Novelty Assessment Report

Paper: DiffusionNFT: Online Diffusion Reinforcement with Forward Process
PDF URL: https://openreview.net/pdf?id=VJZ477R89F
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-27

Abstract

Online reinforcement learning (RL) has been central to post-training language models, but its extension to diffusion models remains challenging due to intractable likelihoods. Recent works discretize the reverse sampling process to enable GRPO-style training, yet they inherit fundamental drawbacks. These include solver restrictions, forward–reverse inconsistency, and complicated integration with classifier-free guidance (CFG). We introduce Diffusion Negative-aware FineTuning (DiffusionNFT), a new online RL paradigm that optimizes diffusion models directly on the forward process via flow matching. DiffusionNFT contrasts positive and negative generations to define an implicit policy improvement direction, naturally incorporating reinforcement signals into the supervised learning objective. This formulation enables training with arbitrary black-box solvers, eliminates the need for likelihood estimation, and requires only clean images rather than sampling trajectories for policy optimization. DiffusionNFT is up to $25\times$ more efficient than FlowGRPO in head-to-head comparisons, while being CFG-free. For instance, DiffusionNFT improves the GenEval score from 0.24 to 0.98 within 1k steps, while FlowGRPO achieves 0.95 with over 5k steps and additional CFG employment. By leveraging multiple reward models, DiffusionNFT significantly boosts the performance of SD3.5-Medium in every benchmark tested.

Core Task Landscape

This paper addresses: **Online Reinforcement Learning for Diffusion Models**

A total of **50 papers** were analyzed and organized into a taxonomy with **15 categories**.

Taxonomy Overview

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

- **Policy Gradient Methods for Diffusion Model Fine-Tuning**
- **Flow Matching and Forward Process Optimization**
- **Diffusion as Generative Components in RL Systems**
- **Application-Specific Diffusion RL**
- **Theoretical Foundations and Algorithmic Innovations**
- **Offline-to-Online and Hybrid Learning Paradigms**

Complete Taxonomy Tree

- Online Reinforcement Learning for Diffusion Models Survey Taxonomy
- Policy Gradient Methods for Diffusion Model Fine-Tuning
  - Denoising Diffusion Policy Optimization (3 papers)
  - [22] Training diffusion models with reinforcement learning (Black, 2023) View paper
  - [23] Maximum Entropy Reinforcement Learning with Diffusion Policy (Dong, 2025) View paper
  - [37] Diffusion-based Reinforcement Learning via Q-weighted Variational Policy Optimization (Shutong Ding, 2024) View paper
  - Efficient Score Matching and Gradient Guidance (2 papers)
  - [1] Efficient Online Reinforcement Learning for Diffusion Policy (Ma, 2025) View paper
  - [27] Maximum entropy inverse reinforcement learning of diffusion models with energy-based models (Himchan Hwang, 2024) View paper
  - Group Relative Policy Optimization for Diffusion (2 papers)
  - [17] Consolidating Reinforcement Learning for Multimodal Discrete Diffusion Models (Ma Tianren, 2025) View paper
  - [44] TreeGRPO: Tree-Advantage GRPO for Online RL Post-Training of Diffusion Models (Zheng Ding, 2025) View paper
- Flow Matching and Forward Process Optimization
  - Forward Process Reinforcement Learning ★ (1 papers)
  - [0] DiffusionNFT: Online Diffusion Reinforcement with Forward Process (Anon et al., 2026) View paper
  - Optimal Transport-Guided Policy Learning (1 papers)
  - [14] Score-Based Diffusion Policy Compatible with Reinforcement Learning via Optimal Transport (Sun Ming-Yang, 2025) View paper
- Diffusion as Generative Components in RL Systems
  - Diffusion-Based Data Augmentation and Replay (6 papers)
  - [9] Enhancing Sample Efficiency in Online Reinforcement Learning via Policy-Guided Diffusion Models (Yixuan Dong, 2024) View paper
  - [11] Continual offline reinforcement learning via diffusion-based dual generative replay (Liu Jin-mei, 2024) View paper
  - [24] Stable continual reinforcement learning via diffusion-based trajectory replay (Chen Feng, 2024) View paper
  - [30] Learning from random demonstrations: Offline reinforcement learning with importance-sampled diffusion models (Fang, 2024) View paper
  - [32] Offline-to-Online Reinforcement Learning with Classifier-Free Diffusion Generation (Huang Xiao, 2025) View paper

