General Articulated Objects Manipulation in Real Images via Part-Aware Diffusion Process

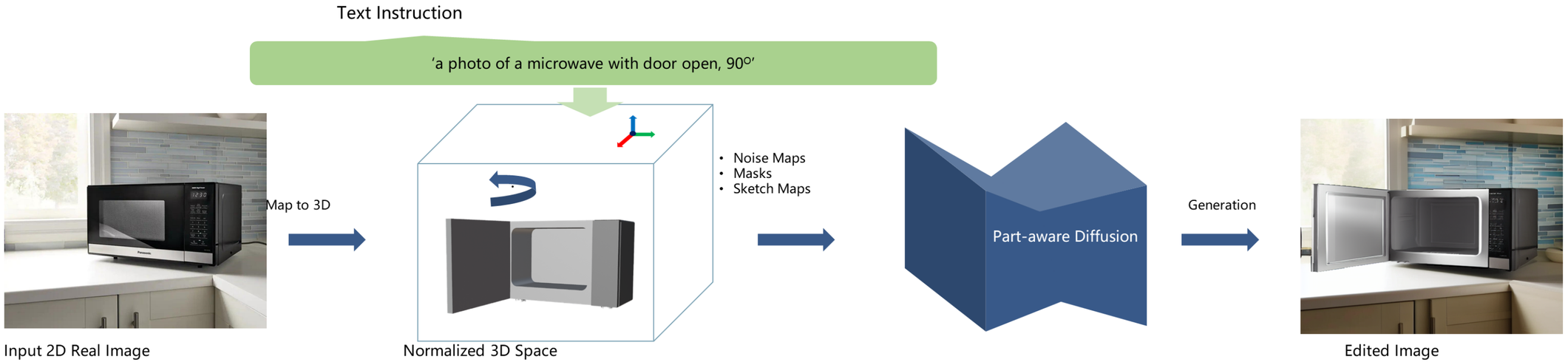
Zhou Fang, Yong-Lu Li*, Lixin Yang, Cewu Lu*

