Novelty Assessment Report

Paper: Efficient Reinforcement Learning by Guiding World Models with Non-Curated Data

PDF URL: https://openreview.net/pdf?id=oBXfPyi47m

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-27

Abstract

Leveraging offline data is a promising way to improve the sample efficiency of online reinforcement learning (RL). This paper expands the pool of usable data for offline-to-online RL by leveraging abundant non-curated data that is reward-free, of mixed quality, and collected across multiple embodiments. Although learning a world model appears promising for utilizing such data, we find that naive fine-tuning fails to accelerate RL training on many tasks. Through careful investigation, we attribute this failure to the distributional shift between offline and online data during fine-tuning. To address this issue and effectively use the offline data, we propose two techniques: i) experience rehearsal and ii) execution guidance. With these modifications, the non-curated offline data substantially improves RL's sample efficiency. Under limited sample budgets, our method achieves a 102.8% relative improvement in aggregate score over learning-from-scratch baselines across 72 visuomotor tasks spanning 6 embodiments. On challenging tasks such as locomotion and robotic manipulation, it outperforms prior methods that utilize offline data by a decent margin.

Core Task Landscape

This paper addresses: **Leveraging Non-Curated Offline Data for Sample-Efficient Reinforcement Learning**

A total of **50 papers** were analyzed and organized into a taxonomy with **27 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Offline-to-Online Reinforcement Learning**
- **Model-Based Methods with Non-Curated Data**
- **Skill and Primitive Discovery**
- **Learning from Unstructured Play Data**
- **Imitation Learning from Diverse Data**
- **Purely Offline Reinforcement Learning**
- **Cross-Domain and Transfer Learning**
- **Specialized Offline RL Settings**
- **Policy Evaluation and Selection**
- **Meta-Learning and Multi-Task Offline RL**
- ... and 1 more categories

Complete Taxonomy Tree

- Leveraging Non-Curated Offline Data for Sample-Efficient Reinforcement Learning Survey Taxonomy
- Offline-to-Online Reinforcement Learning
  - Policy Constraint and Regularization Methods (3 papers)
  - [6] Awac: Accelerating online reinforcement learning with offline datasets (Nair, 2020) View paper
  - [12] An offline-to-online reinforcement learning approach based on multi-action evaluation with policy extension (Xuebo Cheng, 2024) View paper
  - [16] Behavior regularized offline reinforcement learning (Wu, 2019) View paper
  - Generative Model-Based Approaches (2 papers)
  - [3] Energy-guided diffusion sampling for offline-to-online reinforcement learning (Liu Xu Hui, 2024) View paper
  - [9] GTA: Generative Trajectory Augmentation with Guidance for Offline Reinforcement Learning (Jaewoo Lee, 2024) View paper
- Model-Based Methods with Non-Curated Data
  - World Model Pre-Training and Fine-Tuning ★ (3 papers)
  - [0] Efficient Reinforcement Learning by Guiding World Models with Non-Curated Data (Anon et al., 2026) View paper
  - [1] Generalist World Model Pre-Training for Efficient Reinforcement Learning (Y Zhao, 2025) View paper
  - [7] Efficient Reinforcement Learning by Guiding Generalist World Models with Non-Curated Data (Zhao Yi, 2025) View paper
  - Personalized and Heterogeneous Agent Methods (1 papers)
  - [18] Persim: Data-efficient offline reinforcement learning with heterogeneous agents via personalized simulators (Agarwal, 2021) View paper
- Skill and Primitive Discovery
  - Offline Skill Extraction for Acceleration (2 papers)
  - [2] Opal: Offline primitive discovery for accelerating offline reinforcement learning (Ajay, 2020) View paper
  - [19] EXTRACT: Efficient Policy Learning by Extracting Transferrable Robot Skills from Offline Data (Jesse Zhang, 2024) View paper
  - Safe Skill Acquisition (1 papers)
  - [35] SAFER: Data-Efficient and Safe Reinforcement Learning via Skill Acquisition (Slack, 2022) View paper

