Drag Your Noise: Interactive Point-based Editing via Diffusion Semantic Propagation
- Supplementary Material -

Haofeng Liu[1,2*] Chenshu Xu[1*] Yifei Yang[1] Lihua Zeng[2] Shengfeng He[1†]

[1]Singapore Management University [2]South China Normal University

(a) Input (b) DragNoise (c) DragDiffusion (d) Input (e) DragNoise (f) DragDiffusion

Figure 1. Visualization of optimization trajectories. For optimal viewing of animated graphics, it is recommended to use Adobe Acrobat. Additionally, video versions are included in the accompanying video.

1. Optimization Trajectories

To demystify the trajectories in our diffusion semantic optimization, we present animated graphics illustrating intermediate results along with the final results. As shown in Fig. 1, we compare the optimization trajectories of our method with DragDiffusion [1]. Specifically, we generate images with intermediate-optimized bottleneck features for our DragNoise and intermediate-optimized latent maps for DragDiffusion, respectively.

These visualization results demonstrate that the optimization trajectories of our approach exhibit superior continuity and enhanced stability. The observation aligns with the analysis in Sec.4.2 in the main paper that DragNoise more swiftly determines the optimization direction. This can be attributed to that the bottleneck feature serves as a more optimal semantic representation. Moreover, our method is better aware of the semantics. For instance, the sliced cake in the 1st row of Fig. 1b can be optimized towards closure with DragNoise and the robot's leg can be reattached in the 2nd row of Fig. 1e. In contrast, the optimization process of DragDiffusion is more erratic, resulting in more unstable dragging editing effects.

2. Choice of the Feature for Supervision

We conduct an ablation analysis on the feature for supervision, as illustrated in Fig. 2. The feature size progressively increases from Decoder Block 1 to 4. Features from Decoder Block 1, lacking detailed pixel-level information, fail to achieve the desired outcome. While Decoder Block 2 offers some supervision, it compromises identity preservation in the edited image. Conversely, Decoder Block 4's output lacks distinctive information, impeding bottleneck feature updates. Hence, Decoder Block 3 represents the optimal balance between semantic richness and precise editing control.

*The first two authors contributed equally.
†Corresponding author (*shengfenghe@smu.edu.sg*).

(a) Input (b) Decoder Block 1 (c) Decoder Block 2 (d) Decoder Block 3 (e) Decoder Block 4

