

FIND: Fine-tuning Initial Noise Distribution with Policy Optimization for Diffusion Models

Appendix

ACM Reference Format:

. 2024. FIND: Fine-tuning Initial Noise Distribution with Policy Optimization for Diffusion Models Appendix. In . ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

A DERIVATIONS

A.1 The proof of Eq.8

$$\begin{aligned}
 \nabla_{\theta} \mathbb{E}_{\pi_{\theta}} [f_r(\mathbf{c}, \mathbf{z}_0)] &= \nabla_{\theta} \int \pi_{\theta}(\mathbf{z}_T) f_r(\mathbf{c}, \mathbf{z}_0) d\mathbf{z}_T \\
 &= \int \nabla_{\theta} \pi_{\theta}(\mathbf{z}_T) f_r(\mathbf{c}, \mathbf{z}_0) d\mathbf{z}_T \\
 &= \int \pi_{\theta}(\mathbf{z}_T) \nabla_{\theta} \log \pi_{\theta}(\mathbf{z}_T) f_r(\mathbf{c}, \mathbf{z}_0) d\mathbf{z}_T \\
 &= \mathbb{E}_{\pi_{\theta}} [f_r(\mathbf{c}, \mathbf{z}_0) \nabla_{\theta} \log \pi_{\theta}],
 \end{aligned} \tag{1}$$

where the second to last equality is from the interchangeability of gradients and integration (summation), along with the log-derivative trick: $\nabla_{\theta} \log \pi_{\theta}(\mathbf{z}_T) = \frac{\nabla_{\theta} \pi_{\theta}(\mathbf{z}_T)}{\pi_{\theta}(\mathbf{z}_T)}$.

A.2 The proof of Eq.12

$$\begin{aligned}
 \nabla_{\theta} \mathbb{E}_{\pi_{\theta_{\text{old}}}} \left[r^* \frac{\pi_{\theta}}{\pi_{\theta_{\text{old}}}} \right] &= \mathbb{E}_{\pi_{\theta_{\text{old}}}} \left[r^* \frac{\nabla_{\theta} \pi_{\theta}}{\pi_{\theta_{\text{old}}}} \right] \\
 &= \mathbb{E}_{\pi_{\theta_{\text{old}}}} \left[r^* \frac{\pi_{\theta}}{\pi_{\theta_{\text{old}}}} \nabla_{\theta} \log \pi_{\theta} \right]
 \end{aligned} \tag{2}$$

B MORE EXAMPLES ON DRAWBENCH

In this section, we present additional visualization effects based on prompts from the Drawbench [2] dataset. As illustrated in Fig.2, our method effectively ensures that the images generated on the Drawbench dataset are consistent with the prompts. It is worth mentioning that our method optimizes the initial distribution, so after a single optimization, we can sample multiple images, all of which will have greater consistency with the prompts.

C EXPERIMENT ON COMPLEX PROMPT

To further validate the effectiveness of our approach, we designed an evaluation to assess the performance of our model under complex prompt conditions.

Unpublished working draft. Not for distribution.

Permission to make digital or hard copies of all or part of this work for personal or professional use, is granted by ACM, provided that the copies are distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
 Conference'17, July 2017, Washington, DC, USA
 © 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
 ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

Table 1: Quantity results on baseline method and SOTA method and ours. Both ImageReward and Aesthetic Score are such that higher values indicate better performance.

	ImageReward	Aesthetic	Time(min)
Baseline	0.05	5.38	-
DPOK	-0.02	5.40	920
Ours	0.11	5.43	621

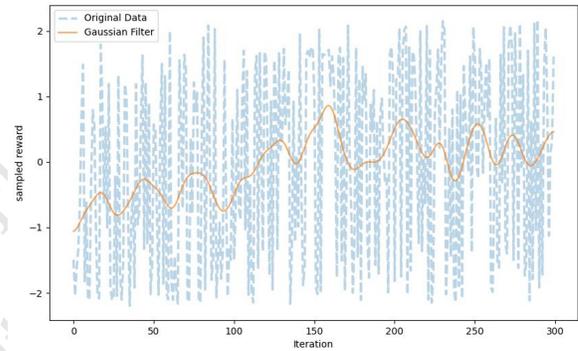


Figure 1: The line graph of the sample rewards during the optimization process, where the blue dashed line represents the original reward graph, and the yellow solid line represents the curve after Gaussian smoothing.

