T-Stitch: Accelerating Sampling in Pre-Trained Diffusion Models with Trajectory Stitching

Zizheng Pan [1], Bohan Zhuang [1], De-An Huang [2], Weili Nie [2], Zhiding Yu [2]
Chaowei Xiao [2 3], Jianfei Cai [1], Anima Anandkumar [4]

[1]Monash University [2]NVIDIA [3]University of Wisconsin, Madison [4]Caltech

1. Background - Generative Models

Text-to-Image

Stable Diffusion 3 [1]

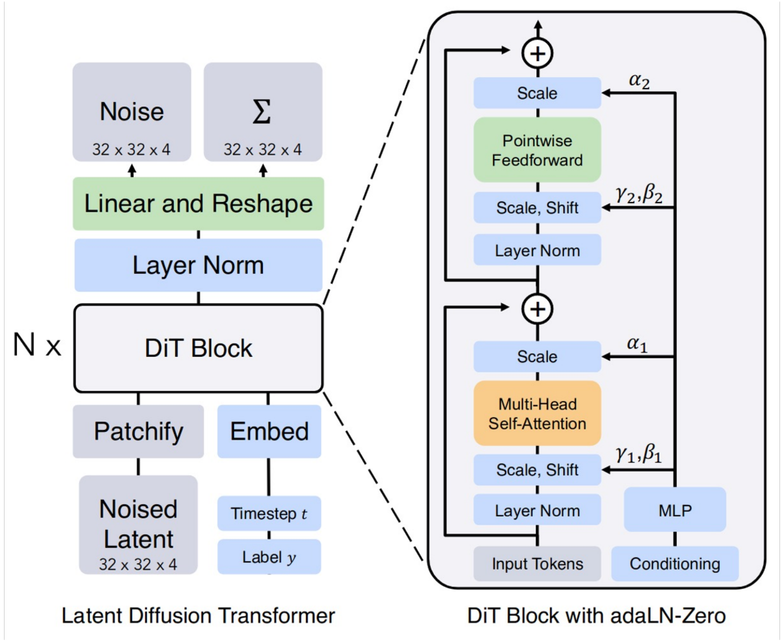
Text-to-Video

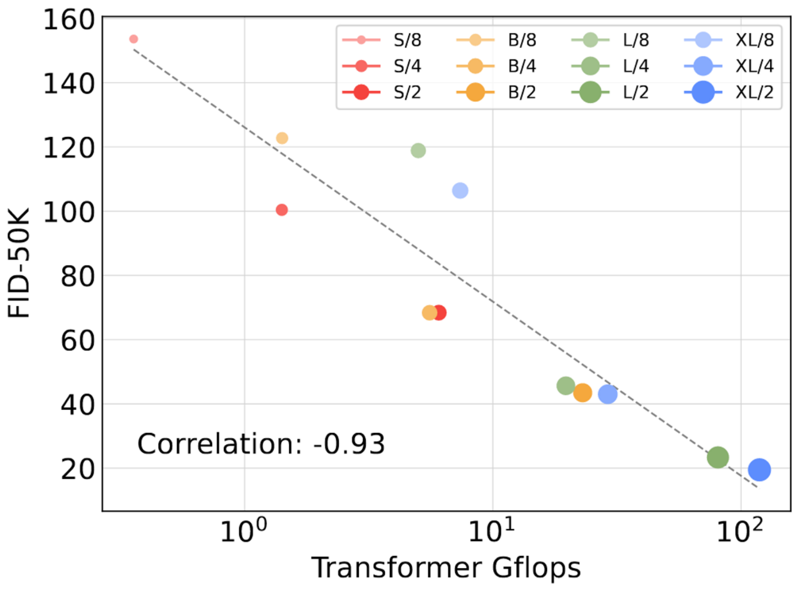
Sora [2]

[1] https://stability.ai/news/stable-diffusion-3
[2] https://openai.com/sora

1. Background - Diffusion Transformer

Behind the scene





Diffusion Transformer **(DiT)**

Larger model, **better** quality

Peebles, William, and Saining Xie. "Scalable diffusion models with transformers." *ICCV*. 2023.

