StyleGAN-Fusion: Diffusion Guided Domain Adaptation of Image Generators (Supplementary Materials)

Kunpeng Song[1] Ligong Han[1] Bingchen Liu[2] Dimitris Metaxas[1] Ahmed Elgammal[1,3]
[1]Rutgers University [2]Bytedance Inc. [3]Playform AI

A. Overview

In the supplementary material, we provide the full text prompts used in experiments, the implementation details, and additional quantitative evaluation results. Finally, we provide an extension of our method to DreamBooth [10] (Fig. 11) and additional visual results for multiple models and prompts (Fig. 12,13,14,15,16,17,18).

B. Full-Text Prompts

We provide full-text prompts used in the experiments mentioned in our paper:

1. “3d human face, closeup cute and adorable, cute big circular reflective eyes, Pixar render, unreal engine cinematic smooth, intricate detail, cinematic”
2. “3d cat, closeup cute and adorable, cute big circular reflective eyes, Pixar render, unreal engine cinematic smooth, intricate detail, cinematic”
3. “Joker”
4. “High quality 3 d render of very fluffy cat, highly detailed, unreal engine cinematic smooth, in the style of detective Pikachu blade runner, hannah yata charlie immer, neon light, low angle, uhd 8 k, sharp focus”
5. “An epic fantasy comic book style portrait painting of dog, very expressive, light blue piercing eyes, round face, character design by mark ryden and pixar and hayao miyazaki, unreal 5, daz, hyperrealistic, octane render, cosplay, rpg portrait, dynamic lighting, intricate detail, summer vibrancy, cinematic”
6. “Werewolf”
7. “Photo of a dog/hamster/badger/fox/otter/lion/ bear/pig”
8. “Cinematic portrait of brutal epic dark dog, concept art, artstation, glowing lights, highly detailed”
9. “Photo of a car, TRON wheel”
10. “Sketch of a car, pen and ink sketch”
11. “A masterpiece ultrarealistic ultradetailed portrait of a incredibly beautiful human face, baroque renaissance, in the night forest. medium shot, intricate, elegant, highly detailed. trending on artstation, digital art, by stanley artgerm lau, wlop, rossdraws, james jean, andrei riabovitchev, marc simonetti, yoshitaka amano. background by james jean and gustav klimt, light by julie bell, 4 k, porcelain skin.
12. “Very beautiful portrait of an extremely cute and adorable face, smooth, perfect face, fantasy, character design by mark ryden and pixar and hayao miyazaki, sharp focus, concept art, harvest fall vibrancy, intricate detail, cinematic lighting, hyperrealistic, 3 5 mm, diorama macro photography, 8 k, 4 k”
13. “Charcoal pencil sketch of human face, lower third, high contrast, black and white”
14. “A very beautiful anime girl, full body, long braided curly silver hair, sky blue eyes, full round face, short smile, casual clothes, ice snowy lake setting, cinematic lightning, medium shot, mid-shot, highly detailed, trending on Artstation, Unreal Engine 4k, cinematic wallpaper by Stanley Artgerm Lau, WLOP, Rossdraws, James Jean, Andrei Riabovitchev, Marc Simonetti, and Sakimichan”

The prompt IDs used in experiments in main text are:

- Figure 1: Face – prompt 1, Cat – prompt 2.
- Figure 3: (Face) → Joker – prompt 3, (Cat) → Pikachu Cat – prompt 4, (Dog) → Comic Dog – prompt 5, (Face) → Werewolf – prompt 6, (Cat) → Dog – prompt 7, (Dog) → Epic Dark Dog – prompt 8.
- Figure 4, 7: prompt 7
- Figure 5, 8, 9, 10, 11: prompt 1.
- Figure 6, 11: prompt 2.
- Table 1, 2: prompt 7. Table 3: prompt 1.

	Ours			**NADA**
	T750	T500	T300	
Dog	165.0258	155.1359	**150.7622**	206.9277
Fox	55.1440	54.1920	**51.5124**	90.4007
Lion	59.4057	35.1491	**30.3362**	153.8199
Tiger	19.9225	**17.0460**	19.2870	115.4611
Wolf	66.3091	**42.5760**	45.3286	139.6573

Table 3. FID scores of Cat → Animals. Our method outperforms baseline by a large margin.

	Ours			**NADA**
	T750	T500	T300	
Cat	**115.1011**	130.2553	124.7166	139.3474
Fox	65.3474	67.3951	**61.0955**	129.5795
Lion	67.3678	55.3808	**52.5201**	173.8075
Tiger	**26.9477**	28.1698	31.1502	223.3331
Wolf	133.7088	81.0359	**71.2905**	159.9959

Table 4. FID scores of Dog → Animals. Our method achieves significantly better FIDs in all T_{SDS} settings.

