JIVE: JACOBIAN-INFORMED VOLUME EXPANSION FOR DIVERSE GENERATIVE SAMPLING

Anonymous authors
Paper under double-blind review

ABSTRACT

Generative models often suffer from mode collapse and limited sample diversity. While prior works attempt to mitigate this by jointly generating a batch of samples and repelling the trajectories apart, these heuristics do not explicitly maximize the diversity of the resulting endpoints. We introduce JIVE, a training-free framework that enhances generative diversity by injecting velocity perturbations aligned with the leading right singular subspace of the generator's endpoint Jacobian. By leveraging this local geometric structure, JIVE provably maximizes endpoint diversity while preserving sample quality. To maintain practical efficiency, we compute these perturbation directions via matrix-free iterations rooted in classical numerical linear algebra, requiring only a small computational overhead. Across different benchmarks, JIVE boosts both pixel and feature-level diversity in few-step and one-step generation. [1]

1 INTRODUCTION

Diffusion and flow-based generative models are capable of producing high fidelity and prompt-aligned images (Ho et al., 2020, Song et al., 2021, Lipman et al., 2023, Liu et al., 2023). A large body of effort has been put towards improving the fidelity and speed of such models, yet much less attention is given to the *diversity* of generated outputs. For a given prompt, different noise seeds often produce images with similar composition and semantic features. Low diversity limits the model's usefulness in downstream applications such as data augmentation, content creation, and exploration. This issue is especially problematic in few-step models (Sauer et al., 2024, Black Forest Labs, 2024, Gandikota & Bau, 2026), where fewer sampling steps leave less opportunity for trajectories to diverge or for iterative diversity guidance to take effect.

To add diversity into the generation process, a natural solution is to inject a perturbation to the velocity at a sampling point. This has the effect of perturbing the trajectory and thus the endpoint at the target distribution. Most prior methods (Wu et al., 2026, Morshed & Boddeti, 2025, Corso et al., 2024) sample multiple seeds at once, and at any time point push these samples away from each other to maximize their difference. We refer to these as batch-repulsion methods. Intuitively, letting samples repulse against each other at intermediate time steps diversifies the trajectories and generation results.

In this paper, we argue that these batch-repulsion methods are only **implicitly** diversifying the endpoints of the trajectories. To maximize the diversity of endpoints, we should directly find velocity perturbations that maximally change the estimated endpoint. To find such velocity perturbations, we propose to interrogate the generator's endpoint Jacobian, which effectively measures how all local perturbations impact the estimated endpoint. In particular, we sample our perturbations from the leading right singular subspace of the endpoint Jacobian, which is the subspace spanned by the top-k right singular vectors corresponding to the top-k singular values. **Our approach finds perturbations that explicitly maximize endpoint diversity.** See Figure 1.

Our method selects perturbations based only on how they affect a single trajectory. This raises the question of how our method could guarantee overall diversity of generated results. We answer this with a theoretical result (Theorem 1) that our method increases the pairwise distance between endpoints of multiple generations, and that maximizing local endpoint change also maximizes the increase in spread among generated points.

[1] Anonymous Code is available at https://anonymous.4open.science/r/JIVE-6E50/.

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.

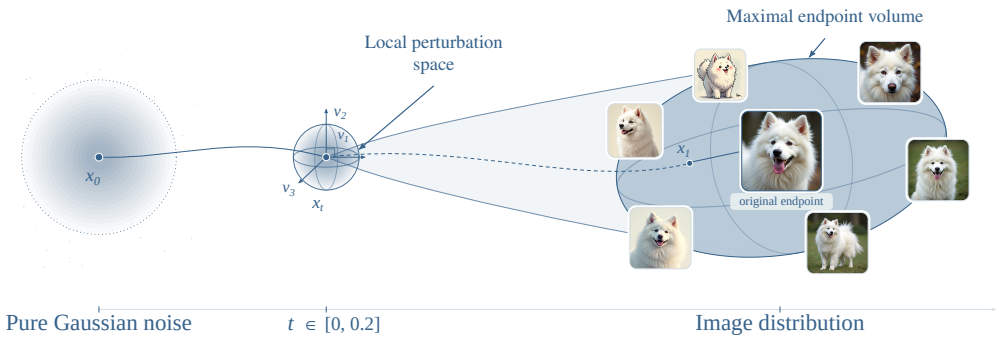


Figure 1: At an intermediate trajectory point x_t, JIVE injects a local velocity perturbation sampled from the **leading right singular subspace of the endpoint Jacobian**. This explicitly maximizes **endpoint diversity**, as theoretically guaranteed by Theorem 1.



Figure 2: **Qualitative comparison from matched initial latents.** Visual comparison of eight samples generated with 4-step FLUX.1-schnell using the same initial seeds for the prompt "***A scientist***". The samples of the original method (unmodified) have similar compositions, whereas JIVE produces distinct subjects and backgrounds.

Algorithmically, we make our method efficient by avoiding explicitly forming the full Jacobian matrix or its singular value decomposition. We instead iterate a random vector towards a top right singular vector with Jacobian and Jacobian-transpose vector products. We also show our method works when power iteration (Golub & Van Loan, 2013) is applied only a single time instead of requiring full convergence towards a top singular vector. This ensures our method is efficient in actual deployable settings.

In summary, our paper consists of the following contributions.

- We propose a novel method to inject diversity into modern flow matching generative models. Unlike previous methods, we explicitly perturb the velocity to maximize the endpoint change by sampling directions from the top singular value space of the endpoint Jacobian.
- We provide a theoretical guarantee for the diversity improvement of the endpoints.
- We provide efficient algorithms to compute the velocity perturbation even though the Jacobian itself can have millions of entries.

Across different standard benchmarks (Yu et al., 2022, Ghosh et al., 2023), JIVE consistently enhances generation diversity across multiple metrics, including Feature Vendi and Pixel Vendi. It demonstrates superior or on par performance compared with baseline flow matching models and existing state-of-the-art methods. Figure 2 illustrates the diversity introduced by our method while maintaining the original semantics from the text prompt.

2 METHOD

Preliminaries. Flow matching learns a velocity field $u_\theta(x, t)$ that transports Gaussian noise to data (Lipman et al., 2023, Liu et al., 2023). To train the model, we interpolate a noise sample x_0 and a data

sample x_1 as $x_t = (1-t)x_0 + tx_1$, and minimize the loss $\mathbb{E}_{x_0,x_1,t}\|u_\theta(x_t,t) - (x_1 - x_0)\|_2^2$. During sampling, we start from $x_0 \sim \mathcal{N}(0, I_d)$ and integrate the velocity field $\mathrm{d}x_t = u_\theta(x_t,t)\,\mathrm{d}t$ from noise at $t = 0$ to data at $t = 1$. Although sampling is iterative, we can find a quick approximation of a trajectory's endpoint using the *one-step predictor* $F_t(x) = x + (1-t)u_\theta(x,t)$. We use this predictor to analyze local generator behavior in the following sections.

2.1 Motivation: independent direction selection

JIVE finds perturbation directions that change a trajectory's endpoint the most. Prior methods mostly achieve this goal by sampling a batch of seeds together and pushing their trajectories apart during sampling (Wu et al., 2026, Morshed & Boddeti, 2025). For example, OSCAR (Wu et al., 2026) maximizes the volume spanned by trajectory endpoints. We demonstrate with our toy model in Figure 3 how this strategy might fail to maximally spread endpoints.

The toy example in Figure 3 consists of sixteen Gaussian modes whose centers are placed on the boundary of a ring. These modes serve as the target distribution of a flow matching model. We sample initial points from a Gaussian in the center of the plane. We equip the example with the velocity field that minimizes the flow matching loss. In this example, the field points toward the closest Gaussian mode at any location in this plane. We initialize eight starting points near the center and have them travel along the velocity field. Using the original flow matching velocity field, many starting points end up in the same mode due to their similar initial positions (Figure 3(a)), showing low diversity.

To perturb points towards new modes we can push them using directions that rotate the point along the ring. This changes the point's closest Gaussian mode and redirects the trajectory towards a different endpoint. For this example, OSCAR's volume maximizing objective (Figure 3(b)) generates perturbation directions (red arrows) that point directly towards the outer circle, and remain the same across all update steps. This is because the volume gradient mainly points outwards from the mean of the batch of points.

We argue that useful perturbation directions can be derived using local geometry. The endpoint predictor's Jacobian provides a tool for analyzing local geometry since it tells how local input directions are transformed into changes in the output. For example, in Figure 3 the rotational direction is amplified by the Jacobian matrix. By multiplying a random vector with the Jacobian, the rotational component of the random vector survives while other components become discarded. JIVE uses this idea to select directions independently for each sample. Figure 3(c) shows how JIVE is able to place most points into new modes and how perturbations are along rotational components. The next section formalizes our direction selection criterion and describes how JIVE estimates the perturbation subspace.

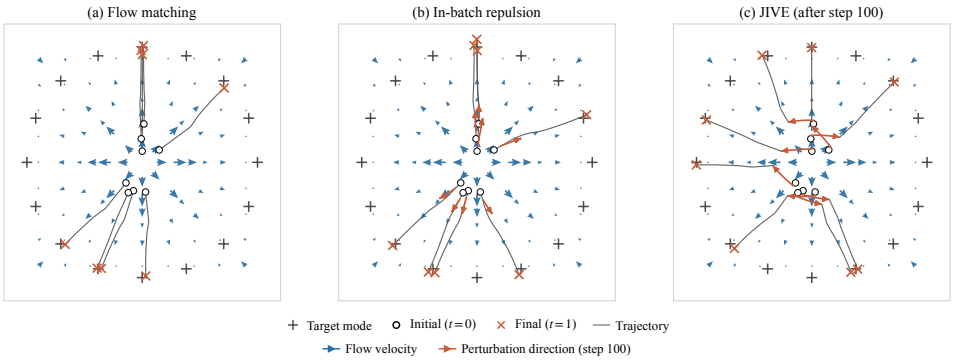


Figure 3: In a toy example, we compare unmodified flow matching (Fig.(a)), a batch repulsion method (OSCAR) (Fig.(b)) and JIVE (Fig.(c)). Eight samples flow from the center toward sixteen Gaussian modes arranged on a ring. Blue arrows are the velocity field, orange arrows are velocity perturbations, and gray curves are the trajectories of the samples. In panel (a), many points flow to the same few modes. In panel (b), batch repulsion pushes trajectories further apart, but the repulsion direction does not maximize endpoint spread. In panel (c), JIVE finds perturbations maximizing endpoint diversity and pushes points towards new modes. See Appendix A.1 for details.

2.2 ALGORITHM

We seek velocity perturbations that produce the greatest change in the endpoint of a sampling trajectory. We therefore use the *leading right singular subspace* of the endpoint predictor Jacobian $J_t = \nabla F_t(x_t)$, i.e., the subspace spanned by the right singular vectors with the k largest singular values. Perturbation vectors sampled from such a subspace cause the *greatest endpoint change* to the generator. Formally, let $Q \in \mathbb{R}^{d\times k}$ be an orthonormal basis for the leading right singular subspace at a trajectory point x_t. We sample a perturbation δ uniformly from the sphere of radius ε in $\mathrm{span}(Q)$, and resume sampling from $x_t + \delta$. In practice, we only need to perturb the trajectory once.

