

Spatiotemporal Constraint-Aware Assistance Motion Generation and Replanning for Human–Robot Collaborative Manipulation Tasks

Takuya Kiyokawa, Taichi Kato, Kensuke Harada



1 Background and Motivation



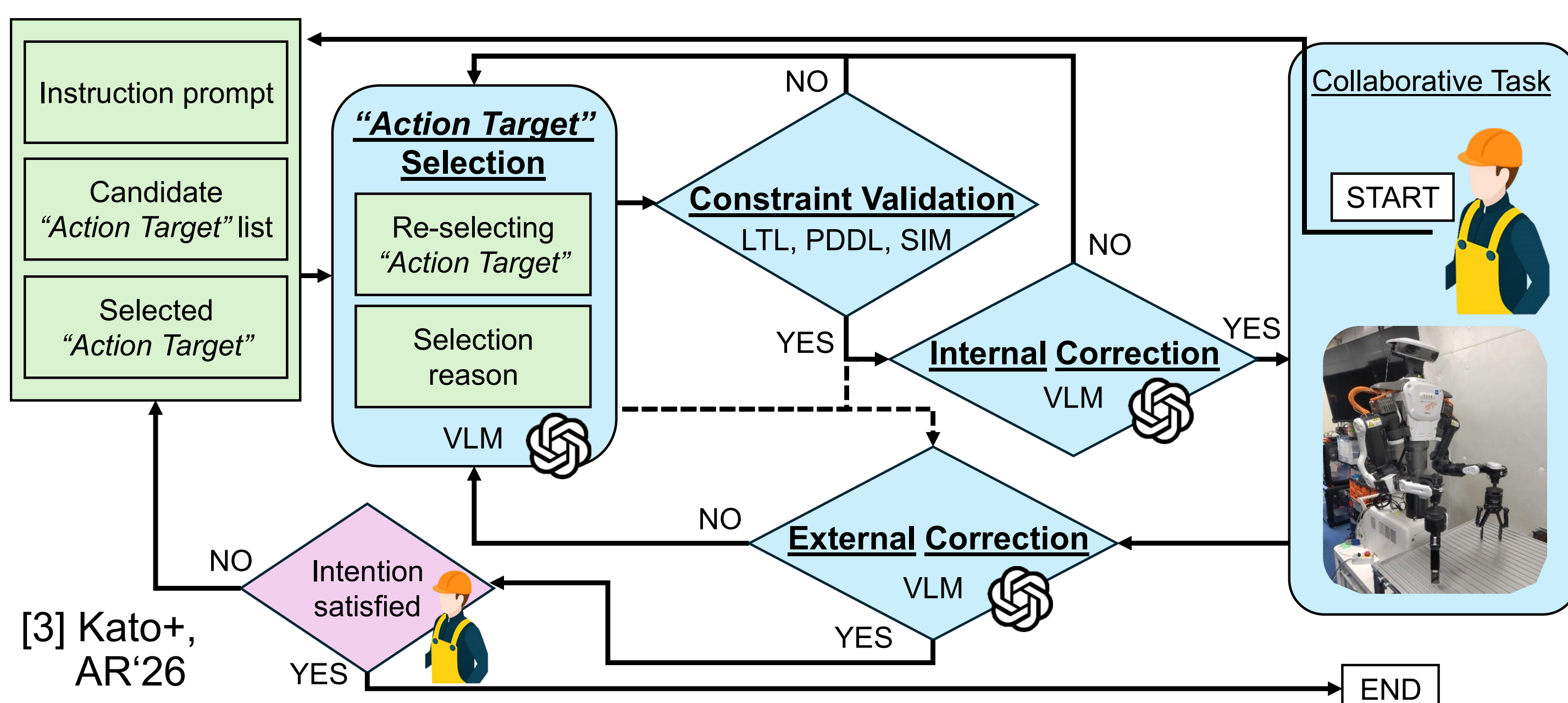
Challenges

- Ambiguous Instructions
- Spatiotemporal Correction
- Text-Only LLM Parsing

[1] Lim+, CASE'24 [2] Ji+, Sci. Rep.'24

Existing systems rely on humans to correct errors

2 Overview of Proposed Framework



[3] Kato+, AR'26

Infers intent from ambiguous inputs and checks spatial feasibility and temporal validity

