Environmental Understanding Vision-Language Model for Embodied Agent

Jinsik Bang, Jaeyeon Bae, Donggyu Lee, Siyeol Jung, Taehwan Kim
Artificial Intelligence Graduate School, UNIST

IMML Lab
Interactive Multimodal Machine Learning Lab

CVPR JUNE 3-7, 2026 DENVER COLORADO

Background & Motivation

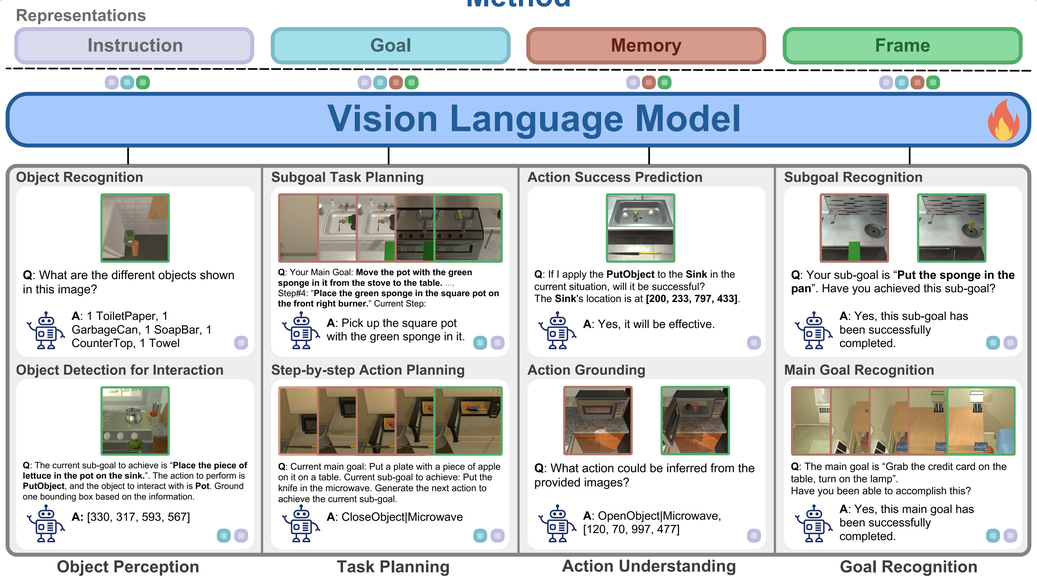
- **Environmental Understanding**: Embodied agents require environmental understanding to perceive, interpret, and interact with their surroundings.
- Embodied agents face key limitations: 1) **modular pipelines**, 2) **metadata-dependent** model, and 3) **limited explicit environmental interpretation**.

Can a single unified VLM achieve end-to-end decision-making through environmental understanding?

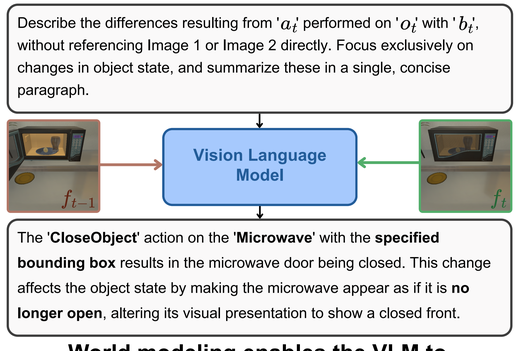
Contributions

- **EUEA Framework**: We **integrate four core skills into a single VLM** for end-to-end environmental understanding.
- **Recovery Step & GRPO Refinement**: Sampling-based recovery **(no extra training)** and GRPO refinement of inconsistent skill predictions yield **additional performance gains.**
- **Skill Dataset & Benchmark**: We build **1.24M (ALFRED) and 3.7M (LangR) skill samples** and an evaluation benchmark.

Method



We augment EUEA skill data from trajectories of ALFRED and LangR, and *Random Exploration Pipeline* curates action failure data.



World modeling enables the VLM to anticipate environmental changes.

Recovery Step

Samples alternative actions via the **Step-by-step Action Planning (*SAP*)** skill when an action fails.

