BOUNDING THE INITIAL-NOISE DENSITY RATIO KEEPS EVERY IMAGE UNDER REWARD ALIGNMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Reward fine-tuning raises the score of text-to-image models under a reward model but makes many correct images of a prompt hard to sample: on Stable Diffusion 3.5-Medium, DRaFT raises PickScore by 1.41 while the share of base-model images its outputs cover falls from 94.8% to 34.7%. Penalties on KL divergence, diversity, or update size control averages, which a rare composition barely moves. We instead require every set of images for a prompt to keep its probability within fixed factors $[\ell, u]$ of the base model. Our key observation is that a frozen sampler maps each initial noise to one image, so the probability of an image set is the Gaussian mass of its noise preimage, and bounding the density ratio of the initial noise bounds every image set at once. RATIO learns a prompt-conditioned invertible noise transform whose density ratio stays in $[\ell, u]$ at every point, for any latent dimension, throughout training: Gaussian-preserving rotations choose directions at no cost, and monotone maps in Gaussian-CDF coordinates move probability under one budget shared by all coordinates. At the same PickScore gain, RATIO keeps 85.7% recall against 69.2–80.6% for seven fine-tuning and noise-space methods, and tests on 5,051 GenEval image events find no violation of its bounds, against 179–386 for the controls. Across three generators and three rewards, the bound fixes how much probability moves, and the reward decides which images receive it.

1 INTRODUCTION

Text-to-image diffusion and flow models such as Stable Diffusion 3.5 (SD3.5), FLUX.1, and SANA are routinely fine-tuned toward reward models of human preference, such as PickScore and HPS, or of prompt correctness, such as GenEval (Stability AI, 2024; Black Forest Labs, 2024; Xie et al., 2025; Kirstain et al., 2023; Wu et al., 2023; Ghosh et al., 2023). A prompt such as "a red cup to the left of a blue bowl" has many correct images, seen from above or from the side, on a kitchen table or a studio backdrop (Figure 1 shows two prompts). Fine-tuning raises the average reward and also changes how often each of these images is drawn. On SD3.5-Medium, DRaFT (Clark et al., 2024) raises PickScore from 21.98 to 23.39, while PRDC recall, the share of independent base-model images that its outputs still cover (Naeem et al., 2020), falls from 94.8% to 34.7% (Table 1). Many correct images that the base model offered for a prompt are no longer produced.

Existing remedies regularize averages: a KL divergence to the base model (Fan et al., 2023; Uehara et al., 2024), a diversity reward (Liu et al., 2025a), the clipping range of a policy ratio (Wang et al., 2025a), the sharpness of the reward (Kim & Sim, 2026), or the scale of the weight update. Averages are dominated by frequent outputs. If overhead compositions make up 1% of the base model's images for a prompt, an aligned model can make them ten times rarer while $D_{\mathrm{KL}}(Q \,\|\, P)$ grows by less than 0.007 nats. In our experiments, an affine noise transform trained with a KL penalty to the same mean noise KL as our method removes 232 prompt-required GenEval events that our method keeps (Section 5.2).

We therefore require a guarantee for every set of images. For a prompt c and any measurable set A of images, such as all correct overhead compositions of the cup-and-bowl prompt, the aligned probability $Q_c(A)$ must stay between $\ell P_c(A)$ and $uP_c(A)$, where P_c is the image distribution of the base model. The lower factor ℓ keeps a fixed share of everything the base model can draw, however rare, and the upper factor u limits concentration. The requirement concerns a density-free image distribution and infinitely many sets never named during training.

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.

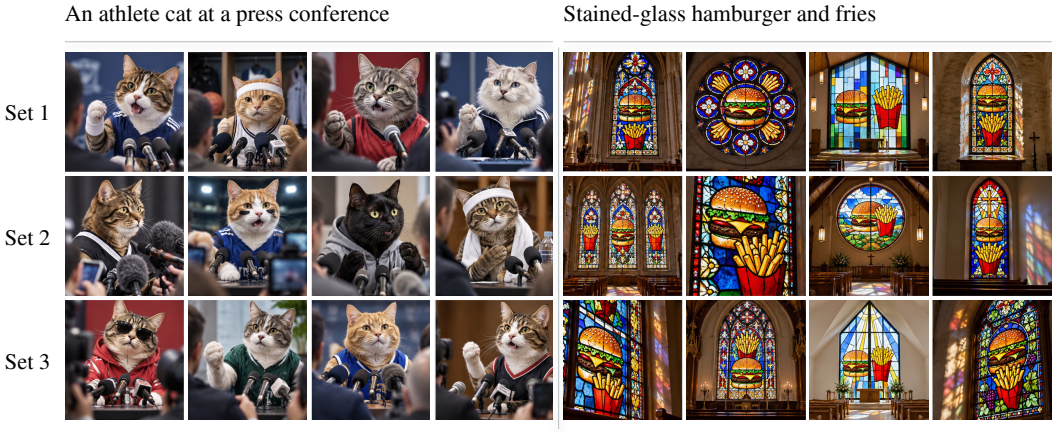


Figure 1: **One prompt, many correct images.** Twelve SD3.5-M outputs for each of two DrawBench prompts: an athlete cat at a press conference, and a church with stained-glass depictions of a hamburger and french fries. Pose, viewpoint, lighting, and layout vary within each prompt. RATIO keeps the probability of every set of such images within fixed factors of its base probability.

Our key observation is that the generator makes this requirement tractable. With a fixed deterministic sampler and decoder, SD3.5 turns each initial noise z into one image $G(z, c)$, so the probability of an image set A is the Gaussian mass of its noise preimage $\{z : G(z, c) \in A\}$. If we freeze the generator and replace only the Gaussian noise by a learned density $q(z \mid c)$, a pointwise bound $\ell \le q(z \mid c)/\gamma(z) \le u$ on the noise density ratio bounds all image sets at once, including sets that nobody specified. A requirement on images becomes a requirement on one function in noise space, which an invertible transform computes exactly.

Existing noise-space methods move the noise by per-sample optimization or residual maps $z \mapsto z + f(z, c)$ (Eyring et al., 2024; 2025), and a Gaussian shifted by a already has the unbounded density ratio $\exp(a^\top z - \|a\|^2/2)$; normalizing flows compute exact densities without bounding them (Dinh et al., 2017; Durkan et al., 2019). We build RATIO (**R**atio-bounded **A**lignment **T**hrough **I**nitial-noise **O**ptimization), a noise transform designed for this constraint. It reweights coordinates in Gaussian-CDF coordinates, where the density ratio is the inverse derivative of a monotone map on $[0, 1]$ and stays bounded in the tails; it limits all coefficients of a block by one shared budget, so the bound holds for the 65,536 latent coordinates of SD3.5 as for any dimension; and it chooses the directions to reweight with prompt-conditioned rotations, which leave the Gaussian unchanged and consume none of the budget. One scalar ρ sets the bounds: at $\rho = 2$, every image set keeps at least 17.04% and at most 7.39 times its base probability, at every training step.

The bounds also determine the attainable reward: the reward-optimal distribution within them keeps a copy ℓP_c of the base model and moves the remaining mass to the highest-reward images (Proposition 2). On SD3.5-Medium with PickScore, RATIO keeps 85.7% recall at a gain of 0.596, while seven fine-tuning and noise-space methods reach 69.2–80.6% at the same gain; tests on 5,051 GenEval image events find no violation for RATIO and 179–386 for the controls. Our contributions are (i) retention as a two-sided bound on every image set, which a frozen generator reduces to a pointwise bound on the initial-noise density ratio; (ii) RATIO, a noise transform whose bound is exact, dimension-free, and valid throughout training, with its attainable reward; and (iii) tests of the bound on coverage, image events, three generators, and three rewards.

2 Related Work

Reward fine-tuning of text-to-image models. Early work fine-tuned diffusion models on reward-weighted samples (Lee et al., 2023); two families then became standard. Policy-gradient methods treat denoising as a sequential decision process, as in DDPO and DPOK (Black et al., 2024; Fan et al., 2023); Flow-GRPO, DanceGRPO, and DiffusionNFT extend this to flow models such as SD3.5 and FLUX with group-relative or forward-process objectives (Liu et al., 2025b; Xue et al., 2025; Zheng et al., 2025). Gradient-based methods backpropagate a differentiable reward through sampling, as in ReFL, AlignProp, and DRaFT (Xu et al., 2023; Prabhudesai et al., 2023; Clark et al., 2024), or fit

a reward-tilted target by stochastic optimal control (Domingo-Enrich et al., 2025); Diffusion-DPO learns from preference pairs directly (Wallace et al., 2024). These methods raise reward efficiently on billion-parameter generators, and they share a failure mode, over-optimization, in which the model concentrates on the images the reward ranks highest (Zhang et al., 2024). The standard remedies regulate averages. KL regularization targets the tilted distribution $P\exp(r/\beta)/Z$, whose ratio to the base is controlled only through β and the reward range, and the trained model inherits no bound (Fan et al., 2023; Uehara et al., 2024); DiverseGRPO adds a diversity reward, GRPO-Guard regulates ratio clipping, and RSA-FT penalizes reward sharpness (Liu et al., 2025a; Wang et al., 2025a; Kim & Sim, 2026). The bottleneck is shared: these methods control how much the model changes on average, while what disappears in practice is the probability of particular images. RATIO keeps the reward objective of this line and replaces the average penalty with a bound on every image set that holds at every step.

Initial noise as the alignment interface. A second line keeps the generator weights frozen and changes its input. The initial noise strongly affects layout and prompt adherence (Qi et al., 2024; Guo et al., 2024), so test-time methods optimize each noise by backpropagating through the sampler, as in DOODL, D-Flow, and ReNO (Wallace et al., 2023; Ben-Hamu et al., 2024; Eyring et al., 2024), or search over candidate noises (Ma et al., 2025a). Amortized methods train a network that maps a Gaussian noise to a better one, such as golden-noise predictors and noise hypernetworks (Zhou et al., 2025; Eyring et al., 2025). This line preserves the pretrained generator, but it treats the noise as a variable to improve. Per-sample optimization replaces the Gaussian by a set of optimized points, and residual maps $z \mapsto z + f(z, c)$ shift the Gaussian tails, so the induced density ratio is unbounded and the probability of an image set can fall arbitrarily far. Invertible networks from the normalizing-flow literature compute exact densities (Dinh et al., 2017; Kingma & Dhariwal, 2018; Durkan et al., 2019), yet no existing architecture bounds its density ratio to the Gaussian at every point, and per-coordinate bounds deteriorate with the latent dimension. RATIO supplies such a transform, with an exact, pointwise, dimension-free ratio bound, and uses the noise interface to carry a probability guarantee to every image.

3 Keeping every image set within fixed probability factors

Setup. Let c be a prompt drawn from a distribution ν. A pretrained text-to-image model with a fixed sampler, guidance weight, and decoder defines a map $G(\cdot, c) : \mathbb{R}^D \to \mathcal{X}$ from initial noise to images; for SD3.5-Medium at 512^2 resolution, $D = 16 \times 64 \times 64 = 65{,}536$. The base model draws $Z \sim \gamma = \mathcal{N}(0, I_D)$, and $\gamma(z)$ also denotes its density. Its image distribution is $P_c = G(\cdot, c)_{\#}\gamma$, the law of $G(Z, c)$. RATIO keeps G frozen and replaces Z by $T_\phi(Z, c)$, where $T_\phi(\cdot, c)$ is invertible with noise density $q_\phi(\cdot \mid c)$; the aligned image distribution is $Q_c = G(\cdot, c)_{\#}q_\phi(\cdot \mid c)$. For a reward $r(x, c)$ we solve

$$\max_{\phi} \ \mathbb{E}_{c\sim\nu}\, \mathbb{E}_{X\sim Q_c}\big[r(X, c)\big] \quad \text{s.t.} \quad \ell\, P_c(A) \le Q_c(A) \le u\, P_c(A) \ \ \forall c,\ \forall A \in \mathcal{B}(\mathcal{X}), \tag{1}$$

with $0 < \ell \le 1 \le u$, where $\mathcal{B}(\mathcal{X})$ is the set of measurable image sets. We call such a set an *event*: all images in which a property holds, such as a composition (correct overhead views), a count (exactly three balls), or a local attribute (a purple cup) (examples in Figure 10, Appendix E.3). The constraint is equivalent to $\ell \le dQ_c/dP_c \le u$ almost everywhere (Appendix A.3), so it protects a rare event exactly as it protects a common one. Divergence constraints lack this property: Appendix A.4 constructs distributions whose KL and Rényi divergences in both directions are arbitrarily small while one rare event grows by any factor M and another shrinks by $1/M$.

From images to noise. The ratio dQ_c/dP_c cannot be evaluated for a diffusion model, but the frozen generator relates it to a quantity that can. Write $\Lambda_\phi(z \mid c) = q_\phi(z \mid c)/\gamma(z)$ for the noise density ratio and $\mathcal{A}_c = \{z : G(z, c) \in A\}$ for the noise preimage of an event. Because both models push noise through the same map,

$$P_c(A) = \gamma(\mathcal{A}_c), \qquad Q_c(A) = \int_{\mathcal{A}_c} \Lambda_\phi(z \mid c)\, \gamma(z)\, dz. \tag{2}$$

Hence $\ell \le \Lambda_\phi \le u$ on $\mathbb{R}^D$ implies (1) for every event simultaneously. The identity requires only a measurable G, so it covers every sampler that is deterministic given the initial noise; Appendix A.3 extends it to samplers that inject noise from a fixed distribution. RATIO thus reweights images that

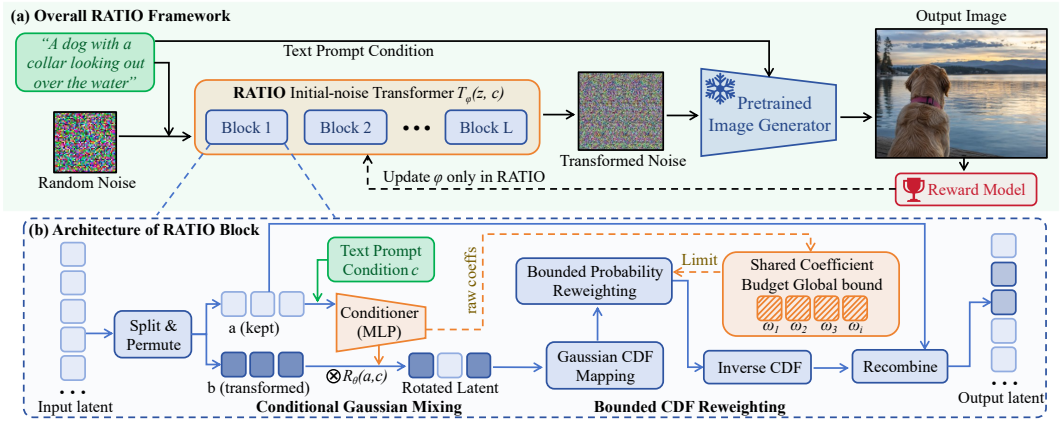


Figure 2: **RATIO changes only the initial noise, and each block keeps the noise density ratio bounded.** (a) A Gaussian draw passes through L invertible blocks before the frozen generator; reward gradients update only ϕ. (b) Each block keeps a part a of its input, rotates the other part b with angles predicted from a and the prompt (conditional Gaussian mixing, which leaves the Gaussian unchanged), and reweights the rotated coordinates in Gaussian-CDF coordinates under one shared coefficient budget. The block bounds multiply into (5).

the base model already produces, which makes a guarantee over all events possible, and Proposition 2 gives the reward that remains available.

4 RATIO: A NOISE TRANSFORM WITH A BOUNDED DENSITY RATIO

By (2), retaining every event reduces to designing an invertible, prompt-conditioned noise transform whose density ratio stays in $[\ell, u]$, with three requirements. (R1) The ratio must be bounded at every z, including the Gaussian tails. (R2) The bound must hold uniformly in the dimension D. (R3) The transform must still move probability along the directions that change the reward, which are generally combinations of many latent coordinates. RATIO composes L invertible blocks, $T_\phi = F_L \circ \cdots \circ F_1$ (Figure 2); each block splits its input into a kept part a and a transformed part $b \in \mathbb{R}^{d_b}$, rotates b, and reweights the rotated coordinates. Sections 4.1–4.3 derive each operation from one requirement.

4.1 REWEIGHTING IN GAUSSIAN-CDF COORDINATES BOUNDS THE RATIO AT EVERY POINT

Residual and affine maps violate (R1) because they change the Gaussian tails: a shift by a multiplies the density by $\exp(a^\top z - \|a\|^2/2)$, and no Gaussian other than γ has a two-sided bounded ratio to γ (Appendix A.4). Gaussian-CDF coordinates give a nonlinear map that preserves the relative weight of the tails. If $\widetilde{b}_j \sim \mathcal{N}(0,1)$, then $v_j = \Phi(\widetilde{b}_j)$ is uniform on $[0,1]$; an increasing bijection S of $[0,1]$ turns this uniform density into $1/S'$, and mapping back with Φ^{-1} returns a coordinate whose density ratio to $\mathcal{N}(0,1)$ is exactly $1/S'(v_j)$, because the Gaussian densities introduced by Φ and Φ^{-1} cancel. Block m uses a sine series,

$$S_{mj}(v) = v + \sum_{k=1}^{K} \omega_{mjk} \frac{\sin(\pi k v)}{\pi k}, \qquad b'_j = \Phi^{-1}\big(S_{mj}(\Phi(\widetilde{b}_j))\big), \tag{3}$$

whose derivative $S'_{mj}(v) = 1 + \sum_k \omega_{mjk} \cos(\pi k v)$ lies in $[1 - W_{mj}, 1 + W_{mj}]$ with $W_{mj} = \sum_k |\omega_{mjk}|$. The sine terms vanish at 0 and 1, so both tails keep their relative weight; if $W_{mj} < 1$, the map is invertible and the coordinate's ratio lies in $[(1 + W_{mj})^{-1}, (1 - W_{mj})^{-1}]$ for every $\widetilde{b}_j$. Odd frequencies move probability across the median, and even frequencies move it between the center and both tails.