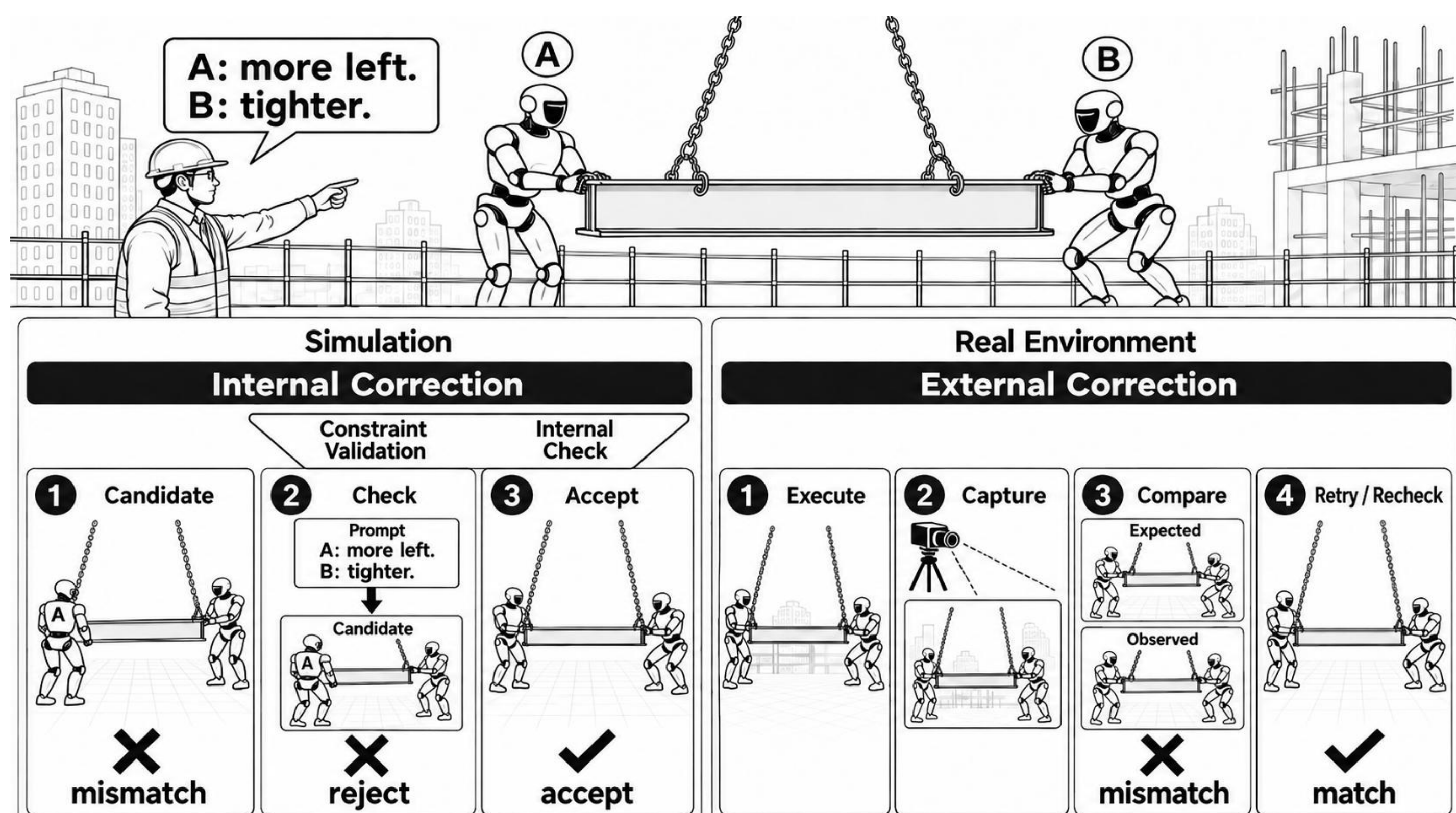
3 Filtering Before Intent Checks

Target Selection and Constraint Validation

- Select one grasp, tool, or part for the request
- LTL: task order; no grasp in the human work region
- PDDL: action preconditions, e.g., tool availability
- SIM: arm reachability via inverse kinematics.