To pick an appropriate norm of the velocity perturbation, we use the heuristic that the distribution of $x_t + \delta$ should not be too different from the distribution of x_t. We choose a bound for the norm of δ. For $x_t \sim \mathcal{N}(0, I_d)$, the norm of x_t is almost always near $\sqrt{d}$. The standard deviation of the norm is about $1/\sqrt{2}$. We want the norm of $x_t + \delta$ to stay near $\sqrt{d}$. We found that δ and x_t are usually nearly orthogonal to each other, so we can write $\|x_t + \delta\|_2^2 = \|x_t\|_2^2 + \|\delta\|_2^2$. We try to not have $\|x_t + \delta\|_2 = \sqrt{\|x_t\|_2^2 + \|\delta\|_2^2}$ be too much greater than $\sqrt{d} + c/\sqrt{2}$ where c is some constant representing the number of standard deviations by which we allow the norm to deviate. On Flux.1-schnell which has latent dimension $64 \times 64 \times 16$ at 512×512 resolution and $c = 2$, the bound for the norm with our heuristic is 26.

2.3 THEORETICAL GUARANTEE OF GAINED DIVERSITY

The preceding analysis selects directions using only the local endpoint change of individual trajectories. We now show how local perturbations affect global spread among many independently generated outputs, measured by expected pairwise squared endpoint distance. Perturbations with zero mean conditional on the latent add covariance to each linearized endpoint, yielding the following connection to our local objective.

Theorem 1 (Perturbations increase expected pairwise distance)**.** *Let X be a random latent, $J(X) = \nabla F_t(X)$, and $Q(X)$ have k orthonormal columns. Conditional on X, draw Z uniformly on the unit sphere in $\mathbb{R}^k$ and set $\delta = \varepsilon Q(X)Z$ so that $\|\delta\|_2 = \varepsilon$. Define $Y = F_t(X)$ and $\widetilde{Y} = Y + J(X)\delta$. The perturbed endpoints have expected squared pairwise distance*

$$\mathbb{E}_{X_1,X_2}\|\widetilde{Y}_1 - \widetilde{Y}_2\|_2^2 = \mathbb{E}_{X_1,X_2}\|Y_1 - Y_2\|_2^2 + \xi_{\epsilon,k}, \textit{where}$$

$$\xi_{\epsilon,k} = \frac{2\varepsilon^2}{k}\,\mathbb{E}_X\,\mathrm{Tr}\big[Q(X)^\top J(X)^\top J(X) Q(X)\big]\,. \tag{1}$$

This theorem means that the proposed velocity perturbation will increase the expected spread of linearized endpoints by $\xi_{\epsilon,k}$. Note the gap $\xi_{\epsilon,k}$ is controlled by the perturbation magnitude ϵ and increases quadratically with it.

Figure 4 depicts a concrete picture of how perturbations can increase pairwise distances. The trace term of Theorem 1 connects directly to our endpoint displacement objective. Consider the value $\|F_t(x_t + \delta) - F_t(x_t)\|_2^2$, which captures how far a perturbation δ causes the endpoint to change. If we uniformly draw perturbations δ with norm ε from a k dimensional subspace Q, under a linear approximation this term becomes

$$\mathbb{E}_\delta\|F_t(x_t + \delta) - F_t(x_t)\|_2^2 \approx \mathbb{E}_\delta\|J_t\delta\|_2^2 = \frac{\varepsilon^2}{k}\,\mathrm{Tr}(Q^\top J_t^\top J_t Q) \tag{2}$$

The same trace term appears in the equation as in Theorem 1. Due to the expectation term of Equation (1), optimizing Equation (2) per individual x_t also optimizes overall spread increase. The top right singular subspace maximizes Equation (2) due to its trace-maximization property.

2.4 EFFICIENT COMPUTATION OF VELOCITY PERTURBATIONS

In this section, we describe the computation of the velocity perturbation. Recall we need to sample a random direction from the top-k leading right singular subspace of the endpoint Jacobian. Direct computation of the Jacobian is infeasible. Instead, we propose an efficient algorithm inspired by the classic power method from numerical linear algebra (Golub & Van Loan, 2013). At a latent trajectory

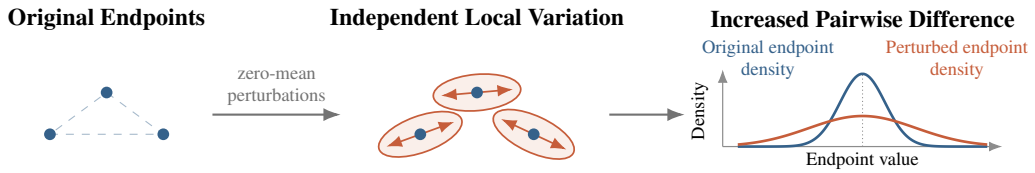


Figure 4: **From local variation to set-level diversity.** Three endpoints (blue) are perturbed within centered neighborhoods (orange). The rightmost curves show the original and perturbed endpoint densities. They have the same endpoint mean, but the perturbed density is broader, which increases the expected pairwise spread.

state x_t, we directly sample the velocity perturbation without explicitly computing the Jacobian $J_t = \nabla F_t(x_t)$, in which $F_t(x_t) = x + (1-t)u_\theta(x_t, t)$ is the one-step predictor of the endpoint.

Velocity Perturbation Computation via Power Method. To sample a velocity perturbation, we draw an isotropic random vector w and multiply it repeatedly by the matrix $J_t^\top J_t$. Take its singular value decomposition $J_t = U\Sigma V^\top = \sum_{i=1}^d \sigma_i u_i v_i^\top$. We can decompose w into the basis of v_i as $w = \sum_{i=1}^d \langle w, v_i\rangle v_i$. Since $J_t w = \sum_{i=1}^d \sigma_i \langle w, v_i\rangle u_i$, multiplying by J_t scales each component of w along the right singular vector v_i by σ_i and maps it onto the left singular vector u_i. The subsequent product with $J_t^\top$ then pulls this response back into the input space, mapping the components back onto the right singular vectors via $J_t^\top J_t w = \sum_{i=1}^d \sigma_i^2 \langle w, v_i\rangle v_i$. Consequently, inputs aligned with the leading right singular vectors experience quadratic amplification (σ_i^2) along their original directions.

We avoid explicitly forming $J_t^\top J_t$ by relying on matrix-free operations. Specifically, we approximate the Jacobian-vector product $J_t w$ via finite differences, $J_t w \approx [F_t(x_t + \epsilon w) - F_t(x_t)]/\epsilon$, which requires only two forward network evaluations. For the transpose product $J_t^\top(\cdot)$, we compute an exact vector-Jacobian product using reverse-mode automatic differentiation (Baydin et al., 2018). When computing the leading k-dimensional subspace ($k > 1$), we employ block power iteration over these alternating products. Once converged, the resulting singular vectors form an orthonormal basis $Q \in \mathbb{R}^{d\times k}$. We then sample a random unit vector $r \sim \mathbb{S}^{k-1}$, construct the perturbation $\delta = \varepsilon Q r$, and resume the integration trajectory from the perturbed state $x_t + \delta$.

One-Iteration Approximation of Block Power Iteration. Computing the full leading singular subspace requires repeated $J_t^\top J_t$ evaluations, which introduces significant computational overhead. To mitigate this cost, we consider two lower-budget approximations for determining the perturbation direction. In both variants, we draw a single isotropic probe w and transform it into a candidate direction δ. We then resume sampling from the perturbed state $x_t + s\delta$, where $s \in \{-1, +1\}$ is a Rademacher sign variable sampled with equal probability.

Solution 1: $J_t^\top J_t$ Product. We form $\delta = \varepsilon J_t^\top J_t w / \|J_t^\top J_t w\|_2$, which amounts to a single $J_t^\top J_t$ product iteration. If $w = \sum_i a_i v_i$, then $J_t^\top J_t w = \sum_i \sigma_i^2 a_i v_i$ so that the factor σ_i^2 amplifies components along high gain right singular directions.

Solution 2: Jacobian Vector Product (JVP). We form $\delta = \varepsilon J_t w / \|J_t w\|_2$ by estimating $J_t w$ with finite differences. This solution is much cheaper as it only costs two forward evaluations to compute.

As shown in Table 1, a single JVP is highly efficient, requiring roughly twice the runtime of a standard forward pass. While evaluating $J_t^\top J_t$ doubles this cost, empirical evaluations (Appendix Table 5) demonstrate its substantially superior accuracy in approximating the top-k singular subspace. Nonetheless, JVP consistently outperforms isotropic random velocity perturbations. In practice, $J_t^\top J_t$ delivers the most reliable performance, while JVP provides an effective, low-cost alternative.

3 Experiments

We evaluate JIVE's diversity gains in both its full and low-cost settings, test its combination with an existing method, and apply it to inversion-based generation. We apply our intervention at initial time $t = 0$ in all experiments. All finite difference estimates use probe perturbations with norm 4 unless otherwise specified.

Table 1: Computational efficiency of the proposed methods ($J_t^\top J_t$ vs. JVP) with one matrix iteration. This is evaluated on FLUX.1-schnell at 512×512 resolution on an A100.

Methods	Seconds	Time×	FLOPs×	Mem. (GB)
One forward pass	0.307	1.00	1.00	22.3
JVP	0.617	2.01	2.00	22.3
$J_t^\top J_t$	1.302	4.24	5.04	35.4

3.1 Benchmark diversity

We first evaluate whether JIVE increases diversity across prompts and complements existing diversity guidance. We compare JIVE with OSCAR (Wu et al., 2026), which repels samples within a batch, and CADS (Sadat et al., 2024), which adds noise to text conditioning and gradually reduces the noise level during sampling.

Metrics. We measure diversity using the Vendi score (Friedman & Dieng, 2023) in both DINOv2 (Oquab et al., 2024) feature space (F-Vendi) and pixel space (P-Vendi), together with mean pairwise L_2 distance in pixel space. We evaluate prompt alignment with CLIPScore (Hessel et al., 2021), image quality with CLIP-IQA (Wang et al., 2023), and human preference alignment with HPSv2 (Wu et al., 2023).

Table 2: **Diversity and plug-in composition on PartiPrompts.** We evaluate both JIVE(JVP) and JIVE($J_t^\top J_t$) on FLUX.1-schnell at four evaluation steps. We use prompts from PartiPrompts at 16 images per prompt. Reported values are mean±SE over prompts. The norms used are 12 on both JIVE variants when alone, and 4 when combined with CADS ($\tau_2 = 1.2$).