C. Implementation Details

Architecture We use the StyleGAN2 PyTorch [8] config-f implementation by [5]. Checkpoint resolutions are: FFHQ 1024×1024, AFHQ-Cat/Dog 512×512. We adapt the StableDiffusion v1.4 [9] model from Diffusers [2]. Training code are built upon Stable-DreamFusion [6] and StyleGAN-NADA [3]

Training The latent mapping MLP layer, all ToRGB layers, and bias are frozen and we only update weights in Conv layers. We used Adam optimizer with default parameters and a learning rate of 5×10^{-4}. All models were trained for 2000 training steps with batch size of 1. Training takes about 20 minutes on an A100 GPU with a memory cost of 14.7G for 1024 resolution and 12.5G for 512 resolution. To increase the stability of training, we clamp the gradient with respect to the generated image $\mathbf{x}$ by its 95% quantile.

D. More Quantitative Evaluations

Fréchet Inception Distance To further quantitatively measure the quality of generated images, we calculate and compare the clean Fréchet Inception Distance score [7] for the animal experiments. Ground truth images are needed when calculating FID. We use the AFHQ dataset [1] and manually extract ground truth images for Fox/Lion/Tiger/Wolf from its "wild" subclass. We use the default batch size of 256 for all FID calculations.

Tab. 3 shows the FID scores for Cat-to-Animals. And Tab. 4 for Dog-to-Animals. Notice, none of these models are exposed to a single ground truth image and the scores are attained in a zero-shot manner.

Our method achieved significantly better FID scores than the baseline: StyleGAN-NADA [3] in all cat/dog-to-animals experiments.

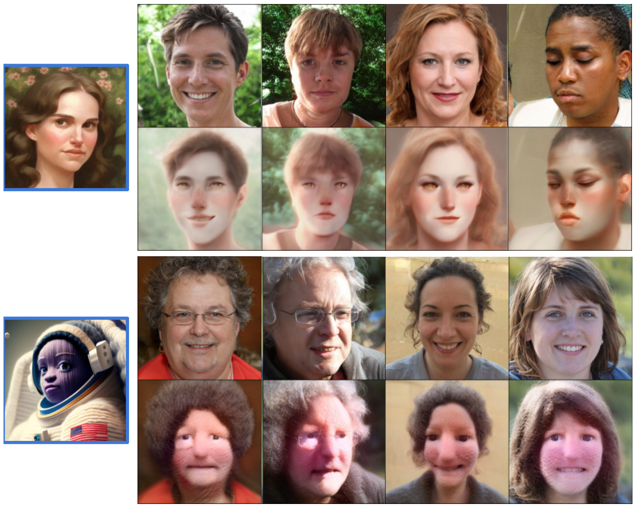


Figure 11. Results from DreamBooth guided models. We show the original DreamBooth StableDiffusion samples in blue boxes.

E. Extension to DreamBooth

We additionally extend our method to DreamBooth [10] where the StableDiffusion model is finetuned on a few personalized images. We tried public available DreamBooth checkpoints "Wa-vy" style [11] and "Woolitize" style [4]. Fig. 11 shows our results using the text prompt "wa-vy style painting close up face" and "woolitize close up face". The mesh geometries are smoother with larger eye areas as requested by the text prompt.

