Same Meaning, Different Pictures: Finding Missing Generated Pictures

Amitabh Mahapatra David Forsyth
University of Illinois Urbana-Champaign

Figure 1. Same meaning, different pictures. Starting from the same anchor, all displayed images are accepted under the same SigLIP2 semantic tolerance. A seed sweep produces a conservative family. Directed conditional exploration finds images that remain semantically close to the anchor while varying more strongly in DINOv3 or SAM-based feature space.

Abstract

Text-to-image generators are commonly queried by fixing a prompt and varying the random seed. This passive query policy produces prompt-aligned images, but it can expose only a narrow subset of the images compatible with the condition. We study this failure mode by defining a SigLIP2 semantic neighborhood around an anchor image and comparing seed sweeps with directed conditional exploration under the same final SigLIP2 tolerance. Directed exploration uses feedback from accepted samples to preserve semantic proximity while increasing DINOv3 or SAM3D-based variation. Qualitative grids, cross-model examples, and conditional diversity frontiers show conditional under-coverage: current generators can produce semantically valid pictures that ordinary seed variation misses. Query policy therefore matters when generated images serve as visual priors or hypothesis sets for perception, simulation, and active sensing.

1. Introduction

A standard way to explore a text-to-image generator is to fix a prompt and vary the random seed. This procedure is simple and often useful, but it is a passive query policy: each candidate is generated independently of the images already observed. We show that this policy can be conservative even after enforcing a matched semantic constraint. The generator can produce semantically valid pictures that ordinary seed variation misses.

The issue is relevant when generative visual models are used as priors or hypothesis generators. Active sensing emphasizes query choice: observations depend on where a system looks, which measurement it takes, or which hypothesis it tests. A passive seed sweep provides independent samples from one prompt. A feedback-guided query can instead use previous accepted samples to search for additional plausi-

ble appearances, layouts, and viewpoints under the same high-level condition. If the passive query returns only one narrow visual family, downstream systems inherit a narrow hypothesis set.

We study this question in image generation, where query policy can be isolated from environment dynamics. Given an anchor image, we define a semantic neighborhood using SigLIP2 and ask whether the generator can remain inside this neighborhood while moving far in other feature spaces. We use DINOv3 as an appearance-sensitive probe and SAM3D-based features as a structure-sensitive probe. These embeddings are operational measurements, not ground-truth labels for semantics, appearance, or geometry.

Figure 1 shows the core phenomenon. The seed-sweep row produces plausible images, but they remain close to a narrow visual template. Directed conditional exploration finds images that satisfy the same SigLIP2 tolerance while changing background, appearance, viewpoint, or structure more substantially. These samples are coverage witnesses: semantically valid generated images that are reachable by the model and missed by passive seed variation.

Contributions. We make four contributions. First, we identify conditional under-coverage in seed-based sampling of text-to-image generators under matched SigLIP2 tolerance. Second, we introduce directed conditional exploration, a noise-space search that preserves SigLIP2 proximity while increasing DINOv3 or SAM3D-based spread. Third, we show qualitative and quantitative evidence across prompts and generators, including Stable Diffusion 3.5 Large, FLUX.2 klein, and Qwen Image. Fourth, we connect the finding to passive versus feedback-guided querying of generative visual models used as visual priors or active-sensing components.

Our work builds on diffusion-based image generation [1, 2, 8], controllable generation [12], and recent evidence that seeds in text-to-image models can have systematic effects [3, 6, 10]. We differ from seed selection work by studying coverage of the image family compatible with a fixed semantic neighborhood. We use modern image embeddings as probes, including SigLIP2 [9, 11], DINOv3 [7], and SAM3D [5].

2. Directed Conditional Exploration

Let G denote a text-to-image generator. We treat generation as a deterministic map from a random parameterization z to an image $\mathcal{I} = G(z)$. The parameterization may be an initial noise tensor or a full sampler noise stack. For an anchor image $\mathcal{I}^* = G(z^*)$, let $s(\cdot)$ be the SigLIP2 embedding and $s^* = s(\mathcal{I}^*)$. We define the SigLIP2-feasible set in generator input space as

$$\mathcal{Z}_{\tau_s}(s^*) = \{z : \|s(G(z)) - s^*\|_2 \leq \tau_s\}. \tag{1}$$

All embeddings are ℓ_2-normalized before distances are computed. We write $\Delta S(x) = \|s(x) - s^*\|_2$ for SigLIP2 distance from the anchor, ΔD for DINOv3 distance, and ΔM for SAM3D-based distance.