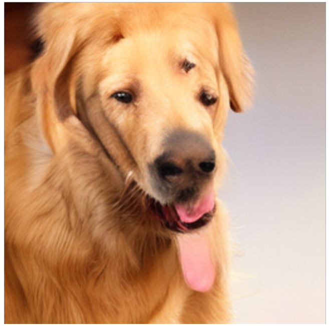
1. Background - Diffusion Transformer

However, large model comes with high computational cost.

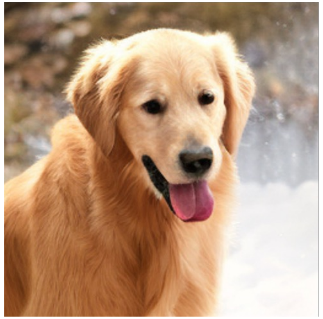
The speed-quality trade-off

Name	Params	FID-50K	Time Cost
DiT-XL	675M	2.27	43s
DiT-S	33M	21.47	4s

- **Sampling Steps:** 250
- **Images:** 8
- **GPU:** RTX 3090



DiT-S



DiT-XL

2. Related Works

How existing works accelerate image diffusion models?

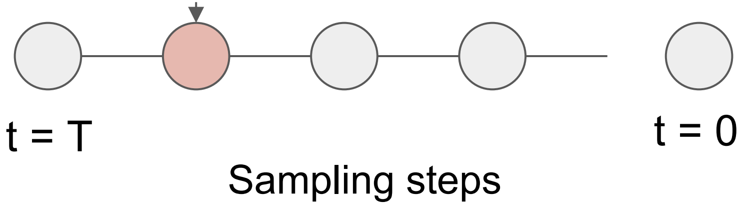
1. Reducing costs per step

- Model quantization.
  E.g. Q-diffusion [1]

- Network pruning.
  E.g., Structured pruning. [2]

- Lightweight architecture design.
  E.g., SnapFusion [3]

- Cache-based method.
  E.g. DeepCache [4]



[1] Li, Xiuyu, et al. "Q-diffusion: Quantizing diffusion models." *ICCV* (2023).
[2] Fang, Gongfan, Xinyin Ma, and Xinchao Wang. "Structural pruning for diffusion models." *NeurIPS* (2024).
[3] Li, Yanyu, et al. "Snapfusion: Text-to-image diffusion model on mobile devices within two seconds." *NeurIPS* (2024).
[4] Ma, Xinyin, Gongfan Fang, and Xinchao Wang. "Deepcache: Accelerating diffusion models for free." *CVPR* (2024).

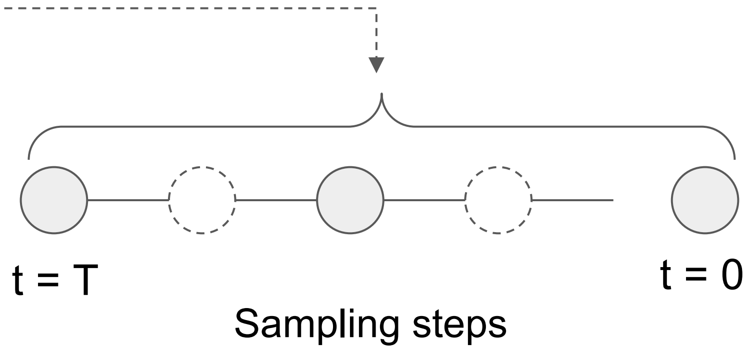
2. Related Works

How existing works accelerate image diffusion models?

2. Reducing total sampling steps

- Advanced samplers.
  E.g., DPM-Solver [1].

- Distilling into fewer steps.
  E.g., Progressive step distillation [2].

