

# Curiosity-Bottleneck: Exploration by Distilling Task-Specific Novelty

Youngjin Kim <sup>1 4</sup>, Wontae Nam <sup>3</sup>, Hyunwoo Kim <sup>1</sup>  
Jihoon Kim <sup>2</sup> and Gunhee Kim <sup>1</sup>



SEOUL NATIONAL UNIV.  
VISION & LEARNING

2

Clova

3

ML2

4

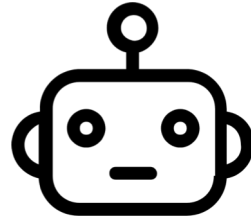
nalbi

Code available at: <http://vision.snu.ac.kr/projects/cb>

# Motivation: Exploration under Distraction



(a) Known Place



Navigating City

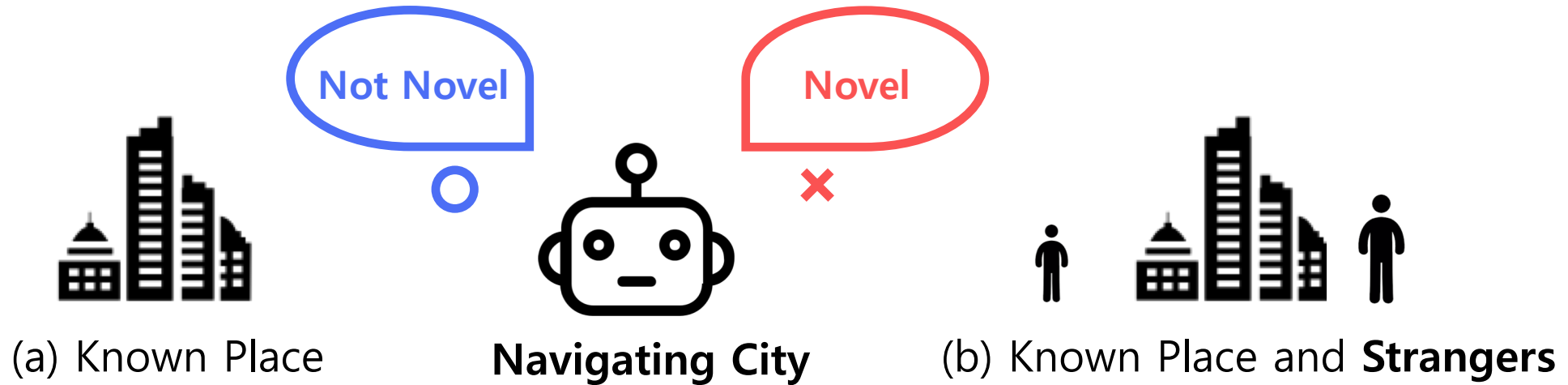


(b) Known Place and **Strangers**

## 1. Distractive Environments are Widespread

- Real-world observations often contain novel but task-irrelevant information.

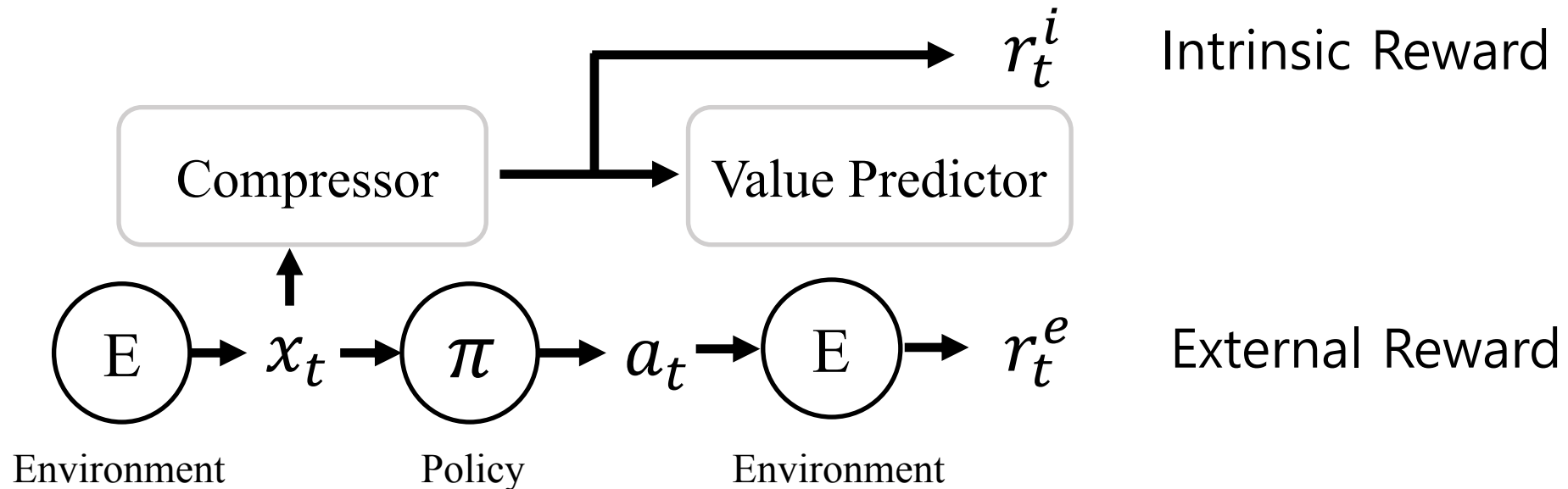
# Motivation: Exploration under Distraction



