



PDFactor

Learning Tri-Perspective View Policy Diffusion Field for Multi-Task Robotic Manipulation

Jingyi Tian¹ Le Wang¹ Sanping Zhou¹ Sen Wang¹
Jiayi Li¹ Haowen Sun¹ Wei Tang²

¹ Xi'an Jiaotong University

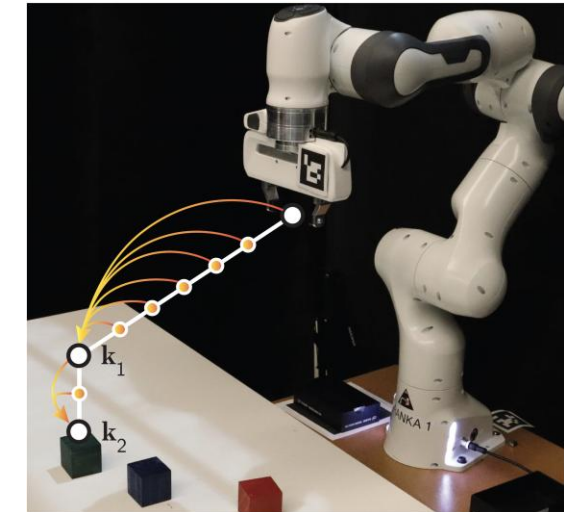
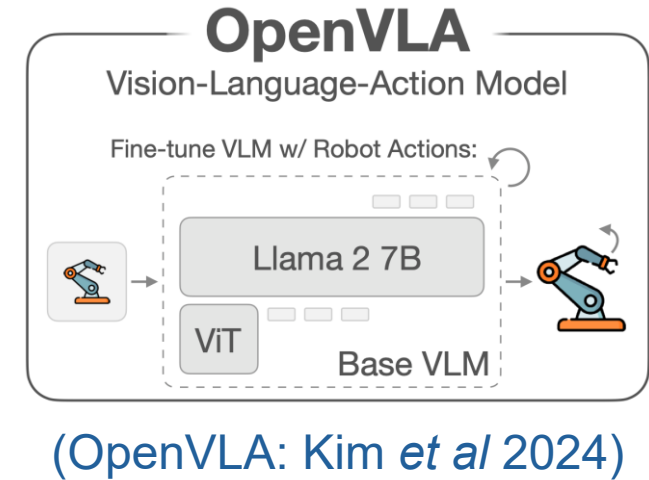
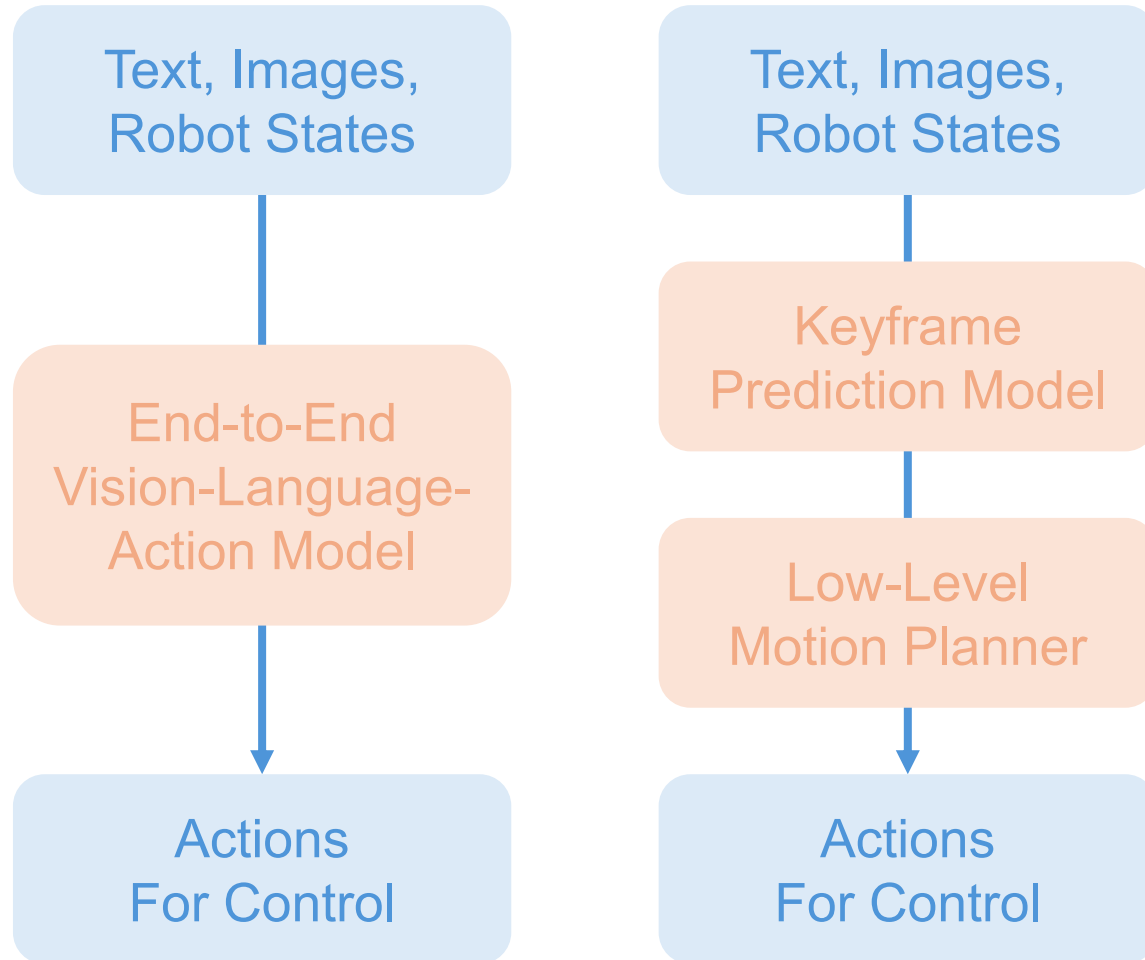
² University of Illinois at Chicago

