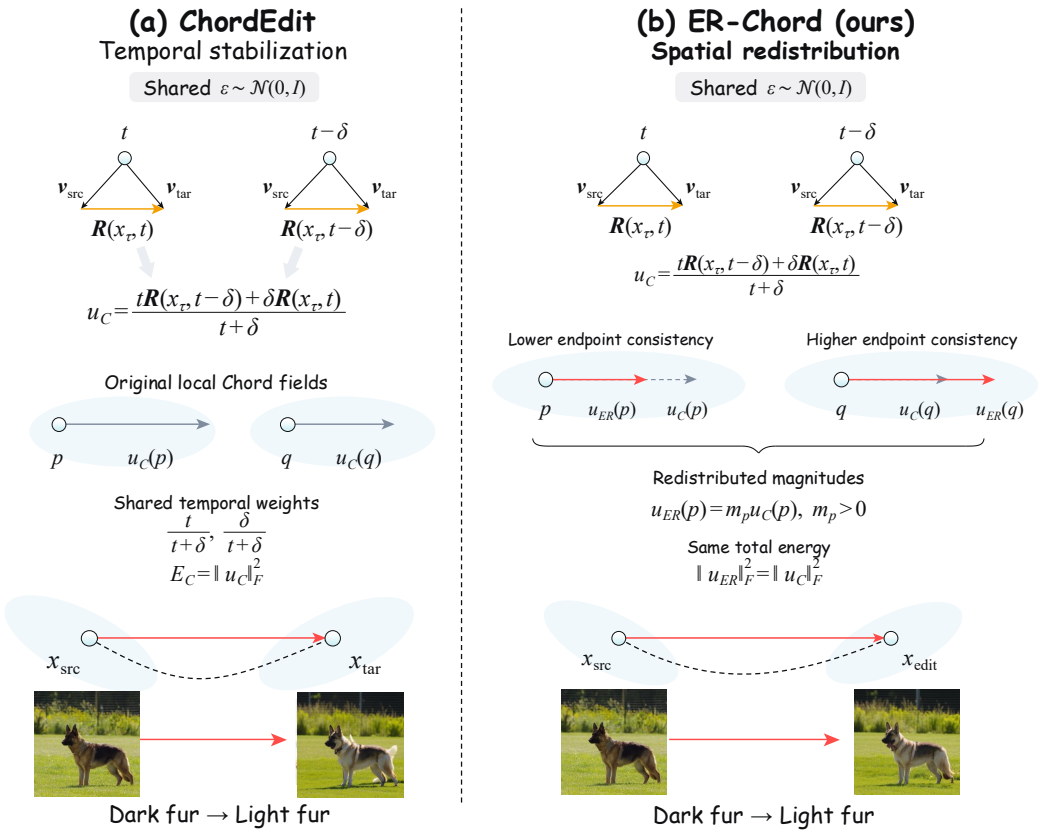

Figure 2: **Temporal vs. spatial control in one-step editing.** (a) ChordEdit temporally aggregates endpoint fields with uniform weights across space, which fails to reflect local differences in prediction consistency and disperses edit energy into irrelevant backgrounds. (b) ER-Chord measures local endpoint consistency as a reliability signal and redistributes control energy ($u_{\mathrm{ER}}(p) = m_p u_{\mathrm{C}}(p)$) under an exact conserved energy budget ($\|u_{\mathrm{ER}}\|_F^2 = \|u_{\mathrm{C}}\|_F^2$), concentrating updates onto semantically intended target regions without background artifacts.

- We demonstrate on PIE-Bench, TEdBench, and three generative backbones that ER-Chord consistently improves source preservation while maintaining semantic edit fidelity, requiring no external guidance, training, or additional denoiser calls.

2 Related Work

Real-image editing and inversion. Text-guided editing must introduce the requested semantics without losing the identity and layout of the source image. SDEdit starts reverse diffusion from a partially noised source, while DiffEdit estimates an edit mask from differences between conditional predictions (Meng et al., 2022; Couairon et al., 2023). Prompt-to-Prompt controls cross-attention, and MasaCtrl and PnP reuse self-attention or intermediate features from a source branch to preserve structure (Hertz et al., 2023; Cao et al., 2023; Tumanyan et al., 2023). Inversion-based methods instead seek a latent trajectory that accurately reconstructs the source. Null-text Inversion optimizes unconditional text embeddings, EDICT introduces coupled transformations for exact inversion, and subsequent methods reduce the mismatch accumulated along inversion and denoising trajectories (Mokady et al., 2023; Wallace et al., 2023; Ju et al., 2024; Huberman-Spiegelglas et al., 2024; Yang & Ye, 2026). These approaches improve fidelity by modifying the inversion path, internal features, or spatial masks. ER-Chord instead operates after a one-step source-to-target control field has been formed and changes only its local magnitudes, without reconstructing an inversion trajectory or modifying network attention and features.

Few-step and one-step editing. Consistency-based and flow-based models have made high-quality synthesis possible with one or a few neural function evaluations (Song & Dhariwal, 2024; Luo et al., 2023; Lipman et al., 2023; Liu et al., 2023). This development has motivated editors designed specifically for low-step samplers. TurboEdit and InstantEdit pair specialized inversion with few-step reconstruction, whereas SwiftEdit learns a one-step inversion model and uses mask-guided attention rescaling for localized editing (Deutch et al., 2024; Gong et al., 2025; Nguyen et al., 2025). BiFM jointly learns forward and inverse flows, FlowEdit constructs an inversion-free transport from pretrained flow models, and EEdit accelerates inversion-based editors by exploiting redundant computation across space and time (Dai et al., 2026; Kulikov et al., 2025; Yan et al., 2025). ChordEdit is the closest starting point for our work. It is training-free and inversion-free, and stabilizes a one-step update by aggregating source-to-target fields across nearby noise levels (Lu et al., 2026). A subsequent analysis interprets its chord window largely as an effective timestep shift and separates its procedure into coarse transport and fine alignment (Li et al., 2026). ER-Chord leaves this temporal estimator and cleanup unchanged, introduces no additional model evaluations, and only redistributes the resulting field magnitude over space.

Spatial localization and magnitude allocation. Spati-ally precise editors generally identify where an edit belongs and then restrict generation in that region. DiffEdit derives a mask from conditional prediction differences, while Learnable Regions and ZONE obtain regions through learned boxes or segmentation (Couairon et al., 2023; Lin et al., 2024; Li et al., 2024). FoI, LIME, and LOCATEdit extract or refine attention-based regions and use them to constrain the denoising process (Guo & Lin, 2024; Simsar et al., 2025; Soni et al., 2025). ALE-Edit further limits attribute leakage through object-restricted text embeddings, region-guided attention masking, and background blending (Mun et al., 2025). Localization has also been incorporated into one-step editing. SwiftEdit uses mask-guided attention rescaling, while WhereEdit discovers editable regions from internal features and applies local modulation (Nguyen et al., 2025; Hu et al., 2026). These methods use semantic or structural evidence to decide where a modification should occur. ER-Chord neither predicts an editable region nor changes the semantic direction of the edit. It uses endpoint consistency only to rescale the magnitudes of fixed local Chord directions under an unchanged per-image energy budget. Localization methods determine where an edit should act, whereas ER-Chord adjusts the strength of the existing Chord field across locations.

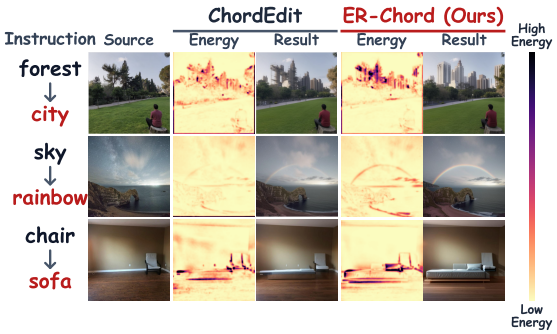

Figure 3: **Spatial energy redistribution under conserved total energy.** Under identical conserved control energy ($1.00\times$), ChordEdit disperses energy into backgrounds causing artifacts, while ER-Chord concentrates energy on target regions for clean edits.

3 METHOD

We develop ER-Chord in three steps. We first revisit the temporally aggregated Chord field and identify spatial magnitude allocation as a degree of freedom that is not explicitly controlled by the temporal estimator. We then use local endpoint consistency as a reliability signal and formulate spatial reweighting as an equal-energy redistribution problem, which preserves both the original control directions and the total control strength. Finally, we show that this seemingly coupled optimization admits a simple direction-preserving solution determined by a single shared scalar, enabling an efficient implementation with negligible additional computation.