## 2. Degeneration of Prior Novelty-Based Exploration Strategies

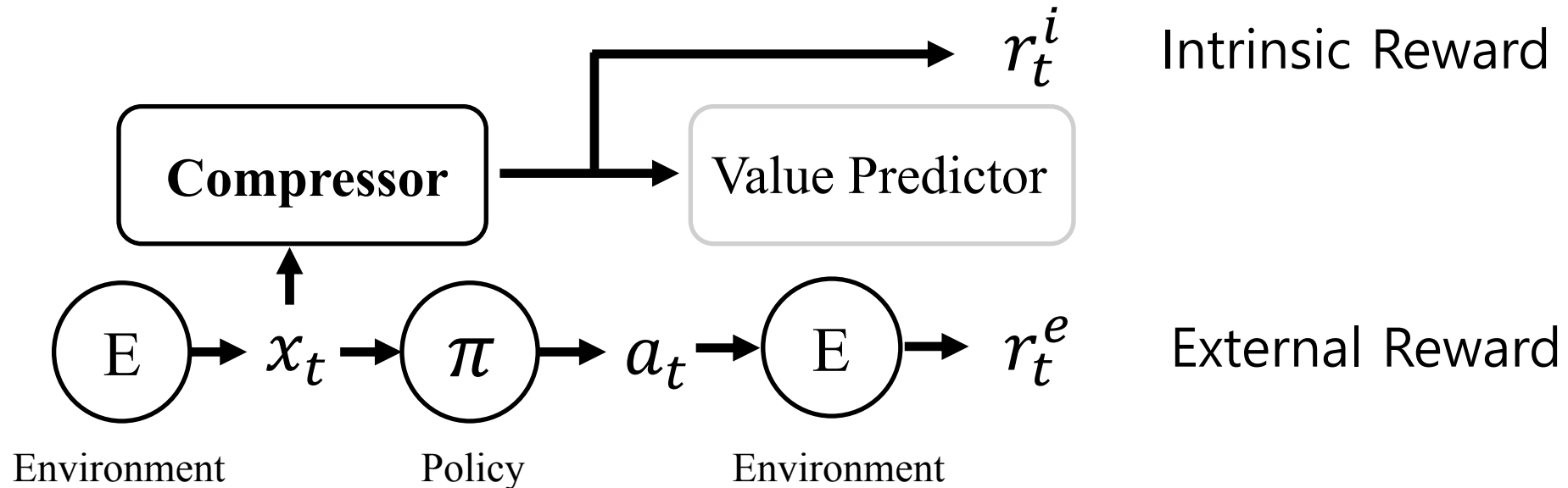
- Due to task-agnostic intrinsic reward
- Need mechanisms to prioritize task-related novelty

