

# Denoising as Path Planning: Training-Free Acceleration of Diffusion Models with DPCache

## Supplementary Material

The supplementary materials include two files:

- *supp.pdf* provides additional experimental settings, quantitative and qualitative results, ablation studies, and discussion on limitations of DPCache.
- *video.mp4* shows qualitative comparisons of different acceleration methods applied to HunyuanVideo on VBench dataset, as well as side-by-side visualizations of the original and DPCache-accelerated outputs on Wan-2.1-I2V-14B on VBench-I2V benchmark.

### 1. Experimental Settings

To ensure reproducibility and fair comparison across methods, Table 1 provides a detailed overview of the hyperparameter settings used for each method in Section 4.2 of the main paper.

Table 1. Hyperparameters used for each method in our experiments.

Model	Method	Hyperparameters
FLUX.1-dev	TeaCache <sup>1</sup>	$l = 0.7$
	TeaCache <sup>2</sup>	$l = 1.0$
	TaylorSeer <sup>1</sup>	$\mathcal{N} = 5, O = 2$
	TaylorSeer <sup>2</sup>	$\mathcal{N} = 8, O = 2$
	SpeCa <sup>1</sup>	$\tau = 1.0, \beta = 0.2$
	SpeCa <sup>2</sup>	$\tau = 8.5, \beta = 0.4$
HunyuanVideo	TeaCache <sup>1</sup>	$l = 0.4$
	TeaCache <sup>2</sup>	$l = 0.5$
	TaylorSeer <sup>1</sup>	$\mathcal{N} = 5, O = 1$
	TaylorSeer <sup>2</sup>	$\mathcal{N} = 8, O = 1$
	SpeCa <sup>1</sup>	$\tau = 1.0, \beta = 0.1$
	SpeCa <sup>2</sup>	$\tau = 2.0, \beta = 0.1$
DiT-XL/2	TaylorSeer <sup>1</sup>	$\mathcal{N} = 9, O = 2$
	TaylorSeer <sup>2</sup>	$\mathcal{N} = 10, O = 2$
	SpeCa <sup>1</sup>	$\tau = 1.0, \beta = 0.1$
	SpeCa <sup>2</sup>	$\tau = 2.0, \beta = 0.3$

### 2. Comparison with State-of-the-Art Methods

#### 2.1. Analysis of Sampling Trajectories

To visually demonstrate the superiority of our method, we visualize the sampling trajectories of FLUX.1-dev under different acceleration strategies, as shown in Figure 1. Specifically, we run 50 samples, compute the average output features at each timestep, apply PCA to project them

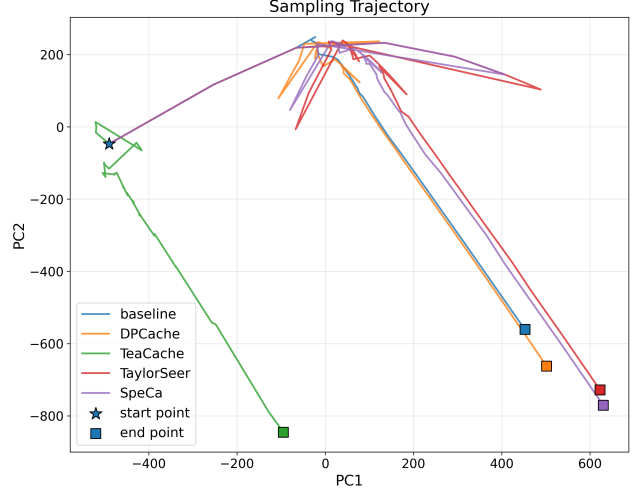


Figure 1. Visualization of sampling trajectories when applying different acceleration methods to FLUX.1-dev.

into a two-dimensional space, and visualize the resulting trajectories in a 2D coordinate system. DPCache’s trajectory closely follows that of the baseline, confirming that our method minimizes global trajectory deviation, which is consistent with Figure 1 of the main paper. Moreover, we notice that prediction timesteps often deviate significantly from the baseline trajectory, but subsequent computation steps effectively steer the trajectory back toward it. By selecting computation timesteps from a global perspective, DPCache ensures that the accelerated model remains consistently close to the original sampling trajectory throughout inference.

#### 2.2. Quantitative Results

**VBench Score for all dimensions.** Figure 2 presents a comprehensive comparison of HunyuanVideo’s performance across all VBench dimensions with various acceleration methods. Under different acceleration settings, our proposed DPCache consistently achieves competitive performance across all dimensions, and notably outperforms other methods in aesthetic quality, imaging quality, and multiple objects. Our method demonstrates more pronounced advantages under high-speedup settings, highlighting the potential of our globally optimized sampling schedule to maintain quality while enabling aggressive acceleration.

