Supplementary Material

Table 1. Comparison of mainstream reward models in four aspects: Scope, Evaluation Dimensions, utilized Feature Extractor (FE) and Timestep Awareness. **EvalDim** denotes evaluation dimension, **Time-aware** denotes timestep awareness. **Aes** denotes aesthetics, **Align** denotes text-image alignment, **Fid** denotes fidelity.

Method	EvalDim	Space	Time-aware	FE
PickScore [6]	Pixel	Fid	×	CLIP [15]
ImageReward [20]	Pixel	Aes+Fid	×	BLIP [7]
MPS [23]	Pixel	Aes+Align+Fid	×	CLIP
HPSv2 [19]	Pixel	Align+Fid	×	CLIP
SPM [8]	Pixel	Aes+Align	√	CLIP
LRM [24]	Latent	Aes+Align	√	SDXL/SD1.5 [14]
VisionReward [21]	Pixel	Align+Fid+Safety	×	Qwen2.5-VL [1]

1. Latent Reward Model: Selection Rationale

As noted earlier, the previous practice in online preference optimization [13, 16] involves applying reward models [6, 8, 9, 19–21, 23] in the final denoising stage, leading to reward hacking. A natural question arises: *why not deploy optimization across a more diverse set of timesteps?* Two primary factors explain this limitation. Computationally, reward evaluation would necessitate VAE decoding at multiple steps, which would increase GPU memory and computation requirements. More fundamentally, their reward models lack timestep sensitivity, and the training paradigms are time-agnostic, having been developed on static image datasets without diffusion process context. We empirically investigate the prevalent reward models and systematically summarize their characteristics in the Tab. 1, including scope space, evaluation dimensions, and consideration of timestep. In light of the analysis above, LRM [24] is the most suitable choice. It can evaluate noisy latent at any timestep without switching to pixel space. Specifically, it takes a noisy latent and feeds it into a pretrained diffusion model, leveraging the model's native understanding of latent representation across all noise levels.

2. Additional ablation studies.

Significance of EMA in Score Estimator. As described in Sec.**??**, we employ an Exponential Moving Average (EMA) strategy to help the score estimator more accurately track the generator's distribution, especially under high-frequency updates. To validate the effectiveness of this approach, we conduct an ablation study comparing performance with and without the EMA strategy. As shown in Fig. 2, the model with EMA achieves higher ImageReward and Pickapic scores in later training stages. The HPSv2 scores remain nearly identical. This confirms that the EMA strategy enhances visual quality and human preference without compromising text alignment.

Phase 2: Other Ablation Experiments on Reinforcement Learning. *1. Online training VS Post-training:* We compare online training with post-training by LPO [25] alone. Our method demonstrates superior performance over Post-Train LPO, which validates the advantage of our proposed training paradigm. *2.High-noise VS all noise:* Training only on high-noise steps achieves better results than training on all-noise steps. *3. Including PixelGAN loss:* Furthermore, the incorporation of an additional Pixel-Gan objective yields a positive gain, resulting in a marginal improvement in the metrics. Fig.3 supplements the results with and without GAN loss across 1,000 to 5,000 iterations. We selected the 5k-step model with GAN loss as our final enhanced version, as it delivered the best performance. The training cost for this model was 12 H20 GPU hours.

3. Algorithm of Flash-DMD

Algorithm 1 provides an overview of Flash-DMD . At stage 1, we propose an advanced step distillation method based on distribution matching that converges quickly. Timestep-aware strategy, together with Pixel-GAN constraint, effectively alleviates mode-seeking and improves the realism of generation. At stage 2, we carefully select and incorporate a latent reward model into the training paradigm, further enhancing the image fidelity and image aesthetics.

4. Flash-DMD on 8-steps SDXL

To demonstrate the effectiveness of Flash-DMD on SDXL [14], we extend our method to distill the full SDXL model into an 8-step generative model. In the first stage, we select the timesteps $[999, 874, 749, 629, 499, 374, 249, 124]$ and apply Pixel-GAN at the final timestep. We adopt the Two-Time-Scale Update Rule (TTUR) with a score estimator update frequency of 2, training on the LAION dataset [18] with a batch size of 48. Two variants are trained for 3k and 6k iterations, respectively, denoted as `TTUR2-3k` and `TTUR2-6k`. In the second stage, we construct win–lose preference pairs using samples generated from the timesteps $[999, 874]$. We initialize this stage from the `TTUR2-3k` model (i.e., the 3k-step checkpoint from Stage 1) and continue training for an additional 3k and 6k steps, respectively. The results for Stage 1 and Stage 2 are reported in Tab. 2 and Tab. 3, respectively. Results show that Flash-DMD keeps outperforming other distillation and reinforcement learning methods at 8-steps generative task, highlighting the advantage of Flash-DMD. In terms of sub-

Figure 1. Qualitative comparisons with other models.