3.1 FROM TEMPORAL AGGREGATION TO SPATIAL ALLOCATION

Let I_{src} denote the source image, and let c_{src} and c_{tar} denote the source and target conditions, respectively. Given a latent state $\mathbf{x}_\tau$, we follow ChordEdit Lu et al. (2026) and denote by $\mathbf{R}(\mathbf{x}_\tau, t)$

the source-to-target proxy field constructed at noise time t. In particular, for the two neighboring endpoints $t-\delta$ and t, we write $\mathbf{R}_{t-\delta} := \mathbf{R}(\mathbf{x}_\tau, t-\delta)$ and $\mathbf{R}_t := \mathbf{R}(\mathbf{x}_\tau, t)$.

ChordEdit temporally aggregates these two endpoint fields to obtain a stabilized control field. At each spatial location $p \in \Omega$, its control vector can be written as

$$\mathbf{u}_{\mathrm{C}}(p) = \frac{t}{t+\delta}\mathbf{R}_{t-\delta}(p) + \frac{\delta}{t+\delta}\mathbf{R}_t(p). \tag{1}$$

The coefficients in (1) are global temporal scalars and do not depend on the spatial location p. The corresponding derivation from the Chord surrogate is provided in Appendix A.1.

Equation (1) provides a temporally stabilized editing vector at every spatial location. However, the resulting spatial magnitude $\|\mathbf{u}_{\mathrm{C}}(p)\|_2$ is only an implicit consequence of the temporal aggregation. In particular, the aggregation rule does not explicitly compare the magnitude assigned to location p with the agreement between $\mathbf{R}_{t-\delta}(p)$ and $\mathbf{R}_t(p)$. Therefore, two locations with substantially different endpoint consistency are still governed by the same temporal mixing rule.

This observation suggests separating temporal stabilization from spatial magnitude allocation. We retain the local directions supplied by $\mathbf{u}_{\mathrm{C}}$ and treat their spatial magnitudes as the remaining degree of freedom. Our goal is to redistribute these magnitudes according to local endpoint consistency while preserving the total control strength of the original Chord field. Section 3.2 develops this reliability-guided equal-energy redistribution.

3.2 Reliability-Guided Equal-Energy Redistribution

We next determine how the spatial magnitudes of the Chord field should be reallocated. The key signal is the agreement between the two endpoint fields. At a spatial location $p \in \Omega$, we use $s_p := \|\mathbf{u}_{\mathrm{C}}(p)\|_2$ to denote the magnitude of the temporally aggregated Chord control and $d_p := \|\mathbf{R}_t(p) - \mathbf{R}_{t-\delta}(p)\|_2$ to measure the disagreement between the two endpoint estimates. Since the same absolute endpoint discrepancy can have different significance at locations with different control magnitudes, we normalize the disagreement relative to the local Chord signal.

Specifically, we define the local endpoint-consistency score as

$$c_p = \mathrm{clip}_{[0,1]}\left(\frac{s_p}{\max\{s_p + \kappa d_p, \varepsilon_c\}}\right), \qquad \kappa \geq 0,\ \varepsilon_c > 0. \tag{2}$$

A larger c_p indicates stronger agreement between the two endpoint fields relative to the local control magnitude. In particular, c_p approaches one when the endpoint disagreement is small compared with s_p, and decreases as the relative disagreement grows. This score measures local temporal consistency of the estimated control field rather than semantic correctness.

We use this reliability signal only to redistribute spatial magnitude. The local Chord directions remain unchanged as the reference editing directions, and the total control strength is kept fixed. The latter constraint is important because otherwise improved source preservation could be obtained trivially by globally weakening the edit. Let $E_{\mathrm{C}} := \|\mathbf{u}_{\mathrm{C}}\|_F^2$ denote the total pre-cleanup control energy of the Chord field. We seek a redistributed field that remains close to $\mathbf{u}_{\mathrm{C}}$, while assigning a larger cost to control magnitude placed at locations with lower endpoint consistency.

These design requirements lead to the following equal-energy redistribution:

$$\begin{aligned} \mathbf{u}_{\mathrm{ER}} \in \arg\min_{\mathbf{u}} \quad & \frac{1}{2}\sum_{p\in\Omega}\|\mathbf{u}(p) - \mathbf{u}_{\mathrm{C}}(p)\|_2^2 + \frac{\mu}{2}\sum_{p\in\Omega}(1-c_p)^2\|\mathbf{u}(p)\|_2^2 \\ \text{subject to} \quad & \|\mathbf{u}\|_F^2 = E_{\mathrm{C}}, \qquad 0 \leq \mu < 1. \end{aligned} \tag{3}$$

The first term preserves fidelity to the original Chord field, while the second term discourages control energy from remaining at locations with low endpoint consistency. The equality constraint keeps the total pre-cleanup energy exactly unchanged. Consequently, any reduction of control magnitude at one subset of spatial locations must be compensated by an increase elsewhere.

Although (3) couples all spatial locations through a global energy constraint, its structure is considerably simpler than a generic constrained field optimization. In the next subsection, we show that its unique global solution preserves every local Chord direction and reduces the entire redistribution to a one-dimensional scalar solve.

3.3 SCALAR-ROOT SOLUTION AND PROPERTIES

The equal-energy constraint in (3) couples all spatial locations through a single global budget. Nevertheless, the optimization admits a particularly simple structure. For each $p \in \Omega$, define $a_p := 1 + \mu(1 - c_p)^2$, which represents the local quadratic penalty induced by the endpoint-consistency score. The spatial dependence is entirely contained in a_p, while the global energy constraint is enforced by one shared scalar.

Theorem 3.1 (Unique direction-preserving solution)**.** *Assume $0 \le \mu < 1$ and $E_{\mathrm{C}} > 0$. Then the optimization problem* (3) *admits a unique global minimizer. In particular, there exists a unique scalar $\eta^\star \in [-\mu, 0]$ such that*

$$\mathbf{u}_{\mathrm{ER}}(p) = \frac{\mathbf{u}_{\mathrm{C}}(p)}{a_p + \eta^\star}, \qquad p \in \Omega, \tag{4}$$

where $\eta^\star$ is the unique root of

$$\sum_{p \in \Omega} \frac{\|\mathbf{u}_{\mathrm{C}}(p)\|_2^2}{(a_p + \eta)^2} = E_{\mathrm{C}}. \tag{5}$$

Theorem 3.1 reduces the spatially coupled optimization to a one-dimensional scalar solve. The local reliability scores determine the spatially varying denominators a_p, while the single shared variable $\eta^\star$ adjusts all locations jointly so that the original control energy is exactly preserved. Since $a_p + \eta^\star > 0$, the solution in (4) rescales each local Chord vector without changing its direction. The proof of existence, uniqueness, and global optimality is provided in Appendix A.2.

To make the redistribution explicit, let $m_p := (a_p + \eta^\star)^{-1}$ denote the local spatial multiplier. Then $\mathbf{u}_{\mathrm{ER}}(p) = m_p \mathbf{u}_{\mathrm{C}}(p)$. The following properties summarize the behavior of the resulting field.

Corollary 3.2 (Energy and direction preservation)**.** *Under the assumptions of Theorem 3.1, the ER-Chord solution satisfies*

$$\|\mathbf{u}_{\mathrm{ER}}\|_F^2 = \|\mathbf{u}_{\mathrm{C}}\|_F^2. \tag{6}$$

Moreover, $m_p > 0$ for every $p \in \Omega$, and

$$\frac{1}{1+\mu} \le m_p \le \frac{1}{1-\mu}.$$

If $c_p > c_q$ and $\mu > 0$, then $m_p > m_q$.

The corollary clarifies the role of the redistribution. ER-Chord neither rotates the local Chord directions nor changes the total pre-cleanup control strength. Instead, it modifies only the relative spatial magnitudes. Locations with higher endpoint consistency receive a larger multiplicative adjustment than locations with lower endpoint consistency. This ordering applies to the multipliers themselves rather than to the absolute final energy shares, which also depend on the original Chord magnitudes.

The scalar $\eta^\star$ can be obtained efficiently by bisection over $[-\mu, 0]$. For a latent field of size $C \times H \times W$ and K bisection iterations, the additional computation is $\mathcal{O}(K \cdot C \cdot H \cdot W)$ and consists only of channel reductions and element-wise arithmetic. ER-Chord reuses the endpoint fields already required by ChordEdit and therefore introduces no additional denoiser evaluations.