Manipulation



(RT-1: Brohan *et al* 2022)

Existing Approaches



(PerAct: Shridhar *et al* 2022)

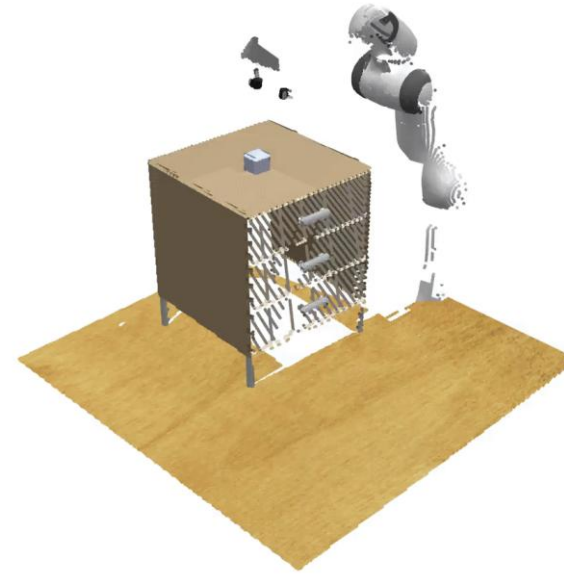
Action Value Map

Pros:

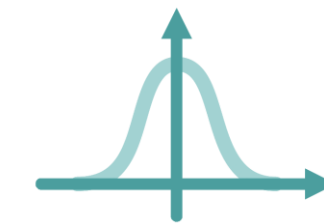
- Explicit alignment with the perception space.
- Fast inference speed.

Cons:

- Approximated distribution.



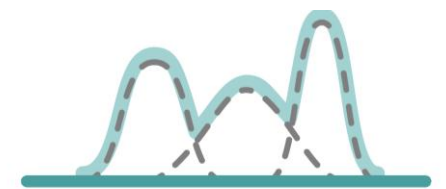
Action Value Map



Regression



Category



Gaussian
Mixture

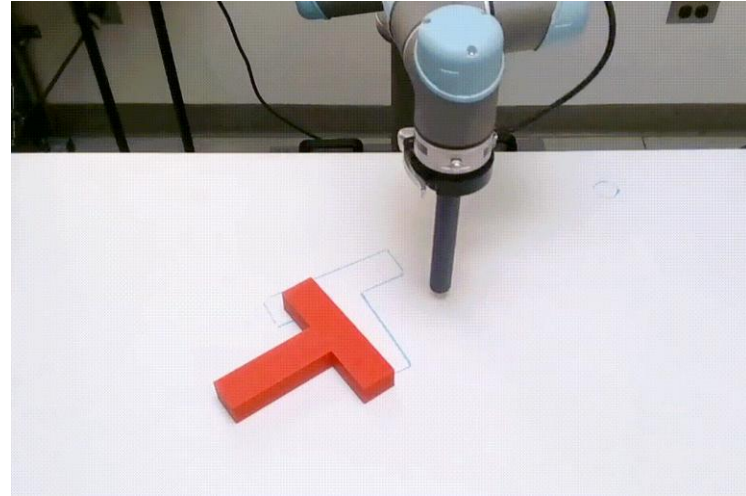
Implicit Policy

Pros:

