SOURCE PORTFOLIO GUIDANCE FOR REWARD-ALIGNED DIFFUSION GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Test-time guidance for text-to-image synthesis seeks to align a pretrained generator with task-specific preferences while retaining diverse outputs, without updating model parameters. Gradient-free source-space methods enable optimization with black-box rewards, but their active populations are designed to continue search rather than to represent all useful discoveries. Consequently, reward-ranked output selection may discard distinct realizations encountered along the search trajectory and fill a small returned set with similar high-scoring samples. We propose Source Portfolio Guidance (SPG), which explicitly separates candidate discovery from finite-set construction under a reward constraint. Trajectory Support Consolidation (TSC) combines multiscale reward-driven proposals with a persistent history, preserving evaluated alternatives beyond the compact parent archive used for exploration. Anchored Support Compression (ASC) retains the best observed candidate and applies winner-anchored farthest-first coverage to a reward-qualified pool, selecting complementary representatives without additional generator evaluations. For fixed evaluated records, this construction preserves peak reward and provides a classical coverage bound when the eligible pool can fill the output set. Across Stable Diffusion 1.4, 1.5, and XL, SPG improves best-of-four ImageReward by 0.105–0.176 and CLIP diversity by 25.2–27.2% over previous state-of-the-art method Source Parallel Tempering (SPT) under the same budget of 420 candidates.

1 INTRODUCTION

Diffusion and flow-based generative models transform simple source distributions into structured data through denoising, score estimation, or flow matching (Ho et al., 2020; Song et al., 2020b; Lipman et al., 2022), with latent-space formulations and text conditioning extending these approaches to high-resolution synthesis (Rombach et al., 2022; Saharia et al., 2022; Podell et al., 2024). Application preferences may nevertheless differ from the training objective, motivating guidance toward external criteria. Whereas classifier and classifier-free guidance exploit learned conditioning signals (Dhariwal & Nichol, 2021; Ho & Salimans, 2022), preference fine-tuning modifies the generator through reward feedback or paired comparisons (Black et al., 2024; Fan et al., 2023; Wallace et al., 2024). With fixed model parameters and an external criterion, test-time guidance instead seeks a reward-reweighted distribution (Wang et al., 2026),

$$\widetilde{p}_\theta(x \mid c) = \frac{p_\theta(x \mid c)\exp\{\lambda r(x,c)\}}{Z_\lambda(c)}, \qquad \lambda > 0, \tag{1}$$

where $p_\theta(x \mid c)$ is the pretrained distribution conditioned on prompt c, r encodes the desired preference, and $Z_\lambda(c)$ normalizes the density. By separating the inference-time criterion from model training while retaining the generative prior, reward reweighting frames preference-aligned text-to-image synthesis as the allocation of limited computation to reward-favored realizations.

Effective guidance requires reaching favorable regions poorly explored under the pretrained distribution, especially with non-differentiable rewards. Although gradient-based methods propagate steering signals through generation (Bansal et al., 2024; Yu et al., 2023), the required gradients need not be available from a black-box reward, motivating source-space methods that evaluate completed generations and update their random inputs. For example, Source Parallel Tempering (SPT) couples local proposals with exchanges across reward temperatures to explore separated regions

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.

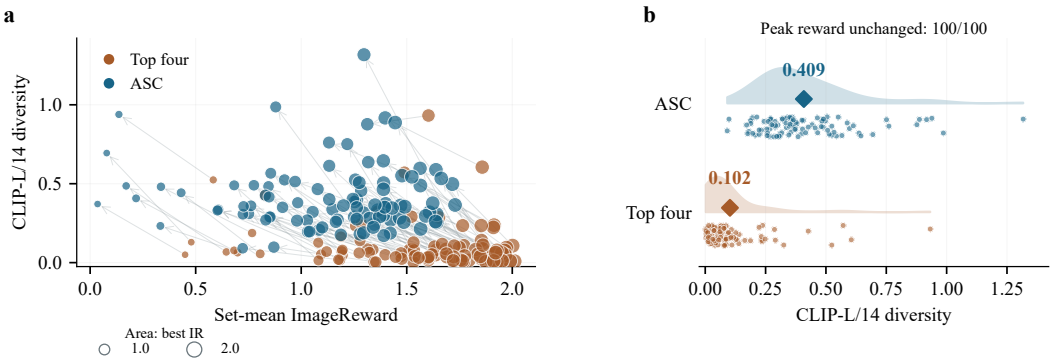


Figure 1: Comparison of top-four selection and ASC on identical candidate pools for 100 prompts using SD 1.4. Panel (a) shows paired reward–diversity changes; bubble area represents best ImageReward. Panel (b) shows CLIP diversity distributions, with dots for prompts and diamonds for means.

(Wang et al., 2026), while evolutionary methods combine selection and mutation with population diversity (He et al., 2025). Even when such mechanisms improve exploration, a finite returned set may lose discovered realizations, because candidates discarded as search parents can still represent distinct outputs, whereas highly ranked descendants may occupy the same feature neighborhood. Translating reward-driven exploration into a limited output set therefore requires explicit quality and representation criteria, rather than assuming population diversity carries over to returned images.

To preserve discovered alternatives within a finite output budget, we introduce Source Portfolio Guidance (SPG), integrating source-space exploration with quality-constrained output construction. Through Trajectory Support Consolidation (TSC), evaluated candidates remain available beyond the reward-ranked parent archive as multiscale proposals refine promising sources, decoupling output selection from the search population. Anchored Support Compression (ASC) then fixes the best observed candidate and selects representatives from a reward-qualified pool through anchored farthest-first coverage (Gonzalez, 1985), thereby preserving peak reward without additional generator evaluations during selection. These guarantees concern fixed evaluated records and feasible eligible-pool coverage, rather than convergence to Equation 1. Under the same evaluation setting, SPG improves best-of-four ImageReward by 0.105–0.176 and CLIP diversity by 25.2–27.2% over the previous state-of-the-art (SOTA) SPT across three Stable Diffusion backbones. Whereas the reward-only control raises peak reward but reduces diversity by 54–67%, the identical-pool comparison in Figure 1 isolates selection from search, retaining peak reward at a measured mean-reward cost.

Our contributions can be summarized as follows:

- We propose Source Portfolio Guidance (SPG), a test-time generation framework that uses Trajectory Support Consolidation to retain multiscale search histories beyond the active parent archive, preventing parent selection from discarding alternatives needed for output construction.
- We introduce Anchored Support Compression (ASC), which selects complementary representatives from reward-qualified candidates through winner-anchored coverage, reducing output redundancy while preserving peak reward without additional generator evaluations.
- On the public SPT benchmark with Stable Diffusion 1.4, 1.5, and XL, SPG improves best-of-four ImageReward by 0.105–0.176 and CLIP diversity by 25.2–27.2% over SPT under matched evaluation settings and candidate budgets.

2 Related Work

2.1 Reward-Guided Generation and Test-Time Search

Reward alignment may modify model parameters, denoising dynamics, or the source distribution, depending on where preference feedback is available and whether the generator can be updated. Within parameter optimization, DDPO formulates denoising as sequential decision-making and

applies policy gradients (Black et al., 2024), whereas DPOK regularizes online reward maximization with a KL penalty relative to the pretrained model (Fan et al., 2023). By differentiating through sampling, DRaFT optimizes rewards directly, with truncated-backpropagation variants reducing its cost (Clark et al., 2024); AlignProp combines reward backpropagation with low-rank adapters and gradient checkpointing (Prabhudesai et al., 2023). Rather than requiring a separately trained reward model, Diffusion-DPO derives an evidence-lower-bound-based preference objective from paired comparisons (Wallace et al., 2024). Although these approaches incorporate preferences into model parameters, a frozen generator requires the alignment mechanism to operate during sampling.

When parameter updates are unavailable, inference-time guidance can inject conditioning signals into denoising dynamics: classifier guidance uses gradients of a noise-conditioned classifier (Dhariwal & Nichol, 2021), while classifier-free guidance combines conditional and unconditional score estimates without an external classifier (Ho & Salimans, 2022). To accommodate more general task functions, Universal Guidance evaluates denoised estimates and introduces forward and backward corrections (Bansal et al., 2024); FreeDoM constructs energies from pretrained networks and revisits diffusion steps through time travel (Yu et al., 2023). TFG unifies training-free guidance within a shared design space and supplies a systematic hyperparameter-search strategy (Ye et al., 2024); for inverse problems, diffusion posterior sampling blends diffusion updates with measurement-dependent gradients instead of imposing exact measurement projections (Chung et al., 2022). Despite their flexibility, gradient-based corrections require information beyond completed-image reward queries, making direct search over generator inputs relevant when only black-box feedback is available.

By relocating guidance to the generator's inputs, source-space search accommodates black-box reward feedback without differentiating through generation. For a deterministic generator $G_\theta(\cdot, c)$ that transports a source law q to $p_\theta(\cdot \mid c)$, the corresponding reward-guided law can be written as

$$\widetilde{q}_{\theta,\lambda}(z \mid c) \propto q(z) \exp\{\lambda r(G_\theta(z, c), c)\}, \qquad (G_\theta(\cdot, c))_{\#}\widetilde{q}_{\theta,\lambda} = \widetilde{p}_\theta(\cdot \mid c), \tag{2}$$

where $\#$ denotes pushforward and the reward normalizer is assumed finite (Wang et al., 2026). To communicate between exploratory and reward-concentrated chains, SPT combines prior-preserving pCN proposals with reward-dependent acceptance and exchanges between tempered chains (Cotter et al., 2013; Earl & Deem, 2005; Wang et al., 2026). Whereas SPT derives diversity gains from sampling this tilted law, SPG constructs a finite output set without preserving Equation 2. Evolution strategies estimate search directions from rewards of random perturbations (Salimans et al., 2017); by contrast, EvoSearch evolves denoising trajectories through selection and mutation tailored to stochastic diffusion dynamics, retaining a population of competing candidates (He et al., 2025). Feynman–Kac correctors derive particle weights from the evolution equations of target distributions, using sequential Monte Carlo resampling to correct annealed or composed sampling paths (Skreta et al., 2025). While these mechanisms govern candidate generation, support retention and compression determine which evaluated candidates represent the search under an explicit output budget.

2.2 Quality-diversity and finite-set construction

Finite-set methods differ in whether diversity is defined through predefined niches, probabilistic repulsion, or geometric coverage. MAP-Elites retains high-performing solutions across behavioral niches (Mouret & Clune, 2015), while determinantal point processes favor subsets through determinants of a positive-semidefinite similarity kernel (Kulesza & Taskar, 2012). Metric K-center instead summarizes a finite feature set by minimizing its worst-case distance to selected representatives, for which farthest-first traversal gives a classical approximation (Gonzalez, 1985). Unlike niche-based retention or an unconstrained subset distribution, SPG restricts admissibility through a reward threshold and requires the best observed candidate to belong to the selected set, with TSC supplying the evaluated history and ASC applying coverage within the qualified domain. The methodological distinction therefore concerns quality and winner constraints, rather than a new generative architecture or K-center theorem. Because feature coverage does not directly measure human preference, reward and independent diversity are evaluated separately; Appendix E provides further background.

3 Methodology

Source Portfolio Guidance (SPG) separates two decisions that are often conflated in test-time guidance: which candidates should be explored and which of the discovered candidates should be returned.

