GDrag: Towards General-Purpose Interactive Editing with Anti-ambiguity Point Diffusion

Xiaojian Lin[1*], Hanhui Li[1,2*], Yuhao Cheng[3], Yiqiang Yan[3], Xiaodan Liang[1,2,4†]

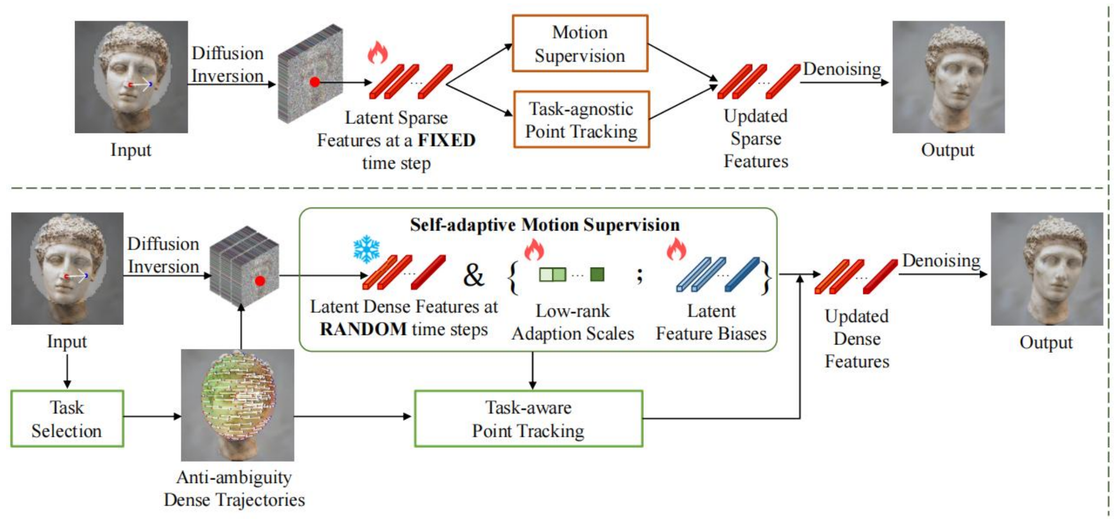
[1]Shenzhen Campus of Sun Yat-Sen University [2]Guangdong Key Laboratory of Big Data Analysis and Processing
[3]Lenovo Research [4]Peng Cheng Laboratory

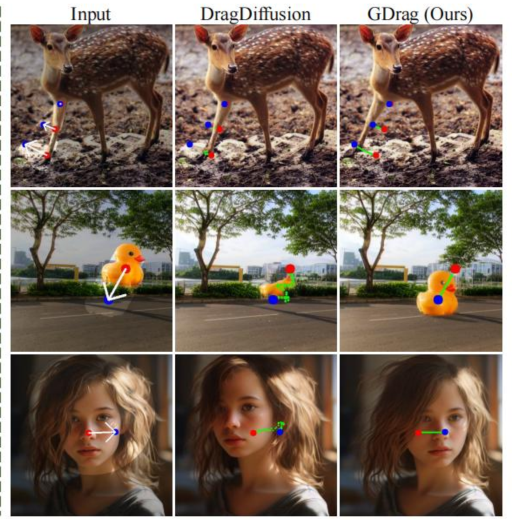
Introduction

- We propose GDrag, the first optimization-based framework that explicitly models editing tasks to **handle ambiguities**.
- Based on different task types, we propose the **ADT** method to construct a dense dragging point set and calculate their trajectories, which can offer more comprehensive and reasonable prior knowledge for image editing.
- We propose the **SMS** method that introduces task-aware, fine-grained optimization parameters to refine latent features. This allows us to address content ambiguities and improve the quality of edited images.
- Project Page: https://github.com/DaDaY-coder/GDrag

