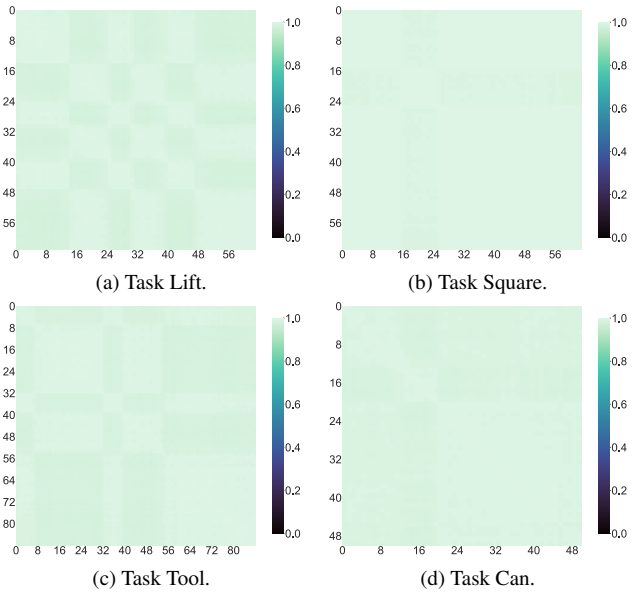
Test-time Sparsity for Extreme Fast Action Diffusion

Supplementary Material

A. Details on Rollout Similarity

Figure 9 provides additional details on rollout-level feature similarity across the Lift, Square, Tool, and Can tasks. We observe that features from different rollout iterations exhibit consistently high similarity on all tasks, reinforcing our insight that cached features from multiple directions, particularly those from rollout iterations, can be effectively leveraged to constrain large pruning errors under aggressive pruning rates.



(a) Task Lift. (b) Task Square. (c) Task Tool. (d) Task Can.

Figure 9. Similarity between features from different rollouts on Lift, Square, Tool, and Can tasks.

B. Details on Ablation Study

For a more comprehensive evaluation, we further include ablation studies on additional tasks, namely Kitchen, Lift$_{ph}$, and the MH dataset. The results in Tables 7 and 8 echo the findings from the main paper: single-direction reuse strategies consistently fail to maintain high success rates under aggressive sparsity across all tasks. In contrast, the proposed omnidirectional caching strategy consistently achieves the best trade-off between performance and efficiency across both tasks and datasets.

C. Training Efficiency

We assess training efficiency by reporting the success rate achieved with varying numbers of sampled trajectories, as shown in Figures 10 and 11. Remarkably, our method reaches performance comparable to the original model under 93% sparsity using only two trajectories, and its performance continues to improve as more trajectories are provided. These results highlight the strong data efficiency of our training pipeline.

Method	**Success Rate (%, ↑)**			
	Can	Transport	Lift	Square
Dense	94	56	100	74
Forward-direction	64	0	2	32
Timestep-direction	72	48	98	66
Rollout-direction	82	0	98	48
Omini-direction	94	62	100	82

Table 7. **Ablation Study.** We report the success rate when only reusing features from single directions on MH dataset under 93% sparsity. Best performances are highlighted.

Method	**Success Rate (%, ↑)**				
	Kit$_{p1}$	Kit$_{p2}$	Kit$_{p3}$	Kit$_{p4}$	Lift
Dense	100	100	100	100	100
Forward-direction	64	0	2	32	6
Timestep-direction	72	48	98	66	4
Rollout-direction	82	0	98	48	8
Omini-direction	100	100	100	100	100

Table 8. **Ablation Study.** We report the success rate when only reusing features from single directions on Kitchen and Lift$_{PH}$ under 93% sparsity. Best performances are highlighted.

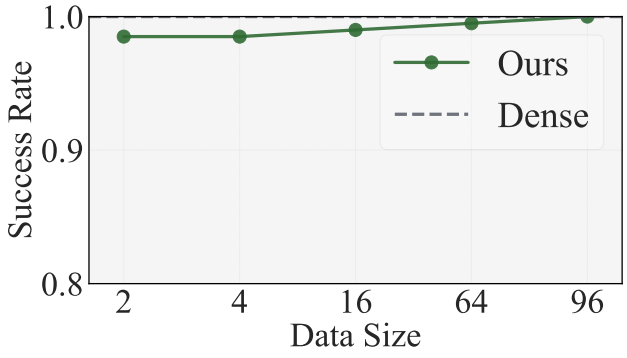


Figure 10. Success rates on the Kitchen task with different numbers of training trajectories.