- Learning from Unstructured Play Data
  - Goal-Conditioned Policy Learning from Play (2 papers)
  - [4] Goal-conditioned imitation learning using score-based diffusion policies (Reuss, 2023) View paper
  - [20] Playfusion: Skill acquisition via diffusion from language-annotated play (Chen LiLi, 2023) View paper
  - Conditional Behavior Generation (3 papers)
  - [15] Emergent agentic transformer from chain of hindsight experience (Liu Hao, 2023) View paper
  - [22] From play to policy: Conditional behavior generation from uncurated robot data (Cui, 2022) View paper
  - [29] Goal-driven transformer for robot behavior learning from play data (Congcong Wen, 2024) View paper
- Imitation Learning from Diverse Data
  - Learning from Sub-Optimal and Mixed-Quality Demonstrations (3 papers)
  - [8] Learning from imperfect demonstrations with self-supervision for robotic manipulation (Kun Wu, 2025) View paper
  - [11] Distance weighted supervised learning for offline interaction data (Hejna, 2023) View paper
  - [40] Good data is all imitation learning needs (Amir Samadi, 2024) View paper
  - Auxiliary Data Integration (2 papers)
  - [17] Robust offline imitation learning from diverse auxiliary data (Ghosh, 2024) View paper
  - [37] Fast imitation via behavior foundation models (M Pirotta, 2023) View paper
- Purely Offline Reinforcement Learning
  - Conservative and Uncertainty-Based Offline RL (3 papers)
  - [10] Uncertainty-Based Offline Reinforcement Learning with Diversified Q-Ensemble (An, 2021) View paper
  - [33] Adaptive pessimism via target Q-value for offline reinforcement learning. (Jie Liu, 2024) View paper
  - [50] Efficient Offline Reinforcement Learning With Relaxed Conservatism (Longyang Huang, 2024) View paper
  - Representation and Metric Learning for Offline RL (2 papers)
  - [23] Offline reinforcement learning with pseudometric learning (Dadashi, 2021) View paper
  - [28] Offline rl with observation histories: Analyzing and improving sample complexity (Hong, 2023) View paper
  - Sample-Efficient Offline RL Techniques (3 papers)
  - [26] Pretraining a Shared Q-Network for Data-Efficient Offline Reinforcement Learning (Park Jongchan, 2025) View paper
  - [27] Efficient experience replay architecture for offline reinforcement learning (Longfei Zhang, 2023) View paper
  - [36] Pushing the Limit of Sample-Efficient Offline Reinforcement Learning (P Cheng, 2025) View paper
- Cross-Domain and Transfer Learning
  - Cross-Domain Offline RL (1 papers)
  - [34] DmC: Nearest Neighbor Guidance Diffusion Model for Offline Cross-domain Reinforcement Learning (Nguyen Minh Hoang, 2025) View paper
  - Broad Data Utilization for Generalization (2 papers)
  - [5] Digirl: Training in-the-wild device-control agents with autonomous reinforcement learning (Hao Bai, 2024) View paper
  - [24] Generalization with lossy affordances: Leveraging broad offline data for learning visuomotor tasks (Fang Kuan, 2023) View paper
- Specialized Offline RL Settings
  - Safe Reinforcement Learning with Offline Data (2 papers)
  - [13] Datasets and benchmarks for offline safe reinforcement learning (Zu-xin, 2023) View paper
  - [32] Semi-gradient DICE for Offline Constrained Reinforcement Learning (Kim Woosung, 2025) View paper
  - Offline RL with Non-Markovian and Structured Rewards (2 papers)
  - [47] Using Reward Machines for Offline Reinforcement Learning With Non-Markovian Reward Functions (Wang, 2025) View paper
  - [48] Offline Reinforcement Learning from Datasets with Structured Non-Stationarity (Ackermann, 2024) View paper
  - Offline RL for Specialized Domains (3 papers)
  - [21] Offline reinforcement learning for customizable visual navigation (D Shah, 2022) View paper
  - [25] Offline reinforcement learning for visual navigation (Shah, 2022) View paper
  - [41] Leveraging Factored Action Spaces for Efficient Offline Reinforcement Learning in Healthcare (Tang, 2023) View paper
- Policy Evaluation and Selection
  - Offline Policy Selection (2 papers)
  - [14] When is Offline Policy Selection Sample Efficient for Reinforcement Learning? (Liu, 2023) View paper
  - [44] Towards Practical Offline Reinforcement Learning: Sample Efficient Policy Selection and Evaluation (Liu, 2024) View paper
  - Preference and Reward Learning (1 papers)
  - [42] Preference Elicitation for Offline Reinforcement Learning (Pace, 2025) View paper
- Meta-Learning and Multi-Task Offline RL
  - Meta-RL with Demonstrations (1 papers)
  - [30] Enhanced Meta Reinforcement Learning using Demonstrations in Sparse Reward Environments (Rengarajan, 2022) View paper
  - Multi-Agent and Game-Theoretic Offline RL (1 papers)
  - [31] Sample-Efficient Tabular Self-Play for Offline Robust Reinforcement Learning (Na Li, 2025) View paper
  - Factored Action Spaces (1 papers)
  - [39] Provably (more) sample-efficient offline RL with options (X Hu, 2023) View paper
- Theoretical Foundations and Benchmarking
  - Sample Complexity and Theoretical Analysis (1 papers)
  - [43] SOReL and TOReL: Two Methods for Fully Offline Reinforcement Learning (Fellows, 2025) View paper
  - Benchmarking and Evaluation (2 papers)
  - [45] Interaction-Efficient Reinforcement Learning: Matching the Real World Data Availability (Jiang, 2025) View paper
  - [49] Evaluation of sample efficiency in offline reinforcement learning (Sujit, 2023)
  - Causal Inference and Active Sampling (1 papers)
  - [46] Can Active Sampling Reduce Causal Confusion in Offline Reinforcement Learning? (Gupta, 2023) View paper
  - Extremum Estimation and Flow Matching (1 papers)
  - [38] Extremum Flow Matching for Offline Goal Conditioned Reinforcement Learning (Rouxel, 2025) View paper