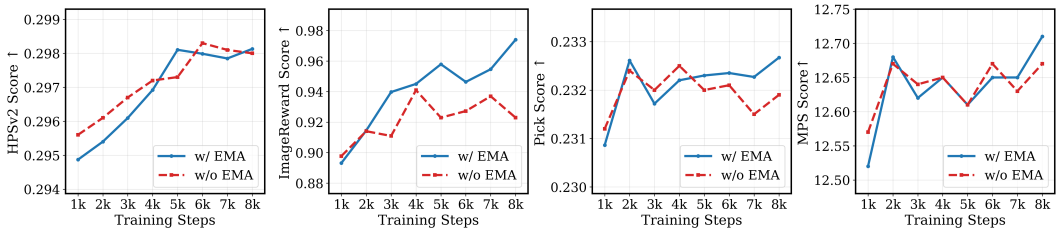

Figure 2. Evaluation results of Flash-DMD (ours) with or without EMA on ImageReward, PickScore, and HPSv2. The training steps range from 1,000 to 8,000. Both models are trained with a two-time scale update rule (TTUR). The generator and the score estimator are updated at a rate of 1:2, i.e., TTUR=2.

jective results, Fig. 4 and Fig. 6 show the results of the 4-step inference at stages 1 and stage 2 of Flash-DMD based on SDXL, respectively. Fig. 7 and Fig. 8 show the results of the 8-step inference at stages 1 and 2 of Flash-DMD , respectively. Our Flash-DMD can generate high-quality results with both realism and aesthetic appeal under a small number of steps. Furthermore, we compare our results with SDXL, SDXL-Lighting [10], SDXL-Turbo [17], Hyper-SDXL [16], DMD2 [22], LPO [25], Realism version of NitroSD [2], PSO [13]. Fig. 1 demonstrates that our model not only surpasses other distillation models but also outperforms the teacher model in refining image quality.

Algorithm 1 Flash-DMD Training Algorithm

Require: pretrained teacher model μ_{real}, real dataset $\mathcal{D}_{\text{real}}$, generator is updated with the ratio `TTUR`, inference steps K, timestep set $S = \{\tau_1, \ldots, \tau_k\}$ and its noisy set S_{noisy}, high noisy timesteps T_{noisy}, Pixel-level discriminator D_ω, VAE decoder layers $\mathcal{V}$, latent reward model R, training stage flag `FLAG`.

Ensure: trained few-step generator $\mathcal{G}_\theta$

$\mathcal{G}_\theta \leftarrow \text{copyWeights}(\mu_{\text{real}})$ ▷ Initialize generator

$\mu_\phi \leftarrow \text{copyWeights}(\mu_{\text{real}})$ ▷ Initialize score estimnator of generator

$D_\omega \leftarrow \text{initializeTrainableHeads}()$ ▷ Initialize trainable heads of discriminator

for iteration $= 1$ to max_iters **do**

  $z \sim \mathcal{N}(0, I)$

  Sample τ_i from S ▷ Pick timestep for current iteration

  Sample $x_{\text{real}} \sim \mathcal{D}_{\text{real}}$

  $x_{\tau_i} \leftarrow \text{backwardSimulation}(z, \tau_k \rightarrow \tau_i)$ ▷ Use backward simulation to get noisy image

  $x_{\tau_1} \leftarrow \text{backwardSimulation}(x_{\tau_i}, \tau_i \rightarrow \tau_1)$ ▷ Use backward simulation to get clean image

  $x \leftarrow \mathcal{G}_\theta(x_{\tau_i}, \tau_i)$

  $p_{\text{real}}, p_{\text{fake}} \leftarrow \mathcal{D}_\omega(\mathcal{V}(x_{\tau_1}))$ ▷ Get real or fake probability

  if iteration mod `TTUR` $== 0$ **then**

    $t_j \leftarrow T_{\text{noisy}}$

    $\mathcal{L}_{\text{DMD}} \leftarrow \text{distributionMatchingLoss}(\mu_{\text{real}}, \mu_\omega, x, t_j)$ ▷ Use Eq. (**??**) for faster convergence in noisy timesteps

    $\mathcal{L}_{\text{adv}} \leftarrow \text{generatorAdversarialLoss}(p_{\text{real}})$ ▷ Use Eq. (**??**) for enhanced realism and details

    $\mathcal{L}_{\mathcal{G}_\theta} \leftarrow \mathcal{L}_{\text{DMD}} + \lambda \cdot \mathcal{L}_{\text{adv}}$ ▷ Final loss function for generator

    $\mathcal{G}_\theta \leftarrow \text{update}(\mathcal{G}_\theta, \mathcal{L}_{\mathcal{G}_\theta})$