Method	F-Vendi↑	P-Vendi↑	Pairwise L_2 ↑	CLIP↑	CLIP-IQA↑	HPSv2↑
Standard ($n = 504$)						
Unmodified sampler	3.76±0.10	5.00±0.05	323.69±2.80	30.25±0.13	0.668±0.003	28.68±0.14
OSCAR	3.80±0.10	4.98±0.05	327.43±2.81	30.27±0.13	0.664±0.003	28.66±0.14
JIVE(JVP)	4.73±0.11	5.49±0.03	455.83±2.12	30.00±0.13	0.608±0.003	25.98±0.14
JIVE($J_t^\top J_t$)	5.14±0.11	**5.60±0.02**	**458.20±1.75**	29.83±0.12	0.602±0.003	25.62±0.13
CADS (τ_2=1.2)	6.10±0.16	5.24±0.04	344.42±2.14	28.96±0.15	0.678±0.003	27.70±0.15
JIVE(JVP)+CADS	6.31±0.16	5.54±0.04	391.89±2.14	28.97±0.15	0.661±0.003	26.87±0.15
JIVE($J_t^\top J_t$)+CADS	**6.37±0.16**	5.55±0.03	385.30±1.91	28.91±0.15	0.663±0.003	26.92±0.15
Intermediate ($n = 518$)						
Unmodified sampler	2.95±0.06	4.65±0.04	294.19±2.64	32.88±0.13	0.679±0.003	30.07±0.15
OSCAR	2.96±0.06	4.63±0.04	297.08±2.67	32.93±0.13	0.675±0.003	30.10±0.14
JIVE(JVP)	3.61±0.07	5.27±0.03	412.30±2.67	32.33±0.13	0.631±0.003	27.66±0.16
JIVE($J_t^\top J_t$)	3.99±0.07	**5.47±0.03**	**430.17±2.01**	32.09±0.12	0.620±0.004	27.28±0.14
CADS (τ_2=1.2)	4.32±0.08	5.11±0.04	330.63±2.31	31.72±0.13	0.687±0.003	29.29±0.14
JIVE(JVP)+CADS	4.42±0.08	5.33±0.04	365.81±2.50	31.64±0.13	0.674±0.003	28.64±0.14
JIVE($J_t^\top J_t$)+CADS	**4.46±0.08**	5.41±0.04	369.10±2.31	31.56±0.13	0.676±0.003	28.72±0.14
Challenging ($n = 610$)						
Unmodified sampler	3.03±0.06	4.74±0.04	311.35±2.51	33.22±0.13	0.665±0.004	30.14±0.14
OSCAR	3.06±0.06	4.72±0.04	314.71±2.53	33.27±0.13	0.660±0.004	30.15±0.14
JIVE(JVP)	3.78±0.07	5.35±0.03	427.71±2.47	32.74±0.13	0.622±0.004	27.90±0.15
JIVE($J_t^\top J_t$)	4.14±0.07	**5.52±0.03**	**437.78±1.98**	32.59±0.13	0.614±0.004	27.55±0.14
CADS (τ_2=1.2)	4.88±0.09	5.23±0.04	339.85±2.17	31.66±0.13	0.673±0.003	28.79±0.14
JIVE(JVP)+CADS	4.92±0.09	5.44±0.03	380.16±2.23	31.64±0.12	0.659±0.003	28.13±0.14
JIVE($J_t^\top J_t$)+CADS	**5.01±0.09**	5.52±0.03	378.32±2.01	31.62±0.12	0.663±0.003	28.21±0.14
All ($n = 1632$)						
Unmodified sampler	3.23±0.04	4.79±0.03	309.72±1.55	32.19±0.08	0.671±0.002	29.66±0.08
OSCAR	3.26±0.04	4.77±0.03	313.05±1.56	32.23±0.08	0.666±0.002	29.67±0.08
JIVE(JVP)	4.02±0.05	5.37±0.02	431.50±1.48	31.76±0.08	0.621±0.002	27.23±0.09
JIVE($J_t^\top J_t$)	4.40±0.05	**5.53±0.01**	**441.67±1.15**	31.58±0.08	0.612±0.002	26.87±0.08
CADS (τ_2=1.2)	5.08±0.07	5.20±0.02	338.34±1.28	30.84±0.08	0.679±0.002	28.61±0.09
JIVE(JVP)+CADS	5.19±0.07	5.44±0.02	379.23±1.35	30.82±0.08	0.664±0.002	27.90±0.09
JIVE($J_t^\top J_t$)+CADS	**5.26±0.07**	5.49±0.02	377.55±1.22	30.76±0.08	0.667±0.002	27.97±0.08

Setup. We generate 16 images for each of the 1,632 PartiPrompts prompts (Yu et al., 2022) using four-step FLUX.1-schnell (Black Forest Labs, 2024), with matched initial latents across methods. Both JIVE(JVP) and JIVE($J_t^\top J_t$) estimate a four-dimensional subspace using 10 iterations, and apply a single perturbation of norm 12 before sampling. Table 2 reports the effectiveness of our method across different prompt difficulties, along with their aggregate score.

Observation 1: JIVE improves diversity over OSCAR. Both JIVE variants improve F-Vendi, P-Vendi, and pairwise L_2 across every prompt difficulty level, with modestly lower text alignment and a larger cost in image quality and preference. Relative to CADS, standalone JIVE yields greater pixel-space diversity but lower Feature-space Vendi, motivating their combination which we explore in the JIVE+CADS rows.

Observation 2: JIVE adds diversity to CADS. Adding JIVE to CADS with a smaller perturbation norm of 4 improves all three diversity measures over CADS in every band, with small decreases in quality and alignment. This shows that JIVE composes effectively with existing diversity guidance and can serve as a flexible plugin method.

3.2 Low cost direction estimation

One-step generation has become increasingly important for practical generative models (Geng et al., 2025, Zhang et al., 2026a, Geng et al., 2026). However, this efficiency often comes at the cost of diversity (Zhang et al., 2026b). We therefore evaluate JIVE on one-step FLUX.1-schnell using a single estimated direction and a single estimation iteration. We also test this minimal budget variant on four-step SD3.5-Large-Turbo (Sauer et al., 2024, Esser et al., 2024), as an additional model family for comparison. Figure 5 compares the diversity-fidelity tradeoff curves of both JIVE variants, CADS, and OSCAR, and Table 3 reports representative norms from these curves. Both JVP and $J_t^\top J_t$ variants use a finite difference for the Jacobian-vector product. For SD3.5-Large-Turbo, we use a finite difference probe norm of 10 due to the higher latent space dimension.

We evaluate on 553 GenEval prompts (Ghosh et al., 2023), generating four images per prompt. We report additional fidelity metrics which are GenEval score for compositional prompt adherence, and PickScore (Kirstain et al., 2023) for human preference alignment. Appendix A.7 provides the corresponding tradeoff curves for GenEval and PickScore.

Observation 3: $J_t^\top J_t$ gives a stronger tradeoff than JVP. Across both models, $J_t^\top J_t$ generally achieves higher Vendi-DINO at comparable CLIPScore or HPSv2. This is consistent with $J_t^\top J_t$ placing more perturbation energy in high-gain right singular directions, thereby producing greater endpoint diversity for a similar quality cost. The JVP variant remains a useful cheaper alternative for finding a perturbation direction.

Observation 4: JIVE remains effective in one-step generation. On one-step FLUX.1-schnell, JIVE($J_t^\top J_t$) achieves a more favorable diversity-CLIPScore and GenEval tradeoff than CADS. CADS retains a stronger HPSv2 tradeoff and remains competitive under PickScore, showing that the relative

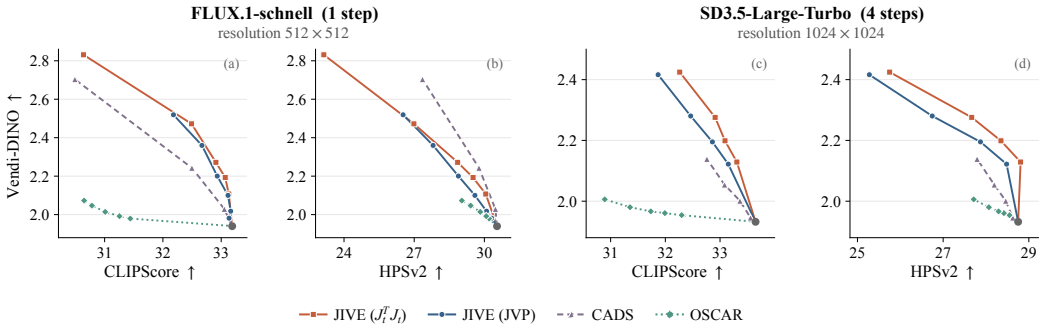


Figure 5: **Diversity-fidelity tradeoffs on GenEval.** We show tradeoff curves for CLIPScore and HPSv2 versus Vendi-DINO, averaged over all 553 GenEval prompts at 4 images per prompt. Additional tradeoff curves for PickScore and GenEval score are in Appendix Figure 9.

Table 3: Diversity-fidelity comparison on GenEval (553 prompts, 4 images per prompt). Both JIVE variants use one direction-estimation iteration with finite differences. JVP/$J_t^\top J_t$ use norms 26/8 on FLUX and 26/16 on Turbo, respectively. Bold indicates the highest diversity score within each model.

Method	F-Vendi ↑	P-Vendi ↑	CLIP ↑	HPSv2 ↑	PickScore ↑	GenEval ↑
FLUX.1-schnell (1 step, 512 × 512)						
Unmodified sampler	1.940±0.019	1.683±0.013	33.189±0.138	30.540±0.109	23.330±0.044	0.700
CADS (τ_2=1.2)	2.243±0.024	1.779±0.013	32.497±0.134	29.770±0.120	23.026±0.043	0.583
OSCAR	2.073±0.021	1.680±0.013	30.648±0.113	29.037±0.118	22.776±0.045	0.667
Ours (JVP, FD)	2.200±0.021	1.964±0.018	32.932±0.132	28.890±0.119	22.945±0.044	0.684
Ours ($J_t^\top J_t$, FD)	**2.272**±0.023	**2.080**±0.016	32.909±0.132	28.858±0.115	22.910±0.043	0.672
SD3.5-Large-Turbo (4 steps, 1024 × 1024)						
Unmodified sampler	1.932±0.021	1.369±0.007	33.650±0.127	28.751±0.116	23.112±0.043	0.708
CADS (τ_2=1.1)	2.139±0.025	1.370±0.007	32.760±0.114	27.793±0.119	22.824±0.041	0.582
OSCAR	2.006±0.022	1.367±0.008	30.892±0.111	27.713±0.121	22.783±0.042	0.701
Ours (JVP, FD)	2.195±0.023	1.964±0.015	32.860±0.125	27.871±0.119	22.677±0.043	0.657
Ours ($J_t^\top J_t$, FD)	**2.276**±0.023	**1.977**±0.013	32.910±0.126	27.667±0.121	22.638±0.041	0.674

advantage depends on the quality metric. Both JIVE variants offer stronger diversity-quality tradeoffs than OSCAR over the evaluated settings.

Observation 5: JIVE transfers effectively to SD3.5-Large-Turbo. On four-step SD3.5-Large-Turbo, JIVE($J_t^\top J_t$) exhibits the strongest observed diversity-quality tradeoff across CLIPScore, HPSv2, PickScore, and GenEval. The JVP variant also generally improves on CADS and OSCAR while requiring less computation for direction estimation. Compared with FLUX, the advantage over CADS is more consistent across quality metrics, supporting the effectiveness of JIVE's minimal estimation budget on a second generator.