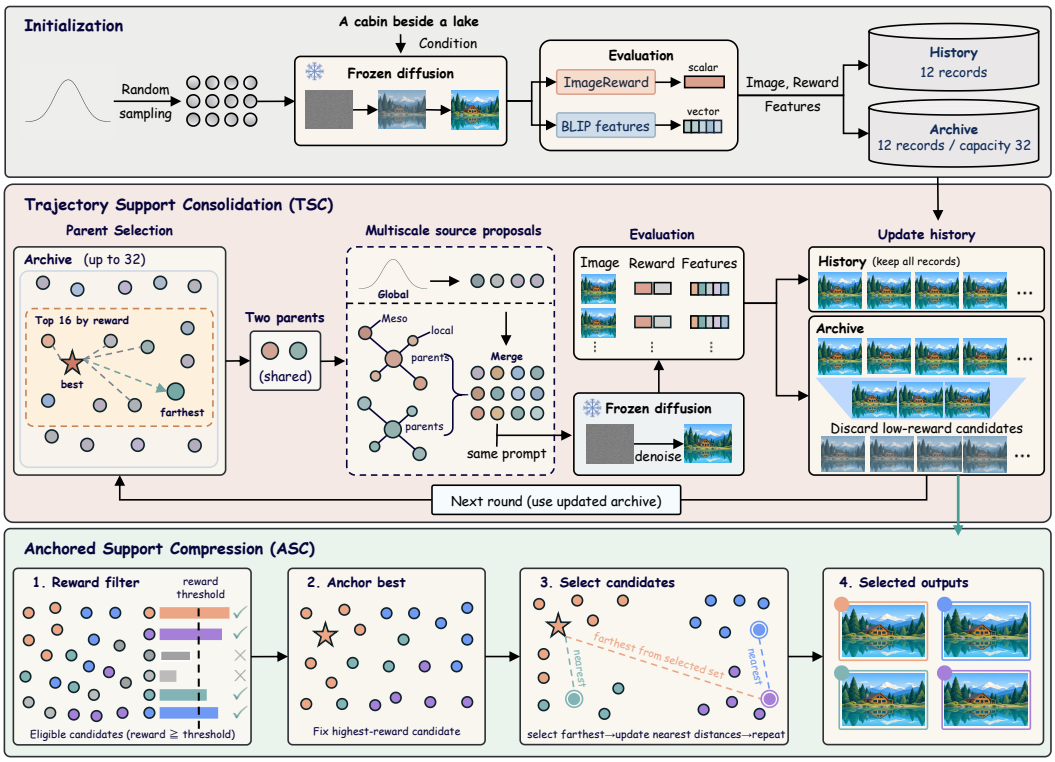


Figure 2: Overview of the SPG framework. Initialization evaluates random sources with a frozen diffusion model and records rewards and BLIP features. TSC generates multiscale proposals from a reward-ranked parent archive, retaining all evaluated candidates in history while updating the compact archive for subsequent rounds. ASC filters the history by reward, anchors the highest-reward candidate, and iteratively selects the candidate farthest from the current selected set in feature space to construct four complementary outputs.

TSC preserves the support accumulated during black-box search, whereas ASC compresses its reward-qualified portion into a finite output set under quality and cardinality constraints (Figure 2). We distinguish the finite-set decision from distributional reward guidance and then formalize the objectives used by the two components.

3.1 Reward guidance and finite-output objectives

Reward reweighting assigns higher probability to preferred individual generations, whereas practical text-to-image systems must return only a small number of images after a finite number of reward evaluations. The resulting finite-output problem therefore separates candidate discovery through search from the allocation of available output slots among the discovered candidates. Even when search reaches high-reward regions, selecting the top-ranked candidates can produce a redundant set if several of them lie in the same feature neighborhood. We therefore formulate SPG as a finite-budget problem with two stages: constructing an evaluated support and selecting a quality-constrained representation of that support.

Let a frozen generator $G_\theta \colon \mathbb{R}^d \times \mathcal{C} \to \mathcal{X}$ map a source variable $z \sim q = \mathcal{N}(0, I)$ and prompt c to an image, thereby inducing $p_\theta(x \mid c)$. We define the corresponding source-space reward as $r_c(z) = r(G_\theta(z, c), c)$. Equation 1 specifies the reward-reweighted target distribution over individual generations (Wang et al., 2026). Rather than sampling from $\widetilde{p}_\theta$, SPG constructs a quality-constrained set from finitely many evaluated candidates. Therefore, the finite-set objective does not require the search archive or the output selector to preserve $\widetilde{p}_\theta$. After B evaluations, let $\mathcal{P}_B = \{z_b\}_{b=1}^{B}$ denote the indexed collection of evaluated source variables, with each z_b stored together with its reward

and descriptor. The strongest observed candidate is defined as follows:

$$z^+ \in \arg\max_{z \in \mathcal{P}_B} r_c(z), \qquad r^+ = r_c(z^+). \tag{3}$$

The value r^+ is therefore the best reward observed under the evaluation budget rather than a global optimum over $\mathbb{R}^d$. Independent best-of-B sampling and adaptive search differ in how they construct $\mathcal{P}_B$. The former draws source variables independently, whereas SPT uses pCN transitions and parallel tempering (Cotter et al., 2013; Earl & Deem, 2005; Wang et al., 2026). Once $\mathcal{P}_B$ is fixed, finite-output construction selects a subset $S \subseteq \mathcal{P}_B$ with $|S| = K$.

The peak score $Q(S) = \max_{z \in S} r_c(z)$ measures the quality of the strongest selected candidate, while the within-set dispersion

$$D_\psi(S) = \frac{2}{K(K-1)} \sum_{i<j} \|\psi(G_\theta(z_i, c)) - \psi(G_\theta(z_j, c))\|_2^2 \tag{4}$$

measures pairwise separation among the selected images, where ψ denotes normalized evaluation features. Although a large $D_\psi(S)$ indicates that the selected images differ, the final set can only represent alternatives that remain available after search. We therefore first consider how evaluated support is retained.

3.2 Trajectory Support Consolidation (TSC)

TSC separates the memory used to continue search from the candidate history used for final selection. Because a compact high-reward archive provides only a lossy summary of the discovered realizations, it serves as proposal memory, while an append-only history preserves evaluated support. Here, trajectory denotes the sequence of evaluated source proposals rather than intermediate diffusion states. The reward-ranked archive $\mathcal{A}_t$ supplies parents, while the history evolves as

$$\mathcal{P}_t = \mathcal{P}_{t-1} \cup \{z_b\}_{b \in I_t}, \tag{5}$$

Here I_t indexes newly evaluated candidates, whose rewards and descriptors are cached with their source variables. The history retains candidates even after their reward ranks fall below the archive cutoff. Its append-only structure also preserves the historical winner, although winner retention alone does not resolve redundancy in a finite output set.

To explore at multiple resolutions, TSC combines fresh Gaussian sources with mesoscopic and local mirrored pCN moves. Given archive parents z_a, z_b and Gaussian directions ϵ_a, ϵ_b, these moves take the form

$$z'_{a,\pm} = \sqrt{1-\beta^2} z_a \pm \beta \epsilon_a, \qquad z'_{b,\pm} = \sqrt{1-\beta^2} z_b \pm \beta \epsilon_b. \tag{6}$$

The archive winner and a high-reward candidate distant from it in source space define complementary parent locations. Fresh Gaussian sources extend the explored support, larger pCN moves explore the parent neighborhoods, and smaller moves refine them. The archive remains reward-ranked throughout, so coverage affects output construction rather than proposal generation. Because rewards and descriptors are recorded during discovery, ASC operates on the retained history without additional generator evaluations. Parent selection, crossover, proposal schedules, and evaluation budgets are detailed in Appendix A.

3.3 Anchored Support Compression (ASC)

TSC determines which evaluated candidates remain available, while ASC determines how the retained support is represented by K outputs. Under a finite output budget, reward-only ranking may allocate several slots to the same feature neighborhood, whereas unconstrained separation may retain low-reward outliers. ASC addresses both cases by imposing reward eligibility before applying anchored coverage. For a reward tolerance $\tau \geq 0$, the eligible pool is defined as follows:

$$\mathcal{P}_\tau = \{z \in \mathcal{P}_B \mid r_c(z) \geq r^+ - \tau\}. \tag{7}$$

Using the Euclidean distance $d_\phi(z,u) = \|\phi(G_\theta(z,c)) - \phi(G_\theta(u,c))\|_2$ between normalized selection features ϕ, ASC approximates

$$\begin{aligned} \min_{S \subseteq \mathcal{P}_\tau} \quad & \rho_\phi(S; \mathcal{P}_\tau) := \max_{z \in \mathcal{P}_\tau} \min_{u \in S} d_\phi(z,u) \\ \text{subject to} \quad & |S| = K, \qquad z^+ \in S. \end{aligned} \tag{8}$$

The reward threshold excludes low-scoring outliers, the anchor retains the best observed candidate, and the covering radius measures the largest distance from an eligible candidate to its nearest selected representative. When $|\mathcal{P}_\tau| < K$, the constrained problem in Equation 8 is infeasible. The selection rule then relaxes the eligibility constraint and defines S as the K highest-reward candidates in $\mathcal{P}_B$. Consequently, $z^+ \in S$ and $Q(S) = r^+$, although the reward-floor condition need not hold for every selected candidate.

Starting from $S_1 = \{z^+\}$, anchored farthest-first traversal fills the remaining slots according to

$$z_j \in \arg\max_{z \in \mathcal{P}_\tau \setminus S_{j-1}} \min_{u \in S_{j-1}} d_\phi(z, u), \qquad S_j = S_{j-1} \cup \{z_j\}. \tag{9}$$

Selection uses normalized visual CLS descriptors from ImageReward's BLIP encoder (Li et al., 2022). Because these descriptors and their rewards are recorded during discovery, ASC requires no additional generator evaluations. The benchmark parameters, distance implementation, and computational costs are detailed in Appendix A.

ASC provides two guarantees on a feasible observed pool. Because the anchor remains in the returned set, $Q(S_K) = r^+$. Every other selected candidate satisfies the reward threshold, giving

$$\overline{r}(S_K) \geq r^+ - \frac{K-1}{K}\tau. \tag{10}$$

Anchored farthest-first traversal also satisfies $\rho_\phi(S_K; \mathcal{P}_\tau) \leq 2\rho^*_\phi$, where ρ^*_ϕ is the optimal anchored K-center radius (Gonzalez, 1985). The bounds apply to fixed search-time records in the observed pool. Neither bound guarantees preservation of mean reward relative to reward-only ranking or establishes improvements in human preference or prompt fidelity. Moreover, because D_ψ and ρ_ϕ describe different set properties and may use different feature representations, a larger D_ψ does not imply a smaller covering radius. Appendix B provides the complete proofs.

The coverage guarantee applies only to candidates retained in $\mathcal{P}_\tau$ and cannot account for realizations absent from the search history. To separate missing support from compression loss, let $\mathcal{T}$ be a fixed reference set of candidates measured under the same feature distance and define

$$\delta(\mathcal{T}, \mathcal{P}_\tau) = \max_{x \in \mathcal{T}} \min_{p \in \mathcal{P}_\tau} d_\phi(x, p). \tag{11}$$

The triangle inequality then gives

$$\rho_\phi(S; \mathcal{T}) \leq \delta(\mathcal{T}, \mathcal{P}_\tau) + \rho_\phi(S; \mathcal{P}_\tau), \tag{12}$$

where $\delta(\mathcal{T}, \mathcal{P}_\tau)$ measures support unavailable to the selector, whereas $\rho_\phi(S; \mathcal{P}_\tau)$ measures compression of the eligible pool. Together, TSC preserves discovered alternatives and ASC selects complementary representatives, reducing candidate loss and output redundancy while retaining the best observed reward.

4 Experiments

To evaluate preference alignment and diversity in the text-to-image guidance setting of Wang et al. (2026), the experiments compare end-to-end performance before isolating the contributions of TSC and ASC and the resulting quality trade-off.

