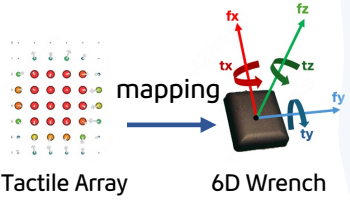

wrist camera


fisheye camera


global camera

Prompt



Stage I : Tactile-Aware VLA Training

VLM

Vision Token

Reconstruction Loss

[Action Chunk (Hx7) | Future Wrench (Hx12)]

Wrench Aware Action Expert

T&VL Fused Feature

T&VL-MoE

Exp

Ex

Ex

Expert M

Router Network

T&VL attention

Gated Bypass

Wrench

Tactile Sequence Token

Stage II : Stage-Specific Online Reference Refinement

Stage Estimator

Current Stage 2

Vision Token

Action

Wrench Context

AC1

Actor-Critic-2 (Selected)

Task-Critic

Ref

Goal

Action Progress

Reward propagation

Intervention-Critic

Action Progress

Human Intervention

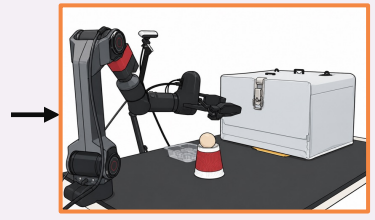
Actor π_θ

...

Refined Action

human-intervention

Replay Buffer



Stage III : Real-world Deployment & Evaluation

