Efficient Policy Adaptation with Contrastive Prompt Ensemble for Embodied Agents

Wonje Choi, Wookyung Kim, Seunghyun Kim, Honguk Woo

Department of Computer Science and Engineering, Sungkyunkwan University

wjchoi1995, kwk2696, kimsh571, hwoo@skku.edu

Overall approach

1. Prompt-based Contrastive Learning

- For domain-invariant representations with respect to a specific domain factor
- Propose a prompt-based optimization technique using a few expert data across various domain factors

2. Guided-Attention-based Prompt Ensemble

- To generate robust state representations for complex domain variations
- Propose a cosine similarity-guided attention-based prompt ensemble architecture

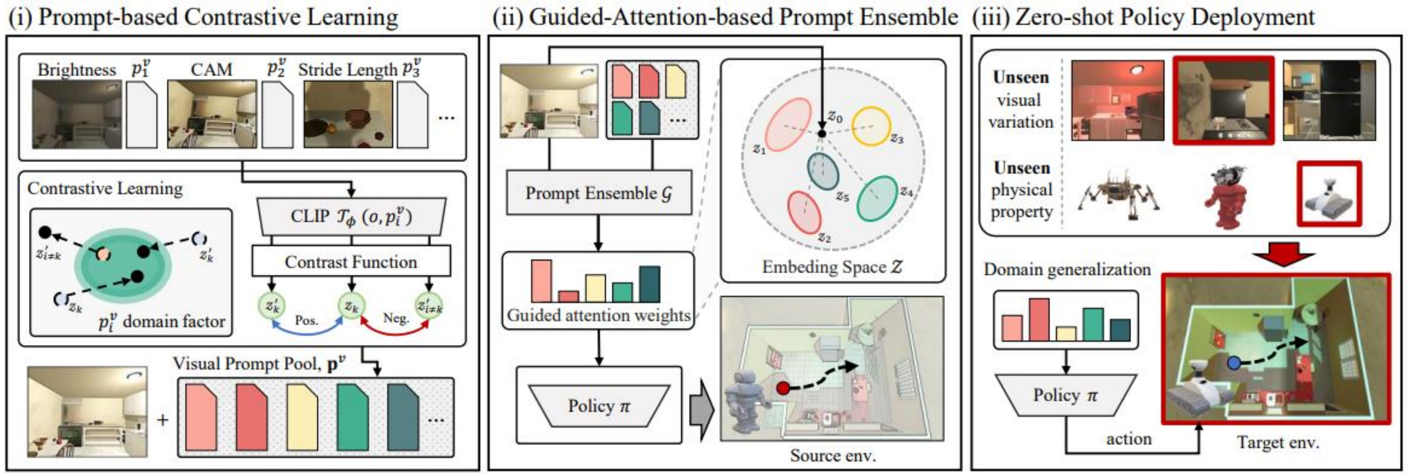

Figure 1. CONPE Framework

Prompt-based Contrastive Learning

- Objective

  - To construct domain-invariant representations for a specific domain factor

- Procedure

1. Adopt several contrastive tasks for visual prompt learning
2. Conduct prompt-based contrastive learning on different domain factors
3. Each visual prompt encapsulates domain-invariant knowledge for a specific domain factor
4. Obtain a visual prompt pool

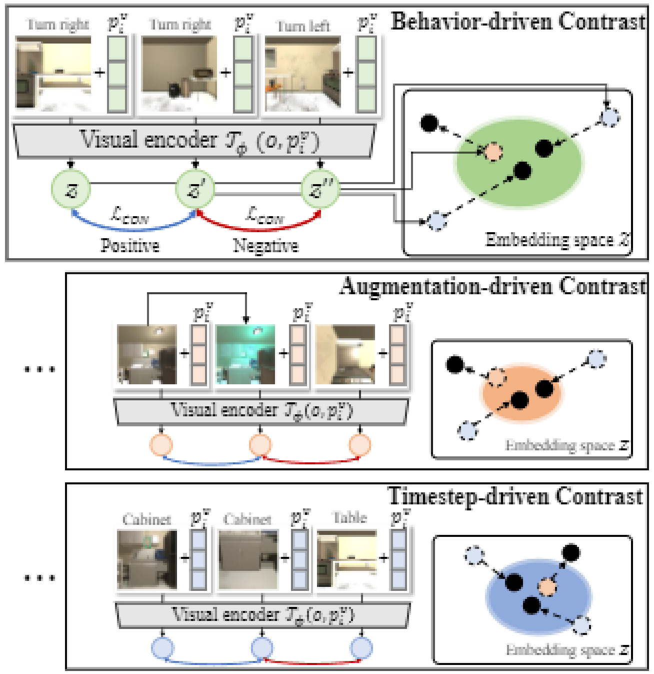

Figure 2. Prompt-based Contrastive Learning with Different Contrastive Tasks

Guided-Attention-based Prompt Ensemble

- Objective

- To integrate individual prompted embeddings into a task-specific state representation