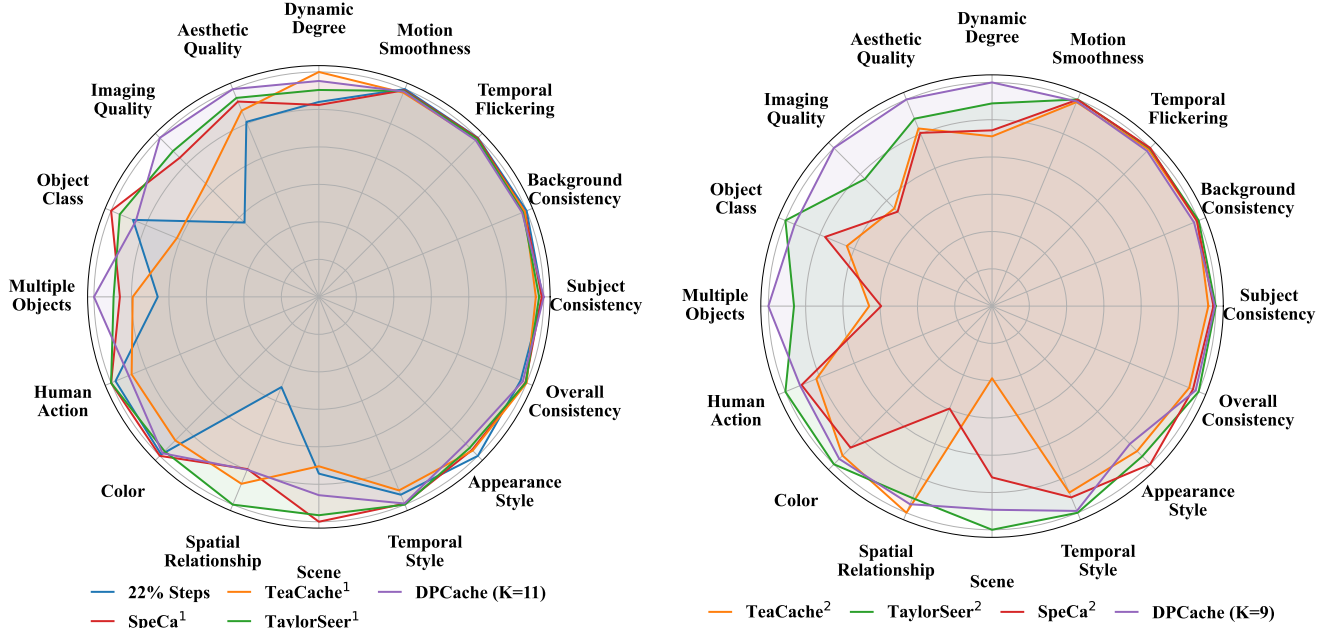


Figure 2. VBench performance of HunyuanVideo across all dimensions with various acceleration methods under different acceleration settings: (left) low speedup, (right) high speedup. Scores are normalized per dimension for improved visual comparison.

### 2.3. Qualitative Results

**Text-to-image generation.** We provide additional qualitative comparisons of various acceleration methods applied to FLUX.1-dev on the DrawBench dataset in Figure 3. Existing approaches often exhibit noticeable artifacts, such as blurriness, geometric distortions, or misalignment with the input text prompt, whereas our proposed DPCache maintains high visual fidelity and closely resembles the output of the unaccelerated baseline.

**Text-to-video generation.** To further validate the effectiveness of DPCache, we apply it to Wan-2.1-I2V-14B [2], which is a larger, higher-resolution model with state-of-the-art performance in video generation. As shown in Figure 4, we sample some inputs from VBench-I2V [1] benchmark and compare the generation results of the baseline and DPCache at a resolution of  $832 \times 832 \times 81$ . DPCache achieves  $4.15 \times$  speedup while preserving visual fidelity: the accelerated output maintains sharpness and detail, even in complex scenes, without introducing noticeable artifacts.