Evaluation dataset. For complex prompt conditions, we select 45 complex prompts from the showcase-sharing platform ¹. For each prompt, we generate 10 images for testing.

Quality Evaluation As demonstrated in Fig.3 and Fig.4, we selected 12 prompts from the evaluation dataset to prove the effectiveness of our method. The results show that the images generated after optimization using our approach exhibit stronger consistency with the prompts compared to those generated by the baseline model and DPOK [1].

Quantity Evaluation Similar to the setting in Section 5.2, we use the ImageReward [4] to assess the consistency between the generated content and the prompts, the Aesthetic Score [3] to evaluate the quality of the generated images, and Time as the total duration used to train and optimize the evaluation dataset. For DPOK [1], we employ a multi-prompt setting, where a single model is trained with multiple prompts over 10,000 iterations. As shown in Table 1, our method achieves significant improvements in both consistency and aesthetics compared to the baseline model. DPOK shows weaker learning capabilities for multiple complex prompts, leading

¹<https://www.midjourney.com/showcase?tab=hot>

59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116

to performance degradation compared to the baseline. Additionally, our method also uses less training time than DPOK.

Reward Visualization To further validate the effectiveness of our optimization method, we visualized the reward values sampled during a single optimization process. As shown in Fig.1, we visualized the reward trajectory for the optimization process of the *strawberry poster, 5 strawberries falling into the water, water splashes, black solid background, illustration, small details 8k* prompt. The background blue dashed line represents the reward curve obtained from sampling during each optimization process, and the yellow line is the reward curve after Gaussian smoothing. We observe that after 150 iterations, the upward trend of the reward slows down. Therefore, considering the time consumption, we set the number of optimization steps to 150.

D EXPERIMENT ON CUSTOM MODELS

To validate the generalization performance of our method, we selected three different baseline models from the open-source diffusion model website Civitai²: DreamShaper³, SemiRealistic⁴, and Serenity⁵, for generating images and aligning them with prompts using the same settings as described in Section 5.1. As shown in Fig.5, our method is also capable of effectively optimizing the initial distribution for different custom models, ensuring alignment with the prompts.

E DISCUSSION, LIMITATION, AND FUTURE WORK

Our method treats diffusion models as a black box, optimizing their initial distribution directly based on their outputs. Hence, it can be applied to any diffusion model, whether it's for generating images, videos, 3D content, etc. However, from another perspective, the generative capability of our approach is highly dependent on the baseline diffusion model. Essentially, our method seeks to unlock the latent potential of large-scale pre-trained models. In other words, all examples provided in the text could potentially be obtained through extensive sampling of the baseline model. The core of our method is to increase the likelihood of generating images that are consistent with the user-provided prompts. If content cannot be generated by the baseline diffusion model, it would still be unachievable even after optimizing the initial distribution through our method. Another limitation of our method lies in the reward function used. Since our reward function is a pre-trained network rather than a clearly defined logical value, the performance of the reward model employed for optimization is closely related to the effectiveness of the optimization.

In our future work, we plan to focus on two main aspects. The first aspect involves experimenting with more baseline models; besides the text-to-image and text-to-video mentioned in the article, we will explore optimizing the initial distribution for various diffusion models, including text-to-3D and robot control, among

others. The second aspect is to experiment with using large language models as our reward model, such as ChatGPT⁶, to see if we can achieve more precise control.

REFERENCES

- [1] Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. 2024. Reinforcement learning for fine-tuning text-to-image diffusion models. *Advances in Neural Information Processing Systems* 36 (2024).
- [2] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. 2022. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in neural information processing systems* 35 (2022), 36479–36494.
- [3] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, et al. 2022. Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems* 35 (2022), 25278–25294.
- [4] Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. 2023. ImageReward: Learning and Evaluating Human Preferences for Text-to-Image Generation. arXiv:2304.05977 [cs.CV]

²<https://civitai.com/>

³<https://civitai.com/models/4384/dreamshaper>

⁴<https://civitai.com/models/144120/merge-semirealistic-yesmixon1535percentphoton65percent>