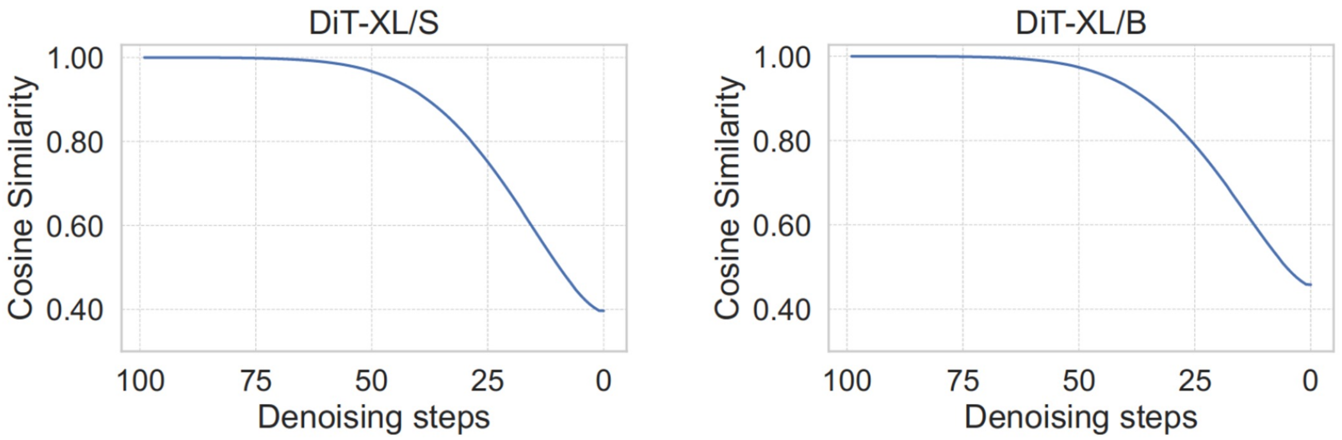


[1] Lu, Cheng, et al. "Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps." *NeurIPS* (2022)
[2] Salimans, Tim, and Jonathan Ho. "Progressive distillation for fast sampling of diffusion models." ICLR (2022).

3. Method - Our Motivation

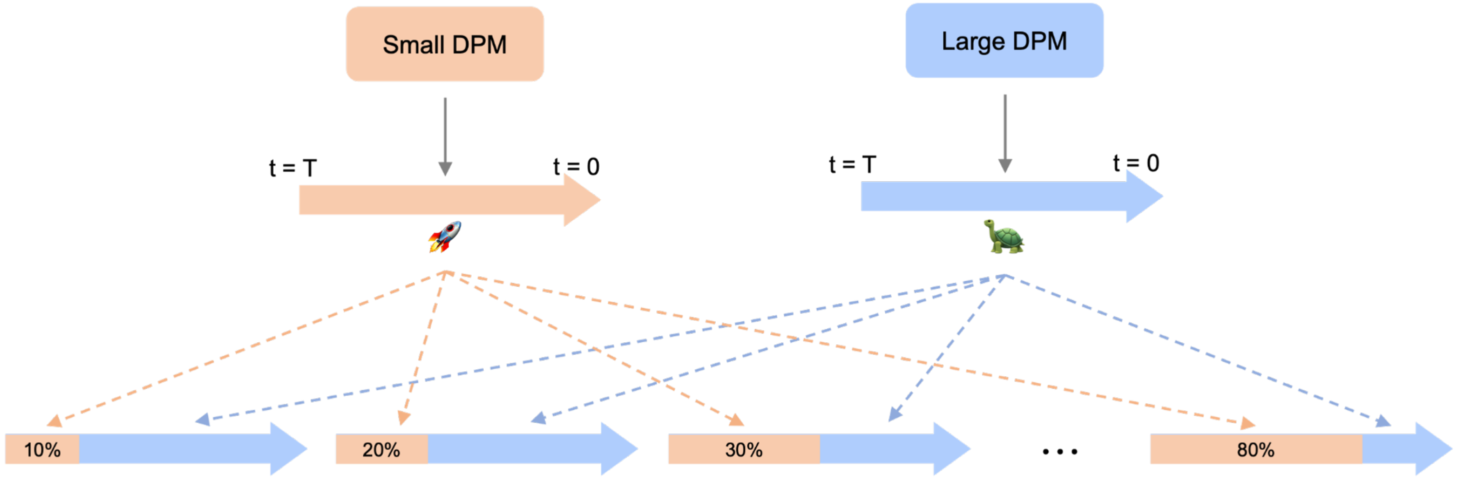
1. Generative models trained on the **same data distribution** share a **common latent space**.

2. **Small** models can generate **highly similar latents at early steps** as the **large** model!



Similarity comparison of latent embeddings at different denoising steps between different DiT models.