3.3 Improving diversity in inversion-based generation

Many image-generation tasks start from an existing image rather than generating from scratch. The goal is to produce new images that remain related to the source while allowing meaningful variations. Diversity is important in this setting because a single source can admit many plausible outputs, and generating multiple alternatives gives users or downstream systems more choices. We consider inversion-based generation (Meng et al., 2022, Mokady et al., 2023, Luzi et al., 2024), where the starting latent is produced by inverting a source image rather than drawing from Gaussian noise. We study this setting by combining JIVE with RF-Inversion (Rout et al., 2025) on stroke-to-image tasks. Given a paint-stroke image, RF-Inversion inverts a controlled ODE to recover a starting latent, then integrates forward to generate an image. During denoising, we inject JIVE($J_t^\top J_t$) after the first reverse-step controller update ($\eta_t = 0$). On FLUX.1-dev, we use 32 source images for paint strokes and generate 16 outputs per source for various different scenes from the LSUN dataset (Yu et al., 2015). Table 4 reports diversity and fidelity and Figure 6 illustrates the workflow.

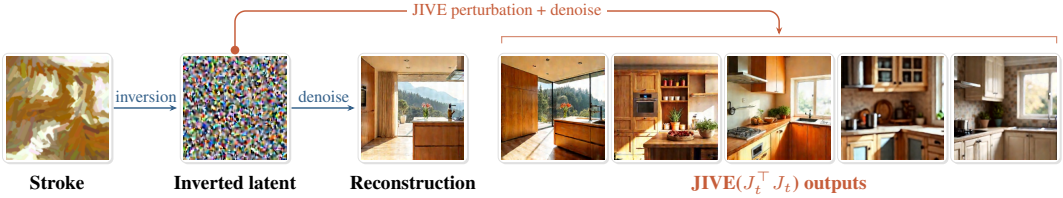


Figure 6: **Diverse generation from inverted latents.** Controlled ODE inversion maps a stroke image to a latent. Denoising this latent returns a single reconstruction, while applying a JIVE($J^\top J$) perturbation after the first denoising step yields distinct outputs from the same latent.

Table 4: **Diverse edits from inverted latents.** RF-Inversion on FLUX.1-dev (30 steps), with JIVE($J_t^\top J_t$) injected after the first reverse-step controller update ($\eta_t = 0$, norm 8). F-Vendi uses DINOv2-base features while P-Vendi uses pixels. Source L_2 is measured relative to the source stroke image. Values are mean±SE over 32 sources. Orange denotes JIVE.

Dataset	Method	Controller	Norm	F-Vendi↑	P-Vendi↑	LPIPS$_d$↑	Source L_2 ↓	DINO$_f$↑	KID×1k↓
classroom	baseline	$\eta=0.5$	N/A	4.67±0.17	2.82±0.06	0.684±0.002	121.68±1.34	0.051±0.002	53.95
	JIVE($J_t^\top J_t$)	$\eta=0.5$	8	**5.37**±0.16	**2.98**±0.06	**0.716**±0.002	126.12±1.34	0.057±0.002	54.90
kitchen	baseline	$\eta=0.5$	N/A	6.78±0.18	2.33±0.07	0.680±0.003	114.09±0.96	0.018±0.003	40.62
	JIVE($J_t^\top J_t$)	$\eta=0.5$	8	**7.40**±0.16	**2.46**±0.07	**0.723**±0.002	121.47±1.02	0.023±0.003	39.99
conference room	baseline	$\eta=0.5$	N/A	4.99±0.16	2.69±0.11	0.654±0.002	119.87±1.60	0.085±0.004	35.12
	JIVE($J_t^\top J_t$)	$\eta=0.5$	8	**5.80**±0.18	**2.88**±0.12	**0.696**±0.003	126.15±1.67	0.086±0.003	25.50
dining room	baseline	$\eta=0.5$	N/A	6.92±0.14	2.62±0.09	0.652±0.003	118.00±1.15	0.065±0.003	11.49
	JIVE($J_t^\top J_t$)	$\eta=0.5$	8	**7.93**±0.12	**2.80**±0.09	**0.700**±0.003	124.32±1.16	0.066±0.003	9.52
restaurant	baseline	$\eta=0.5$	N/A	6.78±0.25	4.01±0.15	0.637±0.004	120.33±1.91	0.048±0.004	44.59
	JIVE($J_t^\top J_t$)	$\eta=0.5$	8	**7.48**±0.24	**4.15**±0.14	**0.681**±0.003	124.36±2.06	0.047±0.004	39.23

Observation 6: diversity gains after inversion often accompany improved quality. JIVE($J_t^\top J_t$) improves F-Vendi, P-Vendi, and LPIPS (Zhang et al., 2018) diversity across all five datasets. KID generally improves on these datasets. These results suggest that JIVE explores meaningful alternative reconstructions rather than introducing arbitrary distortion. The outputs move farther from the source and become more diverse while generally preserving or improving quality.

4 Related Work

Inference-time diversity. OSCAR, DiverseFlow, and Particle Guidance promote diversity through repulsive interactions among sampling trajectories (Wu et al., 2026, Morshed & Boddeti, 2025, Corso et al., 2024). Their diversity updates depend on other samples, whereas JIVE selects each perturbation independently using the generator's Jacobian. Noise optimization (Harrington et al., 2026) iteratively optimizes initial latents using output-diversity objectives and quality rewards. It supports both joint batch optimization and sequential generation, where each new output is optimized relative to previous outputs. JIVE's cheapest variant constructs a direction with a single Jacobian product and requires no auxiliary feature or reward model.

CADS (Sadat et al., 2024) adds noise to text conditioning and gradually anneals it away during sampling. JIVE instead perturbs the sampling latent and can be combined with CADS, as demonstrated in our experiments. STRIDE (Yadav et al., 2026) uses activation PCA to perturb intermediate transformer features at architecture-specific injection layers. JIVE selects latent perturbations through the endpoint-predictor Jacobian without modifying internal features or selecting injection layers. LOCO Edit (Chen et al., 2024) uses leading Jacobian singular directions for localized image editing. By contrast, our goal is to use the Jacobian to find diversity-inducing velocity perturbations.

5 Conclusion

JIVE uses the generator's Jacobian to identify velocity perturbations that alter the trajectory's endpoint. Our analysis connects our local perturbation criterion to greater expected spread among independently generated outputs. Across the evaluated generators, JIVE improves diversity and remains effective in low cost settings. Its application to inversion-based generation further demonstrates its usefulness in applications outside of text-to-image generation.

Limitations and future work Although we provide a theoretical bound on the perturbation norm for latent distributions and test its effectiveness in our experiment, a general norm-selection principle for latents remains open. Developing such a principle is an important direction for future work. Generation diversity also has applications in robotics, particularly generating diverse action sequences and candidate motion plans for downstream selection, and in other domains such as data augmentation.

AI USE STATEMENT

In this work, we used generative AI tools for editing, to generate synthetic prompts used only in Appendix A.9, and to assist in implementing methods and writing proofs. We did not use AI tools to help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, propose or refine hypotheses, design or provide feedback on research methodology or experiments, assist with translation, clean and reformat datasets, support qualitative and thematic data analysis, or interpret results. The authors reviewed the AI-assisted content and take responsibility for the final text, claims, and reported results.

REPRODUCIBILITY STATEMENT

We provide the algorithmic and theoretical details of JIVE in the main paper, with additional derivations, hyperparameters, and ablations in the Appendix. We report the models, datasets, evaluation metrics, and sampling configurations used in our experiments. Code and scripts for reproducing our experiments are available in our anonymous repository.

REFERENCES

Atilim Gunes Baydin, Barak A Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. Automatic differentiation in machine learning: a survey. *Journal of machine learning research*, 18 (153):1–43, 2018.

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

Siyi Chen, Huijie Zhang, Minzhe Guo, Yifu Lu, Peng Wang, and Qing Qu. Exploring low-dimensional subspace in diffusion models for controllable image editing. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL `https://openreview.net/forum?id=50aOEfb2km`.

Gabriele Corso, Yilun Xu, Valentin De Bortoli, Regina Barzilay, and Tommi S. Jaakkola. Particle guidance: non-i.i.d. diverse sampling with diffusion models. In *The Twelfth International Conference on Learning Representations*, 2024. URL `https://openreview.net/forum?id=KqbCvIFBY7`.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, and Robin Rombach. Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first international conference on machine learning*, 2024.

Dan Friedman and Adji Bousso Dieng. The vendi score: A diversity evaluation metric for machine learning. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL `https://openreview.net/forum?id=g97OHbQyk1`.

Rohit Gandikota and David Bau. Distilling diversity and control in diffusion models. In *2026 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1304–1313. IEEE, 2026.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, J Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL `https://openreview.net/forum?id=uWj4s7rMnR`.

Zhengyang Geng, Yiyang Lu, Zongze Wu, Eli Shechtman, J Zico Kolter, and Kaiming He. Improved mean flows: On the challenges of fastforward generative models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 30467–30476, 2026.

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

Gene H Golub and Charles F Van Loan. *Matrix computations*, volume 4. JHU press, 2013.

Anne Harrington, A. Sophia Koepke, Shyamgopal Karthik, Trevor Darrell, and Alexei A Efros. It's never too late: Noise optimization for collapse recovery in trained diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 43124–43134, 2026.

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

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

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

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=XVjTT1nw5z.

Lorenzo Luzi, Paul M Mayer, Josue Casco-Rodriguez, Ali Siahkoohi, and Richard Baraniuk. Boomerang: Local sampling on image manifolds using diffusion models. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=NYdThkjNW1.

Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. SDEdit: Guided image synthesis and editing with stochastic differential equations. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=aBsCjcPu_tE.

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

Mashrur M Morshed and Vishnu Boddeti. Diverseflow: Sample-efficient diverse mode coverage in flows. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23303–23312. IEEE, 2025.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel HAZIZA, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Herve Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=a68SUt6zFt.

Litu Rout, Yujia Chen, Nataniel Ruiz, Constantine Caramanis, Sanjay Shakkottai, and Wen-Sheng Chu. Semantic image inversion and editing using rectified stochastic differential equations. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=Hu0FSOSEyS.

Seyedmorteza Sadat, Jakob Buhmann, Derek Bradley, Otmar Hilliges, and Romann M. Weber. CADS: Unleashing the diversity of diffusion models through condition-annealed sampling. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=zMoNrajk2X.

Axel Sauer, Frederic Boesel, Tim Dockhorn, Andreas Blattmann, Patrick Esser, and Robin Rombach. Fast high-resolution image synthesis with latent adversarial diffusion distillation. In *SIGGRAPH Asia 2024 Conference Papers*, pp. 1–11, 2024.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 2555–2563, 2023.

Jingxuan Wu, Zhenglin Wan, Xingrui Yu, Yuzhe Yang, Bo An, Ivor Tsang, and Yang You. Letting trajectories spread: Quality-preserving control for diverse flow matching. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=LtHNTRg73G.