Shanghai Jiao Tong University

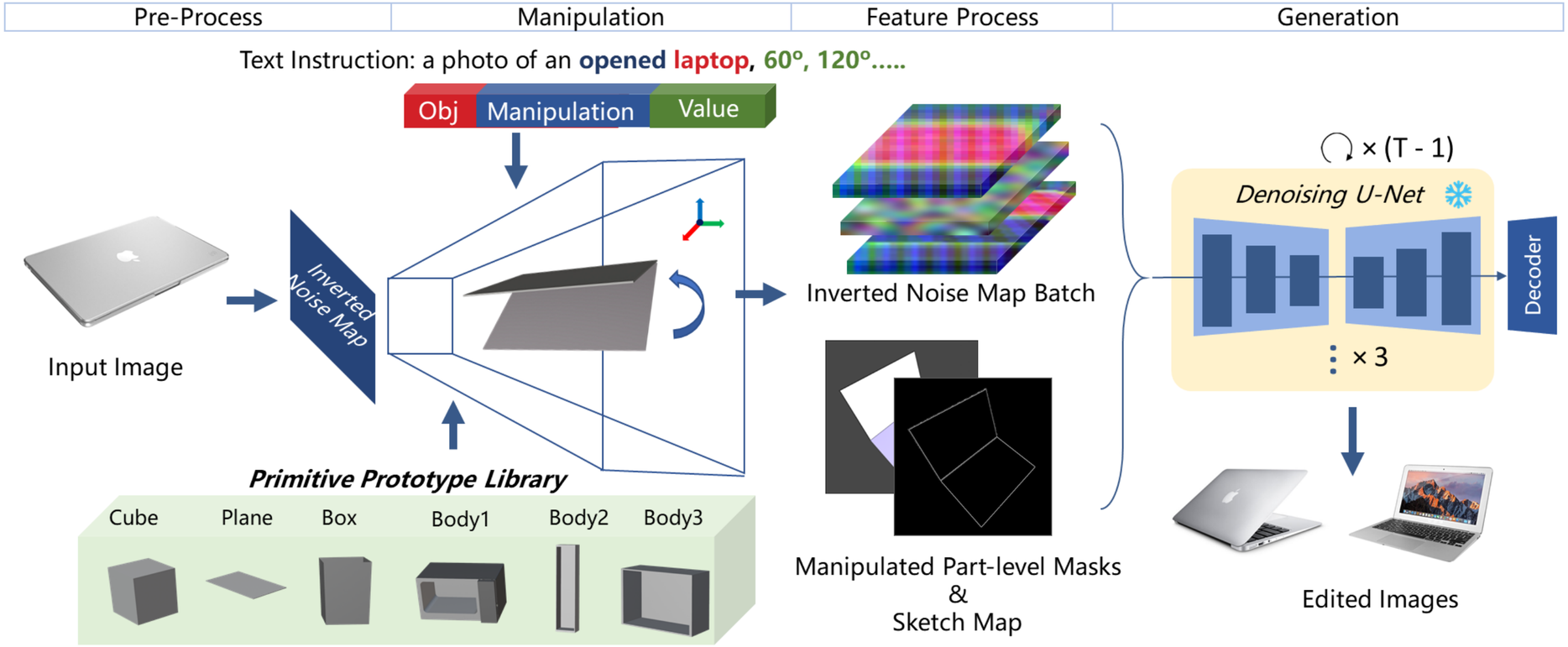
Introduction

Image editing is a fundamental step in computer and robotic vision tasks. However, manipulating **articulated objects** in real images is still challenging.

We introduce **the Part-Aware Diffusion Model** to approach the arbitrary manipulation of articulated objects in real images with text instructions or human interaction.



Method



The overall image editing process

1. Arbitrary Manipulation in 3D Space

- Abstract 3D models of articulated objects are constructed referring to the input 2D real images.
- Arbitrary manipulation can be done in 3D space based on the text instruction or human interaction.