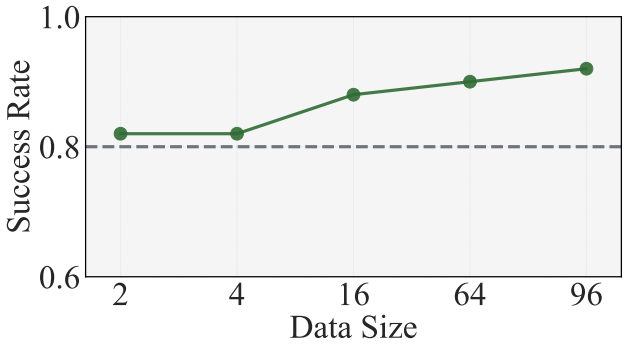


Figure 11. Success rates on the Transport task with different numbers of training trajectories.

D. Details on Qualitative Results

We present the predicted pruning masks $\mathcal{M}$ for different tasks in Figures 12, 13, 14, 15, and 16. These visualizations highlight several key observations.

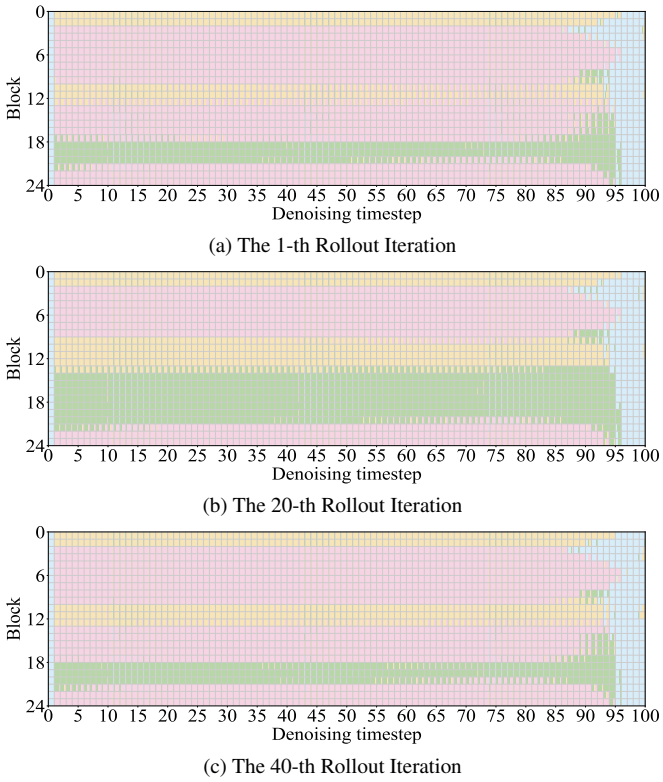
First, the pruning masks differ significantly between tasks. For example, features cached from the forward direction are prioritized in the Kitchen task, whereas features from the rollout direction dominate in the Square task. This verifies that model dynamics evolve based on distinct perceptions in an open environment.

Second, a dominant trend across all tasks is that features cached from the rollout iteration, colored in pink, account for the vast majority of preserved computations. Conversely, the previously adopted temporal-level cached features, colored in yellow, account for a minimal portion. This observation strongly suggests that the rollout direction provides a more effective caching strategy for action diffusion.

Finally, the mask $\mathcal{M}$ exhibits significant variations throughout the rollout iterations, revealing evolving visuomotor dynamics within multi-round interactions. We observe other interesting phenomena, such as computations being heavily concentrated in the final steps of the rollout. We hope these observations inspire future research into efficient visuomotor model dynamics.

E. Generalization to Image Generation

Our design is tailored to the distinct characteristics of action diffusion, specifically the heavy visuomotor conditioning costs and the unexploited redundancy in long-horizon rollouts. However, the underlying principles apply broadly. The Parallelized Inference Pipeline is compatible with standard DiT-style backbones used in image/video diffusion. Moreover, Omnidirectional Reusing Strategy naturally extends to autoregressive video models (e.g., *Rolling Forcing*). It can also adapt to generic image/video generation via a variant that selectively utilizes block and timestep dimensions.



