UNPACKING EVOLVING DYNAMICS THROUGH THE REPRESENTATION PATHWAY

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural networks transform representations along computational pathways, via stacked layers or iterative processes such as diffusion sampling. How these representations evolve and how their evolution relates to model behavior, provides a complementary perspective to analyzing final representations alone. To this end, we introduce Representation Pathway Consistency (REPAC), an image-level analytical metric that quantifies how well relational structure among an image's patch tokens is preserved between successive pathway states and aggregates these adjacent consistencies across the pathway. Using REPAC, we find that higher pathway consistency is generally associated with better performance across discriminative and generative tasks. In generative models, generated images with more consistent pathways tend to lie closer to class-conditional prototypes in representation space, providing a geometric perspective on this association. Moreover, REPAC computed from an initial segment of a generative pathway is predictive of final sample quality. Leveraging this signal, we develop a plug-and-play early-selection strategy that ranks partially generated candidates using REPAC and continues only the highest-ranked ones to completion. This reduces sampling cost while improving the quality of retained samples, without external reward signals or additional training. Together, these findings reveal regularities in representation pathway dynamics and their relationship to downstream behavior, while demonstrating the utility of pathway consistency for quality-aware and efficient selective generation.

1 INTRODUCTION

Visual models are commonly evaluated on a wide range of downstream tasks, from recognition and transfer to image generation. While these evaluations characterize task-specific performance, they primarily measure final predictions or generated outputs and provide limited insight into the sequence of internal transformations that produce them. Within a model, computation unfolds along a *representation pathway*, spanning network layers and, in diffusion models, successive sampling steps (Baldock et al., 2021; Rombach et al., 2022; Peebles & Xie, 2023; Luo et al., 2023). Although prior work has characterized properties of intermediate representations (Zeiler & Fergus, 2014; Yosinski et al., 2014; Alain & Bengio, 2016; Evci et al., 2022) and their dependence on learning objectives (Shekhar et al., 2023; Masarczyk et al., 2023), how these representations evolve for an individual sample, and how that evolution relates to model behavior, remains incompletely understood.

Representation-similarity methods compare relational structures induced by a shared set of observations without requiring correspondence between feature coordinates (Raghu et al., 2017; Morcos et al., 2018; Kornblith et al., 2019). When images serve as observations, these comparisons characterize population-level relationships between layers or models. Example-level measures such as prediction depth track when layer-wise probe decisions stabilize (Baldock et al., 2021), focusing on prediction agreement. To understand changes in the internal structure of an individual image, we ask: *how do relations among its spatial tokens evolve along a representation pathway, and how is this evolution associated with discriminative and generative behavior?*

To address this question, we treat the spatial patches of an individual image as observations and track their relations across pathway states. We introduce Representation Pathway Consistency (REPAC), a training-free statistic that quantifies the preservation of these relations. After aligning the spatial

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.

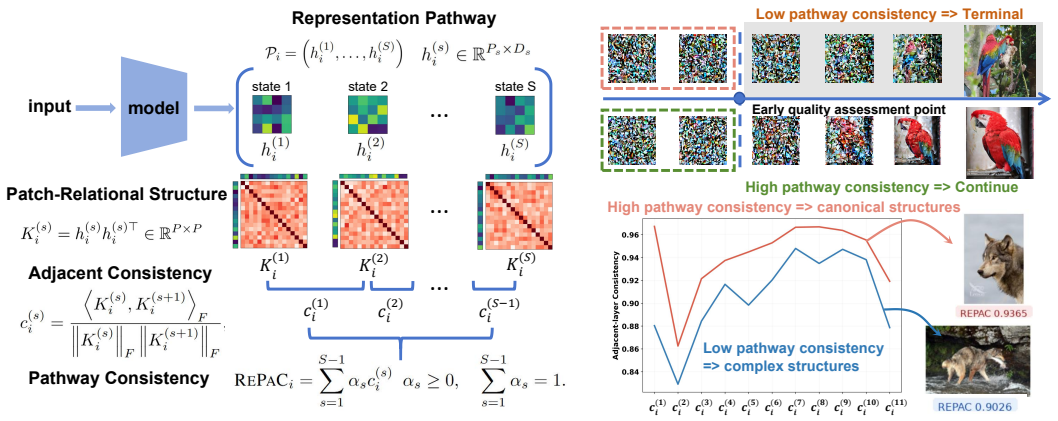


Figure 1: **Overview of representation pathway consistency.** We characterize the evolution of an individual sample's representation along a computational pathway by tracking the preservation of patch-relational structures across intermediate states. The resulting REPAC provides an example-level measure of representation dynamics, where the consistency is associated with visual structures of images, discriminative performance, generative quality, and early quality assessment.

token grids across states, we construct a token-level Gram matrix at each state, compare adjacent matrices using normalized alignment, and aggregate the local transitions along the pathway. The resulting score captures relational consistency even when feature coordinates and dimensionality change across states. The same construction applies across encoder depth, denoiser depth, and denoising time, without labels, trained probes, or a reference population.

Across six visual backbones on ImageNet-1K, groups with higher REPAC generally attain greater linear-probe accuracy and mean ground-truth-class probability. For class-conditional DiT generation, groups with higher temporal consistency over the full sampling trajectory achieve higher Inception Scores (IS). Generated samples with higher REPAC also tend to have greater similarity to class prototypes formed from high-guidance generations in an external Inception-v3 feature space (Szegedy et al., 2016), relating consistency to class typicality. Motivated by these associations, we evaluate partial diffusion pathways for Early Quality Assessment (EQA): candidates are ranked by REPAC within each prompt, and only the selected subset continues to completion. On DPG-Bench with SD3.5-Medium, selecting four of 20 candidates using the uniformly weighted REPAC after 20% of sampling yields an overall score of 85.55, compared with 85.00 for ASD (Wang et al., 2026) and 84.30 for random selection. This selective-continuation protocol requires 36% of the candidate-denoising steps needed to complete all 20 candidates, excluding scoring overhead.

Our contributions are as follows:

- We formulate representation evolution at the level of an individual sample by following patch-relational structure along an ordered computational pathway, and introduce REPAC to quantify its consistency across successive states.
- We identify a shared empirical relationship between pathway consistency and discriminative and generative behavior across diverse visual models spanning different learning paradigms, supported by complementary geometric analysis in representation space.
- We apply REPAC to training-free early quality assessment for generative models, selecting promising candidates for completion while reducing denoising computation.

2 Related Work

Representation formation along network depth. Representation formation in deep networks is inherently layer-wise: successive transformations reorganize visual information, yielding features with different selectivity, transferability, and task accessibility across depth (Zeiler & Fergus, 2014; Yosinski et al., 2014; Alain & Bengio, 2016). Intermediate features can complement final represen-

tations for transfer, while learning objectives shape their layer-wise organization (Evci et al., 2022; Shekhar et al., 2023). Such organization is often nonuniform, forming groups of similar layers or extractor–tunnel structures (Nguyen et al., 2020; Masarczyk et al., 2023). Repeated computation extends this process beyond feed-forward depth: Universal and looped Transformers iteratively update latent states (Dehghani et al., 2018; Yang et al., 2024; Gatmiry et al., 2024; Saunshi et al., 2025; Xu & Sato, 2024), whereas diffusion models evolve representations both denoiser blocks and iterative sampling steps (Baldock et al., 2021; Rombach et al., 2022; Peebles & Xie, 2023; Luo et al., 2023). Motivated by these findings, we study how the relational structure of a single example evolves along such pathways and whether its consistency is systematically related to model behavior.

Representation similarity. Neural representations have been compared using canonical correlation, representational similarity matrices, shape metrics, and kernel alignment (Raghu et al., 2017; Morcos et al., 2018; Williams et al., 2021; Kornblith et al., 2019). These methods characterize the relation between two representation spaces through a set of shared observations, and their interpretation consequently depends on how those observations are defined (Ding et al., 2021; Cui et al., 2022; Klabunde et al., 2025). Most prior analyses use a collection of images as observations to compare layers, models, or learning objectives at the population level. We instead use the patches of a single image as observations. This shift turns representation similarity into an image-level measure of how within-image relational structure evolves between successive states and, by aggregating these transitions, along an entire computational pathway.