Passive seed sweep. The baseline fixes the prompt, varies the seed, and filters generated images using the final SigLIP2 constraint $\Delta S(x) \leq \tau_s$. Filtering is part of the baseline because independent seed variation also changes the semantic embedding. The query is passive because the next generator input is independent of previous accepted images.

Directed exploration. We search for samples that remain feasible under Eq. (1) while increasing diversity in a target feature space. Let $e(\cdot)$ be the feature to vary. In the DINO-directed variant, $e = d$ is DINOv3. In the SAM3D-directed variant, $e = m$ is the SAM3D-based feature. Given an accepted pool P, we favor candidates whose target features are far from previously accepted samples:

$$\max_z \; L_{\text{div}}(z; P) \quad \text{s.t.} \quad \|s(G(z)) - s^*\|_2 \leq \tau_s. \tag{2}$$

In practice, L_{div} is a robust repulsion from a random subset of P in the target feature space.

Projected proposals. Let

$$g_s(z) = \nabla_z \|s(G(z)) - s^*\|_2^2, \qquad g_e(z) = \nabla_z L_{\text{div}}(z; P). \tag{3}$$

We suppress first-order motion in the semantic direction by projecting the diversity gradient away from g_s:

$$p(z) = g_e(z) - \frac{\langle g_e(z), g_s(z)\rangle}{\|g_s(z)\|_2^2 + \epsilon} g_s(z). \tag{4}$$

A candidate is generated by a stochastic step along $p(z)$. The exploration scale is adjusted to maintain a useful acceptance rate. A proposed image is added to P after it passes the final SigLIP2 constraint. During proposal generation we use C-RADIOv4 as an efficient semantic proxy [4]; all final filtering and all reported measurements use the target encoders.

We refer to the three strategies as **Seed**, **DINO-directed**, and **SAM3D-directed**. Seed is the fix-prompt, vary-seed procedure. The latter two are feedback-guided conditional explorations of the same generator under the same semantic tolerance.