⁵<https://civitai.com/models/110426/serenity>

⁶<https://chat.openai.com/>

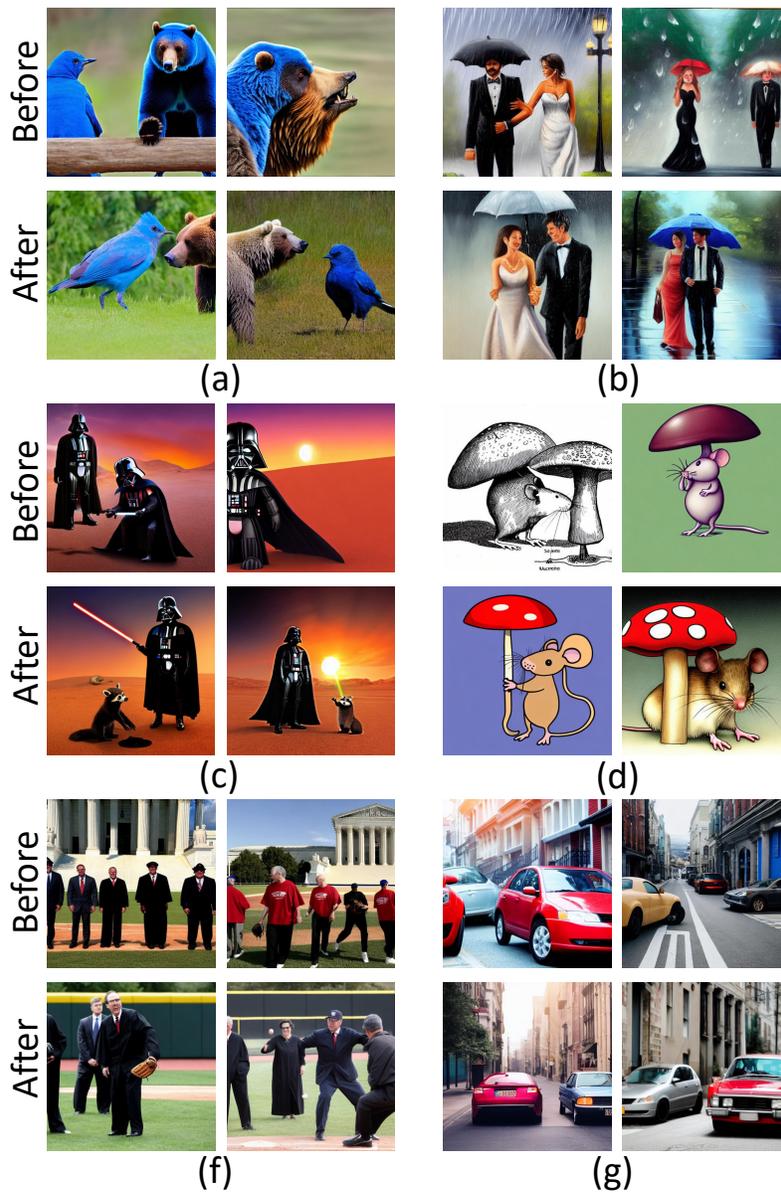


Figure 2: Sample images generated from prompts: (a) a blue bird and a brown bear, (b) An oil painting of a couple in formal evening wear going home get caught in a heavy downpour with no umbrellas, (c) Darth Vader playing with raccoon in Mars during sunset, (d) Illustration of a mouse using a mushroom as an umbrella, (e) Supreme Court Justices play a baseball game with the FBI. The FBI is at bat, the justices are on the field, (f) Two cars on the street

