REFERENCES

Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 17981–17993, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/958c530554f78bcd8e97125b70e6973d-Abstract.html.

David Berthelot, Arnaud Autef, Jierui Lin, Dian Ang Yap, Shuangfei Zhai, Siyuan Hu, Daniel Zheng, Walter Talbott, and Eric Gu. Tract: Denoising diffusion models with transitive closure time-distillation. *arXiv preprint arXiv:2303.04248*, 2023.

Nanxin Chen, Yu Zhang, Heiga Zen, Ron J. Weiss, Mohammad Norouzi, and William Chan. Wavegrad: Estimating gradients for waveform generation. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=NsMLjcFaO8O.

Quan Dao, Hao Phung, Trung Tuan Dao, Dimitris N. Metaxas, and Anh Tran. Self-corrected flow distillation for consistent one-step and few-step text-to-image generation. *CoRR*, abs/2412.16906, 2024. doi: 10.48550/ARXIV.2412.16906. URL https://doi.org/10.48550/arXiv.2412.16906.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

Martin Gonzalez, Nelson Fernandez Pinto, Thuy Tran, Hatem Hajri, Nader Masmoudi, et al. Seeds: Exponential sde solvers for fast high-quality sampling from diffusion models. *Advances in Neural Information Processing Systems*, 36, 2024.

Jiatao Gu, Shuangfei Zhai, Yizhe Zhang, Lingjie Liu, and Joshua M Susskind. Boot: Data-free distillation of denoising diffusion models with bootstrapping. In *ICML Workshop*, 2023.

Ishaan Gulrajani and Tatsunori B. Hashimoto. Likelihood-based diffusion language models. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL http://papers.nips.cc/paper_files/paper/2023/hash/35b5c175e139bff5f22a5361270fce87-Abstract-Conference.html.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *NeurIPS*, 30, 2017.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NeurIPS*, 33: 6840–6851, 2020.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35:26565–26577, 2022.

Dongjun Kim, Chieh-Hsin Lai, Wei-Hsiang Liao, Naoki Murata, Yuhta Takida, Toshimitsu Uesaka, Yutong He, Yuki Mitsufuji, and Stefano Ermon. Consistency trajectory models: Learning probability flow ODE trajectory of diffusion. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024. URL https://openreview.net/forum?id=ymjI8feDTD.

Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *Advances in neural information processing systems*, 34:21696–21707, 2021.

Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. Diffwave: A versatile diffusion model for audio synthesis. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=a-xFK8Ymz5J.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, MIT, NYU, 2009. CIFAR10 and CIFAR100 were collected by Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton.

Yujia Li, Kevin Swersky, and Richard S. Zemel. Generative moment matching networks. In Francis R. Bach and David M. Blei (eds.), *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pp. 1718–1727. JMLR.org, 2015. URL http://proceedings.mlr.press/v37/li15.html.

Tsung-Yi Lin, Michael Maire, Serge J. Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: common objects in context. In David J. Fleet, Tomás Pajdla, Bernt Schiele, and Tinne Tuytelaars (eds.), *Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V*, volume 8693 of *Lecture Notes in Computer Science*, pp. 740–755. Springer, 2014. doi: 10.1007/978-3-319-10602-1_48. URL https://doi.org/10.1007/978-3-319-10602-1_48.

Luping Liu, Yi Ren, Zhijie Lin, and Zhou Zhao. Pseudo numerical methods for diffusion models on manifolds. *arXiv preprint arXiv:2202.09778*, 2022.

Xingchao Liu, Xiwen Zhang, Jianzhu Ma, Jian Peng, et al. Instaflow: One step is enough for high-quality diffusion-based text-to-image generation. In *The Twelfth International Conference on Learning Representations*, 2023.

Aaron Lou, Chenlin Meng, and Stefano Ermon. Discrete diffusion modeling by estimating the ratios of the data distribution. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net, 2024. URL https://openreview.net/forum?id=CNicRIVIPA.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *Advances in Neural Information Processing Systems*, 35:5775–5787, 2022a.

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*, 2022b.

Eric Luhman and Troy Luhman. Knowledge distillation in iterative generative models for improved sampling speed. *arXiv preprint arXiv:2101.02388*, 2021.