4.1 Experimental settings

Following the text-to-image benchmark of Wang et al. (2026), we evaluate Stable Diffusion 1.4, 1.5, and XL on the 100-prompt SPT benchmark, with ImageReward-v1.0 providing black-box guidance and all pretrained model parameters kept frozen. Images are generated using deterministic DDIM sampling ($\eta = 0$) with 50 denoising steps and a classifier-free guidance scale of 7.5. For a matched-budget comparison, SPT, TSC with reward-ranked top-four output, and SPG each evaluate 420 candidates and return $K = 4$ images per prompt, using random seed 42 throughout the primary benchmark. The SPT baseline uses three parallel chains and 34 search iterations, following the original paper's separate equal-budget experiment. Preference alignment is measured by the highest

Table 1: Comparison of best-of-four ImageReward and CLIP diversity across three Stable Diffusion models on 100 prompts. All methods use 420 candidates per prompt. Bold and underline mark the best and second-best results; Δ denotes the difference from SPT.

Backbone	Method	Best-of-four IR ↑	CLIP diversity ↑
SD 1.4	SPT	1.394	<u>0.329</u>
	TSC (top four)	<u>1.567</u>	0.116
	SPG	**1.571**	**0.412**
	Δ vs. SPT	+0.176	+0.083
SD 1.5	SPT	1.389	<u>0.318</u>
	TSC (top four)	**1.552**	0.106
	SPG	<u>1.542</u>	**0.404**
	Δ vs. SPT	+0.153	+0.086
SDXL	SPT	1.408	<u>0.360</u>
	TSC (top four)	**1.520**	0.165
	SPG	<u>1.513</u>	**0.455**
	Δ vs. SPT	+0.105	+0.096

ImageReward among the four returned images, averaged across prompts, while output diversity is measured by their mean pairwise squared distance in normalized CLIP-L/14 feature space (Equation 4). The main comparison averages both metrics over the same 100 prompts, using the same four returned images per prompt. Appendix A details parameter selection and compute accounting; Appendix C.1 additionally reports validation on the 90 prompts not used for parameter selection.

4.2 End-to-end results and qualitative comparison

Table 2: Comparison of SPG and SPT on 90 held-out prompts. Values are relative changes (%), with prompt-level wins in parentheses.

Backbone	Best ImageReward	CLIP diversity
SD 1.4	+12.57% (87/90)	+27.65% (74/90)
SD 1.5	+11.54% (83/90)	+27.44% (73/90)
SDXL	+8.10% (77/90)	+28.61% (75/90)

Across the full 100-prompt benchmark, SPG improves both average best-of-four ImageReward and average CLIP diversity over SPT on all three backbones (Table 1). On SD 1.4, SD 1.5, and SDXL, respectively, reward reaches 1.571, 1.542, and 1.513, while diversity reaches 0.412, 0.404, and 0.455; paired confidence intervals for the mean improvements exclude zero for both metrics on every backbone. Although TSC with reward-ranked top-four output also improves best-of-four ImageReward over SPT, its diversity falls from 0.329 to 0.116, from 0.318 to 0.106, and from 0.360 to 0.165, indicating that stronger reward-driven discovery alone need not produce a more diverse returned set. The reward-only control retains slightly higher average best-of-four scores on SD 1.5 and SDXL, whereas SPG combines higher reward than SPT with greater output diversity on every backbone. Improvements also persist when the ten parameter-selection prompts are excluded (Table 2; protocol details in Appendix C.1). Since the main comparison uses independent search runs, the following same-pool experiment isolates the effect of output selection.

Figure 3 compares independently generated outputs from SPT, TSC, and SPG for the same prompts. Relative to SPT, SPG offers more varied layouts, viewpoints, and visual styles in the displayed cases, with higher CLIP-B/16 separation in all four examples (panel (f)). Panel (e) places these differences in the context of text alignment and perceptual variation, rather than indicating a uniform improvement across evaluators. Whereas TSC's top-four outputs frequently repeat similar compositions, SPG allocates its output slots to complementary alternatives while retaining the highest-reward candidate from its own search history. The corresponding selection effect is established separately by the identical-pool comparison in Figure 1 and Table 3, which confirms peak-reward preservation alongside increased semantic separation.

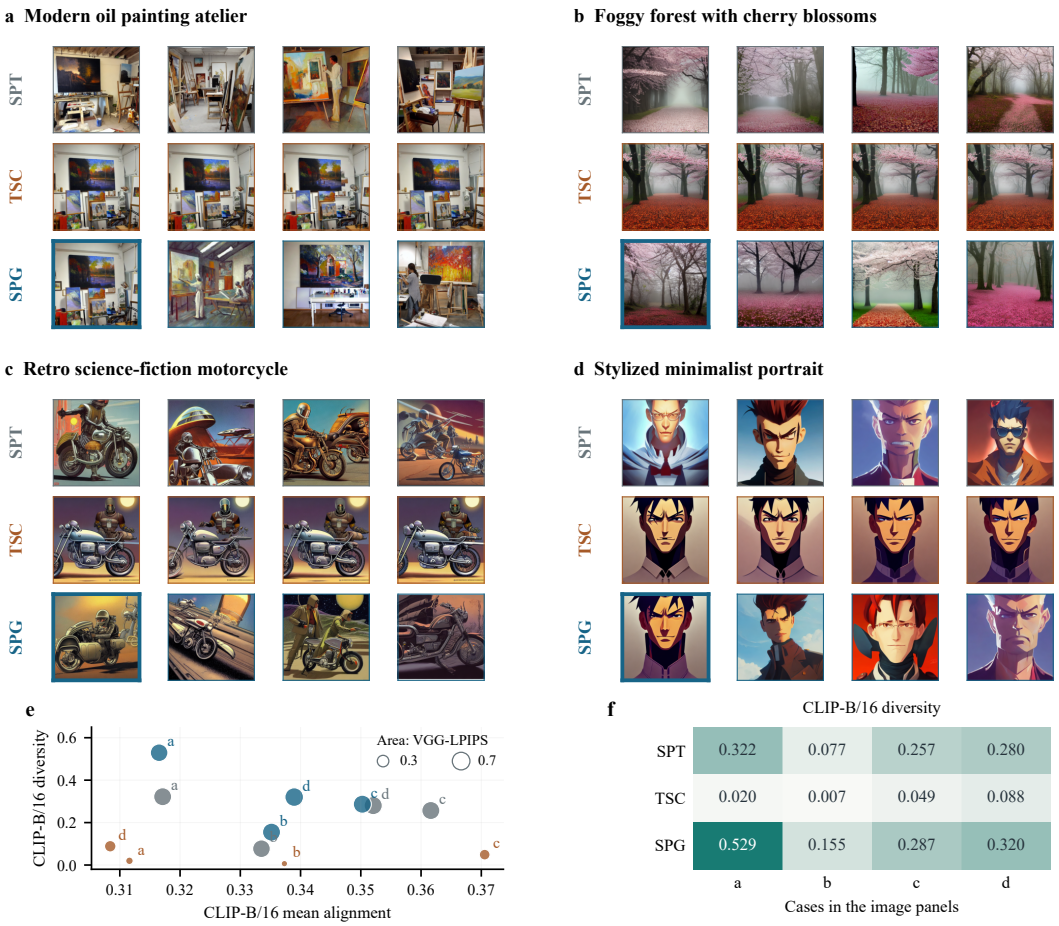


Figure 3: Comparison of samples generated by SPT, TSC with top-four output, and SPG. Panels (a–d) show four illustrative prompts selected after evaluation; outlined images are SPG anchors. Panel (e) shows case-level measurements with bubble area proportional to VGG-LPIPS, and panel (f) reports CLIP-B/16 diversity. Full prompts and additional samples appear in Appendix D.

Table 3: Comparison of TSC and ASC variants on 100 prompts using SD 1.4. Each method evaluates 420 candidates and returns four images; rows 2–3 share candidate pools. Best and mean denote ImageReward. Bold marks column maxima.

TSC support policy	ASC / output rule	Best ↑	Mean ↑	CLIP div. ↑
IID replacement	ASC	1.382	0.963	**0.435**
TSC	w/o ASC: top-four	**1.572**	**1.547**	0.102
TSC	ASC	**1.572**	1.140	0.409

4.3 Ablation Studies

To evaluate the contribution of each component, we compare ASC with top-four selection on identical pools for all 100 prompts (Figure 1, Table 3). Best reward is preserved exactly on every prompt, while CLIP diversity rises from 0.102 to 0.409 (paired gain 0.306; 95% CI [0.274,0.341]; 98/100 wins). Selection requires no additional generator calls, although mean reward falls from 1.547 to 1.140 as complementary candidates enter the set. At $\tau = 0.75$, all selected search records satisfy their reward floors. Tightening τ to 0.25 raises mean reward to 1.430 while reducing diversity to 0.298 (Figure 4). Appendix C.2 provides intervals and fallback diagnostics.

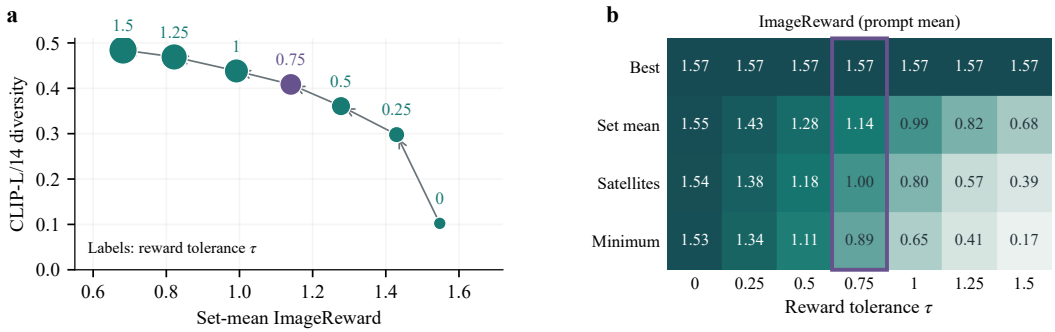


Figure 4: ImageReward and CLIP diversity for different reward tolerances on the same 100 candidate pools. Panel (a) shows the reward–diversity trade-off, with bubble size indicating τ. Panel (b) reports reward statistics; satellites denote non-anchor outputs, and $\tau = 0.75$ is outlined. At $\tau = 0$, selection falls back to the top four.

With persistent storage and ASC held fixed, replacing iid sources by TSC's adaptive proposals improves best ImageReward from 1.382 to 1.572 (paired gain 0.189; 95% CI [0.153,0.228]; 96/100 wins), although CLIP diversity decreases from 0.435 to 0.409. Appendix C.2 distinguishes this discovery contrast from the ASC intervention on identical support, whose effects cannot be added. These controls use a separate SD 1.4 run from Table 1. Together, they support the discovery–representation distinction rather than uniform dominance, while the separate 20-prompt operator study (Appendix C.3) does not establish the refresh/crossover schedule as indispensable.

5 Failure cases and limitations

The evaluation of the main benchmark's 3,600 saved outputs with CLIP-B/16 shows diversity improvements of 31.50%, 32.71%, and 31.44% over SPT across the three backbones, with paired intervals excluding zero throughout. Although these results support the transfer of semantic separation across encoders, mean alignment decreases by 0.89–1.91%, while VGG-LPIPS differences remain inconclusive (Appendix C.4). Together with the decrease in mean reward, these findings reflect the cost of retaining complementary alternatives, whose semantic differences may include nuisance variation or prompt drift rather than useful choices. Since output construction cannot recover alternatives absent from the discovered support, neither improved perceptual diversity nor stronger human preference follows from greater semantic separation alone (Appendix D). Evaluation on 100 inherited prompts and three backbones with one primary seed and a substantial fixed search budget further limits generalization.