Narrative

Core task: leveraging non-curated offline data for sample-efficient reinforcement learning. The field addresses how agents can exploit large, unstructured datasets—often collected without task-specific intent—to accelerate learning when online interaction is costly or risky. The taxonomy reveals a rich landscape spanning multiple methodological branches. Offline-to-Online Reinforcement Learning focuses on warm-starting policies with offline data before fine-tuning online, while Model-Based Methods with Non-Curated Data emphasize learning world models or dynamics from diverse experiences and then planning or adapting policies within those models. Skill and Primitive Discovery extracts reusable behaviors from unstructured trajectories, and Learning from Unstructured Play Data tackles the challenge of goal-agnostic exploration logs. Imitation Learning from Diverse Data and Purely Offline Reinforcement Learning represent contrasting paradigms—one leveraging demonstrations of varying quality, the other operating entirely without environment interaction. Cross-Domain and Transfer Learning, Specialized Offline RL Settings, Policy Evaluation and Selection, Meta-Learning and Multi-Task Offline RL, and Theoretical Foundations and Benchmarking round out the taxonomy, highlighting concerns around generalization, safety constraints, policy choice under uncertainty, and rigorous evaluation.

Within Model-Based Methods, a particularly active line centers on pre-training generalist world models from broad offline corpora and then fine-tuning or guiding them for specific tasks. Guiding World Models[0] exemplifies this approach by steering a pre-trained dynamics model toward task-relevant rollouts, closely aligning with Generalist World Model[1] and Guiding Generalist Models[7], which similarly emphasize adapting large-scale learned simulators. These works contrast with earlier efforts like Opal[2], which focused on narrower model-based offline RL without the generalist pre-training paradigm, and with diffusion-based guidance methods such as Energy Guided Diffusion[3] and Score Based Diffusion Policies[4], which apply similar steering ideas to policy rather than world-model space. The central trade-off across these branches involves balancing the richness and coverage of non-curated data against the risk of distribution shift and the computational cost of large model training, with Guiding World Models[0] positioned among methods that exploit pre-trained generalist representations to achieve sample efficiency in downstream tasks.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. Generalist World Model Pre-Training for Efficient Reinforcement Learning

Authors: Y Zhao, A Scannell, Y Hou, T Cui, L Chen | **Year/Venue**: 2025 | **URL**: View paper

Abstract

â¦ We show that the generalist world model pre-trained on non-curated data can boost RL â¦ our results show that leveraging this non-curated data leads to strong performance. Neverthelessâ¦

⚠ Similarity Notice

These papers appear to be highly similar or the same work, both proposing methods for leveraging non-curated offline data through world model pre-training and fine-tuning with experience rehearsal and execution guidance. The titles, abstracts, methodologies, experimental setups (72 visuomotor tasks, 6 embodiments, same datasets), and core contributions are nearly identical, suggesting they are likely variants or different versions of the same paper.