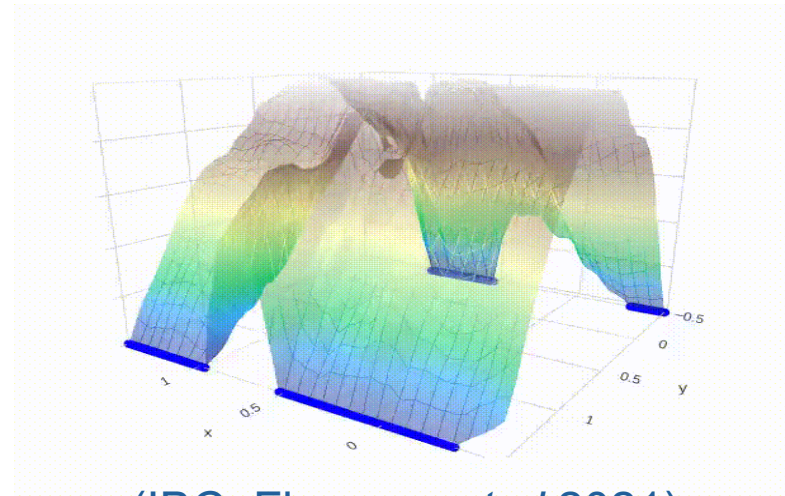
- Any bounded distribution.
- Action space scalability.

Cons:

- Slow inference.
- Perception-action misalignment.



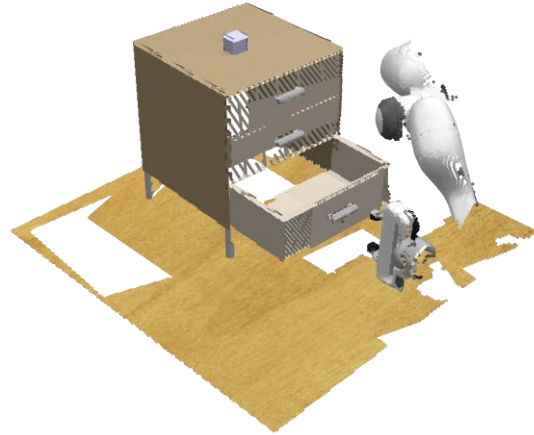
(Diffusion Policy: Chi *et al* 2023)



(IBC: Florence *et al* 2021)

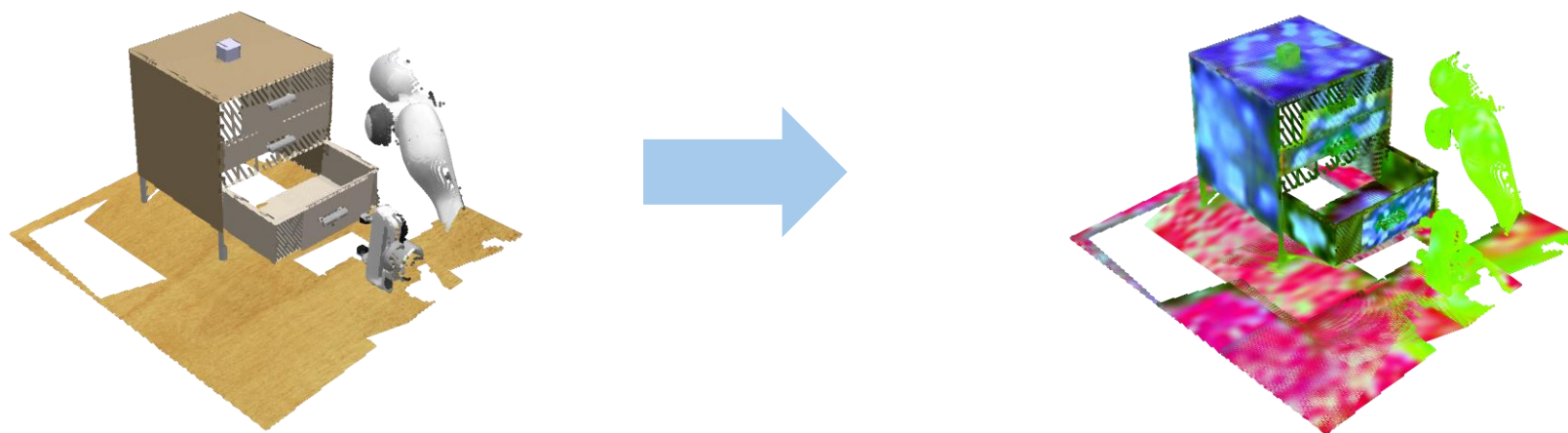
Our Approach

Representing scenes as triplanes.