Example-level behavior and typicality. Example-wise analyses connect internal representations to difficulty and typicality. Prediction depth identifies the earliest layer at which local probe decisions agree with the final prediction (Baldock et al., 2021), while Forte estimates typicality from neighborhoods in an external representation space (Ganguly et al., 2025). Both approaches evaluate an example using information beyond its own representation pathway, such as labeled probe outputs or an external reference population. In contrast, REPAC is computed solely from the example's internal patch representations, requiring neither labels, trained probes, nor reference sets.

Early assessment of diffusion samples. Partial diffusion trajectories can support candidate selection before generation is complete. Probe-Select learns evaluator-specific predictors from early denoiser activations (Guo et al., 2026), whereas CFG-Rejection uses accumulated conditional–unconditional prediction discrepancy (ASD) as a training-free signal (Wang et al., 2026). We instead define the representation pathway for EQA as the states observed from the beginning of sampling to the assessment point and compute REPAC on this pathway. The resulting score remains training-free but captures the evolution of within-image patch relations rather than guidance discrepancy. We compare these signals within condition-matched candidate groups, evaluating whether early pathway consistency identifies better final samples and using the resulting selection to reduce subsequent denoising computation by completing only the selected candidates.

3 REPRESENTATION PATHWAY CONSISTENCY

Visual representations develop through successive transformations. In visual encoders, these transformations unfold across network layers; in diffusion models, they additionally unfold across sampling steps as generation progresses from noise toward an image. We study these representation dynamics, asking how their internal structure changes during computation, whether this evolution exhibits consistent patterns, and how these patterns relate to downstream performance.

3.1 REPRESENTATION PATHWAYS

To study representation evolution, we record its internal spatial features at a chosen sequence of computational positions. We call this ordered sequence a *representation pathway*:

$$\mathcal{P}_i = \left(z_i^{(1)}, \ldots, z_i^{(S)}\right), \qquad z_i^{(s)} \in \mathbb{R}^{P_s \times D_s}, \quad S \geq 2. \tag{1}$$

Here, i identifies one input image or one generation trajectory, and s indexes the recorded states. Each state contains P_s spatial tokens with D_s feature channels; a row corresponds to a patch or spatial region in that state. A pathway therefore follows the collection of spatial tokens for one sample. We consider two state axes to record: network depth and diffusion sampling time.

Layer-wise pathways. For an image x_i processed by an encoder, we record features after selected layers $\ell_1 < \cdots < \ell_S$ in a single forward pass. Let $f^{(\ell)}(x_i)$ denotes spatial features extracted after layer ℓ, the pathway states are $z_i^{(s)} = f^{(\ell_s)}(x_i)$, capturing how the representation evolves with depth.

Pathways in diffusion models. Diffusion sampling provides both a depth axis and a time axis. For one generation trajectory i, let $a_i^{(\ell,t)}$ denotes the spatial features at denoiser block ℓ during sampling step t. We index steps by execution order, $t = 1, \ldots, T$, from initial noise toward the generated image. These features are computed from the current latent at step t; the condition is fixed along the trajectory. We construct pathways by varying one axis while holding the other fixed:

$$\begin{aligned} \mathcal{P}_{i,t}^{\text{layer}} &= \left(a_i^{(\ell_1,t)}, \ldots, a_i^{(\ell_S,t)}\right) \quad \text{(fixed sampling step } t\text{)}, \\ \mathcal{P}_{i,b}^{\text{temp}} &= \left(a_i^{(b,t_1)}, \ldots, a_i^{(b,t_S)}\right) \quad \text{(fixed denoiser block } b\text{)}. \end{aligned} \tag{2}$$

The first follows selected blocks within one denoising step. The second follows the same block across selected sampling steps $t_1 < \cdots < t_S$ of one trajectory. Both instantiate Equation 1; in each case, S counts recorded states and may be smaller than the number of layers or sampling steps. Throughout, *adjacent states* means consecutive entries in the chosen pathway. Model-specific extraction settings are given in Appendix B.

3.2 Patch-Relational Representation Structure

We characterize each state by the pairwise relations among its spatial tokens. This lets us compare how a sample is organized spatially across states with different feature dimensions, without requiring their feature channels to correspond. Token counts can change across a pathway because of downsampling or upsampling. To compare representation across-layers, we use bilinear resampling operator $\mathcal{R}_s$ to align the feature grid to a fixed resolution. We also normalize token channel dimension by ℓ_2-normalization operator $\mathcal{N}$. The resulting representation is $h_i^{(s)} = \mathcal{N}(\mathcal{R}_s(z_i^{(s)})) \in \mathbb{R}^{P \times D_s}$. The patch-relational structure of state s is represented by its linear-kernel Gram matrix

$$K_i^{(s)} = h_i^{(s)} h_i^{(s)\top} \in \mathbb{R}^{P \times P}, \tag{3}$$

where $K_i^{(s)}[u, v]$ is the cosine similarity between spatial tokens u and v. Consequently, $K_i^{(s)}$ captures the patch-relational structure of the state in a shared spatial space $\mathbb{R}^{P \times P}$ independent of D_s.

3.3 Adjacent and Pathway Consistency

With each state characterized by its patch-relational structure, we quantify how well this structure is preserved across successive states. Specifically, we define the adjacent consistency as the normalized Frobenius inner product between adjacent kernel matrices:

$$c_i^{(s)} = \frac{\left\langle K_i^{(s)}, K_i^{(s+1)} \right\rangle_F}{\left\| K_i^{(s)} \right\|_F \left\| K_i^{(s+1)} \right\|_F}, \qquad s = 1, \ldots, S-1. \tag{4}$$

We aggregate these consistencies to define the Representation Pathway Consistency (RePAC):

$$\text{RePAC}_i = \sum_{s=1}^{S-1} \alpha_s c_i^{(s)}, \qquad \alpha_s \geq 0, \quad \sum_{s=1}^{S-1} \alpha_s = 1. \tag{5}$$

The weights α_s determine the contribution of each transition. We consider two variants: RePAC(U) uses uniform weights $\alpha_s = 1/(S-1)$, whereas RePAC(LD) uses linearly decaying weights $\alpha_s = 2(S-s)/[S(S-1)]$ to emphasize earlier transitions, motivated by greater variation in adjacent consistency observed at the early pathway (Section 3.5). Accordingly, $c_i^{(s)}$ describes local structural preservation, while RePAC summarizes consistency over the entire pathway. In Section 4, we examine how this instance-level analytic metric relates to discriminative and generative behavior.

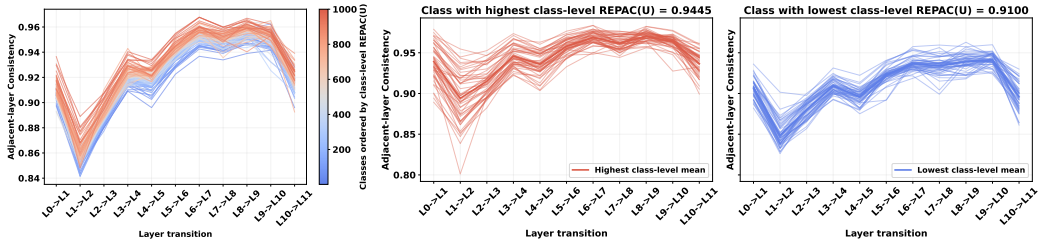


Figure 2: **Consistency dynamics across DINOv3 depth.** Left: class-averaged adjacent consistency trajectories for ImageNet-1K, showing a shared evolution pattern with class-dependent magnitudes. Right: sample-level trajectories from the classes with the highest and lowest REPAC (U)

3.4 EARLY PATHWAYS FOR QUALITY ASSESSMENT

The temporal pathway also makes REPAC available before diffusion sampling is complete. If relational consistency observed early in sampling is informative of the final quality, promising trajectories can be continued while unpromising ones are terminated, enabling early quality assessment. At an intermediate assessment point containing $S_e < S$ states, we compute Equation 5 over only the observed prefix pathway, i.e., $\text{REPAC}_{i,e} = \sum_{s=1}^{S_e-1} \alpha_s c_i^{(s)}$. For each condition, all N candidates are sampled to this point and ranked by $\text{REPAC}_{i,e}$; only the top k are then continued to completion.