    $\mu_\phi \leftarrow \text{EMA}(\theta, \phi, \lambda_{\text{ema1}})$

  end if

  $x \leftarrow x.\text{detach}()$ ▷ Stop gradient

  $t \sim \mathcal{U}(0, 1)$

  $x_t \leftarrow \text{forwardDiffusion}(x, t)$ ▷ Add noise

  $\mathcal{L}_{\text{denoise}} \leftarrow \text{diffusionLoss}(\mu_\phi(x_t, t), x)$

  $\mu_\phi \leftarrow \text{update}(\mu_\phi, \mathcal{L}_{\text{denoise}})$ ▷ Update fake score network μ_{fake}

  $\mathcal{L}_{D_\omega} \leftarrow \text{discriminatorAdversarialLoss}(p_{\text{real}}, p_{\text{fake}})$ ▷ Discriminator's Hinge loss

  $D_\omega \leftarrow \text{update}(D_\omega, \mathcal{L}_{D_\omega})$ ▷ Update discriminator D_ω

  if `FLAG` == `Stage2` **then** ▷ Use reinforcement learning at the second stage

    $s_{\text{pool}} \leftarrow \text{iterativeSample}(S_{\text{noisy}}, K)$ ▷ Sample K image latent for reward model evaluation

    $s_{\text{win}}, s_{\text{loss}} \leftarrow \text{filter}(\mathcal{R}(s_{\text{pool}}))$ ▷ Construct win-loss pairs with $\mathcal{R}$

    $\mathcal{L}_{rl} \leftarrow \text{prefenceOptimization}(s_{\text{win}}, s_{\text{loss}})$ ▷ Use Eq. (**??**) to boost performance

    $\mathcal{G}_\theta \leftarrow \text{update}(\mathcal{G}_\theta, \mathcal{L}_{rl})$

  end if

end for

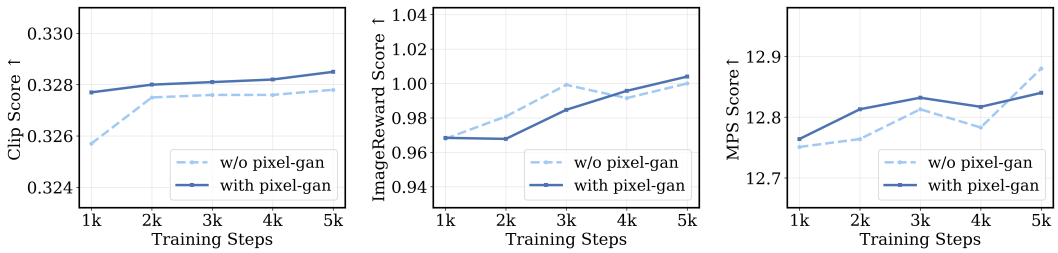

Figure 3. Evaluation results of Reinforcement Learning with and without pixel-GAN. Both models use 5:1 setting.

5. Details of Flash-DMD on SD3-Medium

Instead of training on the LAION dataset [18], we curated a proprietary, high-quality training set of 100,000 instances. It encompasses a diverse range of subjects and scenes, including portraits, architecture, flora and fauna, and food—rendered in a realistic photographic style. Our dataset construction followed a structured methodology.

Table 2. Comparison of Flash-DMD on SDXL under stage 1 with other distillation methods on the COCO-10k dataset. **ImgRwd** denotes ImageReward score. Best performance is highlighted in **bold**, and the second best is underlined.

Method	#NFE	ImgRwd ↑	CLIP ↑	Pick ↑	HPSv2 ↑	MPS ↑	Cost ↓
SDXL	100	0.7143	0.3295	0.2265	0.2865	11.87	-
LCM-SDXL	8	0.6122	0.3247	0.2261	0.2874	11.59	-
SDXL-Lightning	8	0.7187	0.3268	0.2291	0.2900	12.12	-
Hyper-SD	8	0.9119	0.3287	0.2310	0.2977	12.35	-
8-steps Flash-DMD at Stage 1							
TTUR2-3k	8	0.9159	0.3281	**0.2319**	0.2981	12.60	48*3k
TTUR2-6k	8	**0.9416**	0.3284	0.2318	**0.2989**	**12.63**	48*6k

Table 3. Comparison of Flash-DMD under stage2 with other models with reinforcement learning on COCO-10k dataset.

