Accelerating Diffusion Sampling via Exploiting Local Transition Coherence

Overview

The supplementary materials consist of three sections:

- **The first section provides supplementary ablation studies mentioned in ??** (See Appendix A).
- **The second section is the Mathematical Derivation of LTC-ACCEL** (see Appendix B). The overall mathematical derivation consists of two parts:
  - We present the derivation process of w_g in detail (see Appendix B.1).
  - We introduce the derivation process for the error upper bound inequality in detail(see Appendix B.2).
  - We provide additional experimental evidence demonstrating the convergence of w_g and show that this convergence holds across different schedulers (see Appendix B.3).
- **The third section details the experimental setup for the figures** (see Appendix C).
- **The final section focuses on providing detailed experimental settings** (see Appendix D), including three parts:
  - We provide the details of the acceleration interval and the conditions for setting the approximated steps (see Appendix D.1).
  - We specify which experiments utilized the optional algorithm and which did not (see Appendix D.2).
  - We present intuitive experimental visual results to demonstrate the effectiveness of our LTC-ACCEL (see Appendix E).

A. Ablation Studies

In this section, we perform ablation studies to further assess and validate the effectiveness of our method.

Directly Skipping Steps: We compare our approach with the Skipping Steps strategy within the same acceleration framework. Tab. 6 shows that our method improves computational efficiency over the original approach while preserving generation quality better than the Skipping Steps.

Model	Scheduler	Skipping Steps		LTC-ACCEL	
		Steps	ImageReward↑	Steps	ImageReward↑
SD v2	DDIM	7	0.0537	7	0.1472
	DDIM	10	0.2003	10	0.2442
	DDIM	13	0.2812	13	0.3129
SD v2	EDM	7	0.0158	7	0.2018
	EDM	10	0.2003	10	0.3171
	EDM	13	0.2582	13	0.3335

Table 6. Ablation study comparing LTC-ACCEL with the **Skipping Steps** method under the same acceleration framework. The results show that LTC-ACCEL outperforms Skipping Steps, indicating the effectiveness of LTC-ACCEL. The results demonstrate that our method **consistently outperforms** the Skipping Steps strategy in all scenarios.

B. Mathematical Derivation

B.1. Derivation of w_g

To derive the w_g, we have:

$$w_g = \arg\min \left(\|\Delta \mathbf{x}_{t+1,t} - w_g \gamma \Delta \mathbf{x}_{t+2,t+1}\|^2 \right). \quad (9)$$

We expand the objective function in terms of the inner product:

$$\begin{aligned} &\|\Delta \mathbf{x}_{t+1,t} - w_g \gamma \, \Delta \mathbf{x}_{t+2,t+1}\|^2 \\ &= \|\Delta \mathbf{x}_{t+1,t}\|^2 - 2 w_g \gamma \left(\Delta \mathbf{x}_{t+1,t} \cdot \Delta \mathbf{x}_{t+2,t+1} \right) \\ &\quad + w_g^2 \gamma^2 \|\Delta \mathbf{x}_{t+2,t+1}\|^2 . \end{aligned} \quad (10)$$

Taking the derivative of this expression with respect to w_g and setting it equal to zero yields:

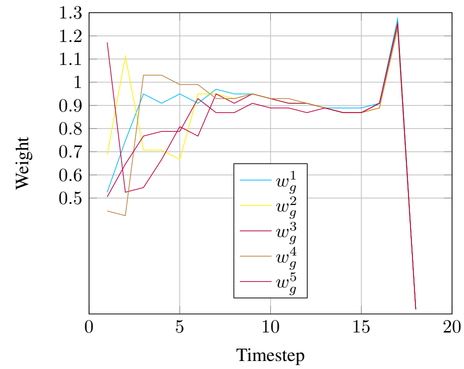
$$\begin{aligned} &\frac{\partial}{\partial w_g} \left[\|\Delta \mathbf{x}_{t+1,t} - w_g \gamma \, \Delta \mathbf{x}_{t+2,t+1}\|^2 \right] \\ &= -2\gamma \left(\Delta \mathbf{x}_{t+1,t} \cdot \Delta \mathbf{x}_{t+2,t+1} \right) + 2 w_g \gamma^2 \|\Delta \mathbf{x}_{t+2,t+1}\|^2 \\ &= 0. \end{aligned} \quad (11)$$

Rearranging this equation, we obtain:

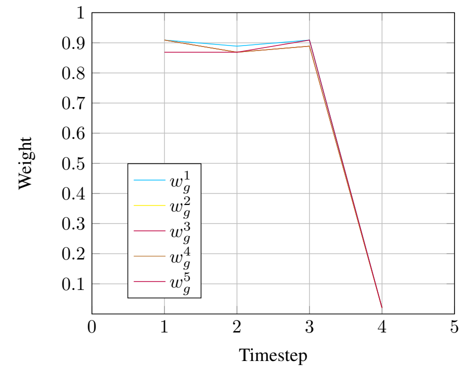
$$w_g \gamma \|\Delta \mathbf{x}_{t+2,t+1}\|^2 = \Delta \mathbf{x}_{t+1,t} \cdot \Delta \mathbf{x}_{t+2,t+1}. \quad (12)$$

Thus, the final expression for w_g is given by:

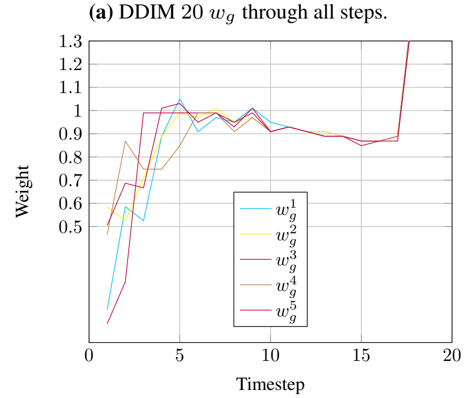
$$w_g = \frac{\Delta \mathbf{x}_{t+1,t} \cdot \Delta \mathbf{x}_{t+2,t+1}}{\gamma \|\Delta \mathbf{x}_{t+2,t+1}\|^2}. \quad (13)$$



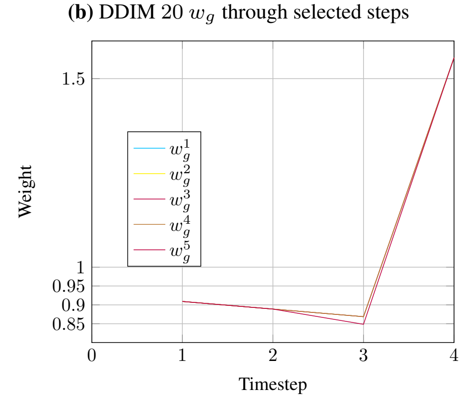
(a) DDIM 20 w_g through all steps.



(b) DDIM 20 w_g through selected steps



(c) DPM 20 w_g through all steps.



(d) DPM 20 w_g through selected steps.

Figure 6. Quantitative results of the variation of w_g. We present our results on DDIM and DPM-Solver++ in 20 steps, with 5 different prompts and latents. Fig. 6a and Fig. 6c demonstrate the original results without acceleration, where w_g achieves convergence after about 12 steps. Fig. 6b and Fig. 6d show the results within the acceleration inverval $[12, 20]$, where different weights are almost the same, indicating strong feature of convergence.

