# LightningDrag: Lightning Fast and Accurate Drag-based Image Editing Emerging from Videos

Yujun Shi[1*] Jun Hao Liew[2*] Hanshu Yan[2] Vincent Y. F. Tan[1] Jiashi Feng[2]
[1]National University of Singapore [2]Bytedance Inc. *Equal Contribution

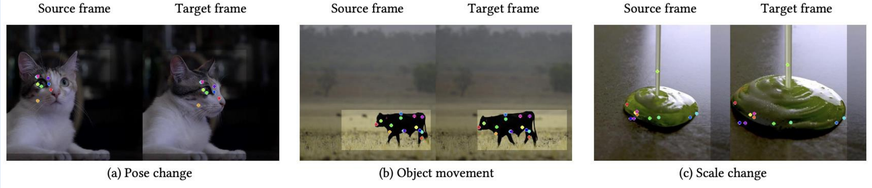
ICML International Conference On Machine Learning

## Summary

Building on the rich motion cues in video frames, we present the first pipeline capable of achieving **ultra-fast (<1s)** drag-based image editing with **high visual quality** on **arbitrary, real-world images**. Larger-scale models built on this pipeline have been integrated into JimengAI (即梦) as a feature called “Local Rotation” (局部旋转).

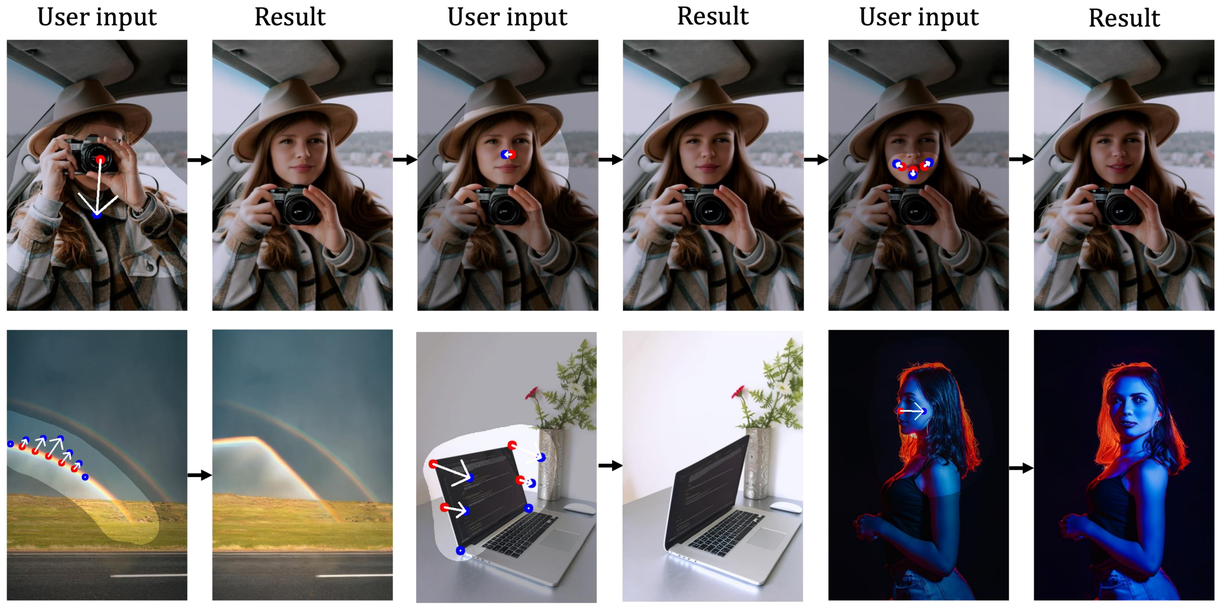
## Video Frames As Supervision Pairs

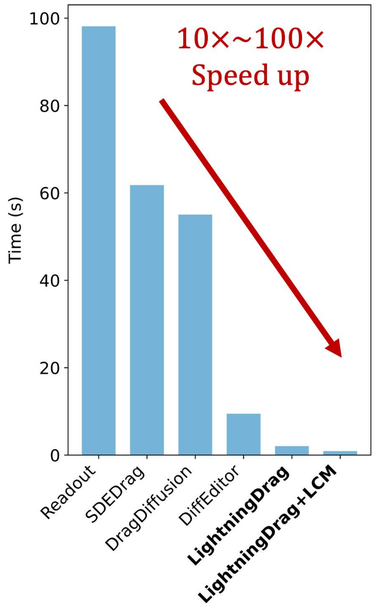
1. We conduct SFM-based and optical-flow-based filtering to obtain videos with static camera motion.
2. We sample paired frames in the video as source and target images.
3. we employ a co-tracker2 to label source and target points in paired frames to obtain diverse supervision pairs for training drag-based image editing



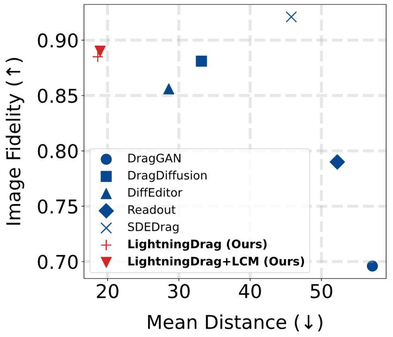

## Architecture

Inspired by prior works, we train a conditional diffusion model, which outputs drag-based editing results given the following conditions: 1) source images; 2) paired source and target points; 3) binary mask specifying editable regions.








The user provides source points (red), target points (blue), and a mask specifying the editable region (brighter area). Our approach (LightningDrag) significantly surpasses the previous methods in terms of both speed and quality.