Simian Luo, Yiqin Tan, Longbo Huang, Jian Li, and Hang Zhao. Latent consistency models: Synthesizing high-resolution images with few-step inference. *arXiv preprint arXiv:2310.04378*, 2023a.

Simian Luo, Yiqin Tan, Suraj Patil, Daniel Gu, Patrick von Platen, Apolinário Passos, Longbo Huang, Jian Li, and Hang Zhao. Lcm-lora: A universal stable-diffusion acceleration module. *CoRR*, abs/2311.05556, 2023b. doi: 10.48550/ARXIV.2311.05556. URL https://doi.org/10.48550/arXiv.2311.05556.

Weijian Luo, Colin Zhang, Debing Zhang, and Zhengyang Geng. Diff-instruct*: Towards human-preferred one-step text-to-image generative models. *CoRR*, abs/2410.20898, 2024. doi: 10.48550/ARXIV.2410.20898. URL https://doi.org/10.48550/arXiv.2410.20898.

Chenlin Meng, Robin Rombach, Ruiqi Gao, Diederik Kingma, Stefano Ermon, Jonathan Ho, and Tim Salimans. On distillation of guided diffusion models. In *CVPR*, pp. 14297–14306, 2023.

Thuan Hoang Nguyen and Anh Tran. Swiftbrush: One-step text-to-image diffusion model with variational score distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7807–7816, 2024.

Viet Nguyen, Anh Nguyen, Trung Tuan Dao, Khoi Nguyen, Cuong Pham, Toan Tran, and Anh Tran. SNOOPI: supercharged one-step diffusion distillation with proper guidance. *CoRR*, abs/2412.02687, 2024. doi: 10.48550/ARXIV.2412.02687. URL https://doi.org/10.48550/arXiv.2412.02687.

Ben Poole, Ajay Jain, Jonathan T Barron, and Ben Mildenhall. Dreamfusion: Text-to-3d using 2d diffusion. *arXiv preprint arXiv:2209.14988*, 2022.

Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. Dreamfusion: Text-to-3d using 2d diffusion. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/forum?id=FjNys5c7VyY.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, pp. 10684–10695, 2022.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. ImageNet large scale visual recognition challenge. *IJCV*, 115:211–252, 2015.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. *arXiv preprint arXiv:2202.00512*, 2022.

Tim Salimans, Thomas Mensink, Jonathan Heek, and Emiel Hoogeboom. Multistep distillation of diffusion models via moment matching. *CoRR*, abs/2406.04103, 2024. doi: 10.48550/ARXIV.2406.04103. URL https://doi.org/10.48550/arXiv.2406.04103.

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

Yang Song, Sahaj Garg, Jiaxin Shi, and Stefano Ermon. Sliced score matching: A scalable approach to density and score estimation. In Amir Globerson and Ricardo Silva (eds.), *Proceedings of the Thirty-Fifth Conference on Uncertainty in Artificial Intelligence, UAI 2019, Tel Aviv, Israel, July 22-25, 2019*, volume 115 of *Proceedings of Machine Learning Research*, pp. 574–584. AUAI Press, 2019. URL http://proceedings.mlr.press/v115/song20a.html.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020b.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. *arXiv preprint arXiv:2303.01469*, 2023.

Zhengyi Wang, Cheng Lu, Yikai Wang, Fan Bao, Chongxuan Li, Hang Su, and Jun Zhu. Prolificdreamer: High-fidelity and diverse text-to-3d generation with variational score distillation. *Advances in Neural Information Processing Systems*, 36, 2024.

Sirui Xie, Zhisheng Xiao, Diederik P Kingma, Tingbo Hou, Ying Nian Wu, Kevin Patrick Murphy, Tim Salimans, Ben Poole, and Ruiqi Gao. Em distillation for one-step diffusion models. *arXiv preprint arXiv:2405.16852*, 2024.

Chen Xu, Tianhui Song, Weixin Feng, Xubin Li, Tiezheng Ge, Bo Zheng, and Limin Wang. Accelerating image generation with sub-path linear approximation model. In Ales Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol (eds.), *Computer Vision - ECCV 2024 - 18th European Conference, Milan, Italy, September 29-October 4, 2024, Proceedings, Part LIII*, volume 15111 of *Lecture Notes in Computer Science*, pp. 323–339. Springer, 2024. doi: 10.1007/978-3-031-73668-1_19. URL https://doi.org/10.1007/978-3-031-73668-1_19.