To further characterize the spatial redistribution, we compare the fraction of the total control energy assigned to each location before and after ER-Chord. For $p \in \Omega$, define the original and redistributed spatial energy shares as $\pi_p^{\mathrm{C}} := \|\mathbf{u}_{\mathrm{C}}(p)\|_2^2 / E_{\mathrm{C}}$ and $\pi_p^{\mathrm{ER}} := \|\mathbf{u}_{\mathrm{ER}}(p)\|_2^2 / E_{\mathrm{C}}$, respectively. Since $\mathbf{u}_{\mathrm{ER}}(p) = m_p \mathbf{u}_{\mathrm{C}}(p)$, these shares satisfy

$$\pi_p^{\mathrm{ER}} = m_p^2 \pi_p^{\mathrm{C}}. \tag{7}$$

Equation (7) makes the redistribution explicit: ER-Chord modifies the original spatial energy allocation only through the local multipliers, while the total energy remains unchanged.

The local consistency scores and the shared scalar $\eta^\star$ play complementary roles: the scores determine the relative spatial penalties, while $\eta^\star$ enforces the common energy budget.

Importantly, the equal-energy constraint does not require every low-consistency location to be attenuated or every high-consistency location to be amplified relative to one. Several multipliers may

Table 1: **Editing results on PIE-Bench** (Ju et al., 2024). MSE and LPIPS are scaled by 10^3. Best, second-best, and third-best results are highlighted with green, yellow, and purple backgrounds, respectively. T-/I-free: training-/inversion-free. Extended metrics appear in Appendix D.

		Background Preservation			Target Alignment		Properties		Efficiency			
Type	Method	PSNR↑	MSE↓	LPIPS↓	Whole↑	Edit↑	T-free	I-free	Steps↓	NFE↓	Time(s)↓	VRAM(M)↓
Multi-step	DDIM + MasaCtrl (2023)	21.25	8.58	106.59	24.13	21.13	✓	✗	50	100	55.20	12272
	Direct Inv. + MasaCtrl (2023)	21.78	7.99	87.38	24.42	21.38	✓	✗	50	100	79.10	12272
	DDIM + PnP (2023)	21.26	8.42	113.58	25.45	22.54	✓	✗	50	100	28.01	9262
	Direct Inv. + PnP (2023)	21.43	8.10	106.26	25.48	22.63	✓	✗	50	100	28.03	9262
	FlowEdit (SD3) (2025)	22.17	7.69	104.81	26.64	23.69	✓	✓	33	33	7.22	17140
Few-step	TurboEdit (SDXL-T) (2024)	21.44	9.49	108.60	24.66	21.79	✓	✓	4	4	2.69	13826
	InfEdit (SD1.4) (2024)	24.14	6.82	55.69	24.89	21.88	✓	✓	4	4	1.41	6502
	InstantEdit (PeRFlow) (2025)	23.80	4.21	60.92	24.97	21.82	✓	✗	4	8	1.30	16270
One-step	SwiftEdit (SB-v2) (2025)	21.71	8.22	91.22	24.93	21.85	✗	✗	1	2	0.54	15060
	ChordEdit (SB-v2) (2026)	22.04	7.13	111.22	25.12	22.58	✓	✓	1	2	0.38	6988
	ChordEdit (w/o clean, SD-T)	23.89	5.05	88.36	24.97	21.87	✓	✓	1	1	0.20	6988
	ChordEdit (SD-Turbo) (2026)	22.20	6.84	128.25	25.58	22.96	✓	✓	1	2	0.38	6988
	ER-Chord (SB-v2)	22.48	6.62	105.10	25.26	22.74	✓	✓	1	2	0.37	6988
	ER-Chord (w/o clean, SD-T)	24.31	4.70	82.82	25.04	21.99	✓	✓	1	1	0.20	6988
	ER-Chord (SD-Turbo)	22.77	6.35	120.23	25.74	23.16	✓	✓	1	2	0.37	6988

lie on the same side of one, provided that their energy-weighted adjustments collectively preserve E_C. Therefore, ER-Chord should be interpreted as a globally coupled spatial reallocation rather than an independent pointwise suppression rule.

The same optimization also yields an additional certificate showing that the redistributed field places no more reliability-weighted energy on inconsistent locations than the original Chord field. We provide the formal statement and proof in Appendix A.4.

4 Experiments

4.1 Evaluation protocol

Benchmarks. PIE-Bench (Ju et al., 2024) contains 700 masked editing examples in ten categories and provides the primary comparison with ChordEdit. We additionally evaluate the original 100-example TEdBench (Kawar et al., 2023) as an external benchmark for text-guided real-image editing. For ChordEdit and ER-Chord, we use the same target-only protocol with identical inputs, target captions, random seeds, and evaluation settings.

Metrics. On PIE-Bench, background preservation uses non-edited-region PSNR, MSE, SSIM, LPIPS (Zhang et al., 2018), and structure distance; semantic alignment uses CLIP-Whole and CLIP-Edited (Radford et al., 2021). On TEdBench, we report LPIPS for perceptual source preservation, CLIP-I for image-level source similarity, CLIP-T for target text–image similarity, and Dir-Sim for directional similarity. Every ER run also records the dual interval, normalized dual residual, equal-energy residual, KKT residual, multiplier bounds, direction residual, and inequality (26).

Implementation and statistics. We retain the ChordEdit configuration: one shared noise sample, $t = 0.90$, $\delta = 0.15$, $\lambda = 1.0$, and cleanup time $t_c = 0.30$. For ER-Chord, we use $\mu = 0.25$, $\kappa = 1.0$, $\varepsilon_c = 10^{-8}$, an energy threshold of 10^{-12}, a dual tolerance of 10^{-6}, and at most 32 bisection iterations. These values are not tuned on the external benchmarks. The comparisons between ChordEdit and ER-Chord are paired by source example. Runtime and peak memory are measured with the same hardware, precision, batch size, and warm-up.

4.2 Comparison on PIE-Bench

Full PIE-Bench evaluation. Table 1 reproduces the comparison from ChordEdit Table 1 and adds ER-Chord under the same SD-Turbo configuration with $(\mu, \kappa) = (0.25, 1.0)$, one shared noise sample, and all 700 PIE-Bench examples (seed 42). The comparison covers all 700 PIE-Bench

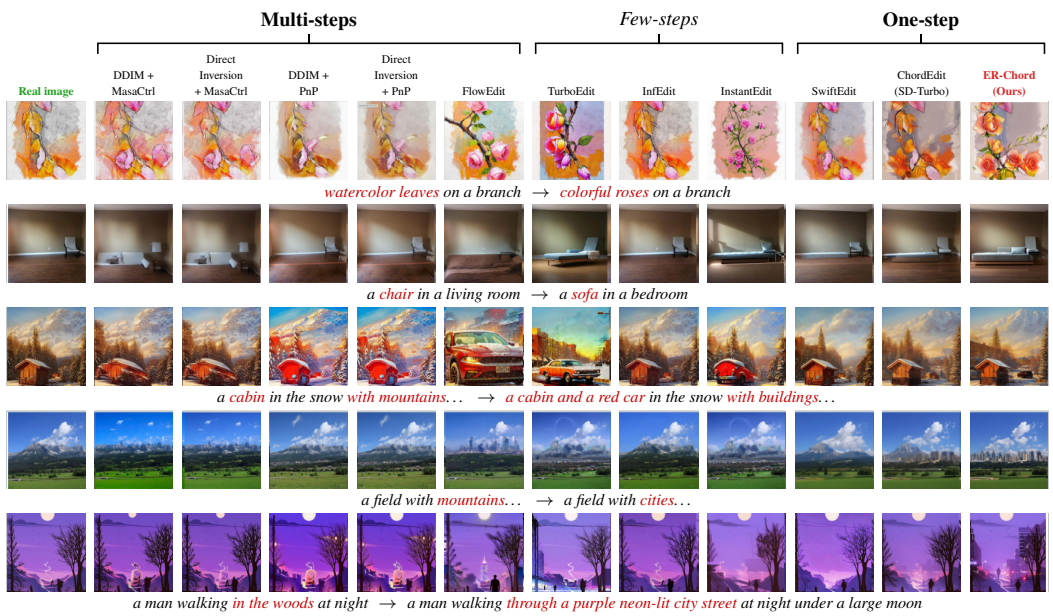

Figure 4: **Qualitative comparison on PIE-Bench.** Real images are in the first column. Multi-step, few-step, and one-step methods are arranged horizontally. Prompts are noted under each row. ChordEdit and ER-Chord use the identical SD-Turbo backbone.

examples across the ten predefined editing categories. As shown in Table 1, the primary benefit of ER-Chord lies in substantial gains in source content preservation. Specifically, ER-Chord improves unedited-region PSNR from 22.20 to 22.77 dB (+0.57 dB), lowers MSE from 6.84 to 6.35, and reduces perceptual distortion LPIPS from 128.25 to 120.23 ($10^3\times$ scaling). Detailed structural preservation metrics (e.g., SSIM and Structure Distance) exhibit concordant improvements in the appendix (Table 6). Crucially, these preservation gains are accompanied by improved target semantic alignment: CLIP-Whole rises from 25.58 to 25.74, and CLIP-Edited from 22.96 to 23.16. The two methods have the same total pre-cleanup control energy by construction. Thus, the gains cannot be attributed to uniformly reducing the control field; they are consistent with a more effective spatial allocation.

