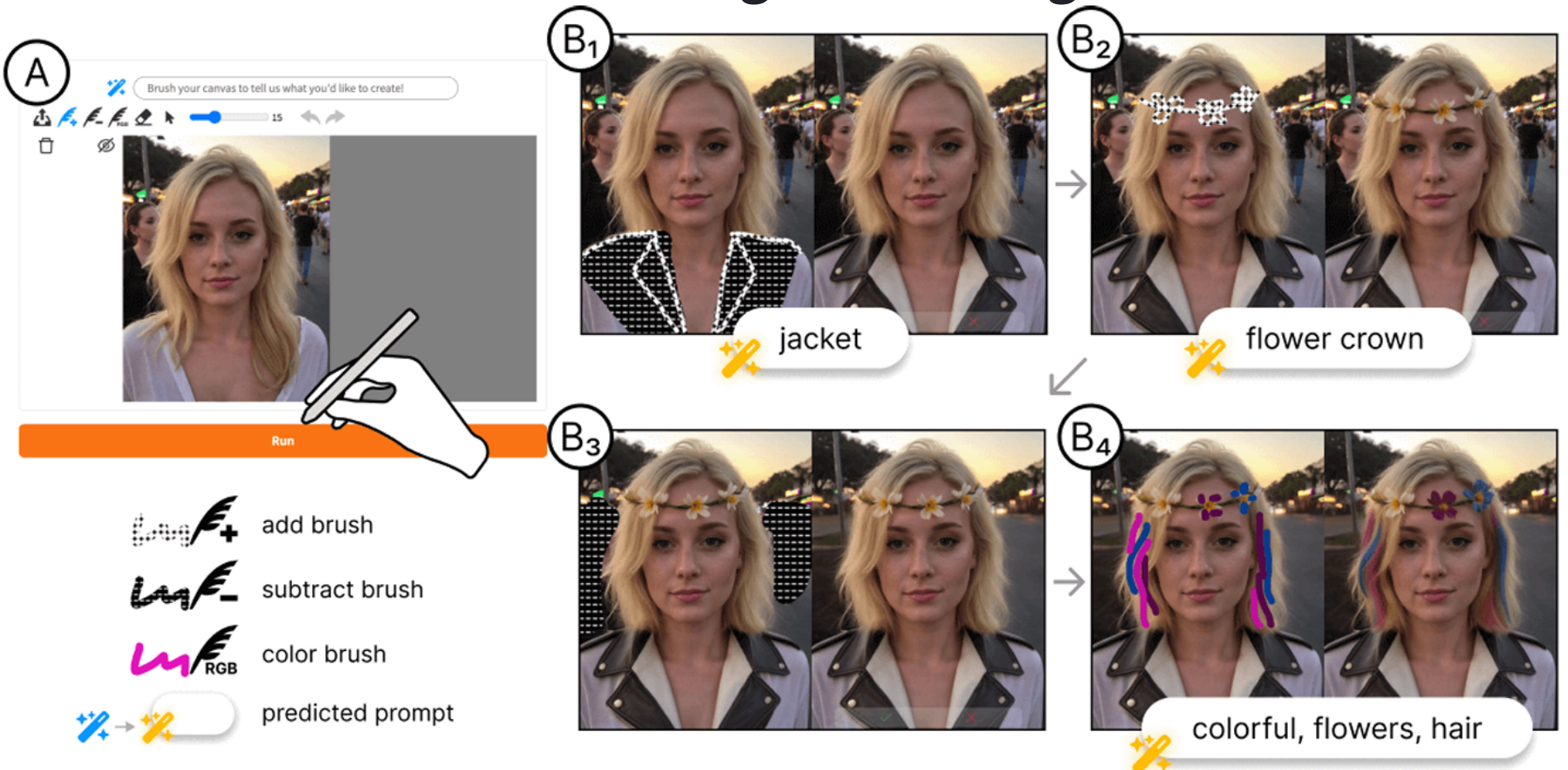
Image Editing with New Content Created by Interactive Generation Model

- Pointed based generated image editing
- General interaction with general image

Wenshuo ZHANG, Ph.D. Student, HKUST

- **Pointed based generated image**

- **General interaction with general image**



Tridictional way(IDE like Photoshop)