(a) The 1-th Rollout Iteration

(b) The 20-th Rollout Iteration

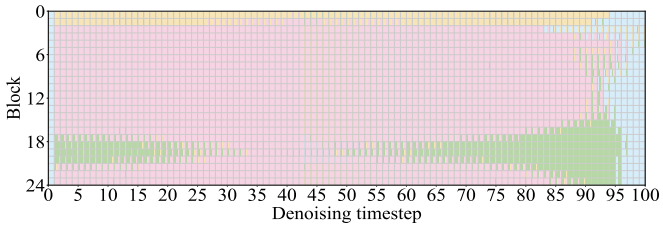
(c) The 40-th Rollout Iteration

Figure 12. Visualization of $\mathcal{M}$ on Can.

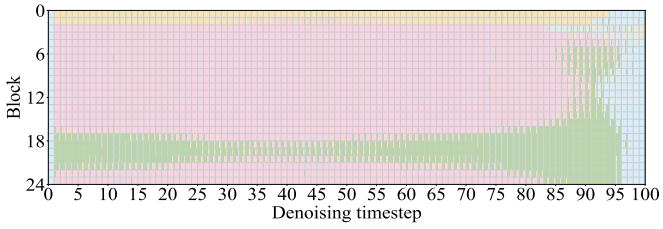
Table 9 demonstrates this strong generalization, showing negligible degradation (FID 2.23) at 80% sparsity.

Table 9. Class-to-image generation on ImageNet with DiT-XL/2.

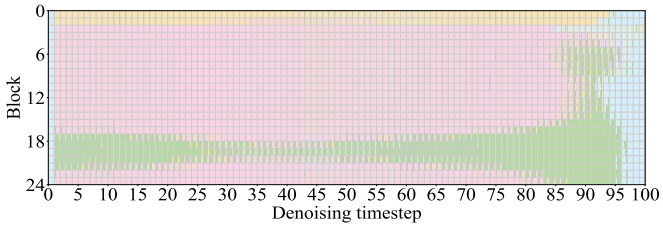
Method	**Sparsity (%)**	**FID(↓)**	**sFID(↓)**	**Speedup(↑)**
DDIM 50 steps	0	2.18	4.21	1.0×
FORA ($\mathcal{N}=5$)	80	6.58	11.29	2.87×
ToCa ($\mathcal{N}=6$)	83	6.55	7.10	2.63×
DuCa ($\mathcal{N}=5$)	80	6.06	6.72	2.78×
TaylorSeer ($\mathcal{N}=5$)	80	2.65	5.36	2.38×
Ours ($\rho=80\%$)	80	**2.23**	**4.37**	2.63×



(a) The 1-th Rollout Iteration

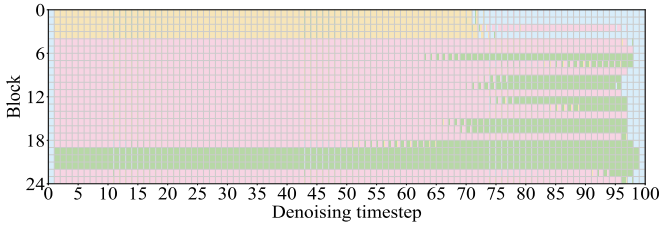


(b) The 20-th Rollout Iteration

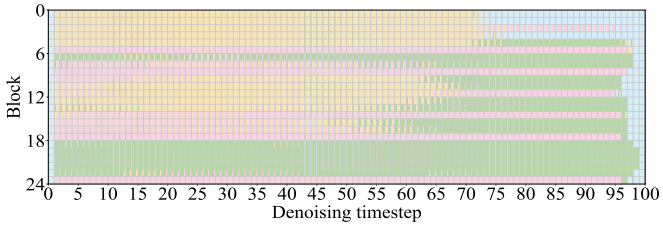


(c) The 40-th Rollout Iteration

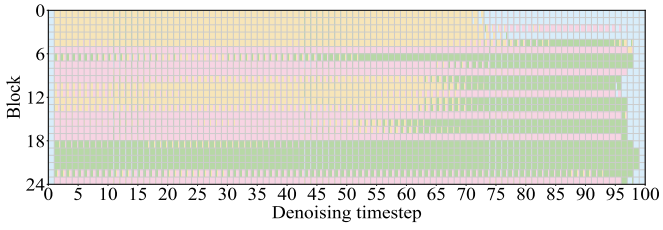
Figure 13. Visualization of $\mathcal{M}$ on Square.