Qualitative comparison. In the examples shown in Figure 4, several multi-step and few-step editors alter the background or blur fine details. ChordEdit also exhibits visible background changes in some cases. With reliability-guided redistribution, ER-Chord produces cleaner target edits in these examples, including the sofa and mountain-to-building transformations, while retaining more of the source background.

4.3 Ablations, Cross-Backbone Transfer, and Efficiency

Transport vs. cleanup ablation. To understand how spatial redistribution interacts with the multi-stage pipeline, Table 2 evaluates ER-Chord with and without the post-hoc cleanup stage. In the no-cleanup setting (NFE $= 1$), the evaluation directly isolates the spatial transport field. Here, ER-Chord improves PSNR from 23.89 to 24.31 dB and CLIP-Edited from 21.87 to 21.99 over the raw Chord field, confirming that reliability reallocation strengthens transport dynamics before cleanup. With cleanup (NFE $= 2$), ER-Chord maintains its advantage (22.77 vs. 22.20 dB PSNR; 23.16 vs. 22.96 CLIP-Edited), orthogonally complementing post-hoc temporal cleanup.

Core mechanism ablations. To isolate the necessity of equal-energy optimization, we test two baselines. First, a heuristic "Reweight-and-Normalize (RN)" directly scales the Chord field by c_p and normalizes total energy, achieving only 22.40 dB PSNR (vs. ER-Chord 22.77 dB). Second, "Spatial Permutation" applies the exact optimal multipliers from ER-Chord but randomly permutes

Table 2: **Transport & cleanup ablation.** PIE-Bench. NFE=1: transport; NFE=2: full. Bold: best.

Configuration		PIE-Bench metrics		
Field	Cleanup	PSNR↑	CLIP-Edited↑	NFE
Naive ($\delta = 0$)	no	21.89	20.83	1
Chord ($\delta = .15$)	no	23.89	21.87	1
ER-Chord	no	**24.31**	**21.99**	**1**
Naive ($\delta = 0$)	yes	21.38	21.96	2
Chord ($\delta = .15$)	yes	22.20	22.96	2
ER-Chord	yes	**22.77**	**23.16**	2

Table 3: **TEdBench comparison** (Kawar et al., 2023). 100 real-image prompts. Bold: best.

Method	Paradigm	LPIPS↓	CLIP-I↑	CLIP-T↑	Dir-Sim↑
SDEdit	Multi-step	0.30	0.85	0.22	0.15
Pix2Pix-Zero	Multi-step	0.29	0.84	0.23	0.14
Imagic	Optimization	0.52	0.86	0.25	0.20
LEDITS++	Inversion	0.23	0.87	0.26	0.21
FlowEdit	Flow	0.22	0.89	0.26	0.22
ChordEdit	One-step	0.21	0.91	0.28	**0.25**
ER-Chord (Ours)	One-step	**0.20**	**0.92**	**0.29**	**0.25**

their spatial assignment. The PSNR drops to 21.05 dB, supporting the importance of matching the multipliers to their original spatial locations.

Transfer to TEdBench. To evaluate out-of-domain generalization, we benchmark on 100 diverse real-image editing cases in TEdBench (Kawar et al., 2023) featuring complex attribute modifications and diverse natural scenes. As shown in Table 3, ER-Chord improves upon ChordEdit in both perceptual source preservation (LPIPS decreases from 0.21 to 0.20) and image fidelity (CLIP-I increases from 0.91 to 0.92). On the reported TEdBench metrics, ER-Chord also improves over the listed multi-step and optimization-based baselines while using one transport update. These results suggest transfer to the tested scenes without benchmark-specific tuning.

Cross-backbone generalization. To assess transfer across backbones, Table 4 evaluates ER-Chord with InstaFlow, SwiftBrush-v2, and SD-Turbo. With fixed hyperparameters ($(\mu, \kappa) = (0.25, 1.0)$), ER-Chord improves PSNR over ChordEdit by $+0.35$ dB on InstaFlow, $+0.44$ dB on SwiftBrush-v2, and $+0.57$ dB on SD-Turbo. CLIP-Edited also improves on all three tested backbones.

Table 4: **Cross-backbone generalization.** Results across backbones. Bold: best per backbone.

Backbone	Method	PSNR↑	CLIP-Edited↑
InstaFlow	Naive	22.05	20.19
	ChordEdit	23.05	21.39
	ER-Chord (Ours)	**23.40**	**21.54**
SwiftBrush-v2	Naive	20.52	21.06
	ChordEdit	22.04	22.58
	ER-Chord (Ours)	**22.48**	**22.74**
SD-Turbo	Naive	21.38	21.96
	ChordEdit	22.20	22.96
	ER-Chord (Ours)	**22.77**	**23.16**

Computational overhead and convergence. Because ER-Chord reuses the ChordEdit endpoints $(\mathbf{R}_t, \mathbf{R}_{t-\delta})$, it adds no network evaluations relative to ChordEdit. The full pipeline uses NFE $= 2$ with cleanup and NFE $= 1$ without it, and the measured peak memory is unchanged at 6988 MiB. The scalar bisection terminates within 32 iterations and takes < 0.01 s CPU time. The measured full-pipeline runtimes are similar: 0.37 s/img for ER-Chord and 0.38 s/img for ChordEdit.

5 Conclusion

This paper investigates spatial magnitude allocation in training-free, one-step image editing. Existing transport methods stabilize updates across noise levels using temporal weights shared across spatial locations, leaving local reliability unaccounted for in the allocation of editing strength. We propose ER-Chord, which uses endpoint agreement to redistribute control energy under an exact pre-cleanup energy constraint. The resulting field preserves the original local directions and has a unique global solution determined by a scalar root, with no additional denoiser evaluations relative to ChordEdit. Experiments on PIE-Bench, TEdBench, and three generative backbones (SD-Turbo, SwiftBrush-v2, and InstaFlow) show improved source preservation alongside target alignment in the tested settings. These results suggest that spatial allocation is a useful complement to temporal stabilization. The equal-energy constraint prevents a trivial improvement from globally weakening the edit: energy reduced in one location must be reallocated to another. The endpoint-consistency signal determines that reallocation, while a shared scalar root keeps the computation small. Because the method rescales the Chord field without changing its local directions, it cannot correct an erroneous direction supplied by the underlying field. Testing how this dependence changes across additional editing tasks and backbones is a useful direction for future work.

AI USE STATEMENT

In this work, we used generative AI tools to improve the clarity, grammar, and presentation of the manuscript. All AI-assisted content was carefully reviewed and verified by the authors. All scientific ideas, methodological designs, experimental results, and conclusions were developed and validated by the authors. We take full responsibility for the final content of this work.

REFERENCES

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 *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 22503–22513. IEEE, 2023.

Guillaume Couairon, Jakob Verbeek, Holger Schwenk, and Matthieu Cord. DiffEdit: Diffusion-based semantic image editing with mask guidance. In *International Conference on Learning Representations (ICLR)*, 2023.

Yasong Dai, Zeeshan Hayder, David Ahmedt-Aristizabal, and Hongdong Li. BiFM: Bidirectional flow matching for few-step image editing and generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23325–23334, 2026.

Trung Dao, Thuan Hoang Nguyen, Thanh Le, Duc Vu, Khoi Nguyen, Cuong Pham, and Anh Tran. Swiftbrush v2: Make your one-step diffusion model better than its teacher. In *European Conference on Computer Vision*, pp. 176–192. Springer, 2024.

Gilad Deutch, Rinon Gal, Daniel Garibi, Or Patashnik, and Daniel Cohen-Or. Turboedit: Text-based image editing using few-step diffusion models. In *SIGGRAPH Asia 2024 Conference Papers*, pp. 1–12, 2024.

Yiming Gong, Zhen Zhu, and Minjia Zhang. Instantedit: Text-guided few-step image editing with piecewise rectified flow. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 16808–16817. IEEE, 2025.