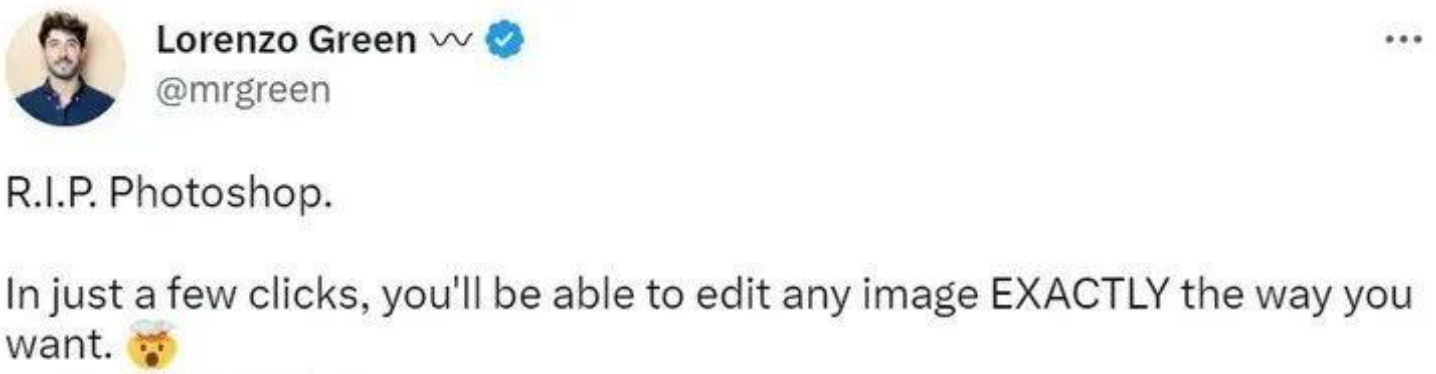
It needs a lot of time and actions

It is impossible for novice user as they lack knowledge

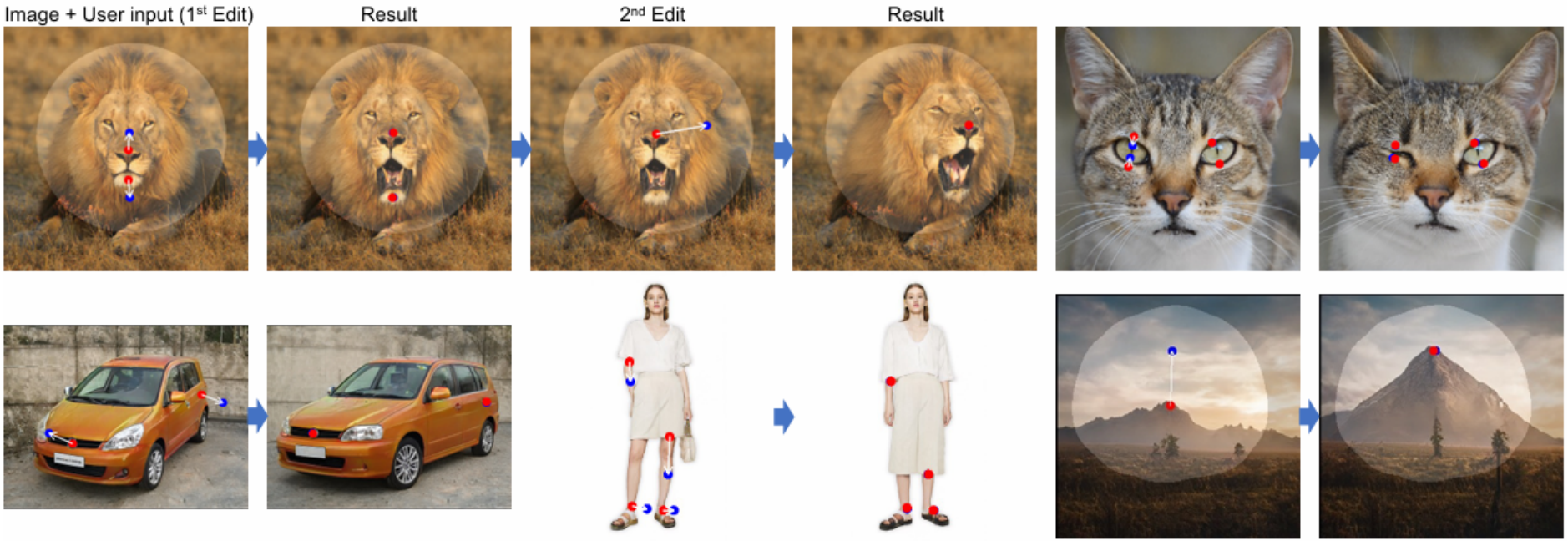
Can we use easier interaction way to edit image with new content created and do finely control on the generation process?

Generation Model is a way(Easy and Powerful).

- **How can generative model be finely controlled?**
- **How to get the editted object consisatant with raw object and user intent?**



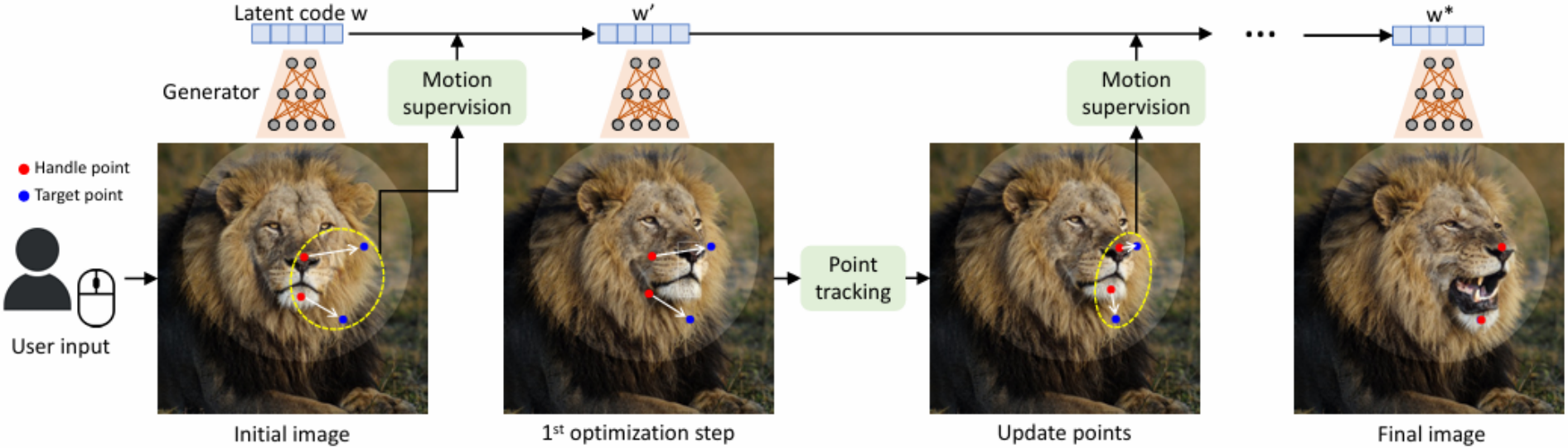
Drag Your GAN: Interactive Point-based Manipulation on the Generative Image Manifold - 18 May 2023