This covers fixation, preparation, transport, and supply

Dual Correction with Constraint Validation



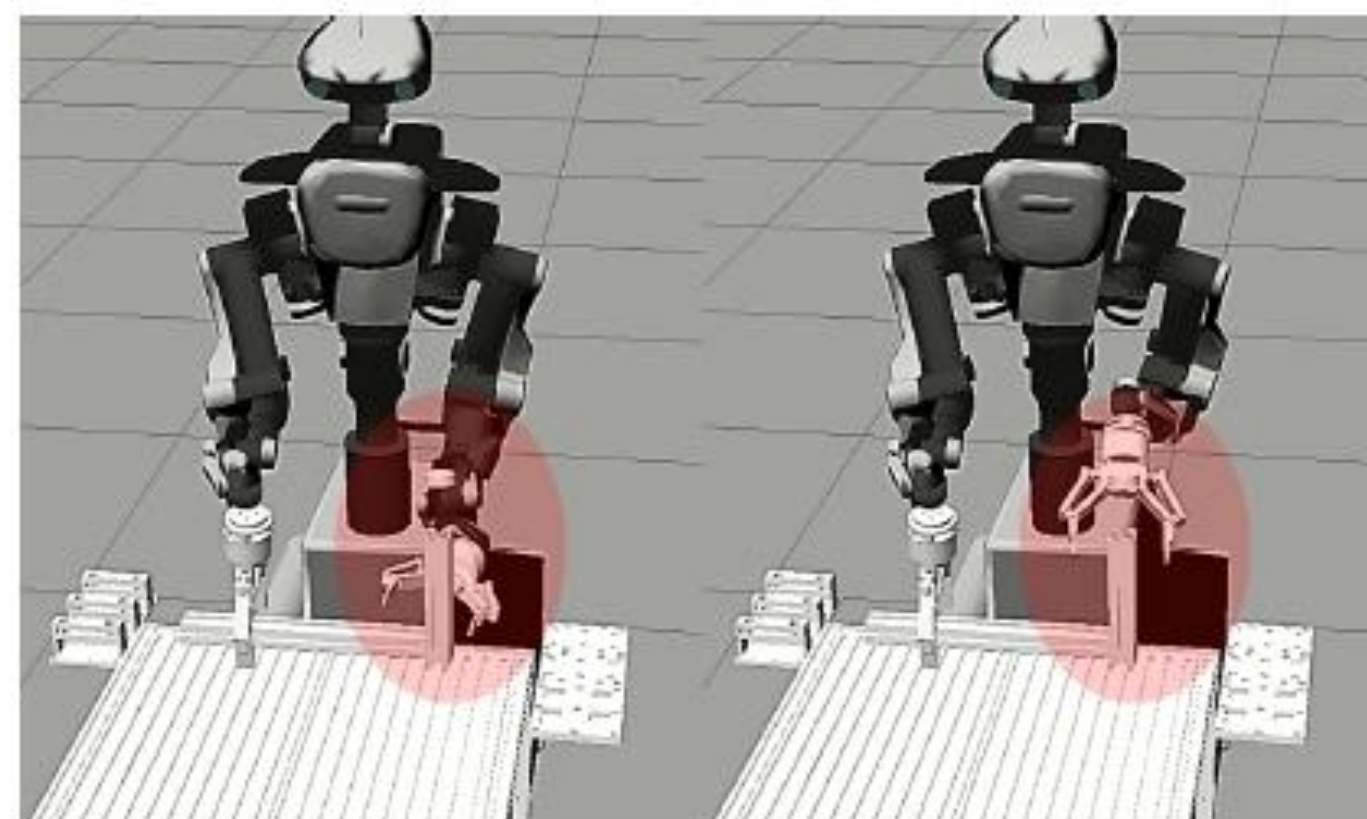
Validate constraints, then check intent internally and outcomes externally