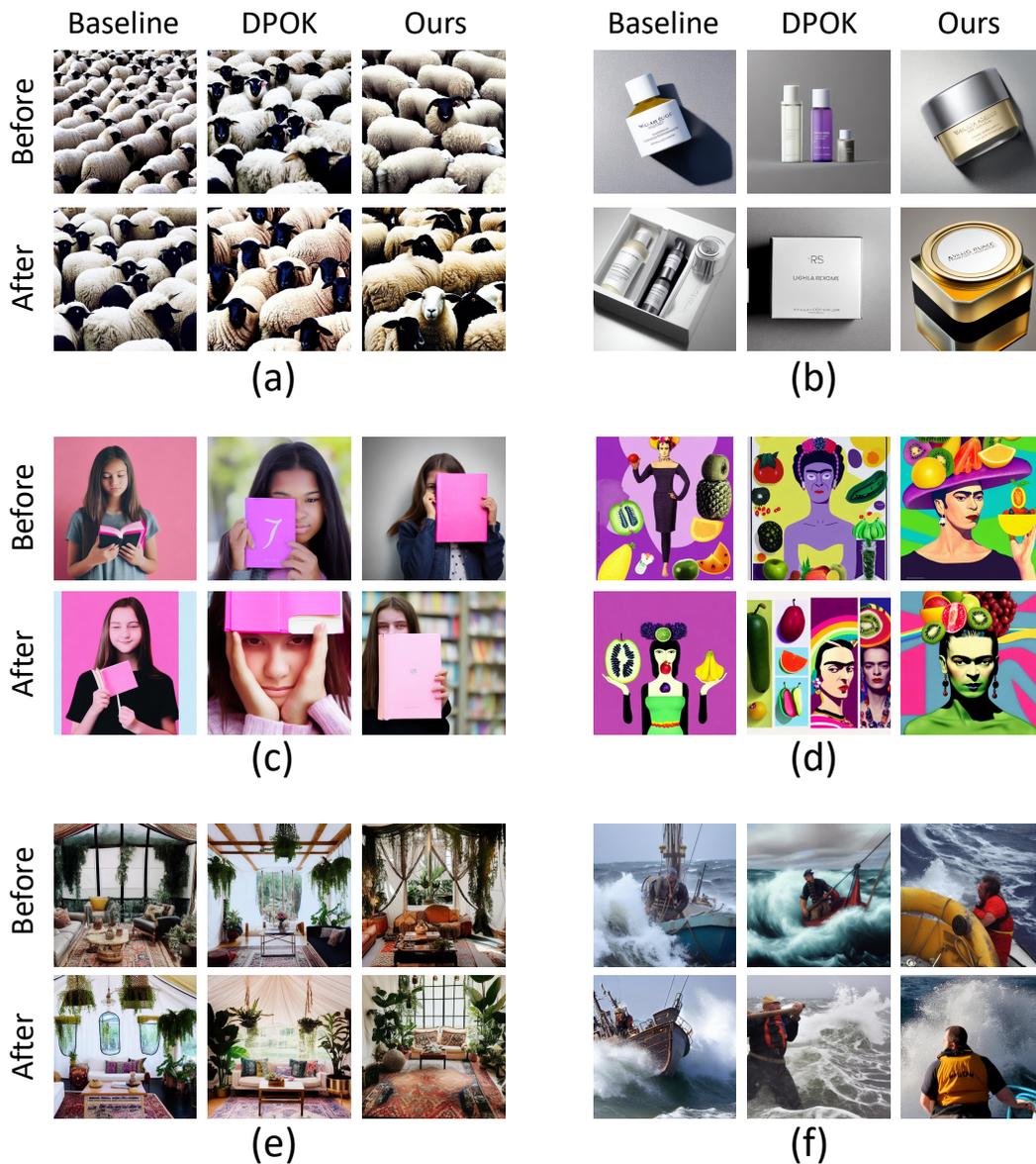


Figure 3: Sample images generated from prompts: (a) A black sheep among a flock of white sheep, raising head as a leader - Concept of standing out from the crowd, of being different and unique with its own identity and special skills among the others, (b) A skin care product packaging, light luxury style, metal lid, unique bottle appearance, made of acrylic, on the front, the brand's logo using aluminum foil hot stamping technology, strong sense of design, award-winning, mainly gold and white Tones, on solid gray background, reflections, shadows, global illumination, professional composition, by William Russell Flint, photorealistic, (c) a teenage girl holding a closed book in front of her face, with the left side of her face visible. The book is on her right side, and its cover is pink. She is slightly smiling, and the cover of the book partially obscures her face, (d) A two dimensional surrealistic-realism-minimal-collage best composition and a perfect scene composition and pop-retro-film-poster, Frida Kalo drooling a rainbow liquid and Carmen miranda instead her conventional fruit hat with a gummy fruits hat, also in the composition is the Hand of Adam's Family, a purple Kremilin, a rolling eye, smashed rotten green tomato, (e) award winning photography, ornate tent mansion modern boho living room, cozy seating, indoor hanging plants, fairy lights, (f) bosun doggedly faces rough seas.