Method	#NFE	ImgRwd ↑	CLIP ↑	Pick ↑	HPSv2 ↑	MPS ↑	GPU Hours
Hyper-SDXL	8	0.9119	0.3287	0.2310	0.2977	12.35	200 A100
LPO-SDXL	40	**1.0417**	**0.3324**	0.2342	0.2965	12.58	92 A100
8-steps Flash-DMD (TTUR2-3k) at Stage 2							
Flash-DMD-3k	8	1.0012	0.3299	0.2338	0.2986	12.75	**12 H20**
Flash-DMD-6k	8	1.0106	0.3290	**0.2343**	**0.2998**	**12.84**	24 H20

First, we generated a set of clear, detailed captions to fully leverage the representational capacity of the SD3 [4] text encoders. These captions were then used to synthesize high-quality images using SD3.5 Large, configured with 28 denoising steps and a CFG scale of 4.5. The final training set was curated through a rigorous manual selection process from the generated candidates. We successfully extend Flash-DMD in SD3-Medium after stage 1, Fig. 9 displays some visual results that highlight the strengths of our algorithm.

6. Clarification about training cost.

We compare the training cost in identical settings (4×8 H20 GPUs, Batch Size 2, TTUR 2, no FSDP). The average iteration time for Flash-DMD is 6.42s compared to 5.97s for DMD2. While the SAM and timestep-aware strategy add a slight overhead (7.5%), the total training cost is dominated by the number of iterations. Ours achieves superior results with much fewer iterations (1k vs 24k).

7. Difference with other methods.

Difference with POSE. Despite similar observations, Flash-DMD and POSE [3] proposed significantly different strategies. Flash-DMD's timestep-aware strategy focuses on decoupling losses across diffusion timesteps (t) to stabilize the score estimator, accelerate convergence, and enhance realism. While POSE designs a stage-aware strategy for different training stages (Stage 1,2,3) of a 1-step video generative model.

Difference with SenseFlow. As to SenseFlow [5], which utilizes EMA but lacks our novel timestep-aware objective decoupling, Flash-DMD better preserves texture and alignment. We clarify that we do not claim EMA as our core contribution. We have credited its adaptation from SenseFlow, and our ablation study (Tab. 2) compares performance with and without EMA, demonstrating that our innovative timestep-aware strategy can stabilize the score estimator independently.

Difference with InstructDiff. The reward hacking phenomenon occurs when reinforcement learning (RL) policies are applied to few-step models with default distribution/parameter constraints. DiffInstruct [11, 12] series also acknowledges this issue and incorporates the KL loss to restrict the shift from initial distribution. In contrast, Flash-DMD unifies distillation and joint RL to constrain the distribution shift from multi-steps teacher model and real images for the first time.

8. Additional Visualizations and Captions

Fig. 4 captions from left to right, top to bottom are :

1. A cat next to a window behind cans and bottles.
2. A bunch of red roses bunched together.
3. A brown teddy bear standing next to bottles of honey.
4. A brown lamb looking up as other sheep eat hay in a field.
5. A brown and white horse walking down a road.
6. A brass-colored vase with a flower bouquet in it.
7. A boy closely examining a frog in his yard.
8. A boy in green shorts and a tie posing in front of a tower.
9. A furry kitten lying on a laptop.
10. A confection of cake, whipped cream, strawberries, and two candles.
11. Two birds standing around a box of birdseed.
12. A soldier riding a red motorcycle down a busy street.

Fig. 5 captions from left to right, top to bottom are:

1. A dog looking up and running to catch a frisbee.
2. A cake designed to resemble a cup.
3. A beautiful red-haired woman holding a cup while wearing a sweater.
4. A row of wooden park benches sitting next to a lake.
5. A close-up of a baseball player bending down with a glove.
6. A bag of strawberries on a table with tomatoes.
7. A crow standing on a plant in a body of water.
8. A black-and-white cat sitting on a bed.
9. A brown-and-white cow standing on a grassy hill.
10. A brown leather couch in a living room.
11. A body of water with an elephant in the background.

Figure 4. Qualitative results from Stage 1 of the 4-step Flash-DMD framework on SDXL. The model is trained with TTUR = 1 for 1,000 steps.

Figure 5. Qualitative results from Stage 1 of the 4-step Flash-DMD framework on SDXL. The model is trained with TTUR = 2 for 4,000 steps.

Figure 6. Qualitative results from Stage 2 of the 4-step Flash-DMD framework on SDXL. The model is initialized from the TTUR1-1k checkpoint and fine-tuned for 5,000 steps.

Figure 7. Qualitative results from Stage 1 of the 8-step Flash-DMD framework on SDXL. The model is trained with TTUR = 2 for 3,000 steps.

