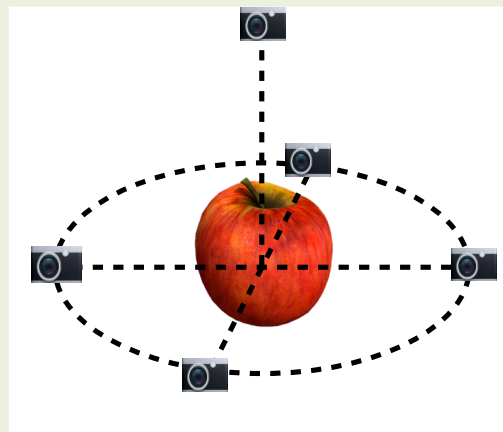
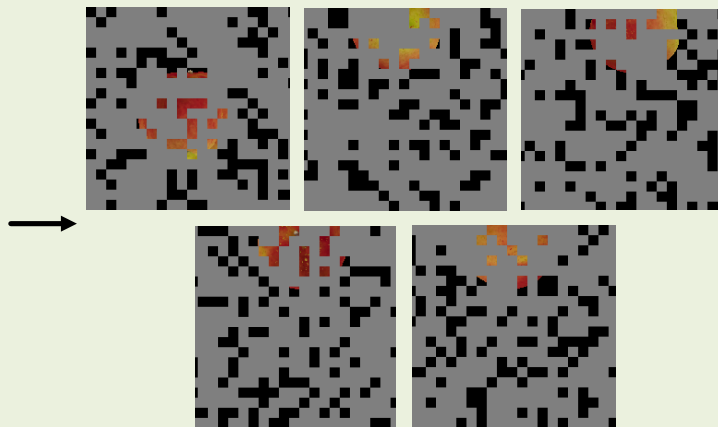


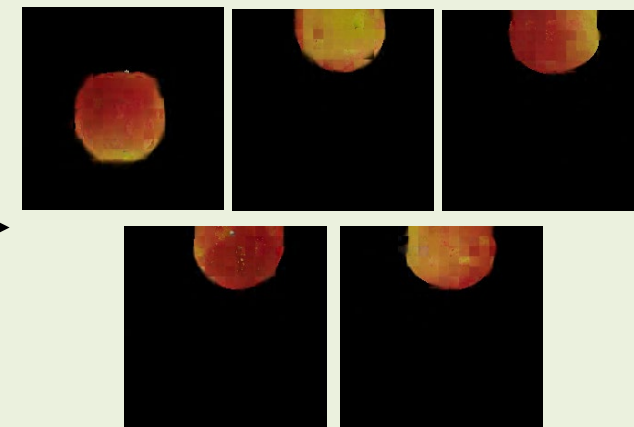
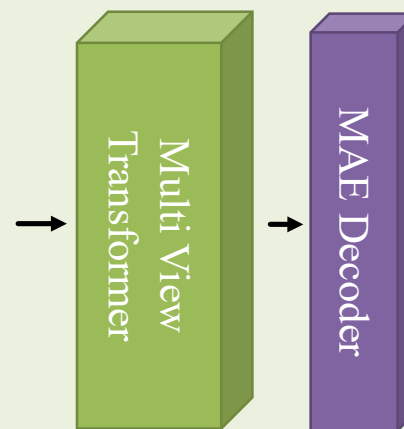
(a) Pretraining



Point Cloud w/ Virtual Cameras

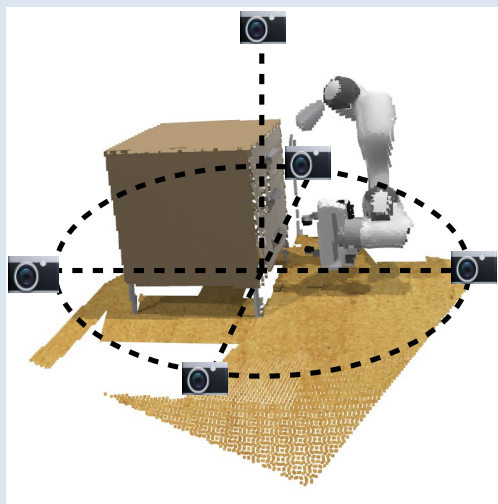


Masked Virtual RGB-D Images

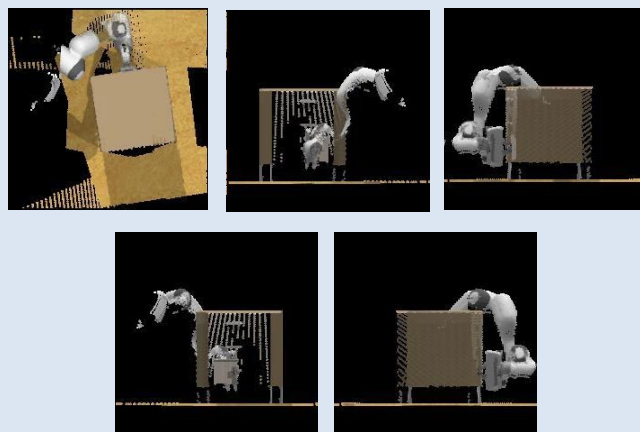


MAE Reconstruction

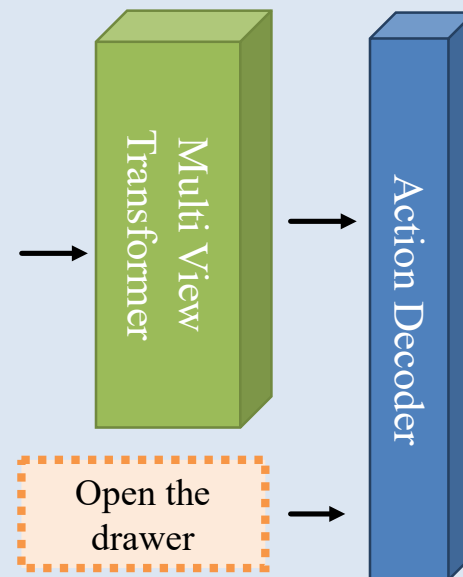
(b) Finetuning



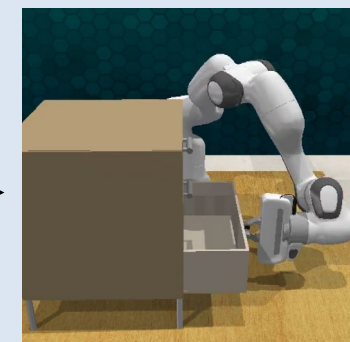
Point Cloud w/ Virtual Cameras



Multiview Virtual RGB-D Images



a_{pos}
 a_{rot}
 a_{open}
Action



Execution