# Approach: Curiosity-Bottleneck



Quantify the *'Degree of Compression'* using  
a **compressive value network**

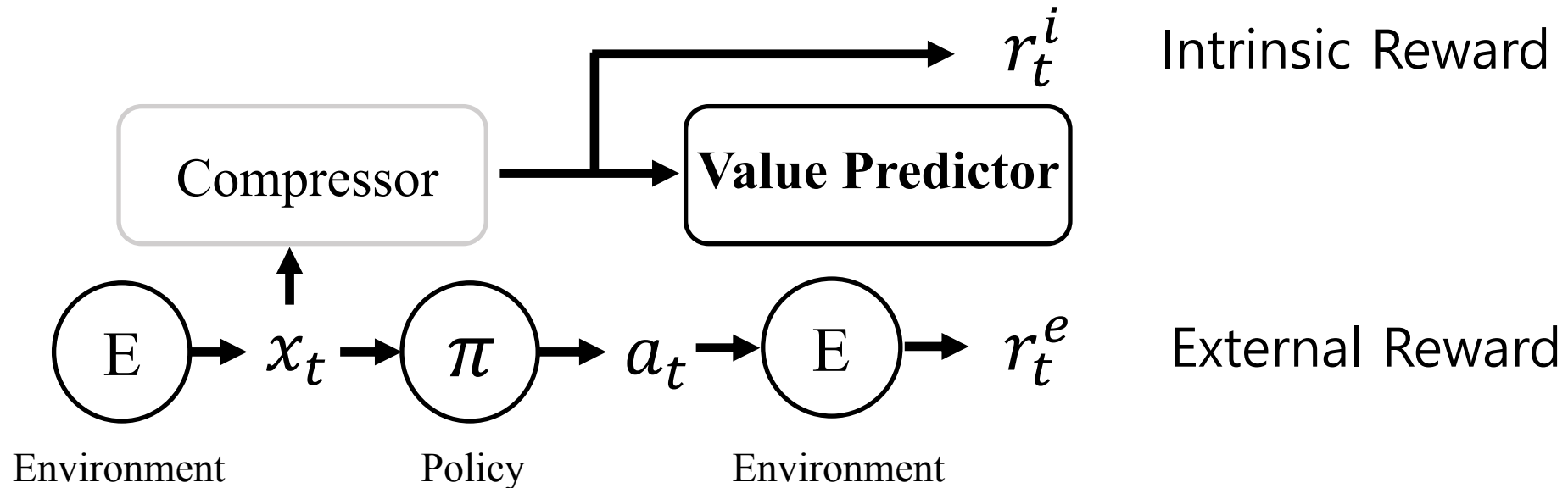
# Approach: Curiosity-Bottleneck



## Compressor

- Encode rare  $x$  to a lengthy code and common  $x$  to a shorter code
- Discard information about  $x$  during compression

# Approach: Curiosity-Bottleneck



## Value Predictor

- Prevent Compressor from discarding task-related information

# Approach: Curiosity-Bottleneck

## 1. Objective Function

- Minimize average code-length of representation  $Z$
- Discard information about observation  $X$

