RegionDrag: Fast Region-Based Image Editing with Diffusion Models

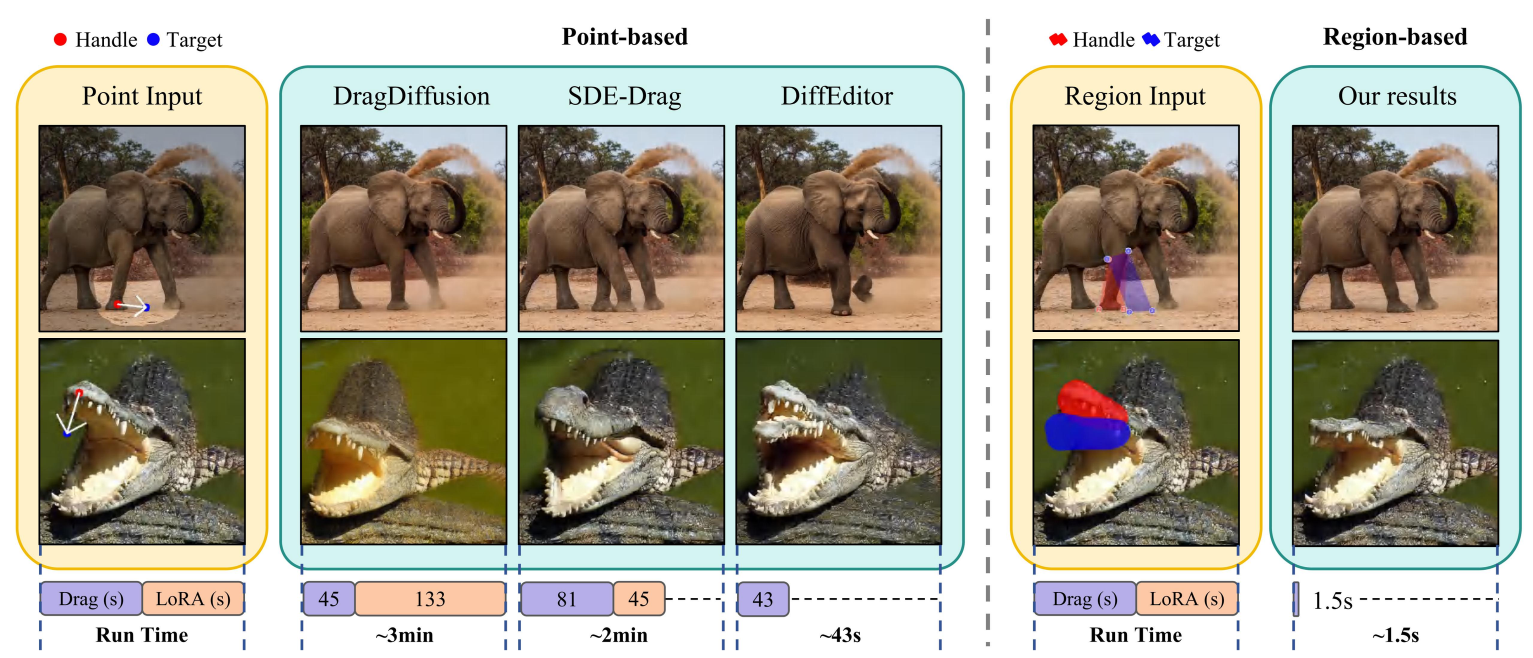
Jingyi Lu[1] Xinghui Li[2] Kai Han[1]✉

[1] Visual AI Lab, The University of Hong Kong [2] Active Vision Lab, University of Oxford

Webpage

Background & Contribution

Goal: Fast and precise image editing with region-based user inputs.



Key Contributions

(1) Region-based image editing method for better user intention alignment.

(2) Gradient-free, single-iteration editing pipeline for fast inference.

(3) Extended datasets with region-based instructions for benchmarking.

Motivation

Why Move Beyond Point-Based Image Editing?

