Investigating Pre-Training Objectives for Generalization In Vision-Based Reinforcement Learning

ICML 2024

Donghu Kim*, Hojoon Lee*, Kyungmin Lee*, Dongyoon Hwang, Jaegul Choo

KAIST AI

Motivation

- Various Pre-training algorithms for Visual-RL exists

Algorithms	Data Type	Formulation
CURL, MAE, …	Image	(s)
ATC, R3M, SiamMAE, …	Video	$(s)_{0:T}$
BC, SPR, IDM, …	Demonstration	$(s, a)_{0:T}$
DT, CQL, …	Trajectory	$(s, a, r)_{0:T}$

Motivation

- **Image-based algorithms** (CURL, MAE)
  - What do they learn? : **<u>Spatial characteristics</u>** of images
  - e.g., Object sizes and shapes
- **Video-based algorithms** (ATC, R3M, SiamMAE)
  - What do they learn? : **<u>Temporal dynamics</u>** of environments
  - e.g., Object movement speed and direction
- **Demonstration / Trajectory based algorithms** (BC, SPR, IDM / DT, CQL)
  - What do they learn? : **<u>Task-relevant information</u>**
  - e.g., Agents, enemies, and reward structure

Motivation

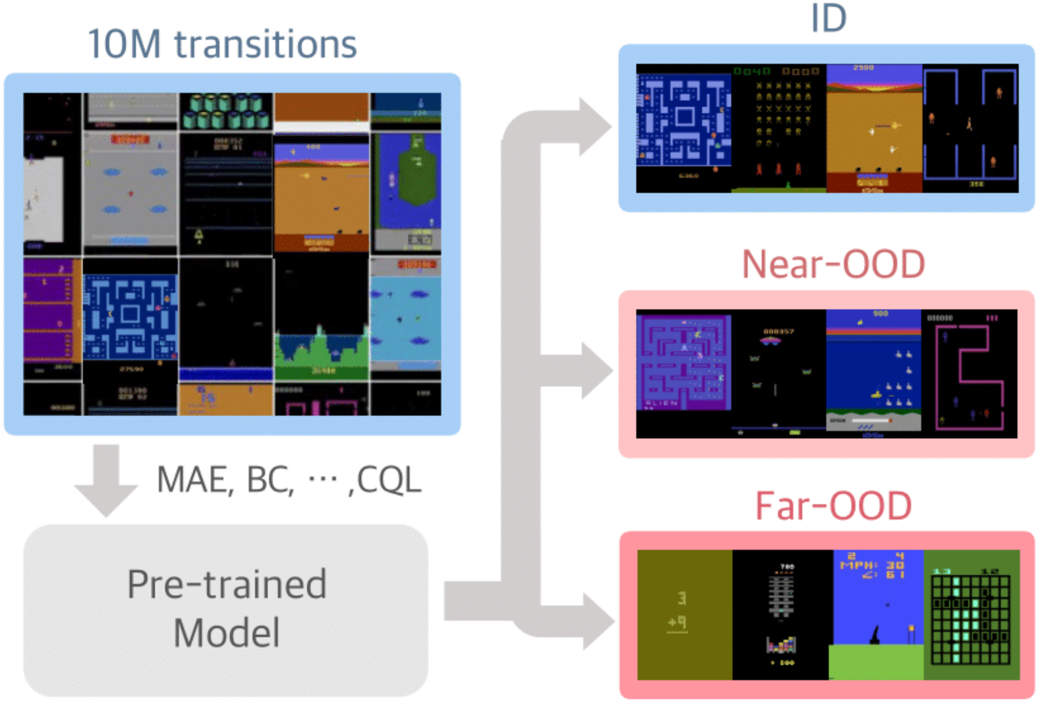
- How do the generalization capabilities of pre-training algorithms differ depending on objectives?

- Before we start this, we need a benchmark

  1. **Unified Protocol**
     1) Data source (Atari)
     2) Same Model Architecture
  2. **Diverse Evaluation Distributions**