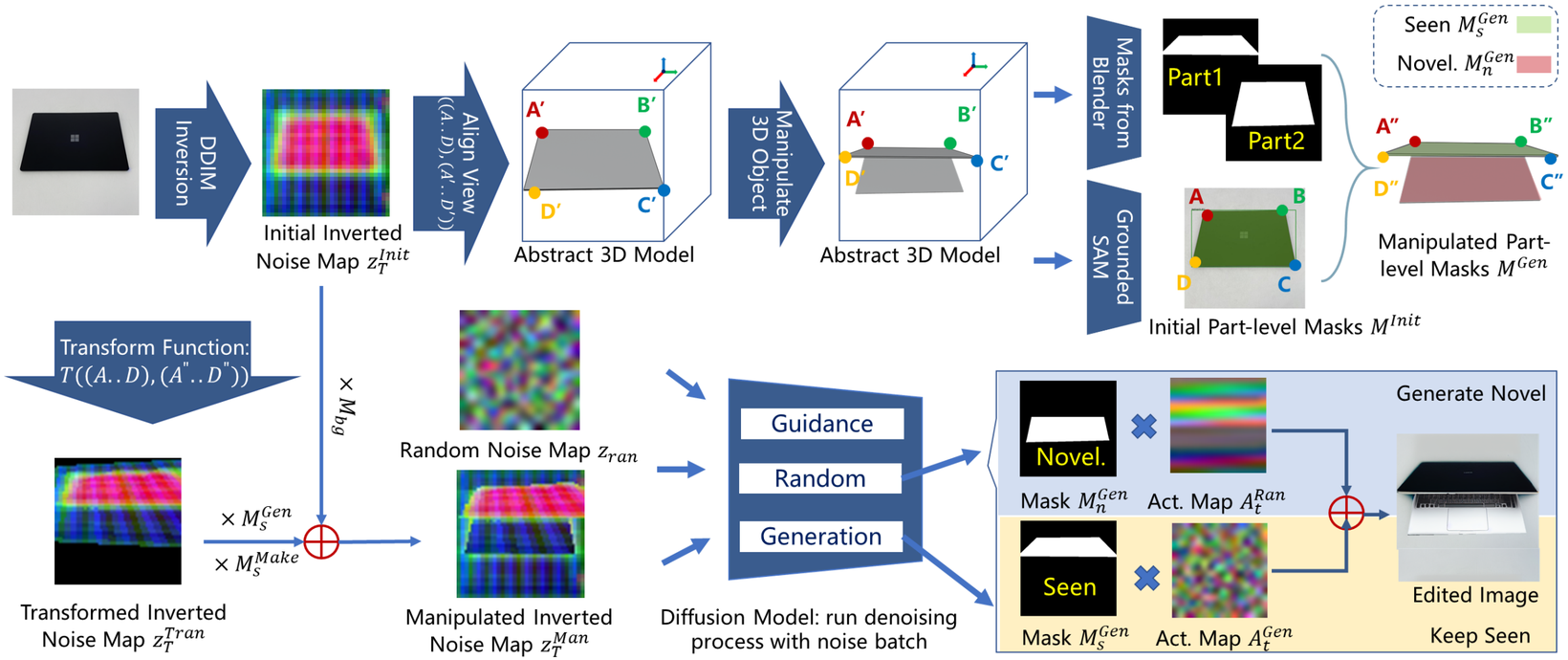
Method

2. Dynamic Feature Maps

- Seen part and novel appearing part masks are defined and exported after manipulation (upper part).
- Diffusion outputs are combined according to previous masks (lower part).

3. Score Function

- Texture Consistency Score Loss: transfer appearance of seen part
- Style Consistency Score Loss: improve consistency of novel appearing part