B.2. Derivation of the Error Upper Bound Inequality

To derive the inequality, we have:

$$\theta = \arccos\left(\frac{\Delta x_{t+1,t} \cdot \Delta x_{t+2,t+1}}{\|\Delta x_{t+1,t}\|^2 \|\Delta x_{t+2,t+1}\|^2}\right) < \tau, \tag{14}$$

which is equivalent to:

$$\frac{\Delta x_{t+1,t} \cdot \Delta x_{t+2,t+1}}{\|\Delta x_{t+1,t}\|^2 \|\Delta x_{t+2,t+1}\|^2} > \cos\tau. \tag{15}$$

Substituting this back into Eq. (9), we get:

$$\|\Delta x_{t+1,t}\|^2 - \frac{(\Delta x_{t+1,t} \cdot \Delta x_{t+2,t+1})^2}{\|\Delta x_{t+2,t+1}\|^2}. \tag{16}$$

Squaring both sides of the inequality and substituting the expressions, we can derive:

$$\begin{aligned}
&\left\|\Delta x_{t+1,t} - w_g \,\gamma\, \Delta x_{t+2,t+1}\right\|^2 \\
&\leq \left\|\Delta x_{t+1,t}\right\|^2 (1 - cos^2\tau) \\
&= \left\|\Delta x_{t+1,t}\right\|^2 sin^2\tau \\
&\leq \left\|\Delta x_{t+1,t}\right\|^2 \tau^2.
\end{aligned} \tag{17}$$

B.3. Experimental Verification of w_g Convergence

We select DPM-Solver++ [5]and DDIM [8] as representative schedulers for our experiments on Stable Diffusion v2. Given that w_g tends to exhibit stronger convergence with more steps, we choose 20 steps as a representative case. Fig. 6a and Fig. 6c illustrate the convergence behavior of the original w_g, while Fig. 6b and Fig. 6d demonstrate that the w_g values obtained through our algorithm also exhibit convergence.

C. Experimental Setup for Figures

In this section, we provide the experimental settings for all figures presented in the main text.

- **Figure 1**
  - The base model is Stable Diffusion v3.5 [1], and the corresponding scheduler is DPM-Solver++.
  - We select the acceleration condition as $t \bmod r = r-1$ and $t > 4$, where the period parameter r is set as $r = 2$.
- **Figure 2**
  - **(a)** The base model is Stable Diffusion v2 [7], and the baseline is DeepCache [6] with 50 steps where $N = 2$. We select the acceleration condition as $t \bmod r = r-3$ and $t > 12$, where the period parameter r is set as $r = 3$ only in experiments about DeepCache.
  - **(b)** The baseline model is Animated-Diff [2] model, and the Distillation model is Animated-Diff-Lightning [4] with 4 steps under the scheduler EDM [3]. We select the acceleration condition as $t \bmod r = r - 1$ and $t > 2$, where the period parameter r is set as $r = 2$.
- **Figure 3**
  - **(a)** The base model is Stable Diffusion v2, and the corresponding scheduler is DDIM. The result is obtained at 40 steps.
  - **(b)** We select the acceleration condition as $t \bmod r = r - 1$ and $t > 12$ based on the original setting, where the period parameter r is set as $r = 2$.
  - **(c)** The base model is Stable Diffusion v2, and the corresponding scheduler is DDIM. The weight values are obtained at 40 steps. Note that the acceleration condition here is $(t > 1)$, and the approximated value is not assigned to x to prevent accumulated errors from deviating the results from the original process.
- **Figure 4**
  - The base model is Stable Diffusion v2, and the corresponding scheduler is DDIM. The result is obtained at 40 steps. We select the acceleration condition as $t \bmod r = r - 1$ and $t > 12$ based on the original setting, where the period parameter r is set as $r = 2$. We give trials to a series of bias from 0.0125 to 0.05.
- **Figure 5**
  - The detailed acceleration settings are available in Tab. 7.

D. Experimental Details

In this section, we give a further insight into relevant details of the experiments mentioned in the paper, including the settings of acceleration intervals, the optional w_g algorithm.

D.1. Acceleration Interval

As our method mentioned, we select certain consecutive timesteps as the acceleration interval if the angles formed between their transition operators are less than a threshold τ. Typically, we prefer to set $\tau = 0.1$. However, the threshold and specific acceleration interval may vary slightly by manual adjustment according to the actual angle plot, in case of a few abnormal angle data samples.

After choosing the acceleration interval, we need to set the period of acceleration, which is defined as the parameter r in the paper. In most cases, the acceleration is applied at timestep $\mathbf{x}_t$ within the acceleration interval if $t \bmod r = r - 1$. Generally, $r = 2$ is applicable for almost all the cases without any other modification manually, which demonstrates the versatility of our method. The acceleration condition may be adjusted if further improvements in generation quality are required.

Text-to-image Synthesis Task

Tab. 7 presents the detailed results of all the experiments we conduct on Stable Diffusion v2 and v3.5-mid in the text-to-image synthesis task, with acceleration settings attached. And Tab. 8 and Tab. 9 show the acceleration settings of each group. Notably, $r = 2$ is across all these experiments, except for the case that $r = 3$ in the experiment with DeepCache in Tab. 8.

Text-to-video Synthesis Task

We evaluate video quality from two perspectives: Frame Consistency and Textual Faithfulness. For Frame Consistency, we compute CLIP image embeddings for every frame of the output video and report the average cosine similarity among all frame pairs. For Textual Faithfulness, we compute the average ImageReward score between each output video frame and its corresponding edited prompt. From the results, we achieve a $1.54\times$ speedup with almost no impact on video generation quality.

Tab. 10 gives a specific view of the acceleration settings for the experiments we conduct on video models in the text-to-video synthesis task. In addition, Tab. 11 presents the acceleration settings in distillation as well. All the experiments keep consistent in $r = 2$.

Ablation Study

The acceleration framework in the ablation experiment with "Skipping Steps" strategy is shown in Tab. 12. In addition, we conduct further ablation studies on Align Your Steps and present results together with acceleration settings in Tab. 13.

D.2. Optional w_g Algorithm

The optional w_g algorithm is designed to further minimize the difference between the w_g we obtain and the optimal one, since in each iteration we just compute an approximate solution. However, most of the experiments demonstrate that even without the optional w_g algorithm our method can achieve promising effects on various models and schedulers, with little bias from the original process. Therefore, we apply the optional algorithm only to DDIM.