Qin Guo and Tianwei Lin. Focus on your instruction: Fine-grained and multi-instruction image editing by attention modulation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6986–6996. IEEE, 2024.

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 (ICLR)*, 2023.

Ming Hu, Mingyu Dou, Jianfu Yin, Miaomiao Zhang, Cong Hu, Yao Wang, Bingliang Hu, and Quan Wang. Whereedit: Mask-aware local latent editing for one-step image editing. *arXiv preprint arXiv:2607.20883*, 2026.

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

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 (ICLR)*, 2024.

Bahjat Kawar, Shiran Zada, Oran Lang, Omer Tov, Huiwen Chang, Tali Dekel, Inbar Mosseri, and Michal Irani. Imagic: Text-based real image editing with diffusion models. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6007–6017. IEEE, 2023.

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

Minghan Li, Jeremy Moebel, and Mengyu Wang. Rethinking one-step image editing through chordedit: Reproduction, simplification, and new insights. *arXiv preprint arXiv:2606.14042*, 2026.

Shanglin Li, Bohan Zeng, Yutang Feng, Sicheng Gao, Xiuhui Liu, Jiaming Liu, Lin Li, Xu Tang, Yao Hu, Jianzhuang Liu, et al. Zone: Zero-shot instruction-guided local editing. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6254–6263. IEEE, 2024.

Yuanze Lin, Yi-Wen Chen, Yi-Hsuan Tsai, Lu Jiang, and Ming-Hsuan Yang. Text-driven image editing via learnable regions. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7059–7068. IEEE, 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 (ICLR)*, 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 (ICLR)*, 2023.

Xingchao Liu, Xiwen Zhang, Jianzhu Ma, Jian Peng, and Qiang Liu. InstaFlow: One step is enough for high-quality diffusion-based text-to-image generation. In *International Conference on Learning Representations (ICLR)*, 2024.

Liangsi Lu, Xuhang Chen, Minzhe Guo, Shichu Li, Jingchao Wang, and Yang Shi. ChordEdit: One-step low-energy transport for image editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14398–14407, 2026.

Simian Luo, Yiqin Tan, Longbo Huang, Jian Li, and Hang Zhao. Latent consistency models: Synthesizing high-resolution images with few-step inference. *arXiv preprint arXiv:2310.04378*, 2023.

Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. SDEdit: Guided image synthesis and editing with stochastic differential equations. In *International Conference on Learning Representations (ICLR)*, 2022.

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

Sunung Mun, Jinhwan Nam, Sunghyun Cho, and Jungseul Ok. Addressing text embedding leakage in diffusion-based image editing. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 16451–16460. IEEE, 2025.

Trong-Tung Nguyen, Quang Nguyen, Khoi Nguyen, Anh Tran, and Cuong Pham. Swiftedit: Lightning fast text-guided image editing via one-step diffusion. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21492–21501. IEEE, 2025.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial diffusion distillation. In *European Conference on Computer Vision*, pp. 87–103. Springer, 2024.

Enis Simsar, Alessio Tonioni, Yongqin Xian, Thomas Hofmann, and Federico Tombari. Lime: Localized image editing via attention regularization in diffusion models. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 222–231. IEEE, 2025.

Yang Song and Prafulla Dhariwal. Improved techniques for training consistency models. In *International Conference on Learning Representations (ICLR)*, 2024.

Achint Soni, Meet Soni, and Sirisha Rambhatla. Locatedit: Graph laplacian optimized cross attention for localized text-guided image editing. *arXiv preprint arXiv:2503.21541*, 2025.

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

Bram Wallace, Akash Gokul, and Nikhil Naik. Edict: Exact diffusion inversion via coupled transformations. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22532–22541. IEEE, 2023.

Zexuan Yan, Yue Ma, Chang Zou, Wenteng Chen, Qifeng Chen, and Linfeng Zhang. EEdit: Rethinking the spatial and temporal redundancy for efficient image editing. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 17474–17484. IEEE, 2025.

Desong Yang and Mang Ye. DirectEdit: Step-level accurate inversion for flow-based image editing. In *International Conference on Machine Learning (ICML)*, 2026.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *2018 IEEE/CVF conference on computer vision and pattern recognition*, pp. 586–595. IEEE, 2018.

A Mathematical Derivations and Proofs

This appendix provides the mathematical details supporting the construction in Section 3. We first derive the temporally aggregated Chord field used as the starting point of our method. We then establish the existence, uniqueness, and global optimality of the equal-energy redistribution solution, followed by the properties implied by its closed-form structure. These results complete the technical arguments omitted from the main text while preserving the method-level presentation in Section 3.

A.1 Derivation of the Temporally Aggregated Chord Field

For completeness, we derive the temporally aggregated Chord field used in Section 3. Following ChordEdit Lu et al. (2026), let $\widehat{\mathbf{u}}_{t-\delta}$ denote the endpoint control estimate entering the temporal surrogate at time $t-\delta$. The Chord estimator considers the quadratic objective

$$\Phi_t(\mathbf{u};\mathbf{x}_\tau) = t\|\mathbf{u}-\widehat{\mathbf{u}}_{t-\delta}\|_F^2 + \int_{t-\delta}^{t} \|\mathbf{u}-\mathbf{R}_\xi\|_F^2 \,\mathrm{d}\xi, \tag{8}$$

where $\mathbf{R}_\xi := \mathbf{R}(\mathbf{x}_\tau,\xi)$ denotes the source-to-target proxy field at noise time ξ.

The objective in (8) is strictly convex with respect to $\mathbf{u}$ because its Hessian is $2(t+\delta)\mathbf{I} \succ 0$. Therefore, its unique minimizer is characterized by the first-order condition

$$t\,(\mathbf{u}-\widehat{\mathbf{u}}_{t-\delta}) + \int_{t-\delta}^{t} (\mathbf{u}-\mathbf{R}_\xi)\,\mathrm{d}\xi = \mathbf{0}. \tag{9}$$

Since the integration interval has length δ, rearranging (9) gives the unique minimizer

$$\mathbf{u}_t^\star = \frac{t}{t+\delta}\widehat{\mathbf{u}}_{t-\delta} + \frac{1}{t+\delta}\int_{t-\delta}^{t} \mathbf{R}_\xi \,\mathrm{d}\xi. \tag{10}$$

ChordEdit then approximates the previous endpoint estimate by $\widehat{\mathbf{u}}_{t-\delta} \approx \mathbf{R}_{t-\delta}$ and evaluates the temporal integral using the endpoint approximation $\int_{t-\delta}^{t} \mathbf{R}_\xi \,\mathrm{d}\xi \approx \delta\mathbf{R}_t$. Substituting these approximations into (10) yields

$$\mathbf{u}_{\mathrm{C}} = \frac{t}{t+\delta}\mathbf{R}_{t-\delta} + \frac{\delta}{t+\delta}\mathbf{R}_t. \tag{11}$$

Restricting (11) to any spatial location $p \in \Omega$ recovers the Chord field used in (1). In particular, the coefficients $t/(t+\delta)$ and $\delta/(t+\delta)$ are determined solely by the temporal construction and are independent of p.

A.2 Proof of the Unique Global Solution

We now prove Theorem 3.1. Recall from Section 3.3 that $a_p := 1+\mu(1-c_p)^2$. Since $c_p \in [0,1]$ and $0 \le \mu < 1$, we have $1 \le a_p \le 1+\mu$ for every $p \in \Omega$.

Proof of Theorem 3.1. Let $J(\mathbf{u})$ denote the objective in (3). Introducing a scalar Lagrange multiplier η for the equal-energy constraint, the Lagrangian is

$$\mathcal{L}(\mathbf{u},\eta) = J(\mathbf{u}) + \frac{\eta}{2}\left(\|\mathbf{u}\|_F^2 - E_{\mathrm{C}}\right). \tag{12}$$

The stationarity condition with respect to the local vector $\mathbf{u}(p)$ is

$$(a_p+\eta)\mathbf{u}(p) = \mathbf{u}_{\mathrm{C}}(p), \qquad p \in \Omega. \tag{13}$$

Hence, whenever $a_p+\eta > 0$ for all $p \in \Omega$, any stationary point must satisfy

$$\mathbf{u}(p;\eta) = \frac{\mathbf{u}_{\mathrm{C}}(p)}{a_p+\eta}. \tag{14}$$

It remains to determine the scalar η from the equal-energy constraint. Substituting (14) into $\|\mathbf{u}\|_F^2 = E_{\mathrm{C}}$ gives

$$\phi(\eta) := \sum_{p\in\Omega} \frac{\|\mathbf{u}_{\mathrm{C}}(p)\|_2^2}{(a_p+\eta)^2} - E_{\mathrm{C}} = 0. \tag{15}$$

For any $\eta \in [-\mu, 0]$, the bound $a_p \ge 1$ implies $a_p+\eta \ge 1-\mu > 0$. Therefore, ϕ is well defined and continuously differentiable on $[-\mu,0]$. Moreover,

$$\phi'(\eta) = -2\sum_{p\in\Omega} \frac{\|\mathbf{u}_{\mathrm{C}}(p)\|_2^2}{(a_p+\eta)^3} < 0, \tag{16}$$

where the strict inequality follows from $E_{\mathrm{C}} > 0$. Hence, ϕ is strictly decreasing on this interval.