This procedure uses $NS_e + k(S - S_e)$ sampling steps instead of the NS steps required to complete all candidates. The resulting theoretical denoising-computation ratio is

$$\rho = \frac{NS_e + k(S - S_e)}{NS}. \tag{6}$$

3.5 PATHWAY CONSISTENCY DYNAMICS

Beyond the aggregated REPAC score, the adjacent consistency trajectory characterizes the dynamics of representation evolving along a pathway. The left panel of Figure 2 shows the class-averaged consistency trajectory for each ImageNet-1K class under DINOv3. Nearly all classes exhibit a common evolution pattern across depth, while differing primarily in magnitude. This shared profile suggests a broadly stable layer-wise transformation pattern whose overall consistency varies across semantic classes. Such class-dependent magnitudes may reflect how the model organizes different visual concepts in its learned representation space, as shaped by the composition of pretraining data.

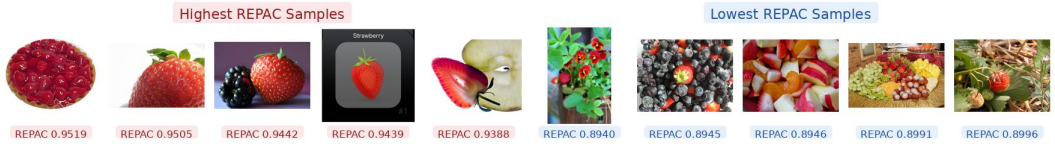
The right panel visualizes sample-level trajectory from the classes with the highest and lowest REPAC (U). In the highest-consistency class, adjacent consistency varies more strongly across early transitions and becomes comparatively stable in later layers. One possible explanation is that early layers perform larger transformations while establishing structured visual information, whereas later layers refine an increasingly stable representation. This concentration of variation near the beginning of the pathway also motivates the linear-decay aggregation for our REPAC (LD). Additional pathway consistency dynamics across different models are provided in Appendix C.

4 EXPERIMENTS

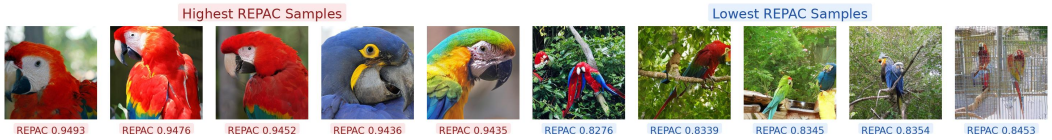
In this section, we conduct experiments across multiple scenarioes to evaluate representation pathway dynamics from three perspectives: (1) characterizing sample-level visual and structural properties captured by pathway consistency; (2) examining its association with discriminative and generative behavior; and (3) demonstrating its utility for early quality assessment in generation.

4.1 EXPERIMENT SETUP

Backbones and benchmarks. We study representation pathway consistency across a diverse set of visual models with different architectures and learning objectives, covering both discriminative and generative settings. For discriminative analysis, we consider four 12-block vision transformers:



(a) Extreme-REPAC ImageNet strawberry samples selected by DINOv3 layer-wise consistency.



(b) Extreme-REPAC DiT-XL/2 macaw generations selected by temporal consistency.

Figure 3: **Representation consistency characterizes distribution centroids.** We visualize within-class samples with extreme REPAC scores under DINOv3 features (real images, top) and DiT-XL/2 (generated images, bottom). Higher REPAC scores highlight exemplar category samples characterized by canonical structures and clean backgrounds, whereas lower scores reflect object occlusion, category mixing, or complex cluttered backgrounds.

DINOv2 ViT-B/14 (Oquab et al., 2023), DINOv3 ViT-B/16 (Siméoni et al., 2025), DistillDIFT with a DINOv2 ViT-B/14 backbone (Fundel et al., 2025), and Pixio ViT-B/16 (Yang et al., 2026). We also include two diffusion backbones: SDXL-Turbo, built on the SDXL U-Net architecture (Podell et al., 2024), and the 28-block DiT-XL/2 (Peebles & Xie, 2023). DiT-XL/2 is further used to analyze pathway consistency during class-conditional generation. For text-to-image early quality assessment, we evaluate Stable Diffusion 3.5 Medium and Large (Esser et al., 2024; Stability AI, 2024b;a). We conduct the discriminative and class-conditional generation experiments on ImageNet-1K (Deng et al., 2009) and evaluate text-to-image generation on DPG-Bench (Hu et al., 2024). Detailed feature extraction, pathway construction, and generation settings are provided in Appendix B.

Evaluation protocol. We group samples globally or within each condition according to REPAC. For condition-wise grouping, samples are ranked separately within each ImageNet class or text prompt, and matched groups are pooled across conditions to control for composition. Unless otherwise specified, samples are divided into five equal-sized groups, from Level 1 (lowest consistency) to Level 5 (highest). We report top-1 accuracy and ground-truth-class probability for ImageNet classification, Inception Score (IS) (Salimans et al., 2016) for DiT generation, and official DPG-Bench metrics for text-to-image generation.

4.2 Visual Evidence of Pathway Consistency

We first examine the visual characteristics of images associated with different REPAC scores. Figure 3 presents samples from the extremes of the REPAC distribution within a single class. For DINOv3, the five highest- and lowest-scoring ImageNet samples are selected using layer-wise pathway consistency. High-REPAC images generally present the class object clearly with simpler composition and less competing visual structure. For DiT-XL/2, applying the same within-class comparison using temporal pathway consistency reveals a similar pattern among generated samples.

These observations suggest that REPAC captures visual regularity reflected in representation evolution. Higher-REPAC samples tend to exhibit more canonical class-specific structures and cleaner compositions, whereas lower-REPAC samples show greater structural complexity and ambiguity.

4.3 Pathway Consistency as a Probe of Discriminative Behavior

To investigate whether pathway consistency reflects discriminative behavior, we conduct linear probing classification experiments on ImageNet-1K using various models. For each backbone, we train a linear classifier on pooled frozen patch features from the ImageNet-1K training set and evaluate it on the validation set. In parallel, we compute REPAC for each validation image. We then rank

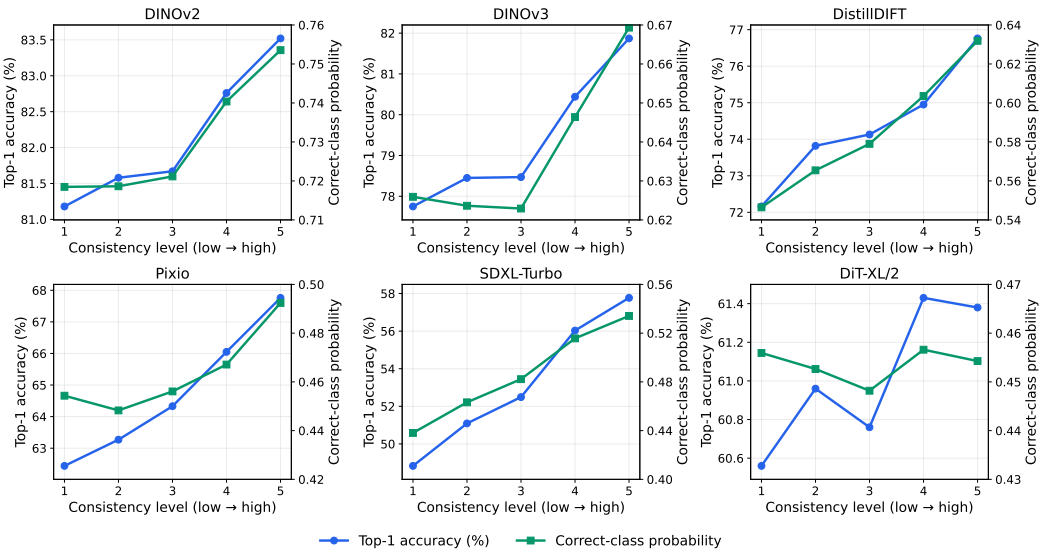


Figure 4: **Classification performance across pathway consistency levels.** Blue curves (left axes) show top-1 accuracy, and green curves (right axes) show the mean softmax probability assigned to the ground-truth class. Both measures generally increase with pathway consistency across models.

all validation images by their REPAC scores and partition them into five equally sized consistency levels, from lowest to highest consistency. For each level, we report the top-1 classification accuracy and the mean softmax probability assigned to the ground-truth class.