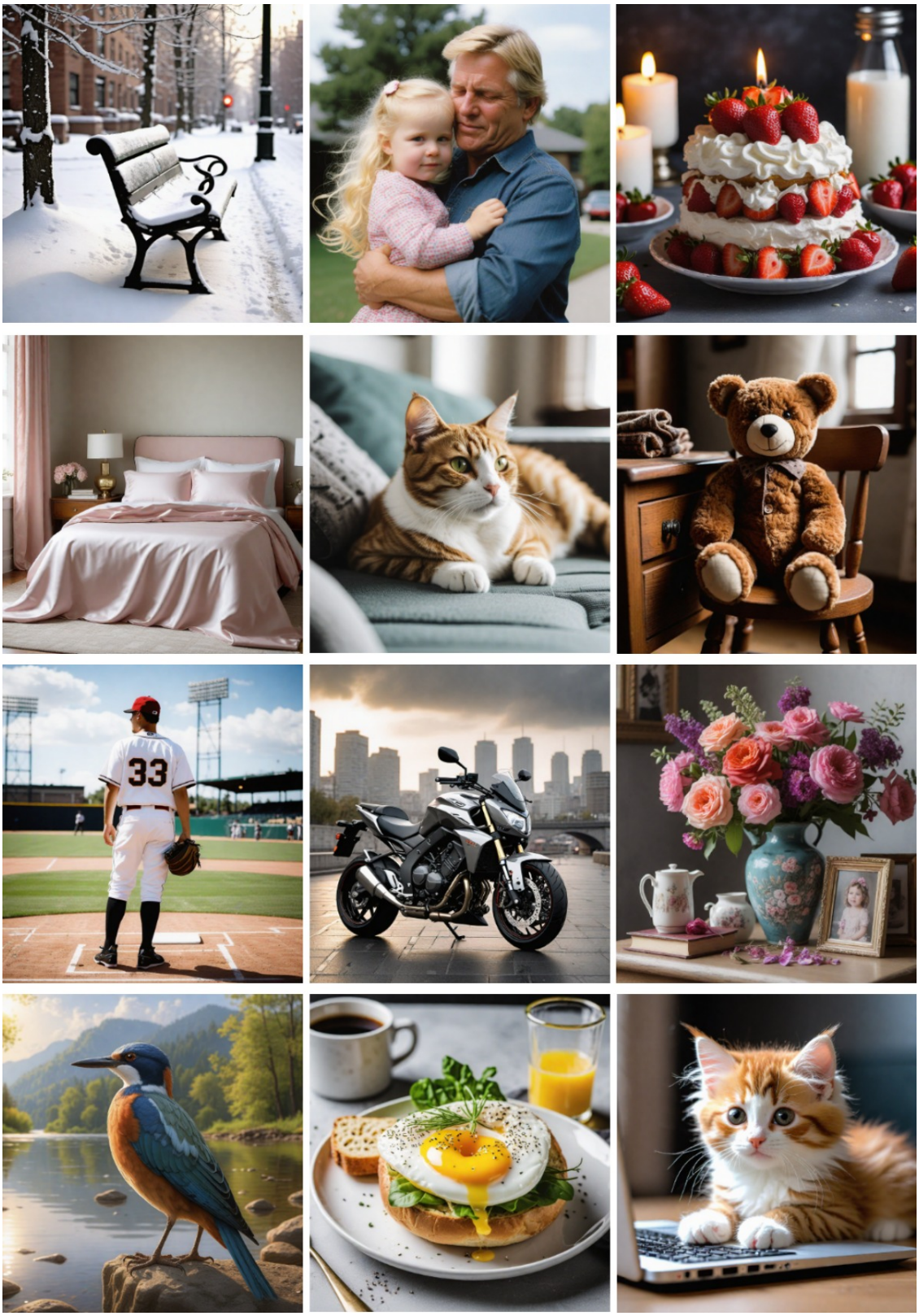

Figure 8. Qualitative results from Stage 2 of the 8-step Flash-DMD framework on SDXL. The model is initialized from the 8-step TTUR2-3k checkpoint and fine-tuned for 3,000 steps.