Generated images are not easy to be modified, it is always one time deal

Drag Gan gives a way to finelly control image editing process with points

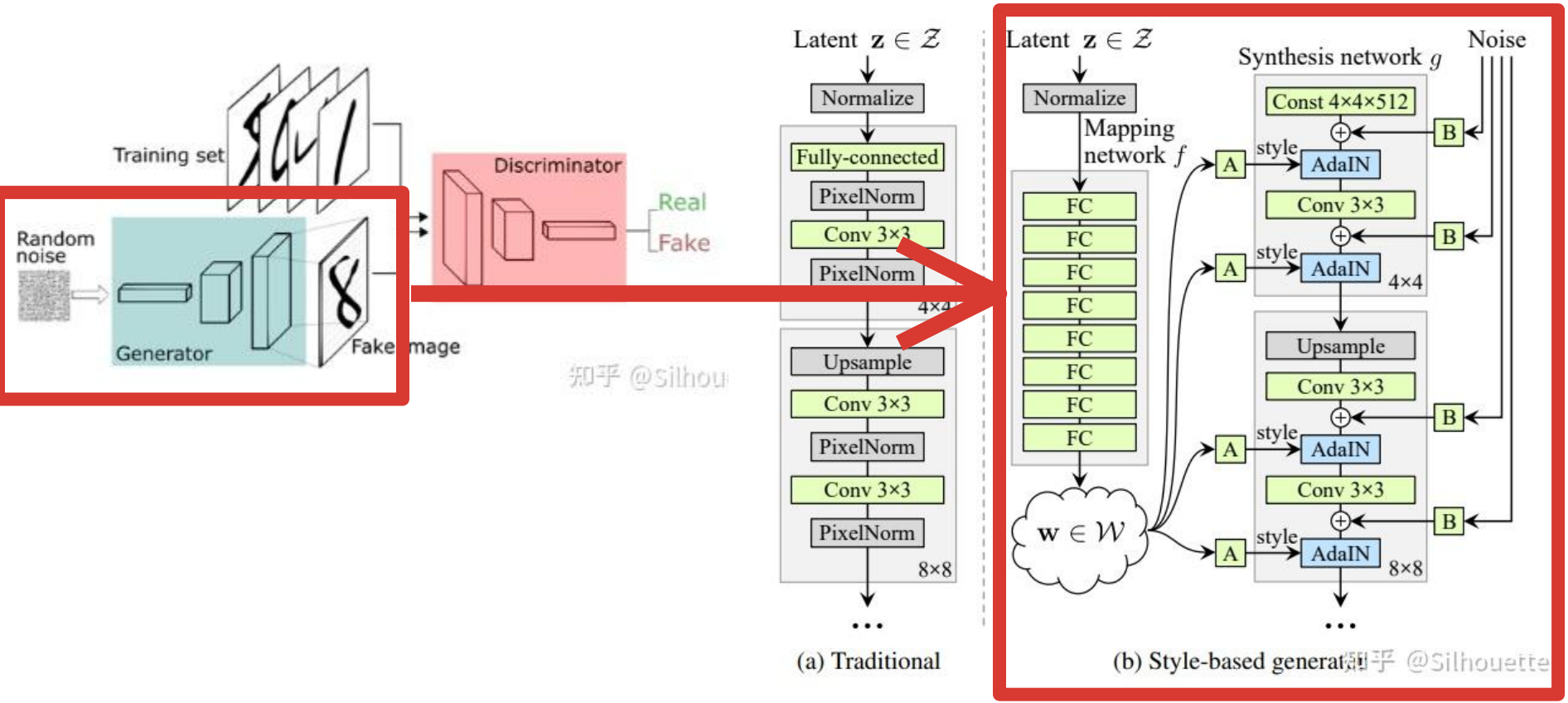
Interaction Method



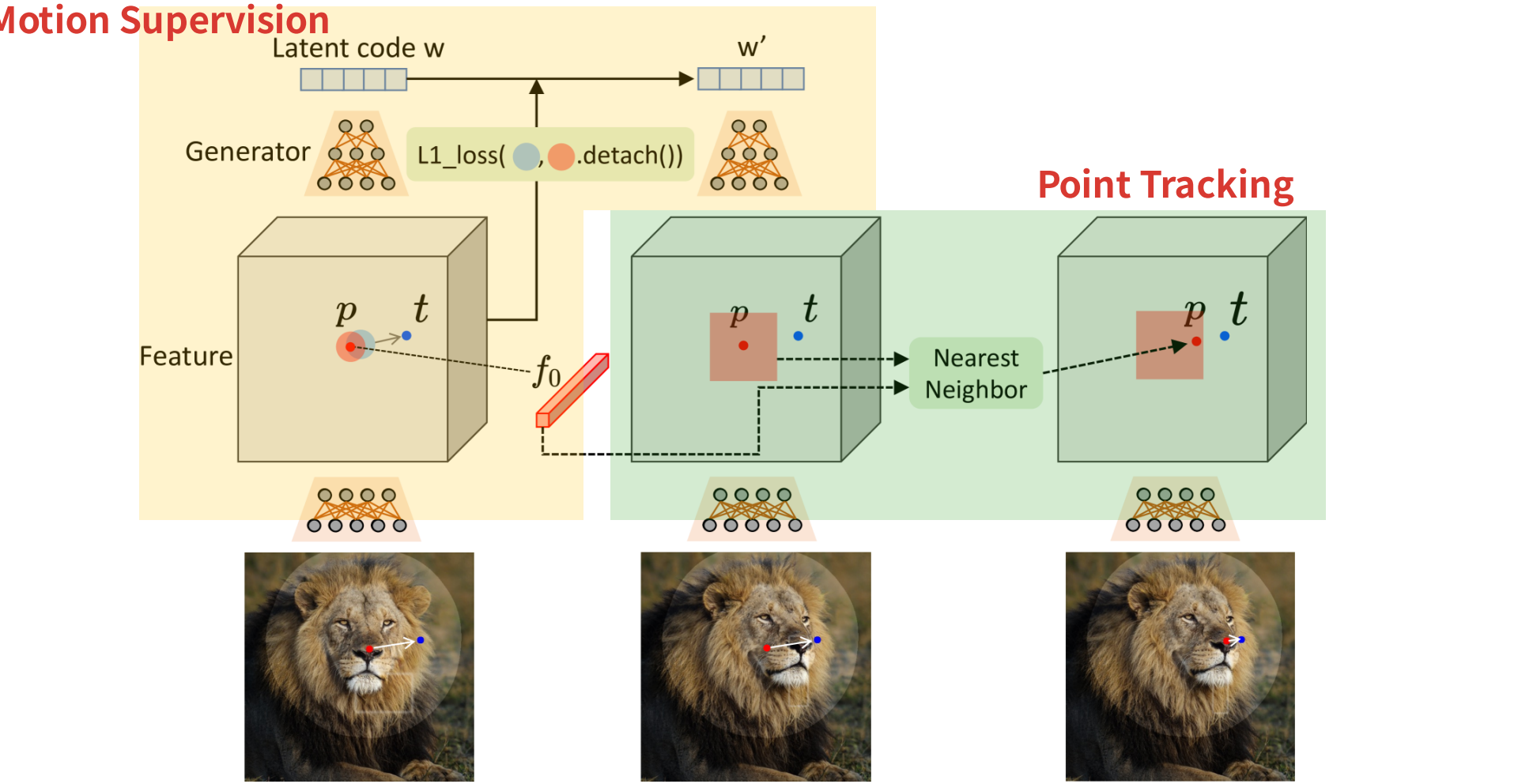
To support this, two tech are given

- Feature-based Motion Supervision
- Point Tracking with Discriminative Generator Features

This is a paper based on Style Gan



Background | **Drag GAN for Generated Image Editing** | Drag Diffusion for Generated Image Editing | MagicQuill for General Imag


Motion Supervision
Latent code w
w'
Generator
L1_loss(, .detach())
Point Tracking
Feature
p
t
f_0
Nearest
Neighbor

Motion Supervision

- Take the feature map after block 6 of StyleGAN
- Supervise a small area around the control point to move around the target point
- A new implicit code is generated
- thereinto$d_i$is a vector to the target point,Frepresentation feature map,$F_0$is the original feature map

$$\mathcal{L}=\sum_{i=0}^{n}\sum_{\boldsymbol{q}_i\in\Omega_1(\boldsymbol{p}_i,r_1)}\left\|\mathbf{F}\left(\boldsymbol{q}_i\right)-\mathbf{F}\left(\boldsymbol{q}_i+\boldsymbol{d}_i\right)\right\|_1+\lambda\left\|\left(\mathbf{F}-\mathbf{F}_0\right)\cdot\left(1-\mathbf{M}\right)\right\|_1$$

Point Tracking

- Update each operation point so that it accurately tracks the corresponding point on the object
- Iteratively change the original control point
- thereinto$F'$for the new feature map,$f_i$is the characteristic of the control point