Figure 4 shows a consistent relationship between pathway consistency and discriminative behavior across backbones. Images in higher-consistency groups generally achieve greater top-1 accuracy and assign higher probability to the ground-truth class, with a pronounced and largely monotonic trend for most backbones. DiT-XL/2 exhibits a weaker and less monotonic association, although its overall trend remains positive. Taken together, these results suggest that REPAC provides a broadly applicable probe of discriminative behavior: representations with greater pathway consistency tend to support more accurate and confident predictions across substantially different model families.

4.4 PATHWAY CONSISTENCY AS A PROBE OF GENERATIVE BEHAVIOR

We evaluate the association between pathway consistency and class-conditional generation performance using DiT-XL/2 on ImageNet-1K. Following the condition-wise grouping protocol in Section 4.1, we rank samples by REPAC within each class, pool matched consistency groups across classes, and compute the Inception Score (IS) for each group. We conduct this analysis for both temporal pathways over full sampling trajectory and layer-wise pathways at selected sampling steps.

Figure 5 reveals a positive association under both pathway constructions. For temporal pathways, IS increases markedly and monotonically across consistency levels. Layer-wise pathways exhibit a sampling-dependent pattern: early in generation, REPAC values remain tightly clustered and the corresponding IS differences are modest; as sampling proceeds, the consistency levels become increasingly separated, accompanied by clearer stratification in IS. These results show that more consistent representation dynamics are associated with stronger generative performance, while the evolving layer-wise pattern shows that this association becomes more pronounced as the generation unfolds.

4.5 EARLY QUALITY ASSESSMENT VIA PATHWAY CONSISTENCY

The association between pathway consistency and final sample quality suggests that REPAC may support early quality assessment. Unlike conventional best-of-N selection, which evaluates candidates only after full generation, we rank partial trajectories by their REPAC scores and continue only the highest-ranked candidates. This procedure requires neither a learned quality predictor nor an external reward model. We compare REPAC (U), REPAC (LD), ASD, and random selection un-

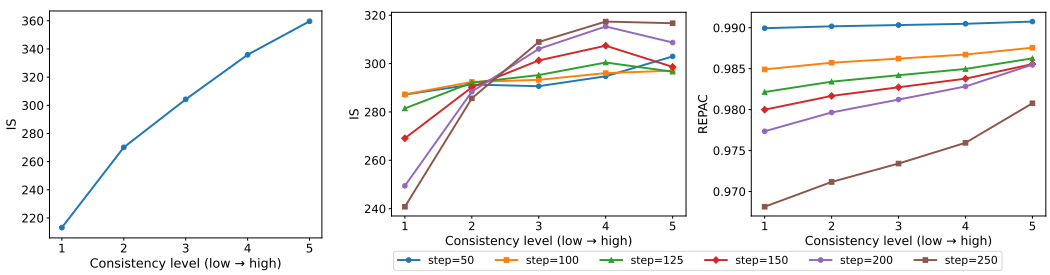


Figure 5: **Inception Score variation across pathway consistency levels.** Left: IS for temporal pathways computed from the middle block over the full sampling trajectory. Center and right: IS and REPAC for layer-wise pathways at selected steps of the 250-step sampling trajectory.

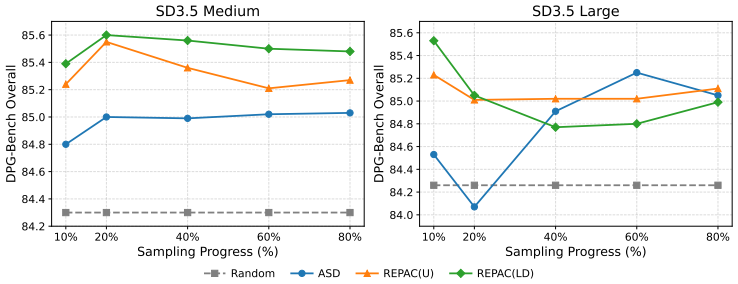


Figure 6: **Overall DPG-Bench performance across the sampling trajectory.** Overall scores for selecting four of 20 candidates from SD3.5-Medium and SD3.5-Large at different assessment points.

der the generation protocol described in Section 4.1. At each assessment point, we rank the partial trajectories within each prompt and continue only the top four candidates to completion. Additional ablation experiments comparing different pathway aggregation strategies, weighting schemes, and layer-wise pathway variants are provided in Appendix E.

Figure 6 shows that both REPAC variants outperform random selection across the two models and all assessment points. The advantage emerges early: after 20% of the sampling trajectory, REPAC (U) and REPAC (LD) achieve overall DPG-Bench scores of 85.55 and 85.60 on SD3.5-Medium, respectively, compared with 85.00 for ASD and 84.30 for random selection. On SD3.5-Large, the corresponding scores are 85.01 and 85.05, compared with 84.07 for ASD and 84.26 for random selection. Thus, pathway consistency measured over only the 20% of the trajectory already identifies candidates with better final text-image alignment across both model scales.

Table 1 provides category-level results at the 10% and 20% assessment points. The best method varies across compositional categories, indicating that no ranking signal dominates every aspect of alignment. Nevertheless, REPAC (LD) achieves the highest overall score for both models at both assessment points. These results establish early pathway consistency as a useful, training-free signal for selective continuation, improving generation quality while reducing computational cost.

4.6 A GEOMETRIC EXPLANATION OF THE BEHAVIORAL ASSOCIATION

The association between pathway consistency and class-conditional generative performance may arise from differences in class-typicality within representation space. To examine this possibility, we construct class prototypes in an independent Inception-v3 feature space. For each ImageNet class y, we generate a high-guidance reference set $\mathcal{R}_y$ with DiT-XL/2 using a CFG scale of 10, and define its prototype as the mean feature of the reference samples. Analyzed samples are generated with the default CFG scale of 1.5. Let $f(\cdot)$ denote the Inception-v3 feature extractor. For a sample x_i conditioned on class y_i, we define its feature z_i, class prototype μ_y, and prototype similarity q_i as

$$z_i = f(x_i), \qquad \mu_y = \frac{1}{|\mathcal{R}_y|} \sum_{x \in \mathcal{R}_y} f(x), \qquad q_i = \frac{z_i^\top \mu_{y_i}}{\|z_i\|_2 \|\mu_{y_i}\|_2}. \tag{7}$$

Table 1: **Detailed DPG-Bench results at early assessment points.** Category-level and overall scores for selecting four of 20 candidates after 10% and 20% of sampling with SD3.5-Medium and SD3.5-Large. Bold indicates the best selection strategy for each model and metric.