D.3. Hyperparameters

In our method, we emphasize that no manual tuning is required during deployment beyond choosing the accelera-

Metric	Model	Scheduler	Original		LTC-ACCEL		Acceleration Condition	Speedup
			Inference Step	Metric Value	Inference Step	Metric Value		
ImageReward	SD v2	DDIM	10	-0.5070	6	0.0261	$t \bmod r = r-1$ and $t > 2$	1.67×
ImageReward	SD v2	DDIM	20	0.3185	12	0.3117	$t \bmod r = r-1$ and $t > 4$	1.67×
ImageReward	SD v2	DDIM	30	0.3578	20	0.3541	$t \bmod r = r-1$ and $t > 10$	1.50×
ImageReward	SD v2	DDIM	40	0.3967	26	0.4009	$t \bmod r = r-1$ and $t > 12$	1.54×
ImageReward	SD v2	DDIM	50	0.4209	30	0.4183	$t \bmod r = r-1$ and $t > 10$	1.67×
ImageReward	SD v2	DDIM	100	0.4266	60	0.4316	$t \bmod r = r-1$ and $t > 20$	1.67×
ImageReward	SD v3.5	DPM-Solver++	12	0.4795	8	0.4796	$t \bmod r = r-1$ and $t > 4$	1.50×
ImageReward	SD v3.5	DPM-Solver++	24	0.9249	16	0.9287	$t \bmod r = r-1$ and $t > 8$	1.50×
ImageReward	SD v3.5	DPM-Solver++	36	1.0254	24	1.0313	$t \bmod r = r-1$ and $t > 12$	1.50×
ImageReward	SD v3.5	DPM-Solver++	48	1.0990	32	1.1016	$t \bmod r = r-1$ and $t > 16$	1.50×
ImageReward	SD v3.5	DPM-Solver++	60	1.0755	40	1.0785	$t \bmod r = r-1$ and $t > 20$	1.50×
ImageReward	SD v3.5	EDM	20	0.9351	13	0.9089	$t \bmod r = r-1$ and $t > 6$	1.53×
ImageReward	SD v3.5	EDM	30	1.0166	19	1.0040	$t \bmod r = r-1$ and $t > 8$	1.58×
ImageReward	SD v3.5	EDM	40	1.0578	26	1.0497	$t \bmod r = r-1$ and $11 \le t \le 37$	1.54×
ImageReward	SD v3.5	EDM	50	1.0725	30	1.0623	$t \bmod r = r-1$ and $t > 10$	1.67×
ImageReward	SD v3.5	EDM	60	1.0766	39	1.0691	$t \bmod r = r-1$ and $15 \le t \le 55$	1.54×
PickScore	SD v2	DDIM	10	20.08	6	21.09	$t \bmod r = r-1$ and $t > 2$	1.67×
PickScore	SD v2	DDIM	20	21.53	12	21.52	$t \bmod r = r-1$ and $t > 4$	1.67×
PickScore	SD v2	DDIM	30	21.60	20	21.59	$t \bmod r = r-1$ and $t > 10$	1.50×
PickScore	SD v2	DDIM	40	21.66	26	21.65	$t \bmod r = r-1$ and $t > 12$	1.54×
PickScore	SD v2	DDIM	50	21.69	30	21.69	$t \bmod r = r-1$ and $t > 10$	1.67×
PickScore	SD v2	DDIM	100	21.73	60	21.73	$t \bmod r = r-1$ and $t > 20$	1.67×
PickScore	SD v3.5	DPM-Solver++	12	21.23	8	21.21	$t \bmod r = r-1$ and $t > 4$	1.50×
PickScore	SD v3.5	DPM-Solver++	24	21.93	16	21.95	$t \bmod r = r-1$ and $t > 8$	1.50×
PickScore	SD v3.5	DPM-Solver++	36	22.19	24	22.21	$t \bmod r = r-1$ and $t > 12$	1.50×
PickScore	SD v3.5	DPM-Solver++	48	22.26	32	22.28	$t \bmod r = r-1$ and $t > 16$	1.50×
PickScore	SD v3.5	DPM-Solver++	60	22.33	40	22.35	$t \bmod r = r-1$ and $t > 20$	1.50×
PickScore	SD v3.5	EDM	20	22.28	13	22.23	$t \bmod r = r-1$ and $t > 6$	1.53×
PickScore	SD v3.5	EDM	30	22.43	19	22.28	$t \bmod r = r-1$ and $t > 8$	1.58×
PickScore	SD v3.5	EDM	40	22.51	26	22.49	$t \bmod r = r-1$ and $11 \le t \le 37$	1.54×
PickScore	SD v3.5	EDM	50	22.53	30	22.52	$t \bmod r = r-1$ and $t > 10$	1.67×
PickScore	SD v3.5	EDM	60	22.53	39	22.52	$t \bmod r = r-1$ and $15 \le t \le 55$	1.54×

Table 7. Text-to-image Synthesis on Stable Diffusion.

Model	DeepCache			LTC-ACCEL			Acceleration Condition	Speedup
	Inference Step	Time	Image Reward	Inference Step	Time	Image Reward		
SD v2	10	264	-0.2246	8	208	-0.2739	$t \bmod r = r-1$ and $t > 4$	1.25×
	20	524	0.2445	16	419	0.2456	$t \bmod r = r-1$ and $t > 8$	1.25×
	50	1411	0.4039	38	1038	0.4096	$t \bmod r = r-3$ and $t > 12$	1.41×
	100	3014	0.4242	75	2171	0.4246	$t \bmod r = r-3$ and $24 < t \le 90$	1.38×

Table 8. Quantitative results of text-to-image, combing our method with Deepcache, where the parameter N mentioned in DeepCache remains $N = 2$.

tion interval—all other key hyperparameters are either automatically computed or empirically robust across different prompts and settings.

1. w_g : The main hyperparameter w_g is automatically computed via **??** and consistently converges within the acceleration interval (**????**), indicating **prompt-agnostic** behavior. In all experiments, w_g is computed once from a single prompt and reused. Its overhead is **comparable to a single forward pass**. The generality of w_g is supported by strong results across diverse prompts (**??** and Fig. 6).

2. $\phi(t)$: $\phi(t)$ defines the relative importance of each timestep and serves as a smooth baseline for computing γ. Though its effect is normalized out in $w_g \cdot \gamma$, it stabilizes the solution of w_g by smoothing temporal weights—a design inspired by ODE solvers. This allows w_g to adapt to model dynamics while keeping ϕ fixed and general. Ablations in Fig. 7d show consistent

Model	Scheduler	Original Inference Step	Original Image Reward	Ays Inference Step	Ays Image Reward	LTC-ACCEL Inference Step	LTC-ACCEL Image Reward	Acceleration Condition	Speedup
SD v1.5	DPM-Solver++	10	0.1111	10	0.1332	8	0.1309	$t \bmod r = r - 1$ and $t > 6$	1.25×

Table 9. Quantitative results of text-to-image, combing our method with Align Your Steps.

Model	Original Inference Step	Original Image Reward	Original Frame Consistency	LTC-ACCEL Inference Step	LTC-ACCEL Image Reward	LTC-ACCEL Frame Consistency	Acceleration Condition	Speedup
epiCRealism	10	0.2439	0.9713	7	0.2426	0.9700	$t \bmod r = r - 1$ and $t > 4$	1.43×
epiCRealism	20	0.3050	0.9729	13	0.2939	0.9732	$t \bmod r = r - 1$ and $t > 6$	1.54×
epiCRealism	30	0.3662	0.9676	19	0.3465	0.9681	$t \bmod r = r - 1$ and $t > 8$	1.58×
realistic-vision	10	0.1142	0.9636	7	0.1135	0.9633	$t \bmod r = r - 1$ and $t > 4$	1.43×
realistic-vision	20	0.2646	0.9676	13	0.2683	0.9672	$t \bmod r = r - 1$ and $t > 6$	1.54×
realistic-vision	30	0.4046	0.9655	19	0.3913	0.9669	$t \bmod r = r - 1$ and $t > 8$	1.58×
CogVideoX-2B	20	-0.1441	0.9442	14	-0.1673	0.9361	$t \bmod r = r - 1$ and $t > 8$	1.43×
CogVideoX-2B	30	0.2302	0.9464	19	0.2320	0.9435	$t \bmod r = r - 1$ and $t > 8$	1.58×
CogVideoX-2B	40	0.3918	0.9514	26	0.3775	0.9511	$t \bmod r = r - 1$ and $t > 12$	1.54×

Table 10. Quantitative results of text-to-video. We present our results on Animated-Diff, and CogVideoX by measuring the Textual Faithfulness and Frame Consistency using 100 prompt-video pairs.

performance across different ϕ, confirming robustness.

