LayerDiffusion: Layered Controlled Image Editing with Diffusion Models

Pengzhi Li, Qinxuan Huang, Yikang Ding and Zhiheng Li
Tsinghua University, Tsinghua Shenzhen International Graduate School

ICML Workshop Artificial Intelligence & Human-Computer Interaction

Introduction

In this paper, we propose LayerDiffusion, a semantic-based layered controlled image editing method. Our method enables non-rigid editing and attribute modification of specific subjects while preserving their unique characteristics and seamlessly integrating them into new backgrounds. We leverage a large-scale text-to-image model and employ a layered controlled optimization strategy combined with layered diffusion training. During the diffusion process, an iterative guidance strategy is used to generate a final image that aligns with the textual description.

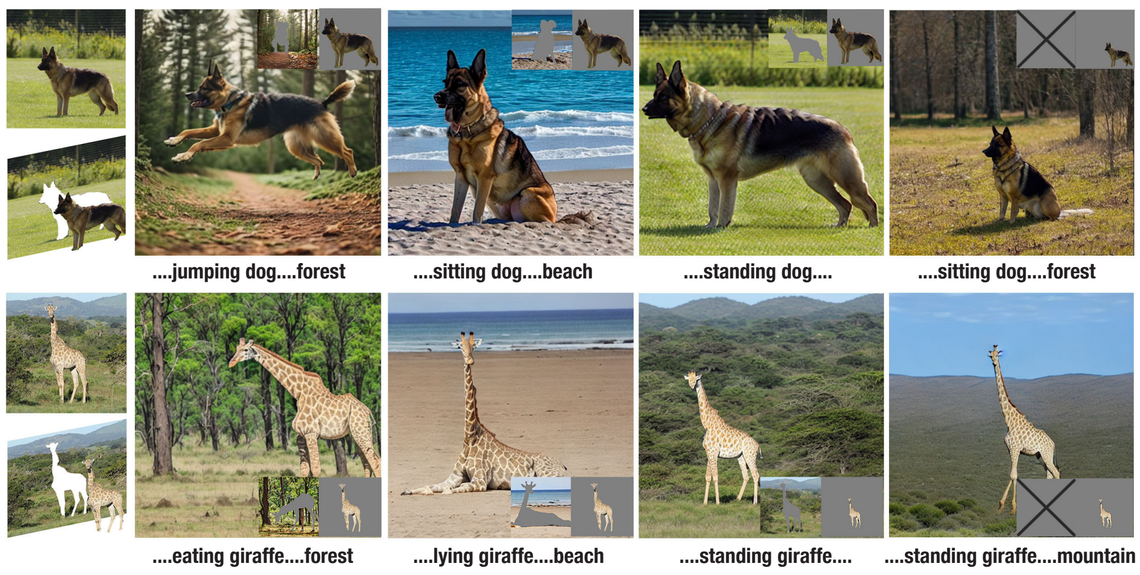
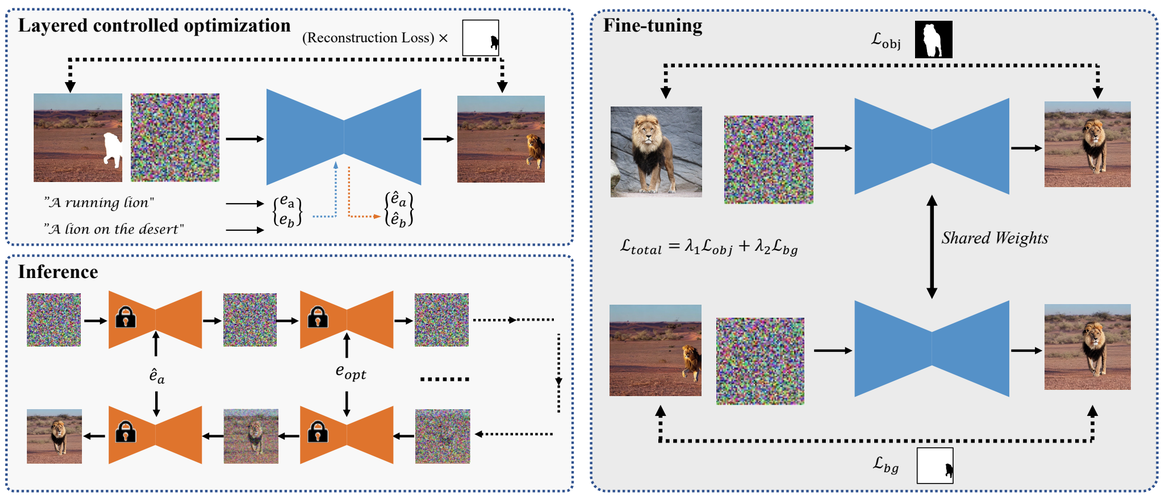


