MagicEraser: Erasing Any Objects via Semantics-Aware Control

Supplementary Material

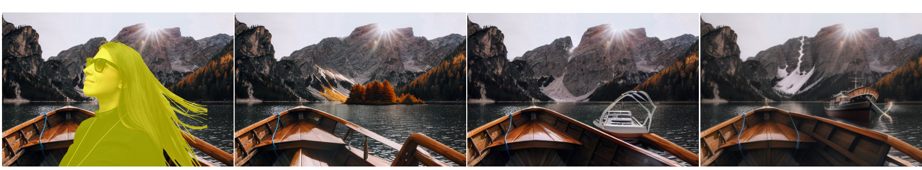
Anonymous ECCV 2024 Submission

Paper ID #4070

In Section 1, we show examples by Stable Diffusion (SD) Inpainting when it is given a common short prompt. In Section 2, more visual comparisons with SD Inpainting are given. Finally, the effectiveness of R_* is explained with two examples in Section 3.

1 SD Inpainting with a Short Prompt

For the example of Fig. 1, SD Inpainting with a long and high-quality prompt, such as "*The boat is on a serene lake surrounded by dramatic mountains with rugged textures. The sun is shining directly above the mountain peaks, creating a flare effect in the camera lens. There's a reflection of the sun on the water, suggesting it's a clear day. The trees on the mountainside are tinged with autumn colors, which adds warmth to the scene*", often obtains better performance than with a short prompt. For example, SD Inpainting with a common short prompt "*A boat on the lake*" tends to generate another boat. Two sampled results with the short prompt are shown in Fig. S1.



Input+Mask | Result with the long prompt | Result 1 with the short prompt | Result 2 with the short prompt

Fig. S1: Erasure results by SD Inpainting with a long prompt and a short prompt.

2 More Comparisons with SD Inpainting

We randomly select more images from the dataset RealHM and then compare with SD Inpainting. LLaVa is applied to generate textual prompts for SD Inpainting. However, the prompts directly generated by LLaVa tends to describe the objects of the given image, which often leads SD Inpainting to generating new objects similar to those in the image. Therefore, we firstly utilize a pre-trained panoptic segmentation network Mask2Former to label the entire image and extract the background tags (e.g, "house" and "snow" in the first example of Fig. S2). Then, we add a prompt (e.g., "Describe the house and snow of the image") to guide LLaVA and obtain the textual description of background. Note that this prompt is obtained automatically according to the tags.

The prompts for MagicEraser are independent of LLaVa, which consist of some panoptic segmentation tags (e.g., "*house*" and "*snow*" in the first example of Fig. S2) and the learned prompt R_* (e.g., "A photo of R_* house" and "A photo of R_* snow"). As shown in Fig. S2, MagicEraser without using LLaVA and extra manual prompt inputs is capable of seamlessly erasing the objects.

3 Effectiveness of R_*

As mentioned in Section 4.3 of the main paper, R_* can be considered as a *universal* "background completion". Two visual results of MagicEraser with and without R_* are shown here in Fig. S3 to demonstrate its effectiveness.

Fig. S2: Visual comparison between SD Inpainting and MagicEraser on images randomly selected from RealHM.

Fig. S3: Erasure results by MagicEraser with and without R_* in the prompts.