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.

Ankit Yadav, Arpit Garg, Ta Duc Huy, and Lingqiao Liu. Stride: Training-free diversity guidance via pca-directed feature perturbation in single-step diffusion models. *arXiv preprint arXiv:2605.11494*, 2026.

Fisher Yu, Ari Seff, Yinda Zhang, Shuran Song, Thomas Funkhouser, and Jianxiong Xiao. Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*, 2015.

Jiahui Yu, Yuanzhong Xu, Jing Yu Koh, Thang Luong, Gunjan Baid, Zirui Wang, Vijay Vasudevan, Alexander Ku, Yinfei Yang, Burcu Karagol Ayan, Ben Hutchinson, Wei Han, Zarana Parekh, Xin Li, Han Zhang, Jason Baldridge, and Yonghui Wu. Scaling autoregressive models for content-rich text-to-image generation. *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856. URL https://openreview.net/forum?id=AFDcYJKhND. Featured Certification.

Guangxun Zhang, Mason Haberle, and Davi Geiger. Stable mean flow: Lyapunov-inspired one-step flow matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9223–9232, 2026a.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018.

Yuan Zhang, Chenyi Li, Haodong Yu, Guoqing Ma, Jiajun Zha, Yuanming Yang, Bo Wang, Wei Tang, Wenbo Li, Haoyang Huang, and Nan Duan. Teacher-feature drifting: One-step diffusion distillation with pretrained diffusion representations, 2026b. URL https://arxiv.org/abs/2605.07327.

A Appendix

A.1 Ring of Gaussians toy experiment

Toy distribution and velocity. Figure 3 starts from a centered two-dimensional Gaussian with standard deviation 0.25. The target is a mixture of sixteen equally likely Gaussian components whose centers are evenly spaced on a circle of radius 2. Each component has standard deviation 0.08. Thus, each black plus sign in the figure is the center of one target Gaussian. For every source-target pair, we use the straight path $X_t = (1-t)Z + tY$. The velocity is $v_t(x) = \mathbb{E}[Y - Z \mid X_t = x]$, which is the average displacement of all pairs whose path could pass through x at time t. Near the center, several target modes are plausible and contribute to this average. As a trajectory approaches the ring, nearby modes receive more weight and determine its final direction. The distribution is analytic, so we evaluate this velocity directly without training a network. Specifically, with $c_j = 2(\cos(2\pi j/16), \sin(2\pi j/16))^\top$, the exact field is

$$q_t = (1-t)^2(0.25)^2 + t^2(0.08)^2, \qquad a_t = \frac{t(0.08)^2 - (1-t)(0.25)^2}{q_t},$$

$$w_j(x,t) = \frac{\exp\big(-\|x - tc_j\|_2^2/(2q_t)\big)}{\sum_{\ell=0}^{15}\exp(-\|x - tc_\ell\|_2^2/(2q_t))}, \quad v_t(x) = a_t x + (1 - a_t t)\sum_{j=0}^{15} w_j(x,t)c_j.$$

Here, $w_j(x,t)$ is the weight assigned to target mode j at the current location and time.

Compared methods. All three panels start from the same eight source samples and use 800 integration steps. The left panel uses ordinary flow matching. The middle panel adds deterministic log-determinant repulsion at every step, with a force that gradually vanishes near the endpoint. This isolates the deterministic repulsive component used to motivate OSCAR (Wu et al., 2026). The right panel applies JIVE only once, after step 100. For each sample separately, JIVE finds the local direction that most strongly changes its predicted endpoint, moves the sample by 0.25 in one of the two signs along that direction, and then resumes ordinary flow matching. JIVE does not inspect or coordinate the other seven samples.

A.2 Proofs of the main results

Proof of Theorem 1. Conditional uniformity on the unit sphere gives

$$\mathbb{E}[\delta \mid X] = 0, \qquad \mathbb{E}[\delta\delta^\top \mid X] = \frac{\varepsilon^2}{k} Q(X)Q(X)^\top.$$

Write $B = J(X)\delta$. Then $\mathbb{E}[B \mid X] = 0$, so the cross-covariance of $Y = F_t(X)$ and B vanishes. Hence by the law of total covariance

$$\mathrm{Cov}(\widetilde{Y}) = \mathrm{Cov}(Y) + \frac{\varepsilon^2}{k}\mathbb{E}[J(X)Q(X)Q(X)^\top J(X)^\top].$$

For independent identically distributed vectors W_1, W_2 with finite second moments, expanding the square yields $\mathbb{E}\|W_1 - W_2\|_2^2 = 2\,\mathrm{Tr}\,\mathrm{Cov}(W_1)$. Applying this identity to $\widetilde{Y}$ and using cyclicity of the trace proves the formula. □

Proposition 1 (Optimal trace subspace)**.** *For $J_t \in \mathbb{R}^{d\times d}$ with singular values $\sigma_1 \geq \cdots \geq \sigma_d \geq 0$ and any $Q \in \mathbb{R}^{d\times k}$ with $Q^\top Q = I_k$,*

$$\mathrm{Tr}(Q^\top J_t^\top J_t Q) \leq \sum_{i=1}^{k} \sigma_i^2,$$

with equality for $Q = [v_1, \ldots, v_k]$.

Proof of Proposition 1. Let $P = QQ^\top$ and write $J_t^\top J_t = \sum_i \sigma_i^2 v_i v_i^\top$. Then

$$\mathrm{Tr}(Q^\top J_t^\top J_t Q) = \sum_i \sigma_i^2 w_i, \qquad w_i = v_i^\top P v_i \in [0,1], \quad \sum_i w_i = k.$$

The largest weighted sum assigns unit weight to the k largest squared singular values. Choosing $Q = [v_1, \ldots, v_k]$ achieves this bound. □

A.3 Direction estimation algorithm

Algorithm 1 is the concrete rank-one implementation used for rectified-flow generation.

Algorithm 1 Jacobian-guided trajectory perturbation for rectified flow

Require: Velocity field $v_\theta(x,t)$, total denoising steps N, initial noise $x_0 \sim \mathcal{N}(0,\mathbf{I})$, power iterations K_{pow}, finite-difference step ϵ, target perturbation norm γ
Ensure: Generated sample x_N

Sample $v^{(0)} \sim \mathcal{N}(0,\mathbf{I})$
for $\ell = 1$ to K_{pow} **do** ▷ Power iteration via finite differences
  $v^{(\ell)} \leftarrow \frac{[x_0+\epsilon v^{(\ell-1)}+v_\theta(x_0+\epsilon v^{(\ell-1)},0)]-[x_0+v_\theta(x_0,0)]}{\epsilon}$
  $v^{(\ell)} \leftarrow v^{(\ell)}/\|v^{(\ell)}\|_2$
end for
Sample $s \sim \text{Uniform}(\{-1,+1\})$
$x_0 \leftarrow x_0 + s\gamma\, v^{(K_{\text{pow}})}$ ▷ Add perturbation vector of norm γ
for $i = 0$ to $N-1$ **do**
  $t_i \leftarrow i/N, \quad \Delta t \leftarrow 1/N$
  $x_{i+1} \leftarrow x_i + \Delta t\, v_\theta(x_i,t_i)$
end for
return x_N

A.4 Eigenvector and singular vector alignment

In this section, we present the observation that the top eigenvector and top singular vector have high alignment in our tested models. Alignment is especially high for multi-step models.

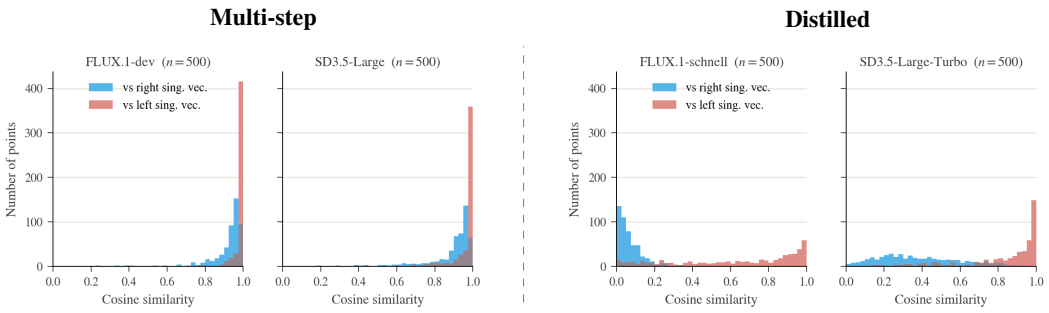


Figure 7: **Eigenvector and singular vector alignment across models.** The histograms show absolute cosine similarity of the leading eigenvector with the leading right and left singular vectors at $t = 0$ ($n = 500$ per model). Blue/red denote right/left singular vectors throughout. Alignment is strong in multi-step models (left) and weaker in distilled models (right), especially with the right singular vector.

A.5 Why random directions don't work.

We observed that random perturbation directions in general fail to meaningfully alter the final image. This can be explained by the endpoint displacement approximation $\|F_t(x_t+\delta)-F_t(x_t)\|_2^2 \approx \|J_t\delta\|_2^2$. We have that

$$\|J_t\delta\|_2^2 = \sum_i \sigma_i^2 \langle \delta, v_i\rangle^2,$$

so for a random direction δ much of its energy is placed in right singular vectors v_i with small singular values. Thus, $\|J_t\delta\|_2^2$ as a whole is small. We verify in Figure 8 that the spectrum in distilled models decays rapidly, so aligning with top singular directions results in meaningfully higher $\|J_t\delta\|_2^2$. We compare tradeoff curves of one J_t iteration with random directions in Figure 12.

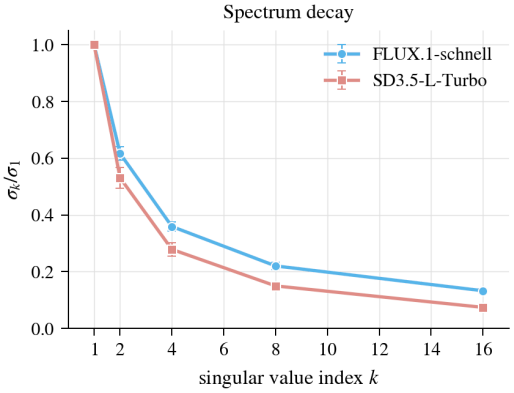


Figure 8: **The leading singular values decay rapidly.** Spectra for two distilled models, normalized by the largest singular value. The spectrum decays rapidly in both models, demonstrating that alignment with top directions allows for more endpoint displacement per unit perturbation

A.6 Runtime and Approximation Power for Perturbation Computation Methods

We compare the approximation power of the two computational methods in Section 2.4: $J_t^\top J_t$ and JVP. The approximation power is demonstrated in Table 5 and the full table of run time is in Table 6.

Based on Table 5, $J_t^\top J_t$ is our stronger low-cost estimator. One product places more than 91% of the direction's energy in the top four right singular directions in both models. Although the JVP direction may not align with a top singular vector, Table 5 shows that it captures substantially more energy in the leading singular subspace than a random direction and achieves greater endpoint diversity gain. This is our cheapest alternative.