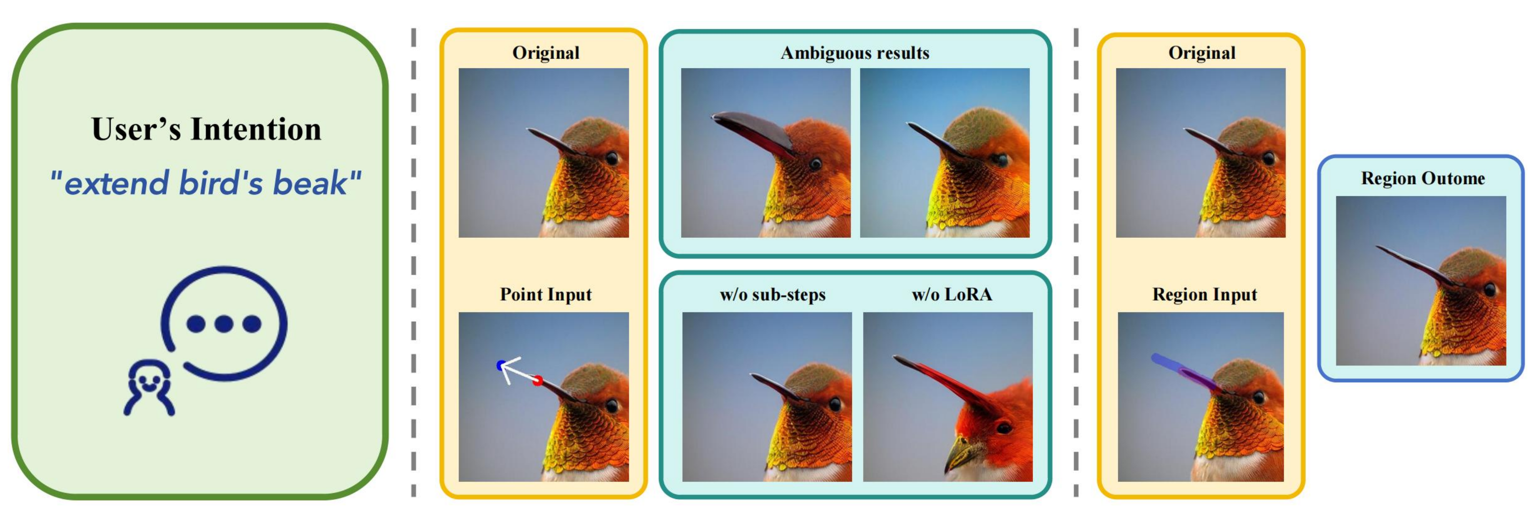
(1) Sparse point inputs often lead to ambiguous interpretations of user intentions.

(2) Point-based methods are slow due to iterative editing and expensive LoRA training.

- Models must infer global image changes from limited point movements.

(3) Region pairs provide richer context and denser mapping compared to sparse point pairs.

- Each region corresponds to a large number of points after dense mapping.



Related papers

[1] Nie, S., Guo, H.A., Lu, C., Zhou, Y., Zheng, C., Li, C.: The blessing of randomness: Sde beats ode in general diffusion-based image editing. arXiv preprint arXiv:2311.01410 (2023)

[2] Shi, Y., Xue, C., Pan, J., Zhang, W., Tan, V.Y., Bai, S.: Dragdiffusion: Har nessing diffusion models for interactive point-based image editing. arXiv preprint arXiv:2306.14435 (2023)

[3] Pan, X., Tewari, A., Leimkühler, T., Liu, L., Meka, A., Theobalt, C.: Drag your gan: Interactive point-based manipulation on the generative image manifold. In: ACM SIGGRAPH (2023)