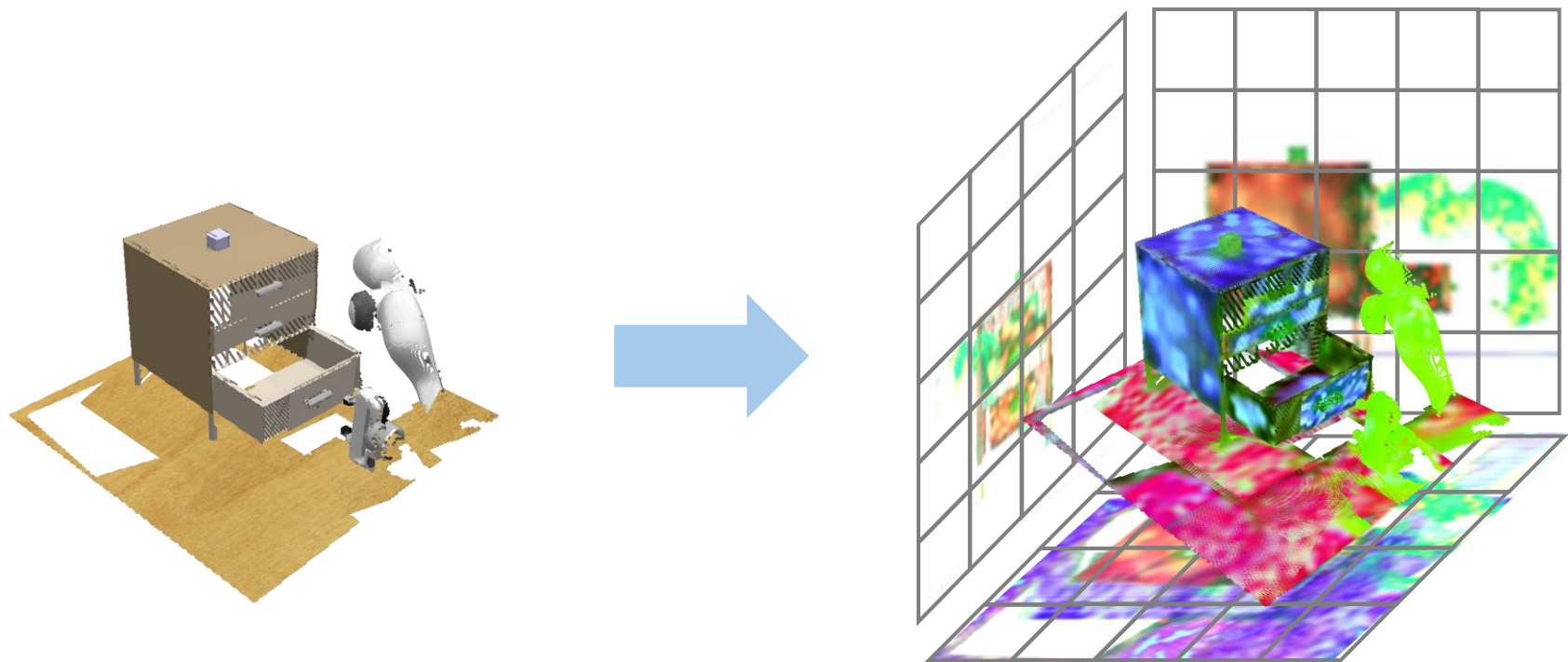
Our Approach

Representing scenes as triplanes.