- [42] Prioritized Generative Replay (Wang, 2024) View paper
- Diffusion World Models for Planning (3 papers)
- [7] Adaptive Online Replanning with Diffusion Models (Zhou, 2023) View paper
- [19] AIGB: Generative Auto-bidding via Diffusion Modeling (Huo, 2024) View paper
- [45] Diffusion model predictive control (Zhou, 2024) View paper
- Diffusion Behavior Models with RL Refinement (4 papers)
- [16] Integrating Diffusion-based Multi-task Learning with Online Reinforcement Learning for Robust Quadruped Robot Control (Qin, 2025) View paper
- [20] Enhancing sample efficiency and exploration in reinforcement learning through the integration of diffusion models and proximal policy optimization (Gao Tianci, 2024) View paper
- [47] Policy Decorator: Model-Agnostic Online Refinement for Large Policy Model (Yuan Xiu, 2024) View paper
- [48] Steering Your Diffusion Policy with Latent Space Reinforcement Learning (Wagenmaker, 2025) View paper

- Application-Specific Diffusion RL
  - Text-to-Image Generation with Human Feedback (7 papers)
  - [2] Reinforcement learning for fine-tuning text-to-image diffusion models (Y Fan, 2023) View paper
  - [3] DPOK: Reinforcement Learning for Fine-tuning Text-to-Image Diffusion Models (Ying, 2023) View paper
  - [4] Large-scale reinforcement learning for diffusion models (Yinan Zhang, 2024) View paper
  - [5] Feedback Efficient Online Fine-Tuning of Diffusion Models (Uehara, 2024) View paper
  - [8] Human-Feedback Efficient Reinforcement Learning for Online Diffusion Model Finetuning (Ayano Hiranaka, 2024) View paper
  - [38] Using Human Feedback to Fine-tune Diffusion Models without Any Reward Model (Kai Yang, 2023) View paper
  - [50] ShieldDiff: Suppressing Sexual Content Generation from Diffusion Models through Reinforcement Learning (Han Dong, 2024) View paper
  - Diffusion RL for Robotics and Control (3 papers)
  - [25] Reducing risk for assistive reinforcement learning policies with diffusion models (Tytarenko, 2024) View paper
  - [33] Stitching sub-trajectories with conditional diffusion model for goal-conditioned offline rl (Sung-Yoon, 2024) View paper
  - [39] Gen-drive: Enhancing diffusion generative driving policies with reward modeling and reinforcement learning fine-tuning (Zhiyu Huang, 2025) View paper
  - Diffusion RL for Language Models (4 papers)
  - [15] d1: Scaling Reasoning in Diffusion Large Language Models via Reinforcement Learning (Siyan Zhao, 2025) View paper
  - [21] Inpainting-Guided Policy Optimization for Diffusion Large Language Models (Zhao Siyan, 2025) View paper
  - [26] Revolutionizing Reinforcement Learning Framework for Diffusion Large Language Models (Wang Yinjie, 2025) View paper
  - [29] Principled and Tractable RL for Reasoning with Diffusion Language Models (Zhan, 2025) View paper
  - Specialized Domain Applications (6 papers)
  - [13] Diffusion-based reinforcement learning for flexibility improvement in energy management systems (Siebe Paesschesoone, 2024) View paper
  - [28] Decision Transformer for IRS-Assisted Systems with Diffusion-Driven Generative Channels (Jie Zhang, 2024) View paper
  - [31] Reinforcement learning with formation energy feedback for material diffusion models. (Jiao Huang, 2025) View paper
  - [36] Diffusion-Based Reinforcement Learning for Edge-Enabled AI-Generated Content Services (Hongyang Du, 2023) View paper
  - [40] Diffusion Models Are Real-Time Game Engines (Valevski, 2024) View paper
  - [41] Multi-objective aerial collaborative secure communication optimization via generative diffusion model-enabled deep reinforcement learning (Chuang Zhang, 2024) View paper
- Theoretical Foundations and Algorithmic Innovations
  - Survey and Tutorial Works (1 papers)
  - [6] Understanding reinforcement learning-based fine-tuning of diffusion models: A tutorial and review (Uehara, 2024) View paper
  - Novel Algorithmic Frameworks (5 papers)
  - [10] Reinforcement Learning with Discrete Diffusion Policies for Combinatorial Action Spaces (Ma, 2025) View paper
  - [34] Enhanced DACER Algorithm with High Diffusion Efficiency (Wang, 2025) View paper
  - [43] EXPO: Stable Reinforcement Learning with Expressive Policies (Dong, 2025) View paper
  - [46] DLPO: Diffusion Model Loss-Guided Reinforcement Learning for Fine-Tuning Text-to-Speech Diffusion Models (Chen Jingyi, 2024) View paper
  - [49] Rl for consistency models: Faster reward guided text-to-image generation (O Oertell, 2024) View paper
- Offline-to-Online and Hybrid Learning Paradigms (3 papers)
  - [12] Augmenting Offline Reinforcement Learning with State-only Interactions (Li, 2024) View paper
  - [18] Safe offline reinforcement learning using trajectory-level diffusion models (R RÃ¶mer, 2024) View paper
  - [35] Offline Reinforcement Learning With Reverse Diffusion Guide Policy (Jia-zhi Zhang, 2024) View paper