2. Efficient Reinforcement Learning by Guiding Generalist World Models with Non-Curated Data

Authors: Zhao Yi, Zhao Wenshuai, Hou Yu-xin, Cui, Tianyu, et al. (13 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Leveraging offline data is a promising way to improve the sample efficiency of online reinforcement learning (RL). This paper expands the pool of usable data for offline-to-online RL by leveraging abundant non-curated data that is reward-free, of mixed quality, and collected across multiple embodiments. Although learning a world model appears promising for utilizing such data, we find that naive fine-tuning fails to accelerate RL training on many tasks. Through careful investigation, we attribut...

⚠ Similarity Notice

This paper is highly similar to the original paper; it may be a variant or near-duplicate. Please manually verify.

Contributions Analysis

Overall novelty summary. The paper proposes leveraging non-curated, reward-free, multi-embodiment offline data to improve online RL sample efficiency through world model pre-training with experience rehearsal and execution guidance. It resides in the 'World Model Pre-Training and Fine-Tuning' leaf, which contains only three papers total, indicating a relatively sparse research direction within the broader taxonomy of 50 papers across 27 leaf nodes. This leaf focuses specifically on pre-training generalist world models from diverse offline data and adapting them for downstream tasks, distinguishing it from model-free offline-to-online methods and skill-based approaches.

The taxonomy reveals neighboring work in adjacent branches: 'Personalized and Heterogeneous Agent Methods' addresses agent-specific simulators rather than generalist models, while 'Offline-to-Online Reinforcement Learning' encompasses policy constraint methods and generative approaches that do not rely on world models. The sibling papers in this leaf—Generalist World Model and Guiding Generalist Models—share the core paradigm of pre-training large-scale dynamics models, but the taxonomy's scope notes clarify that methods without world models or those focused on skill extraction belong elsewhere. This positioning suggests the paper operates in a conceptually distinct but sparsely populated niche.

Among 21 candidates examined, two contributions show potential overlap. Contribution A (realistic setting for non-curated data) examined 6 candidates with 1 refutable match, while Contribution B (experience rehearsal and execution guidance) examined 5 candidates with 1 refutable match. Contribution C (the full NCRL method) examined 10 candidates with no refutations found. The limited search scope—top-K semantic matches plus citation expansion—means these statistics reflect a targeted rather than exhaustive literature review. The two refutable pairs suggest some prior exploration of related problem settings or techniques, though the majority of examined work does not directly overlap.

Given the sparse leaf occupancy and the limited search scope, the work appears to occupy a relatively underexplored intersection of world model pre-training and non-curated data utilization. The analysis covers semantic neighbors and citations but does not claim comprehensive field coverage. The two refutable contributions indicate partial precedent, while the core integrated method shows no direct refutation among the candidates examined, suggesting meaningful novelty within the constraints of this targeted literature search.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Realistic setting for leveraging non-curated offline data

Description: The authors introduce a problem setting where offline data is reward-free, of mixed quality, and collected across multiple embodiments, expanding beyond prior work that assumes curated, reward-labeled, or expert-only data.

This contribution was assessed against **6 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Meta-Controller: Few-Shot Imitation of Unseen Embodiments and Tasks in Continuous Control

URL: View paper

Brief Assessment

Meta Controller[56] focuses on few-shot imitation learning across different robot embodiments using reward-free demonstrations, but does not address the specific problem of leveraging mixed-quality, multi-embodiment offline data for reinforcement learning as proposed in the original paper. The candidate's setting involves behavior cloning from expert demonstrations rather than utilizing non-curated data of varying quality for RL training.