At the right endpoint, $a_p \geq 1$ gives

$$\phi(0) = \sum_{p\in\Omega} \frac{\|\mathbf{u}_{\mathrm{C}}(p)\|_2^2}{a_p^2} - E_{\mathrm{C}} \leq 0. \tag{17}$$

At the left endpoint, $a_p \leq 1 + \mu$ gives $a_p - \mu \leq 1$, and therefore

$$\phi(-\mu) = \sum_{p\in\Omega} \frac{\|\mathbf{u}_{\mathrm{C}}(p)\|_2^2}{(a_p - \mu)^2} - E_{\mathrm{C}} \geq 0. \tag{18}$$

The intermediate value theorem and the strict monotonicity of ϕ therefore imply the existence of a unique $\eta^\star \in [-\mu, 0]$ satisfying $\phi(\eta^\star) = 0$. Substituting $\eta^\star$ into (14) yields (4), and (15) reduces to (5).

We finally establish global optimality over the complete feasible set, which is the energy sphere $\|\mathbf{u}\|_F^2 = E_{\mathrm{C}}$. Let $\mathbf{A}$ denote the block-diagonal matrix whose block associated with location p is $a_p\mathbf{I}$. Then (13) at $\eta = \eta^\star$ can be written as

$$(\mathbf{A} + \eta^\star\mathbf{I})\mathbf{u}_{\mathrm{ER}} = \mathbf{u}_{\mathrm{C}}. \tag{19}$$

Consider any feasible field $\mathbf{z}$ and define $\mathbf{d} := \mathbf{z} - \mathbf{u}_{\mathrm{ER}}$. Since both $\mathbf{z}$ and $\mathbf{u}_{\mathrm{ER}}$ have energy E_{C}, we have $2\langle \mathbf{u}_{\mathrm{ER}}, \mathbf{d}\rangle_F + \|\mathbf{d}\|_F^2 = 0$. Using this identity together with (19), the objective difference satisfies

$$J(\mathbf{z}) - J(\mathbf{u}_{\mathrm{ER}}) = \frac{1}{2}\left\langle \mathbf{d}, (\mathbf{A} + \eta^\star\mathbf{I})\mathbf{d}\right\rangle_F. \tag{20}$$

Since $a_p \geq 1$ and $\eta^\star \geq -\mu$,

$$\mathbf{A} + \eta^\star\mathbf{I} \succeq (1 - \mu)\mathbf{I} \succ \mathbf{0}. \tag{21}$$

Consequently,

$$J(\mathbf{z}) - J(\mathbf{u}_{\mathrm{ER}}) \geq \frac{1 - \mu}{2}\|\mathbf{d}\|_F^2. \tag{22}$$

The right-hand side is strictly positive whenever $\mathbf{z} \neq \mathbf{u}_{\mathrm{ER}}$. Hence, $\mathbf{u}_{\mathrm{ER}}$ is the unique global minimizer of (3). □

A.3 Properties of the Closed-Form Redistribution

We next establish the properties of the closed-form solution stated in Corollary 3.2. Recall that $m_p := (a_p + \eta^\star)^{-1}$, where $a_p = 1 + \mu(1 - c_p)^2$ and $\eta^\star \in [-\mu, 0]$.

Proof of Corollary 3.2. The exact energy preservation

$$\|\mathbf{u}_{\mathrm{ER}}\|_F^2 = \|\mathbf{u}_{\mathrm{C}}\|_F^2 = E_{\mathrm{C}} \tag{23}$$

follows directly from the equality constraint in (3) and the feasibility of the solution constructed in Theorem 3.1.

Since $c_p \in [0, 1]$, we have $1 \leq a_p \leq 1+\mu$. Together with $\eta^\star \in [-\mu, 0]$, this gives $a_p+\eta^\star \geq 1-\mu > 0$. Hence $m_p > 0$ for every $p \in \Omega$. Therefore, $\mathbf{u}_{\mathrm{ER}}(p) = m_p\mathbf{u}_{\mathrm{C}}(p)$ preserves the direction of every nonzero local Chord vector.

The same bounds imply $1 - \mu \leq a_p + \eta^\star \leq 1 + \mu$, and consequently $1/(1 + \mu) \leq m_p \leq 1/(1 - \mu)$.

Finally, suppose that $\mu > 0$ and $c_p > c_q$. Since $(1 - c_p)^2 < (1 - c_q)^2$, we have $a_p < a_q$. The scalar $\eta^\star$ is shared across all spatial locations, so $a_p + \eta^\star < a_q + \eta^\star$. Both denominators are positive, and therefore $m_p > m_q$. □

The spatial energy relation used in Section 3.3 follows immediately from the same multiplicative representation. Recall that $\pi_p^{\mathrm{C}} := \|\mathbf{u}_{\mathrm{C}}(p)\|_2^2/E_{\mathrm{C}}$ and $\pi_p^{\mathrm{ER}} := \|\mathbf{u}_{\mathrm{ER}}(p)\|_2^2/E_{\mathrm{C}}$. Then

$$\pi_p^{\mathrm{ER}} = \frac{m_p^2\|\mathbf{u}_{\mathrm{C}}(p)\|_2^2}{E_{\mathrm{C}}} = m_p^2\pi_p^{\mathrm{C}}. \tag{24}$$

Hence the local multipliers act directly on the original spatial energy shares, while (23) ensures that the shares remain normalized over the entire field. In particular, the ordering $c_p > c_q \Rightarrow m_p > m_q$ concerns the relative multiplicative adjustment and does not imply $\pi_p^{\mathrm{ER}} > \pi_q^{\mathrm{ER}}$, since the final shares also depend on the original Chord allocation.

The closed form also recovers several limiting cases. If $\mu = 0$, then $a_p = 1$ for every p, and the equal-energy condition yields $\eta^\star = 0$, so $\mathbf{u}_{\mathrm{ER}} = \mathbf{u}_{\mathrm{C}}$. More generally, if c_p is constant over the nonzero support of $\mathbf{u}_{\mathrm{C}}$, all active locations share the same positive multiplier. Exact energy preservation then forces this multiplier to equal one. When $E_{\mathrm{C}} = 0$, the constraint admits only the zero field, and thus $\mathbf{u}_{\mathrm{ER}} = \mathbf{0}$.

A.4 Reliability-Weighted Energy Certificate

Beyond the pointwise properties of the closed-form solution, the optimization objective also controls how much energy is assigned to locations with low endpoint consistency. For any nonzero control field $\mathbf{u}$, define its reliability-weighted energy fraction as

$$\mathcal{M}(\mathbf{u}; \mathbf{c}) := \frac{\sum_{p\in\Omega}(1-c_p)^2\|\mathbf{u}(p)\|_2^2}{\|\mathbf{u}\|_F^2}, \tag{25}$$

where $\mathbf{c}$ denotes the collection of endpoint-consistency scores $\{c_p\}_{p\in\Omega}$.