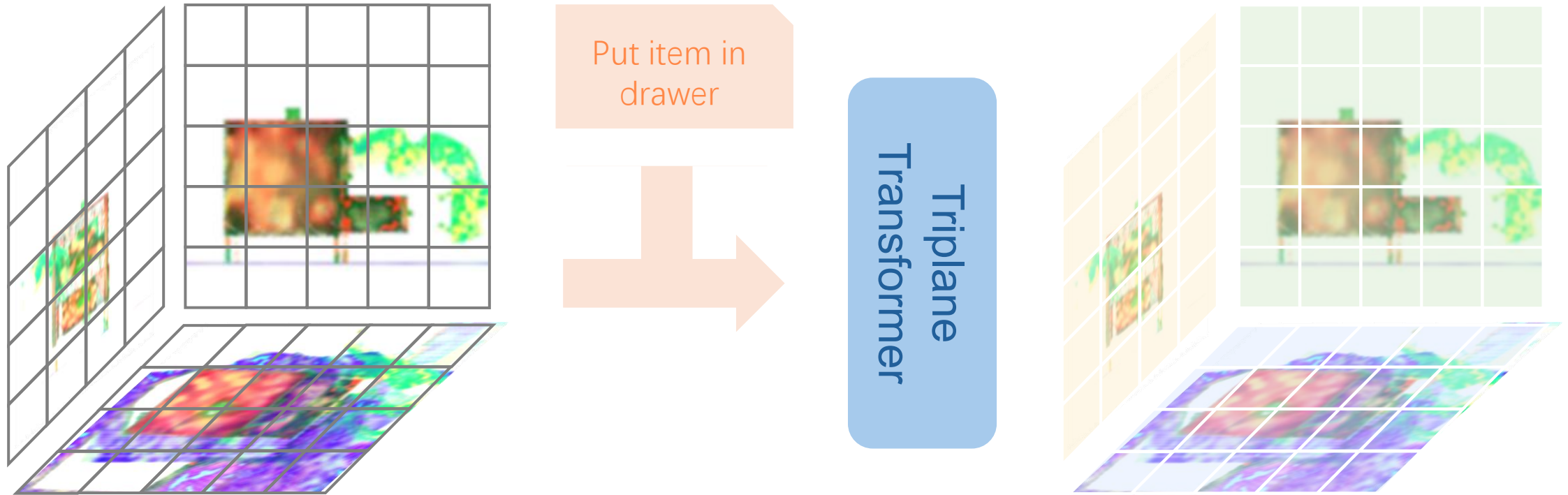
Our Approach

Representing scenes as triplanes.



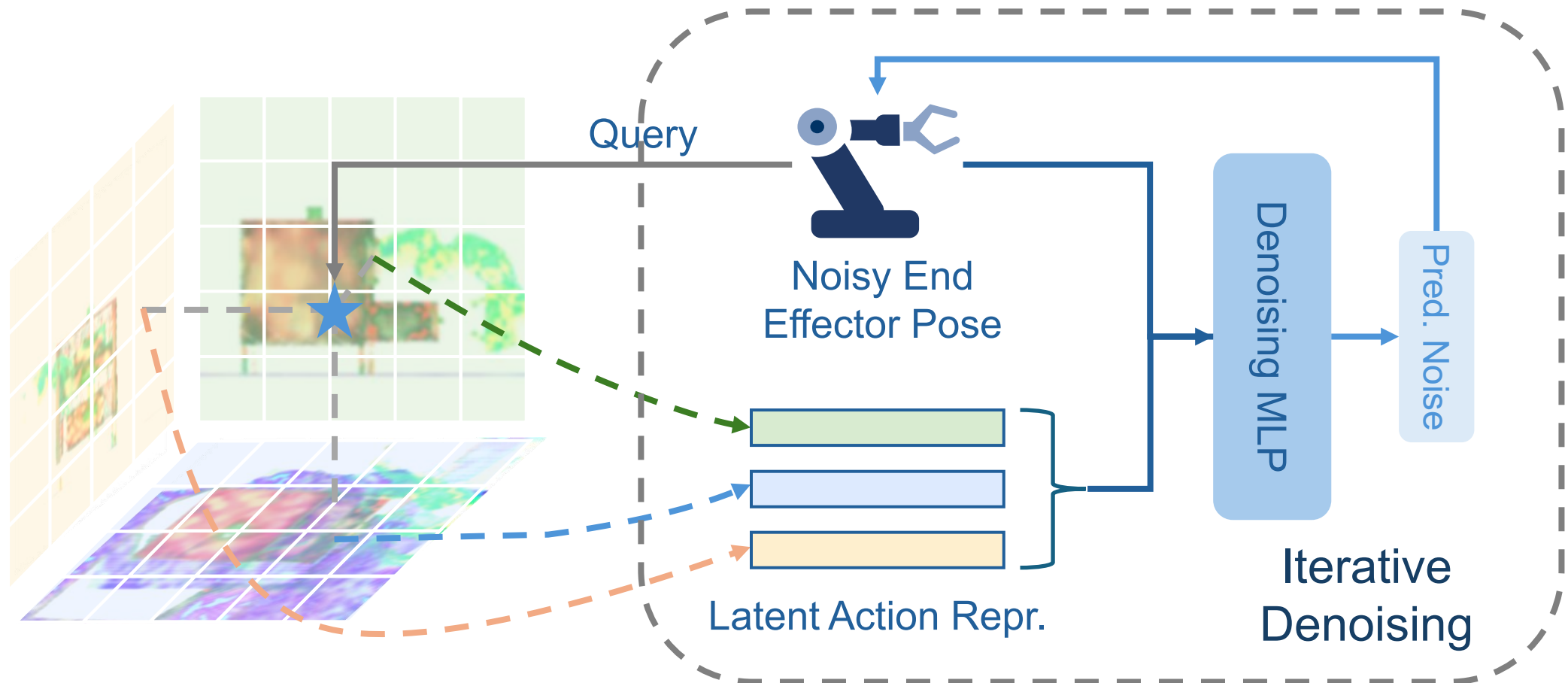
Our Approach

Representing scenes as triplanes.