Table 5: Energy $E_k = \sum_{i=1}^{k} |\langle \delta, v_i \rangle|^2$ in the top-k right singular subspace for unit δ, and gain relative to $\sigma_{\text{rms}} = (\sum_i \sigma_i^2/d)^{1/2}$. The random row is an analytic value (k/d). Both JVP and $J_t^\top J_t$ are evaluated with one iteration.

Model	Direction	E_1	E_2	E_4	E_8	E_{16}	$\|Jw\|_2/\sigma_{\text{rms}}$
FLUX.1-schnell	Random	1.5e-5	3.1e-5	6.1e-5	1.2e-4	2.4e-4	1.0
	JVP	0.033	0.050	0.072	0.102	0.127	30.8
	$J_t^\top J_t$	0.624	0.803	0.915	0.969	0.989	198.3
SD3.5-Large-Turbo	Random	3.8e-6	7.6e-6	1.5e-5	3.1e-5	6.1e-5	1.0
	JVP	0.074	0.108	0.152	0.226	0.277	102.7
	$J_t^\top J_t$	0.648	0.816	0.916	0.981	0.992	321.6

Table 6: Runtime and memory overhead of Jacobian-vector and $J_t^\top J_t$ products on FLUX.1-schnell at 512×512 resolution on an A100. Only the transformer is resident on the GPU. FD is finite difference.

Configuration	Seconds	Time×Base	FLOPs×Base	Peak Mem. (GB)
Baseline	0.307	1.00	1.00	22.3
JVP FD (forward)	0.617	2.01	2.00	22.3
JVP FD (central)	0.619	2.02	2.00	22.3
JVP exact	1.733	5.64	3.00	23.5
$J_t^\top J_t$ FD (forward)	1.302	4.24	5.04	35.4
$J_t^\top J_t$ exact	2.425	7.90	6.03	35.4
$J_t^\top J_t$ FD (forward) w/ grad checkpointing	1.651	5.38	6.04	23.0
$J_t^\top J_t$ exact w/ grad checkpointing	2.790	9.09	7.03	23.6

Table 7: Comparing the low-cost version of JIVE with using extra denoising steps. Experiments were conducted using GenEval prompts at 4 images per prompt.

Method	Diversity		Fidelity			
	F-Vendi ↑	P-Vendi ↑	CLIP ↑	HPSv2 ↑	PickScore ↑	GenEval ↑
FLUX.1-schnell (1 step, 512 × 512)						
Deterministic	1.940±0.019	1.683±0.013	33.189±0.138	30.540±0.109	23.330±0.044	0.700
Extra denoising steps (5)	2.028±0.021	1.710±0.013	32.934±0.134	30.483±0.109	23.249±0.043	0.686
Ours (JVP, FD)	2.200±0.021	1.964±0.018	32.932±0.132	28.890±0.119	22.945±0.044	0.684
Ours ($J_t^\top J_t$, FD)	**2.272**±0.023	**2.080**±0.016	32.909±0.132	28.858±0.115	22.910±0.043	0.672
SD3.5-Large-Turbo (4 steps, 1024 × 1024)						
Deterministic	1.932±0.021	1.369±0.007	33.650±0.127	28.751±0.116	23.112±0.043	0.708
Extra denoising steps (8)	1.998±0.022	1.485±0.010	33.434±0.130	29.465±0.119	23.230±0.043	0.694
Ours (JVP, FD)	2.195±0.023	1.964±0.015	32.860±0.125	27.871±0.119	22.677±0.043	0.657
Ours ($J_t^\top J_t$, FD)	**2.276**±0.023	**1.977**±0.013	32.910±0.126	27.667±0.121	22.638±0.041	0.674

A.7 Additional tradeoff curves

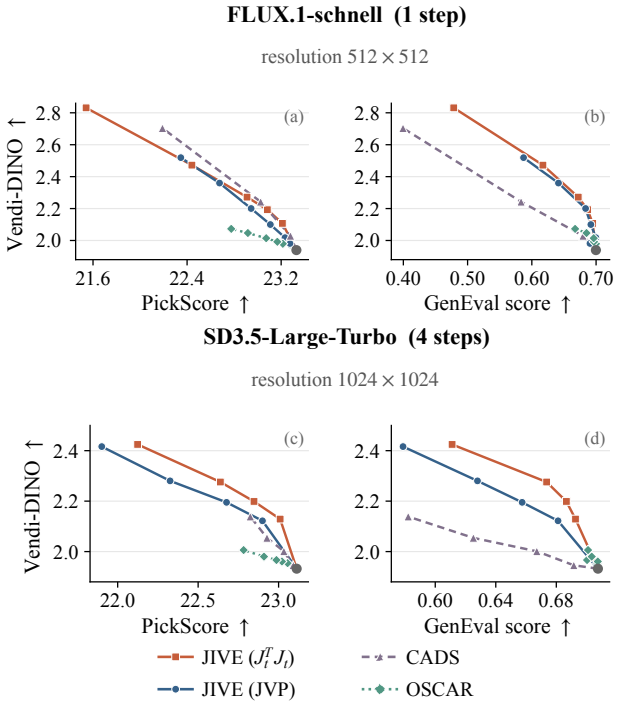


Figure 9: **PickScore and GenEval score tradeoffs on GenEval.** Companion to Figure 5 with the same settings: PickScore (left) and GenEval score (right) versus Vendi-DINO, averaged over 553 prompts. On both FLUX.1-schnell and SD3.5-Large-Turbo, both low-cost variants of JIVE have a better diversity-GenEval tradeoff than CADS or OSCAR.

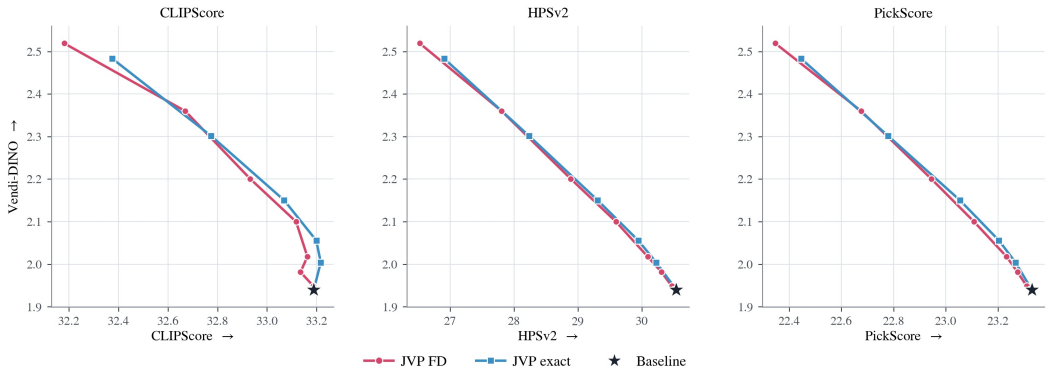


Figure 10: Finite-difference and exact JVP estimates produce similar diversity-quality frontiers. Points are evaluated on Flux.1-schnell at 1 step using GenEval prompts at 4 images per prompt

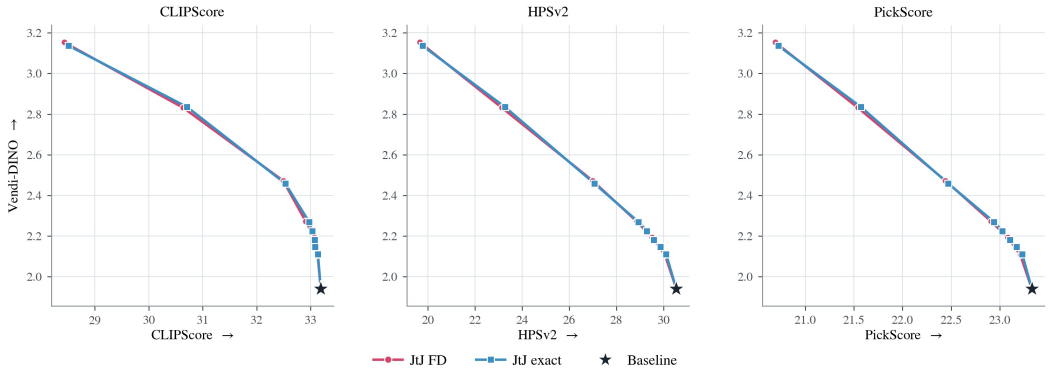


Figure 11: Finite-difference and exact $J^\top J$ estimates produce similar diversity-quality frontiers. Points are evaluated on Flux.1-schnell at 1 step using GenEval prompts at 4 images per prompt

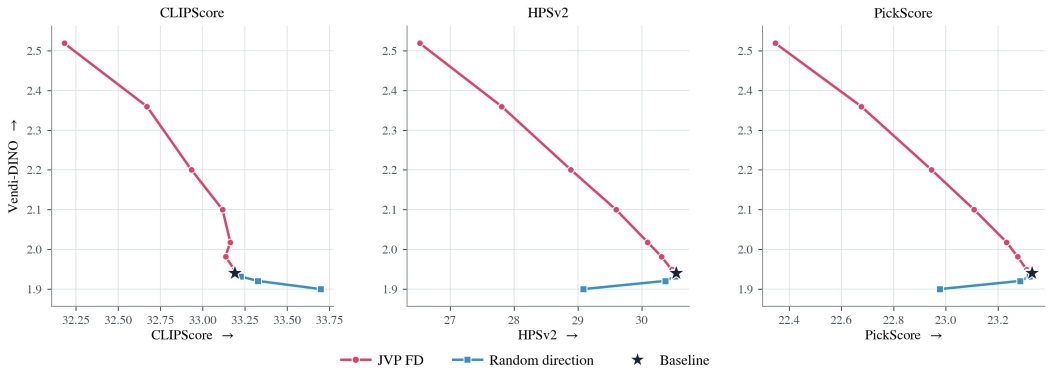


Figure 12: Jacobian-guided perturbation versus random perturbation. Random directions yield almost none of the diversity gain, so the improvement comes from the choice of direction rather than from injected norm. Points are evaluated on Flux.1-schnell at 1 step using GenEval prompts at 4 images per prompt

A.8 ADDITIONAL QUALITATIVE SAMPLES



a woman with sunglasses and red hair



a yellow wall with “Hello, world.” written on it



food



a robot

Figure 13: **Qualitative diversity across twelve prompts.** Each prompt block shows six matched samples from the unperturbed sampler in the upper row and JIVE in the lower row. Both rows use the same initial seeds. Across varied subjects and compositions, JIVE produces broader changes in appearance, viewpoint, and style while preserving the prompt content. Prompts 1-4 are shown here, and prompts 5-12 continue on the following pages.