Narrative

Core task: online reinforcement learning for diffusion models. This emerging field explores how to adapt and optimize diffusion-based generative models through direct interaction with reward signals or feedback mechanisms. The taxonomy reveals several major branches: Policy Gradient Methods for Diffusion Model Fine-Tuning focuses on adapting standard RL algorithms like PPO to the unique structure of diffusion processes, as seen in works such as DPOK[3] and Large-scale RL Diffusion[4]. Flow Matching and Forward Process Optimization investigates alternative parameterizations and training objectives that can simplify or accelerate learning, while Diffusion as Generative Components in RL Systems examines how diffusion models serve as policy representations or world models within broader RL architectures. Application-Specific Diffusion RL targets domains like text-to-image generation (RL Text-to-Image[2]), robotics, and autonomous systems, whereas Theoretical Foundations and Algorithmic Innovations address convergence guarantees, sample efficiency, and novel algorithmic designs. Finally, Offline-to-Online and Hybrid Learning Paradigms bridge pre-trained diffusion models with online fine-tuning strategies, balancing data efficiency and exploration.

A particularly active line of work centers on sample-efficient fine-tuning: methods like Feedback Efficient Finetuning[5] and Human-Feedback Efficient[8] aim to minimize the number of reward queries needed to align diffusion outputs with human preferences or task objectives. Another contrasting direction emphasizes scalability and robustness, with studies such as Efficient Online Diffusion[1] and RL Diffusion Tutorial[6] providing practical frameworks for large-scale deployment. DiffusionNFT[0] sits within the Flow Matching and Forward Process Optimization branch, focusing on reinforcement learning applied directly to the forward diffusion process rather than solely the reverse denoising steps. This approach distinguishes it from reverse-process methods like DPOK[3] and aligns it more closely with forward-process innovations, offering a complementary perspective on where and how RL signals can be injected into the diffusion pipeline to improve generation quality and task alignment.

Related Works in Same Category

No sibling papers were found in the same taxonomy leaf. A taxonomy-subtopic-level comparison will be produced instead.