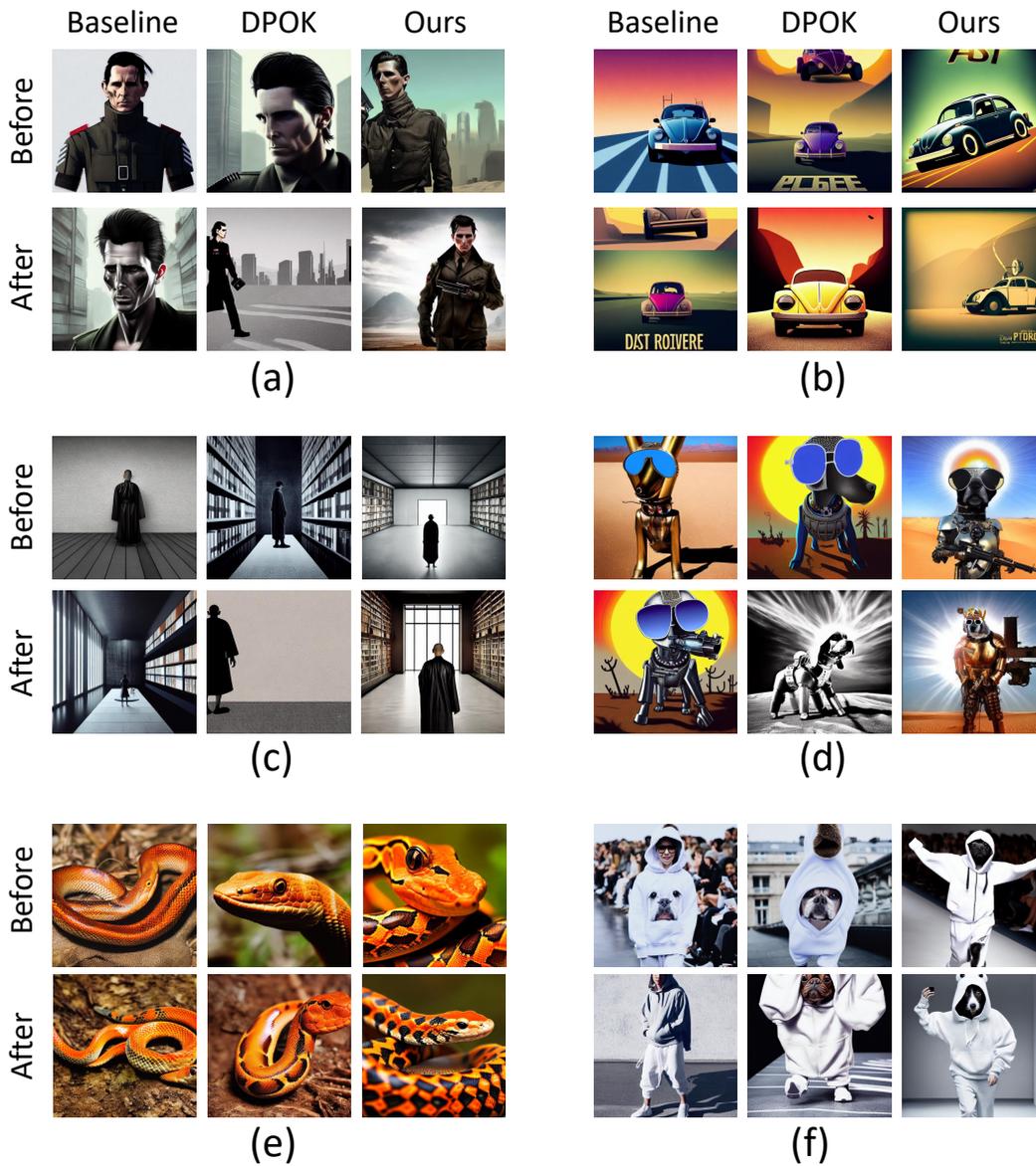


Figure 4: Sample images generated from prompts: (a) dystopian cyberpunk art style image of a handsome soldier with slick black hair stood in a desert. His look is inspired by Christian Bale in American Psycho. He carries a snub nose revolver. He is impeccably groomed wearing clothes that combine military and casual style, (b) Film poster, with a dark gradient towards the bottom, which shows an old VW beetle spinning and creating dust on a paved road. In an adventure and dramatic style. Vintage. Not too detailed. Pixar style, 3d render, lowpoly, isometric, (c) huge long room with geometric lines and forms with water spots on the floor texture details, bold small Monk in black leather robe standing back, matte dark concrete material, black wood, surreal library, highly detailed, intricate texture, minimalistic design, warm light, (d) Humanoid sun dog surrealism, cool dog with sunglasses wearing metallic armor with a large machine gun standing in the desert of an apocalyptic wasteland, (e) orange and dark orange snake, photography, realistic, (f) Vogue magazine cover, cute dog with hoodie and sneakers walking down the runway, in the style of light white and white, appropriation artist, wimmelbilder, hallyu, subtle expressions, surprisingly absurd, Paris Fashion Week, anthropomorphic Full body "VOGUE"