*Qualitative comparisons with other methods on HunyuanVideo, as well as side-by-side visualizations of the original and DPCache-accelerated outputs on WAN-2.1-I2V-14B, are provided in the file named video.mp4.*

**Robustness on out-of-distribution prompts.** In Section 4.3 of the main paper, we quantitatively demonstrate that DPCache is robust to the choice and size of the calibration set. To further validate this robustness, we apply a FLUX.1-dev model calibrated on 10 samples from Draw-

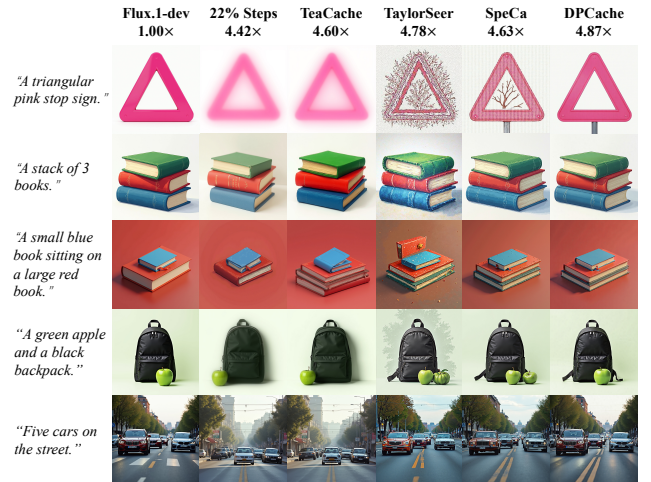


Figure 3. Additional qualitative comparisons of different acceleration methods applied to FLUX.1-dev on DrawBench dataset.

Bench, and evaluate it on out-of-distribution prompts from PartiPrompts [3] that exhibit a significantly different distribution from DrawBench. As shown in Figure 5, we compare different acceleration methods on out-of-distribution prompts.

In the first row, the prompt is short and abstract. The result generated by TeaCache deviates significantly from the baseline, indicating a complete drift from the original sampling trajectory. TaylorSeer and SpeCa produce scenes that are semantically similar to the baseline but introduce spuri-

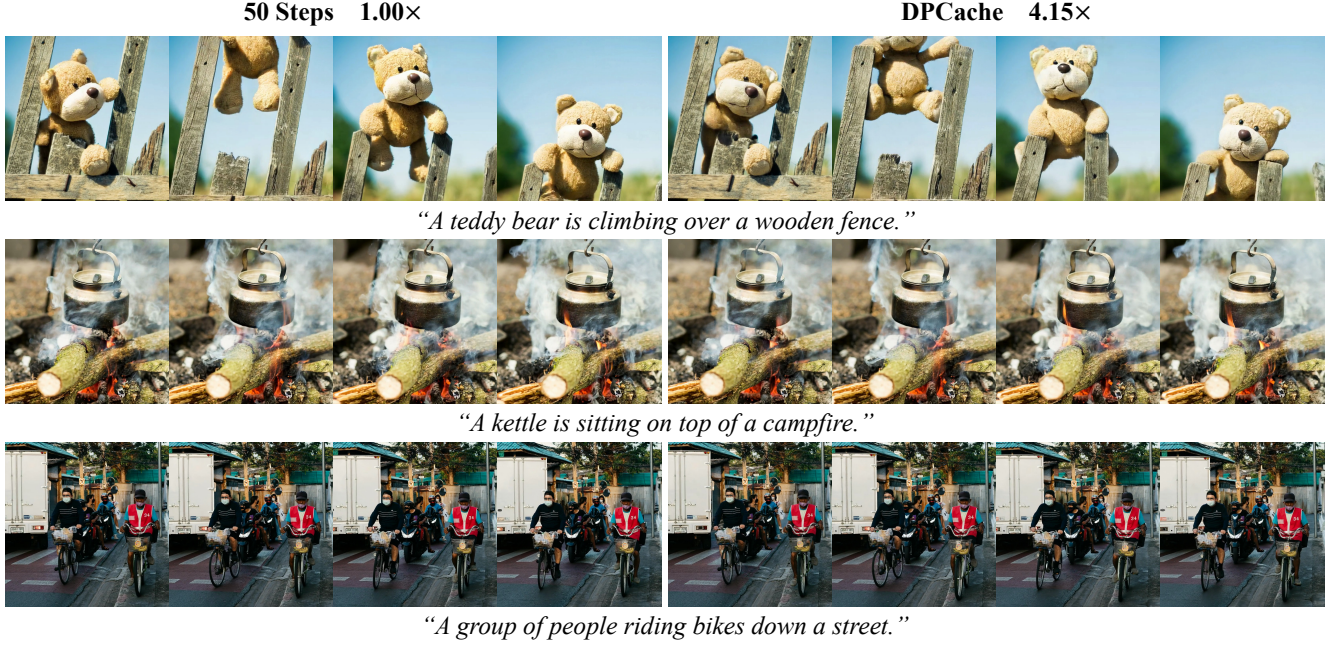


Figure 4. Qualitative comparison between the original (unaccelerated) and DPCache-accelerated Wan-2.1-I2V-14B on the VBench-I2V benchmark.