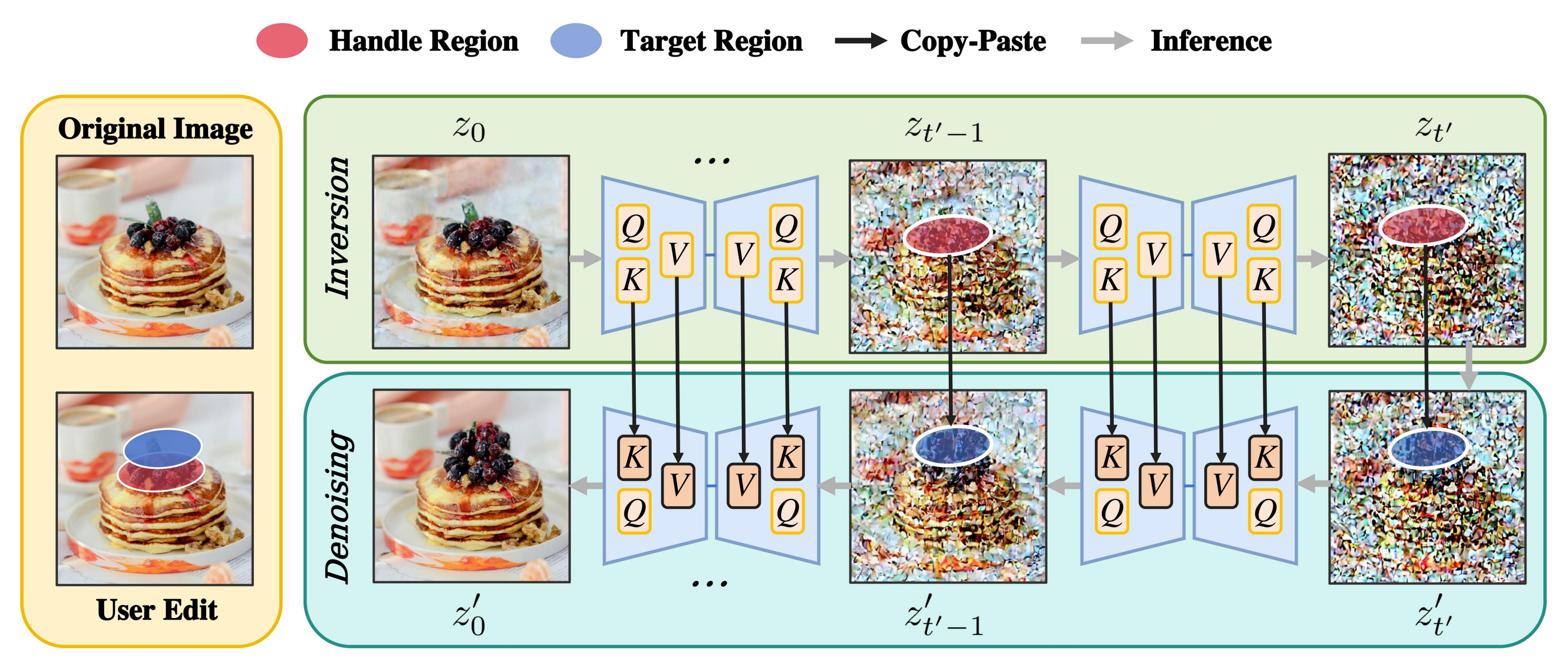
Method

Editing Pipeline

(1) **Region-based Input:** User selects handle and target regions for editing.

(2) **Multi-step Copy-Paste:** Repetitively copy latent representations from handle to target regions during a single inversion-denoising cycle.

(3) **Attention Swapping:** Maintain image consistency using mutual self-attention control.



Dense mapping between user-defined regions

(1) **Flexible Input Methods:** Support both polygon vertices and brush strokes for region selection.

(2) **Mapping Technique:**

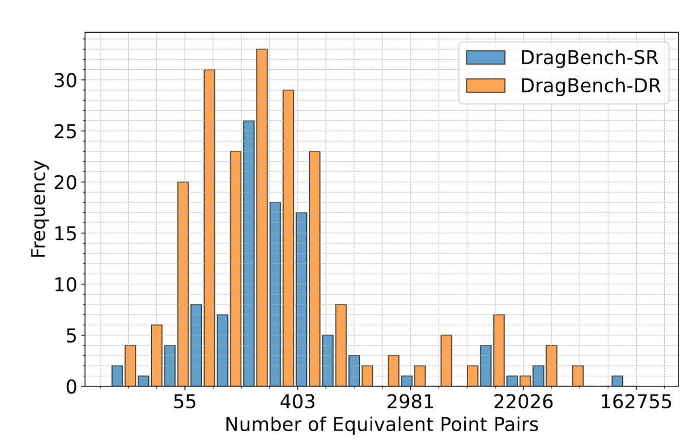
- **For polygons:** Apply affine or perspective transformations.
- **For brush strokes:** Apply horizontal and vertical scaling to map points between handle and target regions.

Datasets

New benchmarks for region-based editing evaluation

DragBench-S [1] and DragBench-D [2] are existing benchmarks for evaluating point-drag methods. We modify these benchmarks to use regions instead of points to reflect user intentions, creating DragBench-SR and DragBench-DR (where R stands for 'Region').