Motivation

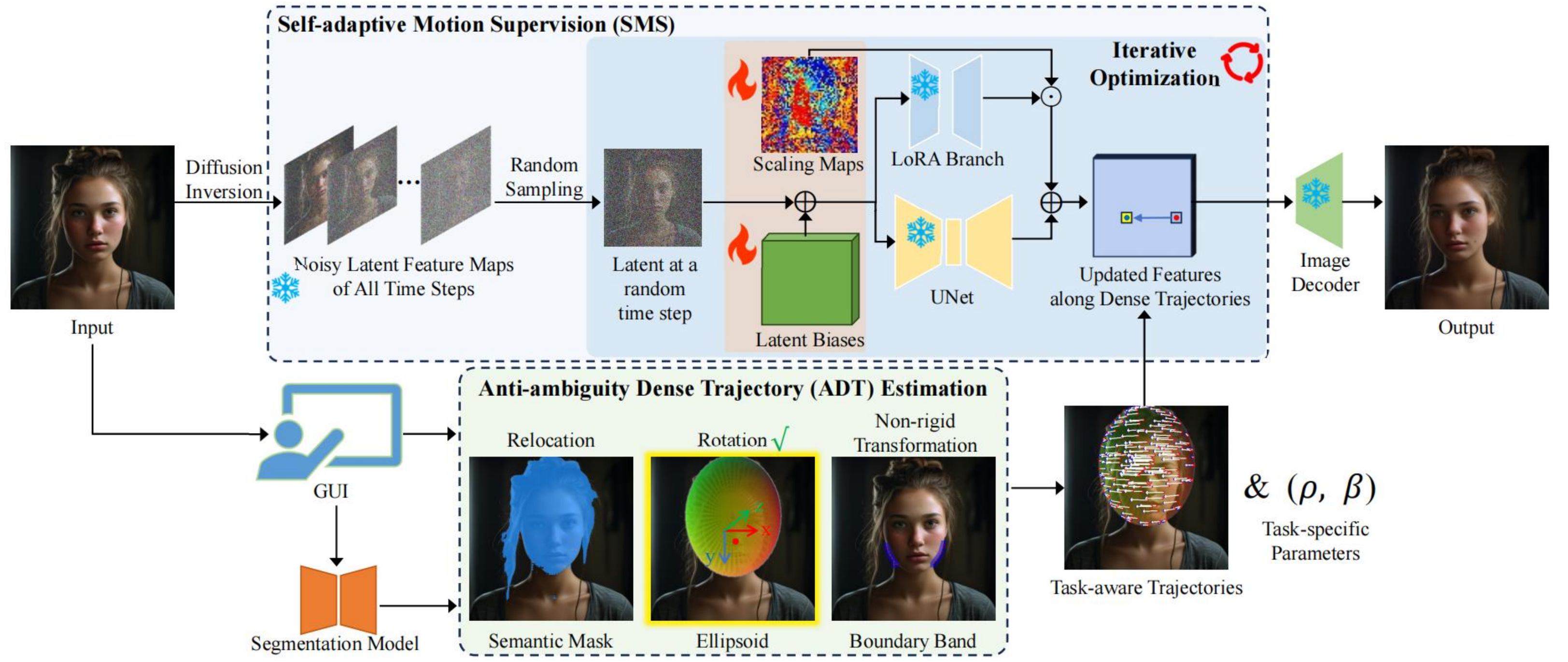
- Although current point-based diffusion methods already achieve relatively good editing effects, they still suffer from two types of ambiguities, i.e., **intention ambiguity** that mixes multiple possible editing tasks into a single trajectory, and **content ambiguity** that fails to identity and preserve the targets.
- To address the above ambiguity issues, we propose a novel task-aware, optimization-based framework for general-purpose interactive editing, named GDrag.
- Paradigm comparison between previous dragging methods (left-top) and the proposed GDrag method (left-bottom). GDrag estimates task-aware dense trajectories and adopts more fine-grained motion supervision. Hence, GDrag obtains smoother, more reasonable trajectories (labeled in green) and appealing results.





Methodology

- The proposed GDrag framework. The key idea of GDrag is to reduce intention and content ambiguities, which is accomplished by the proposed **anti-ambiguity dense trajectory estimation method** (ADT) and the **self-adaptive motion supervision method** (SMS). Given user-specific sparse handle points and the editing task, ADT selects a dense set of points that encode rich contextual information and estimates the corresponding task-aware trajectories. Utilizing these trajectories and a pair of task-specific parameters, the SMS method adjusts the positions and latent features of the dense points by optimizing latent feature biases and scaling maps, thereby achieving fine-grained editing results.



- Visual examples of our generated dense trajectories for diverse tasks, including relocation (top), rotation (middle), and non-rigid transformation (bottom).

Experiments

- We compare GDrag with state-of-the-art methods on the DragBench dataset to validate its effectiveness.

Method	MD↓	LPIPS↓	User Study ↑
DragDiffusion (Shi et al., 2024b)	33.91	0.0940	7.83%
DragonDiffusion (Mou et al., 2024)	31.63	0.1033	10.5%
FreeDrag (Ling et al., 2024)	27.41	0.0996	8.67%
DragNoise (Liu et al., 2024)	29.56	0.1017	12.67%
GDrag (Ours)	**26.49**	**0.0915**	**60.33%**

- Visual comparison between GDrag and state-of-the-art methods. GDrag achieves more precise manipulations and satisfying results across various editing tasks