Taxonomy-Level Summary

Both subtopics address reinforcement learning for diffusion models but differ fundamentally in their theoretical foundations. Forward Process Reinforcement Learning defines RL objectives directly on the forward diffusion process using techniques like flow matching, while Optimal Transport-Guided Policy Learning leverages optimal transport theory to align distributions during policy fine-tuning. The key distinction lies in whether the method operates on the forward process mechanics versus using OT-based distribution alignment.

Similarities: - Both aim to enable RL-based fine-tuning of diffusion models - Both avoid relying solely on reverse process likelihood estimation - Both represent alternatives to standard policy gradient methods for diffusion models

Differences: - Forward Process RL operates directly on forward diffusion dynamics (flow matching, negative-aware contrasts), while Optimal Transport methods use distribution alignment theory - Forward Process RL explicitly excludes reverse process likelihood methods, while Optimal Transport excludes score matching and gradient guidance approaches - The theoretical foundation differs: forward process mechanics versus optimal transport distance minimization

Suggested Search Directions: - Investigate whether optimal transport methods can be combined with forward process objectives - Explore computational efficiency comparisons between forward process RL and OT-guided approaches - Examine whether these methods target different application domains or model architectures

Sibling Subtopics

- **Optimal Transport-Guided Policy Learning** (leaves: 1, papers: 1)
- Scope: Methods integrating optimal transport theory with diffusion policies to align distributions and enable RL fine-tuning.
- Exclude: Excludes pure policy gradient or score matching approaches; see Efficient Score Matching and Gradient Guidance.

Contributions Analysis

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

Contribution 1: Diffusion Negative-aware FineTuning (DiffusionNFT) paradigm

Description: The authors propose DiffusionNFT, a novel online reinforcement learning approach for diffusion models that operates on the forward diffusion process rather than the reverse process. It contrasts positive and negative generations to define an implicit policy improvement direction, naturally incorporating reinforcement signals into the supervised learning objective without requiring likelihood estimation.

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. DFRL-DS: A Diffusion-based Reinforcement Learning Algorithm in Discrete Actions for Base Station Energy-saving Control

URL: View paper

Brief Assessment

DFRL-DS[64] applies diffusion models to base station energy-saving control in discrete action spaces, not to online RL for visual generation. The candidate represents the forward process as policy training and reverse process as decision-making for energy control, which is fundamentally different from DiffusionNFT's forward-process optimization with positive/negative contrast for image generation.

---

2. Human-Feedback Efficient Reinforcement Learning for Online Diffusion Model Finetuning

URL: View paper

Brief Assessment

Human-Feedback Efficient[8] focuses on online human feedback collection during training with feedback-aligned representation learning, not on forward process optimization or negative-aware contrastive learning as in the original paper.

---

3. Step-level Reward for Free in RL-based T2I Diffusion Model Fine-tuning

URL: View paper

Brief Assessment

Step-level Reward[65] focuses on credit assignment for step-level rewards in RL fine-tuning, not on forward process optimization. It addresses reward sparsity through contribution-based redistribution, whereas DiffusionNFT operates on the forward diffusion process using flow matching with positive/negative sample contrasts.

---

4. Diffusion models as optimizers for efficient planning in offline rl

URL: View paper

Brief Assessment

Diffusion Optimizers Planning[66] focuses on offline RL trajectory planning by decomposing diffusion sampling into trajectory generation (via autoregressive models) and optimization phases, not on online RL with forward process optimization using positive/negative contrastive learning as in the original paper.

---

5. DPOK: Reinforcement Learning for Fine-tuning Text-to-Image Diffusion Models

URL: View paper

Brief Assessment

DPOK[3] focuses on policy gradient methods operating on the reverse diffusion process with KL regularization, while the original paper operates on the forward process using flow matching with contrastive positive/negative generation splits. These represent fundamentally different technical approaches to diffusion model reinforcement learning.

---

6. Controllable guidance in reinforcement learning using diffusion models

URL: View paper

Brief Assessment

Controllable Guidance[63] focuses on offline reinforcement learning using diffusion models for planning and policy representation, not online RL with forward process optimization. The candidate addresses sampling from products of distributions with scalar functions, while the original proposes a novel online RL paradigm contrasting positive/negative generations on the forward diffusion process.

