RaSS: Improving Denoising Diffusion Samplers with Reinforced Active Sampling Scheduler

Supplementary Material

1. The Traning Efficiency of RaSS

RaSS is highly efficient and surpasses existing methods that demand additional computational resources such as Progressive Distil [3], DDSS [4] and AutoDiffusion [2] in computational resource requirements. RaSS utilizes a very small model as the sampling scheduler, which significantly reduces the training resource requirements. Table. 1 demonstrates the superior efficiency of RaSS compared to AutoDiffusion, DDSS and PD. The computational resource required by Autodiffusion, DDSS and PD is approximately 1.25×, 3.98× and 377× that of RaSS

Approach	Steps	Method Type	Total cost (GPU Days)
RaSS	5	Training Small model	0.89
AutoDiffusion	5	Genetic Search	1.125
DDSS	5	Reparameterization	3.55
Progressive Distil	4	Distillation	359
Progressive Distil	8	Distillation	314

Table 1. The Training Efficiency comparison of RaSS. We assessed the computational resource demand of RaSS, AutoDiffusion, PD, and DDSS using our reconstructed Improved-Diffusion codebase and ImageNet 64 × 64 on a single V100 GPU.

2. Visualization Results of RASS Using DDIM as the Baseline Sampler at 15 Steps

We visualized RaSS results at 15 steps and were surprised to find that using only half the steps could achieve similar or even higher image quality compared to 25-step results, as shown in Fig. 1. This improvement is due to our careful selection of the state space $\mathcal{S} = \{P, I, \tau\}$ during the data collection phase, which provided essential information for optimizing the sampling scheduler.

3. The visual comparison using RASS on LSUN-Bedroom.

We conducted experiments on the LSUN-Bedroom dataset using the DDPM model [1], and the visualized results (see Fig. 2) for $\alpha = 1$ demonstrate that RaSS achieves higher generation quality with fewer steps compared to methods without RaSS.

References

[1] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 1

[2] Lijiang Li, Huixia Li, Xiawu Zheng, Jie Wu, Xuefeng Xiao, Rui Wang, Min Zheng, Xin Pan, Fei Chao, and Rongrong Ji. Autodiffusion: Training-free optimization of time steps and architectures for automated diffusion model acceleration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7105–7114, 2023. 1

[3] Chenlin Meng, Robin Rombach, Ruiqi Gao, Diederik Kingma, Stefano Ermon, Jonathan Ho, and Tim Salimans. On distillation of guided diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14297–14306, 2023. 1

[4] Daniel Watson, William Chan, Jonathan Ho, and Mohammad Norouzi. Learning fast samplers for diffusion models by differentiating through sample quality. In *International Conference on Learning Representations*, 2021. 1

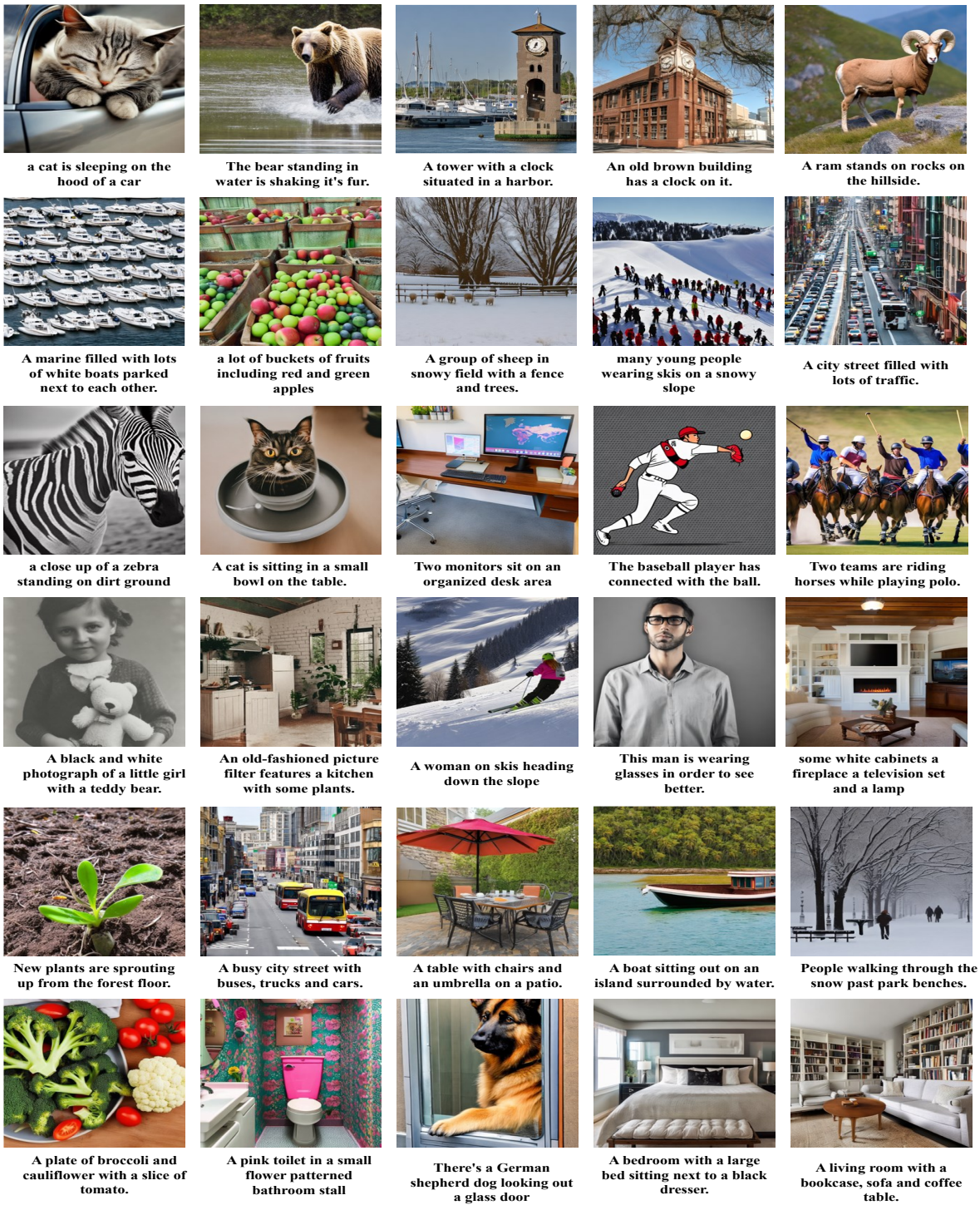

Figure 1. The visualization Results of RASS on MS-COCO.

Figure 2. The visual comparison using RASS on LSUN-Bedroom. The above is DDIM, the following is DDIM+RASS