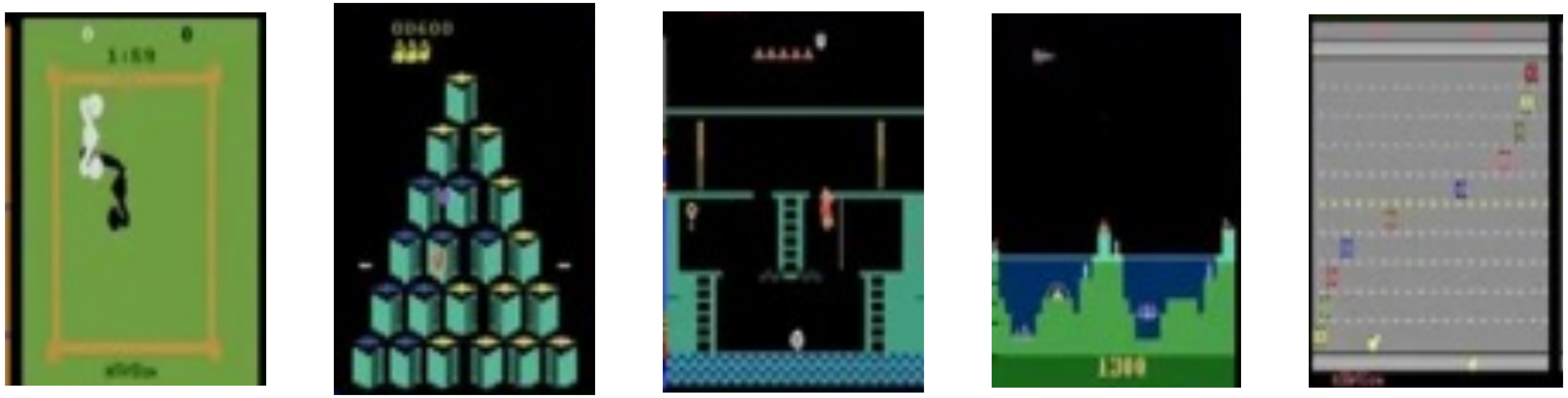
Experimental Setup

- **Dataset**
  - DQN-Replay-Dataset
- **Model**
  - Backbone, Neck, Head
- **Training**
  - Pre-train & Fine-tune
- **Evaluation**
  - In Distribution(ID)
  - Near Out-Of-Distribution(Near-OOD)
  - Far Out-Of-Distribution(Far-OOD)



Experimental Setup

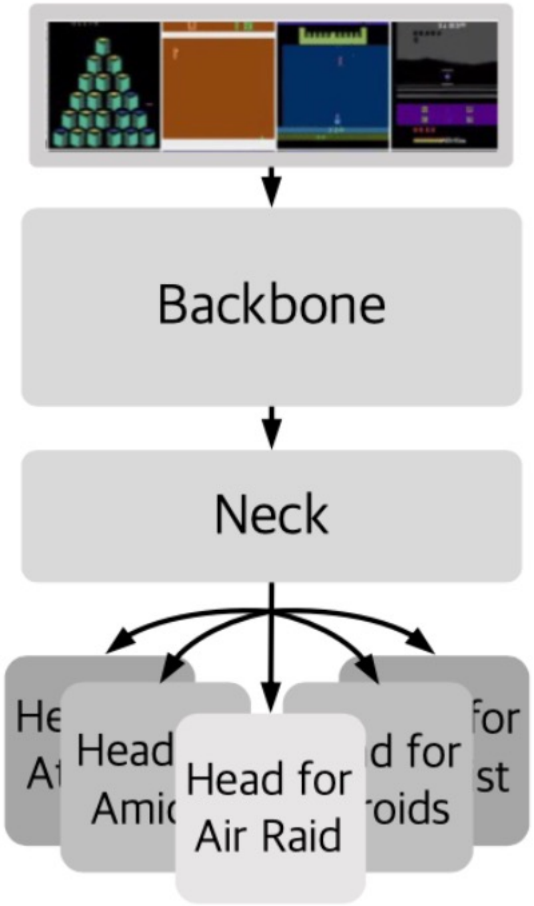
- **Dataset**
  - DQN-Replay-Dataset
  - We chose 10M DQN interactions for offline dataset across 50 Atari games
    - Diverse quality of 200K transitions for each game



Agarwal et al. “An optimisic perspective on offline reinforcement learning.”, ICML 2020.

