CANVAS: Commonsense-Aware Navigation System for Intuitive Human-Robot Interaction





Suhwan Choi¹, Yongjun Cho¹, Minchan Kim¹, Jaeyoon Jung¹, Myunchul Joe¹, Park Yu Been¹, Minseo Kim², Sungwoong Kim², Sungjae Lee², Whiseong Park¹, Jiwan Chung², Youngjae Yu²

¹MAUM Al Inc, ²Yonsei University

HUMANS OFTEN NAVIGATE USING ABSTRACT YET SIMPLE INSTRUCTIONS. HOW CAN WE ENABLE ROBOTS TO DO THE SAME?

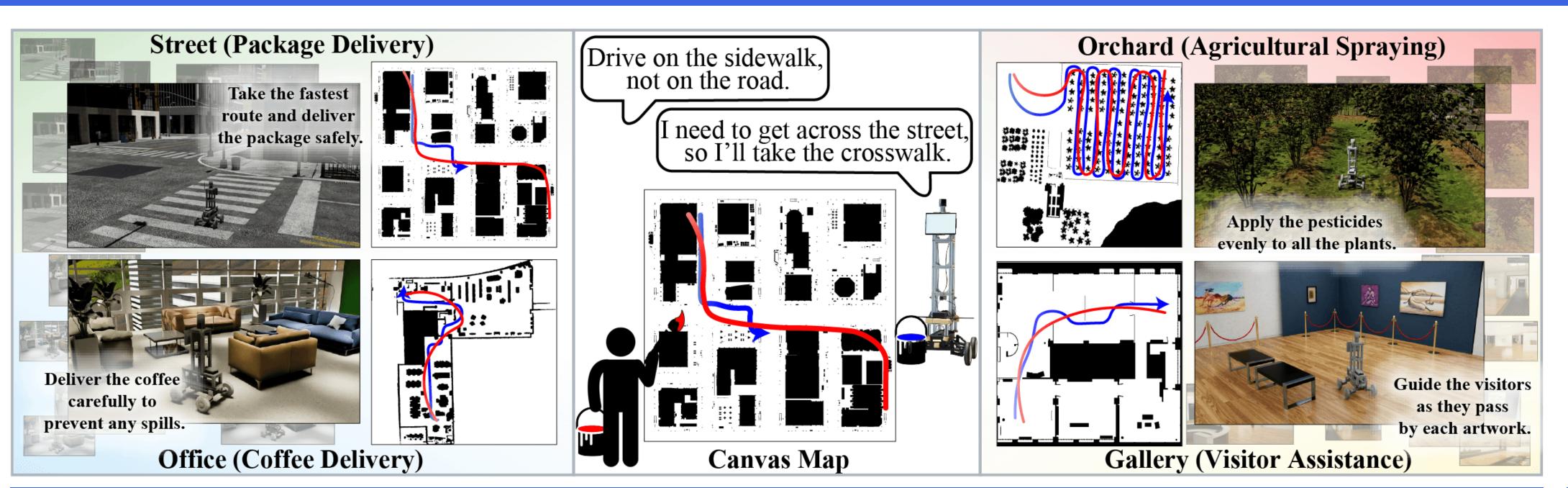


Image X

Sensitive to environmental changes

: Language X

Limited for detailed navigation

✓ Sketch ✓

Front View $X_f(t)$

Canvas Map $X_c(t)$

Language Instruction L

You must drive on the sidewalk

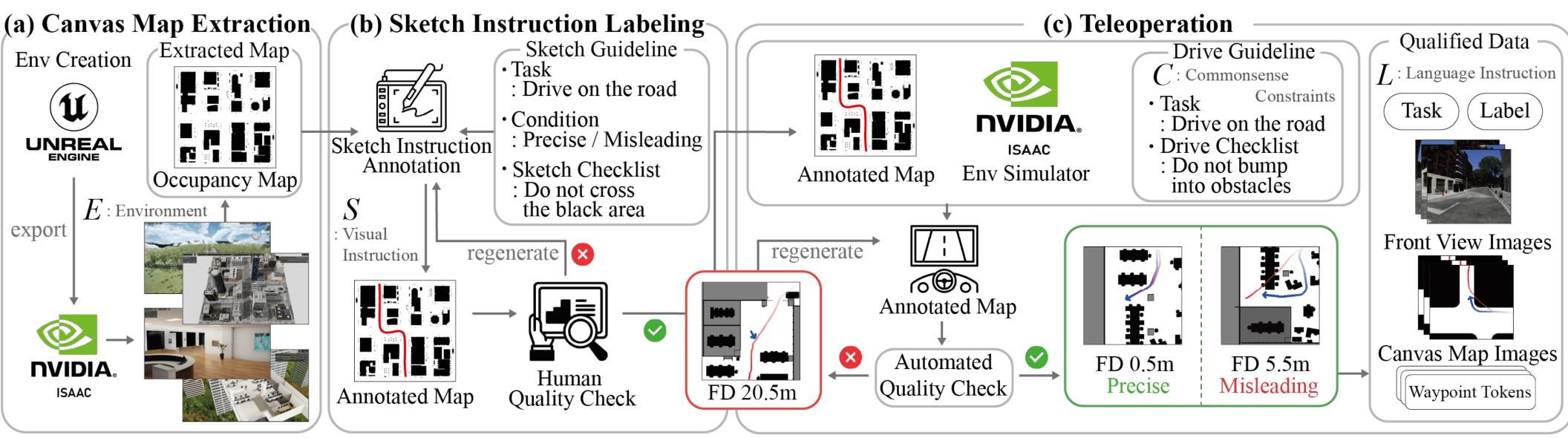
Robust & contains enough detail

But how can we handle with noisy and imprecise instructions?

CANVAS model

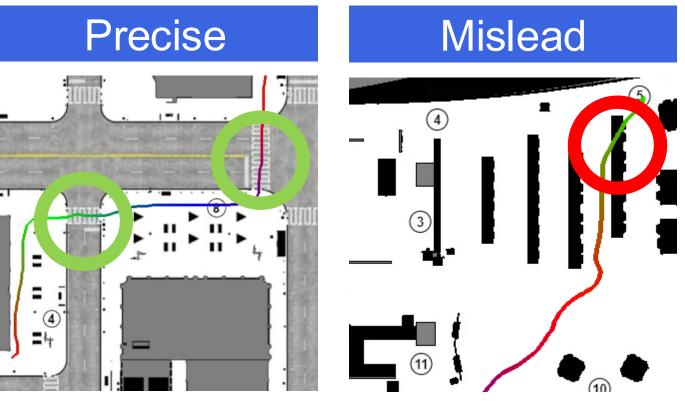
LLM





COMMAND dataset enables robot to use commonsense understanding to transform abstract or noisy human instructions into optimal trajectories.

- All Annotated map collected by human, spanning over 48 hours and 219 km Data worker draw sketch trajectories and teleoperate the robot.
- Precise trajectory is the most efficient route while misleading trajectory includes noises like as passing through a wall or obstacle.
- Use Fréchet distance to check records automatically.
- Fréchet distance is distance between curves used for inspection and labeling.



CANVAS is built upon VLM architecture to achieve common-sense aware.

- Language Instructions are verbal commands which help agent's understanding.
- Canvas Map is an occupancy map extracted from simulated environments.
- Sketch Instruction is trajectory drawn on canvas map.



Street - Sidewalk You are an outdoor last mile delivery robot.

You must follow these driving instructions: 1. You must avoid collisions. 5. You must drive

on the sidewalk. 5.a. **If you need to** cross the road, you must use the crosswalk.