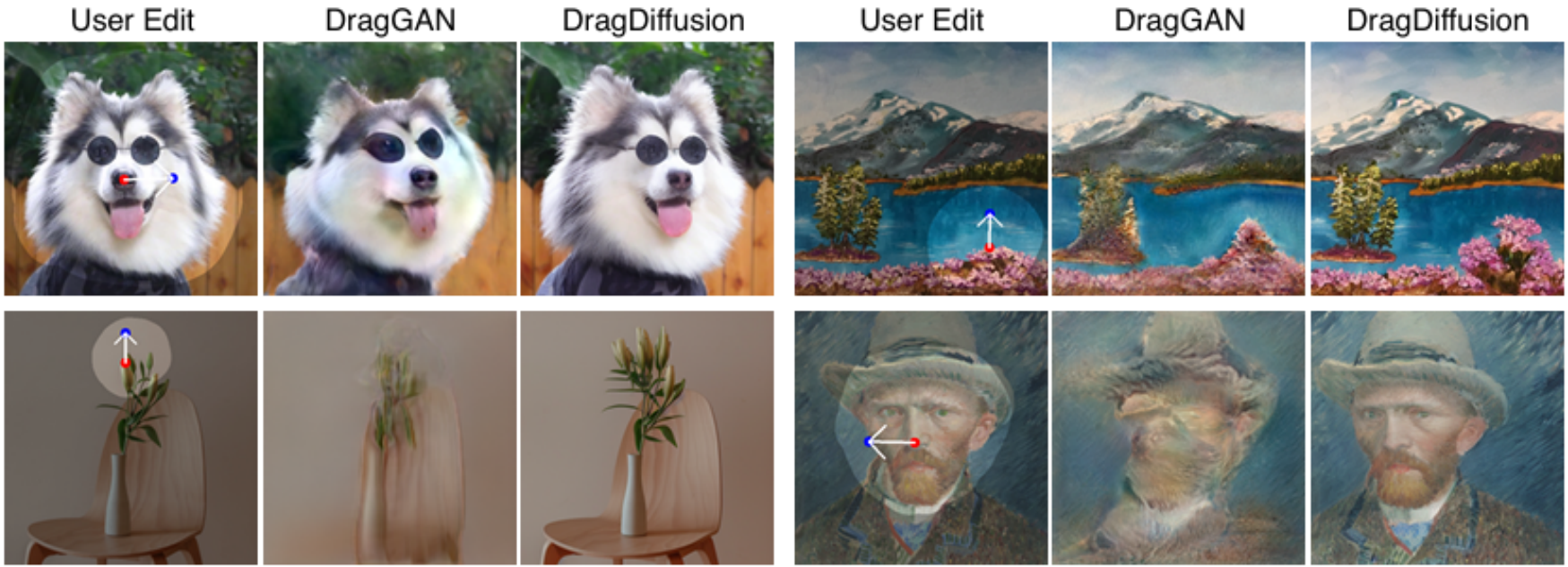
$$\boldsymbol{p}_i:=\underset{\boldsymbol{q}_i\in\Omega_2(\boldsymbol{p}_i,r_2)}{\arg\min}\left\|\mathbf{F}'\left(\boldsymbol{q}_i\right)-f_i\right\|_1$$

Result Seems Fine

However, GAN is not easy to be scale up. Can diffusion model also be easily and finly modified?

Drag Diffusion has better result

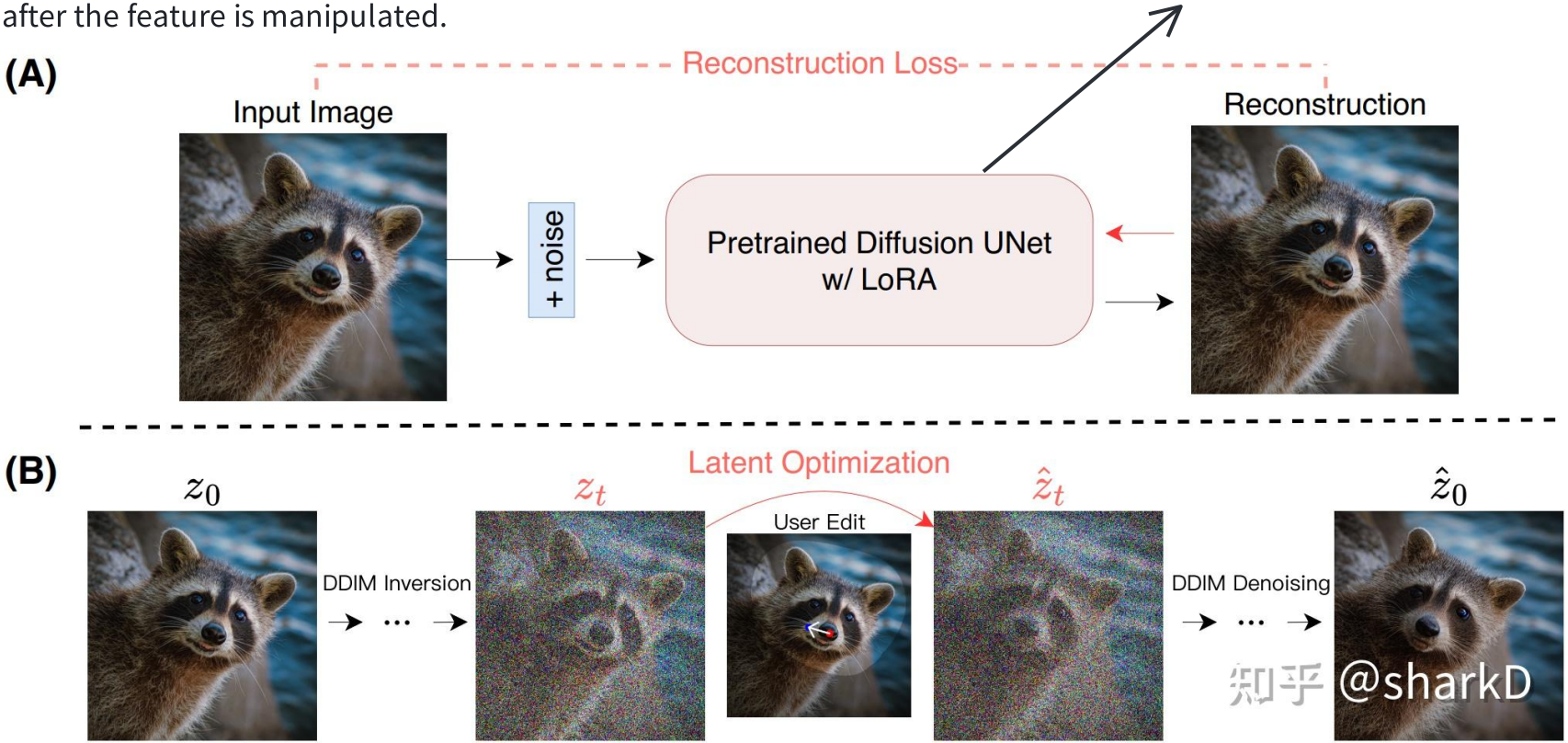
DragDiffusion: Harnessing Diffusion Models for Interactive Point-based Image Editing



Diffusion is much more powerful than GAN.

