Supplementary Material of Powerful and Flexible: Personalized Text-to-Image Generation via Reinforcement Learning

Fanyue Wei[1], Wei Zeng[2], Zhenyang Li[2], Dawei Yin[2], Lixin Duan[1], and Wen Li[1†]

[1]University of Electronic Science and Technology of China
{wfanyue, wzeng316, zhenyounglee, lxduan, liwenbnu}@gmail.com
[2]Baidu Inc
{yindawei}@acm.org

In this supplementary material, we elaborate on the implementation details, the structure of Q_ϕ function, pseudo code of the proposed algorithm and more visualization of the generated images. We also exhibit the flexible ability of our proposed framework for unconditional image generation.

1 Implementation Details

We adopt DreamBooth [2] as our baseline method. Since the official code of DreamBooth [2] is not publicly available, we use the implementation in the popular "diffusers" library[⋆] and the pretrained Stable Diffusion V1.4 for all compared methods for fair comparison. We set the discount rate γ as 0.9986 and the weight of DINO reward λ as 0.1. For Custom Diffusion [1], we reproduce their methods using their official code and the default hyperparameters on the DreamBooth benchmark. The resolution of generated images is 512 × 512. All experiments are performed on a 32G V100 card.

2 Structure of Q-function

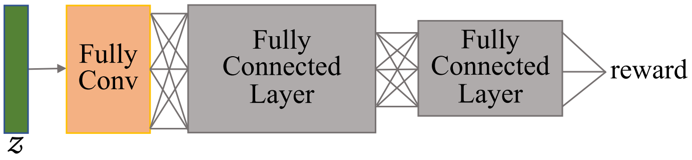

Fig. S1: We illustrate the Q_ϕ network structure in the figure.

In our experiments, we implement the Q-function by a simple neural network to map the latents x_t in the diffusion process in SD into the reward value. Thus,

[†] Corresponding author.

[⋆] https://github.com/huggingface/diffusers we opt for a simple convolutional layer to map the 4 × 64 × 64 latents into 64 × 64. Then the convolutional layer is followed by 2 layer fully-connected network. We use 4096 units in the first and 64 in the second hidden layer in order to predict an unnormalized scalar score. Figure S1 illustrates the structure of our implementation.

For the hyper-parameters in the learning process, we simply set the learning rate as 0.01, optimized by the sophisticated gradient descent algorithm.

3 Pseudo Code of Algorithm with Looking Forward

Besides, the pseudo code of the algorithm, we add the pseudo code of learning to "look forward" in Algorithm 1.

Algorithm 1 DPG Framework for T2I Penalization

```
Input: Policy function $p_\theta$ and Q-function $Q_\phi$, the Time step $T$, the text condition $y$
repeat
    Chose an image from a set of reference images
    Randomly sample $t \sim \{0 \cdots T\}$
    Process $t$ step diffusion, get the latent state $x_t$
    Calculate the reward $r$ based on the definition
    Get the accumulative reward as the target
        $\sum_{i=0}^{t} r(x_i, \tau(y), i)$
    Optimize the Q-function parameters $\phi$ by accumulated steps of gradient descent using
        $\nabla_\phi ||Q_\phi(x_t, \epsilon_\theta(x_t, t, \tau(y))) - \sum_{i=0}^{t} r(x_i, \tau(y), i)||^2$
    Update Q-function with discount rate $\gamma$
        $Q_\phi(x_t, \epsilon_\theta(x_t, t, \tau(y))) = r(x_i, \tau(y), i) + \gamma Q_\phi(x_{t-1}, \epsilon_\theta(x_{t-1}, t-1, \tau(y)))$
    Update the diffusion model $\theta$ by gradient ascent using
        $\nabla_\theta Q_\phi(x_t, \epsilon_\theta(x_t, t, \tau(y)))$
```

4 More Visualization Cases

We present more visualization results for comparisons. As shown in Figure S2, our method is capable to capture the visual details of the personalized subjects with various prompts.