$$\min H(Z) - H(Z|X)$$

- Preserve information related to value estimate  $Y$

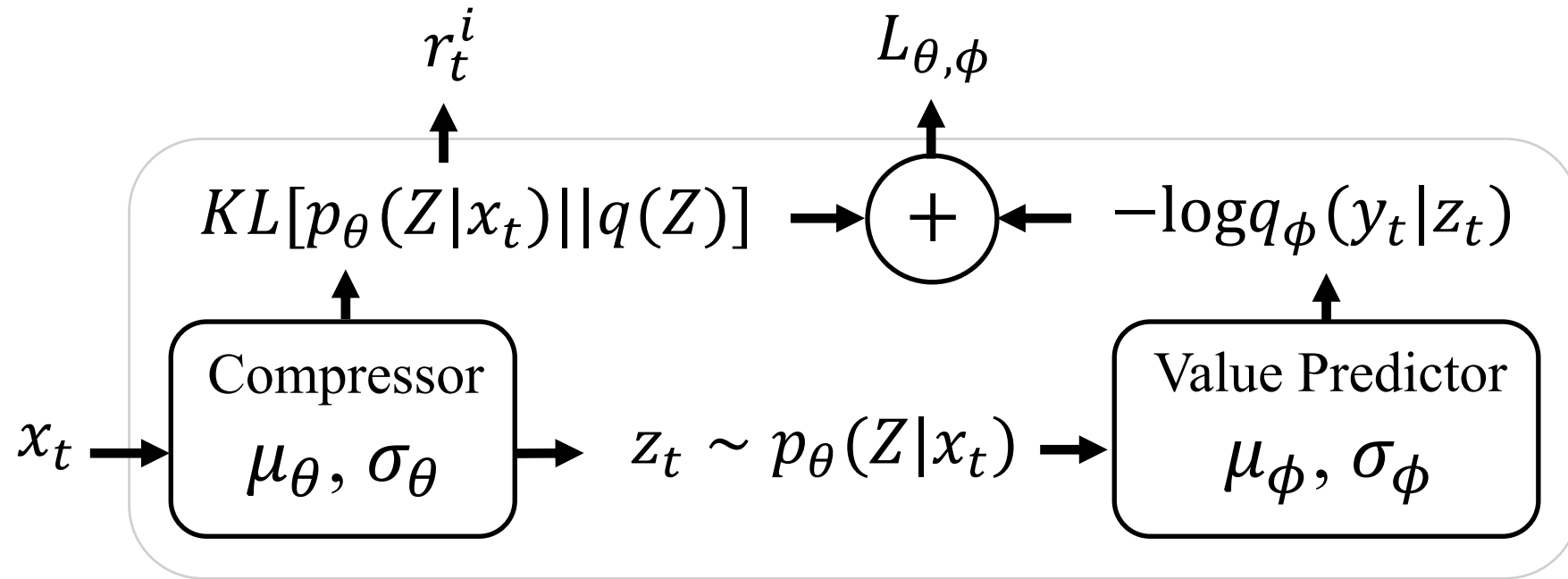
$$\max I(Z; Y)$$

$$L = -I(Z; Y) + \beta I(X; Z)$$

## 2. Intrinsic Reward: Per-instance Mutual Information

$$r^i(x) = \int_{\mathbf{z}} p(\mathbf{z}|x) \log \frac{p(x, \mathbf{z})}{p(x)p(\mathbf{z})} d\mathbf{z}$$