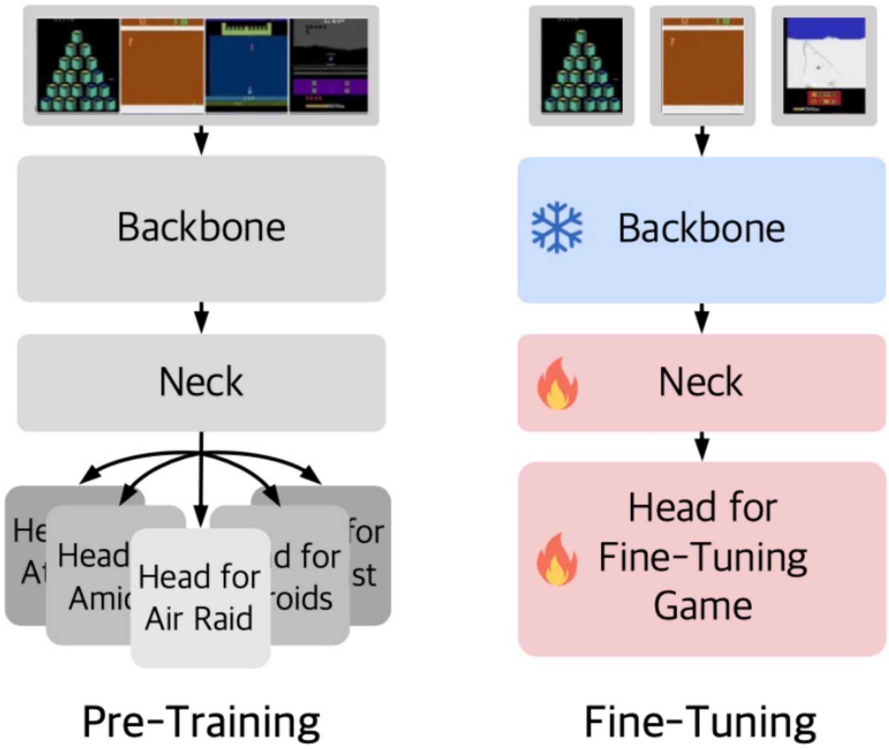
Experimental Setup

- **Model**
  - Backbone:
    - ResNet-50
  - Neck:
    - 2-layer MLP / Game-wise spatial embedding
  - Head:
    - Game-wise Linear Layer



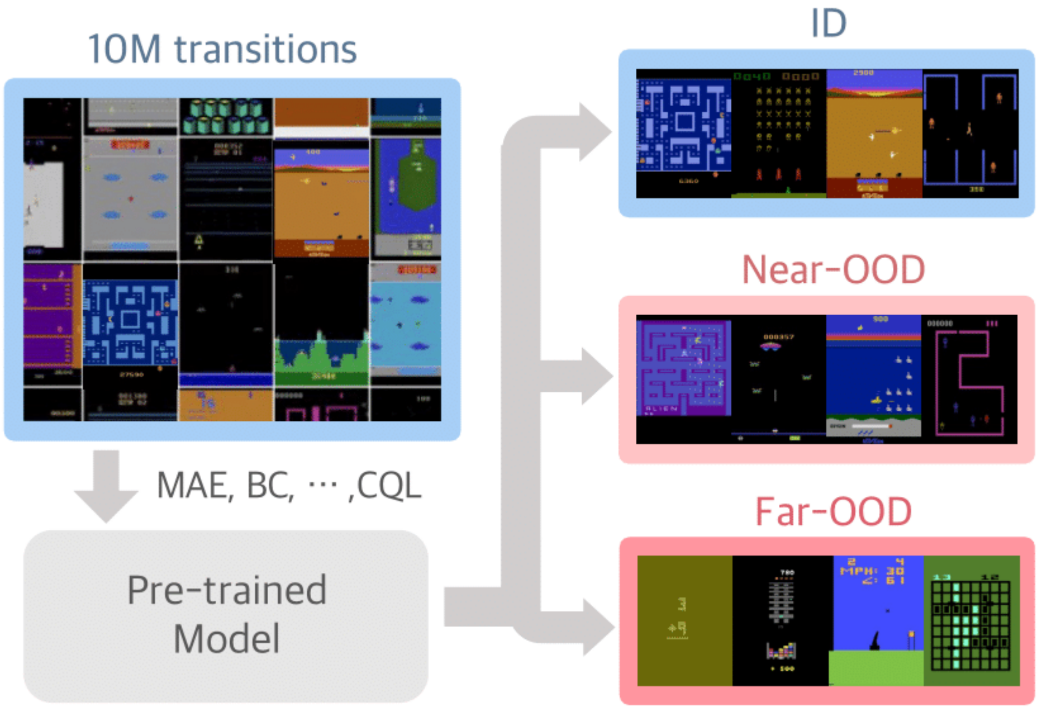
Experimental Setup

- **Training**
  - Pre-training
    - Pre-training with 10M DQN interactions in 50 Atari games
  - Fine-tuning
    - Backbone is kept frozen; others are re-initialized
    - Algorithm: Offline BC, Online RL(Rainbow)



