Ego2Robot Learning

Learning Paradigms

Explicit Embodiment Transfer

→ Robot-format data

- Hand Pose and HOI Reconstruction
- Motion Retargeting
- Ego2Robot Video Translation

Explicit Action and Task Supervision

→ Interpretable Supervision

- Explicit Intermediate Action Representations
- Affordance Representation
- Task-Progress and Reward Inference

Egocentric Video Representation Pretraining

→ Visual Representation

- Temporal Contrastive Learning
- Masked Video Modeling
- Language-Conditioned Pretraining

Latent Action and Dynamics Modeling

→ Predictive Methods

- Latent Action Learning
- World-Action Modeling

Training Data Landscape

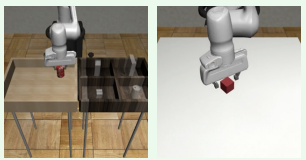
Egocentric Human Videos (EHV)



Robot Demonstration Data



Synthetic and Simulated Data



*Learning paradigms are complementary and can be combined.