# Approach: Curiosity-Bottleneck



## 3. Approximation

Variational Information Bottleneck with Gaussian assumptions

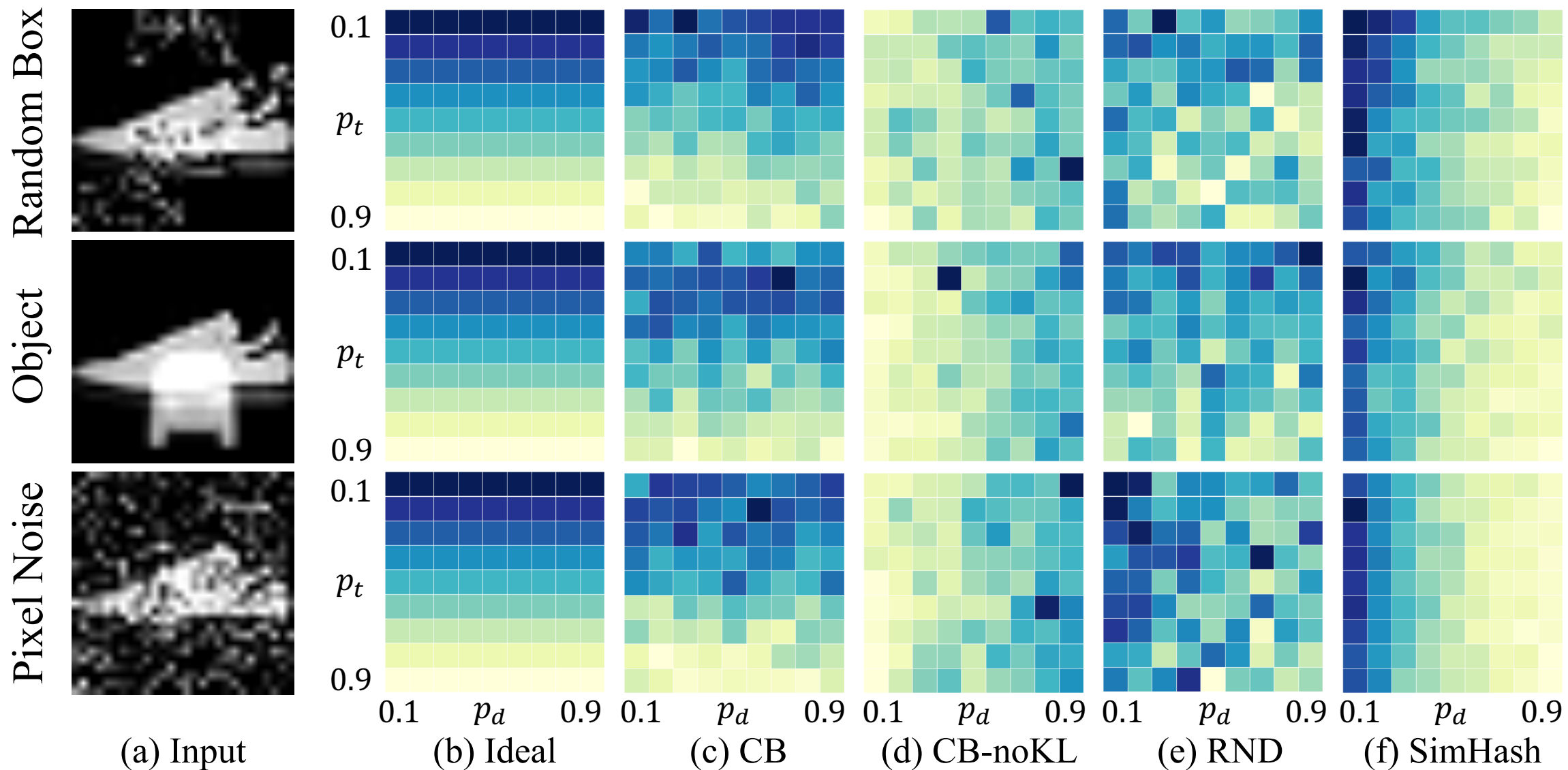
$$L_{\theta,\phi} = E_{x,y}[-\log q_\phi(y|z) + \beta KL[p_\theta(Z|x)||q(Z)]]$$