3. r : We fix $r = 2$ for all settings, except when using caching, where it adapts to reuse intermediate results.

4. τ : τ sets the acceleration interval. Larger values degrade quality via error accumulation. We suggest $\tau < 0.15$ for speed–fidelity trade-off.

5. Bias : The bias promotes global refinement over local updates, as discussed in **??**, with supporting results in **??**. further shows this design yields clear quality gains over greedy baselines."

To further validate the generalizability of these hyperparameters, we evaluate τ, bias, and ϕ as shown in Figures 7c and 7d. Results present the followings: (1) Large τ degrades quality due to unstable denoising; (2) Bias enhances global performance; (3) ϕ's choices minimally affect outcomes, indicating robustness.

D.4. Additional Video Experiments

Beyond the text-to-video experiments presented in the paper, we have conducted additional experiments using novel datasets and evaluation metrics. Specifically, we integrate vBench, a perceptual benchmark into our pipeline. To better reflect real-world scenarios, we use WebVid-style prompts, suited for video generation. We assess video quality via imaging quality and temporal flickering, as shown in Figure 7a. Despite substantial acceleration, flickering increases only marginally ($\approx$ 0.02), confirming that perceptual temporal consistency is largely preserved.

E. Visual Results from Selected Experiments

To provide a more intuitive presentation of our experimental results, we have selected representative images from our experiments for visualization. For video experiments, only the first frame is extracted for comparison. The experimental settings correspond to the displayed images as follows.

- **Fig. 8** Results obtained using DDIM sampling on Stable Diffusion v2.
- **Fig. 9** Results obtained using EDM sampling on Stable Diffusion v3.5.
- **Fig. 10** Results obtained using DPMsolver++ sampling on Stable Diffusion v3.5.
- **Fig. 11** Based on the DeepCache model, the results obtained using DDIM sampling on Stable Diffusion v2.
- **Fig. 12** Based on the Align Your Steps method, we obtained sampling results using DPM-Solver++ on Stable Diffusion v1.5.
- **Fig. 13** Results obtained using DDIM sampling on CogVideoX-2B.
- **Fig. 14** Using EDM sampling on the Animated-Diff model based on epiCRealism.
- **Fig. 15** Using EDM sampling on the Animated-Diff model based on realistic-vision.
- **Fig. 16** Using EDM sampling on the Animated-Diff-lightning model based on epiCRealism.
- **Fig. 17** Using EDM sampling on the Animated-Diff-lightning model based on realistic-vision.

Model	Original			Original			LTC-ACCEL			Acceleration Condition	Speedup
	Inference Step	Image Reward	Frame Consistency	Inference Step	Image Reward	Frame Consistency	Inference Step	Image Reward	Frame Consistency		
epiCRealism	4	0.3662	0.9685	3	0.2913	0.9673	3	0.3550	0.9645	$t \bmod r = r-1$ and $t > 2$	1.33×
epiCRealism	8	0.3371	0.9697	5	0.2978	0.9690	5	0.3493	0.9654	$t \bmod r = r-1$ and $t > 2$	1.60×
realistic-vision	4	0.2412	0.9639	3	0.1249	0.9641	3	0.2156	0.9618	$t \bmod r = r-1$ and $t > 2$	1.33×
realistic-vision	8	0.2469	0.9623	5	-0.0095	0.9635	5	0.2237	0.9598	$t \bmod r = r-1$ and $t > 2$	1.60×

Table 11. Quantitative results of text-to-video, combing our method with Animated-Diff-Lightning (the distilled version of Animated-Diff).

Model	Scheduler	Skipping Steps		LTC-ACCEL		Acceleration Condition
		Steps	ImageReward↑	Steps	ImageReward↑	
SD v2	DDIM	7	0.0537	7	0.1472	$t \bmod r = r-1$ and $t > 4$
	DDIM	10	0.2003	10	0.2442	$t \bmod r = r-1$ and $t > 6$
	DDIM	13	0.2812	13	0.3129	$t \bmod r = r-1$ and $t > 6$
SD v2	EDM	7	0.0158	7	0.2018	$t \bmod r = r-1$ and $t > 4$
	EDM	10	0.2003	10	0.3171	$t \bmod r = r-1$ and $t > 6$
	EDM	13	0.2582	13	0.3335	$t \bmod r = r-1$ and $t > 6$

Table 12. Ablation study comparing LTC-ACCEL with the Skipping Steps method, where Skipping Steps maintains the same acceleration positions as ours. $r = 2$ is consistent in the ablation study.

References

[1] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, Kyle Lacey, Alex Goodwin, Yannik Marek, and Robin Rombach. Scaling rectified flow transformers for high-resolution image synthesis, 2024.

[2] Yuwei Guo, Ceyuan Yang, Anyi Rao, Zhengyang Liang, Yaohui Wang, Yu Qiao, Maneesh Agrawala, Dahua Lin, and Bo Dai. Animatediff: Animate your personalized text-to-image diffusion models without specific tuning. *arXiv preprint arXiv:2307.04725*, 2023.

[3] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *NeurIPS*, 2022.

[4] Shanchuan Lin and Xiao Yang. Animatediff-lightning: Cross-model diffusion distillation. *arXiv preprint arXiv:2403.12706*, 2024.

[5] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models. *arXiv preprint arXiv:2211.01095*, 2022.