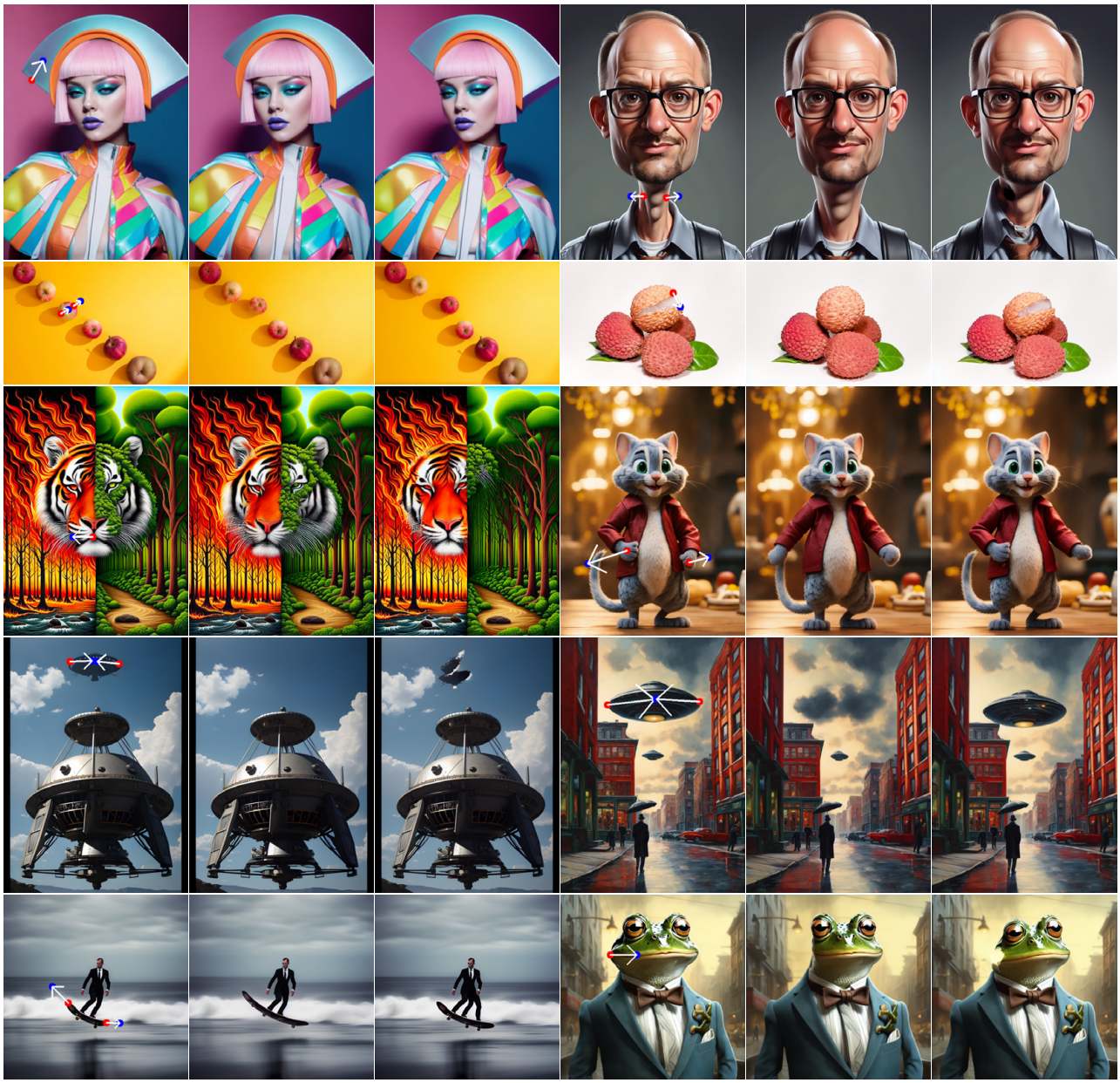
Figure 2. Results of ablation analysis on the feature for supervision.

3. More Results

We present additional results comparing DragNoise with DragDiffusion in Figs. 3 and 4. Our method demonstrates compelling performance across a diverse range of input images. Especially, through carefully positioned anchor points and objective points, DragNoise achieves diverse editing effects. For instance, our method successfully moves an apple while preserving its fidelity, as depicted in the 2nd row of Fig. 3b. Also, DragNoise effectively eliminates UFOs, as evidenced in the 4th row of Figs. 3b and 3e. DragNoise exhibits a heightened awareness of contextual information, yielding editing results that better align with user intent. Noteworthy instances include the preserved hairstyle in the 1st row of Fig. 3b and the retained shape of the tiger in the 3rd row of Fig. 3b. Additionally, we incorporate a demo to illustrate the interactive editing process in the attached video.