Proposition A.1 (Reliability-weighted energy certificate)**.** *Assume $\mu > 0$ and $E_{\mathrm{C}} > 0$. Then the solution of* (3) *satisfies*

$$\mathcal{M}(\mathbf{u}_{\mathrm{ER}}; \mathbf{c}) \le \mathcal{M}(\mathbf{u}_{\mathrm{C}}; \mathbf{c}) - \frac{\|\mathbf{u}_{\mathrm{ER}} - \mathbf{u}_{\mathrm{C}}\|_F^2}{\mu E_{\mathrm{C}}} \le \mathcal{M}(\mathbf{u}_{\mathrm{C}}; \mathbf{c}). \tag{26}$$

Proof. The original Chord field $\mathbf{u}_{\mathrm{C}}$ is feasible for (3) because $\|\mathbf{u}_{\mathrm{C}}\|_F^2 = E_{\mathrm{C}}$. By the global optimality of $\mathbf{u}_{\mathrm{ER}}$ established in Theorem 3.1,

$$J(\mathbf{u}_{\mathrm{ER}}) \le J(\mathbf{u}_{\mathrm{C}}) = \frac{\mu}{2}\sum_{p\in\Omega}(1-c_p)^2\|\mathbf{u}_{\mathrm{C}}(p)\|_2^2. \tag{27}$$

Expanding the left-hand side yields

$$\frac{1}{2}\|\mathbf{u}_{\mathrm{ER}} - \mathbf{u}_{\mathrm{C}}\|_F^2 + \frac{\mu}{2}\sum_{p\in\Omega}(1-c_p)^2\|\mathbf{u}_{\mathrm{ER}}(p)\|_2^2 \le \frac{\mu}{2}\sum_{p\in\Omega}(1-c_p)^2\|\mathbf{u}_{\mathrm{C}}(p)\|_2^2. \tag{28}$$

Both $\mathbf{u}_{\mathrm{ER}}$ and $\mathbf{u}_{\mathrm{C}}$ have total energy E_{C}. Dividing (28) by $\mu E_{\mathrm{C}}/2$ and using (25) gives

$$\mathcal{M}(\mathbf{u}_{\mathrm{ER}}; \mathbf{c}) + \frac{\|\mathbf{u}_{\mathrm{ER}} - \mathbf{u}_{\mathrm{C}}\|_F^2}{\mu E_{\mathrm{C}}} \le \mathcal{M}(\mathbf{u}_{\mathrm{C}}; \mathbf{c}), \tag{29}$$

which establishes the first inequality in (26). The second follows because the additional term in (29) is nonnegative. □

Proposition A.2 (Bounds on the Dual Root)**.** *Let $\eta^\star$ be the optimal dual root, and let $\pi_p = \|\mathbf{u}_{\mathrm{C}}(p)\|_2^2/E_{\mathrm{C}}$. The dual root satisfies*

$$-\mu\mathcal{M}(\mathbf{u}_{\mathrm{C}}; \mathbf{c}) \le \eta^\star \le 0. \tag{30}$$

Proof. For the lower bound, apply Jensen's inequality to the strictly convex function $f(x) = x^{-2}$ on $x > 0$:

$$\begin{aligned} 1 &= \sum_p \pi_p\big(1 + \mu(1-c_p)^2 + \eta^\star\big)^{-2} \\ &\ge \left[\sum_p \pi_p\big(1 + \mu(1-c_p)^2 + \eta^\star\big)\right]^{-2} \\ &= \big(1 + \mu\mathcal{M}(\mathbf{u}_{\mathrm{C}}; \mathbf{c}) + \eta^\star\big)^{-2}. \end{aligned} \tag{31}$$

The argument is positive, so the inequality implies $\eta^\star \ge -\mu\mathcal{M}(\mathbf{u}_{\mathrm{C}}; \mathbf{c})$. The upper bound $\eta^\star \le 0$ follows from the endpoint argument in the proof of Theorem 3.1. These bounds can tighten the bisection interval. □

Proposition A.1 concerns the reliability-weighted control energy defined in (25). It does not by itself imply an improvement in decoded image quality or semantic correctness.

B Implementation Details, Numerical Diagnostics, and Reproducibility

B.1 Algorithmic Flow and Finite-Precision Bisection

For each input image, ER-Chord reuses the two Chord endpoint proxy fields $(\mathbf{R}_t, \mathbf{R}_{t-\delta})$ to construct the base field u_{C}. It computes the channel-reduced signal and disagreement maps, forms the consistency scores c_p and weights a_p, solves for the scalar root $\eta^\star$, and applies the spatial multipliers before the standard Chord update and post-hoc cleanup. For active samples ($E_{\mathrm{C}} > \varepsilon_E$), the normalized dual and energy residuals are $r_\phi = |\phi(\eta^\star)|/E_{\mathrm{C}}$ and $r_E = |\|u_{\mathrm{ER}}\|_F^2 - E_{\mathrm{C}}|/E_{\mathrm{C}}$, respectively.

Table 5 reports numerical solver diagnostics across all 700 PIE-Bench examples. The measured residuals meet the stated tolerances, the multipliers remain within the theoretical bounds, and no violations of the reliability-weighted energy inequality are observed.

Table 5: **Solver checks on PIE-Bench (700 examples, seed 42).** Residuals, multiplier bounds, direction preservation, and the mismatch inequality are computed on the realized ER-Chord field after dtype conversion.

Certificate	Measured (required)
Active samples ($E_{\mathrm{C}} > \varepsilon_E$)	700 of 700
Dual root $\eta^\star$ (mean; min; max)	-0.012; -0.155; -0.001 (within $[-0.25, 0]$)
Normalized dual residual r_ϕ (max)	4.1×10^{-7} ($\leq 10^{-6}$)
Relative energy residual r_E (max)	5.3×10^{-7}
Normalized KKT residual (max)	2.8×10^{-8} ($\leq 10^{-6}$)
Multiplier range (min; max)	0.801; 1.183 (within $[0.8, 4/3]$)
Minimum direction cosine	0.99998
Proxy mismatch before $\mathcal{M}(u_{\mathrm{C}}; c)$	0.0520
Proxy mismatch after $\mathcal{M}(u_{\mathrm{ER}}; c)$	0.0473
Energy–reliability mismatch bound violations	0

B.2 Experimental Setup and Reproducibility Protocol

All ER-Chord experiments strictly follow the ChordEdit (Lu et al., 2026) protocol: one shared noise sample, $t = 0.90$, $\delta = 0.15$, update scale $\lambda = 1.0$, and cleanup time $t_c = 0.30$. For ER-Chord, we set $\mu = 0.25$, $\kappa = 1.0$, $\varepsilon_c = 10^{-8}$, energy threshold $\varepsilon_E = 10^{-12}$, and dual tolerance 10^{-6}. These hyperparameters remain completely fixed across all PIE-Bench categories, TEdBench, and all three generative backbones (SD-Turbo, SwiftBrush-v2, InstaFlow) without any dataset-specific or backbone-specific tuning. Timing and memory benchmarks are measured on the same GPU and software environment with identical precision, batch size, warm-up, and synchronization; one-time model loading is excluded. Under this protocol, ER-Chord averages 0.37 s/img (identical peak VRAM of 6988 MiB to ChordEdit).

C Additional Qualitative Results

Figures 5 and 6 provide additional visual comparisons on PIE-Bench under matched random seeds across representative categories, supplementing Figure 4. In these representative cases, some multi-step inversion methods (DDIM and Direct Inversion with MasaCtrl or PnP) smooth non-target regions or alter fine textures such as smoke plumes and mountain ridges. Some few-step methods (TurboEdit, InfEdit, and InstantEdit) show under-editing or color artifacts. The ER-Chord examples illustrate cleaner target changes with better background preservation. The full pipeline uses one transport update followed by cleanup and takes 0.37 s/img in the reported setup.