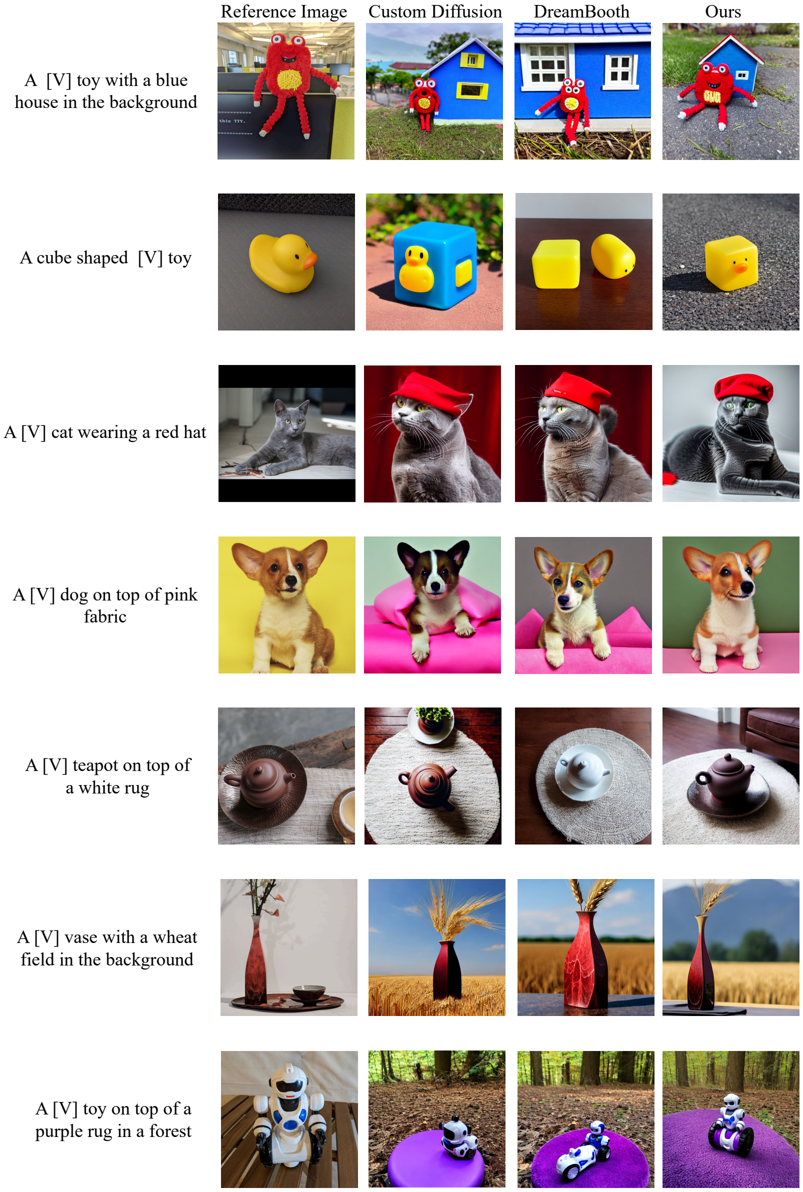
Reference Image
Custom Diffusion
DreamBooth
Ours
A [V] toy with a blue house in the background
A cube shaped [V] toy
A [V] cat wearing a red hat
A [V] dog on top of pink fabric
A [V] teapot on top of a white rug
A [V] vase with a wheat field in the background
A [V] toy on top of a purple rug in a forest

Fig. S2: We present more generation results of the reference image, Custom Diffusion [1], DreamBooth [2] and Ours. As shown in the figure, the generated image given the challenging prompt, Ours preserves the high fidelity of the personalized attributes including color, expressions, texture and *etc*.

5 Training with Other Flexible Reward

Our framework exhibits promising potential with pre-training and can incorporate other metrics such as PSNR, despite our aim in this task to improve the visual fidelity of T2I personalization rather than pre-training or image editing. Due to time constraints, we conduct experiments on unconditional image generation pre-training with PSNR. We perform our experiments on the "church" class on ImageNet. We train the DDPM baseline and that equipped with our DPG framework by PSNR on images at a resolution of 64×64 for 1000 epochs, and evaluate them using a validation set of 50 images from ImageNet on both PSNR and FID. The results in Table S1 demonstrate that our approach has the powerful and flexible capability to improve the generation results.

Table S1: Evaluation on Unconditional Image Generation with PSNR.

Method	PSNR ↑	FID ↓
DDPM baseline	9.083	219.398
Ours with PSNR	9.308	215.965

References

1. Kumari, N., Zhang, B., Zhang, R., Shechtman, E., Zhu, J.Y.: Multi-concept customization of text-to-image diffusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1931–1941 (2023)
2. Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., Aberman, K.: Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22500–22510 (2023)