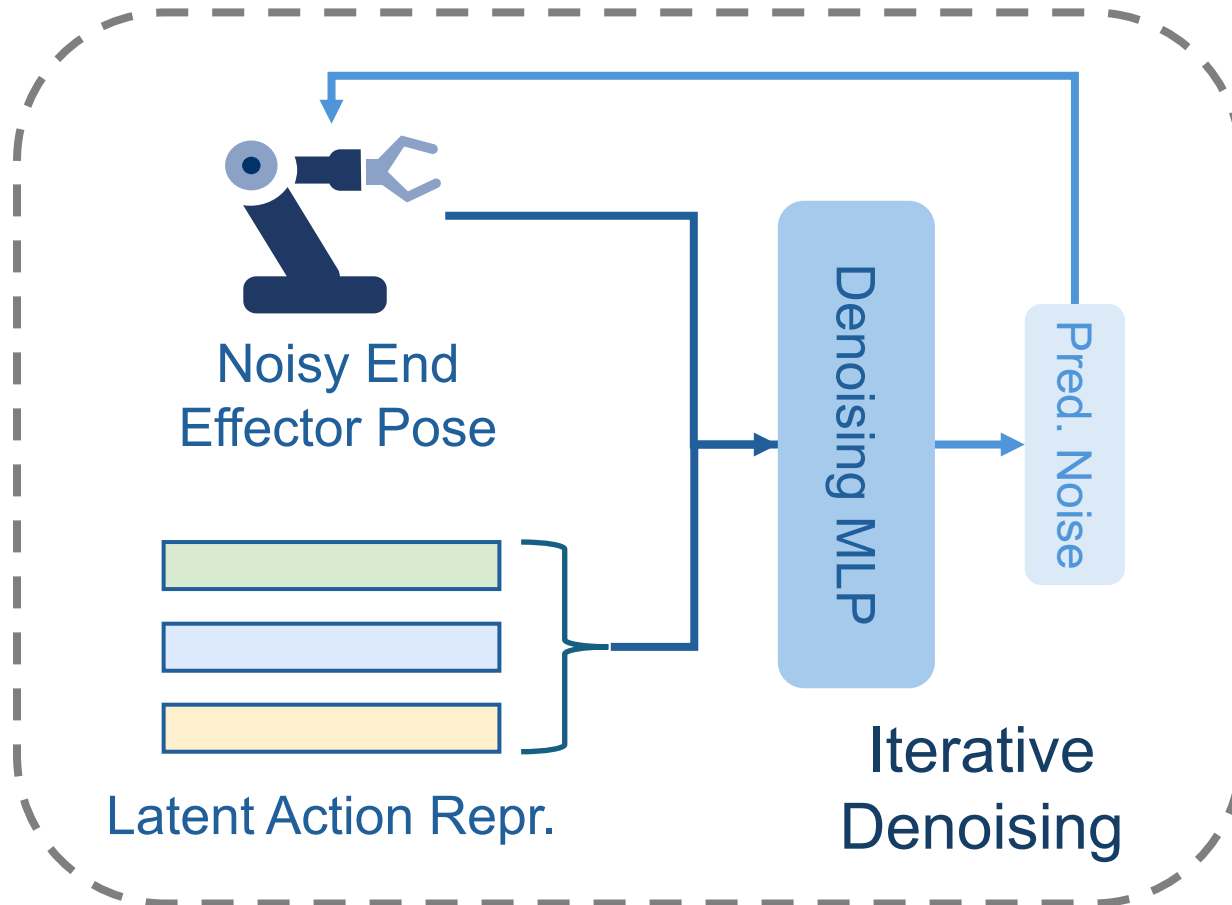
Our Approach

Learning actions via policy diffusion field.