$$r^i(x) = KL[p_\theta(Z|x)||q(Z)]$$



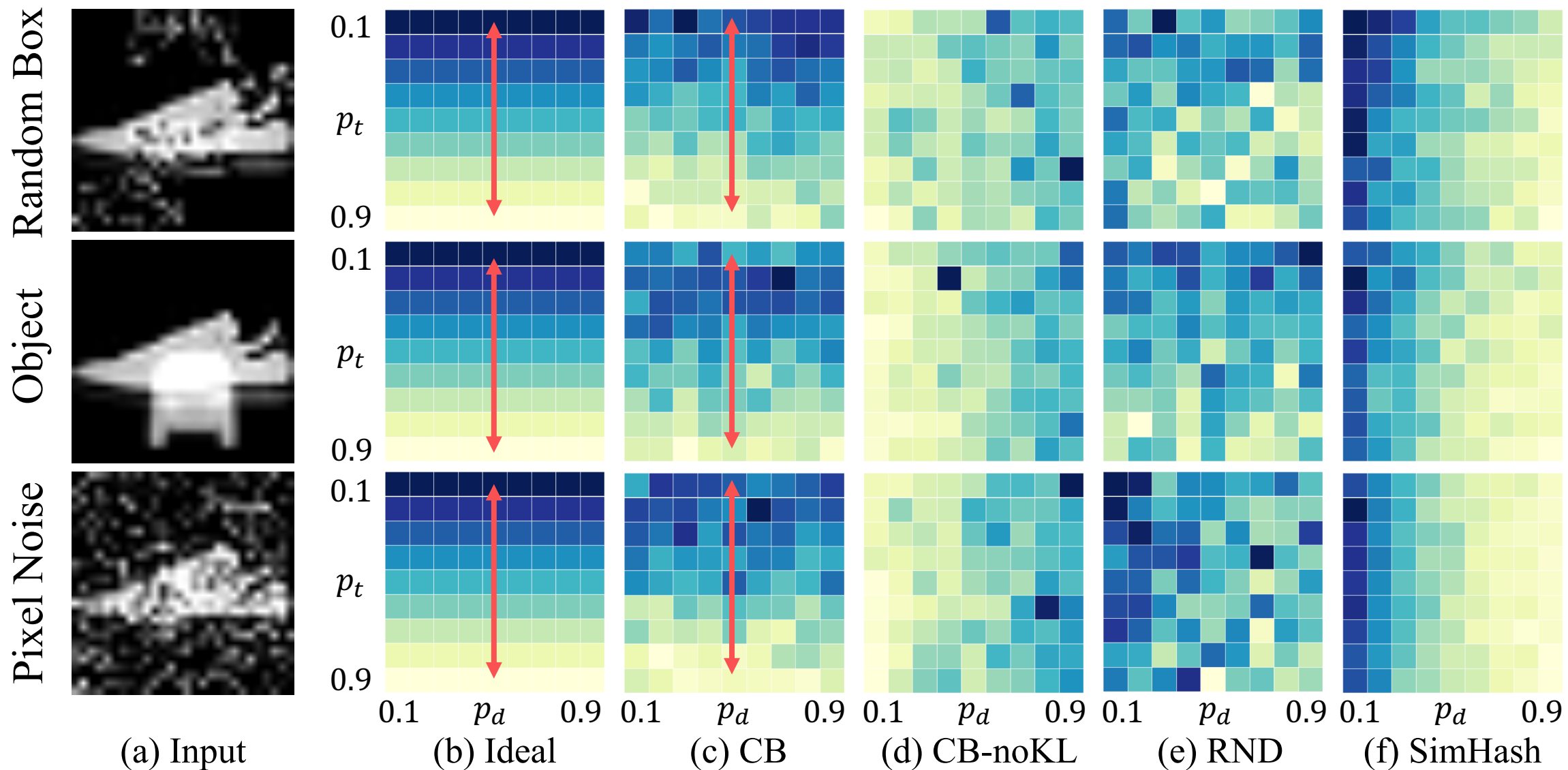
# Experiments: Static Environment

Detects novelty  $p_t(\updownarrow)$  while being robust to distraction  $p_d(\leftrightarrow)$



# Experiments: Static Environment

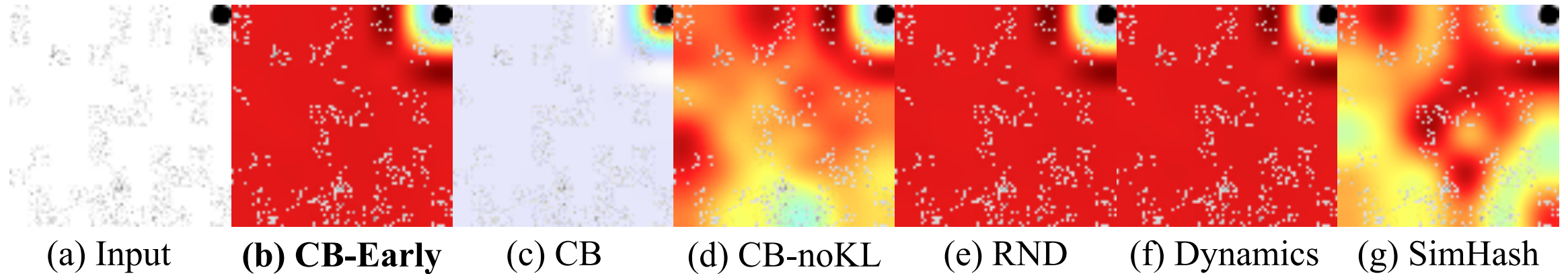
Detects novelty  $p_t(\updownarrow)$  while being robust to distraction  $p_d(\leftrightarrow)$