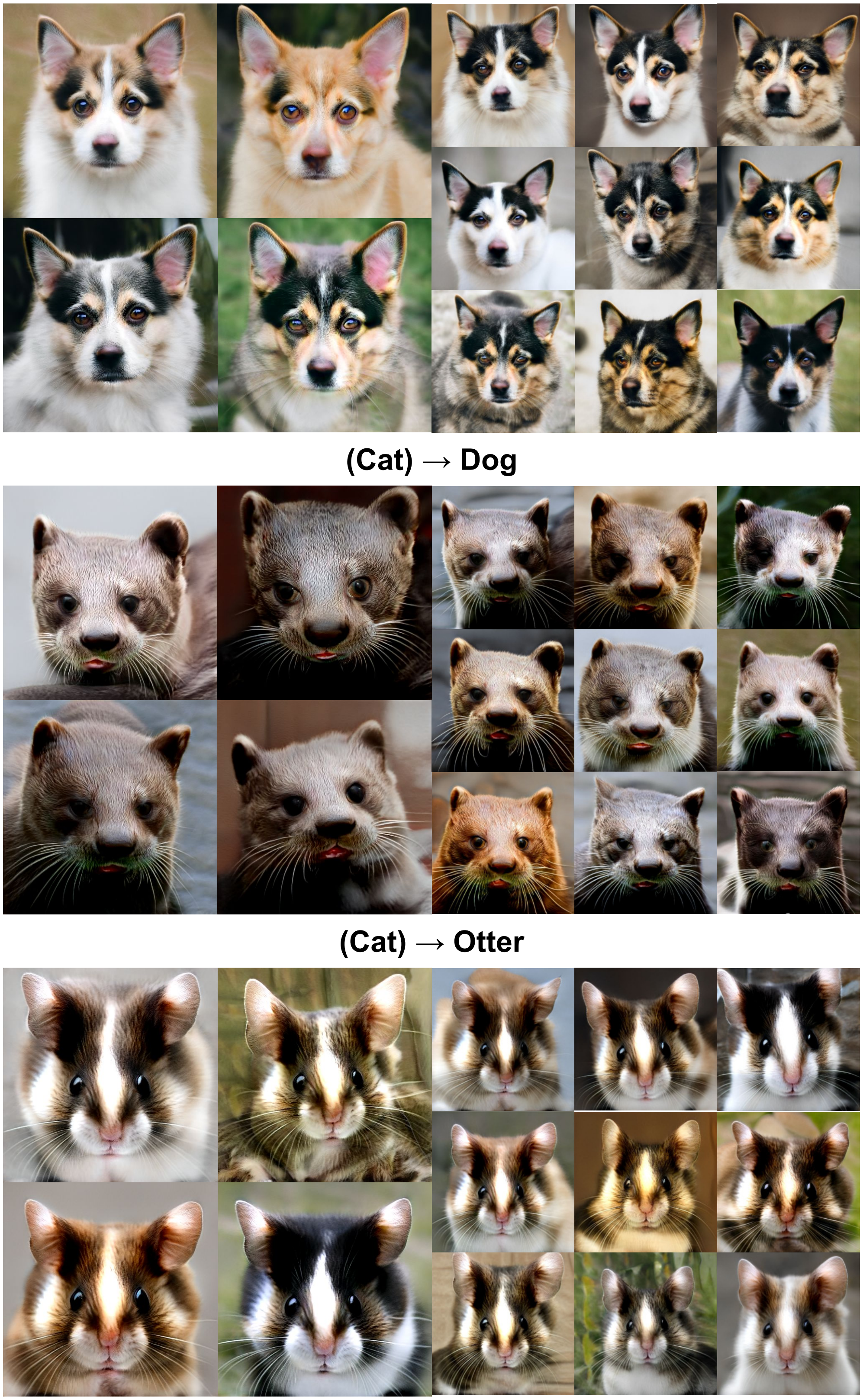
F. Large-Scale Image Galleries

We provide additional visual results for multiple models and prompts. Fig. 12 and Fig. 13 are generated images from our model for the cat-to-animals experiments. Fig. 14 shows the results from baseline: StyleGAN-NADA. Fig. 15 and Fig. 16 are our results from FFHQ face model. Additional results for car and landscape are shown in Fig. 17 and Fig. 18.

References

[1] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. CoRR, abs/1912.01865, 2019. 2

[2] diffusers. Stable diffusion v1-4 model, 2022. https://github.com/rosinality/stylegan2-pytorch, Last accessed on 2022-11-15. 2