Figure 1. Our method achieves layered image editing through text descriptions, enabling simultaneous modifications of backgrounds and specific subjects, such as background replacement, object resizing, and complex non-rigid changes.



$$\mathcal{L}_{obj} = \mathbb{E}_{\boldsymbol{x}_t, \boldsymbol{\epsilon} \sim \mathcal{N}(0,\boldsymbol{I})} \left[\| M_t * (\boldsymbol{\epsilon} - f_\theta(\boldsymbol{x}_t, t, e_{opt})) \|^2 \right]$$

$$\mathcal{L}_{bg} = \mathbb{E}_{\boldsymbol{x}_t, \boldsymbol{\epsilon} \sim \mathcal{N}(0,\boldsymbol{I})} \left[\left\| (\hat{M}_r) * (\boldsymbol{\epsilon} - f_\theta(\boldsymbol{x}_t, t, e_{opt})) \right\|^2 \right]$$

Figure 2. Our method utilizes a layered controlled optimization strategy to refine text embeddings and a layered diffusion strategy to fine-tune the diffusion model. During inference, an iterative guidance strategy is employed to directly generate images aligning with the multiple editing actions described in the input text.

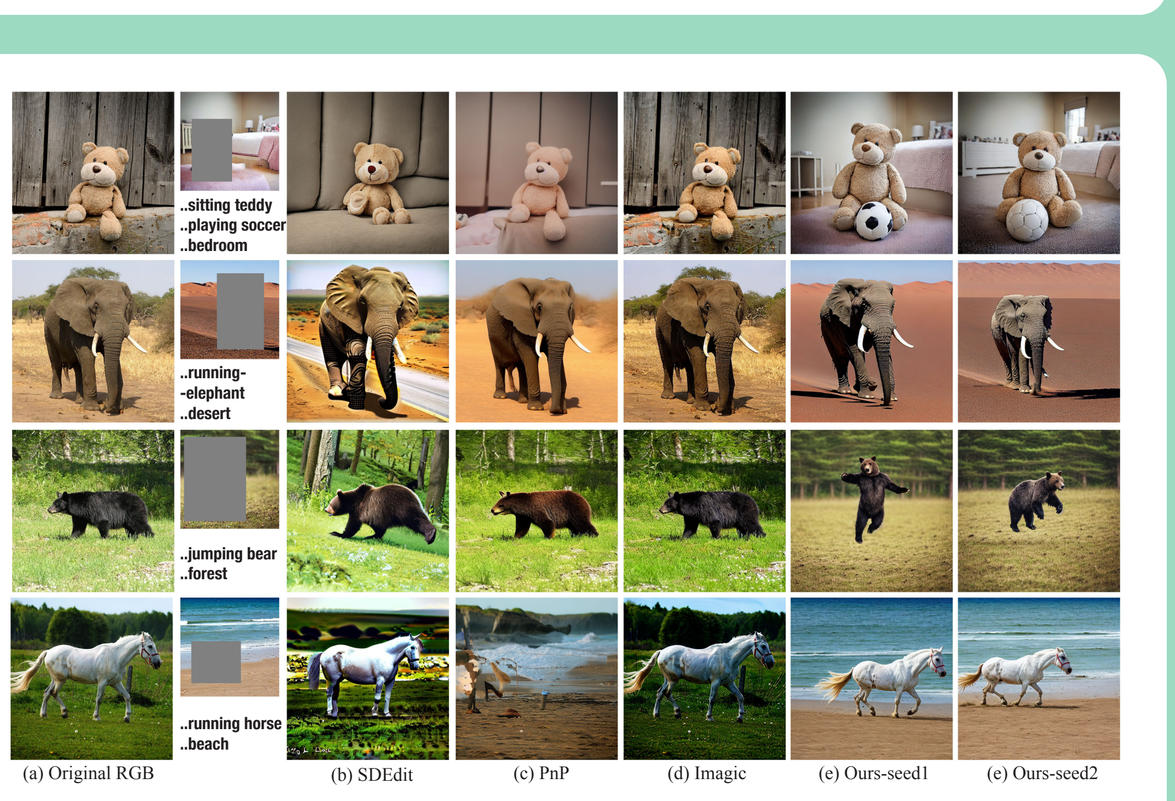


Figure 3. We present several edited images and compare them with similar image editing algorithms