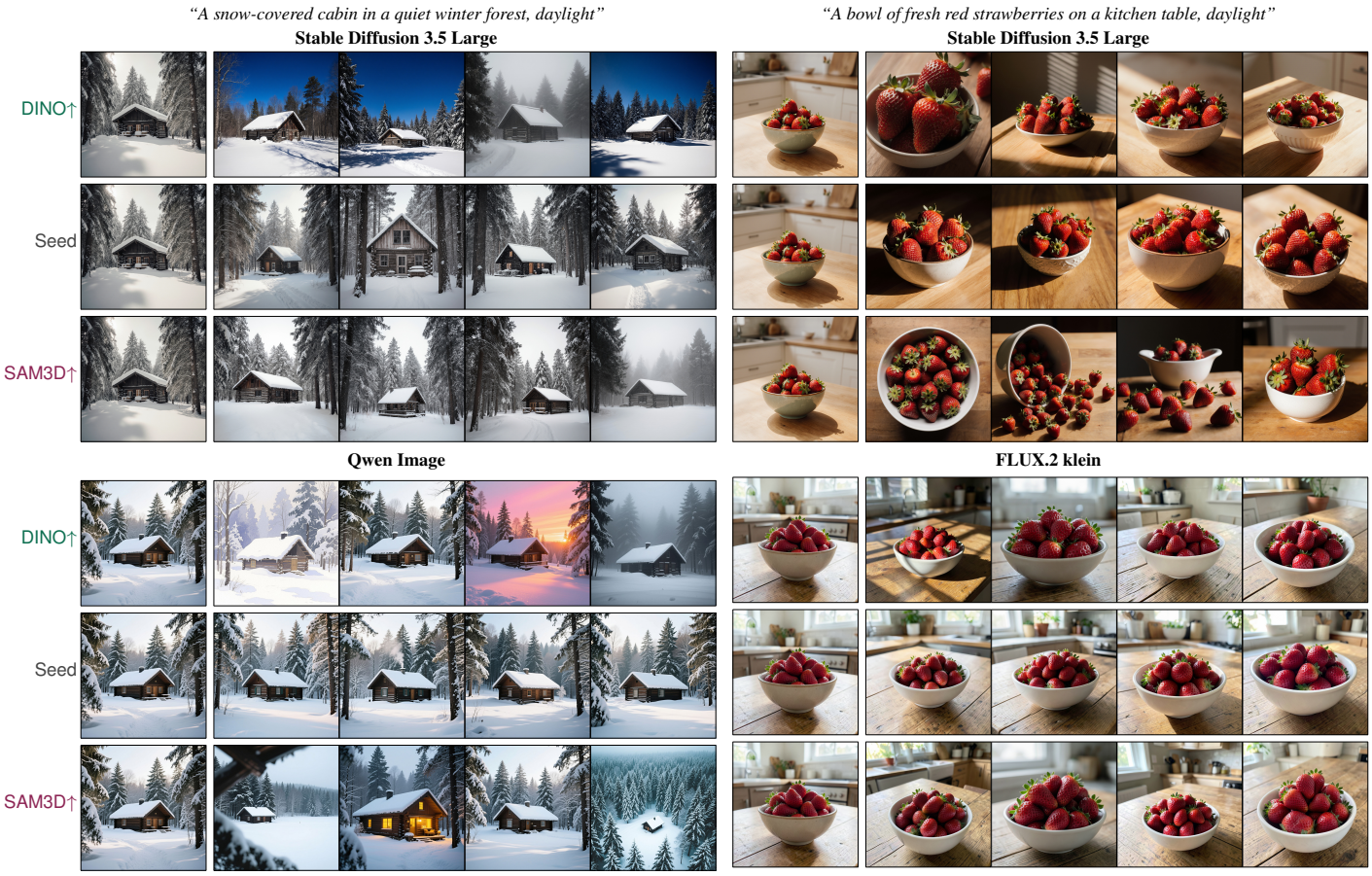


Figure 2. **Cross-model conditional under-coverage.** Each prompt is shown for two generators. Within each model block, all rows use the same prompt and are filtered by the same SigLIP2 semantic tolerance. Seed sweeps produce plausible but conservative image families, while DINO-directed and SAM3D-directed exploration reveal additional semantically valid alternatives under the same final SigLIP2 filter.

3. Experimental Setup

Our primary generator is Stable Diffusion 3.5 Large. For each prompt, we sample an anchor image using a randomly chosen seed. We then compare Seed, DINO-directed, and SAM3D-directed sampling around the same anchor. The main prompt set covers people, animals, objects, scenes, and actions. We include cross-model qualitative checks using FLUX.2 klein and Qwen Image.

The default semantic tolerance is $\tau_s = 0.30$. Each method runs until it produces a sufficient accepted pool, and all statistics are computed on matched subsets of accepted images. This avoids a trivial advantage from producing more accepted samples. The Seed baseline is filtered by the same final SigLIP2 constraint as the directed methods.

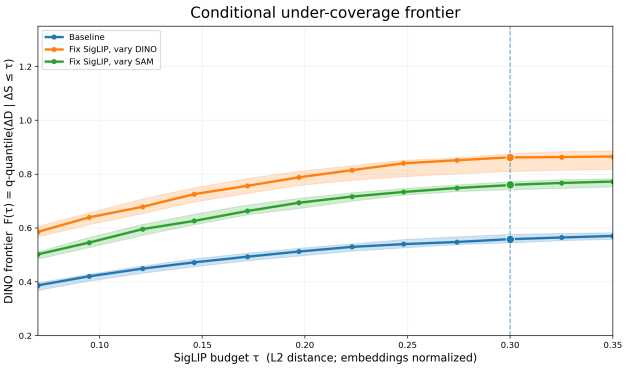
We summarize accepted pools in two ways. Qualitative grids show the anchor and accepted samples from each method under the same semantic tolerance. Conditional diversity frontiers measure how much target-feature variation is reachable for a given semantic budget. For DINOv3, the frontier is