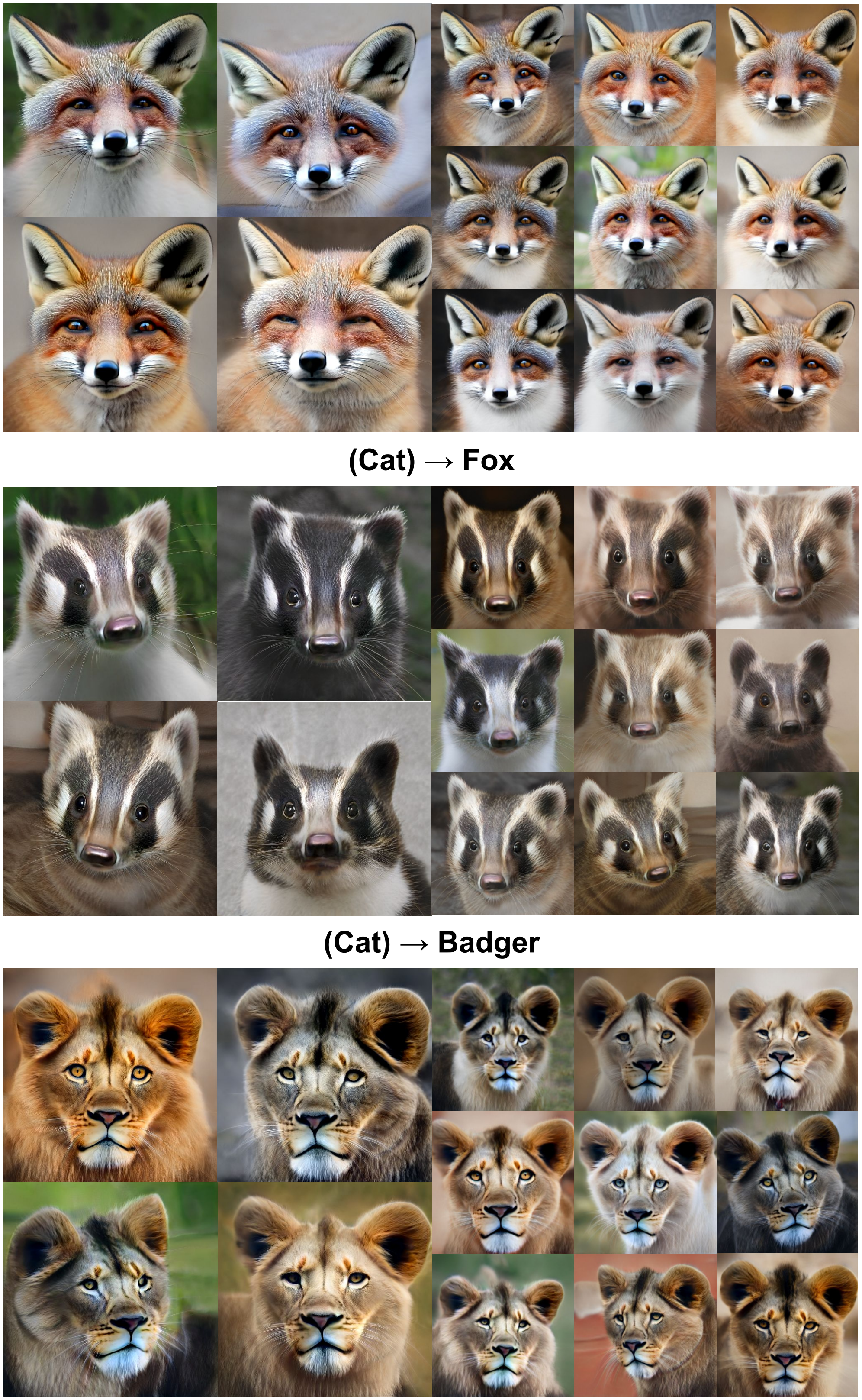


(Cat) → Dog

(Cat) → Otter

(Cat) → Hamster

Figure 12. Our results for AFHQ-cat to other animals.