4.2 ONE BUDGET SHARED BY ALL COORDINATES MAKES THE BOUND DIMENSION-FREE

The ratio of a block is the product $\prod_j S'_{mj}(v_j)^{-1}$ over its d_b coordinates, with $d_b = 32{,}768$ for SD3.5. Separate allowances $W_{mj} \le w$ give the bound $(1 - w)^{-d_b}$; keeping a block within $(1 - \alpha_m)^{-1}$ then

forces $w < 10^{-5}$ for every coordinate at our settings. RATIO instead normalizes all coefficients of a block jointly. From raw conditioner outputs $h_{mjk}(a, c)$,

$$\omega_{mjk} = \frac{\alpha_m \, h_{mjk}}{\varepsilon + \sum_{i=1}^{d_b} \sum_{k'=1}^{K} \sqrt{h_{mik'}^2 + \varepsilon^2}}, \qquad 0 < \alpha_m < 1, \;\; \varepsilon > 0. \tag{4}$$

Then $\sum_j W_{mj} < \alpha_m$, and the inequalities $\prod_j (1 - W_{mj}) \geq 1 - \sum_j W_{mj}$ and $\prod_j (1 + W_{mj}) \leq \exp(\sum_j W_{mj})$ place the block ratio in $[e^{-\alpha_m}, (1 - \alpha_m)^{-1}]$ for any d_b (Lemma 5). The dimension disappears because the allowances are summed before the product is taken, and the conditioner may give almost all of α_m to the few coordinates where the reward changes (Figure 9), a freedom that separate allowances remove.

4.3 Conditional rotations choose the directions at no cost

A CDF map acts on one coordinate, whereas a reward responds to combinations of coordinates; a change of layout spreads over many positions and channels. By Proposition 3, reweighting along a unit direction v changes the expected reward, to first order, by a covariance $\eta_{v,k}$ between the reward and a CDF feature of $v^\top Z$, so reweighting along latent axes can only use $\eta_{e_i,k}$. The standard Gaussian is invariant under orthogonal maps, so a rotation changes the directions that the CDF maps act on while its density ratio stays exactly one: directions are free, and only reweighting consumes the budget.

The rotation may depend on the noise and the prompt as long as the inverse can recover it. After a fixed permutation splits the input into (a, b), a conditioner reads a, pooled prompt features $e(c)$, and the block index m, and returns Givens angles, which define an orthogonal $R_m(a, c)$, and the raw coefficients $h_{mjk}(a, c)$. Given a, the vector $R_m(a, c)b$ is again standard Gaussian, so $(a, b) \mapsto (a, R_m(a, c)b)$ preserves γ; its Jacobian is block-triangular with determinant one, and the inverse recomputes R_m from the unchanged a (Lemma 4). Alternating the split across blocks transforms every coordinate and combines coordinates mixed by earlier blocks. At initialization all coefficients are zero, so T_ϕ preserves γ and $Q_c = P_c$.

4.4 Composing blocks and learning from rewards

The ratio of the full transform factorizes into block ratios, each evaluated against γ at its own input and output, although the intermediate latents are no longer Gaussian (Appendix A.2). Multiplying the block bounds gives

$$\ell := \exp\Big(-\sum_{m=1}^{L} \alpha_m\Big) \;\leq\; \Lambda_\phi(z \mid c) \;\leq\; \prod_{m=1}^{L} (1 - \alpha_m)^{-1} =: u \qquad \text{for all } z, c, \text{ and } \phi. \tag{5}$$

With $\alpha_m = 1 - e^{-\rho/L}$, $u = e^\rho$ and one scalar ρ controls the strength of alignment; $L = 8$ and $\rho = 2$ give $[\ell, u] = [0.1704, 7.389]$. The bound holds for every parameter value, hence at every training step and for every prompt, with no penalty weight to tune.

For a differentiable reward such as PickScore, we backpropagate $r(G(T_\phi(Z, c), c), c)$ through the frozen generator into ϕ. For a black-box reward such as GenEval, inverting the blocks in reverse order gives the exact density $\log q_\phi(y \mid c) = \log \gamma(y) - \sum_{m,j} \log S'_{mj}(v_{mj})$, and we use the score-function gradient $\mathbb{E}_{Y \sim q_\phi}[(r(G(Y, c), c) - b_c) \nabla_\phi \log q_\phi(Y \mid c)]$ with a leave-one-out baseline b_c over eight samples per prompt (Algorithm 1). Generation applies T_ϕ once to one Gaussian draw and runs the unchanged sampler.

4.5 What the bounds guarantee and what they cost

The results assume a measurable G and exact arithmetic; proofs are in Appendices A–C.

Theorem 1 (Event bounds). *Under (5), every prompt c and event A with $p = P_c(A)$ satisfy*

$$\underline{Q}(p) := \max\{\ell p, \; 1 - u(1 - p)\} \;\leq\; Q_c(A) \;\leq\; \min\{up, \; 1 - \ell(1 - p)\} =: \overline{Q}(p), \tag{6}$$

and every value in $[\underline{Q}(p), \overline{Q}(p)]$ is attained by some Q with $\ell \leq dQ/dP_c \leq u$.

The second term of each envelope comes from the complement of A; at $\rho = 2$, an event with $p = 0.1$ stays between 1.70% and 73.9%. The bounds apply to every event, including incorrect outputs such as a wrong object count, so they protect alternatives without knowing which are correct. The next result gives the reward this costs.

Proposition 2 (Reward optimum under the bounds)**.** *Fix c, let $r \in L^1(P_c)$, $\ell < 1 < u$, and $\tau = (1-\ell)/(u-\ell)$. Let P_c^τ be P_c restricted to its highest-reward mass τ, with fractional weight at ties, and renormalized. Among all Q with $\ell \le dQ/dP_c \le u$, the expected reward is maximized by*

$$Q_c^\star = \ell P_c + (1-\ell)P_c^\tau, \qquad \mathbb{E}_{Q_c^\star}[r] - \mathbb{E}_{P_c}[r] = (1-\ell)\big(\mathbb{E}_{P_c^\tau}[r] - \mathbb{E}_{P_c}[r]\big). \tag{7}$$

The optimum keeps a scaled copy of the base model and moves the remaining mass to the top τ fraction of images ($\tau = 11.5\%$ at $\rho = 2$). Figure 3 shows this in image space: the occupancy map $\sum_i w_i M_i(h,w)$ of five dog masks M_i shifts toward the sitting poses that the reward prefers, while every image keeps positive weight, as the lower bound requires. The optimum has three consequences, which Section 5 tests. It bounds the gain of any transform that satisfies (5). It caps a correctness reward at $\overline{Q}(p)$ for a prompt with base accuracy p. It changes any other score g by $(1-\ell)(\mathbb{E}_{P_c^\tau}[g] - \mathbb{E}_{P_c}[g])$, whose sign is set by how r ranks images: the reward decides which images gain probability, and the bound decides how much. For a correctness reward, the top-τ images are the correct ones, so the optimum raises correctness while every incorrect image keeps the fraction ℓ of its base probability; for a preference reward, the top images may also differ from the base in style or text fidelity, and the same formula predicts in which direction such secondary scores move.

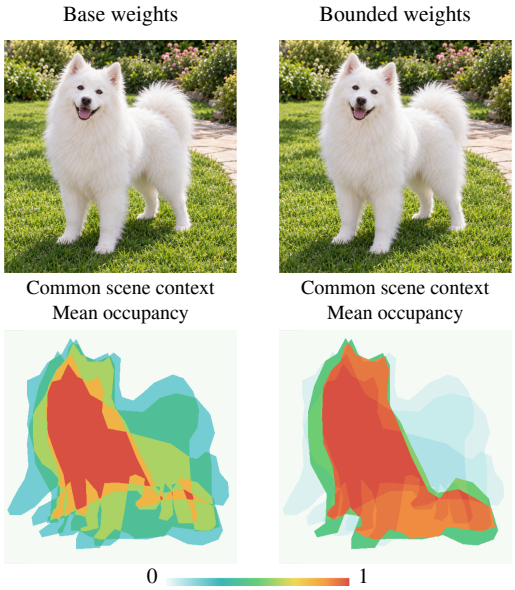


Figure 3: **Where the optimum places probability.** Foreground masks of five dog images, weighted uniformly as under P_c (left) and by $Q_c^\star$ at $\rho = 2$ for a sitting-pose reward (right), on a shared scale.

Proposition 3 (Reward-correlated directions improve reward)**.** *Let $f(z) = r(G(z,c),c) \in L^1(\gamma)$. For a unit vector v and frequency k, let T_t rotate v onto the first coordinate, apply (3) there with the single coefficient $t \in (-1,1)$ at frequency k, and rotate back. Then $\frac{d}{dt}\mathbb{E}_\gamma[f(T_t(Z))]\big|_{t=0} = -\eta_{v,k}$ with $\eta_{v,k} = \mathrm{Cov}_\gamma\big(f(Z), \cos(\pi k \Phi(v^\top Z))\big)$. If $\eta_{v,k} \neq 0$, every sufficiently small coefficient of the opposite sign increases the expected reward.*

Rotations reach directions with large $|\eta_{v,k}|$ at no cost to the bound (Appendix C).

5 Experiments

We test four predictions: the bounds retain more coverage than average controls at a common gain (Section 5.1) and for individual events (Section 5.2); rotations and odd frequencies determine the reward the budget buys (Section 5.3); and the reward, within the bound, decides where probability goes (Sections 5.4–5.5).

Setup. We align SD3.5-Medium (2.5B, 512^2), FLUX.1-dev (12B, 512^2), and SANA-1.5 (1.6B, 1024^2) with their native samplers, using PickScore (differentiable), a partial GenEval score (black-box), and Aesthetic v2 (misaligned with text fidelity) as rewards. RATIO uses $L = 8$, $K = 4$, and $\rho = 2$ unless stated, and trains for 1,000 steps with 16 prompts per step. The controls are LoRA fine-tuning with a differentiable reward (DRaFT, RSA-FT), online reinforcement learning (DiffusionNFT, Flow-GRPO, DiverseGRPO, GRPO-Guard), amortized noise optimization (HyperNoise), and *Affine+KL*, which replaces RATIO's CDF maps with a conditional affine noise map under a KL penalty. Coverage is PRDC recall in DINOv2 features, the share of 5,000 independent base images covered by a method's images (Naeem et al., 2020; Oquab et al., 2023). We evaluate on 500 Pick-a-Pic and 200

Table 1: **Reward, retention, and external correctness in one comparison.** Default columns use $s = 1$ (RATIO: $\rho = 2$); right-hand columns interpolate each prompt set to its own RATIO reward coordinate. All seven external methods use the evaluation protocol in Appendix E; their training exposure is not uniformly matched. Recall, CLIPScore, and GenEval are $\times 100$.

Method	Default operating points						Matched PickScore gain			
	Pick-a-Pic			DrawBench		GenEval	Pick-a-Pic		DrawBench	
	Pick ↑	Rec. ↑	CLIP ↑	Rec. ↑	CLIP ↑	All ↑	Rec. ↑	CLIP ↑	Rec. ↑	CLIP ↑
Stable Diffusion 3.5 Medium (2.5B) (Stability AI, 2024) / PickScore training; guided and CFG-free references										
Guided base	21.98	**94.8**	30.04	**94.4**	30.44	64.2	—	—	—	—
Unguided base	20.19	91.2	25.01	92.8	25.08	24.9	—	—	—	—
DRaFT (Clark et al., 2024)	23.39	34.7	27.89	41.5	28.50	61.6	69.2	29.16	78.3	29.88
RSA-FT (Kim & Sim, 2026)	23.14	59.9	29.00	67.8	29.48	63.7	75.4	29.49	85.5	30.11
DiffusionNFT (Zheng et al., 2025)	23.62	46.0	28.60	53.8	28.99	53.5	75.3	29.34	81.6	29.91
Flow-GRPO (Liu et al., 2025b)	23.49	49.4	29.02	56.8	29.74	63.5	75.6	29.56	86.3	30.20
DiverseGRPO (Liu et al., 2025a)	23.48	53.6	29.23	61.9	29.90	65.2	78.0	29.62	89.0	30.22
GRPO-Guard (Wang et al., 2025a)	**23.65**	52.5	29.34	60.6	29.94	**66.4**	80.6	29.62	88.0	30.21
HyperNoise (Eyring et al., 2025)	22.62	72.7	29.53	82.7	30.07	63.8	75.3	29.59	84.6	30.10
RATIO ($\rho = 2$)	22.58	85.7	**30.08**	90.1	**30.47**	63.6	**85.7**	**30.08**	**90.1**	**30.47**

Pick = PickScore; Rec. = DINOv2 PRDC recall; CLIP = CLIPScore; All = GenEval tag-macro correctness. The Pick-a-Pic gain is 0.60 (22.58 absolute). HyperNoise uses a per-metric path envelope only in matched columns; its default columns are one checkpoint. Dashes mark baseline-only, inapplicable matched-gain cells. Bold marks column maxima, including reference rows; ties at displayed precision are retained.

DrawBench (Saharia et al., 2022) prompts (ten images each) and 553 GenEval prompts (four each), averaging three training runs (Appendix E).

5.1 Retention at the same reward gain

We evaluate each control along its strength path and interpolate its metrics to RATIO's reward on each prompt set. At a PickScore gain of 0.596 on Pick-a-Pic, RATIO covers 85.7% of the base images and the seven controls cover 69.2–80.6% (Table 1); its CLIPScore of 30.08 also exceeds theirs ($\leq$29.62). On DrawBench, RATIO reaches 90.1% recall against 78.3–89.0%. At default strengths the controls gain more, up to 1.67 for GRPO-Guard, at 34.7–72.7% recall.

HyperNoise and Affine+KL isolate the role of the bound: both change only the initial noise of the frozen SD3.5-M, through a residual map and a mean-KL penalty, without a pointwise bound. At the same gain they cover 75.3% and 75.0%, about 10 points below RATIO. The pointwise bound, beyond working in noise space, keeps the alternatives.

The strength paths (Figure 12) explain the numbers. At small gains ($\rho \leq 1$) all methods stay close to the base, and RATIO exceeds the best control by 0.1–0.3 recall points (Table 8). The difference grows to 5.1 points at $\rho = 2$, where the controls reach the gain partly by removing images. At $\rho = 3$ the lower factor falls to $\ell = 0.082$, the guarantee weakens, and the difference shrinks to 0.7 points. The bound matters most when it is tight relative to what the reward would otherwise remove.

5.2 Which image events lose probability?

Theorem 1 predicts more than coverage: no event may leave its bounds. From GenEval metadata we define 5,051 events on SD3.5-M outputs (correctness, colors, counts, and object presence; Figure 10, Appendix E.3), of which 1,125 express what their prompts require (*prompt-valid*). With 256 base and 128 method images per prompt, one-sided Clopper–Pearson bounds reject a deletion when the method's upper bound falls below Q at the base's lower bound, and an amplification when its lower bound exceeds $\overline{Q}$ at the base's upper bound; each test errs with probability at most 2.5% (Appendix E.3).

Table 2: **Direct event tests.** Of 5,051 testable events, 1,125 are prompt-valid. RATIO uses its own ρ band; controls use $\rho = 2$, relative to the guided base. External methods and citations are given in Table 1.

Method	Gain	Reject ↓	Valid ↓	Up ↓
SD3.5-M (2.5B) / 553 GenEval prompts				
RATIO $\rho = 2$	59.6	**0**	**0**	**0**
RATIO $\rho = 3$	**67.1**	**0**	**0**	**0**
DRaFT (Clark et al., 2024)	61.9	343	197	120
RSA-FT (Kim & Sim, 2026)	59.6	179	71	100
DiffusionNFT (Zheng et al., 2025)	60.4	335	196	114
Flow-GRPO (Liu et al., 2025b)	60.8	265	146	105
DiverseGRPO (Liu et al., 2025a)	63.7	226	104	106
GRPO-Guard (Wang et al., 2025a)	66.7	180	68	102
HyperNoise (Eyring et al., 2025)	64.2	350	213	118
Affine + KL	64.0	386	232	124

Gain = 100ΔPickScore; Valid = prompt-valid deletions; Up = amplified events. Counts are first-order event outcomes, not confidence intervals.

RATIO has no rejection at $\rho = 2$ or $\rho = 3$ (Table 2). The external controls have 179–350 rejections against the $\rho = 2$ bounds; DRaFT alone deletes 197 prompt-valid events. The decisive comparison is Affine+KL: it shares RATIO's

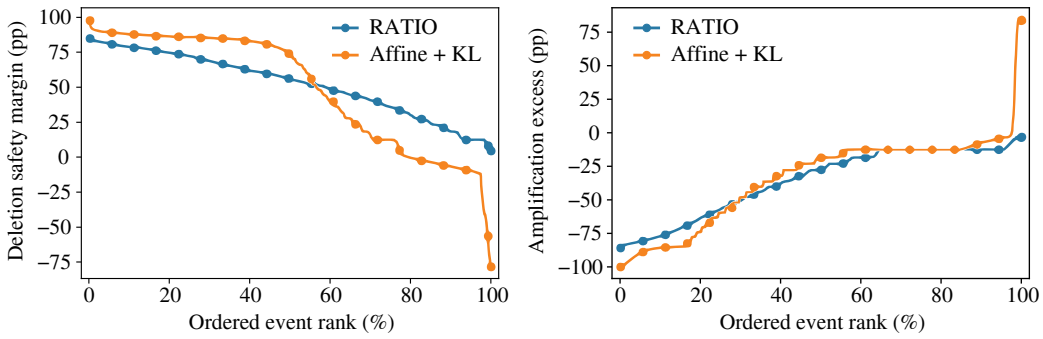


Figure 4: **Similar mean noise KL, different event tails.** SD3.5-M, $\rho = 2$; 256 base and 128 method images per prompt. Left: deletion margin $100[q_U - \underline{Q}(p_L)]$ for the 1,125 prompt-valid events. Right: amplification excess $100[q_L - \overline{Q}(p_U)]$ for all 5,051 events. Each method's events are sorted separately; negative left and positive right values reject the bounds under one-sided 98.75% Clopper–Pearson bounds. At $1.017\times$ RATIO's mean noise KL, Affine+KL has 232 prompt-valid deletions and 124 amplifications; RATIO has none.

conditioner, noise interface, and data, reaches almost the same mean noise KL (0.2945 against 0.2896 nats), and yet deletes 232 prompt-valid events and amplifies 124. Figure 4 shows how: most events change similarly under both transforms, but the affine map concentrates its change on about one fifth of the events, whose margins cross zero. An average constraint lets change accumulate on a few events (Section 3); the pointwise bound excludes this. With 128 draws, deletion is detectable for 1,085 events (Table 9); Theorem 1 covers all.

5.3 What makes the budget productive?

The bound fixes how much probability may move; the architecture decides how much reward it buys. Removing the rotations keeps the bound but lowers the PickScore gain from 0.596 to 0.339 (Table 3). The CDF maps can then use only the reward covariances along the latent axes (Proposition 3), so the same bounds buy less reward, and the smaller change of the image distribution raises recall to 89.9%. Restricting the frequencies to $\{2, 4\}$ halves the coefficients and lowers the gain to 0.51; even frequencies move probability only symmetrically about the median, and their fixed-coefficient features are uncorrelated with the odd part of a reward (Appendix C.4).