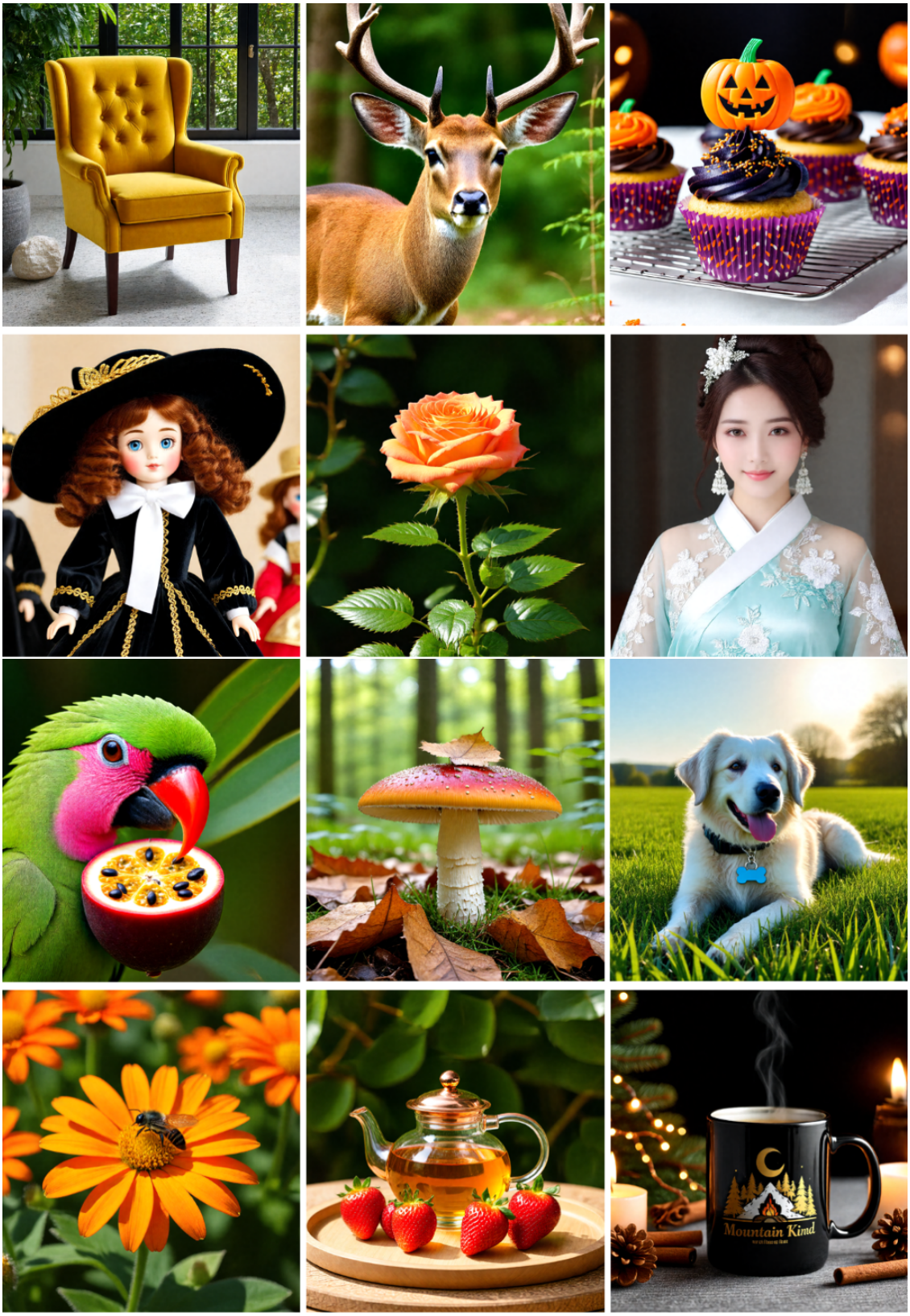

Figure 9. Qualitative results from stage 1 of the 4-step Flash-DMD framework on SD3-Medium. The model is trained with TTUR=2 for 7.000 steps.

12. A case containing a small doll with blue hair, shoes, and clothes.

Fig. 6 captions from left to right, top to bottom are:

1. A bowl of apples and bananas sitting on a woven cloth.
2. A cake decorated with a surfer and palm trees.
3. A close-up of a person eating a doughnut.
4. A brown teddy bear sitting at a table next to a cup of coffee.
5. A close-up of a metallic elephant statue.
6. A baby in a gray jacket eating a piece of pizza crust.
7. A couch and chair sitting in a room.
8. A bench in the park on a rainy day.
9. A brown-and-white cat sitting on a windowsill.
10. A bedsit with a kitchenette featuring white cabinets.
11. A bottle of wine placed next to a glass of wine.
12. A blurred motorcycle against a red brick wall.

Fig. 7 captions from left to right, top to bottom are:

1. A bird on a plank looking at green water.
2. A boy holding a pitcher's mitt at a park.
3. A clock near the front of a house.
4. A bench sitting atop a lush green hillside.
5. A black bear walking through a field of grass and straw.
6. A black motorcycle parked on gray cobblestones.
7. A child in bed wearing a striped sweater and colorful blanket.
8. A brass-colored vase with a flower bouquet in it.
9. A clean bedroom with a dog on the bed.
10. A chef preparing sushi on a countertop.
11. A close-up of a sandwich with French fries.
12. A close-up image of a cat and a keyboard.

Fig. 8 captions from left to right, top to bottom are:

1. A bench along a sidewalk in winter, covered in snow.
2. A man holding a little blond girl.
3. A confection of cake, whipped cream, strawberries, and candles.
4. A bedroom with a silky bedspread and pillows.
5. A cat lying on a sofa next to some pillows.
6. A brown teddy bear seated on a chair beside a wooden drawer.
7. A baseball player standing next to home plate.
8. A sleek motorcycle in a cityscape.
9. A beautiful vase full of flowers, with pictures placed beside it.
10. A beautiful bird standing on the bank of a river.
11. A bagel topped with egg and other ingredients on a plate.
12. A furry kitten lying on a laptop.