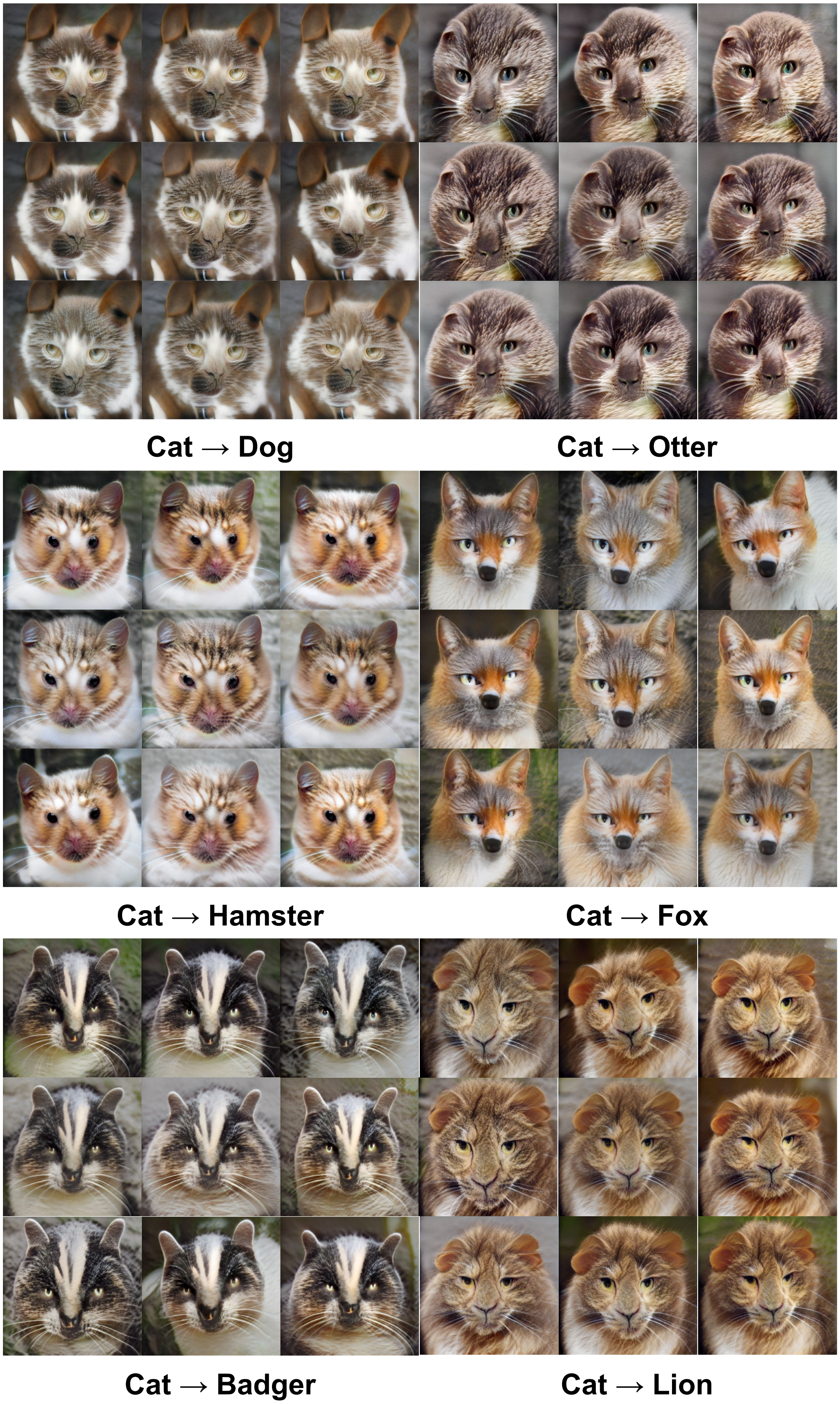


(Cat) → Fox

(Cat) → Badger

(Cat) → Lion

Figure 13. Our results for AFHQ-cat to other animals.



Cat → Dog

Cat → Otter

Cat → Hamster

Cat → Fox

Cat → Badger

Cat → Lion

Figure 14. Results from baseline StyleGAN-NADA for AFHQ-Cat to other animals. Baseline has distorted facial components, unrealistic texture, and lower diversity than ours.

	CLIP ↑				LPIPS ↑			
	T990	T850	T750	T500	T990	T850	T750	T500
dog	0.283	0.281	0.282	0.279	0.407	0.507	0.521	0.557
hamseter	0.302	0.294	0.290	0.288	0.322	0.426	0.433	0.522
badger	0.312	0.304	0.304	0.301	0.407	0.397	0.448	0.517
fox	0.312	0.307	0.301	0.308	0.433	0.469	0.500	0.502
otter	0.304	0.295	0.293	0.297	0.452	0.482	0.501	0.508
bear	0.289	0.293	0.293	0.288	0.421	0.448	0.459	0.501

Table 5. Details about CLIP-LPIPS trade-off.

[3] Rinon Gal, Or Patashnik, Haggai Maron, Amit H Bermano, Gal Chechik, and Daniel Cohen-Or. Stylegan-nada: Clip-guided domain adaptation of image generators. ACM Transactions on Graphics (TOG), 41(4):1–13, 2022. 2

[4] Ahsen Khaliq. Dreambooth woolitize. Hugging Face, 2022. https://huggingface.co/plasmo/woolitize-768sd1-5. 2

[5] Kim Seonghyeon. Stylegan 2 in pytorch, 2020. https://github.com/rosinality/stylegan2-pytorch, Last accessed on 2022-11-15. 2

[6] nanashi kiui. Stable-dreamfusion, 2022. https://github.com/ashawkey/stable-dreamfusion, Last accessed on 2022-11-15. 2

[7] Gaurav Parmar, Richard Zhang, and Jun-Yan Zhu. On buggy resizing libraries and surprising subtleties in FID calculation. CoRR, abs/2104.11222, 2021. 2

[8] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, Advances in Neural Information Processing Systems 32, pages 8024–8035. Curran Associates, Inc., 2019. 2

[9] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models, 2021. 2

[10] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. arXiv preprint arXiv:2208.12242, 2022. 1, 2

[11] wavymulder. Dreambooth wa-vy style. Hugging Face, 2022. https://huggingface.co/wavymulder/wavyfusion. 2