---

7. Understanding reinforcement learning-based fine-tuning of diffusion models: A tutorial and review

URL: View paper

Brief Assessment

RL Diffusion Tutorial[6] is a survey paper reviewing existing RL-based fine-tuning methods for diffusion models, not proposing novel algorithms. It does not present work that predates or refutes DiffusionNFT's novelty claim of forward-process optimization with negative-aware training.

---

8. A simple and effective reinforcement learning method for text-to-image diffusion fine-tuning

URL: View paper

Brief Assessment

Simple Effective RL[62] focuses on comparing PPO and REINFORCE for diffusion fine-tuning efficiency, proposing LOOP as a hybrid method. It does not operate on the forward diffusion process or use contrastive positive/negative generation splitting as DiffusionNFT does.

---

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

URL: View paper

Brief Assessment

Policy-Guided Diffusion[9] focuses on using diffusion models for synthetic experience replay in online RL to augment replay buffers, not on post-training diffusion models themselves via forward process optimization with positive/negative contrasts.

---

10. Offline Reinforcement Learning With Reverse Diffusion Guide Policy

URL: View paper

Brief Assessment

Reverse Diffusion Guide[35] focuses on offline RL using diffusion models as behavior policies, not online RL with forward process optimization and negative-aware contrastive learning.

Contribution 2: Forward-process RL formulation with practical benefits

Description: The forward-process formulation enables training with any black-box solvers (not restricted to first-order SDE samplers), requires only clean images rather than full sampling trajectories for optimization, maintains compatibility with standard diffusion training pipelines, and naturally supports off-policy learning without importance sampling.

This contribution was assessed against **4 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Integrating Diffusion Models into Model-Based Reinforcement Learning for Real-Time Robotic Control A Theoretical Review

URL: View paper

Brief Assessment

Model-Based Robotic Control[61] focuses on integrating diffusion models into model-based RL for robotic control applications, which is a different domain from the original paper's forward-process RL formulation for diffusion model training. The candidate's theoretical review does not address the specific training methodology or the practical benefits claimed in the original contribution.

---

2. Diffusion-BBO: Diffusion-Based Inverse Modeling for Online Black-Box Optimization

URL: View paper

Brief Assessment

Diffusion-BBO[60] focuses on offline-to-online black-box optimization using inverse modeling (mapping objective space to design space), not on forward-process RL formulation for diffusion model training. The candidate addresses uncertainty quantification and acquisition functions for conditional diffusion models in optimization tasks, which is orthogonal to the original paper's contribution of training diffusion models via forward-process RL with arbitrary solvers and without trajectory storage.

---

3. Amortizing intractable inference in diffusion models for vision, language, and control

URL: View paper

Brief Assessment

Amortizing Intractable Inference[58] focuses on posterior inference under diffusion priors using trajectory balance constraints, not on forward-process RL formulation for diffusion model training. The candidate addresses intractable posterior sampling problems rather than online RL training paradigms with black-box solvers.

---

4. Diffusion Model for Data-Driven Black-Box Optimization

URL: View paper

Brief Assessment

Black-Box Optimization[59] focuses on data-driven black-box optimization using diffusion models for structured design variables, not on forward-process RL formulation for diffusion model training. The candidate addresses conditional sampling for optimization problems rather than RL training paradigms.

Contribution 3: Implicit parameterization technique for reinforcement guidance

Description: Instead of learning a separate guidance model and employing guided sampling at inference, the method uses an implicit parameterization that directly integrates reinforcement guidance into a single policy model. This allows continuous RL on one model and eliminates the need for combining multiple models during sampling.

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. Large-scale reinforcement learning for diffusion models

URL: View paper

Brief Assessment

Large-scale RL Diffusion[4] focuses on rolling diffusion for temporal data (video, fluid dynamics) with progressive noise schedules through time, not on reinforcement learning guidance integration for diffusion policy models.