# Experiments: Treasure-Hunt

## Grad-Cam Visualization

The adaptive exploration strategy



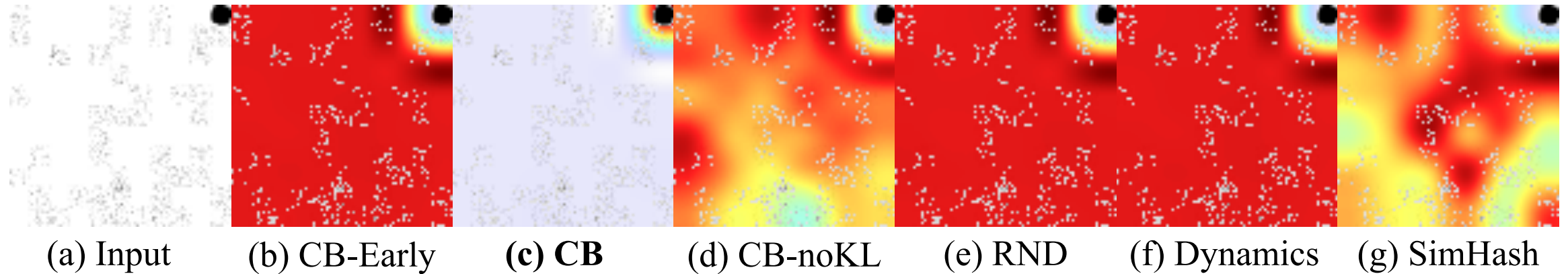
$$KL[p_{\theta}(Z|x) || q(Z)]$$

**Compression loss** term induces **task-agnostic exploration** in early stages

# Experiments: Treasure-Hunt

## Grad-Cam Visualization

The adaptive exploration strategy

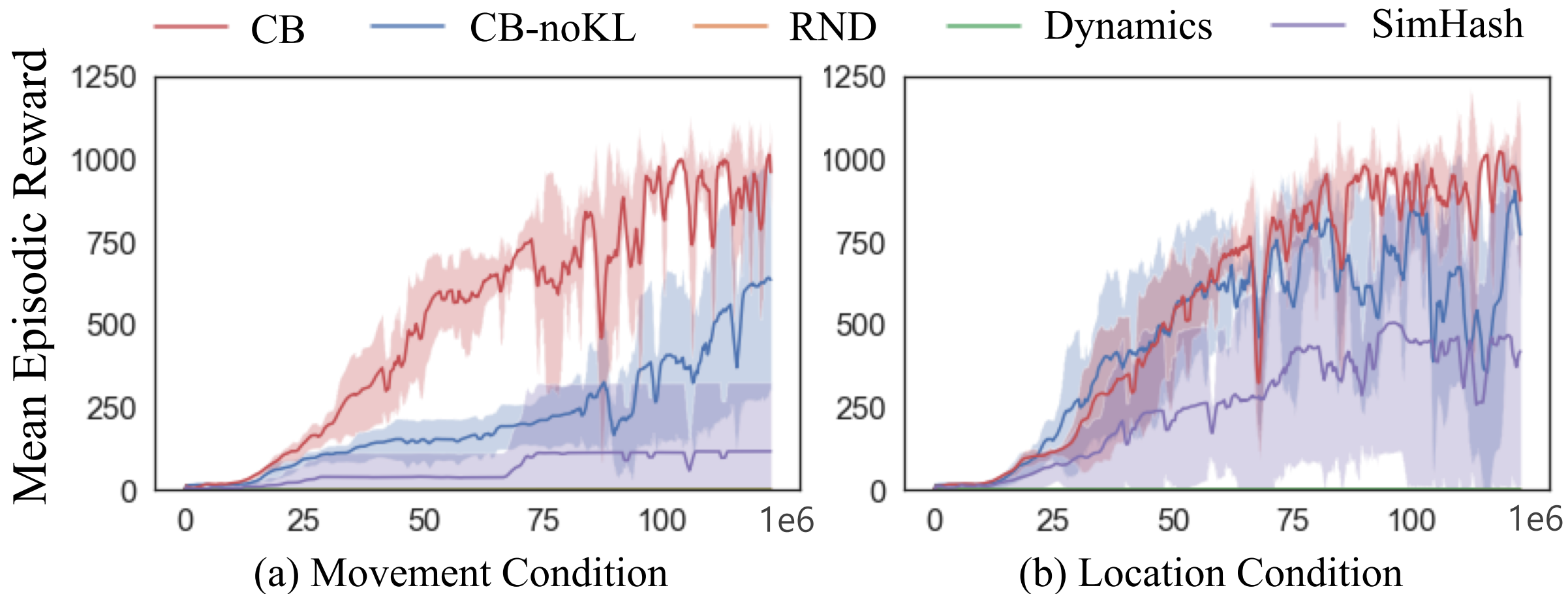


$$-\log q_{\phi}(y|z)$$

**Value prediction loss** term induces **task-specific exploration**  
after collecting external rewards