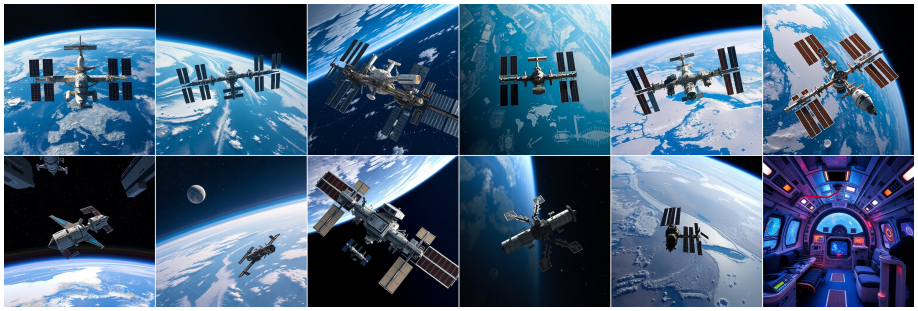
a man



a boy and a tiger



a cat



the International Space Station

Figure 13: **Qualitative diversity across twelve prompts (continued).** Prompts 5-8. Upper rows show unperturbed samples and lower rows show JIVE samples generated from the same initial seeds.



a horse in a forest



a woman



an octopus



an angry man

Figure 13: **Qualitative diversity across twelve prompts (continued).** Prompts 9-12. Upper rows show unperturbed samples and lower rows show JIVE samples generated from the same initial seeds.

A.9 CATEGORY-LEVEL RESULTS

Table 8: Category-level results on the curated prompt set. Each entry is the mean over 10 prompts and 128 images per arm. Panels (a) and (b) cover four categories each. Shading marks the primary Feature Vendi criteria. Boldface denotes the best result within each category and tier. For all metrics, higher values are better.

(a) Objects, Human/Animals, Nature, and Scenes.

		Objects			Human/Animals			Nature			Scenes		
Tier	Metric	Det.	OSCAR	JIVE	Det.	OSCAR	JIVE	Det.	OSCAR	JIVE	Det.	OSCAR	JIVE
GENERAL	Feature Vendi	3.42	3.48	**4.58**	5.10	5.17	**7.41**	3.62	3.69	**6.09**	4.25	4.36	**5.59**
	Feature Vendi (grouped)	2.85	2.89	**3.59**	3.76	3.79	**5.11**	2.94	2.99	**4.54**	3.36	3.43	**4.16**
	Pixel Vendi	**10.35**	10.29	10.29	10.75	10.68	**10.98**	5.55	5.47	**7.01**	11.42	11.37	**12.10**
	Pixel Vendi (grouped)	6.86	6.83	**7.15**	6.80	6.78	**7.34**	4.11	4.06	**5.45**	6.89	6.87	**7.59**
	Mean pairwise L_2	299.9	303.6	**428.2**	286.4	289.1	**412.7**	243.2	246.0	**394.2**	290.3	294.8	**367.7**
	CLIP score	31.83	**31.84**	31.31	32.84	**32.92**	32.38	32.25	**32.26**	30.55	31.64	**31.66**	31.28
	CLIP-IQA	**0.503**	**0.503**	0.502	**0.503**	**0.503**	0.502	**0.503**	**0.503**	0.501	**0.502**	**0.502**	**0.502**
	HPSv2	**0.298**	0.297	0.273	**0.319**	0.318	0.292	0.287	**0.288**	0.249	**0.312**	**0.312**	0.290
DETAILED	Feature Vendi	2.23	2.23	**2.73**	3.55	3.62	**4.19**	2.38	2.39	**3.37**	3.34	3.36	**3.75**
	Feature Vendi (grouped)	1.99	1.99	**2.36**	2.91	2.95	**3.34**	2.10	2.10	**2.79**	2.78	2.79	**3.06**
	Pixel Vendi	9.86	9.79	**10.14**	8.42	8.32	**9.51**	4.86	4.75	**6.60**	10.24	10.20	**10.90**
	Pixel Vendi (grouped)	6.52	6.49	**6.97**	5.58	5.53	**6.46**	3.69	3.62	**5.06**	6.32	6.30	**6.86**
	Mean pairwise L_2	275.3	278.6	**388.2**	255.2	256.9	**351.3**	228.9	231.0	**353.8**	270.5	274.4	**321.4**
	CLIP score	33.23	**33.28**	32.93	34.63	**34.65**	34.36	31.16	**31.19**	30.47	**32.21**	32.20	31.89
	CLIP-IQA	**0.503**	**0.503**	**0.503**	**0.503**	**0.503**	0.502	**0.503**	**0.503**	0.502	**0.502**	**0.502**	**0.502**
	HPSv2	**0.319**	0.318	0.303	**0.321**	0.320	0.308	0.278	**0.279**	0.253	**0.311**	**0.311**	0.298

(b) Characters, Food, Art, and Text.

		Characters			Food			Art			Text		
Tier	Metric	Det.	OSCAR	JIVE	Det.	OSCAR	JIVE	Det.	OSCAR	JIVE	Det.	OSCAR	JIVE
GENERAL	Feature Vendi	3.24	3.23	**4.22**	2.27	2.25	**3.07**	4.21	4.18	**5.37**	4.22	4.42	**7.35**
	Feature Vendi (grouped)	2.70	2.70	**3.38**	2.03	2.02	**2.62**	3.34	3.30	**4.06**	3.26	3.36	**5.12**
	Pixel Vendi	9.97	9.91	**10.75**	9.27	9.20	**10.29**	10.00	10.00	**10.92**	**11.09**	11.04	10.76
	Pixel Vendi (grouped)	6.32	6.29	**7.04**	6.17	6.14	**6.99**	6.24	6.24	**7.04**	6.90	6.88	**7.22**
	Mean pairwise L_2	299.9	303.4	**413.6**	326.4	329.9	**434.8**	306.1	308.3	**401.9**	336.1	340.5	**441.0**
	CLIP score	35.49	**35.55**	34.81	31.30	**31.33**	31.25	33.25	**33.29**	32.16	32.76	**32.79**	32.63
	CLIP-IQA	**0.504**	**0.504**	0.503	**0.503**	**0.503**	0.502	**0.503**	**0.503**	**0.503**	**0.503**	**0.503**	0.502
	HPSv2	**0.338**	**0.338**	0.316	**0.303**	**0.303**	0.281	0.289	**0.290**	0.269	0.294	**0.295**	0.266
DETAILED	Feature Vendi	3.23	3.27	**3.67**	2.06	2.06	**2.26**	3.26	3.21	**3.66**	3.13	3.22	**3.91**
	Feature Vendi (grouped)	2.68	2.71	**3.00**	1.86	1.86	**2.02**	2.70	2.66	**2.98**	2.59	2.65	**3.10**
	Pixel Vendi	9.46	9.40	**10.41**	8.50	8.44	**9.68**	8.98	8.95	**9.99**	10.15	10.09	**10.86**
	Pixel Vendi (grouped)	5.99	5.96	**6.71**	5.63	5.61	**6.53**	5.65	5.64	**6.36**	6.40	6.37	**7.05**
	Mean pairwise L_2	281.4	285.1	**349.2**	303.2	306.4	**385.8**	287.2	288.9	**344.8**	290.0	293.6	**383.9**
	CLIP score	35.09	**35.20**	34.79	**31.24**	31.21	31.15	33.21	**33.25**	32.86	32.57	**32.60**	32.23
	CLIP-IQA	**0.504**	**0.504**	0.503	**0.503**	**0.503**	0.502	**0.503**	**0.503**	0.502	**0.503**	**0.503**	0.502
	HPSv2	**0.329**	**0.329**	0.319	**0.300**	0.299	0.286	0.287	**0.288**	0.273	**0.289**	**0.289**	0.276

A.10 DETAILED PROMPTS

Each entry below is index-matched, within its category, to the correspondingly numbered entry in Appendix A.11: the two describe the same underlying scene and differ only in specificity. This pairing is what makes the tier contrast in Appendix A.9 a controlled comparison.

A.10.1 O — OBJECT

O1. *A glossy red sports car parked on a clean mountain road, viewed from a low front angle in soft morning light.*

O2. *A handmade blue ceramic teapot with a curved handle on a rustic wooden table beside a folded linen cloth.*

O3. *A black vintage film camera beside an open photography book on a dark wooden desk near a window.*

O4. *A vintage brass compass resting on an old nautical map, lit by a flickering candle in a dim study.*

O5. *A stack of three worn leather-bound books with gold lettering, tied with a frayed ribbon, on a dark mahogany shelf.*

O6. *A clear glass vase holding a single white orchid, placed on a marble countertop near a sunlit window.*

O7. *A luxury silver wristwatch with a deep blue dial and steel bracelet, resting on a polished black stone surface under dramatic studio lighting.*

O8. *A well-worn acoustic guitar with a sunburst finish leaning against a vintage leather armchair in a warmly lit music room.*

O9. *A pristine white leather sneaker displayed on a minimalist concrete pedestal in a brightly lit retail showroom.*

O10. *A round red paper lantern glowing warmly, hanging from a dark wooden beam against a deep blue twilight sky.*

A.10.2 H — HUMAN/ANIMAL

H1. *A golden retriever sitting on green grass in a city park, looking toward the camera in warm afternoon sunlight.*

H2. *A young woman wearing a long black coat walking along a busy city sidewalk on an overcast winter afternoon.*

H3. *A small blue bird standing on a thin flowering tree branch against a softly blurred forest background.*

H4. *A young girl in a bright yellow raincoat jumping over a puddle, splashing water on a rainy autumn street.*

H5. *A black cat curled up asleep on a patterned rug in front of a crackling fireplace, with a steaming mug nearby.*

H6. *An elderly man feeding pigeons on a wooden park bench, surrounded by golden autumn leaves in soft morning light.*

H7. *A ballerina in a pale pink tutu stretching at a wooden barre in a sunlit dance studio with tall mirrors.*

H8. *A red fox with a thick winter coat walking through fresh snow in a quiet birch forest, its breath visible in the cold air.*

H9. *A young boy in a striped shirt flying a red diamond kite on a windy grassy hill under a bright blue sky with scattered clouds.*

H10. *A barn owl with speckled feathers perched on a weathered wooden fence post at dusk, with a misty meadow fading into the background.*

A.10.3 N — NATURAL

N1. *A tropical beach with several palm trees, turquoise water, white sand, and scattered clouds at midday.*

N2. *A calm alpine lake surrounded by tall pine trees and snow-covered mountains, with reflections visible in the water.*

N3. *A tall waterfall flowing into a clear pool inside a dense tropical forest covered with moss and broad green leaves.*

N4. *A rolling green meadow dotted with wildflowers, a single large oak tree, and a distant mountain range under a vivid blue sky.*

N5. *A frozen lake under a starry night sky, with snow-covered evergreens lining the shore and the aurora borealis glowing overhead.*

N6. *A dramatic thunderstorm over a vast desert canyon, lightning striking the layered rock formations as dark clouds swirl above.*

N7. *A dense bamboo forest with tall green stalks, narrow sunbeams cutting through the canopy onto a stone path covered in fallen leaves.*

N8. *Sweeping red sand dunes rippled by wind, glowing orange under a low desert sun with long shadows and a cloudless gradient sky.*

N9. *A slow river winding through a misty valley at dawn, with soft pink light on the fog and dark silhouettes of willow trees along the banks.*