Shuchen Xue, Mingyang Yi, Weijian Luo, Shifeng Zhang, Jiacheng Sun, Zhenguo Li, and Zhi-Ming Ma. Sa-solver: Stochastic adams solver for fast sampling of diffusion models. *Advances in Neural Information Processing Systems*, 36, 2024.

Hanshu Yan, Xingchao Liu, Jiachun Pan, Jun Hao Liew, Qiang Liu, and Jiashi Feng. Perflow: Piecewise rectified flow as universal plug-and-play accelerator. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang (eds.), *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024. URL http://papers.nips.cc/paper_files/paper/2024/hash/8f9d1362a386e80a723e98451a8c7564-Abstract-Conference.html.

Tianwei Yin, Michaël Gharbi, Taesung Park, Richard Zhang, Eli Shechtman, Fredo Durand, and William T Freeman. Improved distribution matching distillation for fast image synthesis. *arXiv preprint arXiv:2405.14867*, 2024a.

Tianwei Yin, Michaël Gharbi, Richard Zhang, Eli Shechtman, Fredo Durand, William T Freeman, and Taesung Park. One-step diffusion with distribution matching distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6613–6623, 2024b.

Qinsheng Zhang and Yongxin Chen. Fast sampling of diffusion models with exponential integrator. *arXiv preprint arXiv:2204.13902*, 2022.

Wenliang Zhao, Lujia Bai, Yongming Rao, Jie Zhou, and Jiwen Lu. Unipc: A unified predictor-corrector framework for fast sampling of diffusion models. *Advances in Neural Information Processing Systems*, 36, 2024.

Hongkai Zheng, Weili Nie, Arash Vahdat, Kamyar Azizzadenesheli, and Anima Anandkumar. Fast sampling of diffusion models via operator learning. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 42390–42402. PMLR, 2023a. URL https://proceedings.mlr.press/v202/zheng23d.html.

Kaiwen Zheng, Cheng Lu, Jianfei Chen, and Jun Zhu. Dpm-solver-v3: Improved diffusion ode solver with empirical model statistics. *Advances in Neural Information Processing Systems*, 36: 55502–55542, 2023b.

Mingyuan Zhou, Huangjie Zheng, Zhendong Wang, Mingzhang Yin, and Hai Huang. Score identity distillation: Exponentially fast distillation of pretrained diffusion models for one-step generation. In *Forty-first International Conference on Machine Learning*, 2024.

Mingyuan Zhou, Yi Gu, and Zhendong Wang. Few-step diffusion via score identity distillation. *CoRR*, abs/2505.12674, 2025. doi: 10.48550/ARXIV.2505.12674. URL https://doi.org/10.48550/arXiv.2505.12674.

A Implementation Details

This section presents an overview of our implementation details across CIFAR-10, ImageNet 64×64, and zero-shot COCO 2014.We closely follow the implementation of DMD and DMDv2 throughout our experiments, and open-source our training and evaluation code for reproducibility.

A.1 CIFAR-10

Stage 1 (One-step distillation)

We distill a one-step student model from the EDM Pretrained model Karras et al. (2022), using edm-cifar10-32x32-cond-vp for class-conditional image generation. Training is performed with AdamW optimizer (learning rate of 1e-5, a weight decay 0.01, and beta parameters $(0.9, 0.999)$), a total batch size of 256 on 4 A100 GPUs for 200K iterations. Following the original implementation, we adopt a two-time-update-rule: the adapted network is updated 5 times per a student update. The score identity distillation loss is implemented based on Score Identity Distillation Zhou et al. (2024) with default hyper-parameters: $\alpha = 1.2$ and loss_scaling_generator $= 100$.