$$F_D(\tau) = Q_q \{\Delta D(x) : \Delta S(x) \leq \tau\}, \tag{5}$$

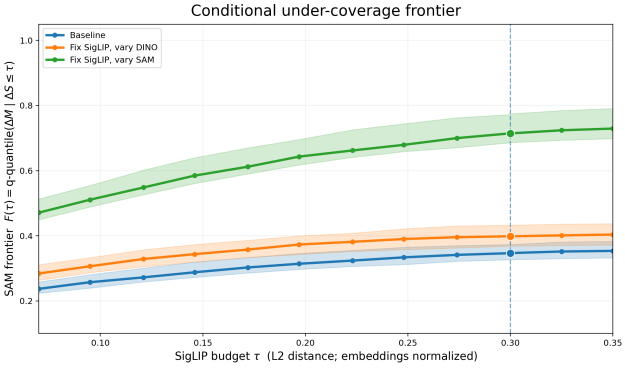
where Q_q is a high quantile. We define $F_M(\tau)$ analogously using ΔM. Unless stated otherwise, we use $q = 0.90$. Qualitative grids show concrete missing pictures, frontier plots summarize the same effect over semantic budgets.

4. Results

Seed sweeps expose a narrow family. Figure 1 gives a qualitative example. The Seed row changes local details, but the samples remain close to a single visual template. The DINO-directed and SAM3D-directed rows satisfy the same SigLIP2 tolerance and show broader changes in appearance, background, viewpoint, and structure. The result illustrates conditional under-coverage: the generator contains feasible images that passive variation misses.



(a) DINOv3 frontier



(b) SAM3D-based frontier

Figure 3. **Conditional under-coverage frontiers.** For each SigLIP2 budget τ, we report a high quantile of target-feature distance among samples satisfying $\Delta S \leq \tau$. DINO-directed exploration reaches larger DINOv3 distances than seed sweeps, and SAM3D-directed exploration reaches larger SAM3D-based distances, under matched semantic tolerance.

Frontier summaries agree with the grids. We summarize the effect over semantic budgets by computing a high quantile of target-feature distance among accepted images satisfying $\Delta S \leq \tau$. All methods are evaluated only on samples that pass the same final SigLIP2 filter, so the frontier measures exposed variation rather than relaxed semantics. The DINOv3 frontier measures appearance-sensitive variation, and the SAM3D frontier measures structure-sensitive variation.

The effect appears across generators. Figure 2 applies the same protocol to two prompts and multiple generators. Although anchors and accepted samples differ across models, passive seed sweeps stay near one iconographic family while directed exploration finds additional feasible images.

Different probes expose different missing pictures. Figures 3 and 4 give the quantitative summary. DINO-directed exploration reaches larger DINOv3 distances at matched SigLIP2 budgets, while SAM3D-directed exploration reaches larger SAM3D-based distances. The two directions are not identical: DINOv3-directed search often changes texture, color, background, and local appearance, while SAM3D-directed search often changes spatial arrangement, silhouette, viewpoint, or object layout. The agreement between grids and frontiers shows the coverage gap both as visible missing pictures and as aggregate feature-space spread.