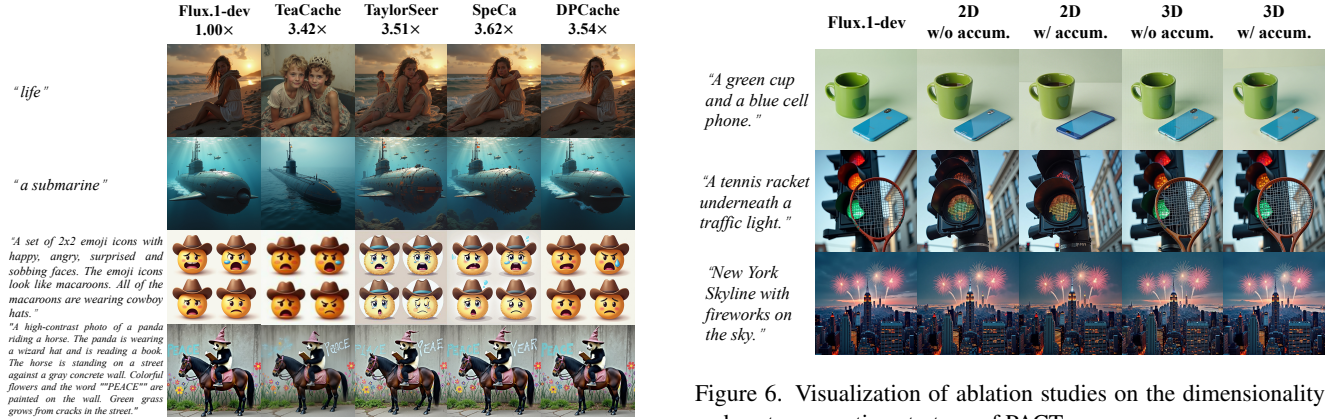


Figure 5. Qualitative comparison of different acceleration methods applied to FLUX.1-dev on out-of-distribution prompts.

ous human figures and distorted structures. In contrast, DPCache reproduces a result nearly identical to the baseline. In the fourth row, the prompt is highly detailed and specifies numerous visual elements. TeaCache and TaylorSeer fail to preserve the text on the wall, while SpeCa generates noticeable artifacts on the wizard hat. DPCache, however, maintains the finest details with the highest fidelity.

These results on out-of-distribution inputs strongly demonstrate the robustness of DPCache to calibration set selection, confirming that our method can generalize well even when calibrated on a small and distributionally mismatched set.

### 3. Ablation Studies

**Path-aware cost tensor.** In Section 4.3 of the main paper, we quantitatively validate the effectiveness of the PACT design. Figure 6 further visualizes results under different cost tensor dimensionalities and cost aggregation strategies.

In the first row, the result using a 2D cost tensor without temporal accumulation exhibits noticeable distortions: the cup shape and the logo on the phone differ from the baseline result. When temporal accumulation is applied to the 2D cost, these errors are further amplified, leading to an even greater deviation in the phone’s appearance. In contrast, both variants of the 3D cost tensor achieve high fidelity to



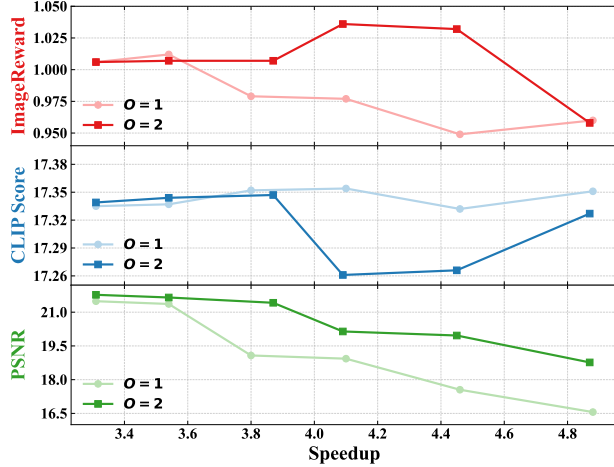


Figure 7. Performance of DPCache under varying acceleration ratios and prediction orders.

the baseline, demonstrating that modeling the influence of the previous key timestep during cost computation better captures the true error of timestep skipping. Moreover, the non-cumulative variants produce outputs with coarser textures and reduced detail. This aligns with our analysis that non-aggregated costs tend to trigger aggressive skipping in the final stages of sampling, impairing fine-grained synthesis. Similar trends are observed in the second and third rows: 2D cost tensors yield lower visual similarity to the baseline, and non-aggregated strategies suffer from significant detail loss.