---

2. Diffusion-dice: In-sample diffusion guidance for offline reinforcement learning

URL: View paper

Brief Assessment

Diffusion-dice[57] uses implicit parameterization for policy transformation in offline RL, not for online RL with forward process optimization. The candidate focuses on transforming behavior policy to optimal policy via diffusion models in offline settings, while the original paper addresses online RL through forward process training without separate guidance models.

---

3. IDQL: Implicit Q-Learning as an Actor-Critic Method with Diffusion Policies

URL: View paper

Brief Assessment

IDQL[51] focuses on implicit Q-learning for offline RL with diffusion policies as actors, not on integrating reinforcement guidance into diffusion policy models during training. The original paper's contribution addresses online RL for diffusion models via forward process optimization, which is a fundamentally different problem setting and methodology.

---

4. Toward multi-task generalization in autonomous navigation: A human-in-the-loop adversarial reinforcement learning with diffusion policy

URL: View paper

Brief Assessment

Adversarial Navigation[56] focuses on autonomous navigation tasks using a combined behavior imitation and RL guidance objective, but does not describe an implicit parameterization technique that integrates reinforcement guidance into a single diffusion policy model as in the original paper.

---

5. Offline Goal-Conditioned Reinforcement Learning with Elastic-Subgoal Diffused Policy Learning

URL: View paper

Brief Assessment

Elastic-Subgoal Diffused[52] uses diffusion models for hierarchical goal-conditioned RL with elastic subgoals, not for integrating reinforcement guidance into a single policy model as in the original paper's forward-process RL approach.

---

6. Advantage Weighted Matching: Aligning RL with Pretraining in Diffusion Models

URL: View paper

Brief Assessment

Advantage Weighted Matching[55] does not use implicit parameterization for reinforcement guidance. Instead, it reweights samples by advantage while maintaining the same score/flow-matching loss as pretraining, which is a fundamentally different approach from the original paper's implicit guidance integration technique.

---

7. Enhancing sample efficiency and exploration in reinforcement learning through the integration of diffusion models and proximal policy optimization

URL: View paper

Brief Assessment

PPO Diffusion Integration[20] uses diffusion models as action priors that generate candidate actions for a separate PPO policy, rather than integrating reinforcement guidance directly into a single policy model through implicit parameterization as described in the original paper.

---

8. Uncertainty-aware multi-objective reinforcement learning-guided diffusion models for 3D de novo molecular design

URL: View paper

Brief Assessment

Molecular Design Diffusion[53] focuses on 3D molecular generation using uncertainty-aware multi-objective RL with surrogate models for property prediction. It does not employ implicit parameterization for integrating reinforcement guidance into diffusion policy models as described in the original paper.

---

9. Offline Reinforcement Learning With Reverse Diffusion Guide Policy

URL: View paper

Brief Assessment

The candidate paper discusses using diffusion models to represent behavior policies in offline RL settings, which differs from the original's implicit parameterization that integrates reinforcement guidance into a single policy model for online training.

---

10. Policy-guided diffusion

URL: View paper

Prior Art Analysis

Policy-guided Diffusion[54] demonstrates prior work that uses implicit parameterization to integrate reinforcement guidance into diffusion models. The candidate paper explicitly describes using 'implicit parameterization technique' to directly integrate guidance into a single policy model, avoiding the need for separate guidance models during sampling. Both papers employ implicit parameterization to combine positive and negative signals within a unified model framework, rather than maintaining separate models for guidance. The candidate's approach of defining implicit positive and negative policies through linear combinations of old and new policies directly parallels the original paper's implicit guidance integration concept.

