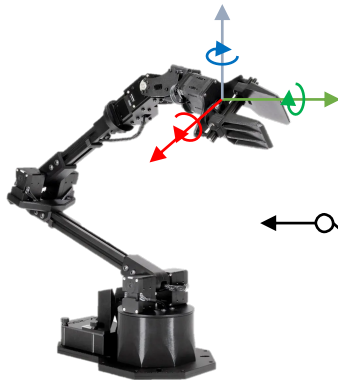
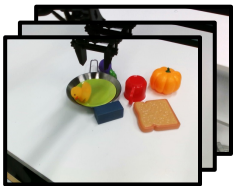


Unimodal Task Train

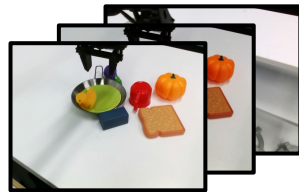
Visual Goal



Robot Policy

Multimodal Task Eval

Visual Goal



eval with seen modality

Textual Goal



- Pick up the duck
- Open the drawer
- Move the pot
- Fold the cloth

eval with **unseen** modality