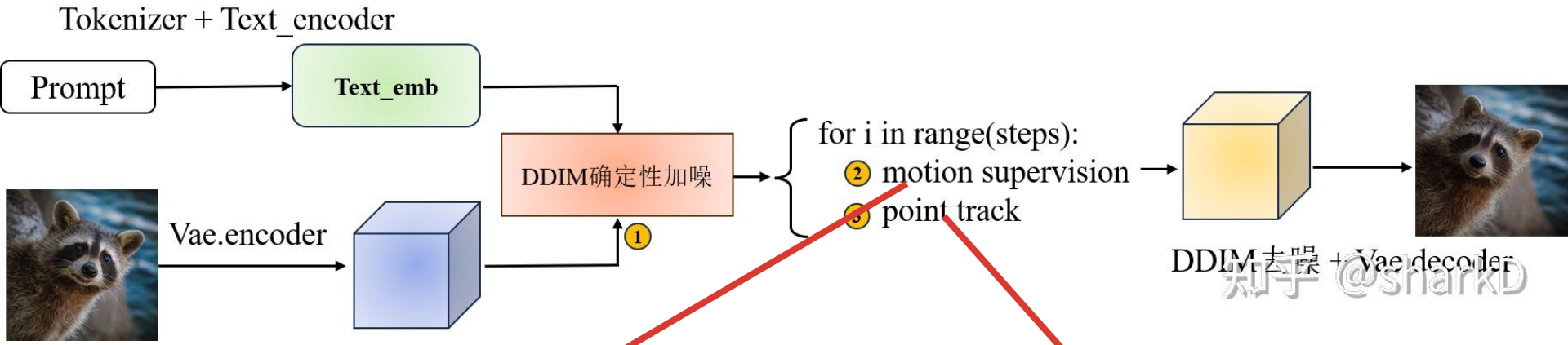
Interaction mathod in Drag GAN can also be involved in diffusion models

The SD (StableDiffusion) is fine-tuned by LoRA and the image you want to edit, **so that the image information can be retained** even after the feature is manipulated.



For part B

StableDiffusionPipline: Tokenizer \ Text_encoder \Vae \ Unet



$$\mathcal{L}(\hat{z}_t^k) = \sum_{i=1}^{n} \sum_{q \in \Omega(h_i^k, r_1)} \left\| F_{q+d_i}(\hat{z}_t^k) - \mathrm{sg}(F_q(\hat{z}_t^k)) \right\|_1 + \lambda \left\| (\hat{z}_{t-1}^k - \mathrm{sg}(\hat{z}_{t-1}^0)) \odot (\mathbb{1} - M) \right\|_1,$$

$$h_i^{k+1} = \arg\min_{q \in \Omega(h_i^k, r_2)} \left\| F_q(\hat{z}_t^{k+1}) - F_{h_i^k}(z_t) \right\|_1 .$$

Drag is not powerful enough for all user needs.

Will there be tool for more flexable image editing?

- User prefer to modify image they provided
- User do not only needs point editing, they prefer more general editing method

Language based editing is a good way for more flexable image editing
But it is not enough for user needs

Inversion-Free Image Editing with Natural Language

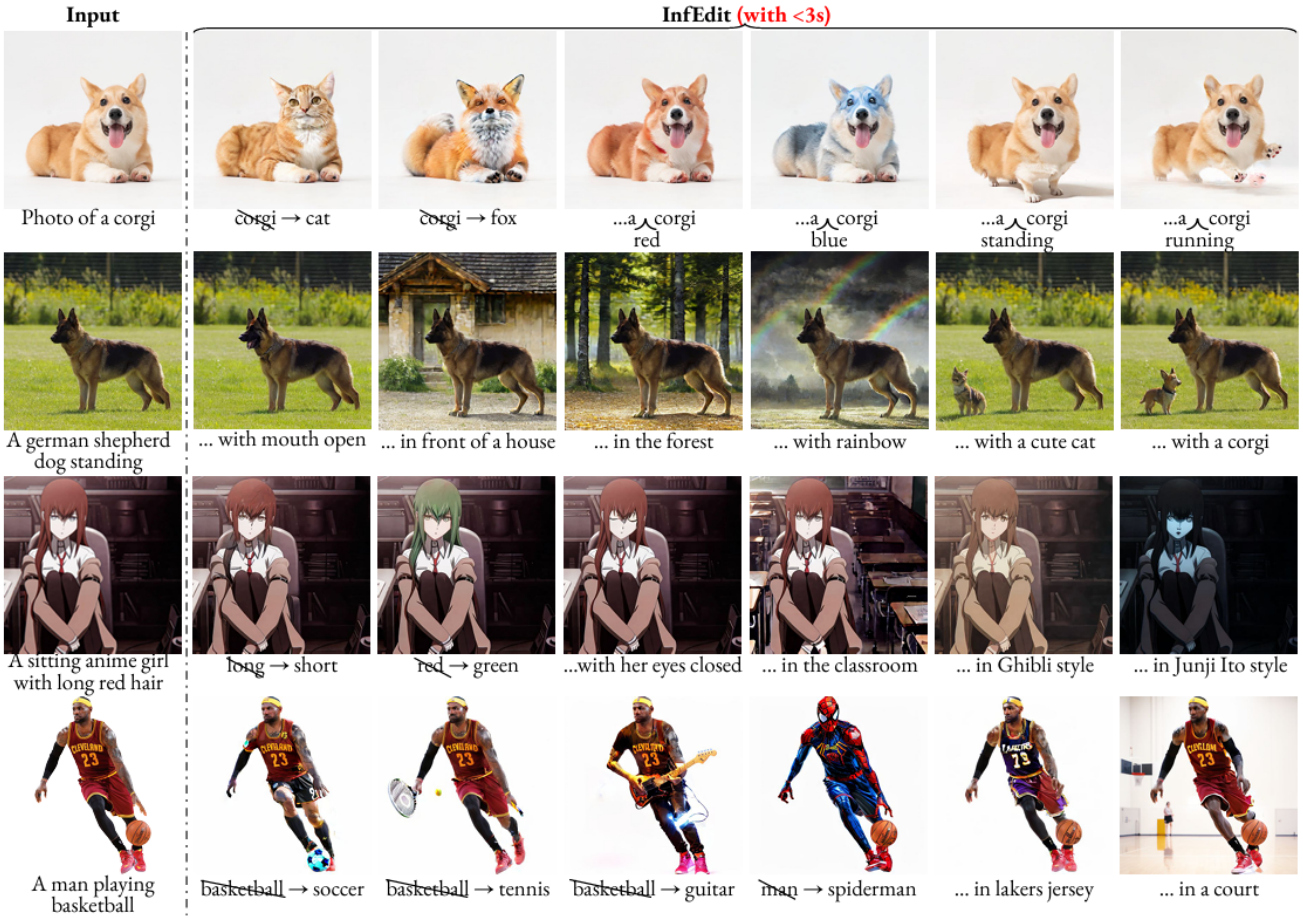
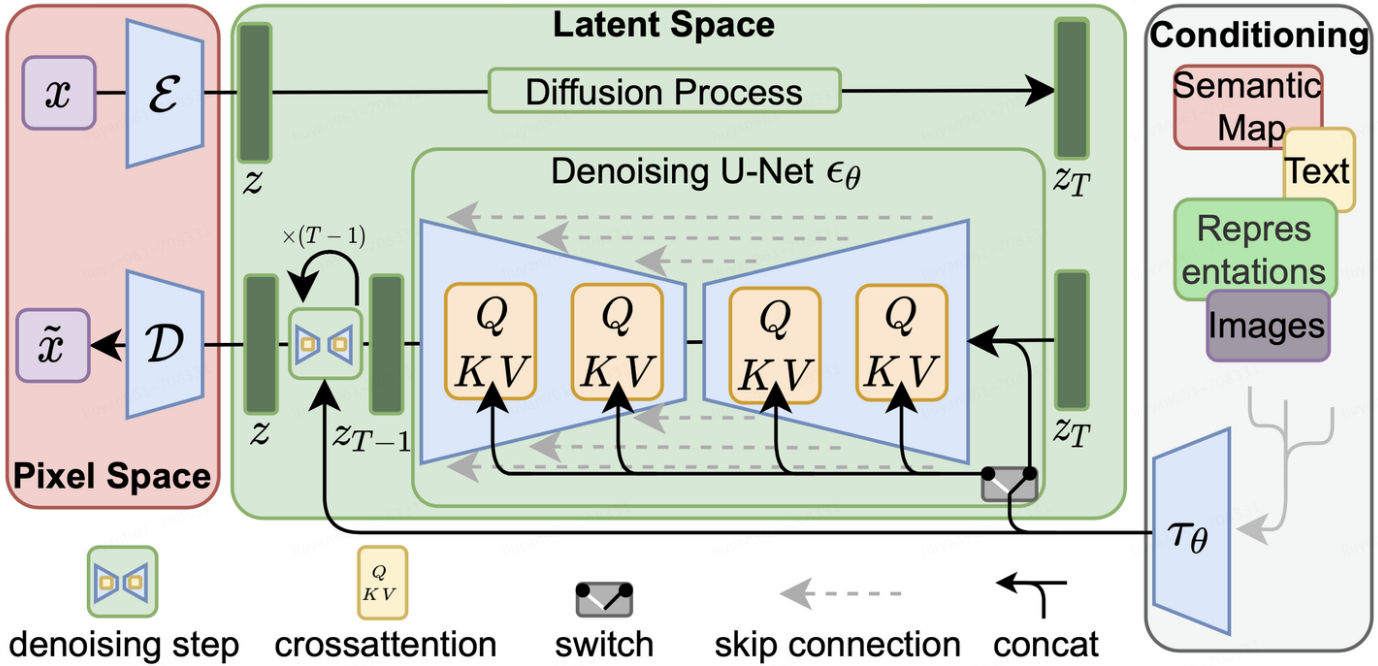