Fig. 9 captions from left to right, top to bottom are:

1. A mustard-yellow armchair with button tufting sits on a speckled white floor. The chair has dark wooden legs. A large potted plant is partially visible to the left. A window with dark frames shows green trees and foliage. A white rock sits on the floor near the chair. The floor is partially shaded by the window.
2. A brown deer with large, curved antlers stands in a forest setting. The deer's coat is a uniform brown color. Its antlers are dark brown and have multiple points. The deer's face is partially visible, showing its nose and eyes. The background is blurred with green foliage. The deer appears to be looking directly at the camera.
3. Several cupcakes are arranged on a wire rack. One cupcake has dark blue frosting, gold sprinkles, and a jack-o'-lantern topper. Other cupcakes have orange frosting and are decorated with orange and purple sprinkles. The cupcakes are in purple and white patterned wrappers. The background is blurred.
4. A doll with brown curly hair, blue eyes, and rosy cheeks wears a large black velvet hat with gold embroidery. The doll's dress is black with gold trim and a white scarf tied around its neck. The doll's face is white with painted features. Other dolls are visible in the background. The doll appears to be wearing a black velvet dress with gold trim. The doll's hat has a wide brim and a decorative band.
5. A vibrant orange rose is the focal point, its petals slightly unfurled. The rose is attached to a green stem with small thorns. The background is a blurred green foliage. The lighting is soft and natural. The rose appears to be in a natural setting.
6. A young woman with fair skin, dark brown hair styled in an updo with a decorative hairpiece, wears a light-colored, sheer, embroidered traditional East Asian garment. She looks directly at the camera with a slight smile. The garment has floral embroidery in white and light blue. She wears small, dangling earrings.
7. A green and pink parrot with a red beak eats a passion fruit. The parrot's head and neck are predominantly green, with a pink band around the neck. The beak is red and orange. The passion fruit is dark purple with a yellow interior and black seeds.
8. A reddish-orange mushroom with a white speckled stem grows in a forest floor covered with fallen leaves and grass. A small, light brown leaf rests on the mushroom's cap. The background features blurred green foliage and trees. The lighting is soft and diffused.
9. A white, fluffy dog lies in a field of green grass. The dog has its tongue out and is wearing a collar with a blue bone-shaped tag. The sun is low in the sky, creating a backlight. The dog's fur appears slightly ruffled. The grass is tall and slightly blurred. Trees are visible in the background. The sky is a gradient of blue and light yel-

low.

10. A bee is perched on a vibrant orange flower. The flower has multiple petals and a yellow center. Several other orange flowers and green leaves are visible in the background, some out of focus. The bee appears to be collecting nectar. The flowers are in a garden setting. The image is brightly lit, highlighting the orange hues of the flowers.
11. A clear glass teapot containing a yellow liquid sits on a round wooden tray. Several ripe red strawberries with green stems are placed on the tray alongside the teapot. The tray rests on a light brown, textured surface. A blurred background of green foliage suggests an outdoor setting. The teapot has a copper-colored handle and lid. The light is bright and natural.
12. A black mug with a gold design sits on a gray textured surface. The design depicts stylized mountains, trees, a crescent moon, and a campfire. The words "Mountain Kind" are written below the design. Steam rises from the mug. A lit candle, pine cones, cinnamon sticks, and a string of lights are also present. The background is blurred and dark.

References

[1] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923*, 2025. 1

[2] Dar-Yen Chen, Hmrishav Bandyopadhyay, Kai Zou, and Yi-Zhe Song. Nitrofusion: High-fidelity single-step diffusion through dynamic adversarial training. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025*, pages 7654–7663. Computer Vision Foundation / IEEE, 2025. 2

[3] Jiaxiang Cheng, Bing Ma, Xuhua Ren, Hongyi Jin, Kai Yu, Peng Zhang, Wenyue Li, Yuan Zhou, Tianxiang Zheng, and Qinglin Lu. Pose: Phased one-step adversarial equilibrium for video diffusion models. *arXiv preprint arXiv:2508.21019*, 2025. 4

[4] 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*. 4

[5] Xingtong Ge, Xin Zhang, Tongda Xu, Yi Zhang, Xinjie Zhang, Yan Wang, and Jun Zhang. Senseflow: Scaling distribution matching for flow-based text-to-image distillation. *CoRR*, abs/2506.00523, 2025. 4