6 Conclusion

Test-time guidance for text-to-image synthesis steers a frozen generator toward task-specific preferences under a limited inference budget. Since practical systems return only a few images, high-reward discovery alone does not ensure a representative output set. Reward-ranked selection may discard distinct realizations and assign multiple slots to similar samples. We introduced Source Portfolio Guidance (SPG) to separate candidate discovery from finite-set construction. TSC retains alternatives discovered through multiscale source-space exploration, while ASC preserves the best observed candidate and selects complementary outputs from the reward-qualified history without additional generator evaluations. Across three Stable Diffusion backbones, SPG improves best-of-four ImageReward by 0.105–0.176 and CLIP diversity by 25.2–27.2% over SPT under matched budgets. Same-pool comparisons further confirm that ASC increases feature separation without changing peak reward, with the reward tolerance controlling the associated mean-reward trade-off. SPG thus converts reward-guided search results into a less redundant, quality-controlled output portfolio. By separating proposal memory from output support, SPG keeps discoveries useful even after they cease to guide search, making finite-output quality and coverage explicit objectives rather than incidental consequences of reward optimization.

AI USE STATEMENT

Generative AI assisted literature organization, prose revision, experiment and visualization code, and manuscript checks. Execution and empirical validation are documented in the research artifact. AI is not an author, and the authors retain responsibility for verifying the references, implementation, experimental results, and final submission.

ETHICS STATEMENT

Pretrained generators and reward models may inherit stereotypes, unsafe content, and reward-model blind spots. Test-time reward optimization does not replace content safeguards or human review in deployment. No human-subject study was conducted and no personal data are released. Failure cases and limitations are retained to document the boundaries of the method and its evaluation.

REPRODUCIBILITY STATEMENT

The supplementary material specifies the prompt protocol, development and held-out splits, random seeds, reward tolerance, generator settings, and evaluation budgets. Algorithm 1 and Appendix A describe the complete TSC–ASC procedure, while Appendix F.1 records artifact checks. The research artifact retains generated images, archived rewards and descriptors, launch scripts, paired-analysis programs, and plotting code to support verification of the reported results.

REFERENCES

Arpit Bansal, Hong-Min Chu, Avi Schwarzschild, Roni Sengupta, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Universal guidance for diffusion models. In *International Conference on Learning Representations*, volume 2024, pp. 51304–51323, 2024.

Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. In *International Conference on Learning Representations*, volume 2024, pp. 4965–4987, 2024.

Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022.

Kevin Clark, Paul Vicol, Kevin Swersky, and David Fleet. Directly fine-tuning diffusion models on differentiable rewards. In *International Conference on Learning Representations*, volume 2024, pp. 4793–4822, 2024.

Simon L Cotter, Gareth O Roberts, Andrew M Stuart, and David White. Mcmc methods for functions: modifying old algorithms to make them faster. *Statistical Science*, pp. 424–446, 2013.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

David J Earl and Michael W Deem. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910–3916, 2005.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first international conference on machine learning*, 2024.

Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. Dpok: Reinforcement learning for fine-tuning text-to-image diffusion models. *Advances in neural information processing systems*, 36:79858–79885, 2023.

Dhruba Ghosh, Hannaneh Hajishirzi, and Ludwig Schmidt. Geneval: An object-focused framework for evaluating text-to-image alignment. *Advances in Neural Information Processing Systems*, 36: 52132–52152, 2023.

Teofilo F Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical computer science*, 38:293–306, 1985.

Haoran He, Jiajun Liang, Xintao Wang, Pengfei Wan, Di Zhang, Kun Gai, and Ling Pan. Scaling image and video generation via test-time evolutionary search. *arXiv preprint arXiv:2505.17618*, 2025.

Jack Hessel, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, and Yejin Choi. Clipscore: A reference-free evaluation metric for image captioning. In *Proceedings of the 2021 conference on empirical methods in natural language processing*, pp. 7514–7528, 2021.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Kaiyi Huang, Kaiyue Sun, Enze Xie, Zhenguo Li, and Xihui Liu. T2i-compbench: A comprehensive benchmark for open-world compositional text-to-image generation. *Advances in Neural Information Processing Systems*, 36:78723–78747, 2023.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35:26565–26577, 2022.

Yuval Kirstain, Adam Polyak, Uriel Singer, Shahbuland Matiana, Joe Penna, and Omer Levy. Pick-a-pic: An open dataset of user preferences for text-to-image generation. *Advances in neural information processing systems*, 36:36652–36663, 2023.

Alex Kulesza and Ben Taskar. Determinantal point processes for machine learning. *Foundations and Trends® in Machine Learning*, 5(2-3):123–286, 2012.

Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved precision and recall metric for assessing generative models. *Advances in neural information processing systems*, 32, 2019.

Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International conference on machine learning*, pp. 12888–12900. PMLR, 2022.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pp. 740–755. Springer, 2014.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.

Jean-Baptiste Mouret and Jeff Clune. Illuminating search spaces by mapping elites. *arXiv preprint arXiv:1504.04909*, 2015.

Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741*, 2021.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pp. 8162–8171. PmLR, 2021.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4172–4182. IEEE, 2023.

Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis. In *International Conference on Learning Representations*, volume 2024, pp. 1862–1874, 2024.

Mihir Prabhudesai, Anirudh Goyal, Deepak Pathak, and Katerina Fragkiadaki. Aligning text-to-image diffusion models with reward backpropagation. 2023.

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.

Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1(2):3, 2022.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 10674–10685. ieee, 2022.

Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in neural information processing systems*, 35:36479–36494, 2022.

Tim Salimans, Jonathan Ho, Xi Chen, Szymon Sidor, and Ilya Sutskever. Evolution strategies as a scalable alternative to reinforcement learning. *arXiv preprint arXiv:1703.03864*, 2017.

Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, et al. Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in neural information processing systems*, 35:25278–25294, 2022.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Marta Skreta, Tara Akhound-Sadegh, Viktor Ohanesian, Roberto Bondesan, Alán Aspuru-Guzik, Arnaud Doucet, Rob Brekelmans, Alexander Tong, and Kirill Neklyudov. Feynman-kac correctors in diffusion: Annealing, guidance, and product of experts. *arXiv preprint arXiv:2503.02819*, 2025.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. pmlr, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020a.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020b.

Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shafiq Joty, and Nikhil Naik. Diffusion model alignment using direct preference optimization. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8228–8238. IEEE, 2024.

Shih-Hsin Wang, Joel Keller, Taos Transue, Drake Benjamin Brown, Thomas Strohmer, and Bao Wang. Test-time guidance for flow-based generative models via parallel tempering on source distributions. In *Forty-third International Conference on Machine Learning*, 2026.

Zijie J Wang, Evan Montoya, David Munechika, Haoyang Yang, Benjamin Hoover, and Duen Horng Chau. Diffusiondb: A large-scale prompt gallery dataset for text-to-image generative models. In *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 1: Long papers)*, pp. 893–911, 2023.

Xiaoshi Wu, Yiming Hao, Keqiang Sun, Yixiong Chen, Feng Zhu, Rui Zhao, and Hongsheng Li. Human preference score v2: A solid benchmark for evaluating human preferences of text-to-image synthesis. *arXiv preprint arXiv:2306.09341*, 2023.

Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. Imagereward: Learning and evaluating human preferences for text-to-image generation. *Advances in Neural Information Processing Systems*, 36:15903–15935, 2023.

Haotian Ye, Haowei Lin, Jiaqi Han, Minkai Xu, Sheng Liu, Yitao Liang, Jianzhu Ma, James Zou, and Stefano Ermon. Tfg: Unified training-free guidance for diffusion models. *Advances in Neural Information Processing Systems*, 37:22370–22417, 2024.

Jiwen Yu, Yinhuai Wang, Chen Zhao, Bernard Ghanem, and Jian Zhang. Freedom: Training-free energy-guided conditional diffusion model. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 23117–23127. IEEE, 2023.

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.

Appendix

A Implementation details

The implementation is organized around the experimental protocol, the TSC–ASC procedure, and the evaluation methodology, so that the source-space formulation can be traced to candidate generation, persistent storage, and final output construction. Unless a same-pool intervention is explicitly specified, methods run independently under matched prompts, generation budgets, and output cardinalities, with fixed configurations governing both candidate records and reported measurements.

A.1 Experimental protocol

Following the 100-prompt benchmark released with SPT, the evaluation guides frozen Stable Diffusion 1.4, 1.5, and XL generators with ImageReward-v1.0 (Wang et al., 2026; Podell et al., 2024). The equal-budget configuration initializes three groups of four sources and performs 34 proposal rounds, yielding 420 evaluated candidates from which each method returns four images. With seed 42, 50 deterministic DDIM steps, and classifier-free guidance scale 7.5 held fixed (Song et al., 2020a; Ho & Salimans, 2022), the resulting 105 batches of four correspond to 5,250 sampler-NFE under the source project's batch-counted convention, or 21,000 candidate-level sampling steps, followed by one common decoding batch for the selected sources. This is the source paper's separate equal-budget setting rather than its 30-iteration main configuration, which is why Table 1 uses the official SPT implementation evaluated under the matched configuration.

Ten fixed development prompts determine $\tau = 0.75$, while the remaining 90 provide disjoint validation. Component and tolerance ablations cover all 100 prompts without retuning, whereas the separate operator study retains the first 20 held-out identifiers fixed before outcomes were inspected. Whereas the three-backbone TSC and SPG comparisons use independent search runs, the same-pool intervention applies top-four ranking and ASC to one TSC history, separating the effect of output construction from the generation of its candidates. Published SPT rewards of 1.409, 1.411, and 1.479 differ from the matched-budget measurements of 1.394, 1.389, and 1.408, and the published SD 1.4 and SDXL diversity scores of 0.346 and 0.351 differ from our three-backbone measurements of 0.329, 0.318, and 0.360. These published values document protocol provenance, while every empirical comparison uses measurements obtained under the common configuration.

A.2 Algorithm implementation

Table 4 and Algorithm 1 specify the complete TSC–ASC implementation, in which source dimensions follow the corresponding Stable Diffusion pipeline (Rombach et al., 2022) and each decoded candidate is stored together with its ImageReward score and normalized BLIP visual descriptor. The initial 12 Gaussian sources initialize both a reward-ranked archive and an append-only history, after which the archive retains at most 32 parent candidates while the history preserves every evaluated record and its identity. Separating these memories allows candidates that leave the high-reward archive to remain available for quality-constrained output construction, with historical identifiers retained in raw files for provenance.

Each proposal round takes the archive winner as its first parent and the source-distant member among the remaining top-16 candidates as its second, then evaluates four global, four mesoscopic, and four local proposals. Global proposals are independent Gaussian draws, whereas mesoscopic and local proposals use the mirrored pCN rule in Equation 6, with block crossover replacing rather than augmenting the mesoscopic group every fifth iteration. Crossover draws two independent 8×8 binary grids, expands each grid to the latent spatial resolution, and forms complementary mixtures of the two parents, thereby producing four proposals without changing the evaluation budget. The scheduled scales in Table 4 move from broader exploration toward local refinement, and each round appends all 12 records before updating the reward archive, giving exactly $12 + 34 \times 12 = 420$ candidate evaluations.

ASC anchors the highest-reward record in the full history and fills the remaining slots by farthest-first selection among candidates satisfying $r \geq r^{+} - \tau$, using cached descriptors rather than further

Table 4: Default inference and SPG search settings. Ablations change only the named proposal or selection rule.