---

2. Generalist World Model Pre-Training for Efficient Reinforcement Learning

URL: View paper

Prior Art Analysis

Generalist World Model[1] demonstrates that the concept of leveraging reward-free, mixed-quality, multi-embodiment offline data was already explored and published. Both papers address the same problem setting with nearly identical characterizations. The candidate paper explicitly defines 'non-curated data' as encompassing reward-free and non-expert data from multiple embodiments, and emphasizes this as a key contribution. The original paper's claim to introduce this 'realistic setting' is directly challenged by the prior publication of Generalist World Model[1], which presents the same problem formulation.

Evidence

Evidence 1 - **Rationale**: Both papers claim to propose a 'more realistic setting' with identical characteristics (reward-free, mixed-quality, multi-embodiment), indicating Generalist World Model[1] already established this problem formulation. - **Original**: we propose a more realistic setting for leveraging offline data that consists of reward-free and mixed-quality multi-embodiment data. - **Candidate**: we focus on the more realistic setting where the offline dataset includes non-curated mixed-quality data consisting of reward-free and non-expert data, collected by multiple agents with varying embodiments.

Evidence 2 - **Rationale**: The candidate paper's contribution statement is nearly identical to the original's claim, demonstrating that Generalist World Model[1] already proposed this 'realistic setting' before the original paper. - **Original**: we propose a more realistic setting for leveraging offline data that consists of reward-free and mixed-quality multi-embodiment data. - **Candidate**: • we propose a new and more realistic setting for leveraging offline data where the offline data consists only of reward-free and non-expert multi-embodiment data.

---

3. Offline Pre-trained Multi-agent Decision Transformer

URL: View paper

Brief Assessment

Multi Agent Transformer[57] focuses on multi-agent reinforcement learning with offline datasets from StarCraft II, not on the single-agent reward-free mixed-quality multi-embodiment setting proposed by the original paper.

---

4. Peac: Unsupervised pre-training for cross-embodiment reinforcement learning

URL: View paper

Brief Assessment

Peac[55] focuses on cross-embodiment unsupervised RL with reward-free environments for pre-training, not on leveraging mixed-quality multi-embodiment offline datasets. The candidate addresses embodiment generalization through online interaction in reward-free environments, while the original paper specifically tackles offline-to-online RL with non-curated, mixed-quality data collected across embodiments.

---

5. Offline Pre-trained Multi-Agent Decision Transformer: One Big Sequence Model Tackles All SMAC Tasks

URL: View paper

Brief Assessment

SMAC Transformer[58] focuses on multi-agent reinforcement learning with offline datasets from SMAC environments, not on the specific problem of reward-free, mixed-quality, multi-embodiment data for single-agent RL as proposed in the original paper.

---

6. Efficient Reinforcement Learning by Guiding Generalist World Models with Non-Curated Data

URL: View paper

Brief Assessment

Guiding Generalist Models[7] addresses the same problem setting with reward-free, mixed-quality, multi-embodiment offline data. Both papers target identical data characteristics and problem formulation.

Contribution 2: Experience rehearsal and execution guidance techniques

Description: The authors develop experience rehearsal (retrieving task-relevant trajectories from offline data to reduce distributional shift) and execution guidance (using a prior policy trained on retrieved data to steer exploration) to address the failure of naive world model fine-tuning.

This contribution was assessed against **5 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Efficient Learning of Goal-Oriented Push-Grasping Synergy in Clutter

URL: View paper

Brief Assessment

Push Grasping Synergy[54] focuses on robotic manipulation tasks (push-grasping in clutter) rather than general RL world model fine-tuning. The paper does not address distributional shift in world model pre-training/fine-tuning contexts.

---

2. Generalist World Model Pre-Training for Efficient Reinforcement Learning

URL: View paper

Prior Art Analysis

Generalist World Model[1] demonstrates prior work on both experience rehearsal and execution guidance techniques. The candidate paper describes experience rehearsal as retrieving task-relevant trajectories to prevent catastrophic forgetting and augment initial states, and execution guidance as using a prior policy to steer exploration. These are the same techniques and mechanisms described in the original paper, indicating that Generalist World Model[1] already developed and published these methods.

