




RegionDrag: Fast Region-Based Image Editing with Diffusion Models –*Supplementary Material*–

Jingyi Lu¹, Xinghui Li², and Kai Han^{1†}

¹The University of Hong Kong ²University of Oxford
lujingyi@connect.hku.hk, xinghui@robots.ox.ac.uk, kaihanx@hku.hk

1 Effectiveness of Region-Based Inputs

In the main text, we quantitatively demonstrate the effectiveness of using more point pairs obtained from the region pairs. To complement our quantitative findings, we present qualitative examples in Fig. 1. The results suggest that increasing the percentage of transformed points used enhances the quality and stability of the editing outcomes.



Fig. 1: Improved editing quality with a higher percentage of transformed points.

2 Discussion of Noise Weight

In the initial denoising step, the latent representations of the handle regions are blended with random noise weighted by α , where α ranges from 0 to 1 (Eq. (5) in main text). As illustrated in Fig. 2, a higher α value retains less of the original content. Nonetheless, the object in the handle region is not guaranteed to be removed even when $\alpha = 1$, as the outcome is influenced by other factors including the denoising process, the attention swapping mechanism, and the text prompt.

[†] Corresponding author.

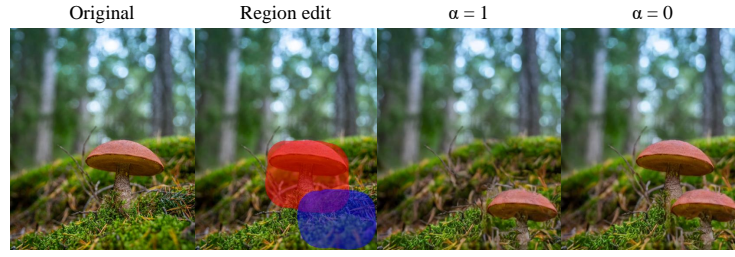


Fig. 2: Results of applying different noise weight α .

3 More Qualitative Results

In Fig. 3, we showcase additional qualitative outcomes of RegionDrag as it processes objects from various domains.

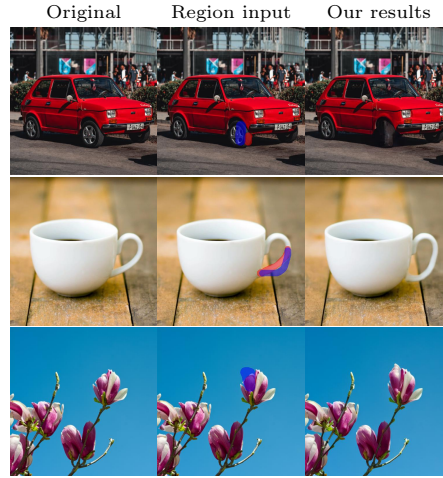


Fig. 3: More qualitative results.

4 More Qualitative Comparisons

Figure 4 presents additional examples comparing point-based editing outcomes with our region-based editing results. The region-based edits maintain object identity, resulting in higher-quality modifications that closely adhere to the desired output.

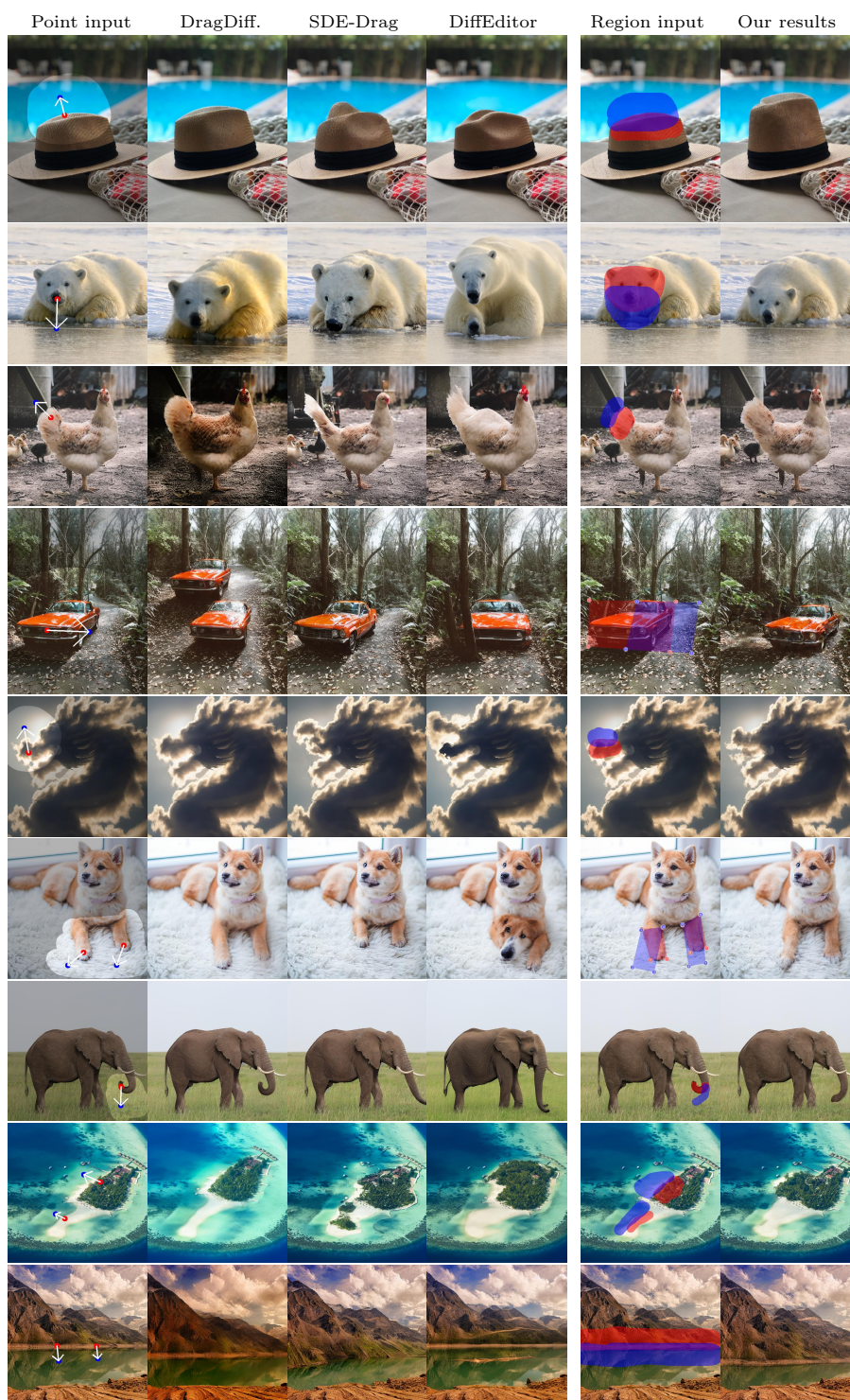


Fig. 4: More qualitative comparisons with baseline methods.