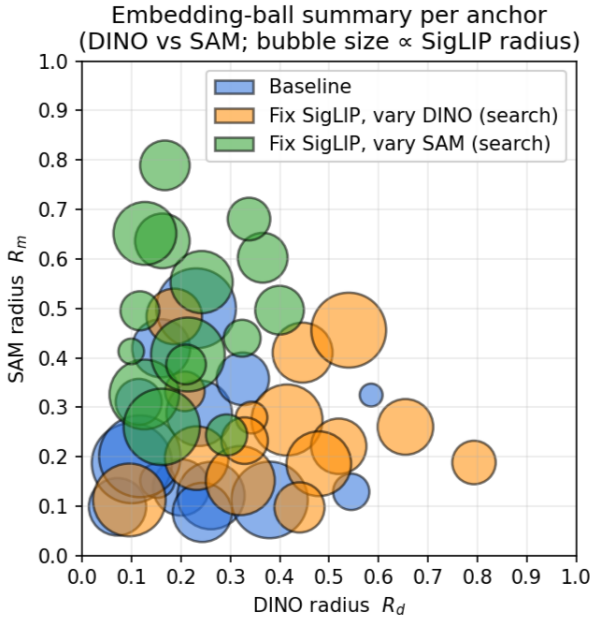


Figure 4. **Different probes expose different missing pictures.** Each point summarizes an accepted pool for a prompt and anchor. The horizontal axis shows DINOv3 spread and the vertical axis shows SAM3D-based spread. DINO-directed and SAM3D-directed searches move accepted pools in different directions, indicating that the two probes reveal different families of semantically feasible images.

5. Discussion

Seed sweeping is a useful default query for text-to-image generators, but it is a passive and incomplete way to expose a conditional image family. Under matched SigLIP2 tolerance, directed exploration finds semantically valid pictures that seed sweeps fail to reveal. The generator supports broader visual alternatives than ordinary seed variation makes visible. This separates generator capacity from query exposure: the missing pictures are reachable by the model, but the passive policy rarely visits them.

When generated images are used as priors, simulators, or hypothesis sets, query choice determines which visual alternatives are exposed. Independent seeds provide one passive view of the condition; feedback-guided exploration uses accepted samples to ask for alternatives that are still semantically consistent. The same model and prompt can therefore expose different hypothesis sets depending on the query policy.

References

[1] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance, 2022. 2

[2] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020. 2

[3] Shuangqi Li, Hieu Le, Jingyi Xu, and Mathieu Salzmann. All seeds are not equal: Enhancing compositional text-to-image generation with reliable random seeds, 2024. 2

[4] Mike Ranzinger, Greg Heinrich, Collin McCarthy, Jan Kautz, Andrew Tao, Bryan Catanzaro, and Pavlo Molchanov. C-radiov4 (tech report), 2026. 2

[5] SAM 3D Team, Xingyu Chen, Fu-Jen Chu, Pierre Gleize, Kevin J. Liang, Alexander Sax, Hao Tang, Weiyao Wang, Michelle Guo, Thibaut Hardin, Xiang Li, Aohan Lin, Jiawei Liu, Ziqi Ma, Anushka Sagar, Bowen Song, Xiaodong Wang, Jianing Yang, Bowen Zhang, Piotr Dollár, Georgia Gkioxari, Matt Feiszli, and Jitendra Malik. Sam 3d: 3dfy anything in images, 2025. 2

[6] Dvir Samuel. Seedselect. GitHub repository, 2025. Accessed 2026-03-03. 2

[7] Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, Francisco Massa, Daniel Haziza, Luca Wehrstedt, Jianyuan Wang, Timothée Darcet, Théo Moutakanni, Leonel Sentana, Claire Roberts, Andrea Vedaldi, Jamie Tolan, John Brandt, Camille Couprie, Julien Mairal, Hervé Jégou, Patrick Labatut, and Piotr Bojanowski. Dinov3, 2025. 2

[8] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models, 2020. 2

[9] Michael Tschannen, Alexey Gritsenko, Xiao Wang, Muhammad Ferjad Naeem, Ibrahim Alabdulmohsin, Nikhil Parthasarathy, Talfan Evans, Lucas Beyer, Ye Xia, Basil Mustafa, Olivier Hénaff, Jeremiah Harmsen, Andreas Steiner, and Xiaohua Zhai. Siglip 2: Multilingual vision-language encoders with improved semantic understanding, localization, and dense features, 2025. 2

[10] Katherine Xu, Lingzhi Zhang, and Jianbo Shi. Good seed makes a good crop: Discovering secret seeds in text-to-image diffusion models, 2024. 2

[11] Xiaohua Zhai, Basil Mustafa, Alexander Kolesnikov, and Lucas Beyer. Sigmoid loss for language image pre-training. *arXiv preprint arXiv:2303.15343*, 2023. 2

[12] Lvmin Zhang and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. *arXiv preprint arXiv:2302.05543*, 2023. 2