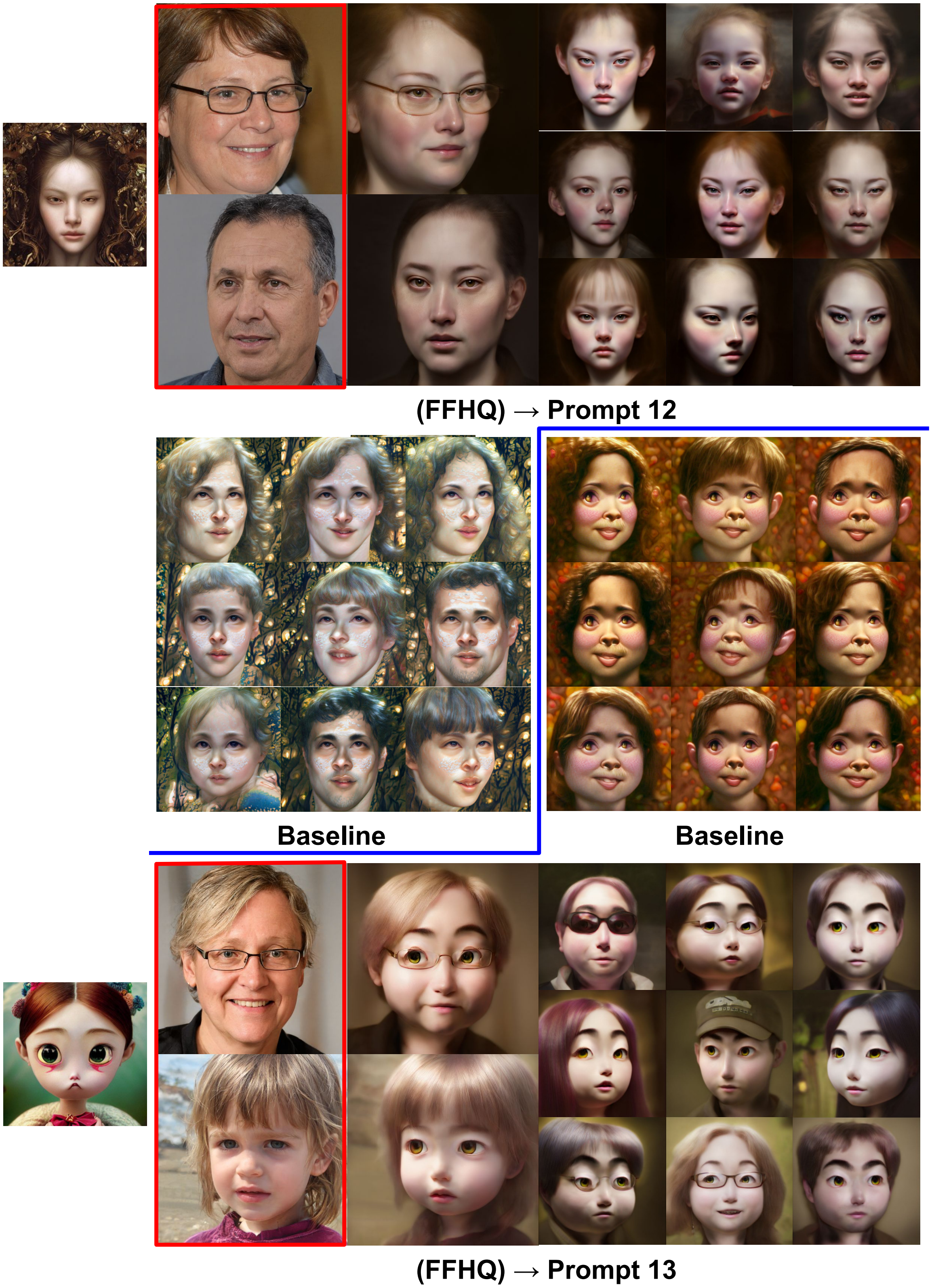


Figure 15. Addtional results from our model and baseline on FFHQ face experiments. Full-text prompts are available in Appendix B. On the left side, we show one sample from StableDiffusion for each prompt. Samples of $\mathcal{G}_{frozen}$ are marked in red boxes.