Frequency distribution of equivalent point pair counts



Quantitative Results

Mean Distance (×100) & LPIPS (×100)

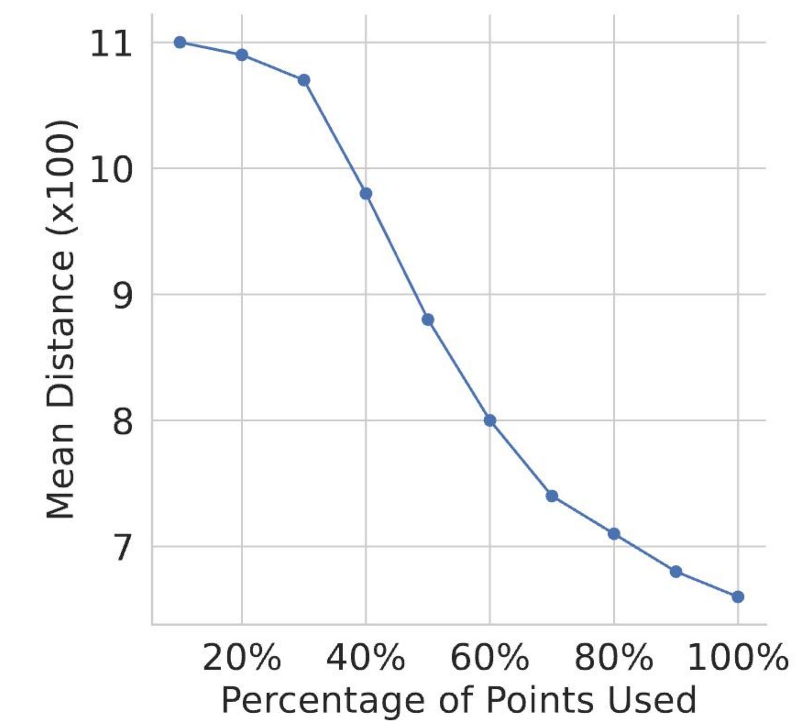
Method		DragBench-S(R)		DragBench-D(R)	
	Time (↓)	MD (↓)	LPIPS (↓)	MD (↓)	LPIPS (↓)
SDE-Drag	126.1	7.5	12.4	8.1	14.9
DragDiffusion	177.7	7.0	18.0	6.7	11.5
DiffEditor	43.1	23.6	17.6	22.1	10.9
Ours	**1.5**	**6.4**	**9.9**	**6.6**	**9.2**

Running Time Comparison (512 × 512 Resolution)

Analysis

Effectiveness of region inputs

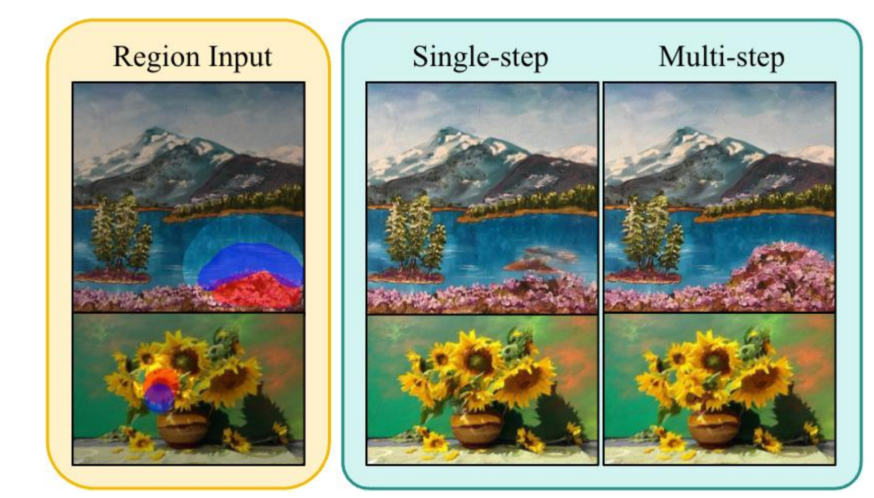
- Randomly sample different percentages of transformed points from each annotated region and conduct inference.



- Region-based inputs lead to superior results by providing stronger constraints than sparse points.

Effectiveness of multi-step copy-paste

- Copy-paste the image's latent representation across either multiple denoising timesteps or a single step.



- Initial single-step edits may be lost in subsequent denoising, leading to unpredictable results.
- Multi-step copy-paste provides guidance at smaller timesteps, preserving image fidelity.

Qualitative Results

RegionDrag achieves targeted modifications while maintaining image coherence.