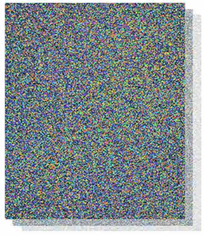
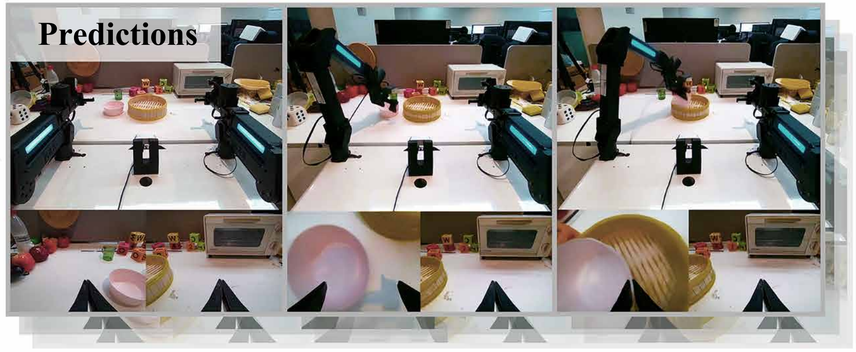
Vidar

Video Diffusion Foundation Model

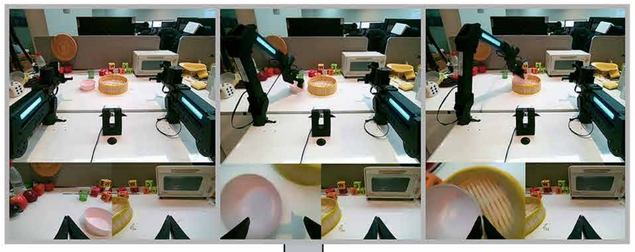


... a fixed rear camera, a movable left arm camera, and a movable right arm camera ... aloha robot ... gently grasp the pink bowl ... position it directly over the light-colored steamer ... lower it ...

Image | **Noise** | **Instruction**



TTS



MIDM

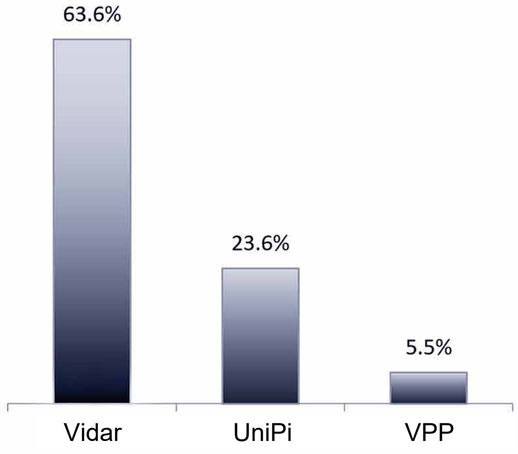
Action | **Action** | **Action**

Dataset

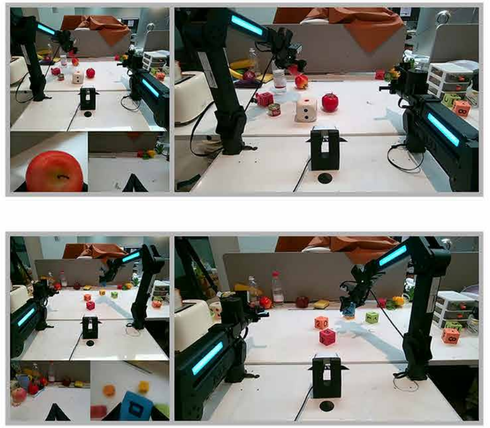
Multi-view Robotics Dataset

81 Tasks ~ 20min Vidar Dataset

Success Rates

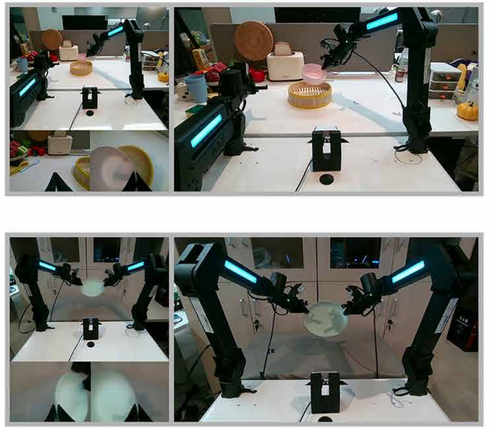


Semantic Tasks



Prediction **Execution**

Unseen Tasks



Prediction **Execution**