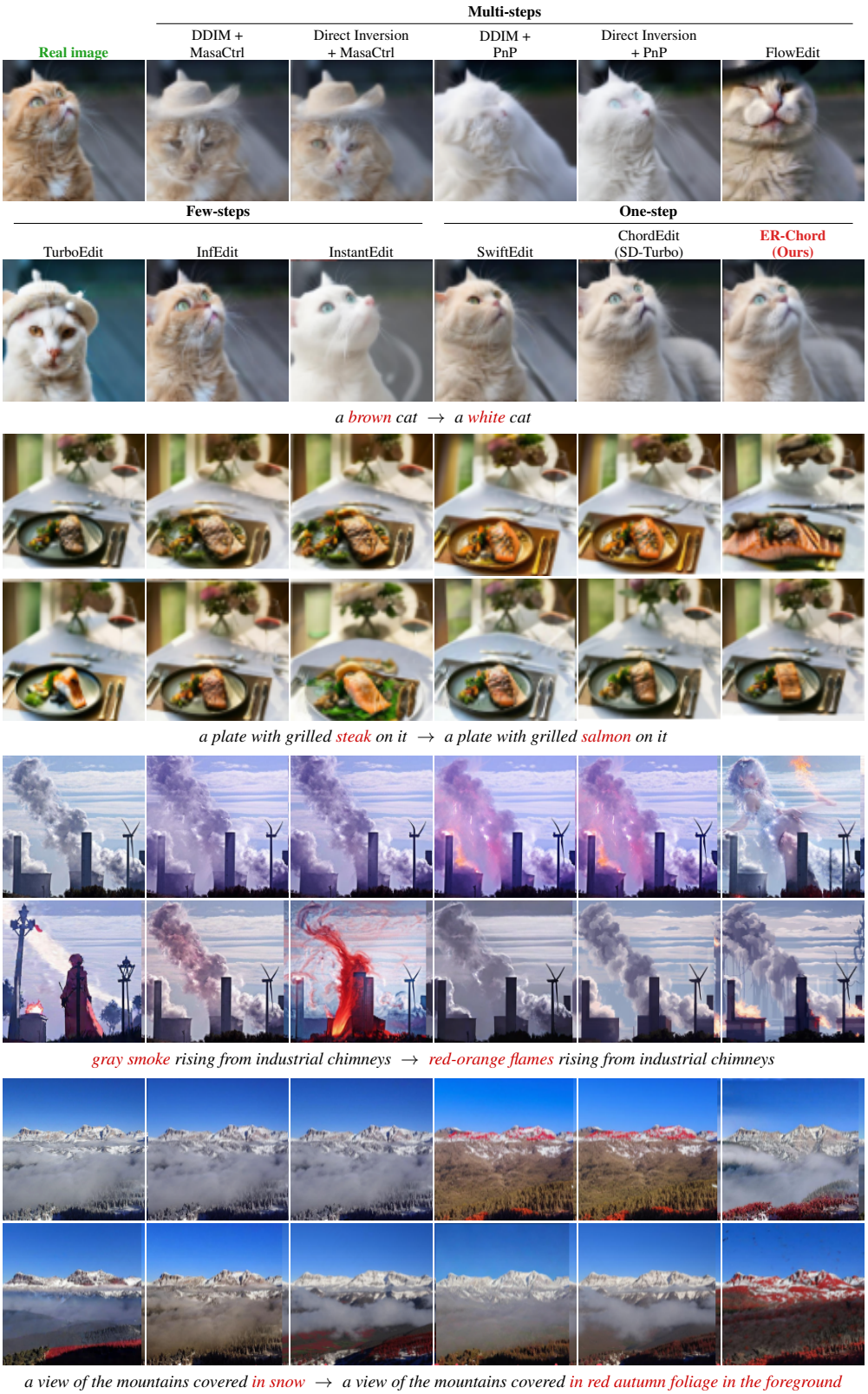

Figure 5: **Additional qualitative comparison on PIE-Bench (Part 1).** Extended visual comparisons across representative categories under matched random seeds, supplementing Figure 4.

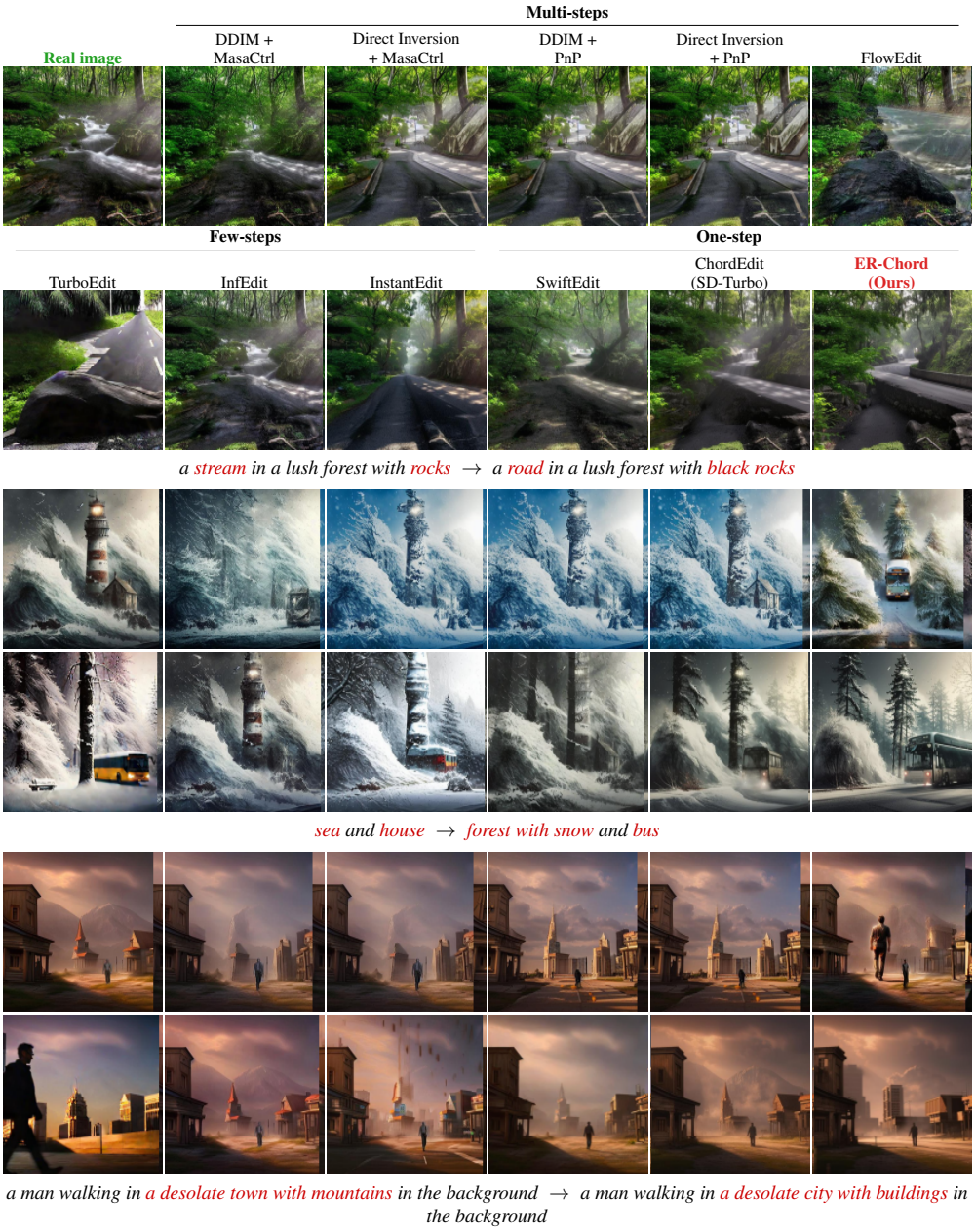

Figure 6: **Additional qualitative comparison on PIE-Bench (Part 2).** Extended visual comparisons across representative categories under matched random seeds, supplementing Figure 4.

D Extended Quantitative Results and Notation

Table 6 reports extended quantitative results on all 700 PIE-Bench examples across the ten editing categories, supplementing Table 1. Relative to ChordEdit, ER-Chord improves SSIM by 0.96 percentage points and reduces structure distance from 16.58×10^{-3} to 15.44×10^{-3}; CLIP-Whole and CLIP-Edited also increase. Table 7 summarizes the main notation.

Table 6: **Additional PIE-Bench metrics (700 examples, seed 42).** ChordEdit values reproduce the extended metrics in Lu et al. (2026). MSE, LPIPS, and structure distance are scaled by 10^3, and SSIM is reported as a percentage, following the source table. Bold denotes the better result.

Metric	ChordEdit	ER-Chord
PSNR↑ (dB)	22.20	**22.77**
MSE↓ ($\times 10^{-3}$)	6.84	**6.35**
SSIM↑ ($\times 10^{-2}$)	75.91	**76.87**
LPIPS↓ ($\times 10^{-3}$)	128.25	**120.23**
Structure dist.↓ ($\times 10^{-3}$)	16.58	**15.44**
CLIP-Whole↑	25.58	**25.74**
CLIP-Edited↑	22.96	**23.16**

Table 7: **Main notation used by ER-Chord.**

Symbol	Meaning
$\mathbf{R}_{t-\delta}, \mathbf{R}_t$	ChordEdit endpoint proxy fields at two nearby noise times.
u_{C}	Original temporally smoothed Chord control field.
$p \in \Omega$	Latent spatial position.
s_p	Chord signal magnitude $\|u_{\mathrm{C}}(p)\|_2$.
d_p	Endpoint disagreement $\|\mathbf{R}_t(p) - \mathbf{R}_{t-\delta}(p)\|_2$.
c_p	Endpoint-consistency score.
μ, κ	Reliability regularization strength and disagreement sensitivity.
E_{C}	Per-image Chord energy $\|u_{\mathrm{C}}\|_F^2$.
η^*	Dual variable enforcing equal energy.
m_p	Local amplitude multiplier $(1 + \mu(1 - c_p)^2 + \eta^*)^{-1}$.
u_{ER}	Equal-energy redistributed control field.
$\mathcal{M}(u; c)$	Inconsistency-weighted fraction of control energy.