Proposition 2 measures how much of the permitted reward the transform realizes. Estimating the optimum from 64 base images per prompt, the learned PickScore gain on SD3.5-M is 67.6% of the optimum at $\rho = 2$ and 59.2% at $\rho = 3$, and 73.8% on SANA-1.5 (Table 13). The optimum grows faster with ρ than the learned gain, so the gap belongs to the finite transform ($L = 8$, $K = 4$).

Table 3: **Design ablations.** SD3.5-M (Stability AI, 2024), Pick-a-Pic. The affine control reaches $1.02\times$ RATIO noise KL; its final row is a matched-gain interpolation. Recall and CLIP are $\times 100$.

Variant	Pick ↑	Rec. ↑	CLIP ↑
SD3.5-M (2.5B) / structural interventions			
Full RATIO	22.58	85.7	30.08
No mixing	22.32	**89.9**	30.06
Even frequencies	22.49	86.2	**30.09**
Affine + KL	**22.62**	73.1	29.50
Affine, same gain	22.58	75.0	29.55

5.4 Transfer across generators and to a black-box reward

With 16,000 training images per model, RATIO raises PickScore by 0.60, 0.35, and 0.54 on SD3.5-M, FLUX.1-dev, and SANA-1.5 at 83.9–87.0% recall (Table 4). DRaFT obtains larger gains (1.41, 0.97, 0.79) at 34.7–49.3% recall and lowers CLIPScore by 1.1–2.2 points, which RATIO keeps within 0.04. GenEval correctness moves by at most 1.2 points under RATIO and falls by 1.4–2.7 under DRaFT, and RATIO stays within 0.4 points of the base on T2I-CompBench++ and DPG-Bench (Tables 11–12).

Through the exact density, RATIO learns from the black-box partial GenEval score: official correctness rises from 64.2% to 76.8%, position accuracy from 23.3% to 52.4%, and recall stays at 85.4% (Table 10). Flow-GRPO and GRPO-Guard reach about 95% correctness at 51.9–54.8% recall. By Theorem 1, a prompt correct with base probability p cannot exceed $\overline{Q}(p)$, which is 86.9% for

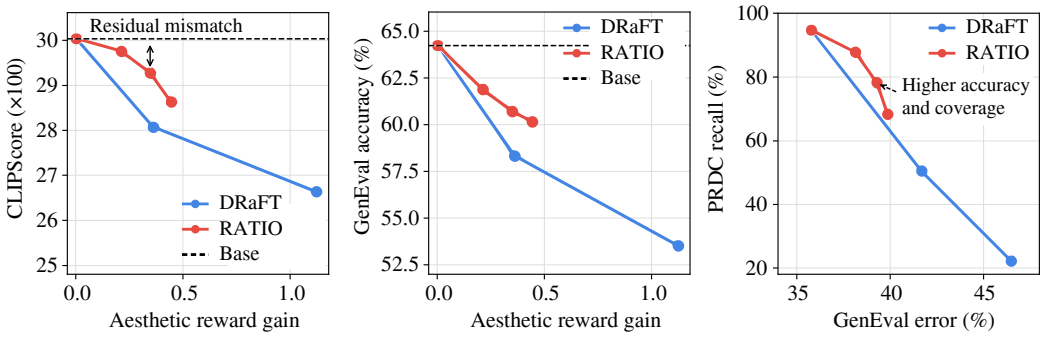


Figure 5: **Aesthetic gains lower text fidelity for both methods, and RATIO keeps more coverage.** SD3.5-M with the guided base, RATIO at $\rho = 1, 2, 3$, and DRaFT at $s = 0.5, 1$. Left and center: CLIPScore and GenEval correctness against Aesthetic v2 gain; dashed lines mark the base. Right: PRDC recall against GenEval error, $100-$correctness. Points are means over three training runs. RATIO at $\rho = 2$ gains 0.347, close to DRaFT at $s = 0.5$ with 0.359.

Table 4: **Cross-generator transfer under the same training exposure.** PickScore optimization, $\rho = 2$ for RATIO and $s = 1$ for DRaFT. GenEval uses its six official tags; HPS v2.1 uses the official four-style prompt set (Ghosh et al., 2023; Wu et al., 2023). All metrics are higher-is-better; apart from PickScore, entries are $\times 100$.

Method	Pick-a-Pic			DrawBench			GenEval							HPS
	Pick	Rec.	CLIP	Pick	Rec.	CLIP	All	Single	Two	Count	Color	Pos.	Attr.	v2.1
Stable Diffusion 3.5 Medium (2.5B) (Stability AI, 2024) / 16,000 training images per learned model														
Base	21.98	**94.8**	30.04	22.34	**94.4**	30.44	**64.2**	**98.4**	77.0	**53.4**	79.8	23.3	**53.4**	28.61
DRaFT (Clark et al., 2024)	**23.39**	34.7	27.89	**23.49**	41.5	28.50	61.6	98.3	71.9	48.9	**80.2**	17.2	52.9	**31.28**
RATIO	22.58	85.7	**30.08**	22.73	90.1	**30.47**	63.6	97.6	**77.5**	50.5	79.2	**25.8**	51.3	29.07
FLUX.1-dev (12B) (Black Forest Labs, 2024) / 16,000 training images per learned model														
Base	22.64	**94.2**	**31.73**	22.83	**94.0**	32.01	65.4	**98.7**	77.4	52.8	82.3	**34.5**	46.7	30.21
DRaFT (Clark et al., 2024)	**23.61**	49.3	30.44	**23.65**	55.2	30.95	64.0	98.0	79.5	49.5	80.2	31.4	45.2	**32.62**
RATIO	22.99	87.0	31.72	23.06	91.1	**32.06**	**66.6**	97.8	**81.7**	**53.3**	**82.4**	**34.5**	**50.1**	30.52
SANA-1.5 (1.6B, 1024px) (Xie et al., 2025) / 16,000 training images per learned model														
Base	21.59	**94.6**	29.09	21.91	**94.1**	**29.49**	**66.0**	**98.7**	**82.6**	**56.8**	**84.0**	**25.5**	**48.3**	27.26
DRaFT (Clark et al., 2024)	**22.38**	46.3	27.96	**22.49**	53.5	28.42	63.3	98.1	77.1	52.3	82.6	22.1	47.4	**29.12**
RATIO	22.13	83.9	**29.13**	22.24	89.9	29.47	65.6	98.5	80.8	**56.8**	**84.0**	25.3	48.2	27.80

GenEval: All is the unweighted tag mean; Single/Two = object categories; Pos. = position; Attr. = color attribution. HPS is evaluated on its own benchmark, not on Pick-a-Pic. Bold is computed independently within each backbone block.

$p = 0.233$ at $\rho = 2$. The lower factor keeps 17% of every incorrect composition along with every valid alternative, a price fixed before training.

5.5 The reward decides where probability goes

Aesthetic v2, whose favorite images are not always faithful to the prompt, separates the two roles in Proposition 2. RATIO raises the Aesthetic score from 5.39 to 5.73, close to DRaFT at $s = 0.5$ (5.75), and keeps 78.3% recall against 50.5% (Table 5). Judges call DRaFT's image the most overprocessed in 42.0% of comparisons and RATIO's in 17.5%. Both lower CLIPScore and GenEval correctness along their strength paths (Figure 5), the sign Proposition 2 predicts when the top Aesthetic images are less faithful to the text. Learned external changes have this predicted sign in 70 of 79 settings (Appendix F.5); judges find RATIO's image groups more diverse than DRaFT's in 69.3% of cases (Appendix F.6).

Table 5: **Reward mismatch.** RATIO $\rho = 2$ and DRaFT $s = 0.5$ (Clark et al., 2024). Recall, CLIP, and GenEval are $\times 100$. Over. is the share judged most overprocessed; 32.4% selected none.

Method	Aest. ↑	Rec. ↑	CLIP ↑	Gen. ↑	Over. ↓
SD3.5-M (2.5B) / Aesthetic v2					
Base	5.39	**94.8**	**30.04**	**64.2**	**8.1**
RATIO	5.73	78.3	29.27	60.7	17.5
DRaFT (Clark et al., 2024)	**5.75**	50.5	28.08	58.3	42.0

6 Conclusion

For a frozen text-to-image generator, keeping every image set within fixed probability factors reduces to a pointwise bound on the initial-noise density ratio, which RATIO enforces exactly; the bound sets how much probability may move, and the reward decides which images receive it.

REFERENCES

Heli Ben-Hamu, Omri Puny, Itai Gat, Brian Karrer, Uriel Singer, and Yaron Lipman. D-Flow: Differentiating through flows for controlled generation. In *International Conference on Machine Learning*, 2024. URL https://proceedings.mlr.press/v235/ben-hamu24a.html.

Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=YCWjhGrJFD.

Black Forest Labs. FLUX.1 [dev], 2024. URL https://huggingface.co/black-forest-labs/FLUX.1-dev.

Kevin Clark, Paul Vicol, Kevin Swersky, and David J. Fleet. Directly fine-tuning diffusion models on differentiable rewards. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=1vmSEVL19f.

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using Real NVP. In *International Conference on Learning Representations*, 2017. URL https://arxiv.org/abs/1605.08803.

discus0434. Aesthetic predictor v2.5, n.d. URL https://github.com/discus0434/aesthetic-predictor-v2-5. Software repository, accessed September 2026.

Carles Domingo-Enrich, Michal Drozdzal, Brian Karrer, and Ricky T. Q. Chen. Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=xQBRrtQM8u.

Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. In *Advances in Neural Information Processing Systems*, 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/hash/7ac71d433f282034e088473244df8c02-Abstract.html.

Luca Eyring, Shyamgopal Karthik, Karsten Roth, Alexey Dosovitskiy, and Zeynep Akata. ReNO: Enhancing one-step text-to-image models through reward-based noise optimization. In *Advances in Neural Information Processing Systems*, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/e31bdea0a93741c2157eea705dd219eb-Paper-Conference.pdf.

Luca Eyring, Shyamgopal Karthik, Alexey Dosovitskiy, Nataniel Ruiz, and Zeynep Akata. Noise hypernetworks: Amortizing test-time compute in diffusion models. *arXiv preprint arXiv:2508.09968*, 2025. URL https://arxiv.org/abs/2508.09968.

Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. DPOK: Reinforcement learning for fine-tuning text-to-image diffusion models. In *Advances in Neural Information Processing Systems*, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/fc65fab891d83433bd3c8d966edde311-Abstract-Conference.html.

Dhruba Ghosh, Hannaneh Hajishirzi, and Ludwig Schmidt. GenEval: An object-focused framework for evaluating text-to-image alignment. In *Advances in Neural Information Processing Systems*, 2023. URL https://arxiv.org/abs/2310.11513.

Xiefan Guo, Jinlin Liu, Miaomiao Cui, Jiankai Li, Hongyu Yang, and Di Huang. InitNO: Boosting text-to-image diffusion models via initial noise optimization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Guo_InitNO_Boosting_Text-to-Image_Diffusion_Models_via_Initial_Noise_Optimization_CVPR_2024_paper.html.

Peter Holderrieth, Uriel Singer, Tommi Jaakkola, Ricky T. Q. Chen, Yaron Lipman, and Brian Karrer. Glass flows: Efficient inference for reward alignment of flow and diffusion models. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=vH7OAPZ2dR.

Xiwei Hu, Rui Wang, Yixiao Fang, Bin Fu, Pei Cheng, and Gang Yu. ELLA: Equip diffusion models with LLM for enhanced semantic alignment. *arXiv preprint arXiv:2403.05135*, 2024. URL https://arxiv.org/abs/2403.05135.

Kaiyi Huang, Chengqi Duan, Kaiyue Sun, Enze Xie, Zhenguo Li, and Xihui Liu. T2I-CompBench++: An enhanced and comprehensive benchmark for compositional text-to-image generation. *arXiv preprint arXiv:2307.06350*, 2025. URL https://arxiv.org/abs/2307.06350v3.

Gabriel Ilharco, Mitchell Wortsman, Ross Wightman, Cade Gordon, Nicholas Carlini, Rohan Taori, Achal Dave, Vaishaal Shankar, Hongseok Namkoong, John Miller, Hannaneh Hajishirzi, Ali Farhadi, and Ludwig Schmidt. OpenCLIP. Open-source software, 2021. URL https://github.com/mlfoundations/open_clip.

Kwanyoung Kim and Byeongsu Sim. Reward sharpness-aware fine-tuning for diffusion models. *arXiv preprint arXiv:2603.21175*, 2026. URL https://arxiv.org/abs/2603.21175.

Diederik P. Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In *Advances in Neural Information Processing Systems*, 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/hash/d139db6a236200b21cc7f752979132d0-Abstract.html.

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. In *Advances in Neural Information Processing Systems*, 2023. URL https://arxiv.org/abs/2305.01569.

Kimin Lee, Hao Liu, Moonkyung Ryu, Olivia Watkins, Yuqing Du, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, and Shixiang Shane Gu. Aligning text-to-image models using human feedback. *arXiv preprint arXiv:2302.12192*, 2023. URL https://arxiv.org/abs/2302.12192.

Henglin Liu, Huijuan Huang, Jing Wang, Chang Liu, Xiu Li, and Xiangyang Ji. DiverseGRPO: Mitigating mode collapse in image generation via diversity-aware GRPO. *arXiv preprint arXiv:2512.21514*, 2025a. URL https://arxiv.org/abs/2512.21514.

Jie Liu, Gongye Liu, Jiajun Liang, Yangguang Li, Jiaheng Liu, Xintao Wang, Pengfei Wan, Di Zhang, and Wanli Ouyang. Flow-GRPO: Training flow matching models via online RL. *arXiv preprint arXiv:2505.05470*, 2025b. URL https://arxiv.org/abs/2505.05470.

Nanye Ma, Shangyuan Tong, Haolin Jia, Hexiang Hu, Yu-Chuan Su, Mingda Zhang, Xuan Yang, Yandong Li, Tommi Jaakkola, Xuhui Jia, and Saining Xie. Scaling inference time compute for diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025a. URL https://openaccess.thecvf.com/content/CVPR2025/html/Ma_Scaling_Inference_Time_Compute_for_Diffusion_Models_CVPR_2025_paper.html.

Yuhang Ma, Yunhao Shui, Xiaoshi Wu, Keqiang Sun, and Hongsheng Li. HPSv3: Towards wide-spectrum human preference score. *arXiv preprint arXiv:2508.03789*, 2025b. URL https://arxiv.org/abs/2508.03789.

Muhammad Ferjad Naeem, Seong Joon Oh, Youngjung Uh, Yunjey Choi, and Jaejun Yoo. Reliable fidelity and diversity metrics for generative models. In *International Conference on Machine Learning*, 2020. URL https://proceedings.mlr.press/v119/naeem20a.html.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jégou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. URL https://arxiv.org/abs/2304.07193.

Mihir Prabhudesai, Anirudh Goyal, Deepak Pathak, and Katerina Fragkiadaki. Aligning text-to-image diffusion models with reward backpropagation. *arXiv preprint arXiv:2310.03739*, 2023. URL https://arxiv.org/abs/2310.03739.

Zipeng Qi, Lichen Bai, Haoyi Xiong, and Zeke Xie. Not all noises are created equally: Diffusion noise selection and optimization. *arXiv preprint arXiv:2407.14041*, 2024. URL https://arxiv.org/abs/2407.14041.

Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S. Sara Mahdavi, Rapha Gontijo Lopes, Tim Salimans, Jonathan Ho, David J Fleet, and Mohammad Norouzi. Photorealistic text-to-image diffusion models with deep language understanding. *arXiv preprint arXiv:2205.11487*, 2022. URL https://arxiv.org/abs/2205.11487.

Christoph Schuhmann. Improved aesthetic predictor, n.d. URL https://github.com/christophschuhmann/improved-aesthetic-predictor. Software repository, ac-

cessed September 2026.

Stability AI. Stable Diffusion 3.5 Medium, 2024. URL https://huggingface.co/stabilityai/stable-diffusion-3.5-medium.

Masatoshi Uehara, Yulai Zhao, Kevin Black, Ehsan Hajiramezanali, Gabriele Scalia, Nathaniel Lee Diamant, Alex M. Tseng, Tommaso Biancalani, and Sergey Levine. Fine-tuning of continuous-time diffusion models as entropy-regularized control. *arXiv preprint arXiv:2402.15194*, 2024. URL https://arxiv.org/abs/2402.15194.

Bram Wallace, Akash Gokul, Stefano Ermon, and Nikhil Naik. End-to-end diffusion latent optimization improves classifier guidance. In *IEEE/CVF International Conference on Computer Vision*, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Wallace_End-to-End_Diffusion_Latent_Optimization_Improves_Classifier_Guidance_ICCV_2023_paper.html.

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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Wallace_Diffusion_Model_Alignment_Using_Direct_Preference_Optimization_CVPR_2024_paper.html.

Jing Wang, Jiajun Liang, Jie Liu, Henglin Liu, Gongye Liu, Jun Zheng, Wanyuan Pang, Ao Ma, Zhenyu Xie, Xintao Wang, Meng Wang, Pengfei Wan, and Xiaodan Liang. GRPO-Guard: Mitigating implicit over-optimization in flow matching via regulated clipping. *arXiv preprint arXiv:2510.22319*, 2025a. URL https://arxiv.org/abs/2510.22319.

Yibin Wang, Yuhang Zang, Hao Li, Cheng Jin, and Jiaqi Wang. Unified reward model for multimodal understanding and generation. *arXiv preprint arXiv:2503.05236*, 2025b. URL https://arxiv.org/abs/2503.05236.

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. URL https://arxiv.org/abs/2306.09341.

Enze Xie, Junsong Chen, Yuyang Zhao, Jincheng Yu, Ligeng Zhu, Yujun Lin, Zhekai Zhang, Muyang Li, Junyu Chen, Han Cai, Bingchen Liu, Daquan Zhou, and Song Han. SANA 1.5: Efficient scaling of training-time and inference-time compute in linear diffusion transformer. In *Proceedings of the 42nd International Conference on Machine Learning*, 2025. URL https://proceedings.mlr.press/v267/xie25b.html.

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. In *Advances in Neural Information Processing Systems*, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/33646ef0ed554145eab65f6250fab0c9-Abstract-Conference.html.

Zeyue Xue, Jie Wu, Yu Gao, Fangyuan Kong, Lingting Zhu, Mengzhao Chen, Zhiheng Liu, Wei Liu, Qiushan Guo, Weilin Huang, and Ping Luo. DanceGRPO: Unleashing GRPO on visual generation. *arXiv preprint arXiv:2505.07818*, 2025. URL https://arxiv.org/abs/2505.07818.