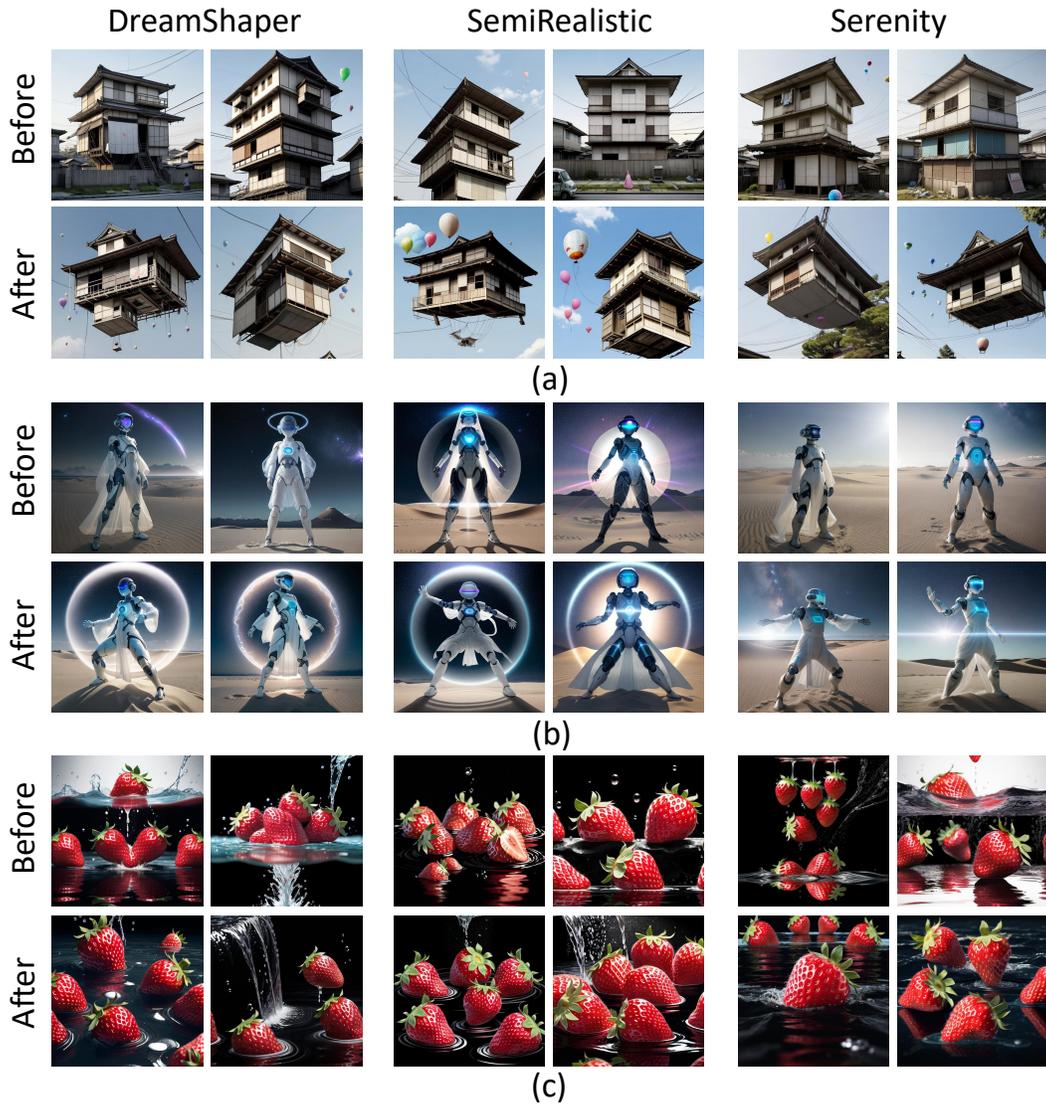


Figure 5: Sample images generated from prompts: (a) *Clear sky. A dilapidated Japanese house flying in the sky with many balloons. A flying house. Outside the house is an old lady doing her laundry. Reality,* (b) *japanese techwear, nebula, dusk, stars, bright, airy, muted colors, posing in powerful Tai-chi poses on a white sand dunescape, controlling the galaxy with arm movement, a holographic halo effect over the head of a pale robot-cyborg with translucent pale-blue seaglass faceshield, billowing translucent high-collar robes,* (c) *strawberry poster, strawberries falling into the water, water splashes, black solid background, illustration, small details 8k*