Quantity	Value
Returned set size K	4
Iterations	34
Proposal groups / candidates per group	3 / 4
Initial candidates / total candidates	12 / 420
Reward archive size	32
DDIM steps / CFG scale	50 / 7.5
Explore (β_m, β_ℓ), iterations 1–10	$(0.75, 0.18)$
Consolidate, iterations 11–24	$(0.45, 0.10)$
Refine, iterations 25–34	$(0.25, 0.05)$
Crossover frequency / block grid	every 5 iterations / 8×8
Reward slack τ	0.75
Full-benchmark seed	42
Search sampler-NFE per prompt	5,250
Final materialization per portfolio	one 50-step batch

Algorithm 1 Source Portfolio Guidance with TSC and ASC

Require: Prompt c, generator G_θ, reward r_c, descriptor ϕ, $T = 34$, $K = 4$, $\tau = 0.75$
Ensure: An anchored K-image portfolio

Set the random seed and draw 12 Gaussian sources Z_0
$\mathcal{P} \leftarrow \mathrm{Evaluate}(Z_0, c)$
$\mathcal{A} \leftarrow \mathrm{TopReward}_{32}(\mathcal{P})$
for $t = 1, \ldots, T$ **do**
  $z_a \leftarrow \arg\max_{z \in \mathcal{A}} r_c(z)$
  $z_b \leftarrow \arg\max_{z \in \mathrm{TopReward}_{16}(\mathcal{A}) \setminus \{z_a\}} \|z - z_a\|_2^2$
  $(\beta_m, \beta_\ell) \leftarrow \mathrm{Schedule}(t)$
  $Z_g \leftarrow$ four independent Gaussian sources
  if $t \bmod 5 = 0$ **then**
    $Z_m \leftarrow \mathrm{BlockCrossover}(z_a, z_b)$
  else
    $Z_m \leftarrow \mathrm{MirroredPCN}(z_a, z_b, \beta_m)$
  end if
  $Z_\ell \leftarrow \mathrm{MirroredPCN}(z_a, z_b, \beta_\ell)$
  $R_t \leftarrow \mathrm{Evaluate}(Z_g \cup Z_m \cup Z_\ell, c)$
  $\mathcal{P} \leftarrow \mathcal{P} \parallel R_t$
  $\mathcal{A} \leftarrow \mathrm{TopReward}_{32}(\mathcal{A} \cup R_t)$
end for
$z^+ \leftarrow \arg\max_{z \in \mathcal{P}} r_c(z)$
$E \leftarrow \{z \in \mathcal{P} \mid r_c(z) \geq r_c(z^+) - \tau\}$
if $|E| < K$ **then**
  $E \leftarrow \mathrm{TopReward}_K(\mathcal{P})$
end if
$S \leftarrow \{z^+\}$
while $|S| < K$ **do**
  $z^* \leftarrow \arg\max_{z \in E \setminus S} \min_{s \in S} \|\phi(G_\theta(z, c)) - \phi(G_\theta(s, c))\|_2^2$
  $S \leftarrow S \cup \{z^*\}$
end while
return $\{G_\theta(z, c) \mid z \in S\}$

generator calls. Squared Euclidean distances between normalized features induce the same farthest-first ordering as Euclidean distances, while the theoretical analysis uses the latter to retain the triangle inequality. If fewer than four records are eligible, the implementation replaces the domain with the top-four candidates before anchored selection, preserving the winner without imposing the original reward floor on this fallback branch. Storing B sources of dimension d and descriptors of dimension m requires $O(B(d+m))$ memory, while the released implementation recomputes distances to selected centers in $O(BK^2m)$ arithmetic and an equivalent nearest-center cache reduces this to $O(BKm)$. With $K = 4$ fixed, both forms reuse the same evaluated records. The additional final decoding batch

contributes 50 batch-counted sampler steps, giving 5,300 including output materialization rather than 5,250 for search alone. Decoding the four selected sources does not create new search candidates, so the search budget remains $B = 420$. Equal candidate budgets control generator evaluation counts, not wall-clock time or memory, which also depend on descriptor storage and selection.

A.3 Evaluation methodology

Best-of-four ImageReward is computed by taking the maximum within each prompt and then averaging over prompts, while set-mean, non-anchor mean, and minimum selected rewards describe the remaining portfolio quality (Xu et al., 2023). For normalized embeddings v_i, both CLIP-L/14 and independent CLIP-B/16 diversity average the six pairwise squared distances among four outputs, with $\|v_i - v_j\|_2^2 = 2(1 - v_i^\top v_j)$, whereas CLIP alignment measures cosine similarity between normalized text and image embeddings (Radford et al., 2021). BLIP descriptors support selection (Li et al., 2022), CLIP-L/14 informs development-set tolerance selection, and independent CLIP-B/16 and VGG-LPIPS are excluded from search and tuning, with LPIPS averaged over the same six image pairs (Zhang et al., 2018; Simonyan & Zisserman, 2014). Search-time records define eligibility and the theoretical reward guarantees, whereas benchmark scores are measured after final decoding. Re-decoding may introduce small numerical differences, so record-level winner preservation does not assert bitwise equality of newly scored images. All confidence intervals use 10,000 paired-prompt bootstrap resamples that retain within-prompt dependence, and absolute reward differences provide the primary reward comparison. Differences are computed before rounding, and prompt-level ties use an absolute tolerance of 10^{-7} in the expanded ablation.

Table 5 shows that all three development settings preserve best reward at 1.525, while increasing τ from 0.40 to 0.75 raises diversity from 0.458 to 0.504 and a further increase to 1.25 yields only 0.511 while changing set-mean reward from 1.111 to 0.818, motivating the selected compromise of $\tau = 0.75$. Figure 5(a) attributes reward improvements over the initial pool to successive search stages, and the mean last-improvement iteration of 27.7 in panel (b) shows that candidate improvements continue into later rounds on these development prompts. Panel (c) separately checks differences between archived and finally decoded rewards, allowing guarantees on evaluated records to be distinguished from measurements on saved images. Hyperparameters remain fixed during held-out evaluation, and post-hoc qualitative examples illustrate behavior without altering quantitative prompt membership.

Table 5: Comparison of reward tolerances on ten development prompts. All settings retain the same highest-reward candidate.

Reward slack τ	Best ImageReward	Set-mean ImageReward	CLIP diversity
.40	1.525	1.285	.458
.75	1.525	1.111	.504
1.25	1.525	.818	.511

B Theoretical proofs

Assume $1 \le K \le B$, $\tau \ge 0$, finite rewards, and well-defined normalized descriptors, with $K \ge 2$ required for pairwise diversity. Each evaluation has a distinct index, so the union defining $\mathcal{P}_t$ preserves separate records even when source values coincide. The pseudocode stores these indexed candidates as source–reward–descriptor tuples. The analysis fixes a prompt and its evaluated candidate records, treating their rewards and descriptors as unchanged throughout each argument and retaining distinct record identities even when feature vectors coincide. Since TSC determines the records available to selection and ASC constructs an anchored set within a fixed eligible domain, reward preservation and geometric coverage can be established separately and then connected through the retained history. Reward-floor and constrained-coverage statements assume $|\mathcal{P}_\tau| \ge K$, whereas anchor preservation also applies to the fallback branch.

B.1 Reward guarantees

Let z_t^+ maximize reward over $\mathcal{P}_t$ and let $S_t \subseteq \mathcal{P}_t$ contain this record. Inclusion of the anchor gives $Q(S_t) \ge r_c(z_t^+)$, while inclusion of the selected set in the pool gives $Q(S_t) \le \max_{z \in \mathcal{P}_t} r_c(z)$,

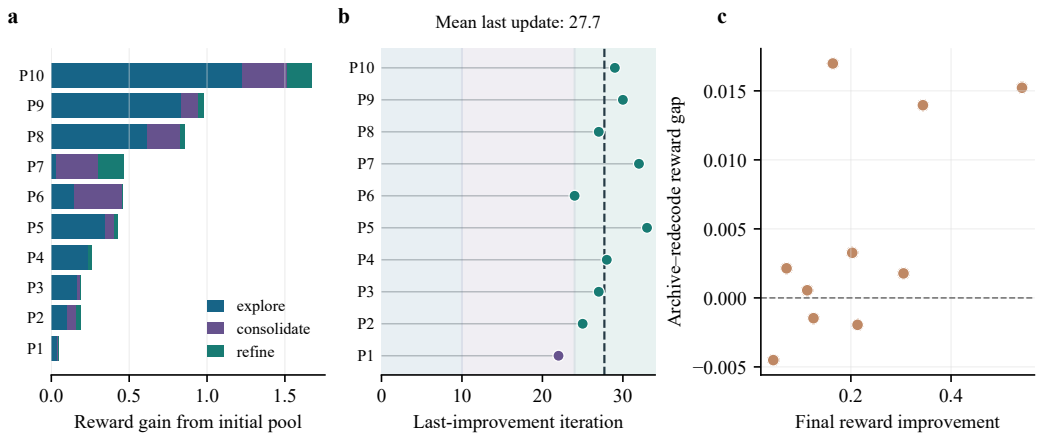


Figure 5: Search diagnostics on ten development prompts. Panels (a–c) show stage-wise reward changes, the last archive-improvement iteration, and reward differences between archived and redecoded outputs. Prompts are ordered by total reward change. Shading identifies search stages.

establishing exact preservation of the observed maximum. Because TSC appends records without deleting earlier evaluations, $\mathcal{P}_t \subseteq \mathcal{P}_{t+1}$ expands the maximization domain and therefore yields

$$Q(S_t) = r_t^+, \qquad Q(S_{t+1}) = \max_{z \in \mathcal{P}_{t+1}} r_c(z) \geq \max_{z \in \mathcal{P}_t} r_c(z) = Q(S_t). \tag{13}$$

The argument does not require independent proposals or acceptance of every candidate into the parent archive, so it applies to adaptive discovery while allowing the non-anchor output slots to change. Monotonicity concerns fixed search-time reward records, rather than assuming numerical identity between archived evaluations and a separate final decoding pass.

When $|\mathcal{P}_\tau| \geq K$, each of the $K-1$ non-anchor selections satisfies $r_c(z) \geq r^+ - \tau$, while the anchor has reward exactly r^+. Summing these inequalities and dividing by K proves Equation 10 and places the selected mean within the interval

$$\min_{z \in S_K} r_c(z) \geq r^+ - \tau, \qquad r^+ - \frac{K-1}{K}\tau \leq \frac{1}{K}\sum_{z \in S_K} r_c(z) \leq r^+. \tag{14}$$

For $K = 4$, the lower bound is $r^+ - 3\tau/4$, making the tolerance an explicit control on admissible reward deviation rather than an implicit weight in a mixed reward–distance objective. On a fixed history, $\tau_1 \leq \tau_2$ also implies $\mathcal{P}_{\tau_1} \subseteq \mathcal{P}_{\tau_2}$, so a larger tolerance broadens the domain of complementary candidates without changing the anchor. If the domain contains fewer than K records, top-K fallback retains the maximum, while the eligibility-based bounds remain restricted to the feasible branch.

B.2 Coverage guarantees

Fix the eligible domain $E = \mathcal{P}_\tau$ and let d_ϕ be Euclidean distance between normalized descriptors, with S_j denoting the first j selected centers and $R_j = \max_{z \in E} \min_{u \in S_j} d_\phi(z, u)$. Adding a center cannot increase a candidate's nearest-center distance, so $R_{j+1} \leq R_j$, while farthest-first traversal chooses the next point at distance R_j from S_j. If $|E| = K$, selecting all records gives zero radius. Otherwise, augment S_K with its farthest remaining point z_{K+1}, for which the same monotonicity ensures that every pair among these $K+1$ points is separated by at least R_K. These properties depend on the fixed candidate geometry rather than the proposal mechanism that produced it.