Zhiyuan You, Xin Cai, Jinjin Gu, Tianfan Xue, and Chao Dong. Teaching large language models to regress accurate image quality scores using score distribution. *arXiv preprint arXiv:2501.11561*, 2025. URL https://arxiv.org/abs/2501.11561.

Ziyi Zhang, Sen Zhang, Yibing Zhan, Yong Luo, Yonggang Wen, and Dacheng Tao. Confronting reward overoptimization for diffusion models: A perspective of inductive and primacy biases. In *International Conference on Machine Learning*, 2024. URL https://proceedings.mlr.press/v235/zhang24ch.html.

Kaiwen Zheng, Huayu Chen, Haotian Ye, Haoxiang Wang, Qinsheng Zhang, Kai Jiang, Hang Su, Stefano Ermon, Jun Zhu, and Ming-Yu Liu. DiffusionNFT: Online diffusion reinforcement with forward process. *arXiv preprint arXiv:2509.16117*, 2025. URL https://arxiv.org/abs/2509.16117.

Zikai Zhou, Shitong Shao, Lichen Bai, Shufei Zhang, Zhiqiang Xu, Bo Han, and Zeke Xie. Golden noise for diffusion models: A learning framework. In *IEEE/CVF International Conference on Computer Vision*, 2025. URL https://openaccess.thecvf.com/

content/ICCV2025/papers/Zhou_Golden_Noise_for_Diffusion_Models_A_Learning_Framework_ICCV_2025_paper.pdf.

AI USE STATEMENT

Generative AI assisted with literature retrieval, research ideation and mathematical checks, drafting paper sections, and language polishing. References were checked against primary sources. The authors take responsibility for all content.

A PROOFS OF THE PROBABILITY BOUNDS

This appendix proves the noise-space bound (5) and Theorem 1. Appendix A.1 analyzes one block, Appendix A.2 composes the blocks and transfers the bound to images, Appendix A.3 collects equivalent forms and extensions, and Appendix A.4 shows why invertibility or small divergences cannot replace a pointwise bound. Throughout, a prompt c is fixed, densities are with respect to Lebesgue measure, γ denotes both the law $\mathcal{N}(0, I_D)$ and its density, Φ is the standard normal CDF and Φ' its density, and all statements hold in exact arithmetic.

A.1 ONE BLOCK

A block F_m applies a fixed permutation, the conditional rotation, the CDF map (3), and the inverse permutation. Permutations preserve γ and have Jacobian determinant ± 1, so it suffices to study the rotation and the CDF map at fixed a.

Lemma 4 (Conditional rotations preserve the Gaussian)**.** *Let a and b be independent standard Gaussian vectors and $R(a, c)$ a measurable orthogonal matrix. Then $(a, R(a, c)b)$ has the same law as (a, b), and for a differentiable conditioner the Jacobian determinant of $(a, b) \mapsto (a, R(a, c)b)$ has absolute value one.*

Proof. Given a, the matrix is fixed, and orthogonal invariance gives $R(a, c)b \mid a \sim \mathcal{N}(0, I_{d_b})$. This conditional law does not depend on a, which proves the claim about the joint law. The Jacobian is $\left(\begin{smallmatrix} I & 0 \\ \partial_a(R(a,c)b) & R(a,c) \end{smallmatrix}\right)$, whose determinant is $\det R(a, c) = \pm 1$. The lower-left block explains why the full Jacobian need not be orthogonal. The inverse uses the unchanged a to recompute R and applies $R^\top$. □

Lemma 5 (Shared budget)**.** *For the coefficients in (4), $\sum_{j,k} |\omega_{mjk}| < \alpha_m$. Every S_{mj} is an increasing bijection of $[0, 1]$, and $1 - \alpha_m \le \prod_j S'_{mj}(v_j) \le e^{\alpha_m}$ for every (a, c, v).*

Proof. The denominator in (4) strictly exceeds $\sum_{j,k} |h_{mjk}|$, so the sum bound holds, including when every coefficient is zero. Differentiating gives $S'_{mj}(v) = 1 + \sum_k \omega_{mjk} \cos(\pi k v) \in [1 - W_{mj}, 1 + W_{mj}]$. The derivative is positive because $W_{mj} < 1$, and the sine terms vanish at 0 and 1, which proves bijectivity. For nonnegative numbers with sum below one, repeated use of $(1-x)(1-y) \ge 1-x-y$ gives $\prod_j (1 - W_{mj}) \ge 1 - \sum_j W_{mj}$, and $\log(1 + w) \le w$ gives $\prod_j (1 + W_{mj}) \le \exp(\sum_j W_{mj})$. The number of coordinates disappears because the allowances are summed before the product is taken. □

Lemma 6 (Ratio of one block)**.** *Let $x = F_m^{-1}(y)$. The Gaussian-reference ratio of F_m at y is*

$$\frac{d(F_{m\#}\gamma)}{d\gamma}(y) = \frac{\gamma(x)}{\gamma(y)\,|\det DF_m(x)|} = \prod_j S'_{mj}(v_j)^{-1} \in \left[e^{-\alpha_m}, (1 - \alpha_m)^{-1}\right],$$

where the v_j are the CDF coordinates of the inverse preimage.

Proof. By Lemma 4, the permutation and the rotation contribute a factor one. For the CDF map at fixed a, differentiating $b'_j = \Phi^{-1}(S_{mj}(\Phi(\widetilde{b}_j)))$ gives $\partial b'_j / \partial \widetilde{b}_j = \Phi'(\widetilde{b}_j) S'_{mj}(v_j) / \Phi'(b'_j)$. The derivative is diagonal in the transformed coordinates because the coefficients depend only on a;

derivatives with respect to a sit in a lower-left block and do not change the determinant. The kept coordinates have equal Gaussian factors at input and output, and each transformed factor $\Phi'(\widetilde{b}_j)/\Phi'(b'_j)$ cancels the same factor of $\gamma(x)/\gamma(y)$. Only $\prod_j S'_{mj}(v_j)^{-1}$ remains, and Lemma 5 bounds it. □

A.2 Composition and event bounds

Write $x_0 = z$, $x_m = F_m(x_{m-1})$, and $y = x_L$. The change-of-variables formula gives

$$\Lambda_\phi(y \mid c) = \frac{\gamma(x_0)}{\gamma(x_L)\prod_m |\det DF_m(x_{m-1})|} = \prod_{m=1}^{L} \frac{\gamma(x_{m-1})}{\gamma(x_m)\,|\det DF_m(x_{m-1})|}. \tag{8}$$

Each factor is the Gaussian-reference ratio of one block at its own input and output, and Lemma 6 bounds it pointwise, although the actual law of x_{m-1} is no longer Gaussian. Multiplying the bounds proves (5).

For a fixed measurable generator, the image likelihood ratio exists and has the representation

$$\frac{dQ_c}{dP_c}(x) = \mathbb{E}_\gamma\big[\Lambda_\phi(Z \mid c) \mid G(Z,c) = x\big] \qquad P_c\text{-almost surely.} \tag{9}$$

Integrating the right-hand side over an event A gives the integral of Λ_ϕ over the preimage, which is $Q_c(A)$ by (2). Conditional expectation preserves upper and lower bounds, so $\ell \le dQ_c/dP_c \le u$. With $p = P_c(A)$ and $q = Q_c(A)$, this gives $\ell p \le q \le up$, and applying it to A^{c} gives $\ell(1-p) \le 1-q \le u(1-p)$. Intersecting the two intervals proves (6).

For sharpness, fix $0 < p < 1$ and any $q \in [\underline{Q}(p), \overline{Q}(p)]$. The density

$$\frac{dQ}{dP_c} = \frac{q}{p}\,\mathbf{1}_A + \frac{1-q}{1-p}\,\mathbf{1}_{A^{\mathsf{c}}} \tag{10}$$

integrates to one, and both constants lie in $[\ell, u]$ exactly when the two event inequalities hold, so it attains $Q(A) = q$, including the endpoints. When $p \in \{0,1\}$, absolute continuity fixes $q = p$.

A.3 Equivalent forms and extensions

For arbitrary probability measures, $\ell P(A) \le Q(A) \le uP(A)$ for all measurable A is equivalent to $\ell \le dQ/dP \le u$ almost surely. The upper inequality gives $Q \ll P$; if the derivative exceeded u on a set of positive measure, integrating over that set would contradict the upper event inequality, and the lower bound follows in the same way. The converse is integration. Two consequences are used in the discussion of divergences: for $0 < p < 1$, the event odds satisfy $\ell/u \le [q/(1-q)]/[p/(1-p)] \le u/\ell$, and integrating the pointwise logarithmic bounds gives $D_{\mathrm{KL}}(Q\|P) \le \log u$ and $D_{\mathrm{KL}}(P\|Q) \le -\log \ell$. The converse implications fail, as Appendix A.4 shows.

Three extensions cover the settings of our experiments. First, mixtures whose weights depend on c but not on the sampled output keep the bound: if every $Q_{c,i}$ satisfies it, so does $\sum_i w_i(c) Q_{c,i}$. This covers the equal-weight mixture of three training runs used in the event tests. Second, a sampler that injects further noise from a fixed distribution defines a Markov kernel $K_c(z, A) \in [0,1]$ from the initial noise to images; replacing the indicator of the preimage in (2) by $K_c(z, A)$ gives the same bounds. Third, the lower bound holds for incorrect outputs as well: a correctness event with base probability p has the sharp ceiling $\overline{Q}(p)$, and when $p = 0$ absolute continuity keeps it at zero.

An event may constrain a whole scene or a small region of it. Figure 6 shows both scales for two prompts; Theorem 1 applies to each, because it only requires the event to be measurable.

A.4 Why invertibility and average divergences do not bound events

For the shift $T(z) = z + a$, change of variables gives $q(z) = \gamma(z - a)$ and $\log(q(z)/\gamma(z)) = a^\top z - \|a\|^2/2$, which tends to $+\infty$ and $-\infty$ along the line $z = ta$. An isometry with an exact inverse can therefore violate both bounds. More generally, for $Q = \mathcal{N}(m, \Sigma)$ with positive definite Σ, the log ratio to γ is a polynomial of degree at most two in z, and such a polynomial is bounded above and

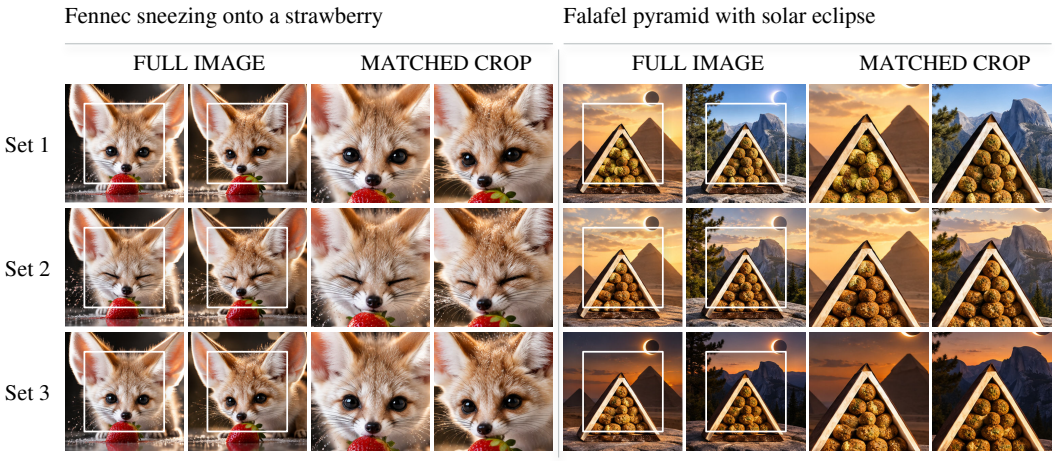


Figure 6: **Events at two scales.** Two prompts, a baby fennec sneezing onto a strawberry and a falafel pyramid beneath a solar eclipse. In each row, two full images precede crops of their white-box regions. An event can describe the full composition, such as the placement of the eclipse, or a local appearance, such as the texture inside the box.

below on $\mathbb{R}^D$ only if its quadratic and linear coefficients vanish, that is, $\Sigma = I$ and $m = 0$. Mean and covariance changes alone cannot satisfy a two-sided bound, which is why RATIO moves probability with a nonlinear map.

Small divergences do not bound events either. Let P be nonatomic, choose disjoint sets B_+ and B_- of probability δ, and for $M > 1$ define

$$\Lambda = M\mathbf{1}_{B_+} + M^{-1}\mathbf{1}_{B_-} + c_\delta \mathbf{1}_{(B_+ \cup B_-)^c}, \qquad c_\delta = \frac{1 - \delta(M + M^{-1})}{1 - 2\delta}, \qquad 0 < \delta < \frac{1}{M + M^{-1}}.$$

Then $\Lambda > 0$ and $\mathbb{E}_P \Lambda = 1$, so $dQ = \Lambda\, dP$ is a probability law under which B_+ grows by M and B_- shrinks by $1/M$. Direct calculation gives

$$D_{\mathrm{KL}}(Q\|P) = \delta(M - M^{-1}) \log M + (1 - 2\delta) c_\delta \log c_\delta, \qquad D_{\mathrm{KL}}(P\|Q) = -(1 - 2\delta) \log c_\delta,$$

and both tend to zero as $\delta \downarrow 0$ for fixed M. For a Rényi order $\beta > 0$, $\beta \neq 1$, the divergence $D_\beta(Q\|P) = (\beta - 1)^{-1} \log \mathbb{E}_P \Lambda^\beta$ has moment $\mathbb{E}_P \Lambda^\beta = \delta M^\beta + \delta M^{-\beta} + (1 - 2\delta) c_\delta^\beta \to 1$, and the reverse divergence uses $\mathbb{E}_P \Lambda^{1-\beta} \to 1$. Choosing M first and then a small enough δ makes any finite collection of these divergences arbitrarily small while the event factors stay M and $1/M$. The introduction's example is the one-sided case: an event of probability 0.01 that becomes ten times rarer, with the rest rescaled, has $D_{\mathrm{KL}}(Q\|P) = 0.0067$ nats.

B The reward attainable under the bounds

Proposition 2 describes the best distribution that any transform satisfying (5) can reach. This appendix proves it (Appendix B.1), explains how we estimate the optimum from samples and what it looks like in image space (Appendix B.2), and bounds how far a proxy reward can mislead an external score under any feasible distribution (Appendix B.3). We fix c and write $P = P_c$ and $P^\tau = P_c^\tau$.

B.1 Proof of Proposition 2

For $\ell < 1 < u$, write a feasible density as $\Lambda = \ell + (u - \ell)w$. The constraints become $0 \le w \le 1$ and $\mathbb{E}_P w = \tau$ with $\tau = (1 - \ell)/(u - \ell) \in (0, 1)$, so maximizing $\mathbb{E}_Q r$ amounts to choosing a weighting of mass τ that maximizes $\mathbb{E}_P[rw]$. Choose a threshold t with $P(r > t) \le \tau \le P(r \ge t)$. If $P(r = t) > 0$, put

$$a_t = \frac{\tau - P(r > t)}{P(r = t)}, \qquad w^\star = \mathbf{1}_{\{r > t\}} + a_t \mathbf{1}_{\{r = t\}},$$

and otherwise $w^\star = \mathbf{1}_{\{r>t\}}$, which already has mass τ. For every feasible w, the product $(r-t)(w-w^\star)$ is nonpositive: above t we have $w^\star = 1$, below t we have $w^\star = 0$, and on the threshold the product vanishes. Since $\mathbb{E}_P(w - w^\star) = 0$,

$$\mathbb{E}_P[r(w - w^\star)] = \mathbb{E}_P[(r - t)(w - w^\star)] + t\,\mathbb{E}_P(w - w^\star) \le 0,$$

where integrability of r justifies the expectations. Hence $w^\star$ is optimal. Defining $dP^\tau/dP = w^\star/\tau$ gives $\Lambda^\star = \ell + (1-\ell)w^\star/\tau$, which is $Q^\star = \ell P + (1-\ell)P^\tau$ and proves (7). Fractional weights at the threshold make the argument valid for atomic output distributions. If $\ell = 1$ or $u = 1$, normalization forces $\Lambda = 1$ and no gain is possible.

For any integrable external score g, linearity gives

$$\mathbb{E}_{Q^\star} g - \mathbb{E}_P g = (1 - \ell)\big(\mathbb{E}_{P^\tau} g - \mathbb{E}_P g\big). \tag{11}$$

The training reward selects the upper tail on which g is evaluated, so a positive reward gain can come with a negative change of g. Two rewards with the same ranking also share the optimizer while producing different numerical gains; for this reason we compare learned and optimal gains only within one model–reward pair (Table 13).

B.2 Estimating the optimum from samples

The tail mean has the variational form $\bar{r}_\tau = \inf_{t\in\mathbb{R}}\{t + \tau^{-1}\mathbb{E}_P(r - t)_+\}$: for any feasible w, $rw \le tw + (r - t)_+$, and the threshold and tie weights above attain equality. For N base samples with sorted rewards $r^{(1)} \ge \cdots \ge r^{(N)}$, let $h = \tau N$, $m = \lfloor h \rfloor$, and $\delta = h - m$; the empirical tail mean is $\widehat{\bar{r}}_{\tau,N} = h^{-1}\big(\sum_{i=1}^{m} r^{(i)} + \delta r^{(m+1)}\big)$, without the last term when $\delta = 0$. To estimate the optimal change of g, we evaluate g on the same reward-selected samples with the same weights. All calculations are performed within each prompt before averaging over prompts, matching the conditional constraint.

The estimate depends on how many samples carry tail weight. With eight blocks, the tail fractions at $\rho = 0.5, 1, 2, 3$ are approximately 0.3719, 0.2618, 0.1149, and 0.0459, so at $\rho = 2$ a pool of $N = 64$ images gives about 7.4 full-weight tail samples per prompt. We therefore compare nested pools of $N = 32$ and $N = 64$ (Table 13); their agreement separates sampling error from the growth of the optimum with ρ.

Figure 3 (Section 4.5) makes the optimum visible in image space. For five fixed dog masks $M_i(h, w)$ with weights w_i, the occupancy map is $O(h, w) = \sum_{i=1}^{5} w_i M_i(h, w)$. The base uses $w_i = 1/5$; the right panel uses the weights of $Q^\star = \ell P + (1 - \ell)P^\tau$ at $\rho = 2$ for a reward that prefers sitting poses, with fractional threshold weights. Every image keeps positive weight, as the lower bound requires, while the occupancy shifts toward the preferred pose. Coordinates are normalized within each frame, and both maps share the scale from 0 to 1. The same picture describes the learned transform in the experiments: it reweights images the base model already draws, and the reward decides which of them gain weight.