Stage 2 and 3 (Progressive multi-step training)

In the subsequent training stages, we initialize the student model with the best FID checkpoint from the previous stage, and finetune with AdamW optimizer. For stage 2, we use a learning rate of 5e-6, a weight decay of 0.01, beta parameters $(0.9, 0.999)$, batch size 256, and train for 35K iterations with 4 A100 GPUs. For stage 3, we reduce the learning rate to 8e-7 and train for 65K iterations under the same setup. We also adopt a progressive training time-split schedule $[[0, 1], [0, 0.5], [0, 0.25]]$, where in the final stage the student learns the mapping from $x_{0.25}$ to x_0. We also evaluate an equal-width schedule $[[0, 1], [0, 0.67], [0, 0.33]]$, and observe similar results. Further exploration of scheduling strategies is left for future work. Our ablational studies show that different unforget weights are marginal effect; therefore, we decide to use an unforget weight of 1.0.

A.2 ImageNet 64×64

Stage 1 (One-step distillation)

We distill a one-step student model from the EDM Pretrained model Karras et al. (2022), using edm-imagenet-64x64-cond-adm for class-conditional image generation. Training is performed with AdamW optimizer (learning rate of 2e-6, a weight decay 0.01, and beta parameters $(0.9, 0.999)$), a total batch size of 72 on 4 A100 GPUs for 400K iterations. Following the original implementation, we adopt a two-time-update-rule: the adapted network is updated 5 times per a student update. The score identity distillation loss is implemented based on Score Identity Distillation with default hyper-parameters: $\alpha = 1.2$ and loss_scaling_generator $= 100$. .

Stage 2 and 3 (Progressive multi-step training)

Subsequent stages progressively extend the distilled student model to support multi-step sampling. Each stage is initialized from the best FID checkpoint of the previous stage and finetuned with AdamW.

- Stage 2: We use a learning rate of 8e-7, a weight decay of 0.01, beta parameters $(0.9, 0.999)$, batch size 72, unforget weight 0.3, and train for 40K iterations with 4 A100 GPUs, which takes approximately 1 day.
- Stage 3: We adopt the same configuration to train the 3-step PMDD model.

A.3 SD v1.5

Stage 1 (One-step distillation)

We distill a one-step student model from the SD v1.5 model Rombach et al. (2022) using prompts from LAION-Aesthetic 6.25+ dataset. Unlike DMD v2, our approach does not require collecting images for the GAN discriminator. Training is performed with AdamW optimizer (learning rate of 5e-6, a weight decay 0.01, and beta parameters $(0.9, 0.999)$), a total batch size of 120 on 3

H100 GPUs for 120K iterations. Following prior work, we adopt a two-time-update-rule where the adapted network is updated 10 times per student update. We also implemented the score identity distillation loss and found that it did not yield meaningful improvements during training.

Stage 2 and 3 (Progressive multi-step training)

Subsequent stages progressively extend the distilled student model to support multi-step sampling. Each stage is initialized from the best FID checkpoint of the previous stage and finetuned with AdamW.

- Stage 2: We use a learning rate of 8e-7, a weight decay of 0.01, beta parameters $(0.9, 0.999)$, batch size 20, unforget weight 0.3, and train for 16K iterations with 1 H100 GPU, which takes approximately 12 hours. To accelerate convergence, we reduce the adapted network updates to 5 per student update.
- Stage 3: Using the same configuration, we train the 3-step PMDD model. Based on ablation findings that high unforget weights overemphasize preservation at the expense of distribution matching, we lower the unforget weight to 0.01 and continue training for another 16K iterations.

B Evaluation Details

For the CIFAR-10 and ImageNet 64×64 datasets, we generate 50,000 images every 1000 iterations, and calculate the FID metric based on EDM's evaluation code Karras et al. (2022). For the zero-shot COCO 2014 text-to-image generation, we closely follow the evaluation process as reported in DMD Yin et al. (2024b), in which we generate 30,000 images using random prompts from the MS-COCO2014 validation set. These images are then downsampled to 256×256, and evaluated against roughly 40,000 real images from the validation set to calculate the FID metric. We also use the OpenCLIP-G backbone to report the CLIP score.