Evidence

Evidence 1 - **Rationale**: The original paper defines implicit positive and negative policies through linear combinations with the old policy. Policy-guided Diffusion[54] similarly applies guidance to shift distributions, demonstrating the prior existence of integrating reinforcement signals into diffusion models through implicit parameterization rather than explicit separate models. - **Original**: theorem 3.2 (policy optimization). consider the training objective: l(θ) = ec,πold(x0|c),t r‖v+ θ (xt, c, t) - v‖2 2 + (1 - r)‖vθ (xt, c, t) - v‖2 2, (5) where v+ θ (xt, c, t) := (1 - β)vold(xt, c, t) + βvθ(xt, c, t), (implicit positive policy) and vθ (xt, c, t) := (1 + β)vold(xt, c, t) - βvθ(xt, c,... - **Candidate**: to achieve this, we train a diffusion model on the offline dataset, from which we can sample synthetic trajectories under the behavior policy. however, while this addresses data sparsity, these trajectories are still off-distribution from our target policy. therefore, inspired by classifier-guided d...

Evidence 2 - **Rationale**: Both papers describe integrating guidance directly into the model rather than using separate guidance models. Policy-guided Diffusion[54] computes policy gradients during denoising to augment the process, which is conceptually equivalent to the

original paper's implicit parameterization approach for integrating reinforcement guidance. - **Original**: implicit guidance integration. intuitively, diffusionnft defines a guidance direction Δ and apply such guidance to the old policy vold (eq. (6)). however, instead of learning a separate guidance model Δθ and employing guided sampling, it adopts an implicit parameterization technique - **Candidate**: inspired by classifier-guided diffusion (section 2.2), we guide the diffusion process using the target policy to move closer to the target distribution. specifically, during the denoising process, we compute the gradient of the action distribution for each action under the target policy, using it to...