B.3 A proxy-error bound for any feasible distribution

The optimum is one feasible distribution; a learned transform is another. For any Q with $\ell \le dQ/dP \le u$, let $e = r - g \in L^1(P)$ and $\Delta_Q h = \mathbb{E}_Q h - \mathbb{E}_P h$. Proposition 2 applied to e gives $\Delta_Q e \le (1 - \ell)(\bar{e}_\tau - \mathbb{E}_P e)$, where $\bar{e}_\tau$ is the mean of e over its own top-τ tail. Since $\Delta_Q g = \Delta_Q r - \Delta_Q e$,

$$\Delta_Q g \;\ge\; \Delta_Q r - (1 - \ell)\big(\bar{e}_\tau - \mathbb{E}_P e\big). \tag{12}$$

An external score can therefore fall under a feasible Q only as far as the upper tail of the proxy error allows; its empirical version sorts the samples by e instead of r. For scores on different scales, we first standardize both with their base means and standard deviations, $\widetilde{r} = (r - \mu_r)/\sigma_r$ and $\widetilde{g} = (g - \mu_g)/\sigma_g$, and apply the bound to $\widetilde{r} - \widetilde{g}$; training keeps the original reward scale.

C Which directions improve the reward

Proposition 3 motivates the rotations: reweighting helps along directions whose CDF features covary with the reward. This appendix constructs the one-direction perturbation (Appendix C.1), proves the

derivative and a finite-step improvement (Appendix C.2), shows that improving directions always exist without architectural limits (Appendix C.3), and explains the role of odd frequencies (Appendix C.4). Fix a prompt and write $f(z) = r(G(z, c), c)$.

C.1 A ONE-DIRECTION PERTURBATION INSIDE THE ARCHITECTURE

Let O be an orthogonal map with $(Oz)_1 = v^\top z$ for a unit vector v, and let C_t change only the first coordinate through the CDF map $S_t(x) = x + t\sin(\pi k x)/(\pi k)$ with $|t| < 1$. Then $T_t = O^{-1} C_t O$ is the perturbation of Proposition 3. It belongs to the architecture whenever the mixing blocks can realize O and O^{-1} and an active block has allowance $\alpha_* > |t|$, with all other coefficients zero; a raw coefficient bias realizes any magnitude below α_*, because its normalized value in (4) varies continuously from zero to α_*.

Write $\psi_{v,k}(z) = \cos(\pi k \Phi(v^\top z))$. If $x' = S_t^{-1}(x)$, then $|x' - x| \le |t|/(\pi k)$, and since the cosine is πk-Lipschitz on $[0, 1]$, $|\cos(\pi k x') - \cos(\pi k x)| \le |t|$. The likelihood ratio at output z is $\Lambda_t(z) = \big(1 + t\cos(\pi k S_t^{-1}(\Phi(v^\top z)))\big)^{-1}$. Using $1/(1+x) = 1 - x + x^2/(1+x)$ with $|x| \le |t|$ gives the uniform remainder

$$\big|\Lambda_t(z) - 1 + t\,\psi_{v,k}(z)\big| \le t^2 + \frac{t^2}{1-|t|} \le \frac{2t^2}{1-|t|} \qquad \text{for all } z, \tag{13}$$

so the expansion is controlled on the entire noise space, including the tails.

C.2 DERIVATIVE AND FINITE-STEP IMPROVEMENT

Let $\mu_f = \mathbb{E}_\gamma f$ and $M_f = \mathbb{E}_\gamma |f - \mu_f| < \infty$. Because $\Phi(v^\top Z)$ is uniform, $\mathbb{E}_\gamma \psi_{v,k} = 0$, so $\eta_{v,k} = \mathrm{Cov}_\gamma(f, \psi_{v,k}) = \mathbb{E}_\gamma[(f - \mu_f)\psi_{v,k}]$. Both Λ_t and its first-order approximation integrate to one, so the remainder integrates to zero; multiplying (13) by $f - \mu_f$ and integrating yields

$$\big|\mathbb{E}_\gamma[f\Lambda_t] - \mu_f + t\,\eta_{v,k}\big| \le \frac{2t^2}{1-|t|} M_f. \tag{14}$$

Dividing by t proves the derivative in Proposition 3 for every integrable f, with no differentiability of the image or the reward. If $\eta_{v,k} \neq 0$, every t with sign opposite to $\eta_{v,k}$ and

$$0 < |t| < \min\Big\{\alpha_*,\ \frac{|\eta_{v,k}|}{2M_f + |\eta_{v,k}|}\Big\}$$

strictly increases the reward; for $f \in [0, 1]$, $M_f \le 1$ gives the simpler threshold $|\eta_{v,k}|/(2 + |\eta_{v,k}|)$.

C.3 IMPROVING DIRECTIONS ALWAYS EXIST

For $f \in L^2(\gamma)$ and a fixed v, let $h_v(x) = \mathbb{E}[f - \mu_f \mid \Phi(v^\top Z) = x]$, a mean-zero element of $L^2(0, 1)$. The functions $\sqrt{2}\cos(\pi k x)$, $k \ge 1$, form a complete orthonormal basis of this subspace, with coefficients $\sqrt{2}\,\eta_{v,k}$, so

$$2\sum_{k\ge 1} \eta_{v,k}^2 = \int_0^1 h_v(x)^2\,dx = \mathrm{Var}\big(\mathbb{E}[f \mid v^\top Z]\big). \tag{15}$$

If all correlations vanish for all directions, then $\mathbb{E}[f - \mu_f \mid v^\top Z] = 0$ for every unit v. For the finite signed measure $d\mu = (f - \mu_f)\,d\gamma$ and any $a \neq 0$, conditioning on $v^\top Z$ with $v = a/\|a\|$ gives $\widehat{\mu}(a) = \mathbb{E}_\gamma[(f - \mu_f)e^{ia^\top Z}] = 0$, and centering gives $\widehat{\mu}(0) = 0$. Uniqueness of the Fourier transform of finite signed measures then forces $f = \mu_f$ almost surely. Every nonconstant square-integrable reward therefore has an improving direction and frequency.

A fixed architecture reaches fewer directions. Let $\mathcal{V}_M$ contain the unit directions whose rotation and inverse can each be realized by at most M mixing blocks, and let $\Gamma_{M,K}(f) = \sup_{v\in\mathcal{V}_M,\,1\le k\le K} |\eta_{v,k}|$. If $\Gamma_{M,K}(f) > 0$, some direction improves the reward within at most $2M + 1$ blocks with an adequate allowance, since blocks with zero coefficients rotate without changing the Gaussian law. The ablations without rotations and with even frequencies only (Table 3) shrink $\mathcal{V}_M$ and the frequency set, respectively.

C.4 Odd frequencies and Gaussian tangent directions

Two further properties explain why the frequency set and the nonlinear CDF directions matter in practice.

Since $\Phi(-z) = 1 - \Phi(z)$, the even-frequency features satisfy $\cos(2\pi k\Phi(-z)) = \cos(2\pi k\Phi(z))$. They are even under $Z \mapsto -Z$, so their correlations with an odd centered reward vanish, while odd frequencies can detect it. Coefficients that depend on the kept part a can break this symmetry, so the restriction concerns the fixed-coefficient features.

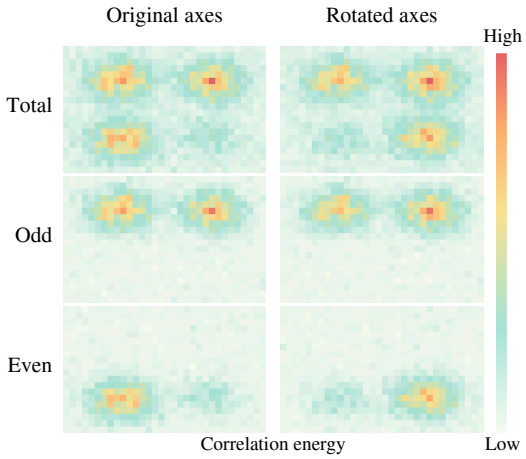


Figure 7: **Reward correlations in original and rotated axes.** Total, odd-frequency, and even-frequency correlation energy on a common scale.

Figure 7 visualizes these correlations. Columns use the original Gaussian coordinates and fixed orthogonal Givens-rotated axes; rows show the total, odd-frequency ($k = 1, 3$), and even-frequency ($k = 2, 4$) correlation energy, $\eta^2_{v,k}$ averaged over channels, so that the total equals odd plus even at each latent location. A brighter location has a larger squared correlation, that is, more first-order gain along that projection; the sign that squaring removes selects the direction of the coefficient in Appendix C.2. The rotated maps evaluate other projections of the same Gaussian law and redistribute the energy across locations and parity groups: rotations change which reward correlations are reachable while the reference distribution stays fixed. Both parity groups carry energy, at different locations, which is consistent with the lower gain of the even-only ablation.

Nonlinear CDF directions also reach rewards that Gaussian reparameterizations miss. Infinitesimal mean and covariance changes have score directions in the span of z_i and $z_i z_j - \delta_{ij}$. Consider

$$f(z) = \frac{1}{2} + \frac{\sin z_1 - \xi \sin(2z_1)}{2(1+\xi)}, \qquad \xi = \frac{e^{3/2}}{2}.$$

This reward lies in $[0, 1]$, is nonconstant, and has an odd centered part. Since $\mathbb{E}[Z_1 \sin(aZ_1)] = ae^{-a^2/2}$, the choice of ξ gives $\mathbb{E}[Z_1(f - \frac{1}{2})] = 0$; other linear correlations vanish by independence and all quadratic ones by odd symmetry. The reward is thus first-order invisible to every Gaussian tangent direction, while (15) still supplies an improving CDF direction.

D Inverse density, training algorithm, and architecture

The black-box training path and the event guarantee both rely on evaluating $q_\phi(y \mid c)$ exactly. Appendix D.1 gives the inverse computation and its gradients, Appendix D.2 states the full algorithm, Appendix D.3 specifies the conditioner, and Appendix D.4 the numerics.

D.1 Inverse density and gradients

For an output noise y, the inverse processes the blocks in reverse order: it undoes the permutation, recomputes the conditioner from the unchanged a, solves each scalar equation $S_{mj}(v_j) = \Phi(b'_j)$ on $[0, 1]$, recovers $\widetilde{b}_j = \Phi^{-1}(v_j)$, and applies $R_m^\top$. Accumulating the block ratios of Lemma 6 gives

$$\log q_\phi(y \mid c) = \log \gamma(y) - \sum_{m,j} \log S'_{mj}(v_{mj}), \tag{16}$$

where every v_{mj} belongs to the inverse preimage in block m.

Figure 8 resolves (16) by latent position. With $a_{m,j} = -\log S'_{mj}(v_{mj})$ mapped back through the permutations, the accumulated map after block b is $C_b(h,w) = \sum_{m \le b} \sum_{\mathrm{ch}} a_{m,\mathrm{ch},h,w}$, and summing C_8 over all positions recovers $\log \Lambda_\phi(y \mid c)$ for that sample. Positive and negative contributions offset each other across blocks, so a bright local value need not mean a large total ratio; the total stays within $[\log \ell, \log u]$ by (5). The maps refer to noise coordinates, and the photographs are the decoded outputs of the same noises.

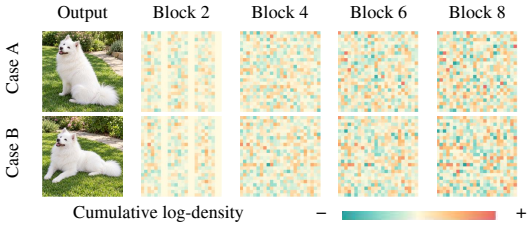


Figure 8: **Accumulation of the noise log-density ratio.** For two samples, cumulative contributions after blocks 2, 4, 6, and 8 ($\rho = 2$, $L = 8$, $K = 4$) and the final images. Each coordinate contributes $-\log S'_{mj}$ at its inverse preimage; contributions are mapped back to the latent grid, summed over channels, and accumulated over blocks on one signed color scale.

The root v of $S(v,\theta) = x$ depends on the coefficients θ and on the right-hand side. Implicit differentiation gives $dv = (dx - \partial_\theta S\, d\theta)/S'(v,\theta)$, with $\partial_{\omega_{mjk}} S_{mj} = \sin(\pi k v)/(\pi k)$; the positive lower bound on S' guarantees a unique root and a finite denominator.

During reverse composition these derivatives propagate through the recomputed conditioner inputs; only the final input y is held fixed. The score-function gradient in Section 4.4 follows from differentiating under the integral, $\nabla_\phi \int f(y) q_\phi(y)\, dy = \int f(y) q_\phi(y) \nabla_\phi \log q_\phi(y)\, dy$, which holds when q_ϕ is differentiable in ϕ and an integrable function dominates the derivative, together with $\mathbb{E}_{q_\phi} \nabla_\phi \log q_\phi = 0$. For independent samples $Y_1, \ldots, Y_M$ at fixed c, the leave-one-out baseline $b_i = (M-1)^{-1} \sum_{j \neq i} f(Y_j)$ is independent of Y_i, so $\mathbb{E}[b_i \nabla_\phi \log q_\phi(Y_i)] = 0$ and each summand remains unbiased; shared baselines only correlate the summands. Rewards, baselines, and samples are held fixed when the score loss is differentiated.

D.2 Training and generation algorithm

Algorithm 1 combines the forward map, the inverse density, both training paths, and generation, in the ruled and numbered format of GLASS Flows (Holderrieth et al., 2026). Uppercase Z, Y, X denote random initial noise, transformed noise, and images, and lowercase letters their values. Indices m, j, k select blocks, transformed coordinates, and frequencies, and t, n, i select iterations, prompts, and samples within a prompt. The permutation Π_m includes the split and Π_m^{-1} restores the layout; $\textsc{Parameters}_\phi(a, c, m)$ evaluates the shared conditioner and biases, returns the Givens rotation, and normalizes the coefficients by (4), so the forward and reverse calls rebuild the same S_{mj} from the unchanged a. The operator sg keeps a value and sets its derivative to zero.

Bisect starts from $(v^-, v^+) = (0, 1)$, sets $v = (v^- + v^+)/2$, replaces v^- by v if $S_{mj}(v) < x_j$ and v^+ otherwise, and stops when $|S_{mj}(v) - x_j| \le \delta$; since $S'_{mj} \ge 1 - W_{mj}$, the root error is at most $\delta/(1 - W_{mj})$ in exact arithmetic. Reported runs use $L = 8$, $K = 4$, $\varepsilon = 10^{-6}$, $\delta = 10^{-10}$, $N = 1000$, $B = 16$, and learning rate 10^{-4}. Black-box training uses $M = 8$, so each baseline averages the other seven rewards of the prompt. Differentiable training passes gradients through the generator's input, and black-box training differentiates the inverse density at detached samples; both update the same transform.

D.3 Conditioner, pairing, and initialization

One convolutional conditioner is shared by the eight blocks. A 3×3 stem maps the kept channels to width 128, two stride-two downsamplers produce widths 256 and 512, and nearest-neighbor upsampling with skip concatenation returns to widths 256 and 128. Each scale uses two 3×3 convolutions, GroupNorm with 32 groups, and SiLU. A 256-dimensional sum of projected text features and a learned block embedding provides FiLM modulation. Separate 1×1 heads predict the rotation angles and the CDF coefficients, and learned per-block coordinate biases are added before normalization.

Pairing applies a factor-two pixel unshuffle and flattens the coordinates. In phase $s \in \{0, 1, 2, 3\}$, the transformed pairs are $(4j + s,\ 4j + s + 1) \bmod D$, and the other two coordinates of each group of four form a; the eight blocks cycle through the four phases twice, and inversion reverses the grouping

Algorithm 1 RATIO: bounded noise transform, inverse density, reward training, and generation

Input: G, r, ν; $L, K, N, B \in \mathbb{N}$, $\rho, \varepsilon, \delta > 0$; fixed permutations $\{\Pi_m\}$;
reward mode $d \in \{\mathrm{diff}, \mathrm{bb}\}$, $M_{\mathrm{bb}} \geq 2$; AdamW configuration $\mathcal{H}$.
Output: ϕ with $\ell \leq q_\phi/\gamma \leq u$ and a one-draw image sampler.

$\alpha_m \leftarrow 1 - e^{-\rho/L}$; $\ell \leftarrow e^{-\sum_m \alpha_m}$; $u \leftarrow e^{\rho}$
Initialize ϕ (Appendix D.3) and optimizer state; set $M = 1$ if $d = \mathrm{diff}$, else $M = M_{\mathrm{bb}}$
function FORWARD($z, c; \phi$)
  for $m = 1, \ldots, L$ **do**
    $(a, b) \leftarrow \Pi_m z$; $(R_m, \omega_m) \leftarrow \text{PARAMETERS}_\phi(a, c, m)$; $\widetilde{b} \leftarrow R_m b$
    $v_j \leftarrow \Phi(\widetilde{b}_j)$; $b'_j \leftarrow \Phi^{-1}(S_{mj}(v_j))$ for all j ▷ Eq. (3)
    $z \leftarrow \Pi_m^{-1}(a, b')$
  end for
  return z
end function
function LOGDENSITY($y, c; \phi$)
  $z \leftarrow y$; $h \leftarrow -\frac{D}{2}\log(2\pi) - \frac{1}{2}\|y\|^2$
  for $m = L, L-1, \ldots, 1$ **do**
    $(a, b') \leftarrow \Pi_m z$; $(R_m, \omega_m) \leftarrow \text{PARAMETERS}_\phi(a, c, m)$
    $x_j \leftarrow \Phi(b'_j)$; $v_j \leftarrow \text{BISECT}(S_{mj}(v) - x_j, [0,1], \delta)$ for all j
    Attach $dv_j \leftarrow (dx_j - (\partial_{\omega_m} S_{mj}) d\omega_m)/S'_{mj}(v_j)$ for all j
    $\widetilde{b}_j \leftarrow \Phi^{-1}(v_j)$; $D_j \leftarrow 1 + \sum_k \omega_{mjk}\cos(\pi k v_j)$ for all j
    $h \leftarrow h - \sum_j \log D_j$; $z \leftarrow \Pi_m^{-1}(a, R_m^\top \widetilde{b})$
  end for
  return h
end function
for $t = 1, \ldots, N$ **do**
  $c_n \overset{\text{iid}}{\sim} \nu$, $z_{ni} \overset{\text{iid}}{\sim} \mathcal{N}(0, I_D)$ for $1 \leq n \leq B$, $1 \leq i \leq M$
  $y_{ni} \leftarrow \text{FORWARD}(z_{ni}, c_n; \phi)$; $r_{ni} \leftarrow r(G(y_{ni}, c_n), c_n)$
  if $d = \mathrm{diff}$ **then**
    $\mathcal{L} \leftarrow -B^{-1}\sum_n r_{n1}$ ▷ Differentiate through $r, G,$ FORWARD
  else
    $b_{ni} \leftarrow (M-1)^{-1}\sum_{j \neq i} r_{nj}$; $a_{ni} \leftarrow \mathrm{sg}[r_{ni} - b_{ni}]$
    $\mathcal{L} \leftarrow -(BM)^{-1}\sum_{n,i} a_{ni}\text{LOGDENSITY}(\mathrm{sg}[y_{ni}], c_n; \phi)$
  end if
  $(\phi, \mathrm{state}) \leftarrow \text{ADAMW}(\phi, \nabla_\phi \mathcal{L}, \mathrm{state}; \mathcal{H})$
end for
function GENERATE($c; \phi$)
  $z \sim \mathcal{N}(0, I_D)$; **return** $G(\text{FORWARD}(z, c; \phi), c)$
end function

and the unshuffle. Initial angle biases alternate between $+\pi/8$ and $-\pi/8$ by block and pair parity, while the angle residual heads, CDF heads, and coefficient biases start at zero, so the initial transform preserves γ. The even-frequency ablation uses $k \in \{2, 4\}$; the ablation without mixing removes the angle parameters; and Affine+KL replaces the CDF head with scale and shift outputs under the same conditional interface.