C Prompts for Figures

We adopt DMD's paper and use the following prompts to generate the figures:

- A DSLR photo of a golden retriever in heavy snow.
- Medium shot side profile portrait photo of a warrior chief, sharp facial features, with tribal panther makeup in blue and red, looking away, serious but clear eyes, 50mm portrait, photography, hard rim lighting photography.
- A hyperrealistic photo of a fox astronaut; perfect face, artstation.
- Create an image that depicts a majestic kingdom with towering castles.
- Transparent vacation pod at dramatic scottish lochside, concept prototype, ultra clear plastic material, editorial style photograph.
- A red sports car parked on a city street at night, photorealistic, sharp focus, HDR, cinematic lighting, ultra-detailed reflections.
- A snowy village with houses decorated with Christmas lights, photorealistic, cinematic lighting, high resolution, ultra-detailed textures.
- A scenic mountain landscape with a lake, photorealistic, ultra-detailed, cinematic lighting, high resolution.
- A dense forest with sunlight filtering through the trees, photorealistic, cinematic composition, sharp details, ultra-detailed textures.
- A plate of pasta on a dining table, photorealistic, food photography style, sharp focus, studio lighting, ultra-detailed textures.

D Additional Qualitative Samples

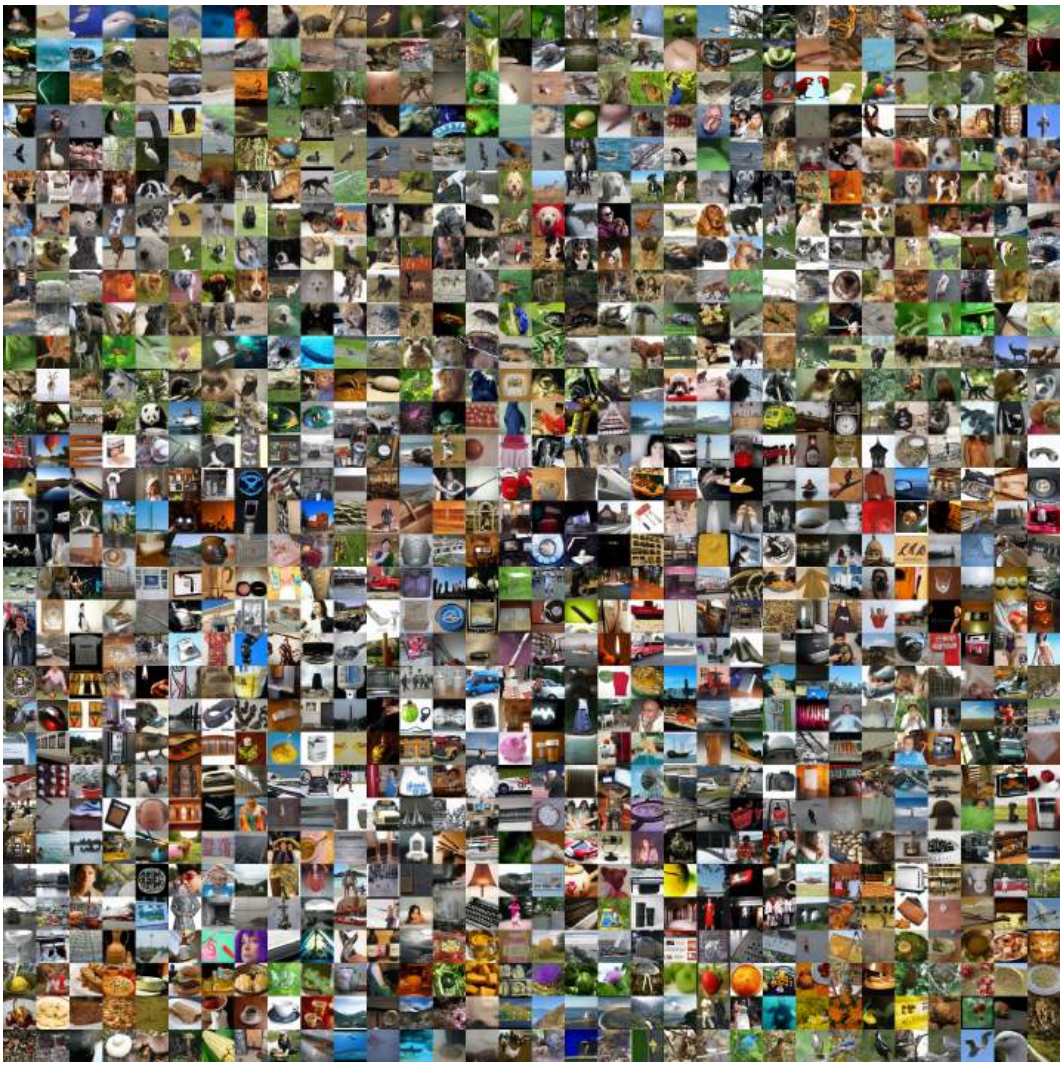

Figure 5: Samples from our 1-step student generator on ImageNet (FID=2.60).