N10. *A rugged volcanic coastline with black sand beaches, crashing white waves, and jagged basalt cliffs under a dramatic overcast sky.*

A.10.4 S — STRUCTURED

S1. *A minimalist modern living room with floor-to-ceiling windows, a gray sofa, a wooden coffee table, and bright natural light.*

S2. *A crowded city street at night with illuminated storefronts, moving cars, wet pavement, and colorful reflections.*

S3. *A historic European town square with a stone fountain in the center, outdoor cafés, and pastel-colored buildings.*

S4. *A steampunk airship docked at a floating Victorian-era platform, with brass gears, riveted panels, and billowing steam in a sunset sky.*

S5. *A futuristic neon-lit alley in a cyberpunk city, with holographic signs, steam vents, and a lone figure in a dark trench coat.*

S6. *A medieval castle interior with a long wooden banquet table set for a feast, torches flickering on stone walls, and colorful stained glass windows.*

S7. *A cozy log cabin with warm golden windows and smoke curling from its stone chimney, nestled among snow-laden pine trees at blue hour.*

S8. *A busy underground subway station during rush hour, with blurred commuters, an arriving train's headlights, and tiled walls lined with advertisements.*

S9. *A serene Japanese temple garden with a five-story red pagoda, a koi pond with a wooden arched bridge, and carefully raked gravel under cherry blossoms.*

S10. *A grand old library with towering wooden bookshelves, wrought-iron spiral staircases, green reading lamps, and dust motes floating in shafts of window light.*

A.10.5 C — COMPOSITIONAL

C1. *A chef slicing vegetables on a wooden cutting board in a professional kitchen, with cooking utensils hanging behind them.*

C2. *A panda wearing a white spacesuit standing on the moon, with a small flag nearby and Earth visible in the background.*

C3. *A blue bicycle leaning against a red brick wall, with a brown basket attached to the front and fallen leaves on the ground.*

C4. *A giant octopus wearing a top hat and monocle, playing a grand piano in an underwater ballroom with glowing jellyfish chandeliers.*

C5. *A tiny astronaut exploring a giant sunflower, with dewdrops like planets and a ladybug resting nearby like a parked spaceship.*

C6. *A steampunk owl with copper feathers and brass goggles, perched on a mechanical branch in a clockwork forest.*

C7. *A giraffe wearing a knitted red scarf riding a vintage mint-green scooter down a cobblestone street, with market stalls blurring past.*

C8. *A small round robot with brass fittings gently watering potted tomato plants on a sunny apartment balcony overlooking a quiet city.*

C9. *A giant blue whale floating serenely above a golden desert at dusk, casting a long shadow over the dunes as birds fly alongside it.*

C10. *A medieval knight in polished armor playing chess with a small emerald dragon at a stone table inside a torch-lit castle tower.*

A.10.6 F — FOOD

F1. *A gourmet cheeseburger with melted cheddar, caramelized onions, and a brioche bun on a rustic ceramic plate beside golden fries.*

F2. *A steaming bowl of ramen with a soft-boiled egg, chashu pork, and scallions, with wooden chopsticks resting on the rim in a cozy shop.*

F3. *A tall stack of fluffy pancakes with melting butter and maple syrup dripping down the sides, topped with fresh blueberries on a white plate.*

F4. *A ceramic cup of latte with intricate rosetta foam art on a wooden café table beside a small vase of dried flowers.*

F5. *A three-layer dark chocolate cake with glossy ganache, topped with fresh raspberries and mint leaves on an elegant cake stand.*

F6. *A glossy fruit tart with kiwi, strawberries, and blueberries arranged in a spiral, displayed on a marble bakery shelf under warm light.*

F7. *An artfully arranged sushi platter with nigiri, maki rolls, pickled ginger, and wasabi on a dark wooden board with chopsticks and soy sauce.*

F8. *A wood-fired margherita pizza with bubbling mozzarella and fresh basil being sliced on a wooden peel in a rustic Italian kitchen.*

F9. *A waffle cone with two scoops of pastel pink strawberry and mint ice cream, beginning to melt, against a bright summer boardwalk backdrop.*

F10. *A morning breakfast table with a French press of coffee, flaky croissants, orange juice, and a folded newspaper in soft window light.*

A.10.7 A — ARTISTIC

A1. *A loose watercolor painting of a quiet harbor at dawn, with soft washes of pink and gray, moored sailboats, and gentle reflections bleeding into the paper.*

A2. *An impressionist oil painting of a sunflower field under a swirling turquoise sky, with thick visible brushstrokes and rich impasto texture.*

A3. *A bold geometric abstract composition of interlocking circles and triangles in terracotta, teal, and cream, with subtle grain texture and balanced asymmetry.*

A4. *An expressive charcoal sketch of a dancer mid-leap, with sweeping gestural lines, smudged shadows, and dynamic movement captured on textured paper.*

A5. *A layered paper-cut artwork of a forest with deer, crafted from sheets of green and gold paper, casting delicate shadows in a lightbox display.*

A6. *A vibrant stained glass window depicting a rising phoenix, with ruby and amber glass segments outlined in dark lead, glowing in transmitted light.*

A7. *An intricate ceramic tile mosaic of a sea turtle swimming through coral, made of hundreds of small turquoise, jade, and sand-colored tiles.*

A8. *A surreal pen-and-ink drawing of a floating city with impossible staircases and hanging gardens, rendered in fine crosshatching on aged paper.*

A9. *A low-poly 3D render of a snow-capped mountain and pine forest at sunrise, with flat-shaded facets in soft gradients of pink, purple, and blue.*

A10. *A vibrant graffiti mural of a roaring lion with a mane of swirling colorful patterns, spray-painted on a brick wall with paint drips.*

A.10.8 T — TEXT/SIGNAGE

T1. *A buzzing red neon sign spelling 'Open' in cursive, mounted on the dark window of a late-night diner with rain outside.*

T2. *A hand-carved wooden sign reading 'Welcome' in rustic lettering, hanging from chains on the door of a countryside bed and breakfast.*

T3. *A retro science-fiction movie poster titled 'Starlight' in bold chrome lettering, showing a rocket soaring past a ringed planet.*

T4. *A cozy café chalkboard menu with 'Today's Special: Pumpkin Soup' written in decorative white chalk lettering, framed by drawn vines.*

T5. *A vintage art-deco travel poster of Paris with 'Visit Paris' in elegant typography above a stylized Eiffel Tower at sunset.*

T6. *A moody book cover titled 'The Last Lighthouse' in embossed serif type, showing a lone lighthouse on a cliff under storm clouds.*

T7. *A large street mural with the word 'Hope' painted in rainbow-gradient block letters across a city wall, with small birds perched on the letters.*

T8. *A pastel birthday cake with 'Happy Birthday' piped in blue icing, surrounded by lit candles and colorful sprinkles on a party table.*

T9. *A split-flap airport departure board displaying 'Tokyo 09:45 On Time' in white letters, glowing in a busy terminal.*

T10. *A classic theater marquee with 'Tonight' in glowing bulb letters, above a red velvet entrance on a rainy evening street.*

A.11 General prompts

The general-tier counterparts of Appendix A.10, matched by category and index.

A.11.1 O — Object

O1. *A red sports car on a road.*
O2. *A ceramic teapot on a table.*
O3. *A vintage camera beside a book.*
O4. *A brass compass on a map.*
O5. *A stack of old books on a shelf.*
O6. *A glass vase with a flower on a counter.*
O7. *A silver wristwatch on a black surface.*
O8. *An acoustic guitar leaning against a chair.*
O9. *A white sneaker on a pedestal.*
O10. *A paper lantern glowing in the dark.*

A.11.2 H — Human/animal

H1. *A golden retriever in a park.*
H2. *A woman wearing a black coat in the city.*
H3. *A bird standing on a tree branch.*
H4. *A girl in a raincoat jumping over a puddle.*
H5. *A cat sleeping on a rug by a fireplace.*
H6. *A man feeding pigeons on a park bench.*
H7. *A ballerina stretching at a barre.*
H8. *A fox walking through snow.*
H9. *A boy flying a kite on a hill.*
H10. *An owl perched on a fence post.*

A.11.3 N — Natural

N1. *A beach with palm trees.*
N2. *A mountain lake surrounded by pine trees.*
N3. *A waterfall in a tropical forest.*
N4. *A meadow with an oak tree and mountains.*
N5. *A frozen lake at night with trees.*
N6. *A thunderstorm over a canyon.*
N7. *A bamboo forest with sunbeams.*
N8. *A desert dune at sunset.*
N9. *A misty river at dawn.*
N10. *A volcanic coastline with black sand.*

A.11.4 S — Structured

S1. *A modern living room with large windows.*
S2. *A busy city street at night.*
S3. *A European town square with a fountain.*
S4. *A steampunk airship at a platform.*
S5. *A futuristic neon alley in a city.*
S6. *A medieval castle banquet hall.*
S7. *A cozy mountain cabin in winter.*
S8. *A subway station during rush hour.*
S9. *A Japanese temple garden with a pagoda.*
S10. *An old library with spiral staircases.*

A.11.5 C — Compositional

C1. *A chef preparing food in a kitchen.*
C2. *A panda wearing a spacesuit on the moon.*
C3. *A blue bicycle against a brick wall.*
C4. *An octopus playing a piano underwater.*
C5. *A tiny astronaut on a sunflower.*
C6. *A steampunk owl on a branch.*
C7. *A giraffe wearing a scarf riding a scooter.*
C8. *A robot watering plants on a balcony.*
C9. *A whale floating above a desert.*
C10. *A knight playing chess with a dragon.*

A.11.6 F — Food

F1. *A cheeseburger on a plate.*
F2. *A bowl of ramen with chopsticks.*
F3. *A stack of pancakes with syrup.*
F4. *A cup of latte with foam art.*
F5. *A chocolate cake with berries.*
F6. *A fruit tart on a bakery shelf.*
F7. *A sushi platter on a wooden board.*
F8. *A pizza being sliced.*
F9. *An ice cream cone with two scoops.*
F10. *A breakfast table with coffee and croissants.*

A.11.7 A — Artistic

A1. *A watercolor painting of a harbor.*
A2. *An oil painting of a sunflower field.*
A3. *A geometric abstract composition.*
A4. *A charcoal sketch of a dancer.*
A5. *A paper-cut artwork of a forest.*
A6. *A stained glass window of a phoenix.*
A7. *A mosaic of a sea turtle.*
A8. *A surreal ink drawing of a floating city.*
A9. *A low-poly 3D render of a mountain.*
A10. *A graffiti mural of a lion.*

A.11.8 T — Text/signage

T1. *A neon sign that says 'Open'.*
T2. *A wooden sign that says 'Welcome'.*
T3. *A movie poster titled 'Starlight'.*
T4. *A café chalkboard menu.*
T5. *A vintage travel poster of Paris.*
T6. *A book cover titled 'The Last Lighthouse'.*
T7. *A street mural with the word 'Hope'.*
T8. *A birthday cake that says 'Happy Birthday'.*
T9. *An airport departure board.*
T10. *A marquee sign that says 'Tonight'.*