SD3.5-M uses $16 \times 64 \times 64$ noise and concatenated 2,048-dimensional pooled CLIP-L/G features. FLUX uses the same latent shape before token packing and a 768-dimensional pooled CLIP-L feature. SANA-1.5 uses $32 \times 32 \times 32$ noise and mean-pooled 2,304-dimensional Gemma-2 features. Text features are projected to 256 dimensions.

Figure 9 shows how the conditioner divides the budget, the freedom that Section 4.2 attributes to a shared allowance. After mapping back to the latent grid, the displayed quantity is

$$B_k(h,w) = \frac{100 \sum_{m,\mathrm{ch}} |\omega_{m,\mathrm{ch},h,w,k}|}{\sum_m \alpha_m}.$$

The four maps together use less than 100% of the allowance by (4). Absolute values record where capacity is allocated, and the signed derivative at a sample determines its density contribution in Figure 8.

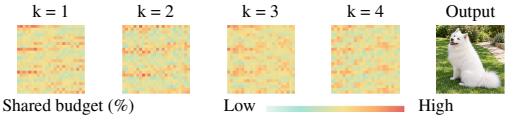


Figure 9: **Allocation of the shared budget.** For Case A of Figure 8, the maps show frequencies $k = 1, 2, 3, 4$ from left to right, followed by the final image. At each latent location, absolute coefficients are summed over blocks and channels, divided by the total allowance $\sum_m \alpha_m$, and multiplied by 100; the maps share one nonnegative scale.

D.4 Numerical evaluation

Generator and conditioner computations use BF16. The Gaussian CDF, its inverse, and log-ratio calculations use FP64 with $\varepsilon = 10^{-6}$, and the CDF inversion uses bracketed iteration on $[0, 1]$ with residual tolerance 10^{-10} and tail-stable evaluations of the log-CDF and log-survival functions. The results of Section 4.5 concern exact arithmetic. At the end of training, the three RATIO runs use 99.50–99.57% of the validation budget, and the mean noise KL is 0.2896 nats for RATIO and 0.2945 for Affine+KL.

E Evaluation protocol

This appendix specifies the data and training runs (Appendix E.1), the metrics and the matched-gain comparison (Appendix E.2), the event tests (Appendix E.3), and the repeats and human studies (Appendix E.4).

E.1 Data and training runs

Differentiable training uses 20,000 unique SFW captions from the Pick-a-Pic v2 training split, with 500 validation captions and 500 disjoint test captions. Rule-reward training uses GenEval prompt templates and a separate set of 500 validation prompts. Training and validation captions are separated from evaluation captions by normalized exact matching, word-order-independent matching, and token-set Jaccard similarity of at least 0.8. Flow-GRPO uses the released models trained by its authors on Pick-a-Pic v1 or GenEval template data.

Table 6: **Training resources.** Generated images (k), trainable parameters (M), and outer iterations. Model citations: Table 4; method citations: Table 1. Counts are per training run. Released Flow-GRPO is not assigned zero historical training cost.

Method	Images (k)	Param. (M)	Iter.
SD3.5-M (2.5B) / PickScore			
RATIO	16.0	13.74	1000
DRaFT (Clark et al., 2024)	16.0	18.78	1000
RSA-FT (Kim & Sim, 2026)	16.0	18.78	1000
DiffusionNFT (Zheng et al., 2025)	2,304.0	18.78	2000
DiverseGRPO (Liu et al., 2025a)	96.0	18.78	1000
GRPO-Guard (Wang et al., 2025a)	1,175.0	18.78	1020
HyperNoise (Eyring et al., 2025)	16.0	75.10	1000
FLUX.1-dev (12B) / PickScore			
RATIO	16.0	13.41	1000
DRaFT (Clark et al., 2024)	16.0	52.30	1000
SANA-1.5 (1.6B) / PickScore			
RATIO	16.0	13.23	1000
DRaFT (Clark et al., 2024)	16.0	22.94	1000
SD3.5-M (2.5B) / rule reward			
RATIO	128.0	13.74	1000
DiffusionNFT (Zheng et al., 2025)	1,152.0	18.78	1000
GRPO-Guard (Wang et al., 2025a)	2,142.7	18.78	1860

No best-value ranking is applied: these are training configurations, not equal-compute performance comparisons.

RATIO uses AdamW at learning rate 10^{-4} for 1,000 steps with 16 prompts per step: one image per prompt for differentiable training (16,000 images) and eight for black-box training (128,000). DRaFT and RSA-FT use rank-32 attention LoRA with scale 64 and the same 16,000-image exposure. HyperNoise uses rank 128, with noise-output scaling and L2 strengths of 0.3, 1, and 3 times its reference value; scaled strengths share trained weights, and each L2 setting is a separate fit.

DiffusionNFT uses groups of 24 images for 48 prompts, with 2,000 iterations for PickScore and 1,000 for GenEval. GRPO-Guard uses the same group sizes for 1,020 and 1,860 iterations. DiverseGRPO uses rank 32, scale 64, $\beta = 0.7$, and $K_{\mathrm{reg}} = 4$, with 1,000 iterations of four prompt groups and 24 images per group. Images per optimizer step equal the prompt batch size times the group size, so

Table 6 reports generated images per run together with trainable parameters and outer iterations. The released Flow-GRPO models carry prior training exposure that the table does not count.

Evaluation prompts, Gaussian noise draws, and evaluators are shared within each backbone. SD3.5-M uses 512^2 images, forty FlowMatchEuler steps, and CFG 4.5; FLUX.1-dev uses 512^2 images, twenty-eight FlowMatchEuler steps, and guidance 3.5; the 1.6B SANA-1.5 checkpoint uses 1024^2 images, twenty native solver steps, and CFG 4.5. DiffusionNFT uses its CFG-free sampler, so both guided and unguided base models are reported.

E.2 Metrics and matched-gain comparisons

PickScore uses the v1 reward model on its native scale. The HPS v2.1 benchmark in Table 4 uses four styles (anime, concept art, paintings, and photos) with 800 prompts each, and reports the equally weighted style mean as 100 times the raw cosine similarity (Wu et al., 2023). CLIPScore is the image–text cosine similarity of OpenCLIP ViT-H/14 trained on LAION-2B (Ilharco et al., 2021), multiplied by 100. Aesthetic v2 is the LAION predictor on CLIP ViT-L/14 features (Schuhmann, n.d.), and Aesthetic v2.5 is a separate evaluator (discus0434, n.d.). The external scores in Figure 18 are HPS v2.1 (Wu et al., 2023), ImageReward v1.0 (Xu et al., 2023), UnifiedReward 2.0 pointwise scores (Wang et al., 2025b), HPS v3 (Ma et al., 2025b), and DeQA-Score (You et al., 2025); each is evaluated on the same images and standardized by its own base-score standard deviation. Image and prompt averages weight every prompt equally.

GenEval uses four images for each of 553 prompts, averages correctness within each of six tags, and reports the unweighted tag mean (Ghosh et al., 2023). The black-box training reward is the nonbinary partial score, whose count, color, and position terms can be negative. T2I-CompBench++ reports eight categories with 300 prompts and ten images each (Huang et al., 2025); numeracy gives half credit for object presence and half for the correct count. DPG-Bench uses all 1,065 prompts, four images per prompt, and 14,392 questions with single-pass prerequisite handling (Hu et al., 2024); its overall score averages dependency-adjusted answers across images, and the category summaries use the evaluator's last-image raw-answer aggregation.

PRDC recall (Naeem et al., 2020) uses DINOv2 ViT-L/14 features (Oquab et al., 2023). It is the fraction of independent base-reference features that fall inside the union of balls around the evaluation features, each ball with radius equal to that feature's fifth-nearest-neighbor distance among the evaluation features. Both pools contain 5,000 images for Pick-a-Pic and 2,000 for DrawBench, and the base self-comparison uses a third independent pool. Recall therefore measures how much of the base distribution a method's outputs still cover.

Table 7: **Matched-gain interpolation inputs.** Target: PickScore 22.58 on SD3.5-M/Pick-a-Pic. w weights the right endpoint. Methods are cited in Table 1; final retention is reported there and in Table 3.

Method	Knob	Endpoints	w
DRaFT (Clark et al., 2024)	s	$0.25 \to 0.5$	0.9249
RSA-FT (Kim & Sim, 2026)	s	$0.5 \to 0.75$	0.0010
DiffusionNFT (Zheng et al., 2025)	s	$0.25 \to 0.5$	0.9941
Flow-GRPO (Liu et al., 2025b)	s	$0.25 \to 0.5$	0.9550
DiverseGRPO (Liu et al., 2025a)	s	$0.25 \to 0.5$	0.8489
GRPO-Guard (Wang et al., 2025a)	s	$0.25 \to 0.5$	0.7546
HyperNoise (Eyring et al., 2025)	$\lambda/\lambda_{\text{ref}}$	$3 \to 1$	0.8382
Affine + KL	λ	$3 \to 1$	0.8406

Each method has a strength parameter. RATIO uses $\rho \in \{0.5, 1, 2, 3\}$. For LoRA methods, s multiplies the learned weight update; for HyperNoise, it scales the residual noise update, $z' = z + sf_\phi(z, c)$. Displayed paths use $s \in \{0.25, 0.5, 0.75, 1\}$. To compare at RATIO's reward, each metric is interpolated linearly between the two evaluated strengths that enclose it, separately for each prompt set, with the weights in Table 7: a right-endpoint weight w gives $y = (1 - w)y_L + wy_R$, applied to each repeat before averaging. HyperNoise uses the largest interpolated value across its scaling and L2 paths for each metric. The Affine+KL penalty is calibrated at step 300, so its final KL and its matched-gain interpolation describe different operating points. The four-decimal weights reproduce the matched means in Tables 1 and 3.

E.3 Event families, sampling, and error control

GenEval metadata define 5,051 events: 553 official correctness events, 2,940 color events, 1,360 count events, and 198 object-presence events. Color events test ten colors on designated objects in 94 color prompts and 100 color-attribute prompts, with missing detection as a separate outcome. Count events test the values 0–16 on 80 prompts, and presence events test both required classes in 99 two-object prompts. Exactly 1,125 events express the required outcomes of their prompts. Figure 10

shows how an image receives a color, count, or presence label. In its cup panel, a manual purple label is compared with a fixed HSV rule (hue 0.70–0.88, saturation ≥ 0.15, value ≥ 0.10, covering at least half of the marked cup region), which returns negative; the other panels mark three sports balls and the presence of both a backpack and a potted plant. Regions and instances are annotated manually, and the color rule is a diagnostic separate from official GenEval scoring. Figure 11 shows the requirement categories that the events cover.

Each prompt has $n_0 = 256$ base images and $n_1 = 128$ method images, drawn in pools separate from the four-image GenEval evaluation. For a method with three training runs, each image comes from a run chosen uniformly at random; the event count is then binomial under the equal-weight mixture, which satisfies the bounds by Appendix A.3. Flow-GRPO uses its single released model. Events of the same prompt share images.

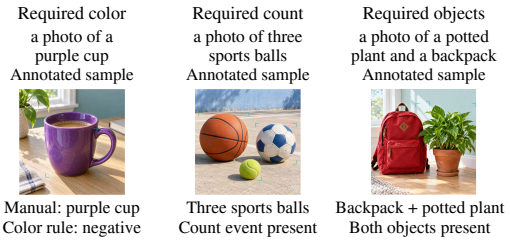


Figure 10: **Events are sets of images defined by a checkable property.** GenEval examples of a color event (purple cup), a count event (three sports balls), and a presence event (backpack and potted plant), with their labels.

Let $L_\delta(k, n) = \mathrm{B}^{-1}(\delta; k, n - k + 1)$ and $U_\delta(k, n) = \mathrm{B}^{-1}(1 - \delta; k + 1, n - k)$ be one-sided Clopper–Pearson bounds with $\delta = 0.0125$, where $\mathrm{B}^{-1}(s; a, b)$ is the s quantile of a beta distribution, $L_\delta(0, n) = 0$, and $U_\delta(n, n) = 1$. Write (p_L, p_U) and (q_L, q_U) for the base and method bounds. A deletion is rejected when $q_U < \underline{Q}(p_L)$ and an amplification when $q_L > \overline{Q}(p_U)$. Because $\underline{Q}$ and $\overline{Q}$ are increasing, a false rejection under true bounds requires one of the two confidence bounds to fail, so the error is at most $2\delta = 2.5\%$ per direction and 5% per event.

Deletion is detectable at these sample sizes when $U_\delta(0, n_1) < \underline{Q}(p_L)$, and amplification when $L_\delta(n_1, n_1) > \overline{Q}(p_U)$; the directional denominators in Table 9 use these conditions, and every one of the 5,051 events is detectable in at least one direction. Paired rejection-rate intervals use 10,000 bootstrap resamples of the 553 prompt blocks, recomputing numerator and denominator together. RATIO is tested against its own $\rho = 2$ or $\rho = 3$ bounds and the controls against the $\rho = 2$ bounds. The choice of reference matters for DiffusionNFT, whose sampler is CFG-free: with the same images, the unguided base gives 737 rejections and 43 prompt-valid deletions, and the guided base 335 and 196.

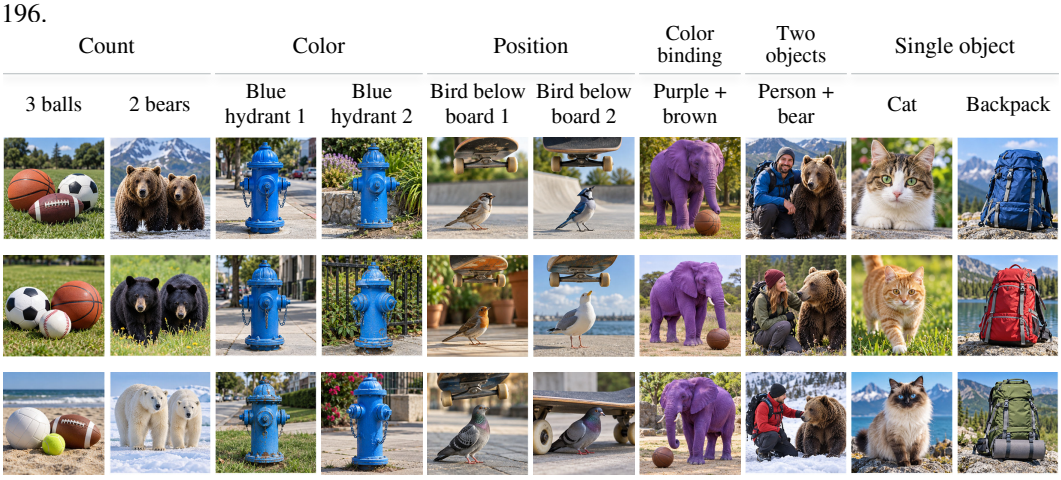


Figure 11: **Requirement categories covered by the event families.** Three rows of examples for counting, color, position, color binding, co-presence, and single objects, with the prompt and category above each column. Different appearances satisfy the same requirement, which is why an event is a set of images; the quantitative tables aggregate official correctness.

E.4 Repeats and human judgments

Score tables report the mean $\bar{x} = \frac{1}{3}\sum_s x_s$ over three repeats: independent training runs for learned methods and three sets of evaluation noise for the released Flow-GRPO and base models. Interpolated

means are conditional on the fixed interpolation weights. Event tables report counts, human tables report outcome counts and decisive-prompt rates, and resource tables report configurations. Derived quantities use unrounded values, and matched-gain and oracle sign comparisons are point-estimate comparisons.

Pairwise preference uses 200 DrawBench prompts. The two images of a pair share the initial noise and appear in random left–right order. Five judges per pair choose the preferred image or indifference, separately for overall preference and prompt matching. A prompt is a win or a loss when at least three judges favor that side and a tie otherwise; win rates are computed over decisive prompts with pointwise 95% Wilson intervals at the prompt level. In Figure 20, a minimum vote margin $m \in \{1, \ldots, 5\}$ additionally requires $|n_A - n_B| \geq m$, so the curves evaluate nested subsets of the 200 prompts.

Set-diversity judgments use 100 DrawBench prompts: judges compare two groups of four images and choose the group that is more diverse while matching the prompt, with five judges per comparison. Figure 1 (Section 1) shows the kind of within-prompt variation that these judgments and PRDC recall measure. The Aesthetic overprocessing task uses 200 Pick-a-Pic prompts: judges see base, RATIO, and DRaFT images with the same initial noise and choose the most overprocessed image or none, five times per prompt. Vote shares are averaged equally over prompts, and the paired DRaFT-minus-RATIO interval uses 10,000 resamples of the 200 prompt blocks.

F Additional results

The results below extend each experiment of Section 5: the strength paths behind the matched comparison (Appendix F.1), the event families and their detectability (Appendix F.2), the black-box GenEval reward (Appendix F.3), compositional and dense prompts (Appendix F.4), reward mismatch and the optimum (Appendix F.5), and human judgments (Appendix F.6).

F.1 Strength paths and coverage

RATIO's ρ bounds the density ratio, a LoRA strength scales a weight update, and a HyperNoise strength scales a noise update; matching the reward gain places these different knobs at a common coordinate. Table 8 compares RATIO with the largest interpolated recall and CLIPScore among the seven controls at each of RATIO's reward levels. On Pick-a-Pic, RATIO's recall exceeds this envelope by 0.1–0.3 points at $\rho \leq 1$, by 5.1 at $\rho = 2$, and by 0.7 at $\rho = 3$; on DrawBench at $\rho = 3$, it is 0.1 points lower. Because every envelope entry takes the best control for that metric, reward level, and prompt set, the comparison is conservative for RATIO. The advantage is largest where the reward gain is substantial and the lower factor is still large ($\ell = 0.17$ at $\rho = 2$), consistent with Figure 12.