Score $s_{t,i} = -\log \pi_\theta(a_{t,i}, o_{t,i} \mid I_{SAP}, m_{t-k:t})$

π_θ: VLM Policy
$a_{t,i}, o_{t,i}$: i-th generated action and object at t step
I: Instruction for skill
m: memory information

GRPO Refinement Stage

Sampling-based RL method that refines inconsistent predictions across **all skills** (except FSC) through rule-based reward functions.

Reward function $R_{total} = R_{OP} + R_{TP} + R_{AU} + R_{GR}$

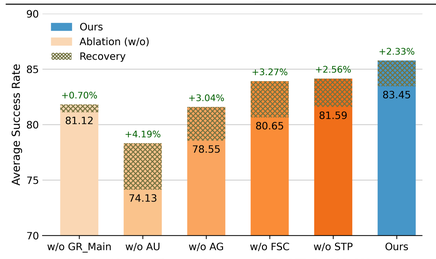
OP: Object Perception
TP: Task Planning
AU: Action Understanding
GR: Goal Recognition

Experiments

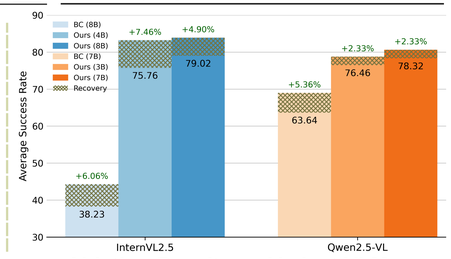
Task Evaluation

VLM Agent	Task Success Rate						
	Avg.	Look	Pick	Pick Two	Clean	Cool	Heat
EMMA* [43]	67.83	66.67	71.95	75.93	65.31	55.56	71.80
Human Performance* [31]	91.00	-	-	-	-	-	-
BC (InternVL3-8B)	74.59	88.89	73.17	57.41	62.24	96.83	75.64
Ours (SFT)	83.45	**90.74**	**86.59**	75.93	65.31	**98.41**	**91.03**
Ours (GRPO)	**85.78**	**90.74**	85.37	**85.19**	**74.49**	**98.41**	87.18

Recovery Method	Success Rate		Goal Condition	
	SFT	GRPO	SFT	GRPO
Ours	83.45	85.78	88.42	90.17
w/ Env feedback	**85.78**	85.78	**90.09**	90.17
w/ Recovery Step	**85.78**	**86.48**	89.74	**90.48**



Ablation Results on EUEA Skills



Ablation Results on Variant VLMs

Skill Evaluation

Model	Object Perception		Goal Recognition		Action Understanding		Task Planning	
	Grounding	Detection	Main	Sub	Prediction	Grounding	Planning	Step-by-step
GPT-5	51.45	24.34	92.20	73.80	81.48	28.86	0.801	71.52
GPT-o3	57.28	29.55	94.80	80.60	86.75	31.76	0.796	77.41
Claude-4.5-Sonnet	38.05	1.54	90.80	65.80	51.60	62.55	0.812	46.15
Gemini-2.5-Pro	63.53	60.75	86.20	80.20	85.43	31.08	0.819	85.76
LLaVA-OneVision-7B [22]	22.19	18.19	69.60	68.00	71.05	44.88	0.695	6.87
InternVL2.5-8B [8]	30.90	8.79	67.00	76.20	68.70	47.39	0.608	0.82
InternVL3-8B [53]	33.70	26.68	71.80	77.40	68.80	48.26	0.742	4.42
Qwen2.5-VL-7B [4]	28.15	1.82	84.00	73.00	48.78	49.61	0.764	39.44
BC (InternVL3-8B)	**78.01**	73.94	71.80	83.00	63.25	9.27	0.630	**98.53**
Ours (InternVL3-8B)	75.84	**81.73**	**99.40**	**98.60**	**96.80**	**89.09**	**0.894**	98.20

Existing VLMs still remain limited in environmental understanding.

Conclusion

- Our EUEA framework **effectively improves task performance,** while our skill evaluation **reveals limited environmental understanding** in existing VLMs.
- **Future Work**: Future directions include extending EUEA to continuous environments and real-world VLAs, expanding the skill set, and integrating explicit reasoning over memory and future states.

Project Page