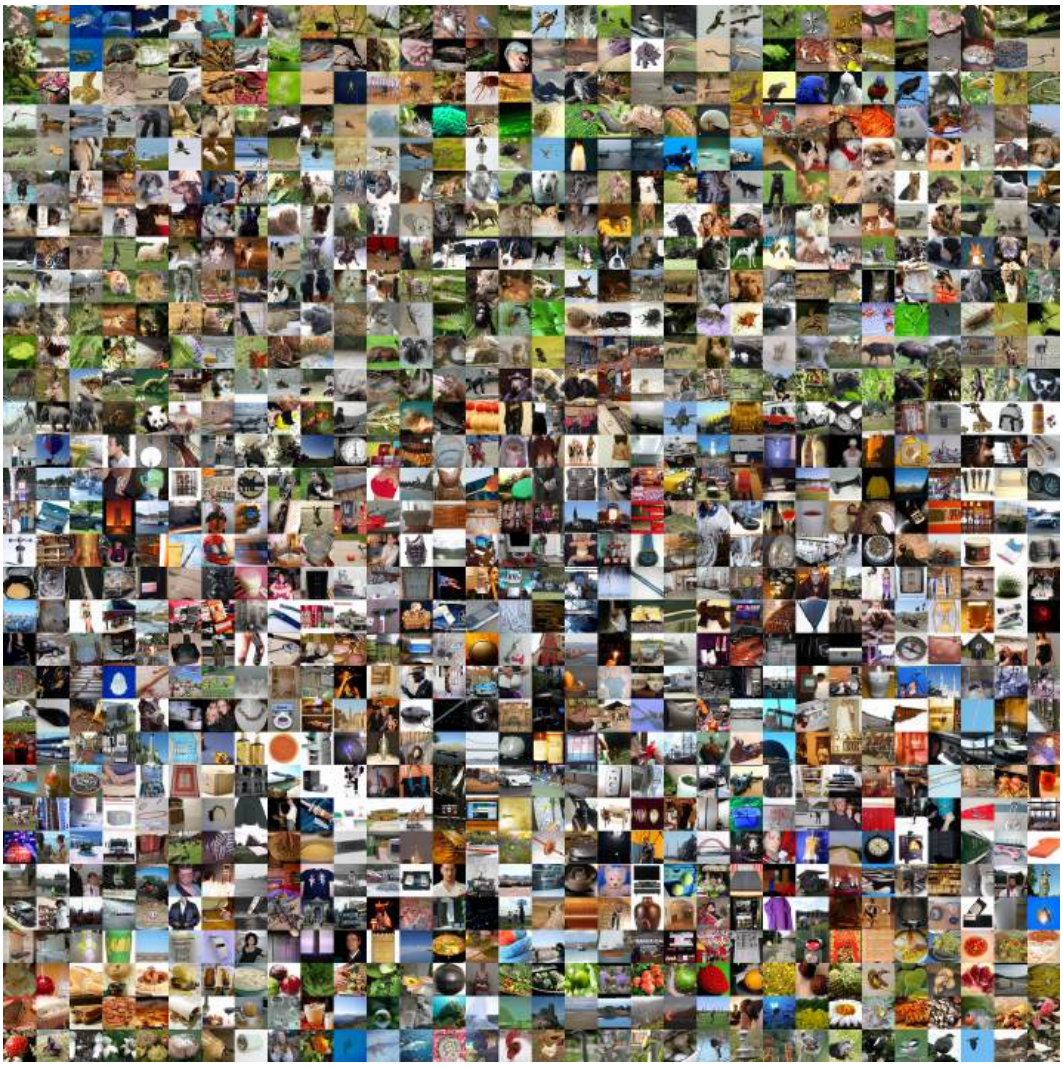

Figure 6: Samples from our 2-step student generator on ImageNet (FID=1.95).