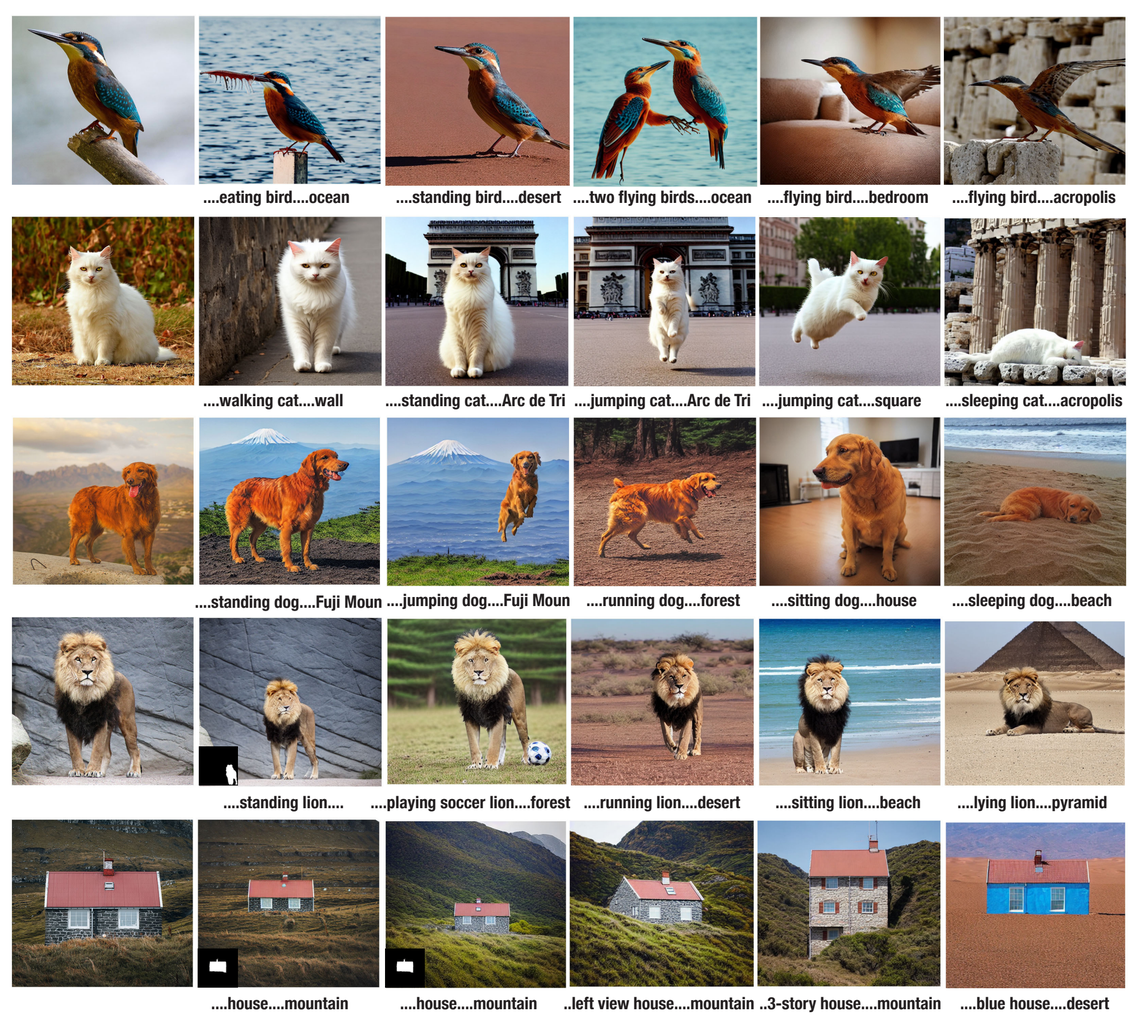


Figure 4. Given a complex text description, the original image (left) is capable of performing multiple editing actions and maintaining similar characteristics of a specific subject. Note that the mask in the bottom left corner is used to change the size of the selected object.