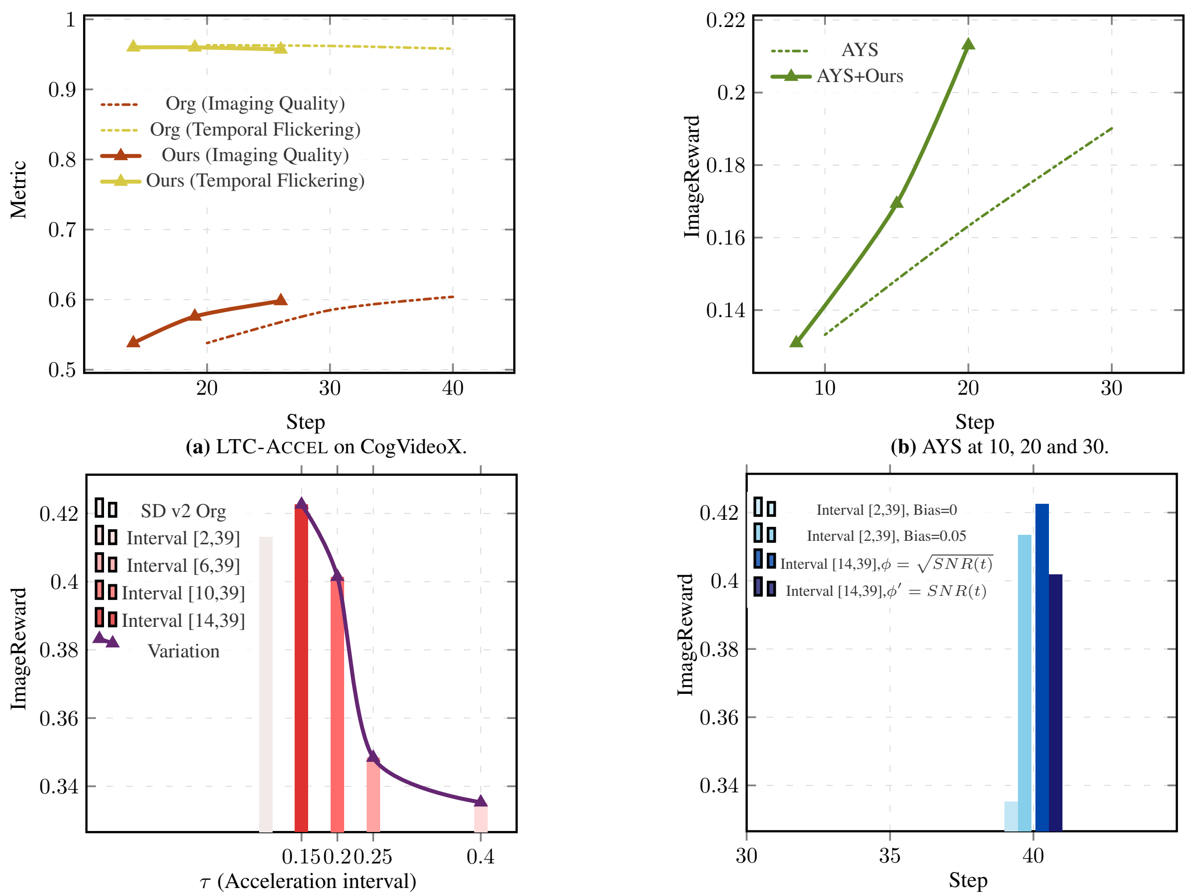
[6] Xinyin Ma, Gongfan Fang, and Xinchao Wang. Deepcache: Accelerating diffusion models for free. *IEEE*, 2023.

[7] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, 2022.

[8] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

Model	Scheduler	Ays + Skip-steps Inference Step	Ays + Skip-steps Image Reward	Ays + LTC-ACCEL Inference Step	Ays + LTC-ACCEL Image Reward	Acceleration Condition
SD v1.5	DPM-Solver++	8	0.0820	8	0.1309	$t \bmod r = r - 1$ and $t > 6$

Table 13. Ablation study comparing our method with Align Your Steps (Ays), where Ays maintains the same acceleration positions as ours. $r = 2$ is consistent in the ablation study.



(a) LTC-ACCEL on CogVideoX.

(b) AYS at 10, 20 and 30.

(c) τ selection at 40.

(d) ϕ and bias at 40.

Figure 7. Comparative experiments on video generations, compatibility with AYS, and ablation studies on interval selection, ϕ and bias.

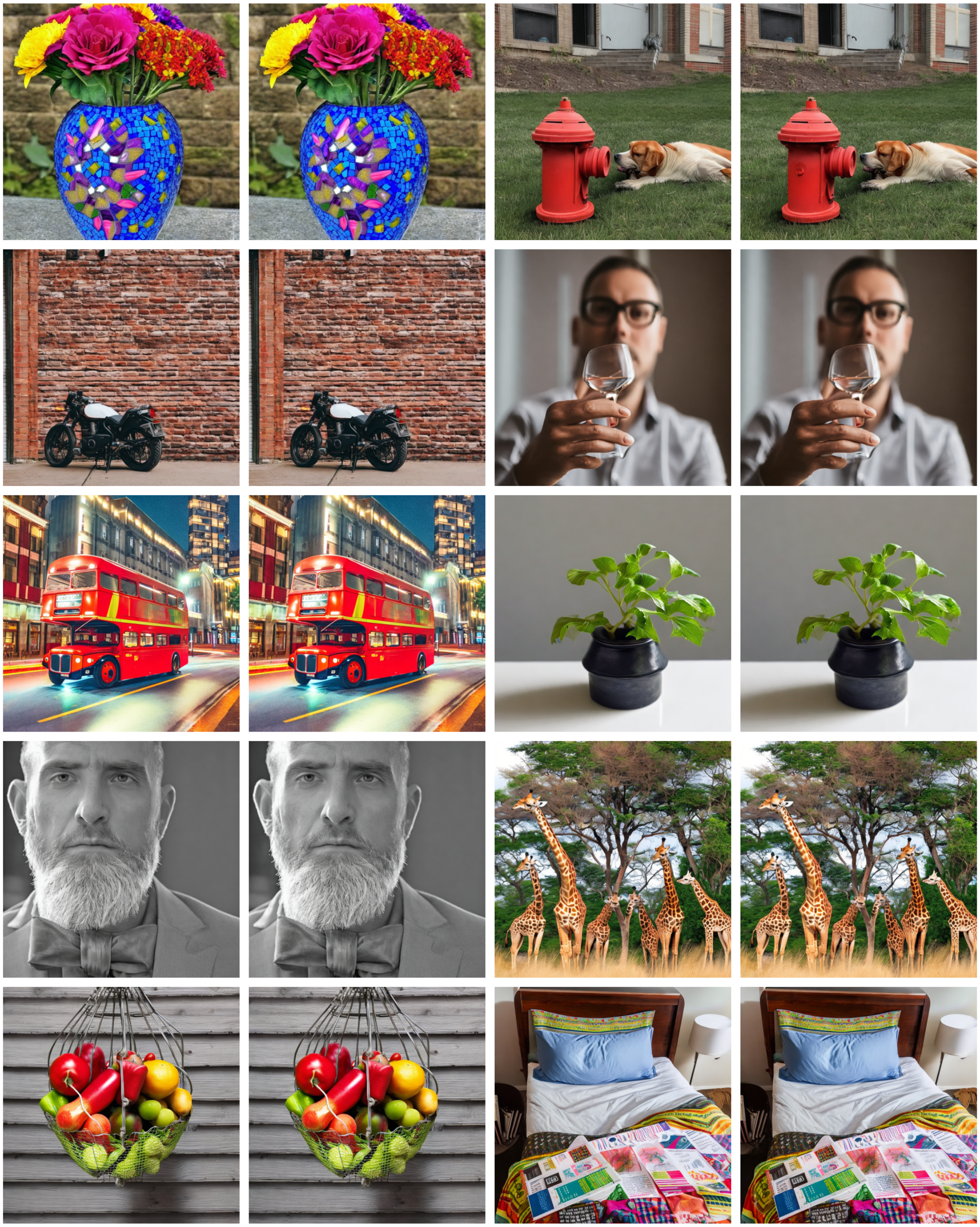


Figure 8. Results obtained using DDIM sampling on Stable Diffusion v2 are shown. In each row, the first and third images correspond to the 50-step outputs from the original process, while the remaining two images display the LTC-ACCEL results achieved in just 30 steps.