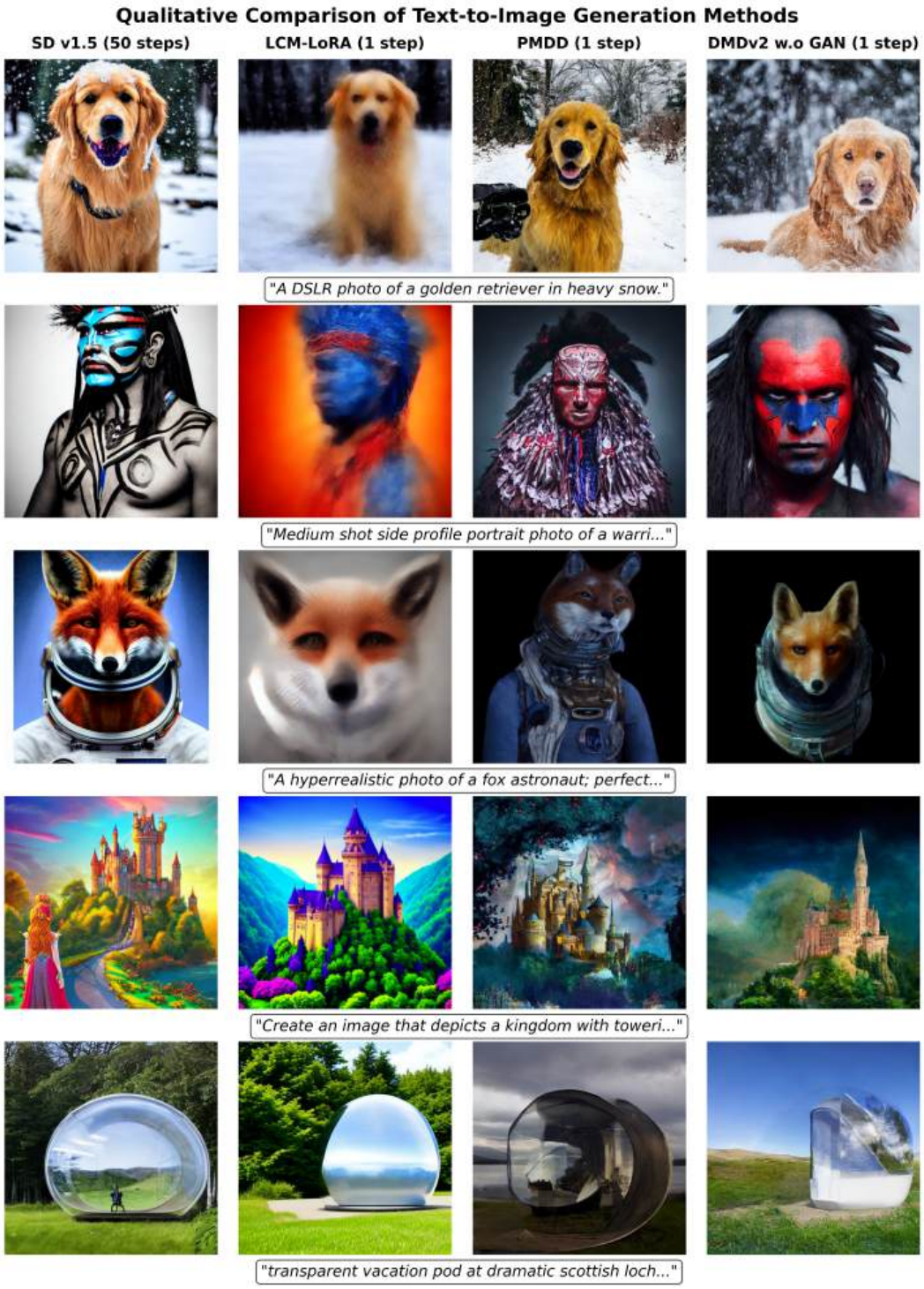

Figure 7: Comparison of text-to-image generation across Stable Diffusion v1.5 (50 steps) and one-step diffusion distillation methods such as LCM-LoRA, PMDD, and DMD v2.