(a) The 1-th Rollout Iteration

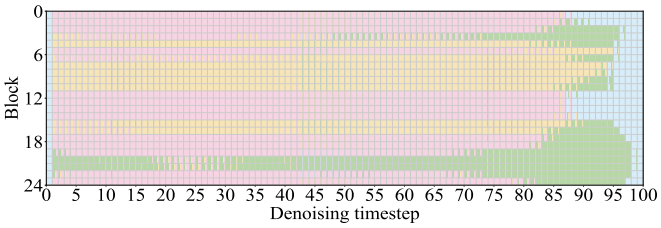


(b) The 20-th Rollout Iteration

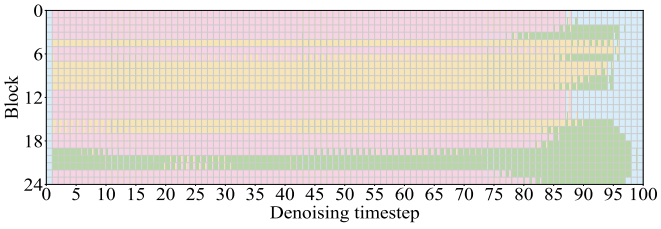


(c) The 40-th Rollout Iteration

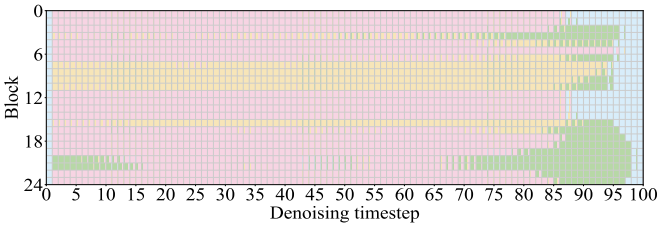
Figure 14. Visualization of $\mathcal{M}$ on Tool.



(a) The 1-th Rollout Iteration

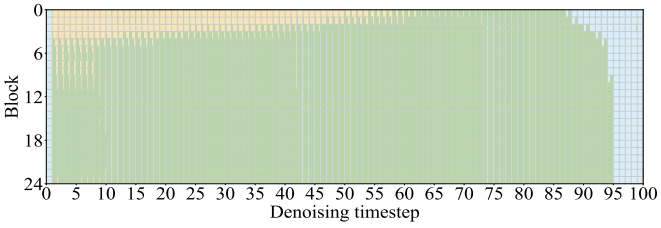


(b) The 20-th Rollout Iteration

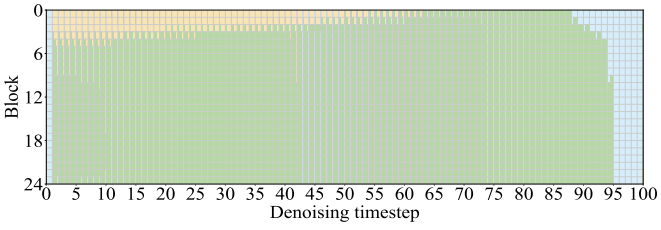


(c) The 40-th Rollout Iteration

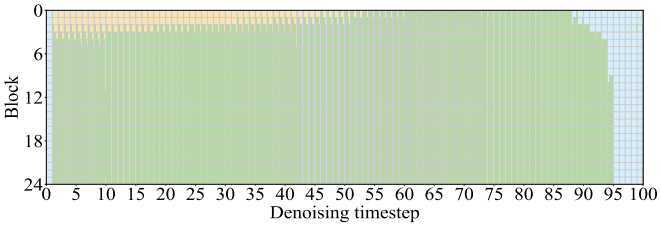
Figure 15. Visualization of $\mathcal{M}$ on Transport.



(a) The 1-th Rollout Iteration



(b) The 20-th Rollout Iteration



(c) The 40-th Rollout Iteration

Figure 16. Visualization of $\mathcal{M}$ on Kitchen.