Metric	SD3.5-Medium				SD3.5-Large			
	Random	ASD	REPAC (U)	REPAC (LD)	Random	ASD	REPAC (U)	REPAC (LD)
10% Sampling Progress								
Attribute	90.22	90.42	90.00	**91.10**	89.69	88.81	89.44	**91.23**
Entity	89.60	**89.87**	89.39	89.50	90.16	**90.18**	90.06	89.03
Global	88.84	90.34	88.78	**91.37**	87.30	85.25	87.77	**90.82**
Other	89.60	88.07	**91.62**	89.12	89.19	90.14	**90.54**	88.55
Relation	89.37	90.64	**91.19**	89.65	91.20	**91.62**	91.00	89.41
Overall	84.30	84.80	85.24	**85.39**	84.26	84.53	85.23	**85.53**
20% Sampling Progress								
Attribute	87.80	89.69	**90.86**	90.80	**90.82**	90.66	88.65	90.72
Entity	89.71	**91.02**	89.80	89.57	88.05	90.18	90.12	**90.29**
Global	**92.19**	90.83	85.94	89.43	88.82	89.76	**90.78**	88.01
Other	90.00	87.97	**91.91**	89.11	**90.02**	86.88	89.32	**90.02**
Relation	91.54	89.78	88.58	**91.68**	89.70	87.38	**91.01**	87.70
Overall	84.30	85.00	85.55	**85.60**	84.26	84.07	85.01	**85.05**

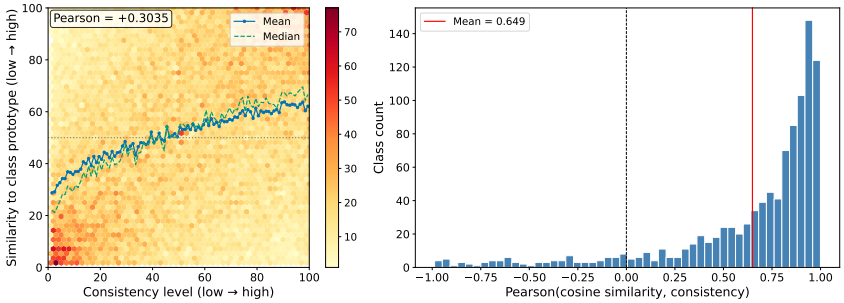


Figure 7: **Pathway consistency as a probe of class-conditional representation geometry.** Left: within-class percentiles of sample–prototype cosine similarity versus REPAC. Right: distribution of class-wise Pearson correlations between the two quantities.

We measure the association between prototype similarity and pathway consistency both across the full sample set and within each class. For class y, the class-wise correlation is

$$\gamma_y = \text{Pearson}\left(\{\text{REPAC}(x_i)\}_{y_i=y}, \{q_i\}_{y_i=y}\right). \tag{8}$$

Figure 7 compares the within-class percentiles of REPAC and prototype similarity. Samples with higher pathway consistency tend to lie closer to their class prototypes in Inception-v3 feature space, yielding an overall Pearson correlation of 0.3035. The association is also predominantly positive within individual classes: the mean class-wise correlation is 0.649, and only a small minority of classes exhibit negative correlations. This geometric pattern helps explain the behavioral association observed in Section 4.4: samples with more consistent internal representation dynamics lie closer to their class prototypes in the feature space of an external image classifier, suggesting stronger class-specific information and, consequently, better class-conditional generative performance.

5 CONCLUSION

We introduced REPAC, an example-level measure of representation pathway dynamics based on the preservation of patch-relational structures across computational states. Across diverse discriminative and generative models, pathway consistency is associated with model behavior, and its relationship with generative performance is further connected to class-conditional representation geometry. Beyond analysis, early-pathway REPAC enables training-free quality-aware generation by selecting promising candidates before completion, reducing denoising computation while maintaining higher-quality outputs. These results establish pathway consistency as a general perspective for understanding representation evolution and its connection to model behavior.

6 AI USE STATEMENT

In this work, we used generative AI tools for manuscript proofreading, language refinement, and improving presentation clarity. We have not used generative AI tools for developing research ideas, designing the method, conducting experiments, analyzing results, or generating scientific claims. All AI-assisted content was reviewed and verified by the authors. The authors take full responsibility for the final manuscript, including all text, claims, experimental results, and artifacts produced with the aid of generative AI.

REFERENCES

Guillaume Alain and Yoshua Bengio. Understanding intermediate layers using linear classifier probes. *arXiv preprint arXiv:1610.01644*, 2016.

Robert Baldock, Hartmut Maennel, and Behnam Neyshabur. Deep learning through the lens of example difficulty. *Advances in Neural Information Processing Systems*, 34:10876–10889, 2021.

Tianyu Cui, Yogesh Kumar, Pekka Marttinen, and Samuel Kaski. Deconfounded representation similarity for comparison of neural networks. *Advances in Neural Information Processing Systems*, 35:19138–19151, 2022.

Mostafa Dehghani, Stephan Gouws, Oriol Vinyals, Jakob Uszkoreit, and Łukasz Kaiser. Universal transformers. *arXiv preprint arXiv:1807.03819*, 2018.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

Frances Ding, Jean-Stanislas Denain, and Jacob Steinhardt. Grounding representation similarity through statistical testing. *Advances in neural information processing systems*, 34:1556–1568, 2021.

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

Utku Evci, Vincent Dumoulin, Hugo Larochelle, and Michael C Mozer. Head2toe: Utilizing intermediate representations for better transfer learning. In *International Conference on Machine Learning*, pp. 6009–6033. PMLR, 2022.

Frank Fundel, Johannes Schusterbauer, Vincent Tao Hu, and Björn Ommer. Distillation of diffusion features for semantic correspondence. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 6762–6774. IEEE, 2025.

Debargha Ganguly, Warren Morningstar, Andrew Yu, and Vipin Chaudhary. Forte: Finding outliers with representation typicality estimation. In *International Conference on Learning Representations*, volume 2025, pp. 23827–23864, 2025.

Khashayar Gatmiry, Nikunj Saunshi, Sashank J Reddi, Stefanie Jegelka, and Sanjiv Kumar. Can looped transformers learn to implement multi-step gradient descent for in-context learning? *arXiv preprint arXiv:2410.08292*, 2024.

Huanlei Guo, Hongxin Wei, and Bingyi Jing. Toward early quality assessment of text-to-image diffusion models. *arXiv preprint arXiv:2603.02829*, 2026.

Xiwei Hu, Rui Wang, Yixiao Fang, Bin Fu, Pei Cheng, and Gang Yu. Ella: Equip diffusion models with llm for enhanced semantic alignment. *arXiv preprint arXiv:2403.05135*, 2024.

Max Klabunde, Tassilo Wald, Tobias Schumacher, Klaus Maier-Hein, Markus Strohmaier, and Florian Lemmerich. Resi: A comprehensive benchmark for representational similarity measures. In *International Conference on Learning Representations*, volume 2025, pp. 18701–18740, 2025.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *International conference on machine learning*, pp. 3519–3529. PMlR, 2019.

Grace Luo, Lisa Dunlap, Dong Huk Park, Aleksander Holynski, and Trevor Darrell. Diffusion hyperfeatures: Searching through time and space for semantic correspondence. *Advances in Neural Information Processing Systems*, 36:47500–47510, 2023.

Wojciech Masarczyk, Mateusz Ostaszewski, Ehsan Imani, Razvan Pascanu, Piotr Miłoś, and Tomasz Trzcinski. The tunnel effect: Building data representations in deep neural networks. *Advances in Neural Information Processing Systems*, 36:76772–76805, 2023.

Ari Morcos, Maithra Raghu, and Samy Bengio. Insights on representational similarity in neural networks with canonical correlation. *Advances in neural information processing systems*, 31, 2018.

Thao Nguyen, Maithra Raghu, and Simon Kornblith. Do wide and deep networks learn the same things? uncovering how neural network representations vary with width and depth. *arXiv preprint arXiv:2010.15327*, 2020.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

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

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

Maithra Raghu, Justin Gilmer, Jason Yosinski, and Jascha Sohl-Dickstein. Svcca: Singular vector canonical correlation analysis for deep learning dynamics and interpretability. *Advances in neural information processing systems*, 30, 2017.

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

Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *Advances in neural information processing systems*, 29, 2016.