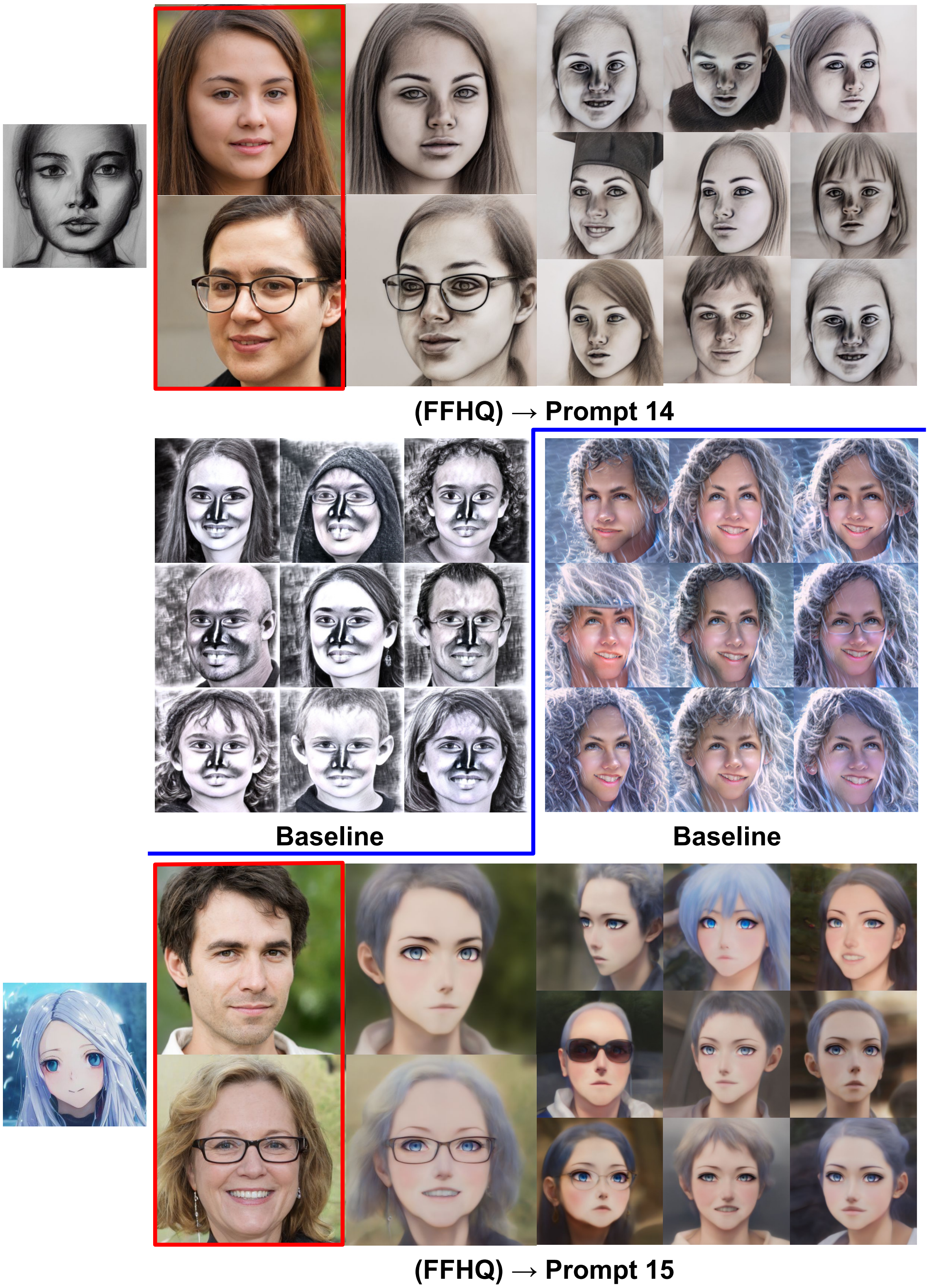


Figure 16. Addtional results from our model and baseline on FFHQ face experiments. Full-text prompts are available in Appendix B. On the left side, we show one sample from StableDiffusion for each prompt. Samples of $\mathcal{G}_{frozen}$ are marked in red boxes.

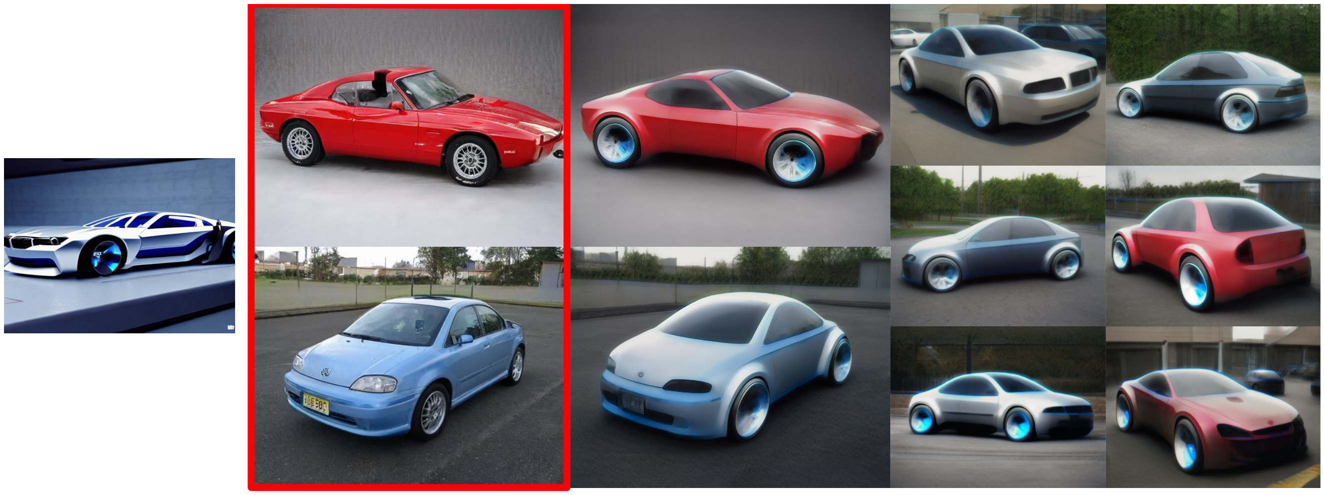


Figure 17. Addtional results from our model on Car to concept car experiments. Full-text prompt: Cyberpunk BMW concept-inspired sports car on the road, futuristic look, highly detailed body, very expensive, photorealistic camera shot, bright studio setting, light reflections, unreal engine 5 quality render. On the left side, we show one sample from StableDiffusion for each prompt. Samples from $\mathcal{G}_{frozen}$ are marked in the red box.