3. Method - Our Approach



The Proposed Trajectory Stitching (T-Stitch)

Core idea: Applying DPMs of different sizes at different denoising steps instead of using the same model at all steps, as in previous works.

4. Experiments - Quick Overview

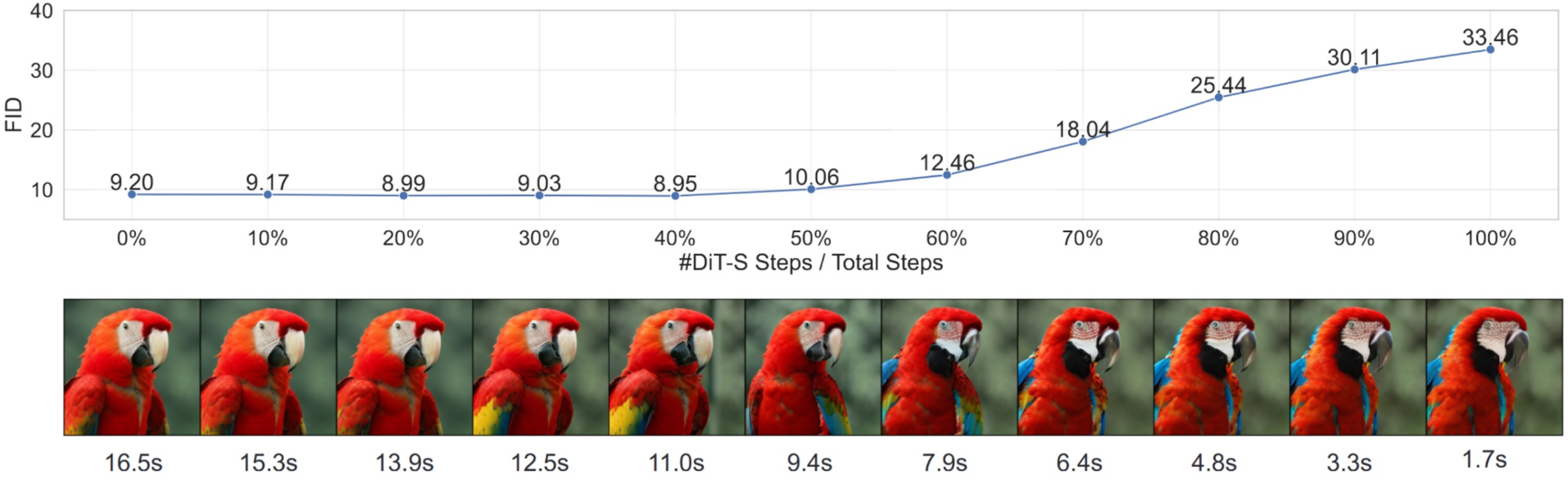
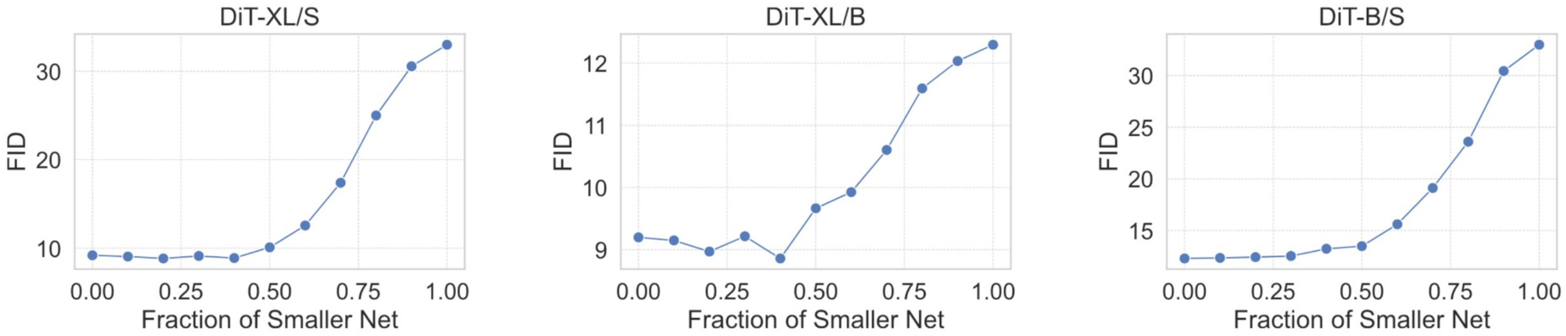