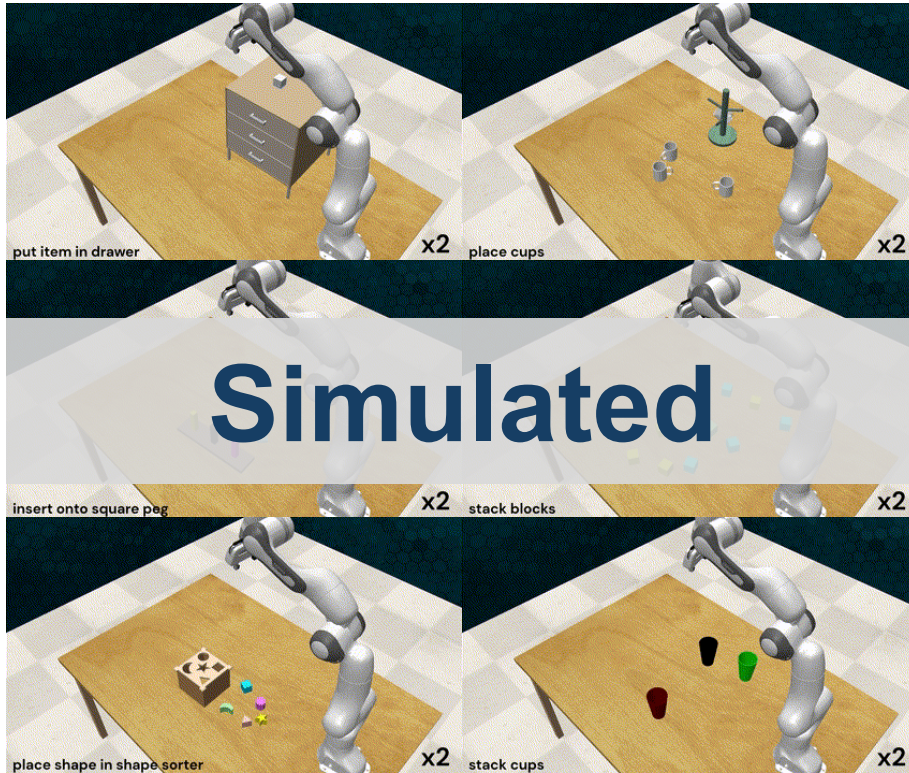
Our Approach

Learning actions via policy diffusion field.

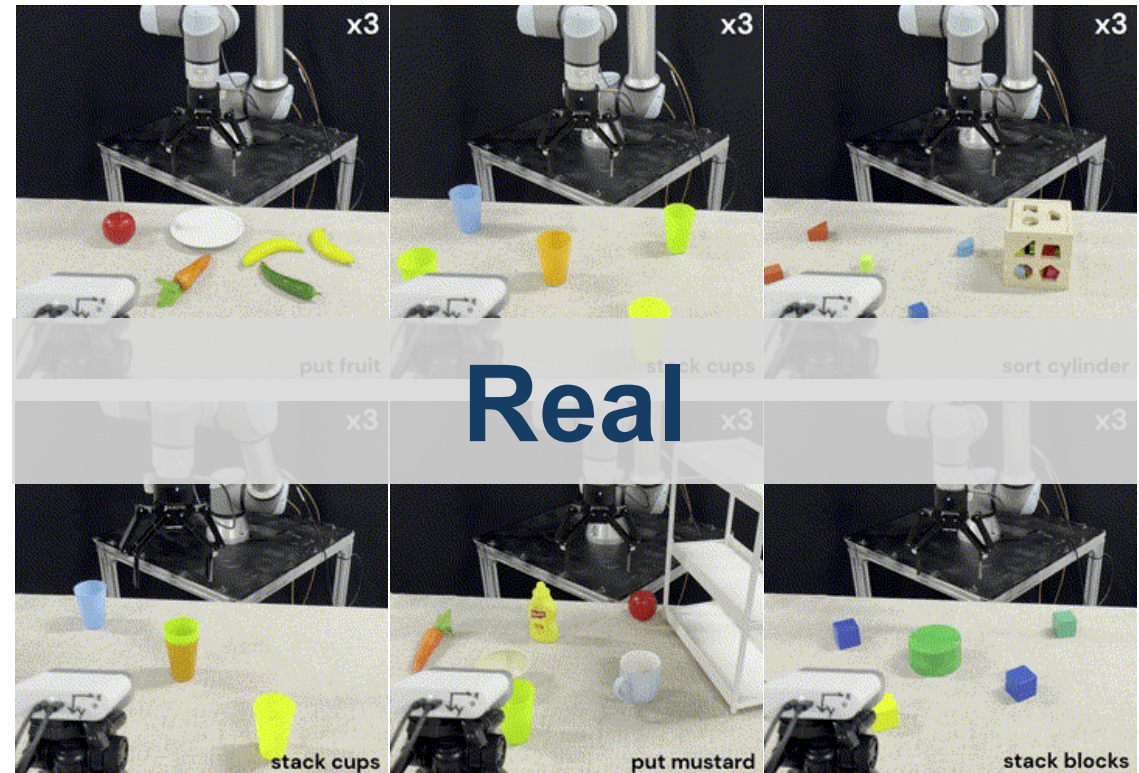


- Explicit alignment with perception space.
- Small denoising MLP, thus fast inference.
- Any bounded, continuous action distribution.