- Procedure

1. Compute guidance score based on the cosine similarity between the observation and visual prompted embeddings
2. Train jointly attention module and policy for the robust state representations

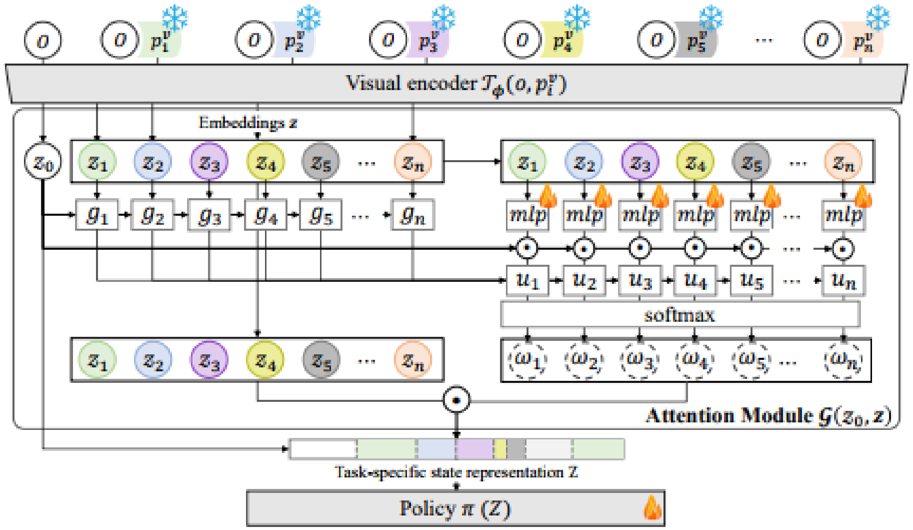

Figure 3. Guided-Attention-based Prompt Ensemble

Evaluation: Simulation Environment

- AI2-THOR
- Evaluate for Visual Navigation Tasks
- Use camera fields of view, stride length, rotation degrees, and illumination as domain factors
- egocentric-Metaworld
- Evaluate for Reaching Tasks
- Use camera positions, gravity, wind speeds, and illuminations as domain factors
- CARLA
- Evaluate for Autonomous Driving Tasks
- Use camera positions, camera field of views, weather conditions, times of day, and control sensitivity as domain factors