Figure 1. **Top:** FID comparison on class-conditional ImageNet when progressively stitching more DiT-S steps at the beginning and fewer DiT-XL steps in the end, based on DDIM 100 timesteps and a classifier-free guidance scale of 1.5. FID is calculated by sampling 5000 images. **Bottom:** One example of stitching more DiT-S steps to achieve faster sampling, where the time cost is measured by generating 8 images on one RTX 3090 in seconds (s).

4. Experiments

T-Stitch is compatible with DiTs and U-Nets.

DiTs on ImageNet-256



U-Net on ImageNet-256

Table 1. T-Stitch with LDM (Rombach et al., 2022) and LDM-S on class-conditional ImageNet. All evaluations are based on DDIM and 100 timesteps. We adopt a classifier-free guidance scale of 3.0. The time cost is measured by generating 8 images on one RTX 3090.

Fraction of LDM-S	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
FID	20.11	19.54	18.74	18.64	18.60	19.33	21.81	26.03	30.41	35.24	40.92
Inception Score	199.24	201.87	202.81	204.01	193.62	175.62	140.69	110.81	90.24	70.91	54.41
Time Cost (s)	7.1	6.7	6.2	5.8	5.3	4.9	4.5	4.1	3.6	3.1	2.9

4. Experiments

T-Stitch is complementary to reducing sampling steps and advanced samplers.

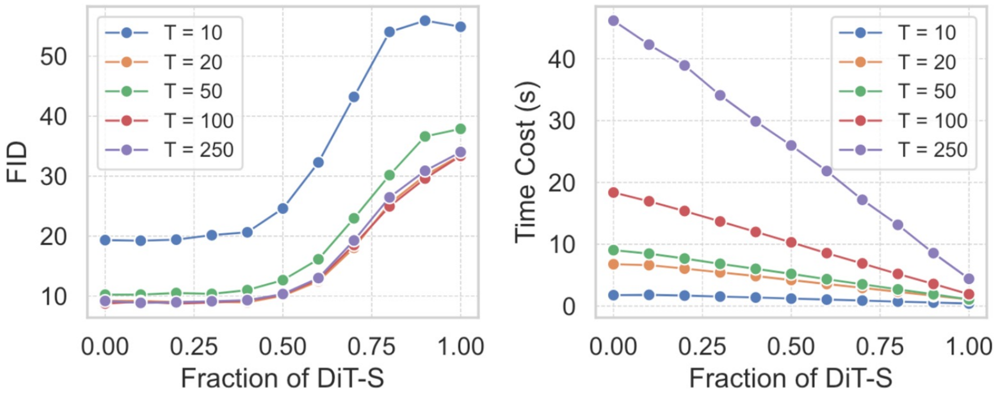
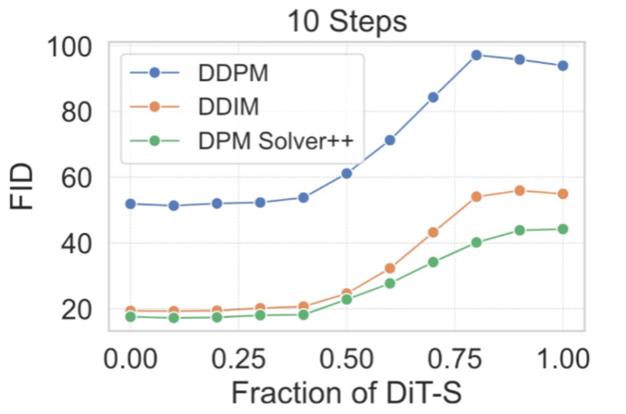


Figure 9. **Left:** We compare FID between different numbers of steps. **Right:** We visualize the time cost of generating 8 images under different number of steps, based on DDIM and a classifier-guidance scale of 1.5. "T" denotes the number of sampling steps.



T-Stitch with different sampler

Thanks!

Paper

Code released! 🌟