# Novelty Assessment Report

**Paper**: One-Step Flow Q-Learning: Addressing the Diffusion Policy Bottleneck in Offline Reinforcement Learning
**PDF URL**: https://openreview.net/pdf?id=60VgwdzxDM
**Venue**: ICLR 2026 Conference Submission
**Year**: 2026
**Report Generated**: 2026-01-07

## Abstract

Diffusion Q-Learning (DQL) has established diffusion policies as a high-performing paradigm for offline reinforcement learning, but its reliance on multi-step denoising for action generation renders both training and inference slow and fragile. Existing efforts to accelerate DQL toward one-step denoising typically rely on auxiliary modules or policy distillation, sacrificing either simplicity or performance. It remains unclear whether a one-step policy can be trained directly without such trade-offs. To this end, we introduce One-Step Flow Q-Learning (OFQL), a novel framework that enables effective one-step action generation during both training and inference, without auxiliary modules or distillation. OFQL reformulates the DQL policy within the Flow Matching (FM) paradigm but departs from conventional FM by learning an average velocity field that directly supports accurate one-step action generation. This design removes the need for multi-step denoising and backpropagation-through-time updates, resulting in substantially faster and more robust learning. Extensive experiments on the D4RL benchmark show that OFQL, despite generating actions in a single step, not only significantly reduces computation during both training and inference but also outperforms multi-step DQL by a large margin. Furthermore, OFQL surpasses all other baselines, achieving state-of-the-art performance in D4RL.



## Core Task Landscape

This paper addresses: **Accelerating Diffusion-Based Offline Reinforcement Learning with One-Step Action Generation**

A total of **27 papers** were analyzed and organized into a taxonomy with **12 categories**.

### Taxonomy Overview

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

- **One-Step Action Generation Methods**
- **Multi-Step Diffusion Policy Methods**
- **Specialized Application Domains**
- **World Model and Latent Space Methods**
- **Theoretical and Survey Contributions**

### Complete Taxonomy Tree

- Accelerating Diffusion-Based Offline Reinforcement Learning with One-Step Action Generation Survey Taxonomy
- One-Step Action Generation Methods
  - Flow Matching-Based One-Step Policies ★ (5 papers)
  - [0] One-Step Flow Q-Learning: Addressing the Diffusion Policy Bottleneck in Offline Reinforcement Learning (Anon et al., 2026) View paper
  - [7] Revisiting Diffusion Q-Learning: From Iterative Denoising to One-Step Action Generation (Nguyen, 2025) View paper
  - [8] Flow-Based Single-Step Completion for Efficient and Expressive Policy Learning (Fleming Cody, 2025) View paper
  - [10] Flow Q-Learning (Park, 2025) View paper
  - [24] One-Step Generative Policies with Q-Learning: A Reformulation of MeanFlow (Zeyuan Wang, 2025) View paper
  - Consistency Distillation for Acceleration (3 papers)
  - [1] Accelerating Diffusion Models in Offline RL via Reward-Aware Consistency Trajectory Distillation (He Yutong, 2025) View paper
  - [5] One-step diffusion policy: Fast visuomotor policies via diffusion distillation (Wang, 2024) View paper
  - [16] Accelerating Diffusion Planners in Offline RL via Reward-Aware Consistency Trajectory Distillation (Xintong Duan, 2025) View paper
  - Unified Generative Policy Frameworks (1 papers)
  - [15] Offline Reinforcement Learning with Generative Trajectory Policies (Xinsong Feng, 2025) View paper
- Multi-Step Diffusion Policy Methods
  - Guidance and Energy-Based Optimization (4 papers)
  - [2] Diffusion-dice: In-sample diffusion guidance for offline reinforcement learning (Liyuan Mao, 2024) View paper
  - [9] Offline Reinforcement Learning With Reverse Diffusion Guide Policy (Jia-zhi Zhang, 2024) View paper
  - [11] Contrastive Energy Prediction for Exact Energy-Guided Diffusion Sampling in Offline Reinforcement Learning (Lu Cheng, 2023) View paper
  - [21] Diffusion Policies with Value-Conditional Optimization for Offline Reinforcement Learning (Yunchang Ma, 2025) View paper
  - Modular and Decoupled Training (2 papers)
  - [12] Diffusion Policies creating a Trust Region for Offline Reinforcement Learning (Tianyu Chen, 2024) View paper
  - [20] Modular Diffusion Policy Training: Decoupling and Recombining Guidance and Diffusion for Offline RL (Chen Zhaoyang, 2025) View paper
  - Entropy Regularization and Exploration (1 papers)
  - [22] Entropy-regularized Diffusion Policy with Q-Ensembles for Offline Reinforcement Learning (Ziwei Luo, 2024) View paper