Figure 1. Our inversion-free editing (InfEdit) method demonstrates strong performance in various complex image editing tasks.

Background | Drag GAN for Generated Image Editing | Drag Diffusion for Generated Image Editing | **MagicQuill for General Image**

Existing DDIM/DDPM is hard for image modification with consistent objects

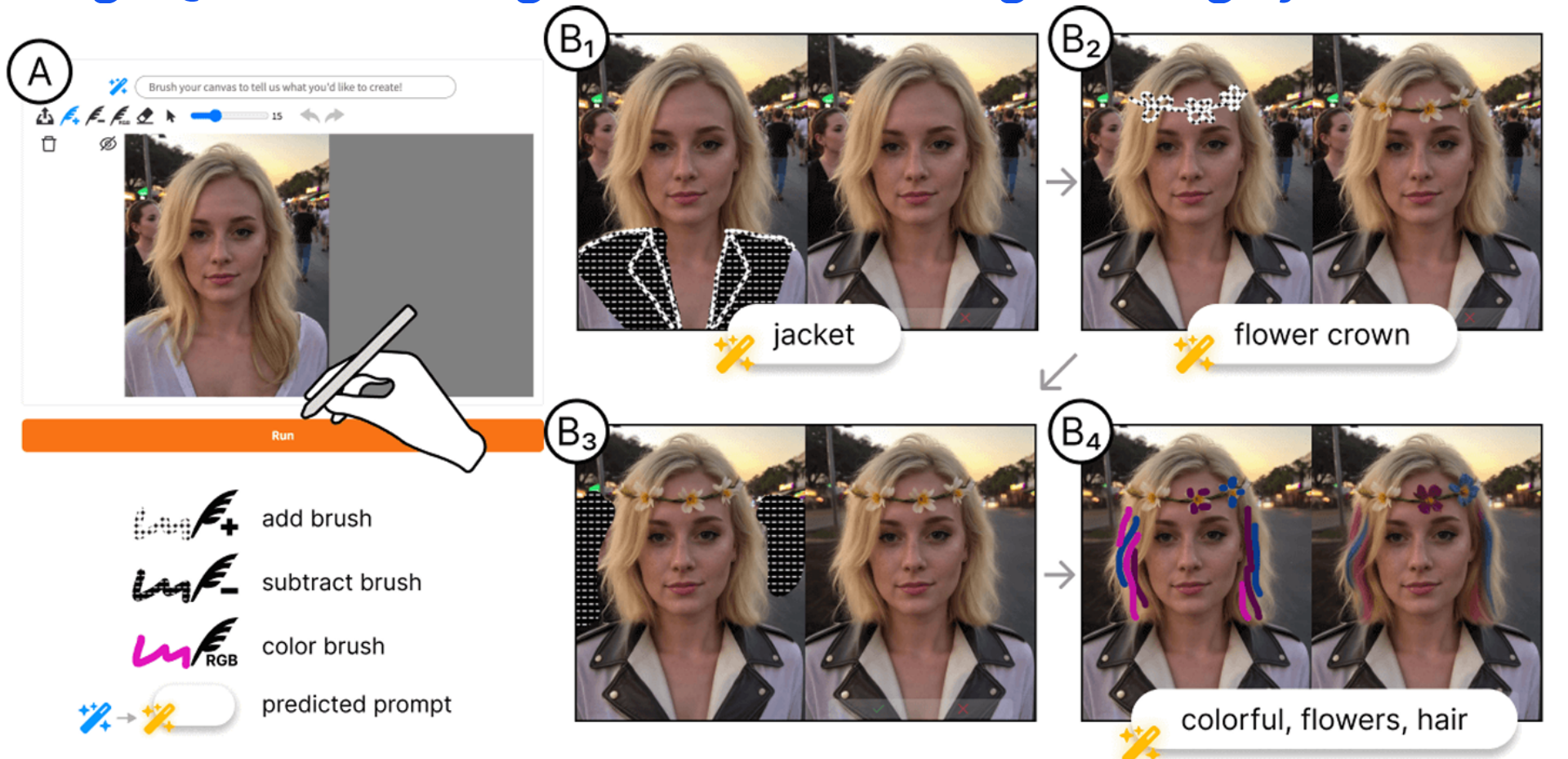


1) The time-consuming nature of the reversal process; **- Inversion-Free Image Editing**