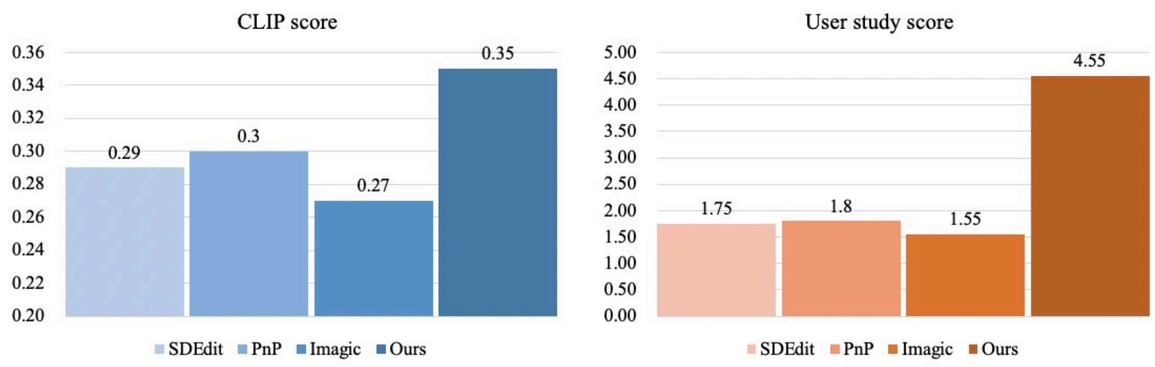


Figure 5. We compare several image editing methods using the CLIP and subjective user perception scores. Our method achieves a relatively higher score.

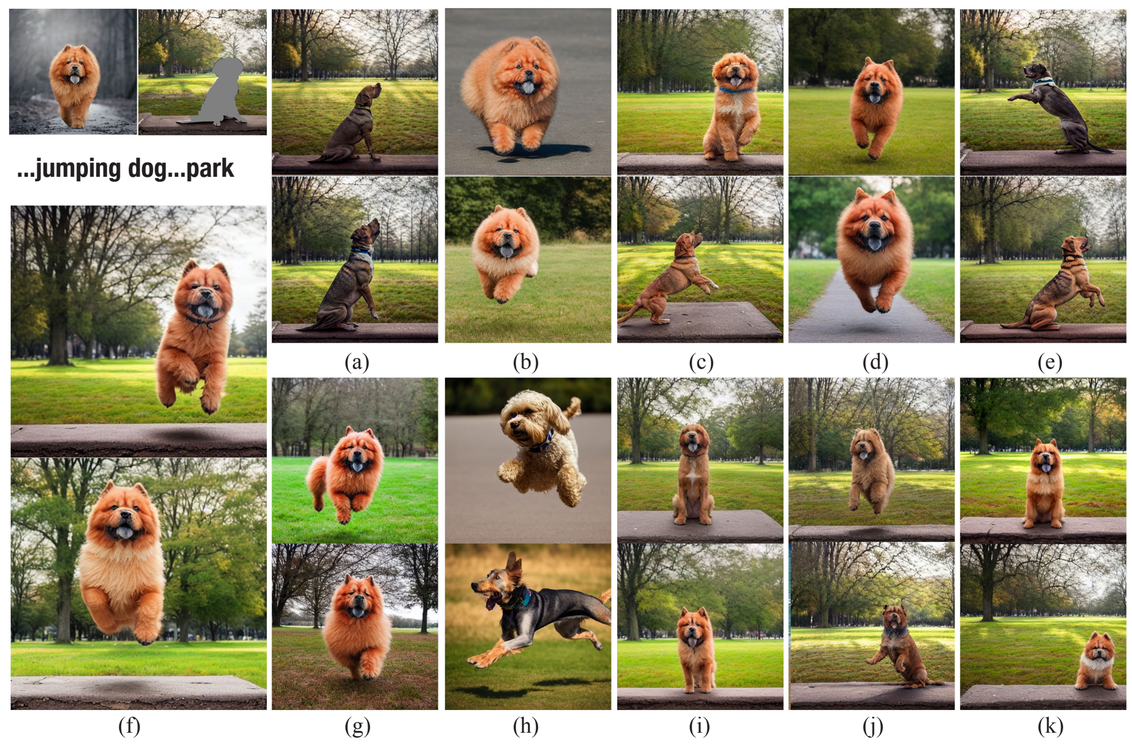


Figure 6. We present the edited images with different settings. For each setting, we show two generated images using different random seeds. (f) illustrates the final edited results.