Evidence

Evidence 1 - **Rationale**: Both papers explicitly name and propose the same two techniques (experience rehearsal and execution guidance), indicating Generalist World Model[1] already developed these methods. - **Original**: we propose two techniques:i)experience rehearsal andii)execution guidance - **Candidate**: we show that generalist world model pre-training (wpt), together with retrieval-based experience rehearsal and execution guidance, enables efficient reinforcement learning (rl)

---

3. Learn the Ropes, Then Trust the Wins: Self-imitation with Progressive Exploration for Agentic Reinforcement Learning

URL: View paper

Brief Assessment

Self Imitation Exploration[53] focuses on self-imitation learning with curriculum scheduling for agentic RL in LLM agents, not on world model fine-tuning with offline data retrieval to address distributional shift in visuomotor control tasks.

---

4. Optimal Volt/Var Control for Unbalanced Distribution Networks With Human-in-the-Loop Deep Reinforcement Learning

URL: View paper

Brief Assessment

Volt Var Control[51] addresses voltage control in power distribution networks using human intervention to modify dangerous actions during training. This is fundamentally different from the original paper's experience rehearsal (retrieving task-relevant trajectories from offline data) and execution guidance (using a prior policy for exploration) techniques designed for general RL fine-tuning with world models.

---

5. Enhancing Sample Efficiency in Online Reinforcement Learning via Policy-Guided Diffusion Models

URL: View paper

Brief Assessment

Policy Guided Diffusion[52] focuses on using diffusion models to generate synthetic samples for replay buffer augmentation in online RL, addressing distribution mismatch between generated and policy data. This is fundamentally different from the original paper's experience rehearsal (retrieving task-relevant trajectories from offline data) and execution guidance (using a prior policy for exploration steering).

Contribution 3: NCRL method leveraging non-curated data in both stages

Description: The authors propose NCRL, a two-stage approach that pre-trains a task-agnostic world model on non-curated data and then fine-tunes it using the proposed techniques, demonstrating substantial improvements in sample efficiency across 72 visuomotor tasks.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Learning View-invariant World Models for Visual Robotic Manipulation

URL: View paper

Brief Assessment

View Invariant Models[62] focuses on learning view-invariant representations for visual robotic manipulation under viewpoint disturbances, not on leveraging non-curated offline data for world model pre-training and fine-tuning across diverse visuomotor tasks.

---

2. Bridging the sim2real gap: Vision encoder pre-training for visuomotor policy transfer

URL: View paper

Brief Assessment

Vision Encoder Pretraining[59] focuses on evaluating pre-trained vision encoders for sim2real transfer in robotics, not on world model pre-training and fine-tuning for visuomotor RL tasks using non-curated offline data.

---

3. Generalist World Model Pre-Training for Efficient Reinforcement Learning

URL: View paper

Brief Assessment

While Generalist World Model[1] presents a similar two-stage approach (pre-training and fine-tuning world models with non-curated data), the specific method name 'NCRL' and its particular implementation details appear to be distinct from the 'WPT' method in the candidate paper.

---

4. Universal visual decomposer: Long-horizon manipulation made easy

URL: View paper

Brief Assessment

Universal Visual Decomposer[65] focuses on visual task decomposition for long-horizon manipulation using pre-trained visual representations, not on world model pre-training and fine-tuning with non-curated offline data for general RL tasks.

---

5. Video-Enhanced Offline Reinforcement Learning: A Model-Based Approach

URL: View paper

Brief Assessment

Video Enhanced Offline[67] focuses on using unlabeled video data to construct world models for visual control tasks, while NCRL specifically addresses reward-free, mixed-quality, multi-embodiment data with experience rehearsal and execution guidance techniques for visuomotor tasks.

---

6. SeMOPO: Learning High-quality Model and Policy from Low-quality Offline Visual Datasets

URL: View paper

Brief Assessment

SeMOPO[66] focuses on handling visual distractors in offline RL through state decomposition and uncertainty estimation, not on leveraging non-curated data across pre-training and fine-tuning stages for world models.