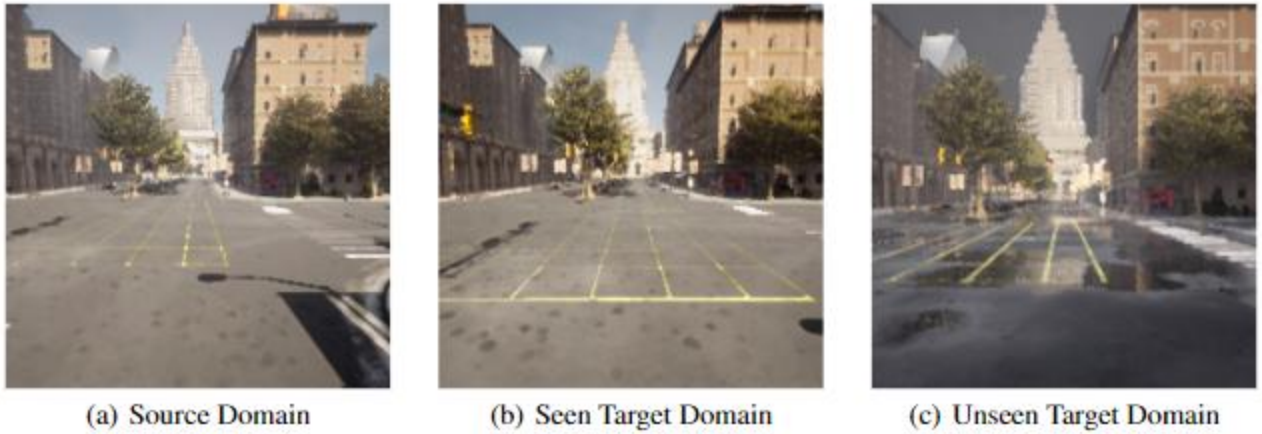

Figure 4. Examples from the Source, Seen Target, and Unseen Target Domains

Evaluation: Performance

- Zero-shot Evaluation
  - Policies of each method are learned on 4 source domains
  - Use 30 seen target domains and 10 unseen target domains
- Experiment results
  - ConPE shows 5.2 ~ 24.0% improvement over other baselines for seen target domains
  - ConPE shows 6.9 ~ 20.0% improvement over other baselines for unseen target domains

(a) Zero-shot Performance in AI2THOR with Object and Point Goal Navigation Tasks

Method	ObjectNav.			PointNav.		
	Source	Seen Target	Unseen Target	Source	Seen Target	Unseen Target
LUSR	53.3±1.1	21.3±1.9	15.1±1.8	85.6±4.6	71.8±3.8	62.4±5.8
CURL	51.3±1.0	8.0±0.1	6.9±1.3	70.8±7.4	55.2±2.7	54.8±3.0
ATC	82.2±9.7	72.3±3.3	51.3±8.6	95.0±3.3	89.1±1.9	81.9±3.6
ACO	55.0±23.8	39.6±21.5	35.8±5.8	91.1±6.3	73.4±2.0	67.5±2.8
EmbCLIP	89.3±3.0	77.6±1.3	59.0±6.4	95.3±4.6	84.5±1.9	77.4±1.4
ConPE	**96.3±1.0**	**83.3±0.3**	**79.7±6.4**	**97.8±1.0**	**89.7±1.6**	**84.3±2.0**

(b) Zero-shot Performance in egocentric-Metaworld with Reach and Reach-wall Tasks

Method	Reach			Reach-Wall		
	Source	Seen Target	Unseen Target	Source	Seen Target	Unseen Target
LUSR	100.0±0.0	46.0±15.1	44.7±2.3	50.0±10.0	33.3±6.1	30.7±6.4
CURL	100.0±0.0	53.3±5.0	46.7±3.1	43.3±15.3	2.0±0.0	0.7±1.2
ATC	100.0±0.0	71.3±8.1	72.0±2.0	66.7±5.8	5.3±1.2	4.0±0.0
ACO	100.0±0.0	52.0±2.0	44.0±3.5	63.3±15.3	8.7±2.3	4.7±1.2
EmbCLIP	100.0±0.0	64.7±6.1	66.7±4.2	**100.0±0.0**	58.0±7.2	49.3±5.0
ConPE	100.0±0.0	**88.7±3.1**	**86.7±3.1**	**100.0±0.0**	**75.3±3.1**	**67.3±2.3**

(c) Zero-shot Performance in CARLA with Different Maps

Method	Map 1			Map 2		
	Source	Seen Target	Unseen Target	Source	Seen Target	Unseen Target
LUSR	2141.9	635.1±606.2	1073.9±212.6	**2279.6**	1173.7±914.3	2159.4±146.5
CURL	945.4	864.2±638.0	1256.0±61.6	1050.1	1089.9±824.0	2190.3±10.2
ATC	2280.5	1684.4±368.2	1073.7±618.8	2272.2	2253.9±218.7	2200.1±307.8
ACO	2265.8	1545.6±596.1	1330.0±144.5	2270.6	2360.9±88.0	2415.5±53.0
EmbCLIP	2235.7	1732.2±588.6	1415.1±669.9	2262.7	2139.1±655.9	2401.3±12.3
ConPE	**2237.5**	**1738.0±163.5**	**1933.4±29.7**	2277.2	**2422.5±79.6**	**2512.9±15.7**

Table 1. Zero-shot Performance

Conclusion

- We presented the CONPE framework, a novel approach that allows embodied RL agents to adapt in a zero-shot manner across diverse visual domains

- The ensemble facilitates domain-invariant and task-specific state representations

- We will adapt the framework with semantic knowledge based on pretrained language models

- Language model improves the policy generalization capability for embodied agents in dynamic complex environments