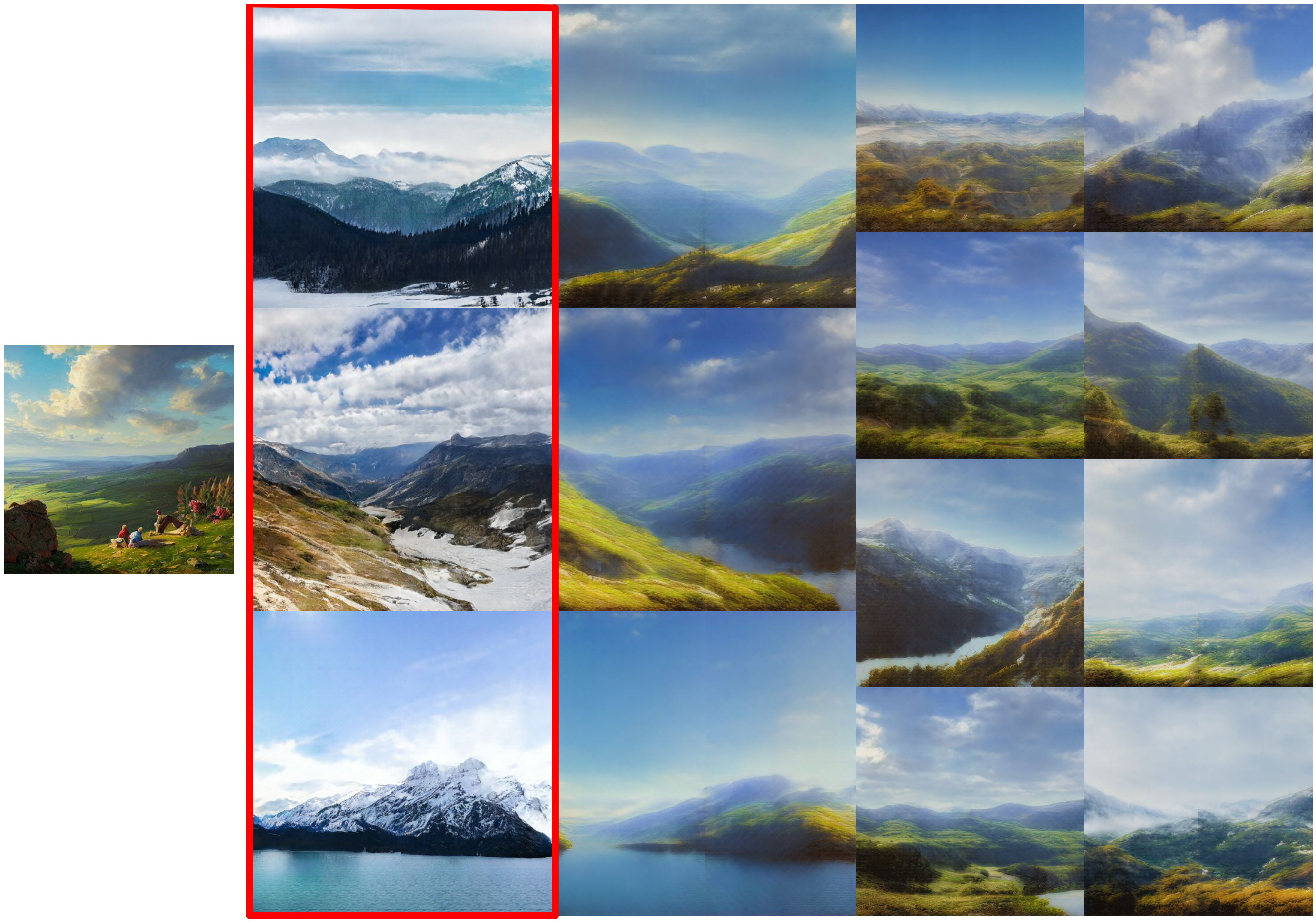


Figure 18. Addtional results from our model on Landscape experiment. Full-text prompt: vast view of landscape by Vladimir Volegov and Alexander Averin and Peder Mørk Mønsted and Adrian Smith and Raphael Lacoste. On the left side, we show one sample from StableDiffusion for each prompt. Samples from $\mathcal{G}_{frozen}$ are marked in the red box.