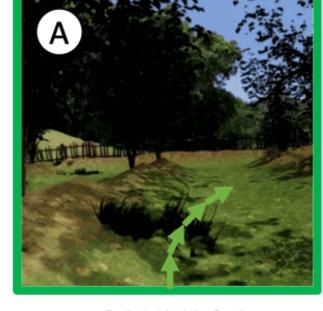


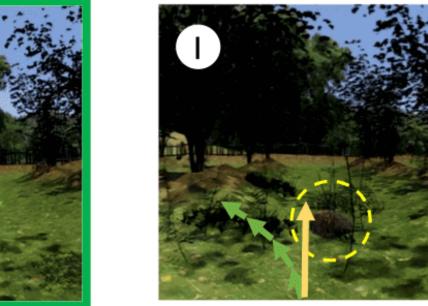
CANVAS-L



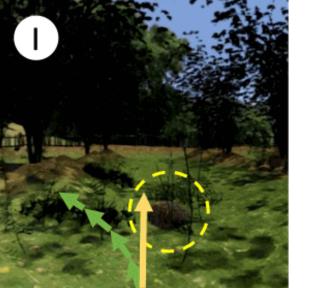
Human Sketch Instruction







CANVAS-S



CANVAS-L

speed-sprayer robot. You must follow

Language Instruction

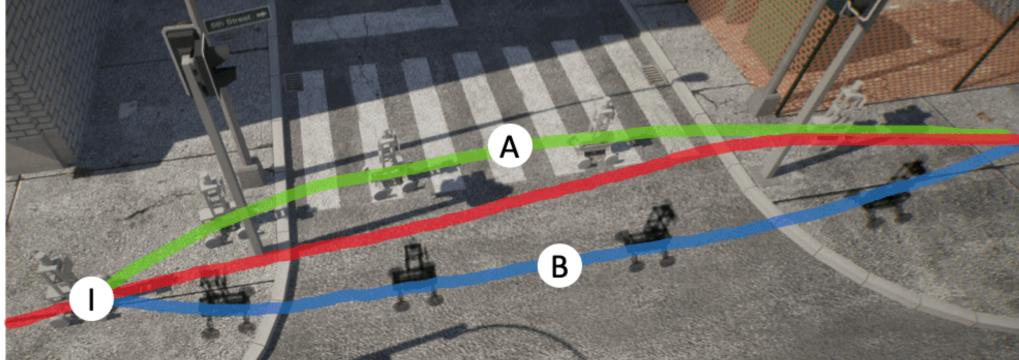
Orchard

You are an outdoor

Waypoints

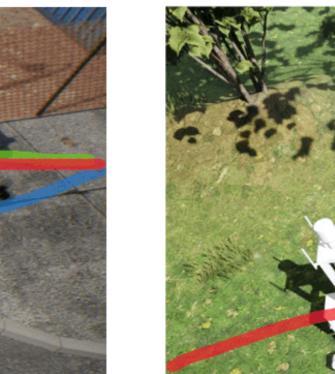
these driving instructions:

3.a. **If the** Trajectory **Instruction cannot** be followed due to any obstacles, you should deviate to bypass the obstacle.