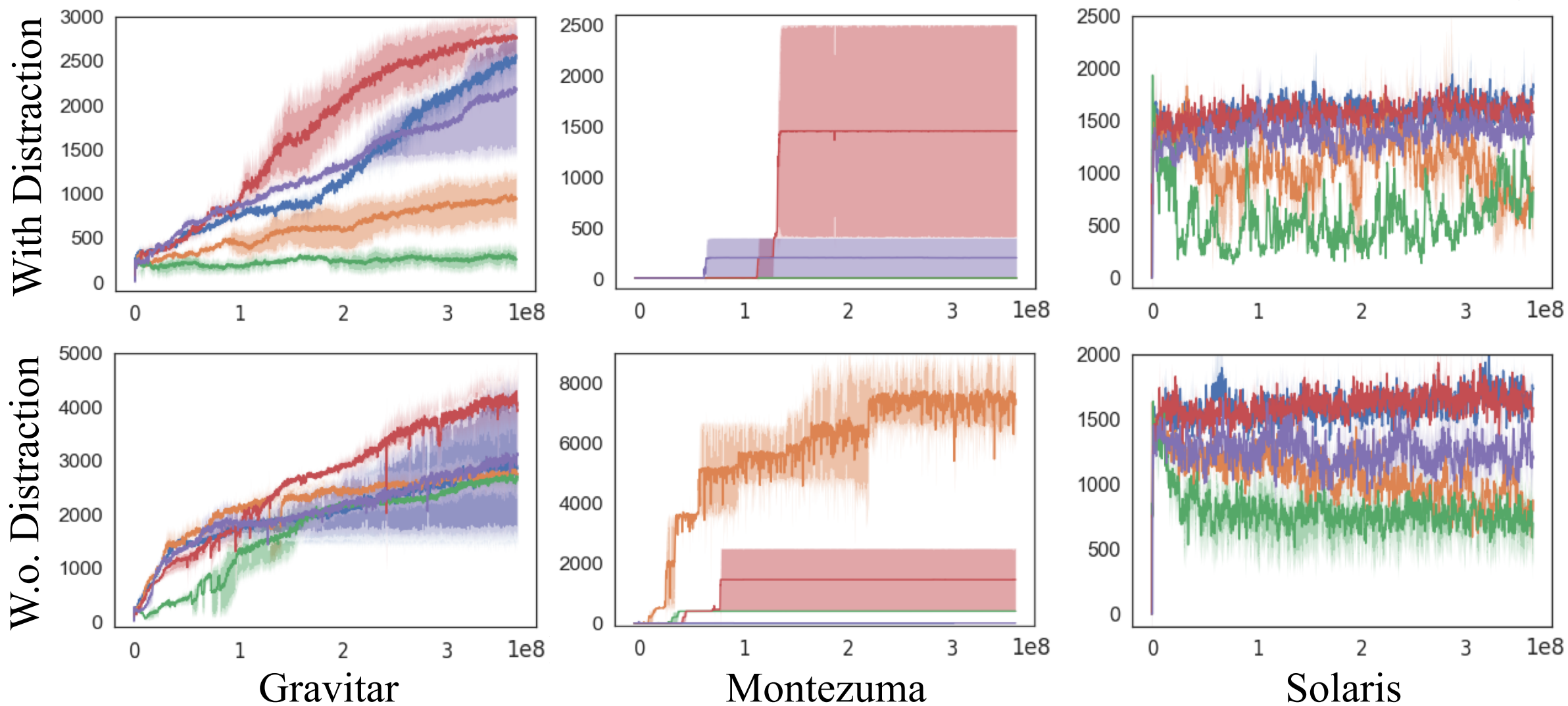
# Experiments: Treasure-Hunt

Consistently outperform baselines on different distraction settings



# Experiments: Atari Hard-Exploration Games

— CB    — CB-noKL    — RND    — Dynamics    — SimHash



# **Curiosity-Bottleneck:**

Exploration by Distilling Task-Specific Novelty

# **Thank You!**

Poster @ Pacific Ballroom #48

Code Available at <http://vision.snu.ac.kr/projects/cb>