

Robot Actions

Un-noised Human Actions

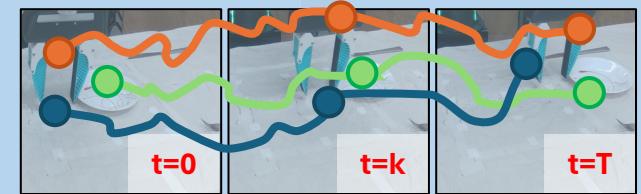
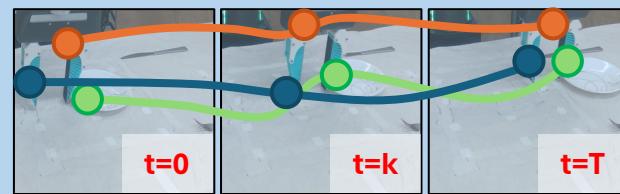
Noised Human Actions

Original trajectories

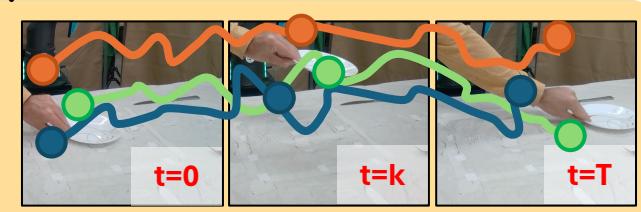
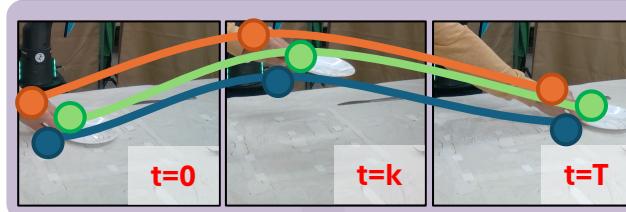
 $+\mathcal{N}(\mu, \sigma)$ 

Noisy trajectories

Robot



Human



Distinguishable across embodiment

Indistinguishable across embodiment

Data



Scalability

Bad 😞

Easy to scale 😊

Easy to scale 😊

Dynamics gap

No gap 😊

Dynamics mismatch 😞

Adaptive dynamics matching 😊

(a) Standard diffusion policy

(b) Naive co-training

(c) Our X-Diffusion