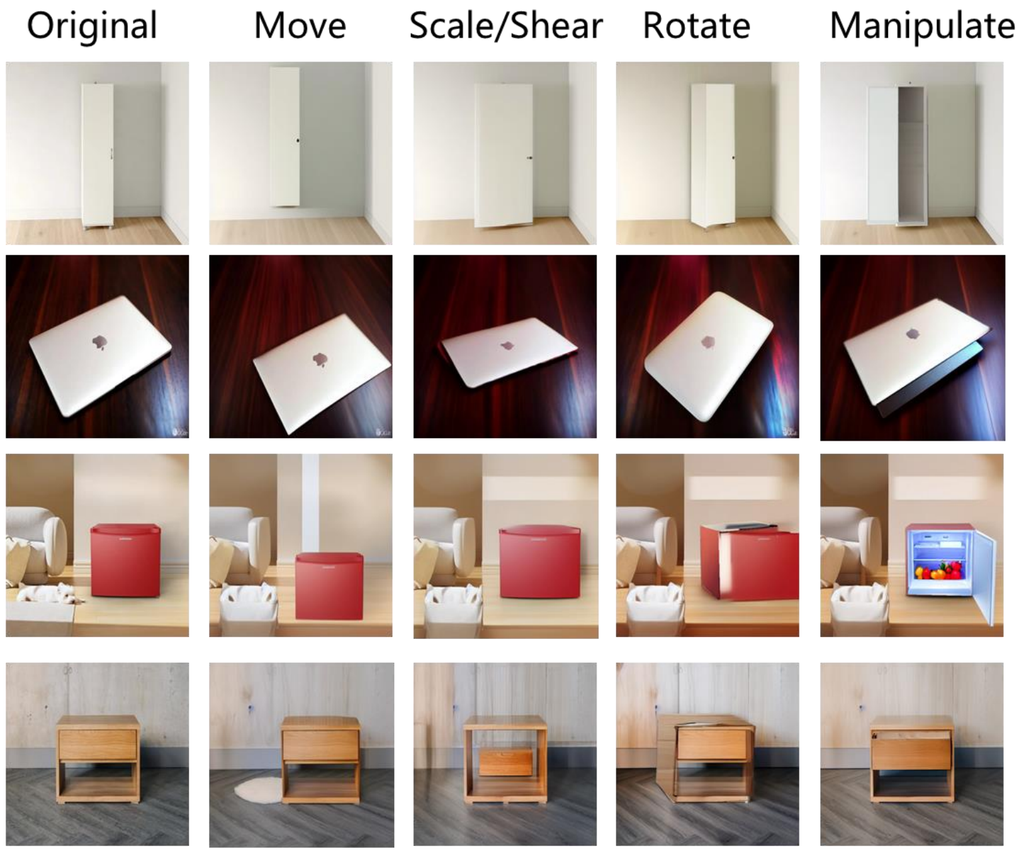


Experiment

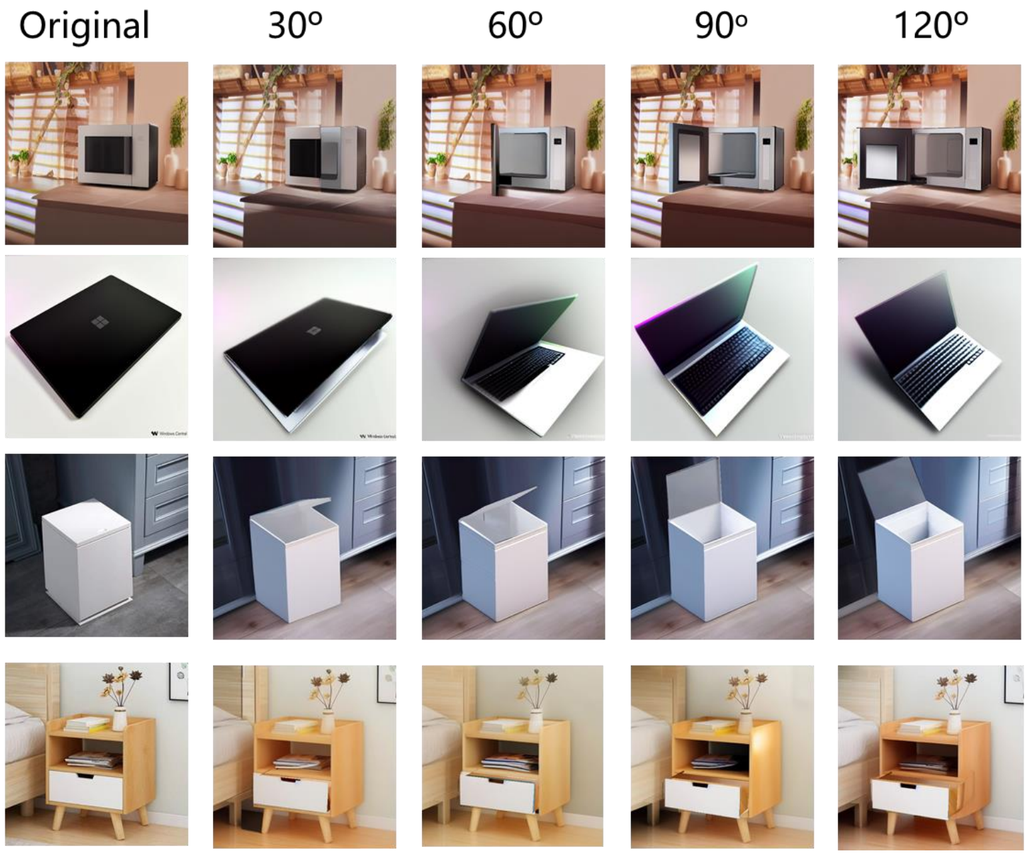
The proposed PA-Diffusion model:

(a) implements several basic manipulation including move, scale/shear. rotate, and manipulate;

(b) opening process from 0^o to 120^o .