Evaluation

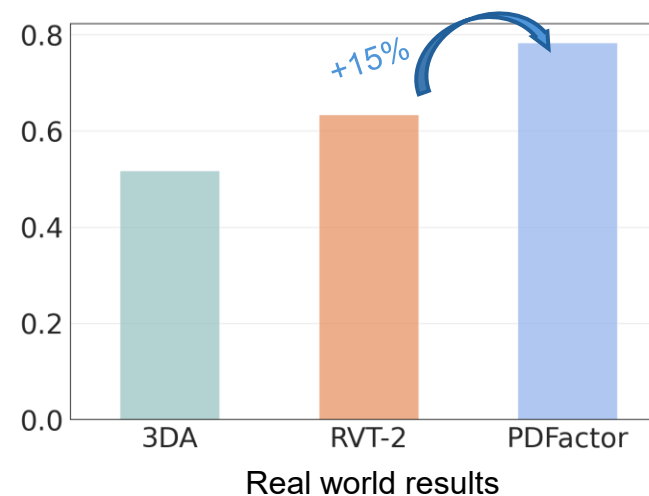
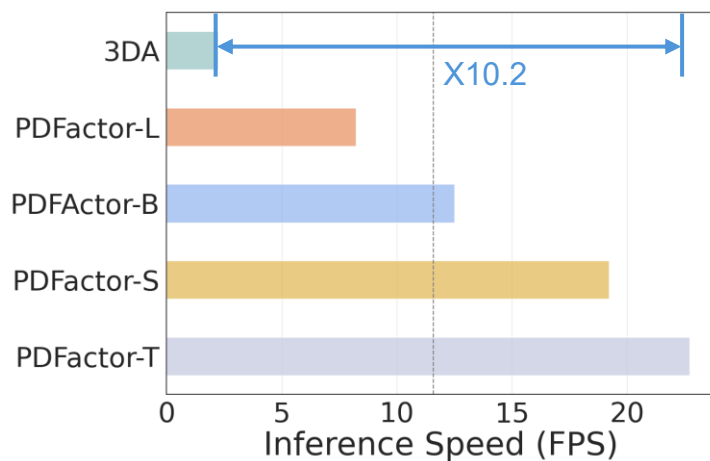
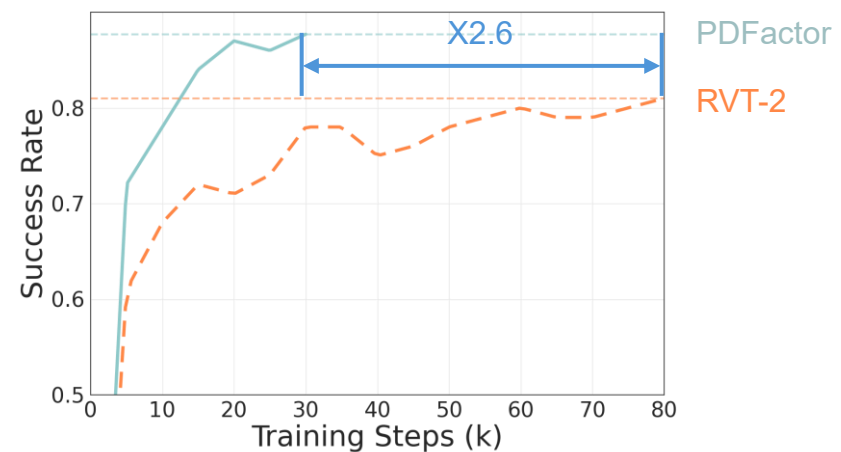
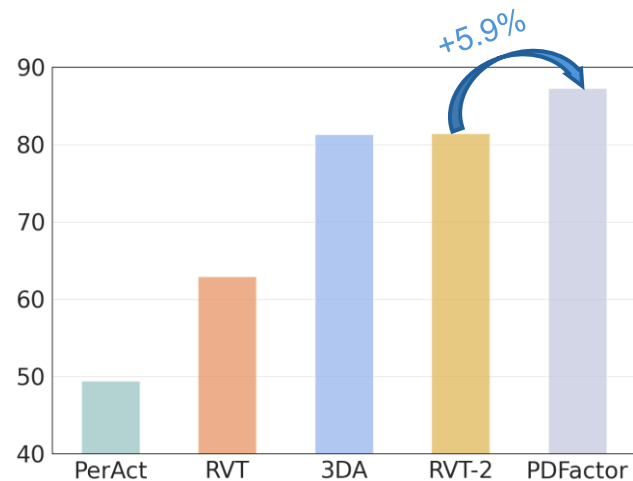


18 tasks



6 tasks

Quantitative Results



Summary

- A hybrid action representation to ground continuous and multimodal action directly in 3D perception space.
- Learning a 6-DoF end-effector pose distribution via policy diffusion field.