Table 8: **Budget sensitivity.** At each RATIO reward coordinate, External is the per-metric maximum over the seven controls in Table 1. The winning control may change across cells. All scores are $\times 100$.

ρ	Recall ↑		CLIP ↑	
	RATIO	External	RATIO	External
Pick-a-Pic / SD3.5-M (2.5B)				
0.5	**93.5**	**93.4**	30.07	**30.02**
1	91.3	91.0	30.07	29.99
2	85.7	80.6	**30.08**	29.62
3	78.2	77.5	29.92	29.55
DrawBench / SD3.5-M (2.5B)				
0.5	**94.0**	**93.9**	30.45	**30.44**
1	93.6	93.1	30.43	30.36
2	90.1	89.0	**30.47**	30.22
3	87.0	87.1	30.28	30.17

Bold marks the maximum down each column within a prompt-set block, not a test of significance.

Along every path, recall falls as the PickScore gain grows. Below a gain of about 0.3 the four paths nearly coincide; beyond it, RATIO's path lies above the others, and at a gain near 0.6 it keeps about 86% recall where DRaFT and HyperNoise keep about 67–72%. DRaFT and Flow-GRPO reach gains above 1 by continuing along steeper paths. In CLIPScore, RATIO stays at the base level up to $\rho = 2$, while every control falls below it as soon as its gain exceeds about 0.3. Recall counts how many base images have a close neighbor among a method's images in DINOv2 space. Figure 13 shows the kind of variation that such coverage reflects: pose, framing, costume, and surroundings vary among the outputs of one prompt. Whole-image proximity can also coexist with different local correspondences, which Figure 14 exposes through patch colors: spatially consistent color transfer indicates coherent appearance matches, and whitened patches mark matches that fail the mutual-neighbor check.

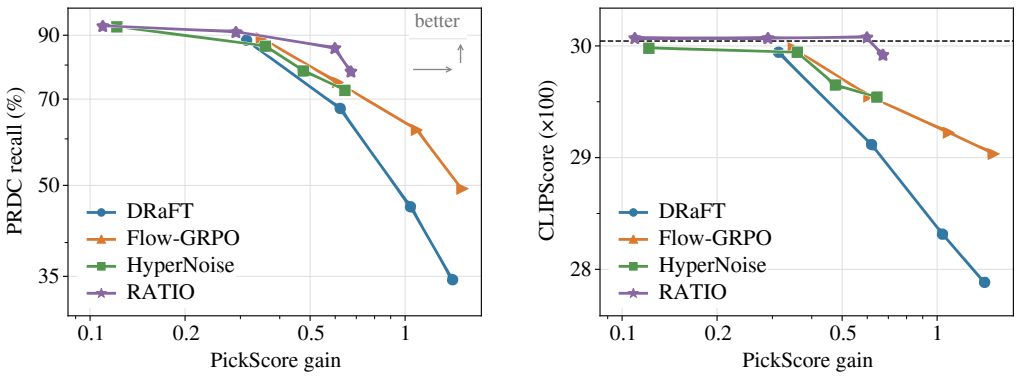


Figure 12: **Reward–retention paths under PickScore training.** SD3.5-M on 500 Pick-a-Pic prompts with ten images each and 5,000 independent base images. RATIO uses $\rho = 0.5, 1, 2, 3$; DRaFT, Flow-GRPO, and HyperNoise use strengths $s = 0.25, 0.5, 0.75, 1$. Left: PRDC recall (%). Right: CLIPScore; the dashed line marks the guided base. Both horizontal axes and the recall axis are logarithmic. Points average three training runs, except Flow-GRPO's three evaluation-noise repeats of its released model.

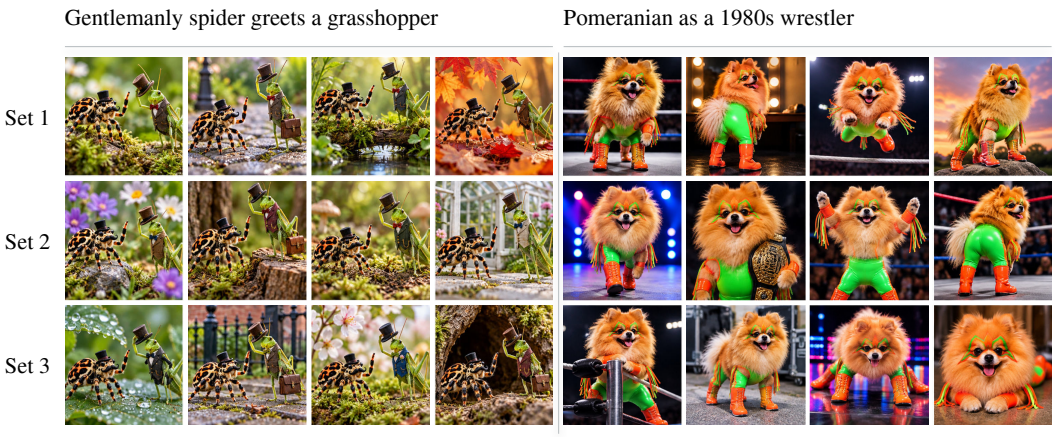


Figure 13: **Variation within unusual prompts.** Two DrawBench prompts, a gentlemanly spider greeting a grasshopper and a Pomeranian dressed as a 1980s professional wrestler, with twelve outputs each in three rows of four. Pose, framing, costume, and surroundings vary within each prompt; PRDC recall measures how much of this variation of the base model remains after alignment.

F.2 Event families and detectability

Whether a violation can be detected depends on the base probability of the event and on the sample sizes. At $\rho = 2$, deletion is detectable for 1,085 of the 5,051 events and amplification for 4,560 (Table 9); among the 1,125 prompt-valid events, 1,010 are deletion-detectable. DRaFT deletes 197 prompt-valid events, 17.5% of the family and 19.5% of its detectable part. The paired difference in total rejection rate, DRaFT minus RATIO, is 6.79 points (95% CI 5.72–7.89) over the 553 prompt blocks. RATIO has no rejection in any family. Affine+KL has 262 rejected deletions, 232 of them prompt-valid, and 124 rejected amplifications at 1.017× RATIO's mean noise KL; Figure 4 displays the valid deletions and all amplification tests.

Table 9: **Directional eligibility.** SD3.5-M, $\rho = 2$. Eligibility depends only on the guided base and sample sizes. RATIO has no rejections; DRaFT (Clark et al., 2024) and affine deletion counts are shown by event family.

Family	Total	Eligible		Deletions	
		Down	Up	DRaFT	Affine
Correctness	553	463	330	81	95
Color	2940	298	2841	73	88
Count	1360	126	1350	49	58
Presence	198	198	39	20	21

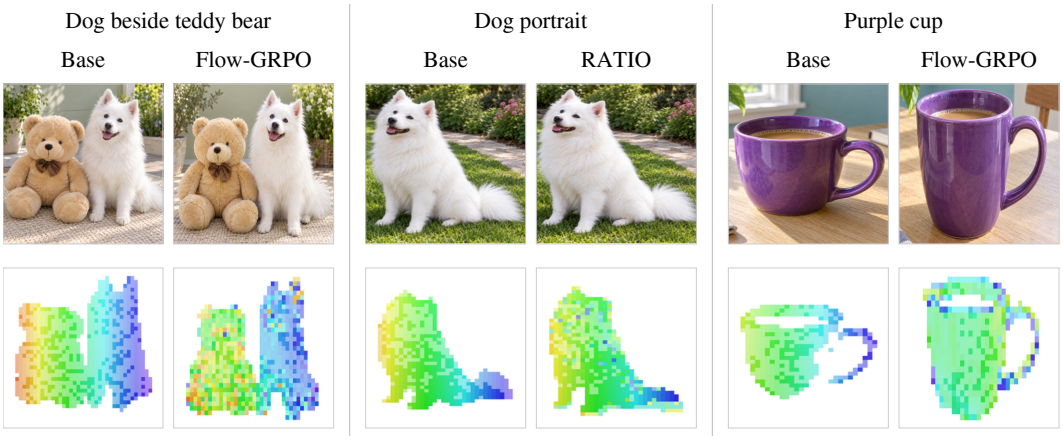


Figure 14: **Local appearance correspondence within paired outputs.** Each group shows two images above their foreground patch correspondences, all at the same square size. From left to right: a dog beside a teddy bear, a dog portrait, and a purple cup; the first and third compare the base with Flow-GRPO, and the middle compares the base with RATIO. Images are resized to 320×320 and split into 8×8-pixel patches (40×40 grid). Nearest patches use a fixed 17-dimensional descriptor of RGB statistics and gradients within fixed foreground masks. Source position determines color, target patches inherit the color of their nearest source patch, and non-mutual matches are whitened by 45%.

Figure 15 groups events by their empirical base probability $\hat{p} = k_0/256$. Starting from 5-point intervals, adjacent bins are merged from low to high until each contains at least 25 events, and a smaller remainder joins the preceding bin, so the bins hold 25–1,944 events. Each point is placed at its bin's mean $\hat{p}$ with height $100(1 - \text{rejected deletions/events})$. DRaFT's deletions concentrate in events of moderate and high base probability, where deletion is detectable, whereas low-frequency color and count events remain unrejected mainly because their deletion cannot be detected with 128 images. The event families test object presence, counts, colors, and correctness; relations and viewpoints enter through the official correctness events. Figure 16 shows why presence alone is insufficient: a scene can contain every named object and still violate a required relation or viewpoint.

F.3 The black-box GenEval reward

Table 10 pairs the partial GenEval training reward with official correctness and retention, including the CFG-free reference for DiffusionNFT. RATIO raises every tag except single objects, whose base accuracy of 98.4% leaves almost no room, and the largest gains occur where the base model is weakest: position rises from 23.3% to 52.4%, counting from 53.4% to 69.3%, and color attribution from 53.4% to 67.6%.

Table 10: **Rule-reward optimization on SD3.5-M** (Stability AI, 2024). Partial is the nonbinary training reward; All is official six-tag macro correctness (Ghosh et al., 2023). All metrics are higher-is-better; values are $\times 100$. Recall uses the guided base reference, including the CFG-free rows, as in the original retention comparison.

Method	**GenEval: training feedback and official correctness**								**Pick-a-Pic**	**DrawBench**
	Partial	All	Single	Two	Count	Color	Pos.	Attr.	Rec.	CLIP
SD3.5-M (2.5B) / guided reference and guided methods										
Guided base	72.4	64.2	98.4	77.0	53.4	79.8	23.3	53.4	**94.8**	30.44
RATIO	86.6	76.8	98.1	87.3	69.3	86.2	52.4	67.6	85.4	**30.57**
Flow-GRPO (Liu et al., 2025b)	**95.5**	95.2	**99.7**	**99.1**	94.6	**93.8**	**98.9**	85.4	51.9	30.10
GRPO-Guard (Wang et al., 2025a)	95.4	**95.5**	99.6	98.4	**95.2**	93.2	97.6	**88.8**	54.8	30.34
SD3.5-M (2.5B) / CFG-free reference and DiffusionNFT										
Unguided base	36.3	24.9	73.4	25.2	14.6	23.9	4.7	7.7	**91.2**	25.08
DiffusionNFT (Zheng et al., 2025)	**96.9**	**98.4**	**99.9**	**99.9**	**98.7**	**99.3**	**98.2**	**94.6**	43.8	**29.73**

Sampler groups are kept separate; bold is computed within each group. GenEval category abbreviations follow Table 4. Training-feedback scores must not be read as binary correctness.

Theorem 1 bounds what any distribution within the $\rho = 2$ bounds can reach on each tag. For a prompt with base accuracy p, the correctness event cannot exceed $\overline{Q}(p) = \min\{up, 1 - \ell(1-p)\}$. Because $\overline{Q}$ is concave, the average of the per-prompt ceilings over a tag is at most $\overline{Q}$ evaluated at the tag's mean accuracy, so the tag means of the base model give upper bounds: 86.9% for position,

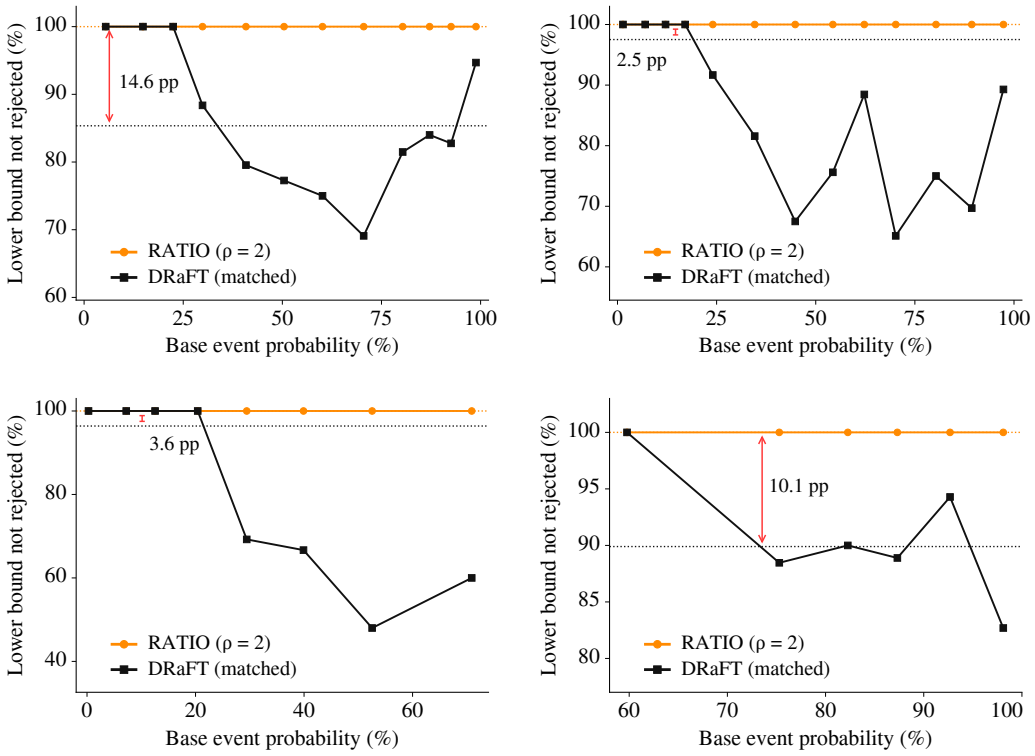


Figure 15: **Deletion rejections by event family and base probability.** SD3.5-M after PickScore training, RATIO ($\rho = 2$) and DRaFT ($s = 0.5$). Panels: correctness (553 events), color (2,940), count (1,360), and object presence (198). Each prompt supplies 256 base and 128 method images. Points report the share of events without a rejected deletion in bins of base probability that contain at least 25 events, placed at the mean base frequency. Dotted lines give whole-family shares, and arrows mark differences of 14.6, 2.5, 3.6, and 10.1 points. Denominators include events whose deletion is undetectable; Table 9 gives directional detectability.



Figure 16: **Relational and viewpoint requirements.** Left: a donkey and an octopus holding a rope while a cat jumps over it, with the donkey holding the rope in its mouth. Right: a tomato on a pumpkin on a kitchen stool, with a fork sticking into the pumpkin, viewed from above. Twelve outputs per prompt. The full prompts separate object presence from the required relations and viewpoint.

92.1% for counting and color attribution, 96.1% for two objects, and 96.6% for color. RATIO covers 37–54% of the distance from the base to these values, and since the true ceilings are lower, at least that share of the available headroom. Flow-GRPO and GRPO-Guard exceed the position, counting,

and two-object values with 97.6–98.9%, 94.6–95.2%, and 98.4–99.1%; no distribution satisfying the $\rho = 2$ bounds can produce these point estimates, and their recall of 51.9–54.8% shows the coverage they give up. The black-box path thus behaves as the bounds predict: it moves probability toward correct compositions until the lower factor, which also keeps 17% of every incorrect one, becomes the limit.

F.4 Compositional and dense prompts

Table 11 resolves PickScore transfer into the eight T2I-CompBench++ categories. DRaFT's larger reward gain comes with lower spatial and numeracy scores and higher color and texture scores, and GRPO-Guard reaches 60.6 on numeracy and 37.8 on complex composition, above RATIO's 59.5 and 36.6; RATIO stays within 0.4 points of the base in every category, as expected from a transform that reweights base images for a preference reward without targeting composition. Table 12 reports DPG-Bench with one checkpoint per method; its overall score is dependency-adjusted and its categories use raw-answer aggregation, so the categories do not average to the overall score.

Table 11: **Compositional transfer across all seven external controls.** SD3.5-M (Stability AI, 2024), PickScore training; default $s = 1$, RATIO $\rho = 2$. T2I-CompBench++ categories (Huang et al., 2025) are separate scores, each $\times 100$.

Method	Attribute binding			Spatial relations		Composition		
	Color	Shape	Texture	2D	3D	Numeracy	Nonsp.	Complex
SD3.5-M (2.5B) / complete default-checkpoint comparison								
Guided base	79.9	56.7	73.3	28.4	37.4	59.4	**31.5**	36.2
Unguided base	56.2	36.9	51.1	10.9	24.1	37.3	27.0	25.3
DRaFT (Clark et al., 2024)	80.8	56.8	**74.5**	25.0	35.4	55.2	30.4	34.5
RSA-FT (Kim & Sim, 2026)	80.2	56.9	74.0	27.2	37.3	57.6	31.1	36.1
DiffusionNFT (Zheng et al., 2025)	79.0	55.0	73.3	23.5	36.2	55.5	30.5	33.7
Flow-GRPO (Liu et al., 2025b)	**81.7**	57.7	**74.5**	28.1	37.9	59.3	31.2	36.9
DiverseGRPO (Liu et al., 2025a)	81.4	57.7	74.1	29.0	38.0	59.7	31.4	37.6
GRPO-Guard (Wang et al., 2025a)	81.4	**58.1**	**74.5**	**29.5**	**38.1**	**60.6**	31.4	**37.8**
HyperNoise (Eyring et al., 2025)	80.2	57.0	73.5	28.5	37.3	58.7	31.2	36.9
RATIO ($\rho = 2$)	80.2	57.0	73.5	28.6	37.2	59.5	**31.5**	36.6

All metrics are higher-is-better; Nonsp. denotes non-spatial relations. The unguided row identifies the CFG-free reference for DiffusionNFT; the guided row is the common visual-retention reference.