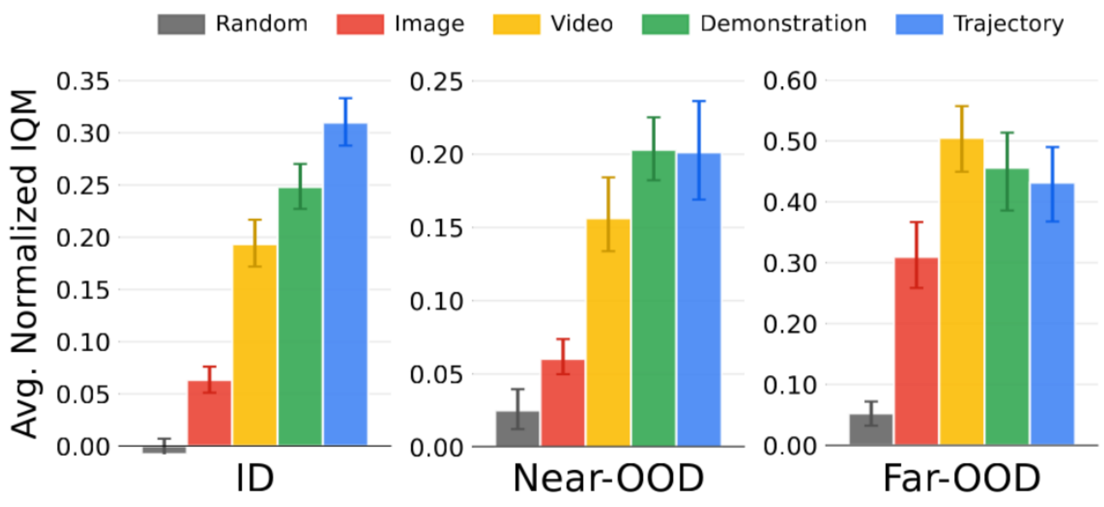
Experimental Setup

- **Evaluation**
  - In-Distribution(ID)
    - 50 games that were used for pre-training
  - Near-OOD
    - 10 games with <u>similar tasks</u> in ID games
  - Far-OOD
    - 5 games with <u>novel tasks</u>



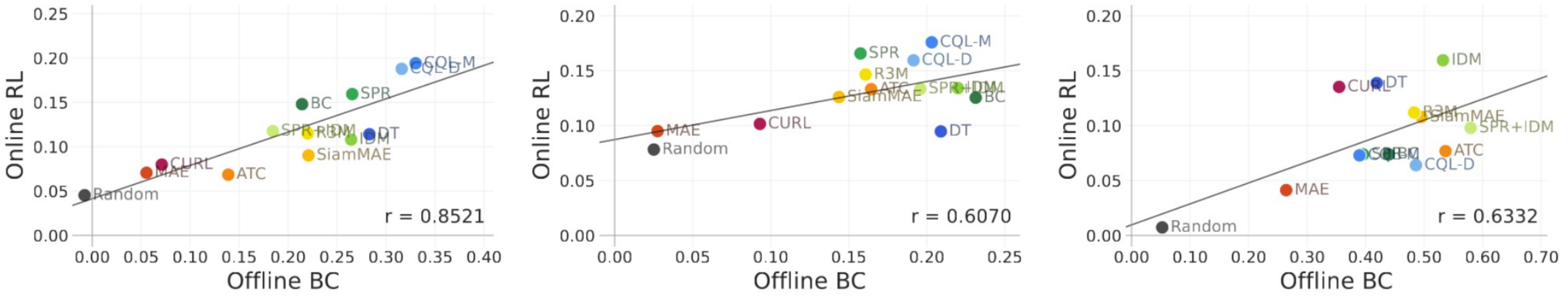
Main Results

1. Learning task-agnostic information from images and videos consistently enhance performance across all environments.
2. Learning task-specific knowledge from demonstrations and trajectories improved performance in ‘familiar’ environments but faltered under stronger distribution shifts.



Main Results

3. Effective adaptation in one scenario correlates to effective adaptation in the other.
   - Strong correlations between Offline BC & Online RL



(a) ID (b) Near-OOD (c) Far-OOD

Takeaways

If your downstream fine-tuning environments…

- Have <u>identical tasks</u> to the pre-training environments:

  → Try trajectory-based algorithms (e.g., DT, CQL, …)

- Have <u>similar tasks</u> to the pre-training environments:

  → Try demonstration-based algorithms (e.g., BC, SPR, IDM)

- Are Unknown / May contain <u>novel tasks</u>:

  → Try video-based algorithms (e.g., ATC, R3M, SiamMAE)

Thank You!