Let S^* be an optimal anchored K-center set with radius ρ_ϕ^*. If $R_K > 2\rho_\phi^*$, assigning the $K+1$ points to their nearest centers in S^* places two points u, v at the same center s^* by the pigeonhole principle, although the triangle inequality requires

$$d_\phi(u, v) \leq d_\phi(u, s^*) + d_\phi(s^*, v) \leq 2\rho_\phi^* < R_K, \tag{15}$$

contradicting their separation and proving $\rho_\phi(S_K; E) \leq 2\rho^*_\phi$. This is the classical farthest-first argument (Gonzalez, 1985) applied to the shared-anchor constraint, with both the greedy and comparison sets containing the same required candidate. Squaring distances in the implementation leaves the traversal unchanged, so the factor-two guarantee holds for the Euclidean radius, while the corresponding factor for its square is four.

To connect compression with discovered support, fix a reference set $\mathcal{T}$ under the same distance, choose a nearest eligible point $p_x \in E$ for each $x \in \mathcal{T}$, and choose a nearest selected center $s_x \in S_K$ for p_x. Although s_x need not be the nearest center to x, it remains an available representative, giving

$$\begin{aligned} \min_{s \in S_K} d_\phi(x, s) &\leq d_\phi(x, s_x) \\ &\leq d_\phi(x, p_x) + d_\phi(p_x, s_x) \\ &\leq \delta(\mathcal{T}, E) + \rho_\phi(S_K; E). \end{aligned} \tag{16}$$

Taking the maximum over x proves Equation 12 and combines with the coverage guarantee to yield $\rho_\phi(S_K; \mathcal{T}) \leq \delta(\mathcal{T}, E) + 2\rho^*_\phi$. The first term measures the distance from reference support to the eligible history and the second describes its representation by ASC, clarifying the complementary responsibilities of discovery and compression. Comparisons across histories require a common reference set and eligibility rule, and the decomposition does not identify pairwise output diversity with an empirical covering radius.

Persistent history enlarges the choices available for representation under a shared quality rule. To make this statement precise, fix a completed history H, its final reward-ranked archive $A \subseteq H$, and the common threshold $q = r^+_H - \tau$. The eligible domains $E_H = \{z \in H : r_c(z) \geq q\}$ and $E_A = A \cap E_H$ satisfy $E_A \subseteq E_H$. For any fixed reference set $\mathcal{T}$ and nonempty E_A, minimizing over the larger domain gives

$$\delta(\mathcal{T}, E_H) \leq \delta(\mathcal{T}, E_A). \tag{17}$$

If the shared anchor belongs to E_A and $|E_A| \geq K$, every anchored K-element subset of E_A is also feasible in E_H. Consequently, the optimal radius on the common reference E_H cannot increase when selection may use the full history rather than only the archive. This inclusion argument explains the representational opportunity created by TSC, whereas ASC supplies a constructive approximation within the available domain. It does not assert that two greedy traversals must be ordered on every instance. Nor does the argument require the archive to lose the winner, since retaining alternatives below its reward cutoff is the relevant distinction.

The reference set and threshold must remain fixed in this comparison. Across search iterations, the best reward may increase and raise $r^+_t - \tau$, so an append-only history need not produce nested eligible domains or a monotonically decreasing support deficit. Likewise, a smaller pool can have a smaller self-covering radius simply because difficult-to-represent candidates have been removed. A direct retention comparison therefore evaluates both selected sets against E_H using the same cached BLIP descriptors, anchor, tolerance, and output cardinality. The resulting radius measures representation of discovered eligible support, whereas CLIP pairwise distance measures separation among returned images. These are complementary quantities, and neither the coverage bound nor the set-inclusion argument alone establishes improved human preference.

C Additional experimental results

C.1 Benchmark validation

The full benchmark and disjoint validation provide complementary evidence that SPG improves both reward quality and semantic separation under a fixed generation budget. Table 1 measures both objectives on the same 100 prompts, where best-of-four reward improves over SPT on 96, 92, and 85 prompts and CLIP diversity improves on 81, 82, and 80 prompts for SD 1.4, SD 1.5, and SDXL, respectively. Table 2 then confirms positive average changes in both metrics after excluding the ten prompts used to select $\tau = 0.75$, so the benchmark-level advantage is also present in the disjoint 90-prompt evaluation without further tuning. Together, the prompt-level counts and held-out means show that the improvements are not confined to the development examples, while preserving the distinction between average performance and per-prompt outcomes.

C.2 COMPONENT ANALYSIS

Table 6: Paired comparisons for TSC and ASC on 100 prompts. Differences are full method minus control; 95% confidence intervals use 10,000 paired-prompt bootstrap samples.

Direct full-minus-control contrast	Δ	95% CI	Wins
TSC vs. IID: best IR	+0.189	[0.153, 0.228]	96/100
ASC vs. top-four: CLIP diversity	+0.306	[0.274, 0.341]	98/100

The controls separate adaptive discovery from output construction. Comparing IID sources with TSC while retaining ASC and persistent storage tests the proposal policy as a whole, whereas applying top-four ranking and ASC to the same TSC history tests the selector without changing discovery. Neither contrast separately isolates history retention from the final archive, and the CLIP statistic is not a measurement of the BLIP covering radius. The set-inclusion result in Appendix B supplies the structural rationale for retaining history without attributing either of these unmeasured effects to the existing comparisons.

The same-pool comparison links the improvement in finite-output diversity to ASC, because candidate generation and the observed reward maximum are identical before selection. Applying ASC instead of top-four ranking raises CLIP diversity from 0.102 to 0.409 while retaining best reward exactly for every prompt (Figure 1, Table 3), with a paired increase of 0.306 and a 95% interval of [0.274,0.341]. Diversity improves on 98 prompts, remains unchanged on one, and decreases on one. Since no additional candidates are generated, the comparison demonstrates that ASC exposes complementary outputs already present in the TSC history, rather than benefiting from a larger search budget. The selected set-mean reward of 1.140 is interpreted together with the tolerance-dependent statistics below, which distinguish anchor preservation from the quality of the remaining outputs.

The tolerance sweep connects the selection effect to an explicit quality control, as Figures 4 and 6(a) trace the frontier on the same 100 evaluated pools. Tightening τ from 0.75 to 0.25 raises set-mean reward from 1.140 to 1.430 while changing diversity from 0.409 to 0.298, whereas increasing it to 1.50 extends diversity to 0.484 with mean reward 0.682. Best reward remains 1.572 across all settings, so the frontier concerns non-anchor output slots rather than replacement of the strongest candidate. Tables 7 and 9 report portfolio statistics and comparisons with SPT, respectively, with the development-selected tolerance held fixed in the primary comparison.

Table 7: Output statistics for different reward tolerances on 100 prompts. Best reward averages 1.572 throughout. Non-anchor and minimum rewards are averaged over prompts. Reward-floor and fallback checks are reported below.

τ	Set mean	Non-anchor mean	Minimum	CLIP diversity
0.00	1.547	1.539	1.531	0.102
0.25	1.430	1.382	1.345	0.298
0.50	1.277	1.179	1.112	0.361
0.75	1.140	0.997	0.894	0.409
1.00	0.992	0.798	0.651	0.438
1.25	0.821	0.571	0.410	0.469
1.50	0.682	0.385	0.172	0.484

Recorded selection indices and search-path rewards verify that eligibility is enforced on individual candidates rather than only in aggregate. At $\tau = 0.75$, the eligible pool contains 5–408 candidates per prompt, every selected record satisfies $r \geq r^{+} - 0.75$, and no fallback is required. At $\tau = 0.25$ and 0.50, one prompt uses top-four fallback and violates the nominal floor, whereas all tested tolerances of at least 0.75 satisfy it within 10^{-6}. At zero tolerance, all 100 prompts use fallback. Search-time and finally decoded selected rewards agree in this ablation batch. These checks connect the implementation to the conditional guarantee in Appendix B, while image-pair distances remain measurements of returned-set separation rather than eligible-pool covering radii.

Table 6 complements the ASC intervention with a discovery comparison that holds storage, selector, and budget fixed, under which replacing IID sources with TSC raises best ImageReward from 1.382 to 1.572 across 100 prompts. Read together with the identical-pool result, the two controls assign

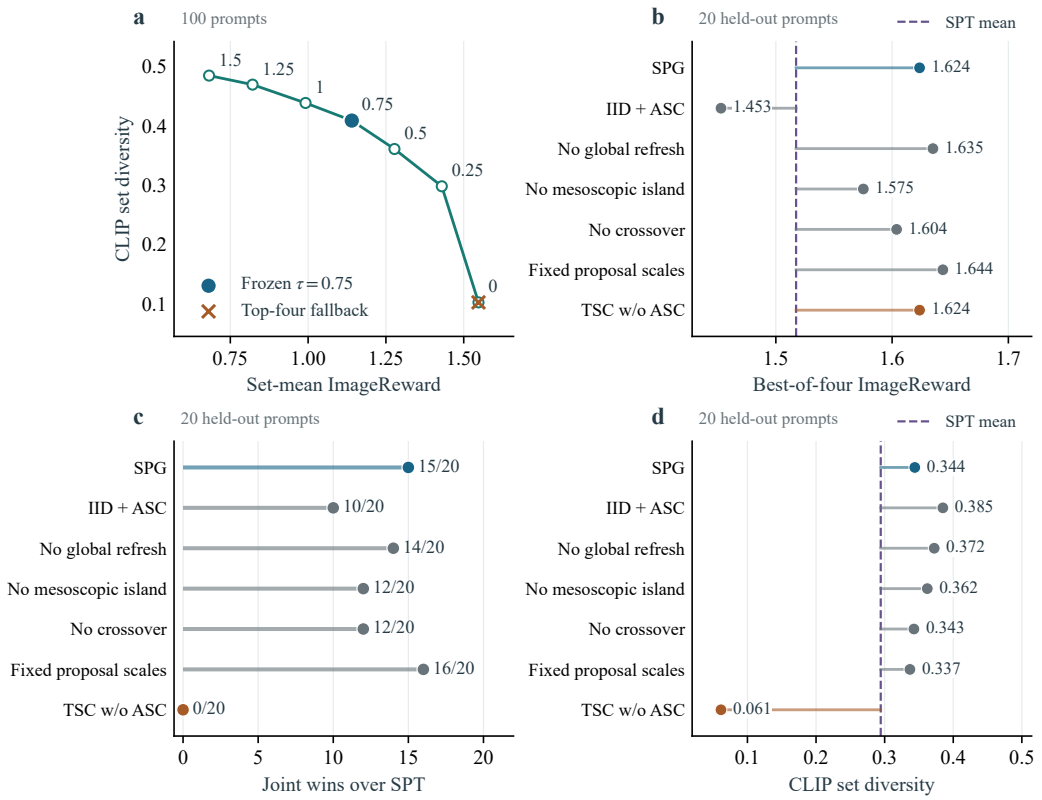


Figure 6: Selection tolerances on 100 prompts and search variants on a separate 20-prompt held-out subset. Panel (a) shows the 100-prompt reward–diversity frontier, with the selected tolerance in blue and top-four fallback marked by a cross. Panels (b–d) use the original 20-prompt runs, reporting metric means relative to SPT and joint wins. Confidence intervals for these variants appear in Table 8.

different empirical roles to the components, with TSC improving the strongest discovered candidate and ASC representing additional alternatives without changing that candidate. Because each contrast fixes the interface relevant to its component, the analysis supports the combined discovery–selection design without adding effects from different comparisons. Confidence intervals use paired-prompt bootstrap resampling without multiple-testing correction. In the proposal ablations, removing global refresh replaces its four Gaussian draws with mirrored pCN moves at $\beta = 0.90$, removing the mesoscopic group replaces it with four Gaussian draws, removing crossover retains mesoscopic pCN moves on every round, and fixing scales uses $(0.45, 0.10)$ throughout. These replacements preserve the 420-candidate budget rather than deleting evaluations.