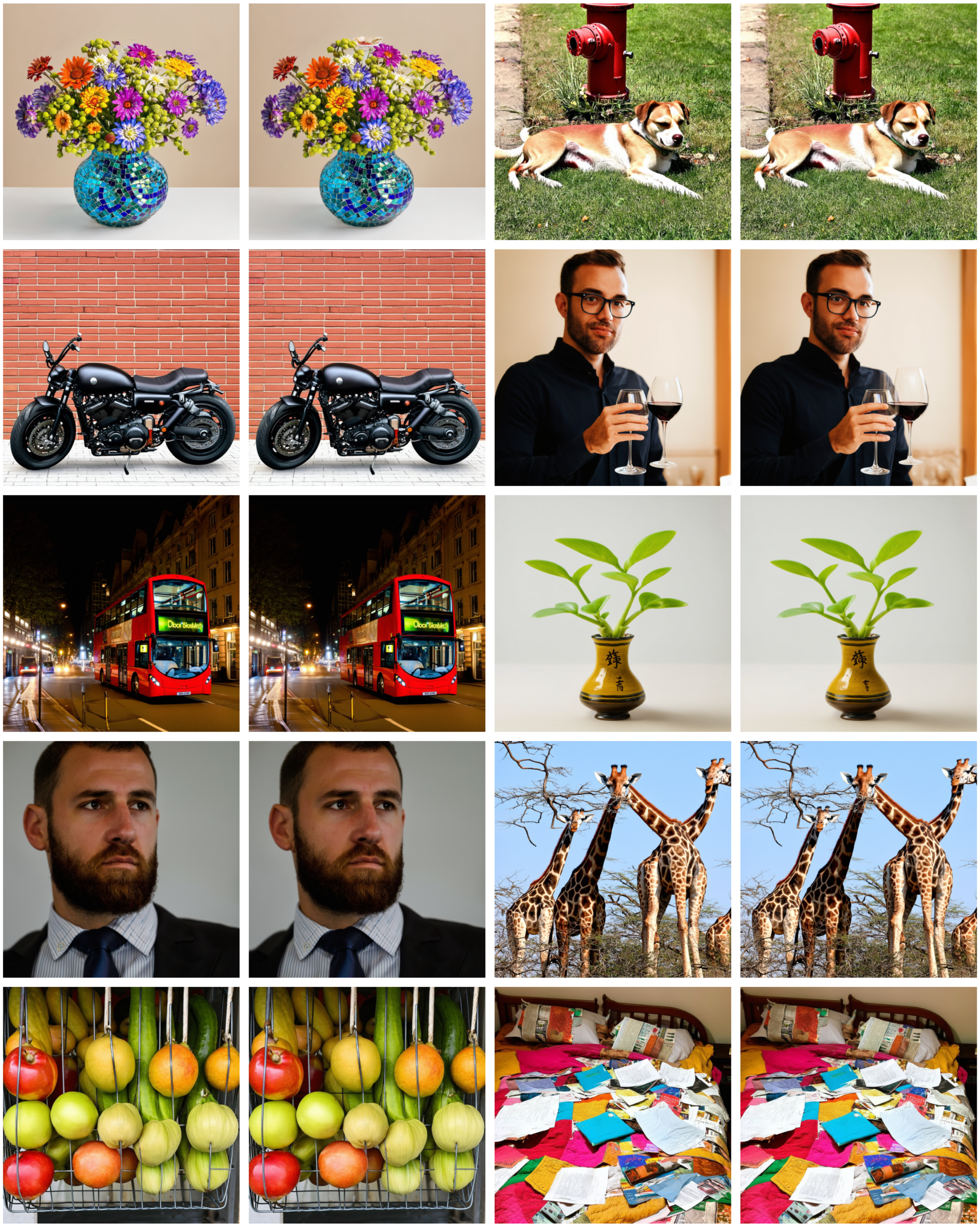


Figure 9. Results obtained using EDM sampling on Stable Diffusion v3.5 are shown. In each row, the first and third images correspond to the 60-step outputs from the original process, while the remaining two images display the LTC-ACCEL results achieved in just 39 steps.

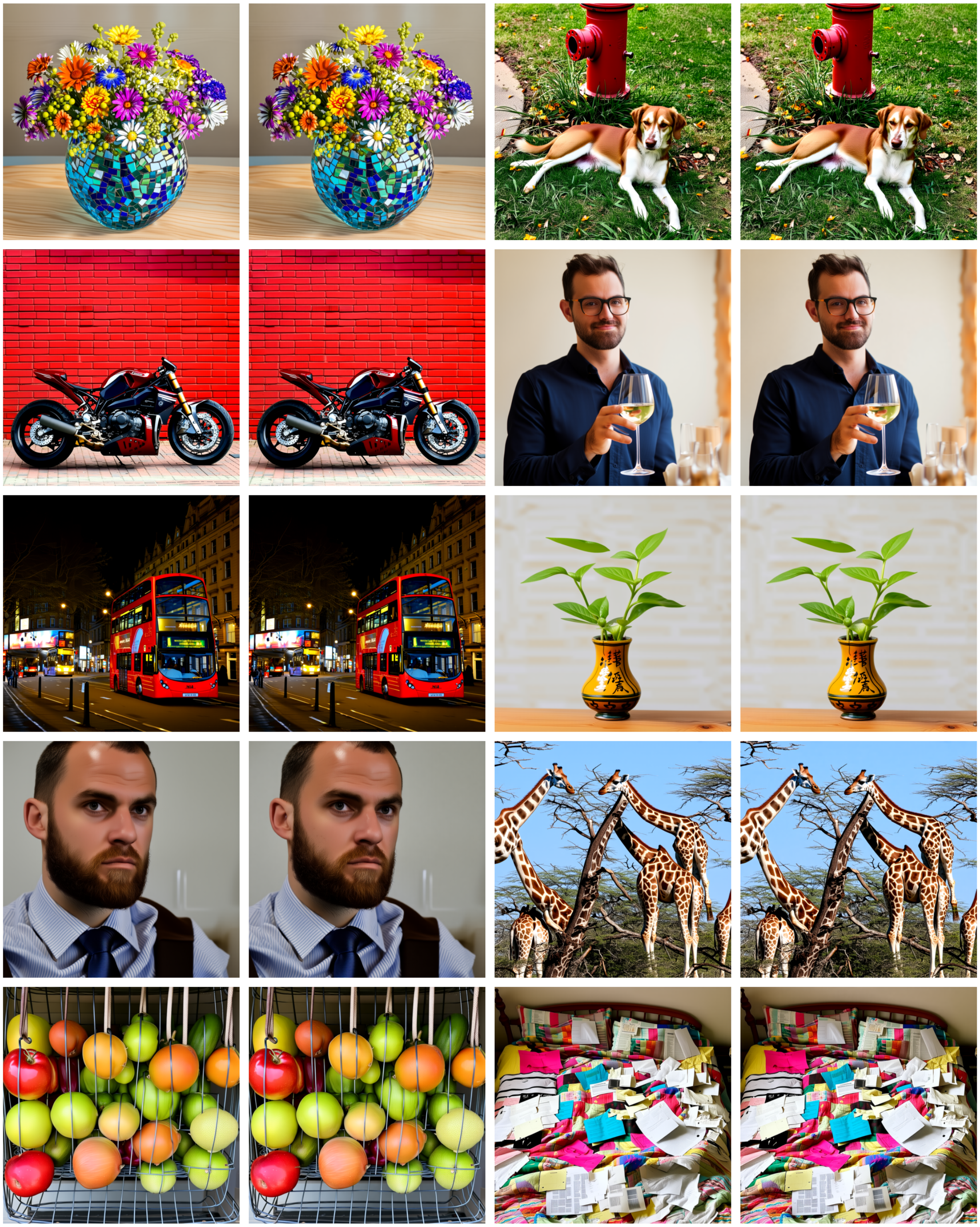


Figure 10. Results obtained using DPMsolver++ sampling on Stable Diffusion v3.5 are shown. In each row, the first and third images correspond to the 60-step outputs from the original process, while the remaining two images display the LTC-ACCEL results achieved in 40 steps.