4 Evaluating Motion Replanning

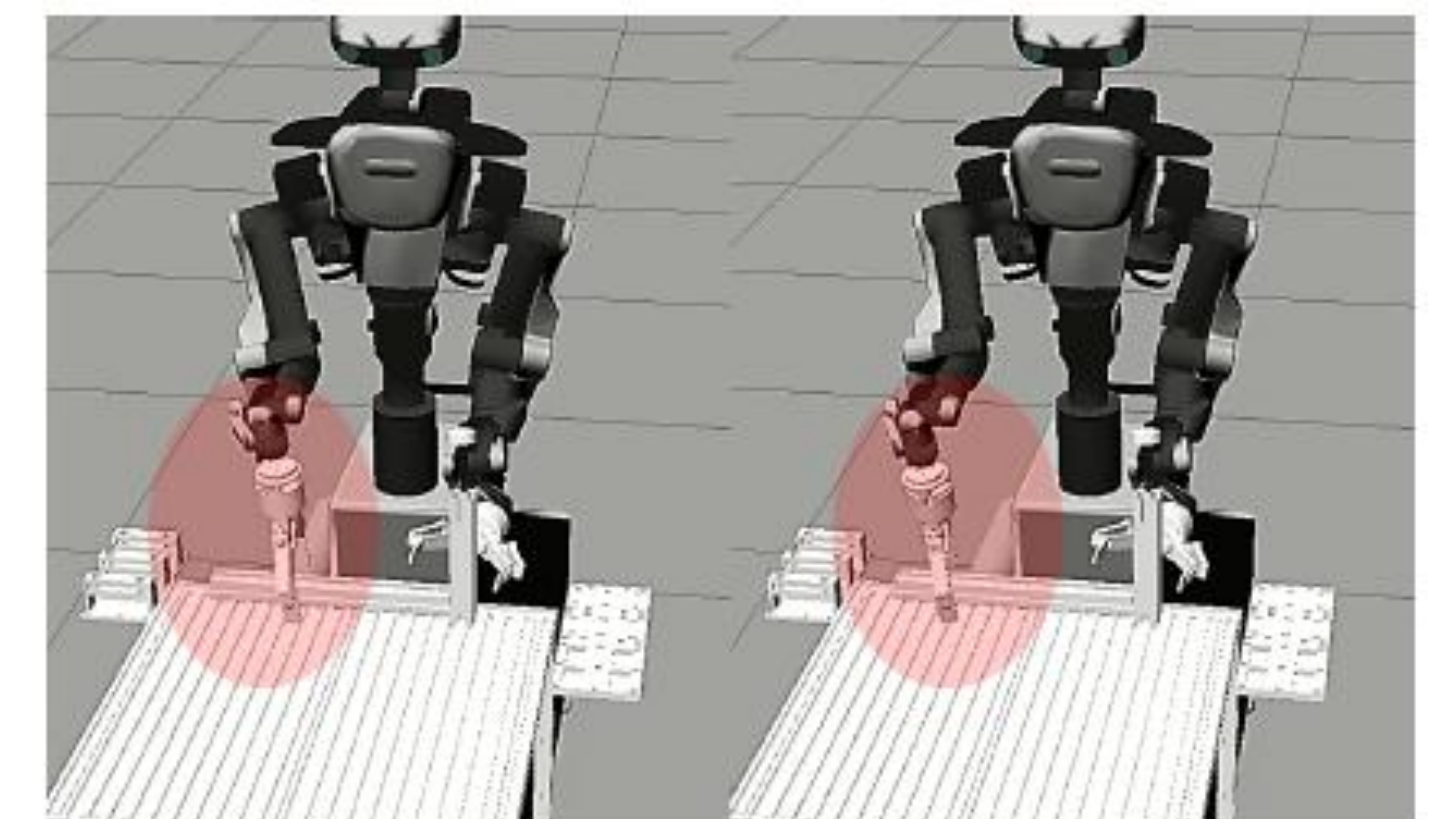
Experimental Setup

- Assistance Tasks: Object Fixation + Tool Preparation
- Collaborative Task: Aluminum-Frame Assembly
- Hardware: Humanoid + Grippers + RGB Camera
- Tools: Gazebo, GPT, Whisper, GraspGen