Nikunj Saunshi, Nishanth Dikkala, Zhiyuan Li, Sanjiv Kumar, and Sashank J Reddi. Reasoning with latent thoughts: On the power of looped transformers. In *International Conference on Learning Representations*, volume 2025, pp. 14855–14881, 2025.

Shashank Shekhar, Florian Bordes, Pascal Vincent, and Ari Morcos. Objectives matter: Understanding the impact of self-supervised objectives on vision transformer representations. *arXiv preprint arXiv:2304.13089*, 2023.

Oriane Siméoni, Huy V Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, et al. Dinov3. *arXiv preprint arXiv:2508.10104*, 2025.

Stability AI. Stable Diffusion 3.5 Large model card. Hugging Face model card, 2024a. URL `https://huggingface.co/stabilityai/stable-diffusion-3.5-large`.

Stability AI. Stable Diffusion 3.5 Medium model card. Hugging Face model card, 2024b. URL `https://huggingface.co/stabilityai/stable-diffusion-3.5-medium`.

Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

Sixian Wang, Zhiwei Tang, and Tsung-Hui Chang. Diffusion sampling path tells more: An efficient plug-and-play strategy for sample filtering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 36030–36039, 2026.

Alex H Williams, Erin Kunz, Simon Kornblith, and Scott Linderman. Generalized shape metrics on neural representations. *Advances in neural information processing systems*, 34:4738–4750, 2021.

Kevin Xu and Issei Sato. On expressive power of looped transformers: Theoretical analysis and enhancement via timestep encoding. *arXiv preprint arXiv:2410.01405*, 2024.

Lihe Yang, Shang-Wen Li, Yang Li, Xinjie Lei, Dong Wang, Abdelrahman Mohamed, Saining Xie, Hengshuang Zhao, Kaiming He, and Hu Xu. In pursuit of pixel supervision for visual pre-training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 31974–31984, 2026.

Liu Yang, Kangwook Lee, Robert Nowak, and Dimitris Papailiopoulos. Looped transformers are better at learning learning algorithms. In *International conference on learning representations*, volume 2024, pp. 42195–42214, 2024.

Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? *Advances in neural information processing systems*, 27, 2014.

Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pp. 818–833. Springer, 2014.

A LIMITATIONS

Our findings establish an empirical association between pathway consistency and model behavior, but do not imply that consistency is causal. REPAC also relies on spatially aligned patch tokens and may be sensitive to the selected states, feature layers, grid alignment, and pathway weighting. The geometric analysis measures proximity to high-CFG class references in Inception-v3 space, not to the real-data manifold, and may favor class-typical samples over diversity.

B IMPLEMENTATION DETAILS

Feature extraction and pathway construction. For discriminative analysis, we average-pool patch features spatially and train a linear probe on the pooled representations. We use the final block for DINOv2, DINOv3, DistillDIFT, and Pixio; block 14 for DiT-XL/2; and the first residual cross-attention unit in the first post-bottleneck upsampling block for SDXL-Turbo. With diffusion backbones, we encode each image into the latent space, perturb the latent and extract denoiser features at the fixed diffusion timestep $t = 100$ under their respective 1,000-step noise schedules: linear for DiT-XL/2 and scaled-linear for SDXL-Turbo. Layer-wise pathways contain the spatial outputs of all transformer blocks, or the selected U-Net blocks for SDXL-Turbo. For DiT generation, we construct these pathways at sampling steps $s \in \{50, 100, 125, 150, 200, 250\}$. Temporal pathways track block 14 of DiT-XL/2 every 10 sampling steps and blocks 12 and 19 of Stable Diffusion 3.5 Medium and Large, respectively, at every step.

Generation setup. We generate 50,000 class-conditional ImageNet samples with DiT-XL/2 at 256×256 resolution using 250 sampling steps and CFG scale 1.5. For DPG-Bench, Stable Diffusion 3.5 Medium and Large generate 1024×1024 images using 30 steps and CFG scale 7.0. For each prompt, we sample 20 candidates, evaluate them at steps $3, 6, 12, 18, 24$, and complete the top four.

C ADDITIONAL PATHWAY CONSISTENCY DYNAMICS

In the main paper, we analyze the layer-wise consistency dynamics of DINOv3 and observe a common evolution pattern across ImageNet classes: although different classes exhibit different consistency magnitudes, their adjacent consistency trajectories follow similar trends along network depth. Here, we further extend this analysis to additional discriminative and generative models.

Figure 8 presents the layer-wise pathway consistency dynamics of additional backbones. Across different architectures, we observe similar behavior to the results reported in Section 3.5. The local consistency trajectory is not random across layers; instead, adjacent transformations exhibit structured evolution patterns shared by different samples and classes, while still maintaining a certain degree of variance that reflects sample-specific and class-dependent diversity.

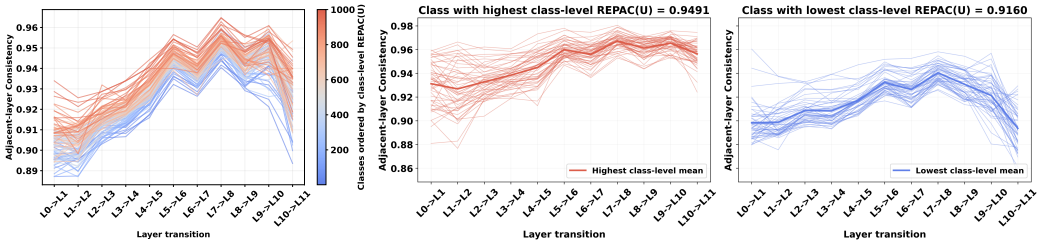
In particular, early pathway transitions generally exhibit larger variations, while later transformations tend to become comparatively stable. This suggests that early layers perform larger representation reorganizations, whereas deeper layers progressively refine already structured representations. The consistency dynamics are therefore not only reflected in the aggregated REPAC (LD) score but also emerge as a general property of representation evolution across different visual models.

D ADDITIONAL ANALYSIS OF REPRESENTATION GEOMETRY

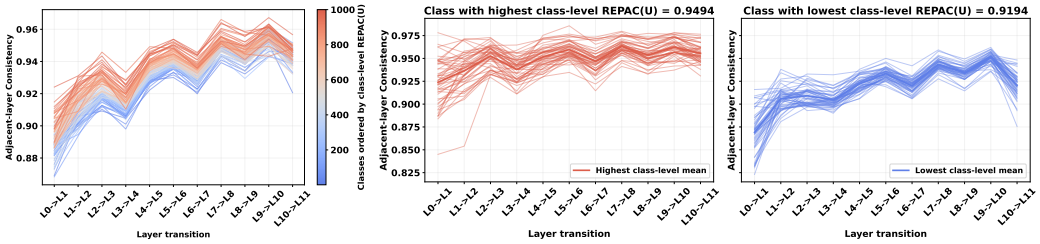
In Section 4.6, we investigate the relationship between pathway consistency and representation geometry using cosine similarity between generated samples and their corresponding class prototypes in an external Inception-v3 feature space. Here, we provide an additional analysis using Euclidean distance instead. For each generated sample, we compute the distance between its feature representation and the corresponding class prototype:

$$d_i = \|z_i - \mu_{y_i}\|_2, \tag{9}$$

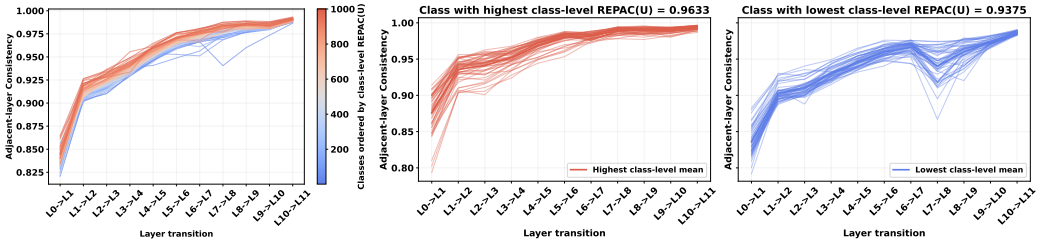
where z_i denotes the Inception-v3 feature of sample x_i, and μ_{y_i} is the prototype of its corresponding class. Unlike cosine similarity, smaller Euclidean distance indicates that a sample is closer to its class prototype.