These results confirm that leveraging a 3D cost tensor with temporal aggregation enables DPCache to estimate skipping-induced error more accurately, thereby preserving structural consistency and fine details during accelerated sampling.

**Hyperparameter analysis.** As shown in Figure 7, we vary the number of sampling steps  $K$  to achieve different speedup ratios, and compare performance under prediction orders  $O = 1$  and  $O = 2$ . Unlike acceleration methods that rely on per-step thresholds or fixed sampling intervals, DPCache directly controls the total number of sampling steps  $K$ , enabling smoother and more predictable management of the speed–quality trade-off. As the speedup ratio increases (*i.e.*,  $K$  decreases), all metrics evolve steadily without abrupt drops, reflecting the stability of the dynamically optimized schedule. With respect to the prediction order, second-order prediction yields higher ImageReward and PSNR than first-order prediction, confirming that higher-order prediction yields more accurate feature extrapolation, which results in generated images of higher quality and closer alignment with the original denoising trajectory. In contrast, first-order prediction achieves a slightly higher CLIP Score, indicating that strict trajectory adherence may come at the expense of alignment with the text prompt.



Figure 8. Failure cases of DPCache.

## 4. Limitations

Although DPCache achieves significant acceleration while preserving generation quality by seeking a globally optimal sampling path, it has certain limitations. Since its optimization objective prioritizes fidelity to the original sampling trajectory, DPCache may sacrifice output diversity and inherit semantic or structural errors from the original model. As shown in the left example of Figure 8, the original model erroneously renders “Image” as “Omage”, and DPCache retains this mistake after acceleration. Moreover, in cases where the unaccelerated model generates plausible but imperfect results, DPCache’s emphasis on trajectory matching can inadvertently amplify minor deviations into visible artifacts. As shown in the right example of Figure 8, the original model introduces spurious debris inside the flowerpot, and DPCache, by closely following the base trajectory, further amplifies this artifact. These observations suggest promising directions for future work. For instance, we could adopt input-adaptive scheduling and integrate learnable predictors to correct failures of the original model during accelerated inference.

## References

- [1] Ziqi Huang, Fan Zhang, Xiaojie Xu, Yinan He, Jiashuo Yu, Ziyue Dong, Qianli Ma, Nattapol Chanpaisit, Chenyang Si, Yuming Jiang, et al. Vbench++: Comprehensive and versatile benchmark suite for video generative models. *arXiv preprint arXiv:2411.13503*, 2024. 2
- [2] Team Wan, Ang Wang, Baole Ai, Bin Wen, Chaojie Mao, Chen-Wei Xie, Di Chen, Fei Wu Yu, Haiming Zhao, Jianxiao Yang, Jianyuan Zeng, Jiayu Wang, Jingfeng Zhang, Jingen Zhou, Jinkai Wang, Jixuan Chen, Kai Zhu, Kang Zhao, Keyu Yan, Lianghua Huang, Mengyang Feng, Ningyi Zhang, Pandeng Li, Pingyu Wu, Ruihang Chu, Ruili Feng, Shiwei Zhang, Siyang Sun, Tao Fang, Tianxing Wang, Tianyi Gui, Tingyu Weng, Tong Shen, Wei Lin, Wei Wang, Wei Wang, Wenmeng Zhou, Wenten Wang, Wenting Shen, Wenyan Yu, Xianzhong Shi, Xiaoming Huang, Xin Xu, Yan Kou, Yangyu Lv, Yifei Li, Yijing Liu, Yiming Wang, Yingya Zhang, Yitong Huang, Yong Li, You Wu, Yu Liu, Yulin Pan, Yun Zheng, Yuntao Hong, Yupeng Shi, Yutong Feng, Zeyinzi Jiang, Zhen Han, Zhi-Fan Wu, and Ziyu Liu. Wan: Open and



advanced large-scale video generative models. *arXiv preprint arXiv:2503.20314*, 2025. [2](#)

- [3] Jiahui Yu, Yuanzhong Xu, Jing Yu Koh, Thang Luong, Gunjan Baid, Zirui Wang, Vijay Vasudevan, Alexander Ku, Yinfei Yang, Burcu Karagol Ayan, et al. Scaling autoregressive models for content-rich text-to-image generation. *arXiv preprint arXiv:2206.10789*, 2(3):5, 2022. [2](#)