---

7. Visual foresight: Model-based deep reinforcement learning for vision-based robotic control

URL: View paper

Brief Assessment

Visual Foresight[60] focuses on self-supervised model-based RL for vision-based robotic control using unsupervised interaction data, but does not address the two-stage pre-training and fine-tuning framework with non-curated offline data that NCRL proposes. The candidate uses continuous autonomous data collection for training predictive models, not a separate pre-training stage with offline data followed by fine-tuning with experience rehearsal and execution guidance techniques.

8. Dino-wm: World models on pre-trained visual features enable zero-shot planning

URL: View paper

Brief Assessment

Dino World Models[61] focuses on offline world model training using pre-trained visual features (DINOv2) for zero-shot planning in visuomotor tasks, but does not address the two-stage pre-training and fine-tuning framework with experience rehearsal and execution guidance that NCRL proposes for handling non-curated, reward-free, multi-embodiment data.

9. Learning to drive by watching youtube videos: Action-conditioned contrastive policy pretraining

URL: View paper

Brief Assessment

YouTube Policy Pretraining[64] focuses on learning action-conditioned visual representations from driving videos for autonomous driving tasks, not on world model pre-training and fine-tuning for general visuomotor control across multiple embodiments.

10. Offline Robotic World Model: Learning Robotic Policies without a Physics Simulator

URL: View paper

Brief Assessment

Offline Robotic World[63] focuses on offline model-based RL with epistemic uncertainty estimation for robotic control without simulators, whereas NCRL addresses offline-to-online RL using non-curated data across multiple embodiments with experience rehearsal and execution guidance techniques.

Appendix: Text Similarity Detection

Textual similarity detection checked 19 papers and found 3 similarity segment(s) across 2 paper(s).

The following **2 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

1. Generalist World Model Pre-Training for Efficient Reinforcement Learning

Detected in: Core Task (sibling), Contribution: contribution_1, Contribution: contribution_2, Contribution: contribution_3

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

2. Efficient Reinforcement Learning by Guiding Generalist World Models with Non-Curated Data

Detected in: Core Task (sibling), Contribution: contribution_1

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

References

- [0] Efficient Reinforcement Learning by Guiding World Models with Non-Curated Data View paper
- [1] Generalist World Model Pre-Training for Efficient Reinforcement Learning View paper
- [2] Opal: Offline primitive discovery for accelerating offline reinforcement learning View paper
- [3] Energy-guided diffusion sampling for offline-to-online reinforcement learning View paper
- [4] Goal-conditioned imitation learning using score-based diffusion policies View paper
- [5] Digirl: Training in-the-wild device-control agents with autonomous reinforcement learning View paper
- [6] Awac: Accelerating online reinforcement learning with offline datasets View paper
- [7] Efficient Reinforcement Learning by Guiding Generalist World Models with Non-Curated Data View paper
- [8] Learning from imperfect demonstrations with self-supervision for robotic manipulation View paper
- [9] GTA: Generative Trajectory Augmentation with Guidance for Offline Reinforcement Learning View paper
- [10] Uncertainty-Based Offline Reinforcement Learning with Diversified Q-Ensemble View paper
- [11] Distance weighted supervised learning for offline interaction data View paper
- [12] An offline-to-online reinforcement learning approach based on multi-action evaluation with policy extension View paper
- [13] Datasets and benchmarks for offline safe reinforcement learning View paper
- [14] When is Offline Policy Selection Sample Efficient for Reinforcement Learning? View paper
- [15] Emergent agentic transformer from chain of hindsight experience View paper
- [16] Behavior regularized offline reinforcement learning View paper
- [17] Robust offline imitation learning from diverse auxiliary data View paper
- [18] Persim: Data-efficient offline reinforcement learning with heterogeneous agents via personalized simulators View paper
- [19] EXTRACT: Efficient Policy Learning by Extracting Transferrable Robot Skills from Offline Data View paper
- [20] Playfusion: Skill acquisition via diffusion from language-annotated play View paper
- [21] Offline reinforcement learning for customizable visual navigation View paper
- [22] From play to policy: Conditional behavior generation from uncurated robot data View paper
- [23] Offline reinforcement learning with pseudometric learning View paper
- [24] Generalization with lossy affordances: Leveraging broad offline data for learning visuomotor tasks View paper
- [25] Offline reinforcement learning for visual navigation View paper
- [26] Pretraining a Shared Q-Network for Data-Efficient Offline Reinforcement Learning View paper
- [27] Efficient experience replay architecture for offline reinforcement learning View paper
- [28] Offline rl with observation histories: Analyzing and improving sample complexity View paper
- [29] Goal-driven transformer for robot behavior learning from play data View paper
- [30] Enhanced Meta Reinforcement Learning using Demonstrations in Sparse Reward Environments View paper
- [31] Sample-Efficient Tabular Self-Play for Offline Robust Reinforcement Learning View paper