C.3 Sensitivity analysis

On the separate 20-prompt held-out subset, Figure 6(b,d) and Table 8 compare the full system with proposal variants that retain ASC, alongside a separate top-four selection control, allowing the reward and diversity consequences of search choices to be read against the same SPT reference. SPG reaches best reward 1.624 and diversity 0.344, with positive paired intervals for both differences from SPT, whereas replacing TSC with IID sources lowers best reward to 1.453 despite retaining the same selector and evaluation budget. The top-four control retains the full system's best reward but returns diversity of only 0.061, illustrating why improved discovery and complementary selection are evaluated as distinct components. Panel (c) supplies a prompt-level joint summary, with SPG exceeding SPT on both metrics for 15 of the 20 prompts. These operator-study values are retained from the original 20-prompt runs and are not pooled with the expanded 100-prompt component study.

Table 8: Comparison of search and selection variants on 20 held-out prompts using SD 1.4. Entries report means, paired differences from SPT, and bootstrap 95% confidence intervals. Joint wins count prompts with higher best reward and CLIP diversity than SPT.

Variant	Best reward	Δ [95% CI]	CLIP diversity	Δ [95% CI]	Joint
SPG	1.624	+.106 [.076,.141]	.344	+.050 [.015,.086]	15/20
IID + ASC	1.453	−.065 [−.153,.011]	.385	+.091 [.061,.124]	10/20
No global refresh	1.635	+.118 [.081,.160]	.372	+.078 [.027,.135]	14/20
No mesoscopic island	1.575	+.058 [−.019,.124]	.362	+.068 [.016,.120]	12/20
No crossover	1.604	+.086 [.021,.143]	.343	+.048 [.018,.080]	12/20
Fixed proposal scales	1.644	+.126 [.091,.169]	.337	+.043 [.015,.068]	16/20
TSC w/o ASC	1.624	+.106 [.076,.141]	.061	−.233 [−.297,−.178]	0/20

Table 9: Comparison of reward tolerances on identical pools for 100 prompts. Best ImageReward averages 1.572 throughout. Joint wins require both best reward and diversity to exceed SPT.

τ	Set-mean ImageReward	CLIP diversity	Joint wins vs. SPT
0.00	1.547	0.102	1/100
0.25	1.430	0.298	46/100
0.50	1.277	0.361	61/100
0.75	1.140	0.409	73/100
1.00	0.992	0.438	78/100
1.25	0.821	0.469	83/100
1.50	0.682	0.484	83/100

C.4 Generalization analysis

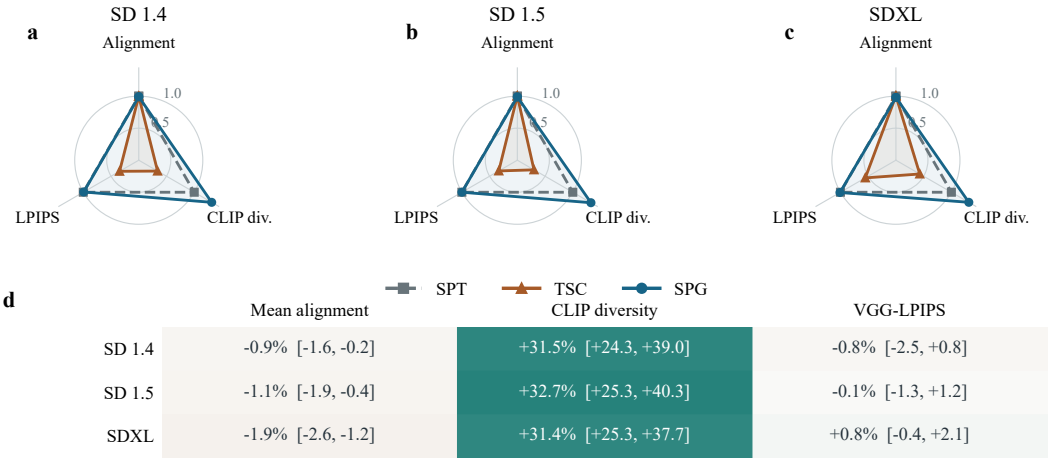


Figure 7: Comparison using independent evaluators on 3,600 main-benchmark outputs from three methods and three backbones, with 100 prompts and four images per method–backbone pair per prompt. Panels (a–c) show metric ratios relative to SPT; the dashed line denotes one. Panel (d) reports relative changes for SPG with paired-bootstrap 95% confidence intervals.

Independent re-evaluation tests whether the increased separation remains visible when the feature encoder changes, rather than assessing outputs only with the primary CLIP-L/14 representation. The evaluated set contains $3 \times 3 \times 100 \times 4 = 3{,}600$ outputs from three methods (SPT, TSC with top-four output, and SPG), three backbones, 100 prompts, and four returned images per prompt, excluding the additional ablation outputs. Across SD 1.4, SD 1.5, and SDXL, CLIP-B/16 diversity exceeds SPT by 31.50%, 32.71%, and 31.44%, respectively, with paired intervals excluding zero throughout (Figure 7), while Table 10 reports alignment and VGG-LPIPS on the same 3,600 saved images. Because these evaluators are not used during search or tolerance selection, their diversity measurements support transfer of semantic separation beyond the primary evaluator. Alignment and perceptual-distance results remain part of the complete evaluation, with their implications discussed together in Appendix D.2.

Table 10: Comparison of CLIP-B/16 alignment, CLIP-B/16 diversity, and VGG-LPIPS across the 3,600 main-benchmark outputs. These evaluators are not used during search or selection. Alignment averages all four outputs. Diversity and LPIPS measure within-set distances. Confidence intervals appear in Figure 7.

Backbone	Method	CLIP-B align. ↑	CLIP-B div. ↑	LPIPS ↑
SD 1.4	SPT	.330	.254	.742
	TSC	.329	.086	.257
	SPG	.327	.335	.736
SD 1.5	SPT	.330	.250	.737
	TSC	.331	.075	.247
	SPG	.326	.331	.737
SDXL	SPT	.327	.273	.734
	TSC	.328	.118	.405
	SPG	.321	.359	.740

Figure 8 resolves the held-out comparison at prompt level, with panels (a,b) showing changes relative to SPT rather than raw metric values and panel (c) plotting reward and diversity changes jointly. Positive mean markers on all three backbones agree with the held-out summary, while the upper-right region of the joint plot identifies prompts on which both objectives improve together. By displaying every held-out prompt alongside the means, the figure connects the aggregate advantage to its distribution across examples rather than relying on selected visual cases. The resulting view supports a positive average shift in both objectives without implying that every individual prompt has the same response.

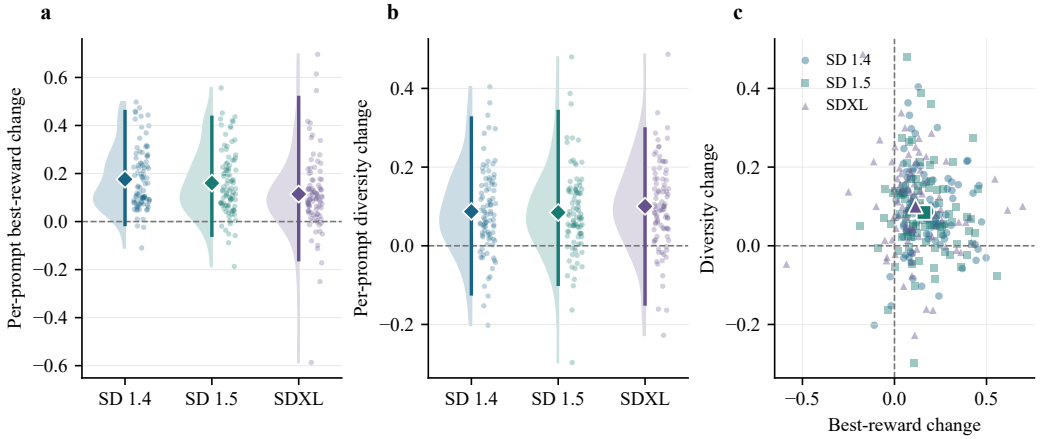


Figure 8: Distribution of prompt-level results on 90 held-out prompts per backbone. Panels (a,b) show prompt-level changes from SPT, with dots for prompts, diamonds for means, and bars for the 2.5–97.5 percentiles. Panel (c) shows paired changes from SPT. Colors and shapes identify backbones, and large white-edged markers denote means.

On the ten development prompts at seed 42, the official-entry-point SPT baseline used in the main benchmark has a mean best-of-four ImageReward of 1.345, compared with 1.557 for TSC under the same 420-candidate budget. The absolute improvement is 0.213 (95% paired-bootstrap interval [0.134,0.309]), corresponding to a 15.81% relative increase, with higher reward on all ten prompts. This subset analysis describes the reward-guided discovery component within the existing benchmark, rather than providing an additional independent evaluation of the complete portfolio method.

D QUALITATIVE RESULTS

D.1 COMPACT COMPLETE GALLERY

Figures 9 and 10 present seven additional prompts and 84 outputs that do not appear in the main qualitative comparison, with each case showing all four outputs from SPT, TSC with top-four output, and SPG. The logo, crow, red-car, and library cases in Figure 9 extend the comparison to graphic design, character rendering, object viewpoints, and architectural scenes, where SPG retains its anchor while providing alternatives in visual treatment or composition. Figure 10 adds food imagery, a full-length portrait, and digital artwork, illustrating how the returned set spans different subject presentations and stylistic realizations. Together, these examples complement the main figure through additional prompt content rather than repeated samples, while the separate failure case appears only in Figure 11. All methods retain their saved output order, the first SPG border marks the anchor, and full prompts and image provenance are recorded in `data/qualitative_compact_manifest.json`. These are post-hoc illustrative cases displayed at their original aspect ratios, without cropping or retouching.

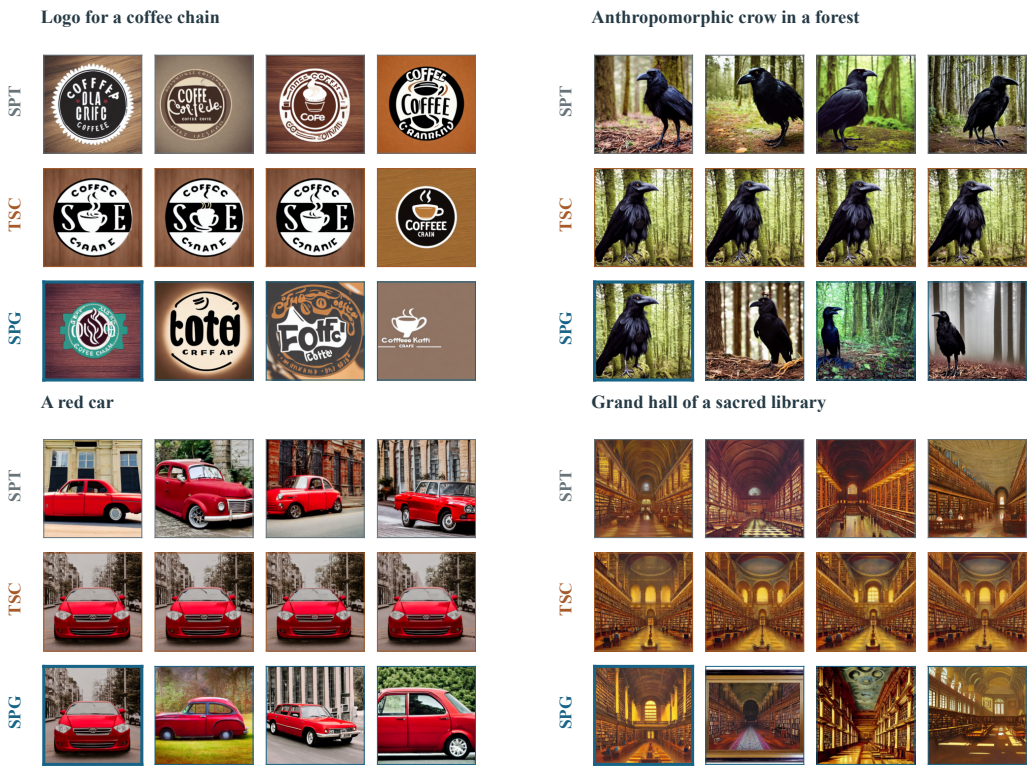


Figure 9: Additional samples generated by SPT, TSC with top-four output, and SPG (part 1 of 2). All four outputs are shown for each method and prompt.