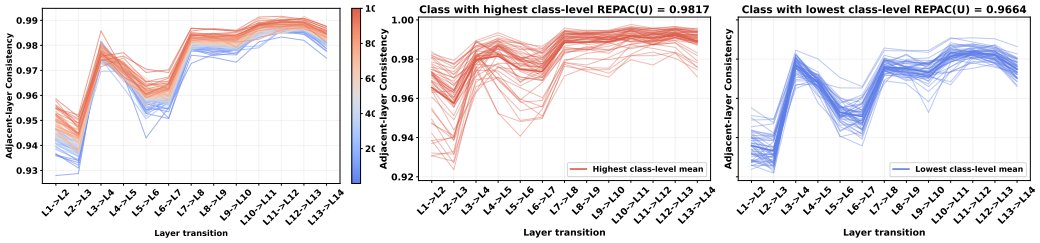
(a) Consistency dynamics across DINOv2 depth.



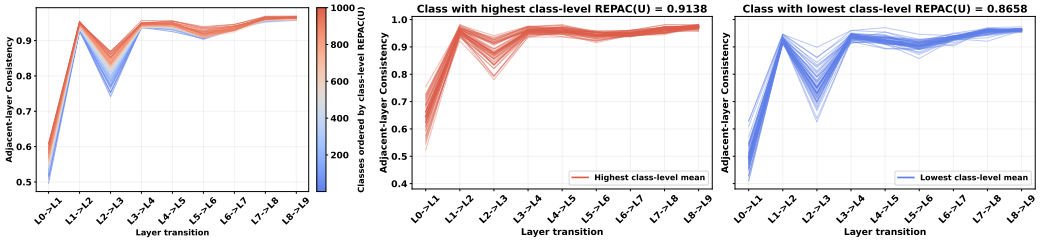
(b) Consistency dynamics across DistillDIFT depth.



(c) Consistency dynamics across Pixio depth.



(d) Consistency dynamics across DiT-XL/2 depth.



(e) Consistency dynamics across SDXL-Turbo depth.

Figure 8: **Additional pathway consistency dynamics across models.** Layer-wise local consistency trajectories exhibit similar evolution patterns to those observed in the main paper. Different models and classes mainly differ in consistency magnitude while maintaining structured depth-wise dynamics. Early transitions also usually exhibit greater variation than later ones.

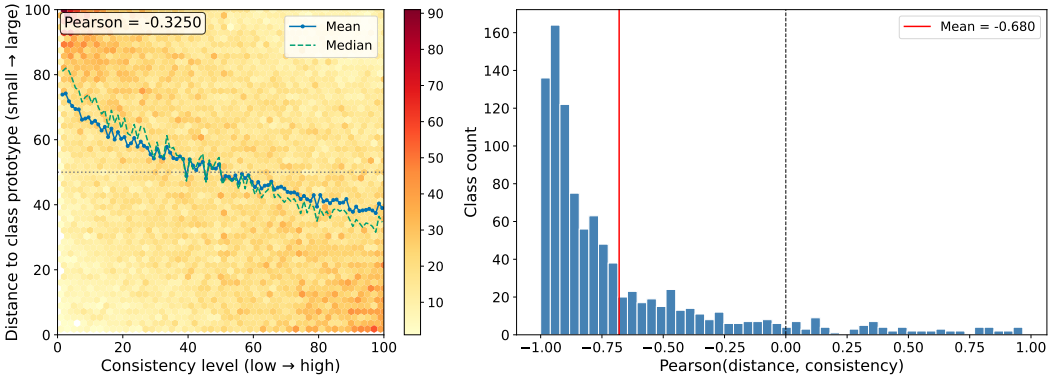


Figure 9: **Representation geometry analysis using Euclidean distance.** Higher pathway consistency corresponds to smaller distances between samples and their class prototypes, consistent with the cosine similarity analysis in the main paper.

Figure 9 shows the relationship between REPAC and prototype distance. The results are consistent with the cosine similarity analysis in the main paper. Samples with higher pathway consistency tend to have smaller distances to their corresponding class prototypes. Therefore, the geometric association is independent of the specific similarity metric: whether measured by increasing cosine similarity or decreasing Euclidean distance, more consistent representation pathways correspond to samples located closer to class-conditional prototype regions.

E ABLATION STUDY FOR EARLY QUALITY ASSESSMENT

We further analyze the design choices of early quality assessment by comparing different pathway aggregation strategies and baseline selection signals. Specifically, we evaluate three temporal REPAC variants with different weighting schemes: uniform weighting (U), linearly decaying weighting (LD), and linearly increasing weighting (LI) to emphasize different stages of transition. In addition, we compare temporal REPAC with a REPAC variant for depth pathway (DP), where consistency is computed along denoiser blocks at a fixed sampling step (assessment point). For temporal REPAC variants, we evaluate internal pathway positions corresponding to $1/2$, $3/4$, and the final (1) relative depth locations. We additionally include two baseline strategies: random selection and ASD. All methods are evaluated under the same candidate generation and selection protocol described in Section 4.5.

Table 2 summarizes the comparison. Temporal REPAC variants consistently provide useful early quality signals, while different weighting strategies exhibit different sensitivities to pathway stages. The uniform and linearly decaying variants generally preserve the strongest association with final sample quality, suggesting that aggregating consistency across the observed trajectory is beneficial. In contrast, linearly increasing weighting emphasizes later transitions and provides a complementary view of pathway evolution. Layer-wise REPAC also demonstrates that internal representations contain quality-related information before generation is completed. However, the temporal pathway directly captures the evolution process across sampling steps and therefore provides a more comprehensive characterization of representation dynamics. Overall, these ablations support the robustness of pathway consistency as a training-free signal for early quality assessment.

F ADDITIONAL VISUAL EXAMPLES OF PATHWAY CONSISTENCY

The main paper provides qualitative examples showing that samples with higher pathway consistency generally exhibit clearer semantic structures and more coherent visual appearance. Here, we extend this visualization to additional models, categories, and generation settings.

Figure 10- 12 presents additional high- and low-REPAC examples from both discriminative and generative tasks. Similar to the observations in Section 4.2, high-consistency samples tend to con-

Table 2: **Ablation study of early quality assessment strategies.** Comparison among temporal REPAC variants, layer-wise REPAC, ASD, and random selection.