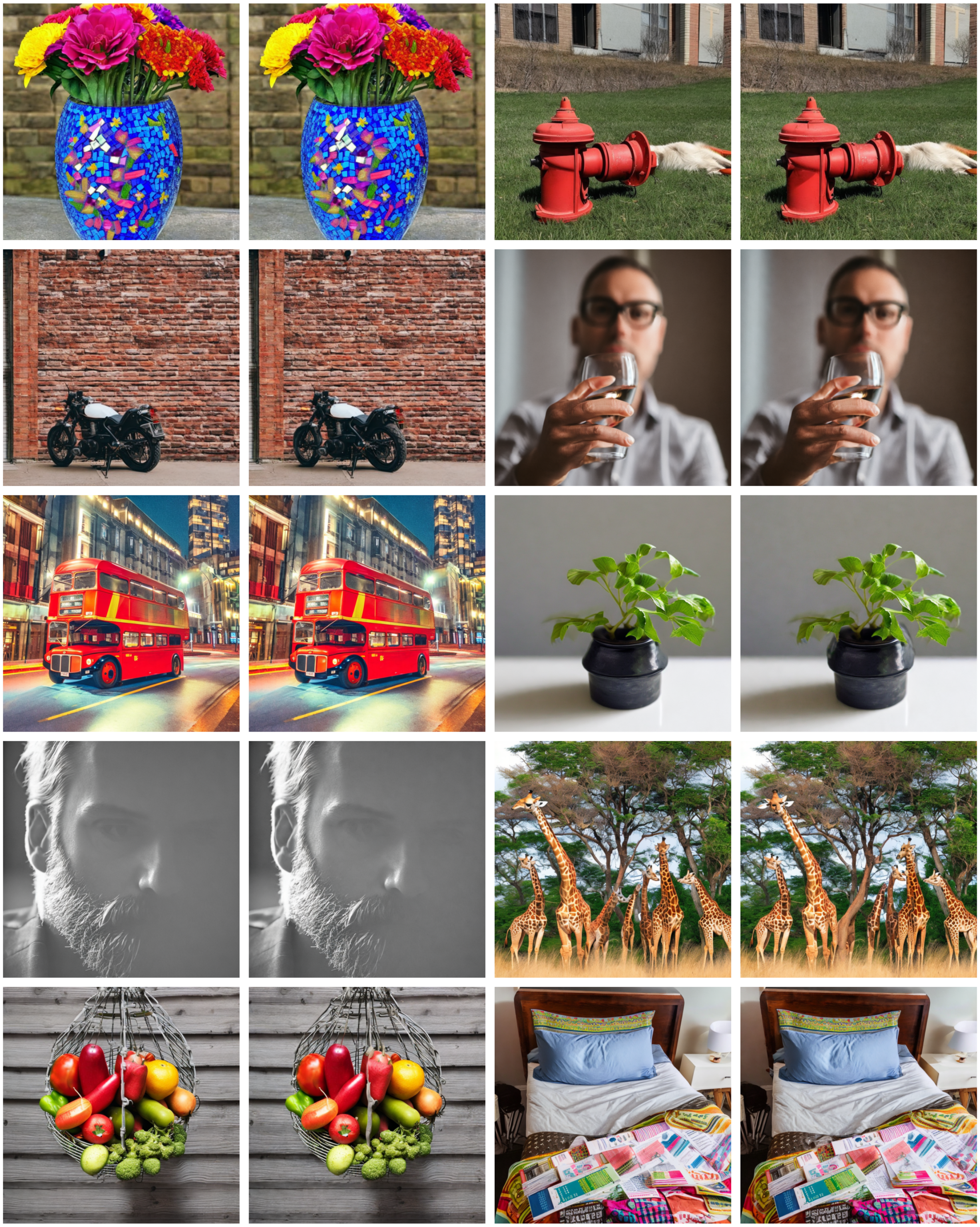


Figure 11. Based on the DeepCache model, the results obtained using DDIM sampling on Stable Diffusion v2 are presented. In each row, the first and third images depict the DeepCache results after 50 iterations, while the remaining two images display the outputs from LTC-ACCEL combined with DeepCache after 38 iterations.