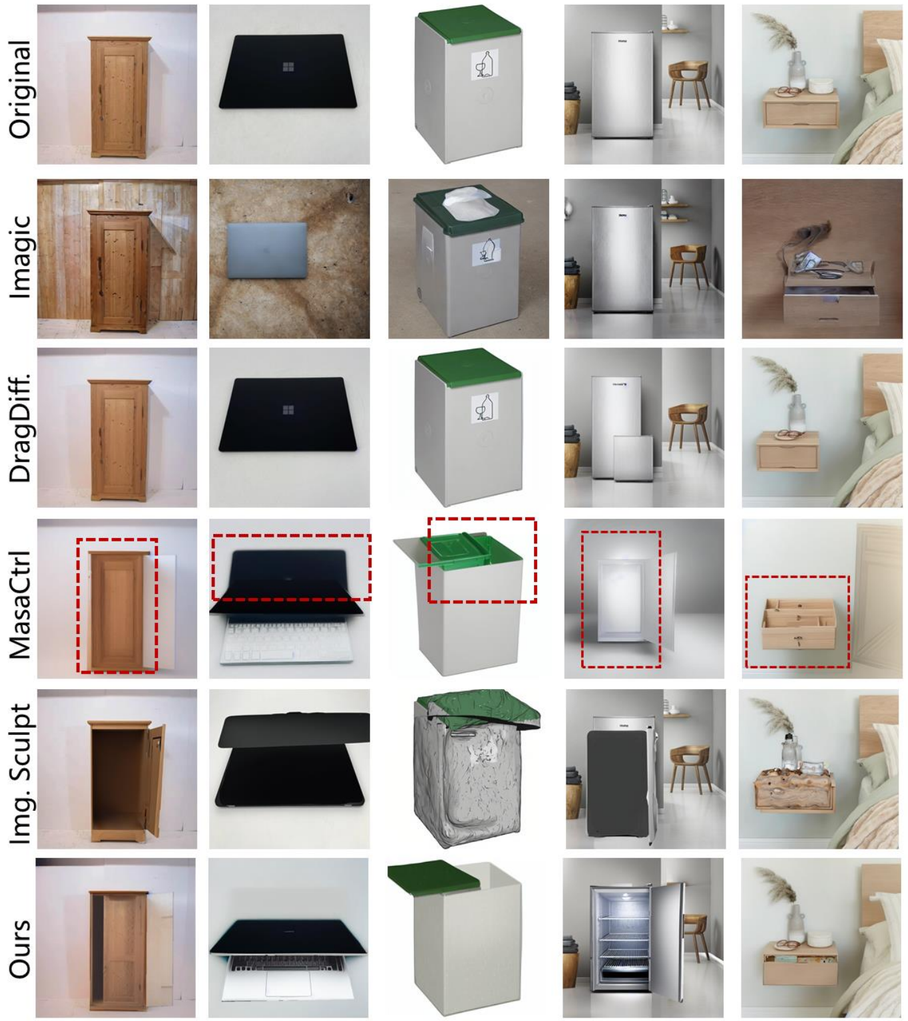
(a)



(b)

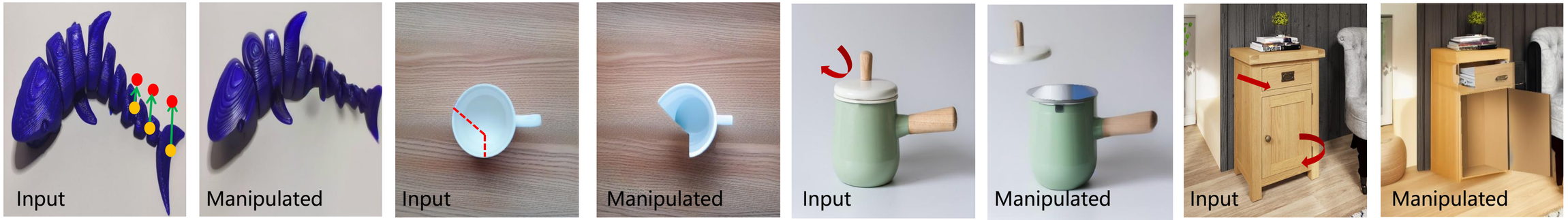
Experiment

- With TCSL and SCSL, the edited images are consistent to the original and more reasonable (left).
- Comparison with SOTA image editing methods, the proposed PA-Diffusion model can successfully manipulate various articulated objects without artifacts (right).



Experiment

Non-rigid objects, ***non-uniform*** shapes, objects with ***weird or multiple joints*** can still be manipulated with the proposed PA-Diffusion model.