[6] Yuval Kirstain, Adam Polyak, Uriel Singer, Shahbuland Matiana, Joe Penna, and Omer Levy. Pick-a-pic: An open dataset of user preferences for text-to-image generation. *Advances in neural information processing systems*, 36:36652–36663, 2023. 1

[7] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International conference on machine learning*, pages 12888–12900. PMLR, 2022. 1

[8] Zhanhao Liang, Yuhui Yuan, Shuyang Gu, Bohan Chen, Tiankai Hang, Mingxi Cheng, Ji Li, and Liang Zheng. Aesthetic post-training diffusion models from generic preferences with step-by-step preference optimization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025*, pages 13199–13208. Computer Vision Foundation / IEEE, 2025. 1

[9] Xinyao Liao, Wei Wei, Xiaoye Qu, and Yu Cheng. Step-level reward for free in rl-based t2i diffusion model fine-tuning. *arXiv preprint arXiv:2505.19196*, 2025. 1

[10] Shanchuan Lin, Anran Wang, and Xiao Yang. Sdxl-lightning: Progressive adversarial diffusion distillation. *arXiv preprint arXiv:2402.13929*, 2024. 2

[11] Weijian Luo. Diff-instruct++: Training one-step text-to-image generator model to align with human preferences. *arXiv preprint arXiv:2410.18881*, 2024. 4

[12] Weijian Luo, Tianyang Hu, Shifeng Zhang, Jiacheng Sun, Zhenguo Li, and Zhihua Zhang. Diff-instruct: A universal approach for transferring knowledge from pre-trained diffusion models. *Advances in Neural Information Processing Systems*, 36:76525–76546, 2023. 4

[13] Zichen Miao, Zhengyuan Yang, Kevin Lin, Ze Wang, Zicheng Liu, Lijuan Wang, and Qiang Qiu. Tuning timestep-distilled diffusion model using pairwise sample optimization. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*. OpenReview.net, 2025. 1, 2

[14] 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. *arXiv preprint arXiv:2307.01952*, 2023. 1

[15] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR, 2021. 1

[16] Yuxi Ren, Xin Xia, Yanzuo Lu, Jiacheng Zhang, Jie Wu, Pan Xie, Xing Wang, and Xuefeng Xiao. Hyper-sd: Trajectory segmented consistency model for efficient image synthesis. *Advances in Neural Information Processing Systems*, 37:117340–117362, 2024. 1, 2

[17] Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial diffusion distillation. In *European Conference on Computer Vision*, pages 87–103. Springer, 2024. 2

[18] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, et al. Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in neural information processing systems*, 35:25278–25294, 2022. 1, 3

[19] Xiaoshi Wu, Yiming Hao, Keqiang Sun, Yixiong Chen, Feng Zhu, Rui Zhao, and Hongsheng Li. Human preference score v2: A solid benchmark for evaluating human preferences of text-to-image synthesis. *arXiv preprint arXiv:2306.09341*, 2023. 1

[20] Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. Imagereward: Learning and evaluating human preferences for text-to-image generation. *Advances in Neural Information Processing Systems*, 36:15903–15935, 2023. 1

[21] Jiazheng Xu, Yu Huang, Jiale Cheng, Yuanming Yang, Jiajun Xu, Yuan Wang, Wenbo Duan, Shen Yang, Qunlin Jin, Shurun Li, Jiayan Teng, Zhuoyi Yang, Wendi Zheng, Xiao Liu, Ming Ding, Xiaohan Zhang, Xiaotao Gu, Shiyu Huang, Minlie Huang, Jie Tang, and Yuxiao Dong. Visionreward: Fine-grained multi-dimensional human preference learning for image and video generation, 2024. 1

[22] Tianwei Yin, Michaël Gharbi, Taesung Park, Richard Zhang, Eli Shechtman, Fredo Durand, and William T Freeman. Improved distribution matching distillation for fast image synthesis. In *NeurIPS*, 2024. 2

[23] Sixian Zhang, Bohan Wang, Junqiang Wu, Yan Li, Tingting Gao, Di Zhang, and Zhongyuan Wang. Learning multi-dimensional human preference for text-to-image generation.

In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8018–8027, 2024. 1

[24] Tao Zhang, Cheng Da, Kun Ding, Kun Jin, Yan Li, Tingting Gao, Di Zhang, Shiming Xiang, and Chunhong Pan. Diffusion model as a noise-aware latent reward model for step-level preference optimization. *CoRR*, abs/2502.01051, 2025. 1

[25] Tao Zhang, Cheng Da, Kun Ding, Huan Yang, Kun Jin, Yan Li, Tingting Gao, Di Zhang, Shiming Xiang, and Chunhong Pan. Diffusion model as a noise-aware latent reward model for step-level preference optimization. *arXiv preprint arXiv:2502.01051*, 2025. 1, 2