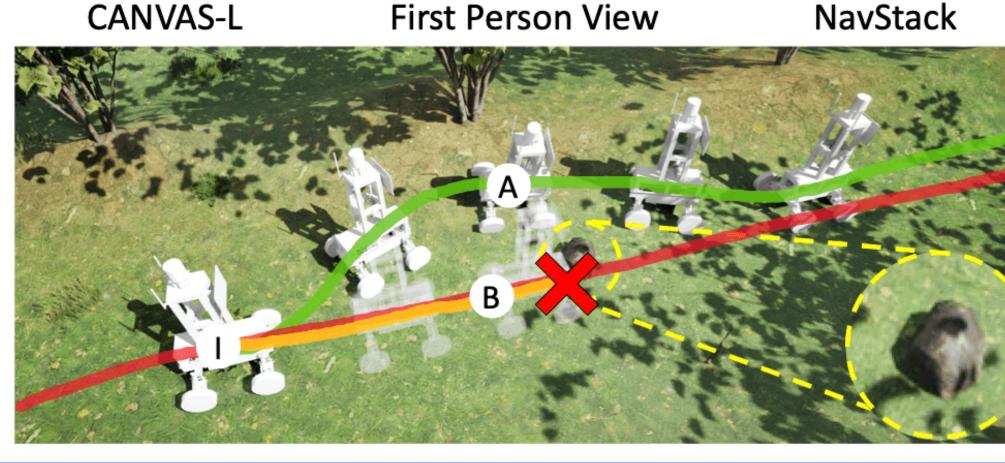


0.856m

67%



NavStack



Regulte on simulated environments

	Re	Suits	on si	mulate	ea en	vironr	nents	
	Method	Precise			Misleading			Total
	Method	SR(↑)	$CR(\downarrow)$	$TDD(\downarrow)$	SR(↑)	$CR(\downarrow)$	$TDD(\downarrow)$	SR(↑)
			S	Seen Environment				
	Office							
	NavStack	87%	13%	0.846m	0%*	100%*	-	-
	CANVAS-S	100%	0 %	0.730m	87%	13%	0.843m	93%
	CANVAS-L	_100 % _	0%	0.802m	100%	0%	0.753m	100%
				Street (Re	oad)			
	NavStack	100%	0 %	1.654m	0%*	100%*	-	-
	CANVAS-S	100%	0 %	1.189m	100%	0 %	1.075m	100%
	CANVAS-L	97%	3%	1.117m	97%	3%	1.236m	97%
	Street (Sidewalk)							
	NavStack	53%	53%	1.450m	0%*	100%*	-	-
	CANVAS-S	60%	40%	1.451m	47%	53%	2.379m	54%
	CANVAS-L	87%	13%	1.394m	53%	47%	1.839 m	70%
Orchard								
	NavStack	0%	87%	-	0%*	100%*	-	-
	CANVAS-S	73 %	60%	1.561m	60%	33 %	1.448m	67%
	CANVAS-L	67%	47 %	1.759m	60%	53%	1.392m	64%
Unseen Environment								
Gallery								
	NavStack	100%	0 %	0.783m	0%*	100%*	-	-
	CANVAS-S	87%	13%	0.773m	33 %	66 %	0.938m	60%

 Outperforms NavStack in environment with many obstacles In orchard, ROS NavStack records 0%, CANVAS achieves 69%

33%

 $0.9 \mathrm{m}$

Enable to navigate unseen environments

CANVAS-L | 100%

Effectively handles misleading scenarios!

Violation rates for commonsense-constraints

Environment	Method	Precise IVR(↓)	Misleading $IVR(\downarrow)$	Instruction Violation Type Right hand traffic (Road)
Street (Road)	NavStack CANVAS-S	7% 0%	100%* 7%	Drive on sidewalk and crosswalk
Succi (Road)	CANVAS-S CANVAS-L	17%	30%	• CANVAS know commonsense rules!
Street (Sidewalk)	NavStack CANVAS-S CANVAS-L	7% 0% 0%	100%* 26% 13%	$IVR = \frac{N_{Instruction\ Violation}}{N}$

		Resu	al environments (Office)			
Method	Precise SR(†)	Misleading SR(†)	Total SR(†)	HESAI QT128 Crossover Ultra-Wide View 16ATF5DEXIPS		
NavStack CANVAS-S	$-\frac{100\%}{77\%}$ -	$-\frac{0\%^*}{60\%}$	-69%	Ultra-Wide View Short-Range LiDAR Multi-touch Monitor		
CANVAS-L	93%	33%	63%	NVIDIA Jetson AGX Orin e-con Systems		
 Trained solely on simulated data Strong Sim2Real transfer capabilities Strong Sim2Real transfer capabilities						

SLAM with FAST-LIO2 was used to find the robot's current position

Ablation study on the effect of VLM pre-training							
Environment	Method	Precise SR(†)	Misleading SR(↑)	Total SR(†)			
Seen - Office	CANVAS-L	100%	100%	100%			
Seen Since	w/o Pre-training	100%	87%	93%			
Unseen - Gallery	CANVAS-L	100%	33%	67%			
Offseen - Ganery	w/o Pre-training	60%	40%	50%			
Real - Office	CANVAS-L	93%	33%	63%			
Real - Office	w/o Pre-training	73%	33%	53%			



Sim Office

Real Office

Project Page

Converter