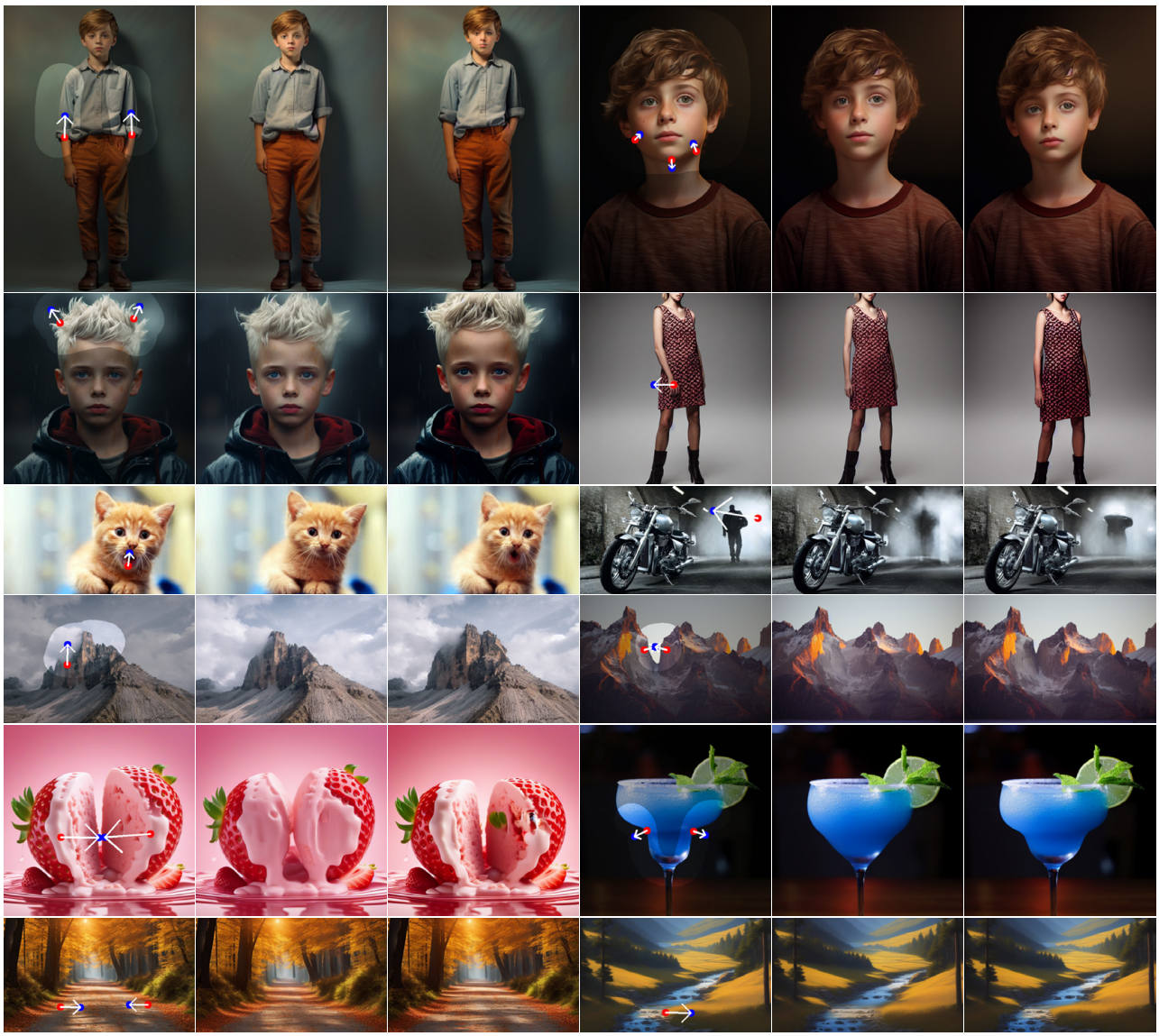
References

[1] Yujun Shi, Chuhui Xue, Jiachun Pan, Wenqing Zhang, Vincent YF Tan, and Song Bai. Dragdiffusion: Harnessing diffusion models for interactive point-based image editing. *arXiv preprint arXiv:2306.14435*, 2023. 1



(a) Input (b) DragNoise (c) DragDiffusion (d) Input (e) DragNoise (f) DragDiffusion

Figure 3. More editing results compared with DragDiffusion.



(a) Input (b) DragNoise (c) DragDiffusion (d) Input (e) DragNoise (f) DragDiffusion

Figure 4. (Continued) More editing results compared with DragDiffusion.