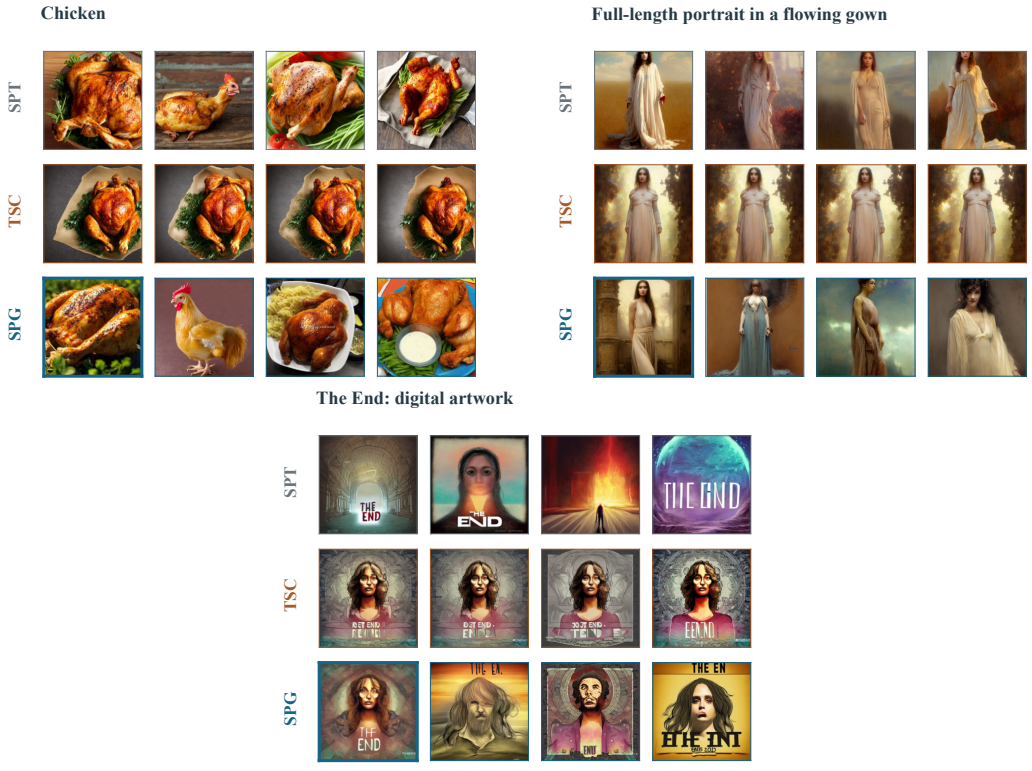


Figure 10: Additional samples generated by SPT, TSC with top-four output, and SPG (part 2 of 2). All four outputs are shown for each method and prompt.

D.2 Failure cases and limitations



Figure 11: Comparison of generated samples on a failure case. Greater feature separation does not necessarily improve prompt adherence or remove rendering artifacts.

Figure 11 illustrates the limits of reward-constrained coverage. In the 100-prompt same-pool ablation, ASC preserves best reward exactly while reducing set-mean reward by 0.407, with a paired 95% interval of $[-0.422, -0.391]$, so anchor preservation does not imply unchanged quality for every output. An anchor can retain an incorrect detail overlooked by the reward model, while descriptor separation can favor color or texture changes over useful semantic alternatives. Selection also remains limited to the support discovered during search. These effects help explain why mean CLIP-B/16 alignment decreases by 0.89%, 1.14%, and 1.91% even as semantic diversity increases. VGG-LPIPS changes of -0.78%, -0.06%, and $+0.83\%$ have intervals containing zero (Figure 7), so semantic separation alone does not establish improved perceptual diversity or human preference.

The evidence covers 100 inherited prompts, three Stable Diffusion backbones, and one primary seed under a substantial fixed search budget. The ten-prompt development analysis evaluates TSC rather than the complete portfolio method. Operator variants also achieve higher values on some metrics (Table 8), so the results support the discovery–selection design without establishing a universally optimal proposal schedule. Coverage guarantees apply to a fixed eligible domain with sufficient candidates, and the top-four fallback preserves the winner but not the nominal reward floor.

E Additional related work

E.1 Generative modeling and preference alignment

Diffusion and flow models supply reusable priors for test-time guidance, with early diffusion formulations learning to reverse gradual corruption (Sohl-Dickstein et al., 2015; Ho et al., 2020) and score-based methods estimating gradients of perturbed data distributions (Song & Ermon, 2019; Song et al., 2020b). Improved diffusion parameterizations and design-space analyses refine sampling efficiency and model quality (Nichol & Dhariwal, 2021; Karras et al., 2022), while DDIM provides a sampling construction that supports deterministic source-to-image mappings (Song et al., 2020a) and flow matching and rectified flow learn transport vector fields (Lipman et al., 2022; Liu et al., 2022). For text-conditioned synthesis, GLIDE studies text-guided generation and editing (Nichol et al., 2021), Imagen strengthens the use of language representations (Saharia et al., 2022), latent diffusion moves generation into a compressed image space (Rombach et al., 2022), and hierar-

chical CLIP-latent generation links text and images through a semantic representation (Ramesh et al., 2022). DiT and rectified-flow transformers further scale the generative backbone (Peebles & Xie, 2023; Esser et al., 2024), whereas SPG addresses how the evaluated candidates of a fixed backbone are organized into a small returned set.

Preference alignment incorporates feedback through different learning interfaces, with DDPO treating denoising as a multistep decision process optimized by policy gradients (Black et al., 2024) and DPOK combining online policy optimization with KL regularization against the pretrained model (Fan et al., 2023). DRaFT differentiates rewards through sampling and reduces optimization cost through truncated backpropagation (Clark et al., 2024), AlignProp similarly aligns text-to-image models through reward backpropagation (Prabhudesai et al., 2023), and Diffusion-DPO derives an objective from pairwise preferences and a variational treatment of diffusion likelihood (Wallace et al., 2024). These methods incorporate preference information into model parameters, whereas SPG leaves both generator and reward model fixed and reuses evaluations made for the current prompt. The distinction concerns where alignment decisions are made, rather than imposing a new structural assumption on the generative prior.

E.2 Test-time guidance and search

Guidance with frozen generators can act on denoising updates or on their input sources, with classifier guidance using noise-conditioned classifier gradients (Dhariwal & Nichol, 2021) and classifier-free guidance combining conditional and unconditional predictions to strengthen conditioning (Ho & Salimans, 2022). Universal Guidance accommodates more general task feedback (Bansal et al., 2024), FreeDoM introduces energy-based conditions from pretrained networks (Yu et al., 2023), and TFG organizes training-free guidance methods into a common design space with a systematic hyperparameter-search strategy (Ye et al., 2024). For inverse problems, diffusion posterior sampling combines measurement feedback with diffusion updates (Chung et al., 2022), addressing constraints that differ from completed-image black-box reward queries. These approaches establish flexible inference-time control, while SPG additionally treats retention of evaluated candidates and the final return rule as explicit parts of a finite-output decision.

Among source-space methods, SPT moves reward reweighting to the generator input and explores the tilted law through pCN proposals and exchanges between tempered chains (Wang et al., 2026), drawing on function-space MCMC and parallel tempering (Cotter et al., 2013; Earl & Deem, 2005). Evolution strategies use rewards of random perturbations to estimate optimization directions (Salimans et al., 2017), whereas EvoSearch evolves image and video denoising trajectories with selection and mutation mechanisms designed to improve quality while preserving population diversity (He et al., 2025). Feynman–Kac correctors instead use weighted simulation and sequential Monte Carlo resampling to correct sampling paths for annealed or composed distributions (Skreta et al., 2025). Table 11 summarizes these differences in search objects and returned objectives, with SPG retaining the evaluated source history and compressing it into K representatives under reward eligibility and a fixed anchor. This objective does not require the returned set to follow SPT's stationary target law and does not presume that prior search methods disregard diversity.

Table 11: Comparison of inference-time guidance methods by their optimization targets.

Method	Search object	Feedback interface	Returned-object objective
Best-of-N	Gaussian sources	terminal score	top-ranked points
SPT (Wang et al., 2026)	Gaussian sources	black-box reward	tempered point samples
EvoSearch (He et al., 2025)	denoising population	evaluator score	high-scoring sample(s)
FK correctors (Skreta et al., 2025)	particle paths	potentials	guided distribution
SPG	full source history	reward + descriptor	anchored constrained set

E.3 Set construction and evaluation

Quality and diversity can be coupled through different set objectives, with MAP-Elites retaining high-performing solutions across user-defined behavioral dimensions (Mouret & Clune, 2015) and determinantal point processes representing repulsion and subset preference through a similarity-kernel determinant (Kulesza & Taskar, 2012). Farthest-first K-center summarizes a finite feature set through

its worst nearest-representative distance, whose classical approximation guarantee underlies the present coverage analysis (Gonzalez, 1985). Compared with behavioral partitioning, probabilistic repulsion, or unconstrained geometric summarization, ASC first defines admissibility through reward and then requires the observed winner, assigning distinct roles to quality eligibility, anchor retention, and geometric representation. TSC supplies the full evaluated history for these decisions, so the methodological contribution concerns applying constrained set construction to generation outputs rather than introducing a new covering theorem.

Evaluation of a finite portfolio distinguishes candidate preference, text association, and variation among outputs. ImageReward learns text-to-image preference scores (Xu et al., 2023), HPS v2 provides a human-preference evaluation signal (Wu et al., 2023), and Pick-a-Pic collects users' pairwise choices between generated images (Kirstain et al., 2023), while CLIP and BLIP provide different vision–language representations (Radford et al., 2021; Li et al., 2022), CLIPScore studies reference-free image–text relevance (Hessel et al., 2021), and LPIPS measures perceptual differences through deep features (Zhang et al., 2018). FID and generative precision and recall instead characterize distribution-level behavior (Heusel et al., 2017; Kynkäänniemi et al., 2019), which differs in granularity from four outputs per prompt. Dataset roles are likewise distinct, with LAION-5B supporting large-scale image–text training (Schuhmann et al., 2022), COCO providing object and scene annotations (Lin et al., 2014), DiffusionDB recording user prompts (Wang et al., 2023), and T2I-CompBench and GenEval testing compositional relations and object-focused alignment (Huang et al., 2023; Ghosh et al., 2023). The present study retains SPT's prompt protocol for matched comparison and uses an independent encoder to assess transfer of set separation, with representation-dependent limitations discussed separately.

F Artifact records

F.1 Reproducibility

The artifact audit verifies 100 score files and 400 images per method–backbone pair, fixed prompt membership, and finite scores, with no recorded failures in the audited SPG runs and byte-identical scores for one replayed prompt. Experiments run on NVIDIA A40 GPUs under Slurm, with software versions, launch scripts, batch-counted search NFE, and wall-clock diagnostics linking the configuration in Table 4 to execution records. Appendix A specifies the prompt split, seed, tolerance, and final decoding procedure, while the artifact retains images, diagnostics, paired-analysis programs, and plotting code. Reward-floor checks use archived selection records rather than new image generation, and historical filenames remain unchanged to preserve provenance.