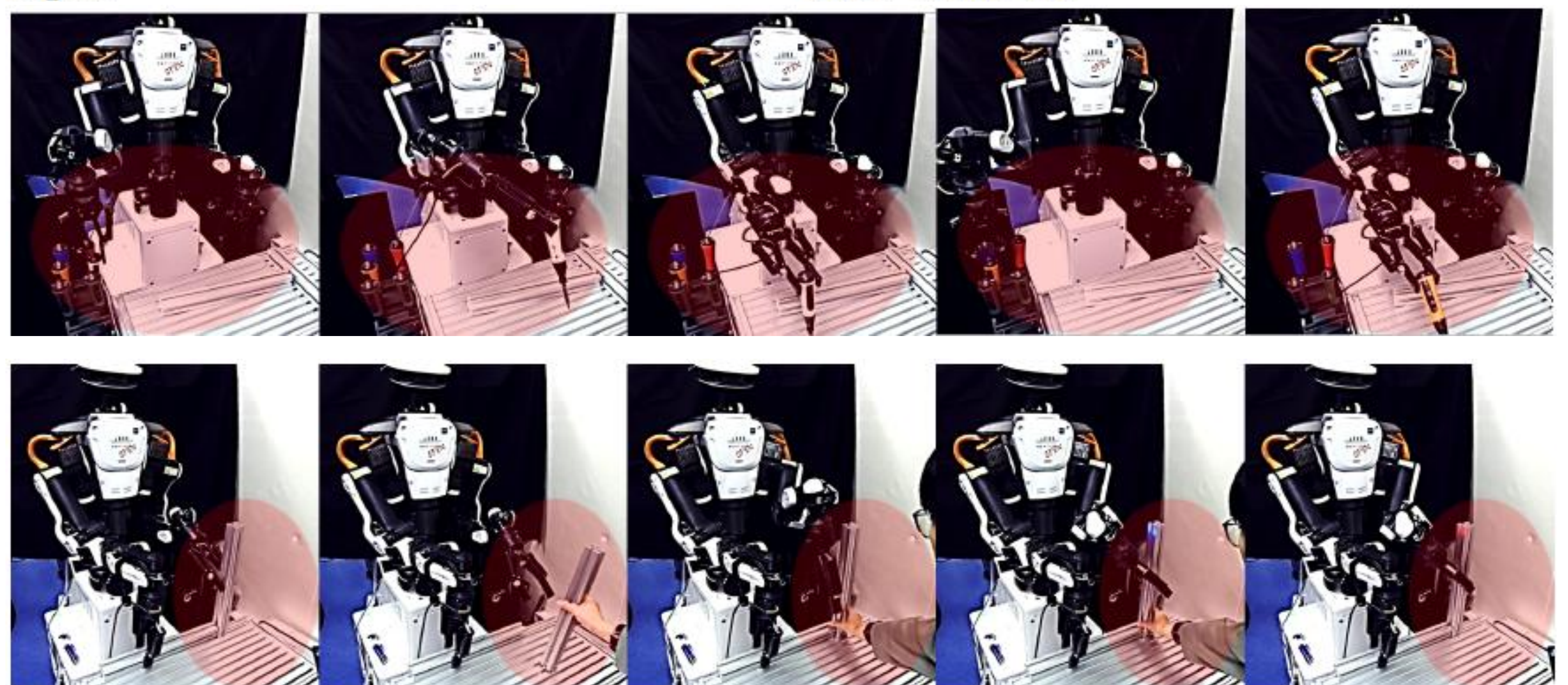
Sim / Real Task Success



(a) Left arm adjustment: "Raise the left a bit higher"



(b) Right arm adjustment: "Move a little more to the left"



Simulation Real-world

Task	Instruction Category	Success Rate	Success Rate
Object Fixation	Position adjustment (Left / Right)	20/20 (100%)	10/15 (66.7%)
Tool Preparation	Initial selection	10/10 (100%)	8/8 (100%)
Tool Preparation	Corrective selection (Replanning)	9/10 (90.0%)	6/8 (75.0%)

Real-robot failures involved the external check

VLM / Module Comparison for Object Fixation

Correction Modules	o4-mini		5.4-mini (medium)		5.4-mini (none)	
	Task Succ.	Sel. Succ.	Task Succ.	Sel. Succ.	Task Succ.	Sel. Succ.
w/o Both (Baseline)	94%	86%	99%	97%	87%	85%
w/o Internal	61%	-	100%	-	83%	-
w/o External	100%	100%	100%	100%	89%	87%
Full (Internal + External)	72%	100%	100%	100%	94%	87%
Avg. Inference Time (Selection)		17.7 s	19.5 s		3.59 s	

It should be selected based on use-case priorities

+ Spatiotemporal Constraint Validation

16 invalid-first cases per task: semantic, temporal, symbolic, physical

Task (5.6-terra)	Object Fixation		Tool Preparation	
	Valid Selection	Total Time	Valid Selection	Total Time
Baseline	2/16 (12.5%)	9.49 s	4/16 (25%)	4.82 s
All w/o Dual Correction	12/16 (75%)	4.33 s	12/16 (75%)	3.36 s
All w/o LTL	10/16 (62.5%)	9.23 s	12/16 (75%)	5.66 s
All w/o PDDL	10/16 (62.5%)	9.45 s	12/16 (75%)	5.37 s
All w/o SIM	10/16 (62.5%)	10.68 s	12/16 (75%)	5.65 s
All (Proposed)	14/16 (87.5%)	9.97 s	16/16 (100%)	5.96 s

Higher success with time comparable to baseline

5 Conclusions and Future Work

- Proposed framework enabled sim/real assistance
- Ours achieved highest success rates of valid selection
- Future: real-robot and humanoid construction trials