Table 12: **Dense-prompt transfer across all seven external controls.** SD3.5-M (Stability AI, 2024), PickScore training; $s = 1$, RATIO $\rho = 2$. DPG-Bench (Hu et al., 2024) uses all 1,065 prompts. All entries retain the native percentage scale.

Method	Overall	Question categories				
	All	Global	Entity	Attribute	Relation	Other
SD3.5-M (2.5B) / complete default-checkpoint comparison						
Guided base	79.8	84.8	89.6	**89.3**	91.6	82.1
Unguided base	60.0	71.8	77.3	75.6	74.9	68.8
DRaFT (Clark et al., 2024)	78.8	86.6	88.8	88.3	88.6	83.7
RSA-FT (Kim & Sim, 2026)	79.1	85.3	88.9	88.6	89.6	81.2
DiffusionNFT (Zheng et al., 2025)	78.4	83.9	88.7	87.5	88.3	80.0
Flow-GRPO (Liu et al., 2025b)	79.9	84.7	89.1	89.2	91.0	84.3
DiverseGRPO (Liu et al., 2025a)	80.1	83.9	89.4	89.2	91.9	81.2
GRPO-Guard (Wang et al., 2025a)	**80.5**	**87.7**	**89.7**	**89.3**	**92.2**	**85.6**
HyperNoise (Eyring et al., 2025)	79.3	84.2	88.8	88.9	90.7	82.9
RATIO ($\rho = 2$)	79.7	83.0	89.2	88.7	91.0	82.8

All metrics are higher-is-better. Overall uses dependency-adjusted aggregation; categories use the evaluator's last-image raw-answer aggregation. Do not average category columns to reconstruct Overall.

Figure 17 compares raw and dependency-adjusted DPG scores on the same outputs. For a prompt with q questions, the raw score counts yes answers over four images and divides by $4q$, and the adjusted score first sets to zero every answer whose prerequisite fails in the evaluator's single-pass order. Scores are averaged within question-count groups and then over repeats. The groups at $q = 10, 20, 30, 39$ contain 517, 40, 3, and 1 prompts, so the estimates for the longest prompts rest on few prompts. The adjusted curves lie roughly ten points below the raw curves for all three models, which shows prerequisite errors propagating through dependent questions. Up to about 28 questions the three models stay within a few points of one another; beyond that, the groups contain at most three prompts and the curves fluctuate.

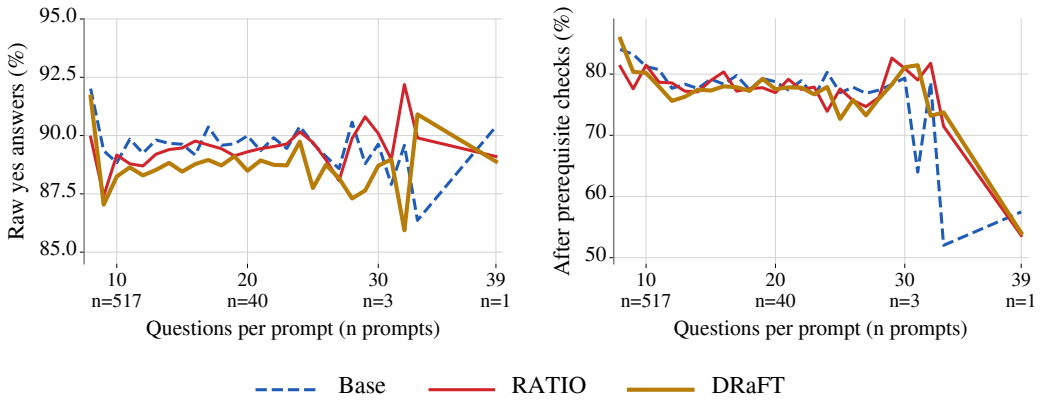


Figure 17: **Prerequisite errors lower DPG scores across question-count groups.** SD3.5-M base, RATIO ($\rho = 2$), and DRaFT ($s = 1$) after PickScore training, on all 1,065 DPG-Bench prompts with four images per prompt and three repeats. Left: raw yes-answer fraction. Right: fraction after single-pass prerequisite checks. Image-level fractions are averaged over prompts and repeats for each question count; all 27 observed counts are shown, with group sizes marked, including three prompts at 30 questions and one at 39.

F.5 Reward mismatch and the optimum

Proposition 2 states that, at the optimum, the reward decides which images gain probability, and (11) predicts the change of any external score from the base samples alone. Figure 18 tests this prediction broadly: for each setting, it compares the optimal change of an external score, computed from 64 base images per prompt, with the change realized by the learned transform. The seven external scores are HPS v2.1, ImageReward v1.0, Aesthetic v2.5, UnifiedReward 2.0, CLIPScore, HPS v3, and DeQA-Score, and the four panels contain:

- SD3.5-M/PickScore: the seven scores at $\rho = 0.5, 1, 2, 3$ on 500 Pick-a-Pic prompts, plus GenEval correctness at $\rho = 2$ on 553 prompts (29 settings);
- FLUX and SANA/PickScore: the seven scores at $\rho = 2$ on Pick-a-Pic (14 settings);
- SD3.5-M/Aesthetic: the seven scores plus PickScore at $\rho = 1, 2, 3$ on Pick-a-Pic, plus GenEval correctness at each budget (27 settings);
- partial GenEval reward: the seven scores plus PickScore and Aesthetic v2 on 553 GenEval prompts at $\rho = 2$ (9 settings).

Base-score standard deviations pool all images within each group. The learned changes follow the predicted sign in 70 of the 79 settings. The learned transform is one feasible distribution and the optimum another, so their values need not coincide: for SD3.5-M/PickScore at $\rho = 2$, the optimal and learned CLIPScore changes are -0.21 and $+0.04$ base standard deviations. Inequality (12) bounds the change of any feasible distribution, including the learned one.

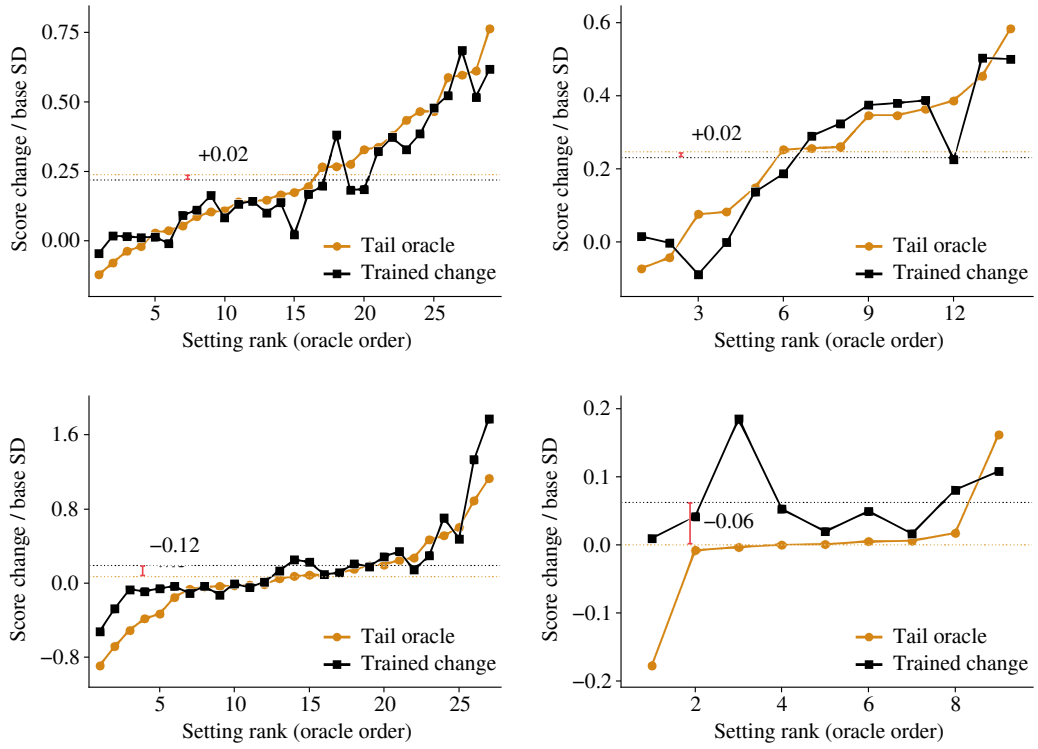


Figure 18: **Optimal and learned external changes across settings.** Panels, in reading order: SD3.5-M/PickScore (29 settings), FLUX and SANA/PickScore (14), SD3.5-M/Aesthetic (27), and SD3.5-M/partial GenEval reward (9). A setting pairs a RATIO budget with one external metric (listed in the text). Optimal changes use 64 base images per prompt and (11); learned changes use three-repeat means. Both are divided by the base-score standard deviation and sorted together by the optimal prediction. Dotted lines are panel means, and annotations give the mean difference, optimal minus learned. Nine pairs have opposite point-estimate signs.

Table 13 compares learned gains with the empirical optimum estimated from nested pools of 32 and 64 base images per prompt. The two pools agree to within 0.04 at every budget, so sampling error in the optimum is small relative to the gap. At $\rho = 2$ the learned SD3.5-M PickScore gain is 67.6% of the 64-image estimate, and at $\rho = 3$ it is 59.2%: widening the bounds raises the optimum faster than the trained transform follows. Reward scales are native to each objective, so gains are compared only within a model–reward block. Across blocks, the realized fraction at $\rho = 2$ lies between 60.2% (FLUX, PickScore) and 73.8% (SANA, PickScore), and for the Aesthetic reward it stays between 65.5% and 70.0% across budgets.

Table 13: **Empirical upper-tail diagnostics.** U_N estimates full-band gain from N base draws per prompt. Gain is learned reward gain, in the reward's native units; Real. is Gain/U_{64} (%). Model references: Table 4.

ρ	U_{32}	U_{64}	Gain	Real.
SD3.5-M (2.5B) / PickScore				
0.5	0.26	0.26	0.11	41.5
1	0.50	0.50	0.29	57.4
2	0.87	0.88	0.60	**67.6**
3	**1.09**	**1.13**	**0.67**	59.2
SD3.5-M (2.5B) / Aesthetic v2				
1	0.30	0.30	0.21	**70.0**
2	0.52	0.53	0.35	65.9
3	**0.65**	**0.68**	**0.44**	65.5
SD3.5-M (2.5B) / partial GenEval				
2	0.22	0.22	0.14	64.4
FLUX.1-dev (12B) / PickScore				
2	0.57	0.58	0.35	60.2
SANA-1.5 (1.6B) / PickScore				
2	0.72	0.73	0.54	73.8

Figure 5 shows the Aesthetic paths behind Table 5. RATIO at $\rho = 2$ and DRaFT at $s = 0.5$ reach Aesthetic gains of 0.347 and 0.359; along both paths, increasing the strength lowers text consistency and correctness, and RATIO keeps higher recall at every displayed operating point. At a comparable Aesthetic gain, the two methods therefore differ mainly in how much of the base

distribution they keep, while the direction of the external change is shared, as (11) predicts for any method that follows the same reward ranking.

A reward ranks images by particular visual evidence, and two scores can rank the same image by different evidence. Figure 19 illustrates this on one image with two simple scores.

The color-contrast score $S_A = N^{-1}\sum_p(0.5R_p - G_p + 0.5B_p)$ emphasizes surfaces, and the luminance-gradient energy $S_B = 32N^{-1}\sum_{p\sim q}(Y_q - Y_p)^2$ emphasizes boundaries, where $Y = 0.2126R + 0.7152G + 0.0722B$ on a 192×192 input and $p \sim q$ ranges over neighboring pixels. The signed maps use a constant gray baseline, with cyan for negative, pale yellow for zero, and coral for positive attributions. The difference map in panel (d) shows where the two scores disagree: a transform trained on S_A would move probability toward images with strong color surfaces, and one trained on S_B toward images with strong edges, even under the same bounds. Learned reward models such as Aesthetic v2 also rank images by specific visual evidence, and the sign of an external change follows this ranking (Section 5.5).

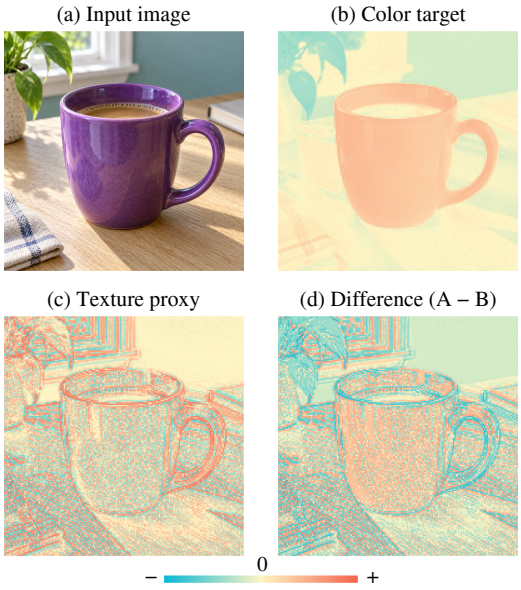


Figure 19: **Two scores use different spatial evidence.** Signed integrated gradients of a color-contrast score and a luminance-gradient score, and their difference, on a shared scale.

F.6 Human judgments

RATIO wins 55.8% of decisive prompts against the guided base (95% CI 47.5–63.8), 49.7% against DRaFT at $s = 0.5$ (42.1–57.3), and 60.1% against DiffusionNFT at $s = 0.5$ (52.0–67.8). The first two intervals include 50%, and the comparison with DiffusionNFT favors RATIO at these operating points. Table 14 lists the outcome counts for both questions.

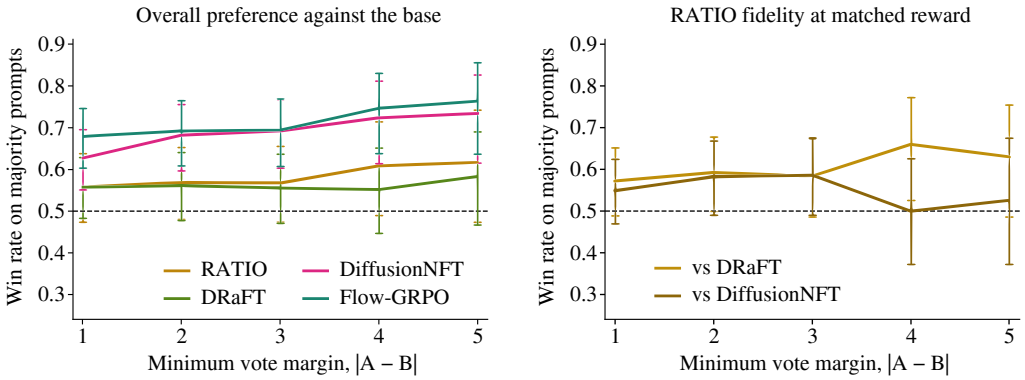


Figure 20: **Human preference conditional on vote agreement.** SD3.5-M after PickScore training; 200 DrawBench prompts, five judges per pair. Left: overall preference for RATIO ($\rho = 2$), DRaFT, DiffusionNFT, and Flow-GRPO ($s = 1$) against the guided base. Right: prompt matching of RATIO against DRaFT and DiffusionNFT ($s = 0.5$). A margin m selects prompts with at least three votes for one side and $|n_A - n_B| \geq m$. Points give decisive-prompt win rates with pointwise 95% Wilson intervals; the dashed line marks 50%, and subset sizes appear below.

Raising the vote-margin threshold in Figure 20 keeps prompts with stronger agreement among judges. From $m = 1$ to $m = 5$, the left-panel counts fall from 138 to 47 for RATIO, 174 to 72 for DRaFT, 169 to 64 for DiffusionNFT, and 159 to 55 for Flow-GRPO, and the right-panel counts fall from 138 to 46 against DRaFT and from 153 to 38 against DiffusionNFT. Each interval describes preference within its selected subset, with prompts as sampling units. In the separate 100-prompt set-diversity task, RATIO receives 69.3% of decisive preferences against DRaFT (59.0–78.0) and 58.3% against DiffusionNFT (47.7–68.3). Pairwise preference ties each comparison to a shared initial noise, while the set task tests variation across several outputs of one prompt.

Table 14: **Human pairwise judgments.** SD3.5-M, 200 DrawBench prompts and five judges per pair. RATIO uses $\rho = 2$; DRaFT and DiffusionNFT (Clark et al., 2024; Zheng et al., 2025) use $s = 0.5$. W/T/L uses a three-vote majority. Win rate excludes tied prompts.

Question	W/T/L	Win (%)
RATIO / guided base		
Preference	77/62/61	55.8
Prompt match	79/57/64	55.2
RATIO / DRaFT $s = 0.5$		
Preference	80/39/81	49.7
Prompt match	79/62/59	57.2
RATIO / DiffusionNFT $s = 0.5$		
Preference	86/57/57	60.1
Prompt match	84/47/69	54.9

Confidence intervals and their interpretation remain in the text and Figure 20, not in table cells. Different questions are not ranked against each other.

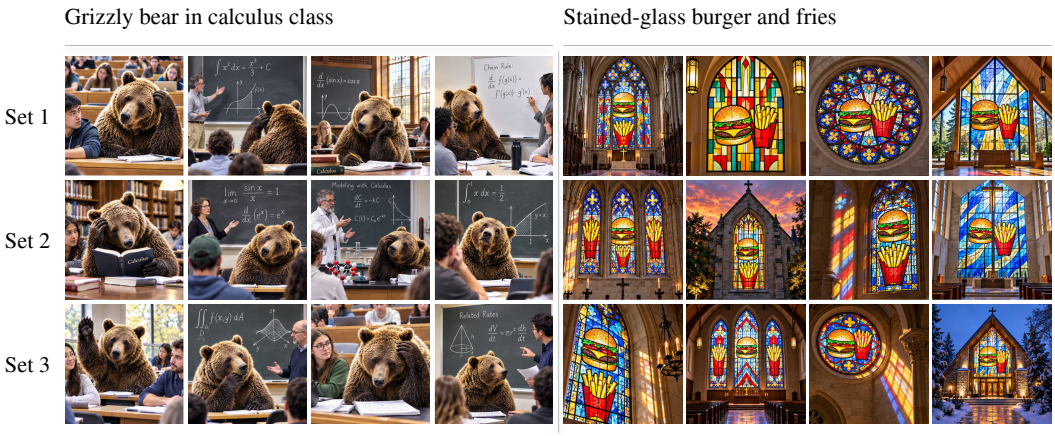


Figure 21: **Outputs for complex DrawBench prompts.** A confused grizzly bear in a calculus class, and a church with stained-glass windows depicting a hamburger and french fries; twelve outputs each in three rows of four. Subject appearance, scene arrangement, and background vary within each prompt, which is the variation that the set-diversity judgments assess.