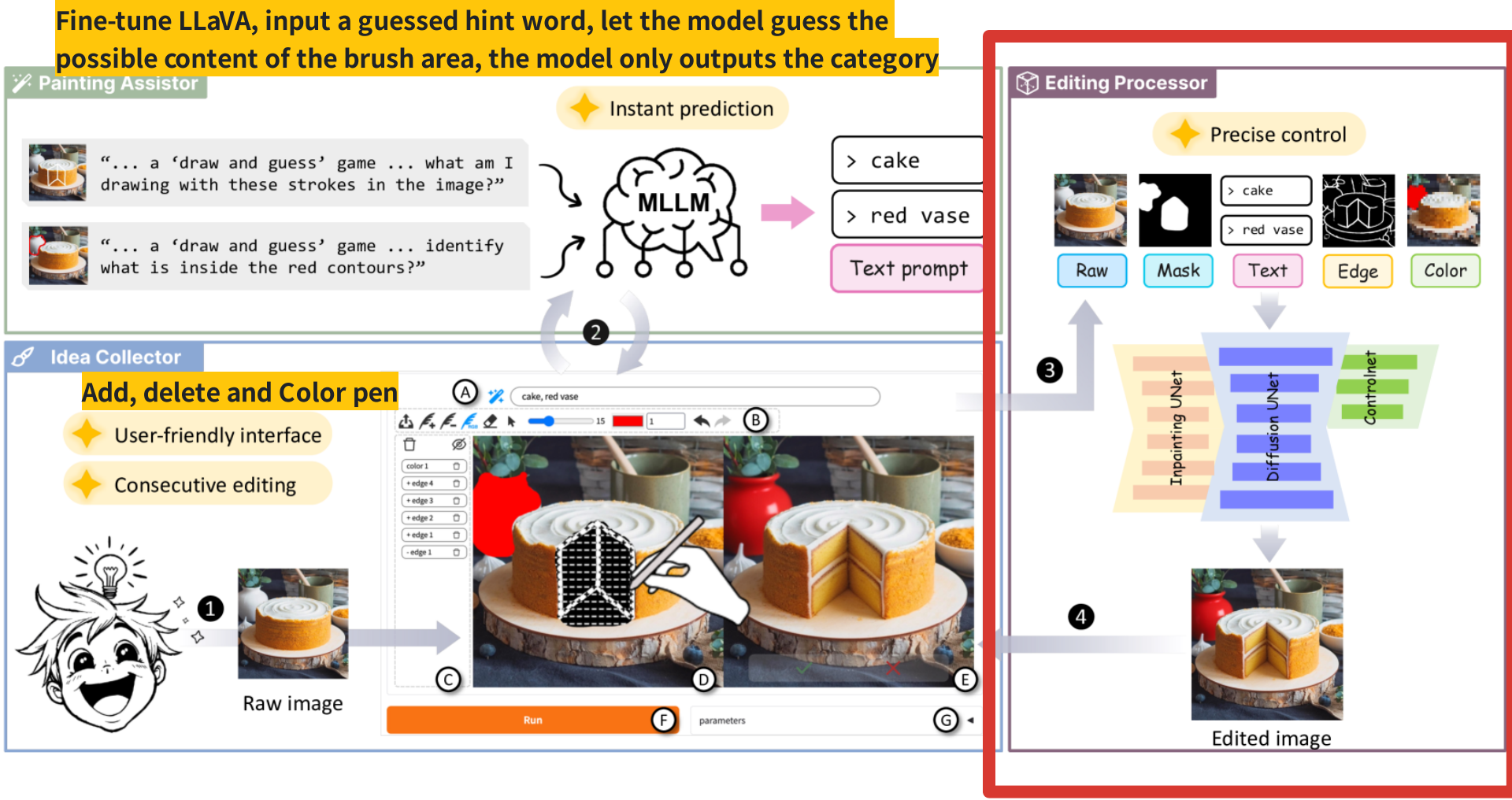
2) The difficulty in balancing consistency and accuracy; **- Denoising Diffusion Consistent Models**

3)They align diffusion model with efficient consistency sampling methods in consistency models.