		SD3.5-Medium						SD3.5-Large					
Method	Depth	Attr.	Entity	Global	Other	Relation	Overall	Attr.	Entity	Global	Other	Relation	Overall
10% Sampling Process													
ASD	–	90.42	89.87	90.34	88.07	90.64	84.80	88.81	90.18	85.25	90.14	**91.62**	84.53
Random	–	90.22	89.60	88.84	89.60	89.37	84.30	89.69	90.16	87.30	89.19	91.20	84.26
REPAC (DP)	–	89.69	90.06	87.72	86.49	90.95	83.97	89.48	89.62	88.05	90.17	90.59	84.07
REPAC (U)	1/2	90.00	89.39	88.78	91.62	91.19	85.24	89.44	90.06	87.77	90.54	91.00	85.23
	3/4	89.62	90.27	91.35	89.24	90.68	85.04	90.19	90.35	83.56	90.52	91.02	85.48
	1	89.62	89.90	90.97	91.45	89.12	84.58	89.87	89.05	**91.70**	88.99	89.66	83.98
REPAC (LD)	1/2	91.10	89.50	91.37	89.12	89.65	**85.39**	**91.23**	89.03	90.82	88.55	89.41	**85.53**
	3/4	89.72	90.24	89.45	89.96	**91.61**	85.22	89.17	**91.04**	87.50	**90.63**	90.03	85.16
	1	**91.26**	90.08	88.47	88.31	85.25	84.63	90.70	89.94	86.87	84.39	88.63	83.79
REPAC (LI)	1/2	90.22	**90.67**	89.90	**91.67**	87.79	85.13	89.81	89.59	91.63	89.97	90.20	84.85
	3/4	90.37	89.88	**91.74**	89.45	89.00	84.87	90.44	90.29	87.18	86.86	90.59	85.03
	1	89.95	89.80	89.18	91.62	89.20	84.29	90.69	89.91	90.62	86.44	89.66	84.51
20% Sampling Process													
ASD	–	89.69	91.02	90.83	87.97	89.78	85.00	90.66	90.18	89.76	86.88	87.38	84.07
Random	–	87.80	89.71	92.19	90.00	91.54	84.30	**90.82**	88.05	88.82	90.02	89.70	84.26
REPAC (DP)	–	89.64	89.40	83.09	88.35	91.67	84.32	89.41	89.99	88.27	88.17	90.67	84.34
REPAC (U)	1/2	**90.86**	89.80	85.94	**91.91**	88.58	85.55	88.65	90.12	90.78	89.32	91.01	85.01
	3/4	90.27	91.14	87.68	89.61	89.98	84.88	89.29	90.25	90.16	89.72	**91.60**	84.67
	1	88.97	**91.32**	90.48	85.56	91.50	85.28	90.34	89.17	90.35	89.32	89.36	84.42
REPAC (LD)	1/2	90.80	89.57	89.43	89.11	91.68	**85.60**	90.72	**90.29**	88.01	90.02	87.70	**85.05**
	3/4	88.77	89.87	87.64	90.72	**92.78**	85.11	89.98	89.28	88.78	**92.02**	90.48	85.00
	1	90.13	89.70	**92.69**	90.60	88.83	84.91	88.88	90.05	**92.88**	87.41	91.11	84.59
REPAC (LI)	1/2	89.19	90.64	92.28	90.88	87.53	85.20	90.48	87.90	91.42	90.08	89.06	84.88
	3/4	90.61	89.68	90.78	87.81	89.24	84.70	90.31	89.64	90.38	89.72	89.04	84.92
	1	90.45	88.03	90.15	89.31	91.67	84.99	90.70	87.75	89.99	89.34	90.93	84.46

tain clearer class-specific structures, while low-consistency samples often exhibit more ambiguous or fragmented visual patterns. For discriminative models, this tendency appears across different architectures and ImageNet categories, suggesting that pathway consistency captures properties of representation formation rather than being specific to a single backbone. For generative models, samples with higher temporal pathway consistency generally show stronger semantic alignment and more coherent object structures.

Together with the quantitative analyses, these qualitative examples further demonstrate that representation pathway consistency provides a consistent perspective for understanding representation evolution across different tasks and model families.



(a) Extreme-REPAC ImageNet strawberry samples selected by DINOv2 layer-wise consistency.



(b) Extreme-REPAC ImageNet strawberry samples selected by DistillDIFT layer-wise consistency.



(c) Extreme-REPAC ImageNet strawberry samples selected by Pixio layer-wise consistency.



(d) Extreme-REPAC ImageNet strawberry samples selected by DiT-XL/2 layer-wise consistency.



(e) Extreme-REPAC ImageNet strawberry samples selected by SDXL-Turbo layer-wise consistency.

Figure 10: **Images at the extremes of pathway consistency.** High-REPAC samples generally show clearer class subjects and simpler, more coherent visual structure than low-REPAC samples.



(a) Extreme-REPAC ImageNet yurt samples selected by DINOv2 layer-wise consistency.



(b) Extreme-REPAC ImageNet yurt samples selected by DistillDIFT layer-wise consistency.



(c) Extreme-REPAC ImageNet yurt samples selected by DistillDIFT layer-wise consistency.



(d) Extreme-REPAC ImageNet yurt samples selected by Pixio layer-wise consistency.

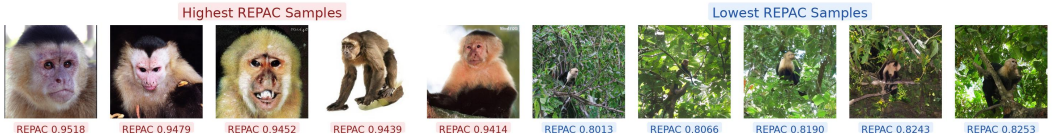


(e) Extreme-REPAC ImageNet yurt samples selected by DiT-XL/2 layer-wise consistency.



(f) Extreme-REPAC ImageNet yurt samples selected by SDXL-Turbo layer-wise consistency.

Figure 11: **Images at the extremes of pathway consistency.** High-REPAC samples generally show clearer class subjects and simpler, more coherent visual structure than low-REPAC samples.



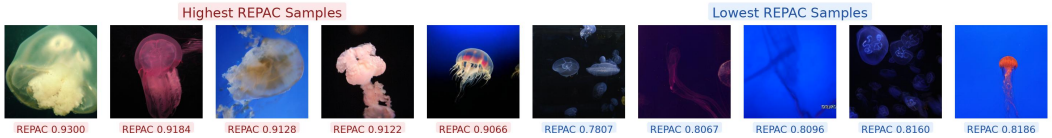
(a) Extreme-REPAC DiT-XL/2 capuchin generations selected by temporal consistency.



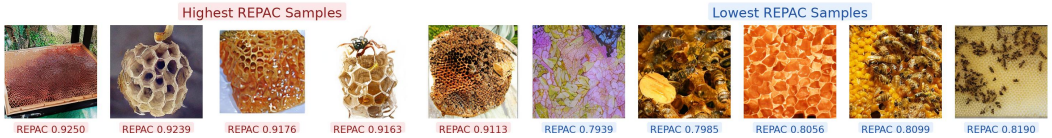
(b) Extreme-REPAC DiT-XL/2 strawberry generations selected by temporal consistency.



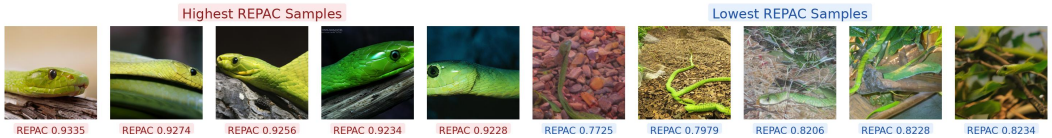
(c) Extreme-REPAC DiT-XL/2 puck generations selected by temporal consistency.



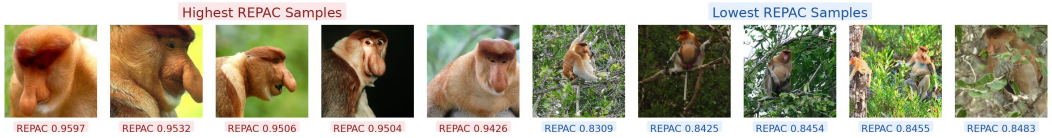
(d) Extreme-REPAC DiT-XL/2 jellyfish generations selected by temporal consistency.



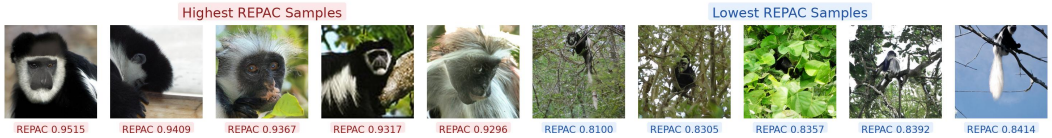
(e) Extreme-REPAC DiT-XL/2 honeycomb generations selected by temporal consistency.



(f) Extreme-REPAC DiT-XL/2 green mamba generations selected by temporal consistency.



(g) Extreme-REPAC DiT-XL/2 proboscis monkey generations selected by temporal consistency.



(h) Extreme-REPAC DiT-XL/2 colobus generations selected by temporal consistency.

Figure 12: **Images at the extremes of pathway consistency.** High-REPAC samples generally show clearer class subjects and simpler, more coherent visual structure than low-REPAC samples.