(a) Manipulation Platform (6 dexterous manipulation tasks)
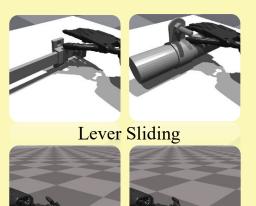
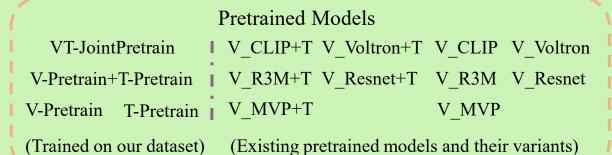BottleCap Turning

Faucet Screwing

Lever Sliding

Table Reorientation

In-hand Reorientation

Bimanual Hand-over

(b) Representation Module (18 benchmarking models)

Pretrained Models

VT-JointPretrain | V_CLIP+T V_Voltron+T V_CLIP V_Voltron

V-Pretrain+T-Pretrain | V_R3M+T V_Resnet+T V_R3M V_Resnet

V-Pretrain T-Pretrain | V_MVP+T   V_MVP

(Trained on our dataset) (Existing pretrained models and their variants)

Non-pretrained Models

V+T   V

T   Base

(c) Manipulation Policy Learning via PPO

Proprioception $\mathcal{P}$

I_{sim} OR C_{sim}

Representation Module $\mathcal{M}_\theta(\cdot)$

h

Manipulation Policy π_θ

Actions