MagicQuill: An Intelligent Interactive Image Editing System

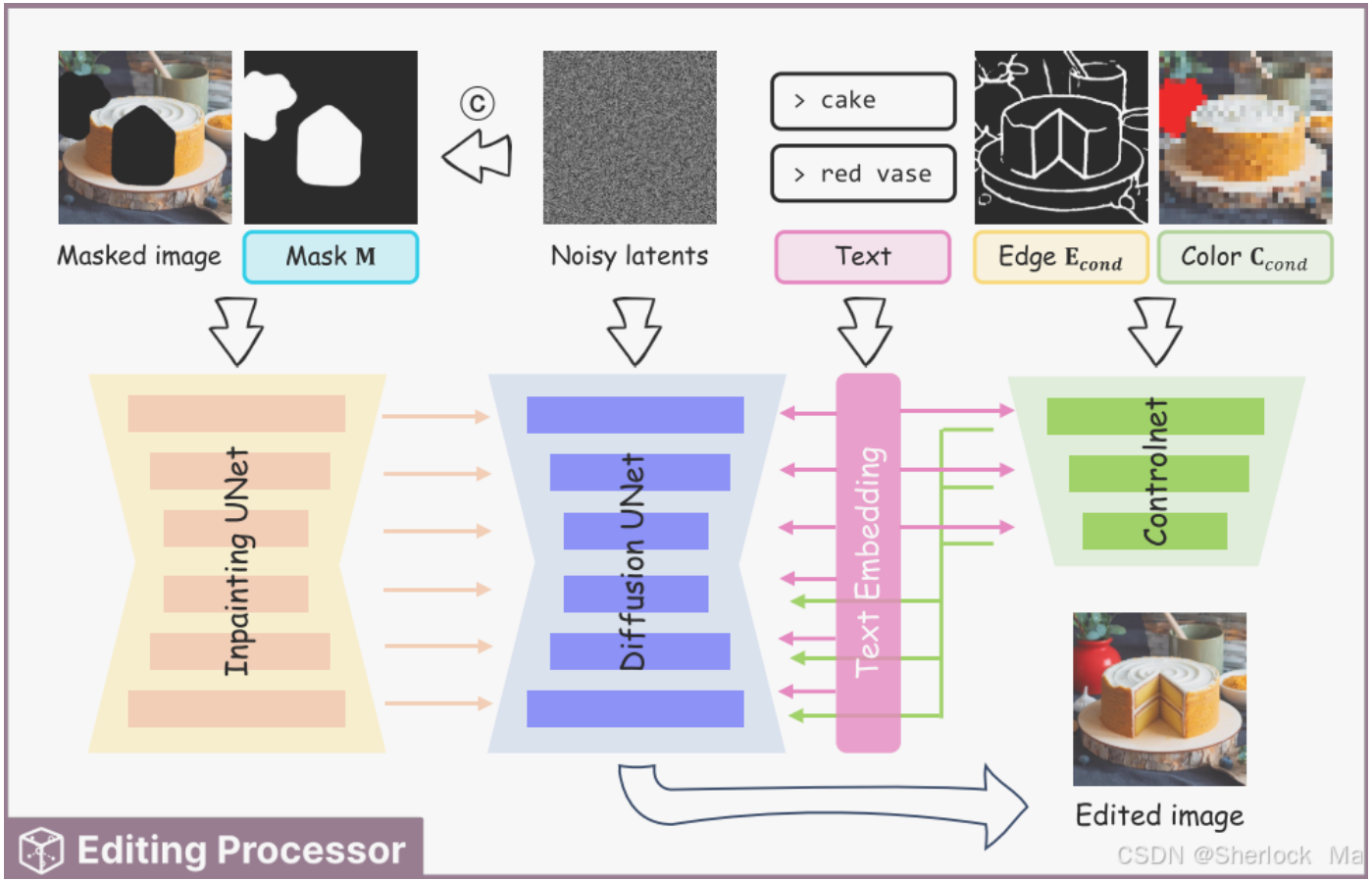


Background | Drag GAN for Generated Image Editing | Drag Diffusion for Generated Image Editing | **MagicQuill for General Image**



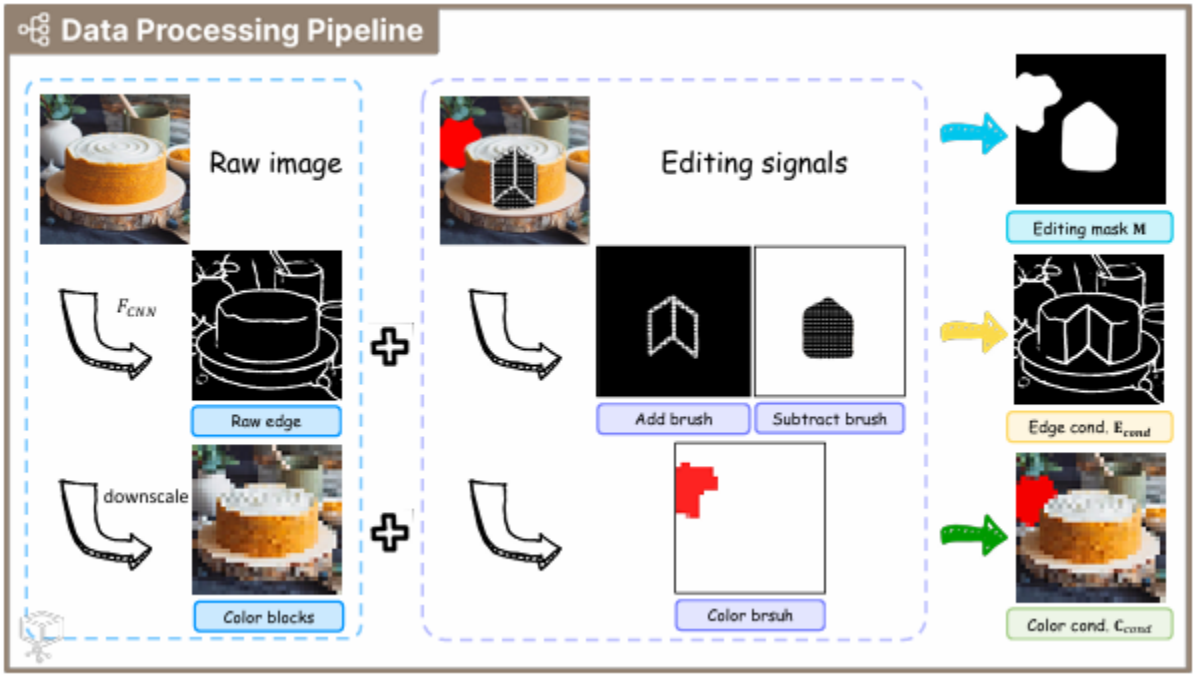
Background | Drag GAN for Generated Image Editing | Drag Diffusion for Generated Image Editing | **MagicQuill for General Image**

Model Structure



Background | Drag GAN for Generated Image Editing | Drag Diffusion for Generated Image Editing | **MagicQuill for General Image**

Standerlize each image to model input structure

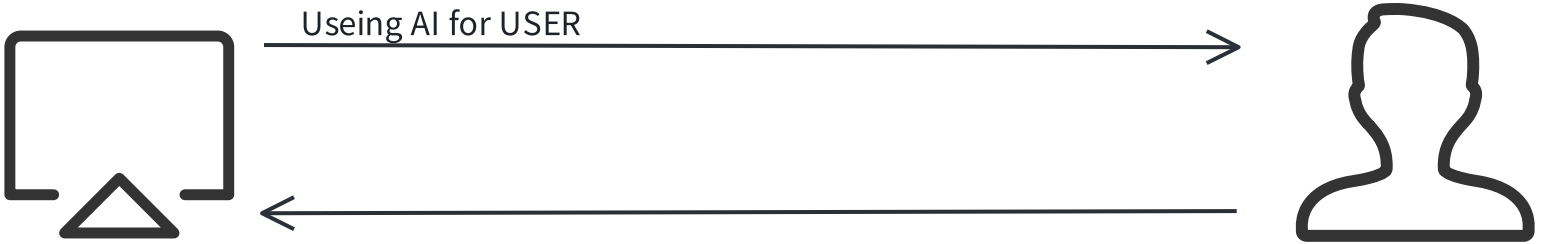


Dataset Generation

(a) Original Image (b) Edge Map (c) Chosen Mask (d) Inpainting Result (e) Edge Overlay

Figure 5. Illustration of dataset construction process. (a) Original images from the DCI dataset; (b) Edge maps extracted from original images; (c) Selected masks (highlighted in purple) with highest edge density; (d) Results after BrushNet inpainting on augmented masked regions; (e) Final results with edge map overlay on selected areas. By overlaying edge maps on inpainted results, we simulate scenarios where users edit images with brush strokes, as the edge maps resemble hand-drawn sketches. The bounding box coordinates of the mask and labels are inherited from the DCI dataset.

Discussion



Enhence AI by user needs