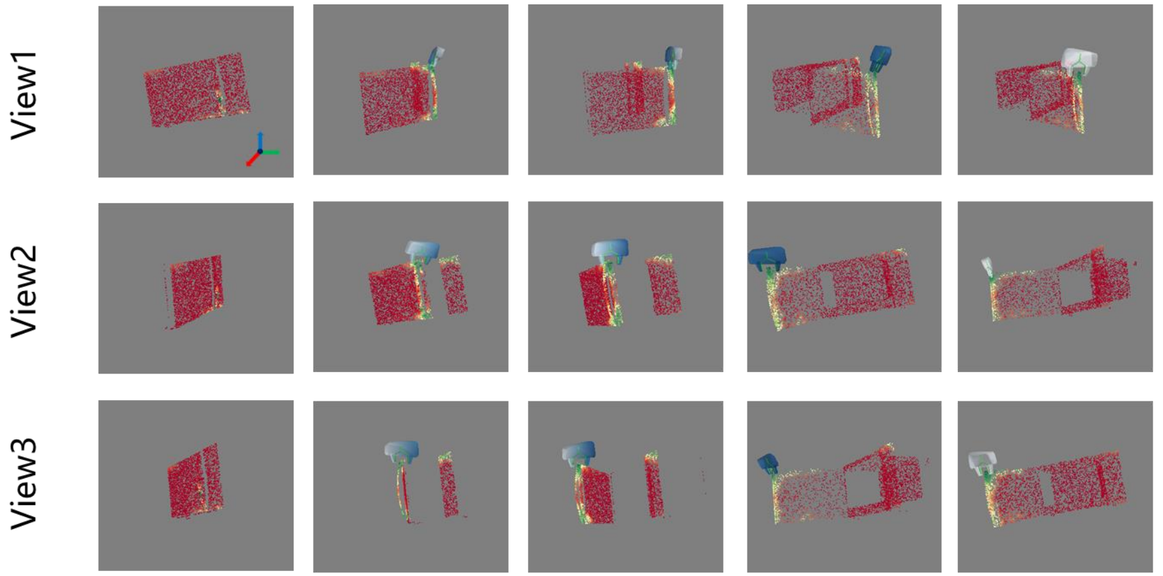
Robotic Application

Our proposed PA-Diffusion model can also be utilized to generate **Real** or **Synthetic** robot manipulation samples as follows:

Step1. Generate video of manipulation process

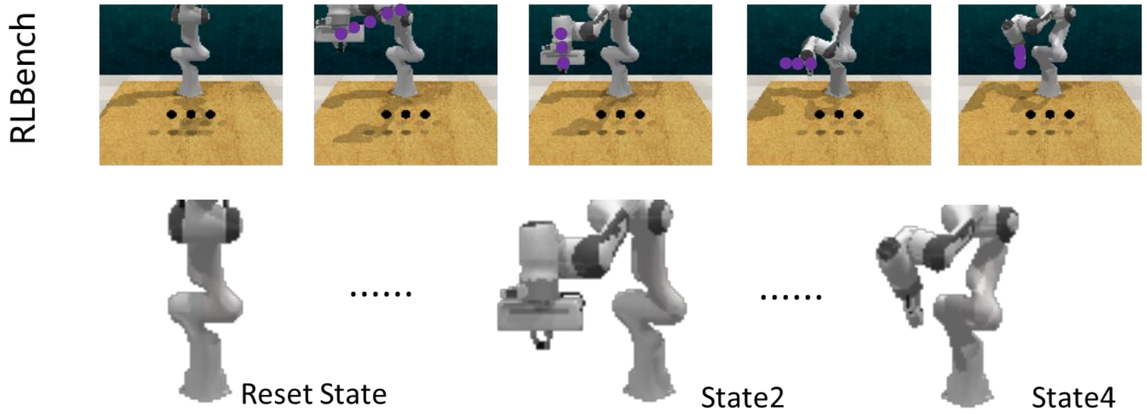
(a) Manipulate the articulated object

Step2. Generate gripper pose for each state



(b) Extract the 3D end-effector pose in each state

Step3. Generate robot arm trajectory in the simulator



(c) Robot arm & end-effector move from reset to the above specified pose in each state

Future works

Next step, we plan to :

1. Cover **more** categories of articulated objects.
2. Extend this method to handle **deformable objects** and **fluid**.
3. Expand edit image dataset to million-scale for supporting various tasks in **robotic scenarios**.

Thank You