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] DiffusionNFT: Online Diffusion Reinforcement with Forward Process View paper
- [1] Efficient Online Reinforcement Learning for Diffusion Policy View paper
- [2] Reinforcement learning for fine-tuning text-to-image diffusion models View paper
- [3] DPOK: Reinforcement Learning for Fine-tuning Text-to-Image Diffusion Models View paper
- [4] Large-scale reinforcement learning for diffusion models View paper
- [5] Feedback Efficient Online Fine-Tuning of Diffusion Models View paper
- [6] Understanding reinforcement learning-based fine-tuning of diffusion models: A tutorial and review View paper
- [7] Adaptive Online Replanning with Diffusion Models View paper
- [8] Human-Feedback Efficient Reinforcement Learning for Online Diffusion Model Finetuning View paper
- [9] Enhancing Sample Efficiency in Online Reinforcement Learning via Policy-Guided Diffusion Models View paper
- [10] Reinforcement Learning with Discrete Diffusion Policies for Combinatorial Action Spaces View paper
- [11] Continual offline reinforcement learning via diffusion-based dual generative replay View paper
- [12] Augmenting Offline Reinforcement Learning with State-only Interactions View paper
- [13] Diffusion-based reinforcement learning for flexibility improvement in energy management systems View paper
- [14] Score-Based Diffusion Policy Compatible with Reinforcement Learning via Optimal Transport View paper
- [15] d1: Scaling Reasoning in Diffusion Large Language Models via Reinforcement Learning View paper
- [16] Integrating Diffusion-based Multi-task Learning with Online Reinforcement Learning for Robust Quadruped Robot Control View paper
- [17] Consolidating Reinforcement Learning for Multimodal Discrete Diffusion Models View paper
- [18] Safe offline reinforcement learning using trajectory-level diffusion models View paper
- [19] AIGB: Generative Auto-bidding via Diffusion Modeling View paper
- [20] Enhancing sample efficiency and exploration in reinforcement learning through the integration of diffusion models and proximal policy optimization View paper
- [21] Inpainting-Guided Policy Optimization for Diffusion Large Language Models View paper
- [22] Training diffusion models with reinforcement learning View paper
- [23] Maximum Entropy Reinforcement Learning with Diffusion Policy View paper
- [24] Stable continual reinforcement learning via diffusion-based trajectory replay View paper
- [25] Reducing risk for assistive reinforcement learning policies with diffusion models View paper
- [26] Revolutionizing Reinforcement Learning Framework for Diffusion Large Language Models View paper
- [27] Maximum entropy inverse reinforcement learning of diffusion models with energy-based models View paper
- [28] Decision Transformer for IRS-Assisted Systems with Diffusion-Driven Generative Channels View paper
- [29] Principled and Tractable RL for Reasoning with Diffusion Language Models View paper
- [30] Learning from random demonstrations: Offline reinforcement learning with importance-sampled diffusion models View paper
- [31] Reinforcement learning with formation energy feedback for material diffusion models. View paper
- [32] Offline-to-Online Reinforcement Learning with Classifier-Free Diffusion Generation View paper
- [33] Stitching sub-trajectories with conditional diffusion model for goal-conditioned offline rl View paper
- [34] Enhanced DACER Algorithm with High Diffusion Efficiency View paper
- [35] Offline Reinforcement Learning With Reverse Diffusion Guide Policy View paper
- [36] Diffusion-Based Reinforcement Learning for Edge-Enabled AI-Generated Content Services View paper
- [37] Diffusion-based Reinforcement Learning via Q-weighted Variational Policy Optimization View paper
- [38] Using Human Feedback to Fine-tune Diffusion Models without Any Reward Model View paper
- [39] Gen-drive: Enhancing diffusion generative driving policies with reward modeling and reinforcement learning fine-tuning View paper
- [40] Diffusion Models Are Real-Time Game Engines View paper
- [41] Multi-objective aerial collaborative secure communication optimization via generative diffusion model-enabled deep reinforcement learning View paper
- [42] Prioritized Generative Replay View paper
- [43] EXPO: Stable Reinforcement Learning with Expressive Policies View paper
- [44] TreeGRPO: Tree-Advantage GRPO for Online RL Post-Training of Diffusion Models View paper
- [45] Diffusion model predictive control View paper
- [46] DLPO: Diffusion Model Loss-Guided Reinforcement Learning for Fine-Tuning Text-to-Speech Diffusion Models View paper
- [47] Policy Decorator: Model-Agnostic Online Refinement for Large Policy Model View paper
- [48] Steering Your Diffusion Policy with Latent Space Reinforcement Learning View paper
- [49] Rl for consistency models: Faster reward guided text-to-image generation View paper
- [50] ShieldDiff: Suppressing Sexual Content Generation from Diffusion Models through Reinforcement Learning View paper
- [51] IDQL: Implicit Q-Learning as an Actor-Critic Method with Diffusion Policies View paper
- [52] Offline Goal-Conditioned Reinforcement Learning with Elastic-Subgoal Diffused Policy Learning View paper
- [53] Uncertainty-aware multi-objective reinforcement learning-guided diffusion models for 3D de novo molecular design View paper
- [54] Policy-guided diffusion View paper
- [55] Advantage Weighted Matching: Aligning RL with Pretraining in Diffusion Models View paper
- [56] Toward multi-task generalization in autonomous navigation: A human-in-the-loop adversarial reinforcement learning with diffusion policy View paper

- [57] Diffusion-dice: In-sample diffusion guidance for offline reinforcement learning View paper
- [58] Amortizing intractable inference in diffusion models for vision, language, and control View paper
- [59] Diffusion Model for Data-Driven Black-Box Optimization View paper
- [60] Diffusion-BBO: Diffusion-Based Inverse Modeling for Online Black-Box Optimization View paper
- [61] Integrating Diffusion Models into Model-Based Reinforcement Learning for Real-Time Robotic Control A Theoretical Review View paper
- [62] A simple and effective reinforcement learning method for text-to-image diffusion fine-tuning View paper
- [63] Controllable guidance in reinforcement learning using diffusion models View paper
- [64] DFRL-DS: A Diffusion-based Reinforcement Learning Algorithm in Discrete Actions for Base Station Energy-saving Control View paper
- [65] Step-level Reward for Free in RL-based T2I Diffusion Model Fine-tuning View paper
- [66] Diffusion models as optimizers for efficient planning in offline rl View paper