- [32] Semi-gradient DICE for Offline Constrained Reinforcement Learning View paper
- [33] Adaptive pessimism via target Q-value for offline reinforcement learning. View paper
- [34] DmC: Nearest Neighbor Guidance Diffusion Model for Offline Cross-domain Reinforcement Learning View paper
- [35] SAFER: Data-Efficient and Safe Reinforcement Learning via Skill Acquisition View paper
- [36] Pushing the Limit of Sample-Efficient Offline Reinforcement Learning View paper
- [37] Fast imitation via behavior foundation models View paper
- [38] Extremum Flow Matching for Offline Goal Conditioned Reinforcement Learning View paper
- [39] Provably (more) sample-efficient offline RL with options View paper
- [40] Good data is all imitation learning needs View paper
- [41] Leveraging Factored Action Spaces for Efficient Offline Reinforcement Learning in Healthcare View paper
- [42] Preference Elicitation for Offline Reinforcement Learning View paper
- [43] SOReL and TOReL: Two Methods for Fully Offline Reinforcement Learning View paper
- [44] Towards Practical Offline Reinforcement Learning: Sample Efficient Policy Selection and Evaluation View paper
- [45] Interaction-Efficient Reinforcement Learning: Matching the Real World Data Availability View paper
- [46] Can Active Sampling Reduce Causal Confusion in Offline Reinforcement Learning? View paper
- [47] Using Reward Machines for Offline Reinforcement Learning With Non-Markovian Reward Functions View paper
- [48] Offline Reinforcement Learning from Datasets with Structured Non-Stationarity View paper
- [49] Evaluation of sample efficiency in offline reinforcement learning
- [50] Efficient Offline Reinforcement Learning With Relaxed Conservatism View paper
- [51] Optimal Volt/Var Control for Unbalanced Distribution Networks With Human-in-the-Loop Deep Reinforcement Learning View paper
- [52] Enhancing Sample Efficiency in Online Reinforcement Learning via Policy-Guided Diffusion Models View paper
- [53] Learn the Ropes, Then Trust the Wins: Self-imitation with Progressive Exploration for Agentic Reinforcement Learning View paper
- [54] Efficient Learning of Goal-Oriented Push-Grasping Synergy in Clutter View paper
- [55] Peac: Unsupervised pre-training for cross-embodiment reinforcement learning View paper
- [56] Meta-Controller: Few-Shot Imitation of Unseen Embodiments and Tasks in Continuous Control View paper
- [57] Offline Pre-trained Multi-agent Decision Transformer View paper
- [58] Offline Pre-trained Multi-Agent Decision Transformer: One Big Sequence Model Tackles All SMAC Tasks View paper
- [59] Bridging the sim2real gap: Vision encoder pre-training for visuomotor policy transfer View paper
- [60] Visual foresight: Model-based deep reinforcement learning for vision-based robotic control View paper
- [61] Dino-wm: World models on pre-trained visual features enable zero-shot planning View paper
- [62] Learning View-invariant World Models for Visual Robotic Manipulation View paper
- [63] Offline Robotic World Model: Learning Robotic Policies without a Physics Simulator View paper
- [64] Learning to drive by watching youtube videos: Action-conditioned contrastive policy pretraining View paper
- [65] Universal visual decomposer: Long-horizon manipulation made easy View paper
- [66] SeMOPO: Learning High-quality Model and Policy from Low-quality Offline Visual Datasets View paper
- [67] Video-Enhanced Offline Reinforcement Learning: A Model-Based Approach View paper