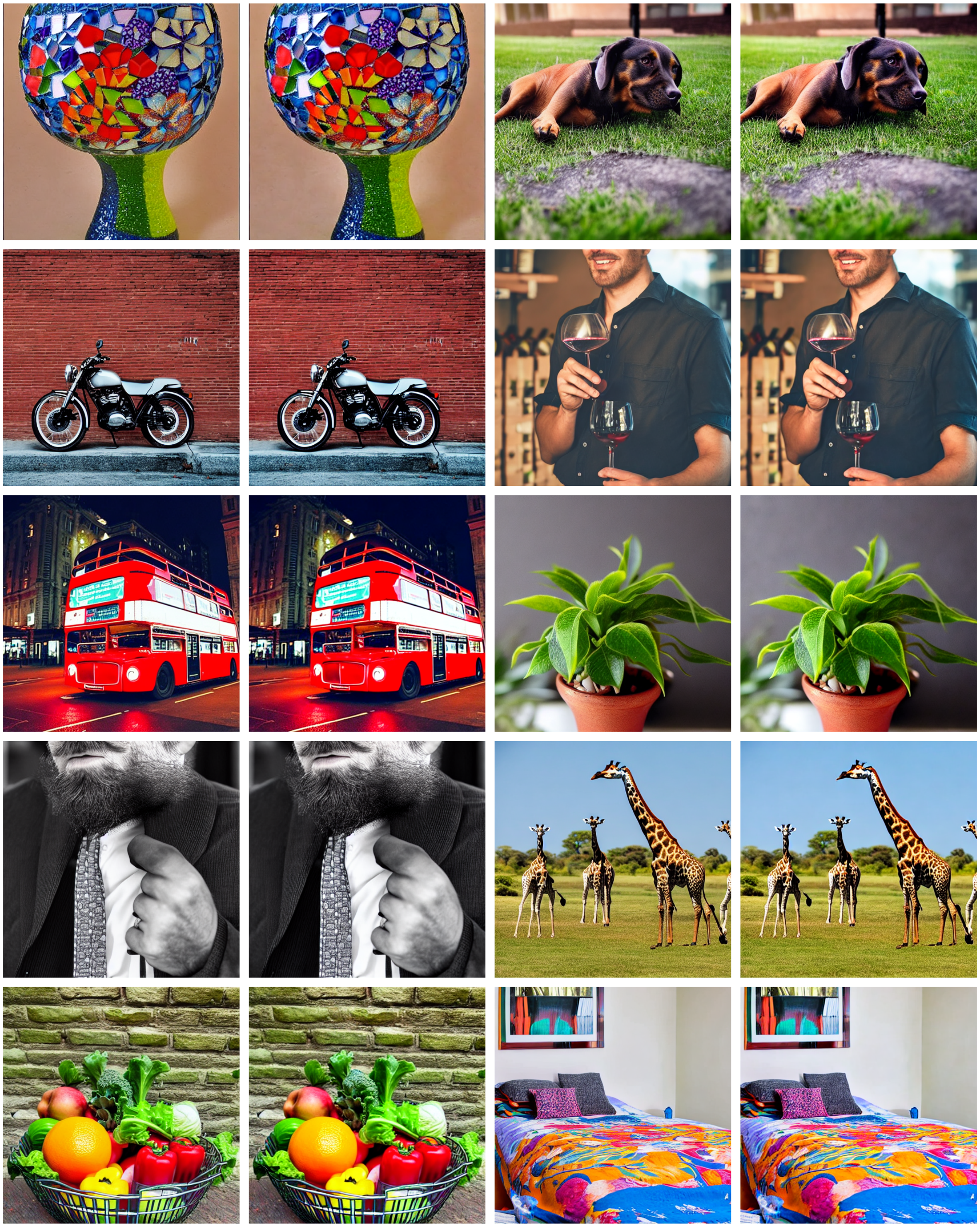


Figure 12. Based on the Align Your Steps method, we obtained sampling results using DPM-Solver++ on Stable Diffusion v1.5. In each row, the first and third images represent the outputs after 10 iterations using only the "Align Your Steps" approach, while the second and fourth images show the results achieved by combining LTC-ACCEL with align your step for 8 iterations.

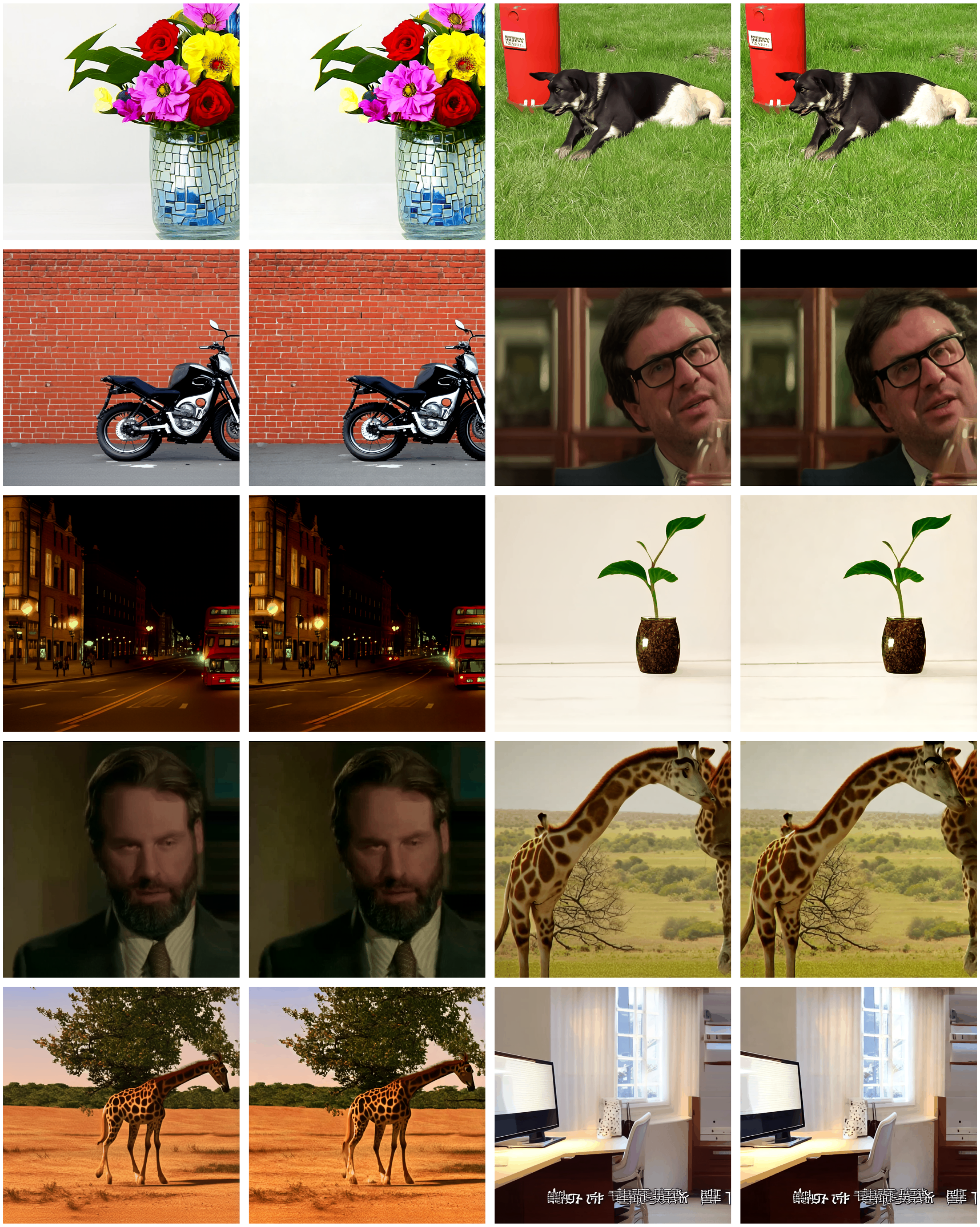


Figure 13. Results obtained using DDIM sampling on CogVideoX-2B, with only the first frame of each video selected. In each row, the first and third images represent the outputs after 40 iterations of the original process, while the remaining two images display the LTC-ACCEL results achieved in 26 iterations.

sec/Figures/Animated_ep_30.png

Figure 14. Using EDM sampling on the Animated-Diff model based on epiCRealism, we obtained results where only the first frame of each video was selected. In each row, the first and third images correspond to the outputs after 30 iterations of the original process, while the remaining two images show the LTC-ACCEL results achieved in 19 iterations.

sec/Figures/Animated_real_30.png

Figure 15. Using EDM sampling on the Animated-Diff model based on realistic-vision, we obtained results where only the first frame of each video was selected. In each row, the first and third images correspond to the outputs after 30 iterations of the original process, while the remaining two images show the LTC-ACCEL results achieved in 19 iterations.

sec/Figures/Animatedlightning_ep_4.png

Figure 16. Using EDM sampling on the Animated-Diff-Lightning model based on epiCRealism, we obtained results where only the first frame of each video was selected. In each row, the first and third images correspond to the outputs after 4 iterations of the original process, while the remaining two images show the LTC-ACCEL results achieved in 3 iterations.

sec/Figures/Animatedlightning_real_4.png

Figure 17. Using EDM sampling on the Animated-Diff-Lightning model based on realistic-vision, we obtained results where only the first frame of each video was selected. In each row, the first and third images correspond to the outputs after 4 iterations of the original process, while the remaining two images show the LTC-ACCEL results achieved in 3 iterations.