    - Efficient Diffusion Parameterization (2 papers)
    - [17] Efficient Diffusion Policies for Offline Reinforcement Learning (Kang, 2023) View paper
    - [19] Streaming Diffusion Policy: Fast Policy Synthesis with Variable Noise Diffusion Models (Du, 2024) View paper
- Specialized Application Domains
    - Multi-Agent Coordination (2 papers)
    - [3] OM2P: Offline multi-agent mean-flow policy (Li ZhuoRan, 2025) View paper
    - [23] Multi-agent Coordination via Flow Matching (Lee dongsu, 2025) View paper
    - Safety and Constraint Satisfaction (1 papers)
    - [6] Safe offline reinforcement learning with feasibility-guided diffusion model (Zheng Yi-nan, 2024) View paper
    - Robustness and Generalization (2 papers)
    - [14] Robust Policy Learning via Offline Skill Diffusion (Kim Woo Kyung, 2024) View paper
    - [25] Diffusion Policies for Out-of-Distribution Generalization in Offline Reinforcement Learning (Suzan Ece Ada, 2023) View paper
- World Model and Latent Space Methods (3 papers)
    - [4] Reasoning with Latent Diffusion in Offline Reinforcement Learning (Venkatraman, 2023) View paper
    - [13] DAWM: Diffusion Action World Models for Offline Reinforcement Learning via Action-Inferred Transitions (Li Zongyue, 2025) View paper
    - [27] DyDiff: Long-Horizon Rollout via Dynamics Diffusion for Offline Reinforcement Learning (H Zhao, n.d.) View paper
- Theoretical and Survey Contributions (2 papers)
    - [18] Toward Fast and Generalizable Decision-Making with Diffusion Models (Duan, 2025) View paper
    - [26] Integrating Diffusion Models into Model-Based Reinforcement Learning for Real-Time Robotic Control A Theoretical Review (AV Chaudhari, n.d.) View paper

### Narrative

Core task: Accelerating diffusion-based offline reinforcement learning with one-step action generation. The field has evolved around a central tension between expressive multi-step diffusion policies and the computational cost of iterative sampling at deployment. The taxonomy reflects this divide through several main branches: One-Step Action Generation Methods seek to distill or directly learn policies that produce actions in a single forward pass, often via consistency models, flow matching, or reward-aware distillation techniques such as Reward-Aware Consistency Distillation[1] and Flow Q-Learning[10]. Multi-Step Diffusion Policy Methods retain the iterative denoising framework but explore efficiency gains through streaming architectures like Streaming Diffusion Policy[19], modular training schemes, or trust-region constraints as in Diffusion Trust Region[12]. Specialized Application Domains address settings such as multi-agent coordination (Multi-agent Flow Matching[23]) or safety-critical tasks (Safe Feasibility-Guided Diffusion[6]), while World Model and Latent Space Methods integrate diffusion with learned dynamics or hierarchical skill representations. Theoretical and Survey Contributions provide broader perspectives on the landscape, as seen in Diffusion Model-Based RL Review[26].

A particularly active line of work centers on flow matching-based one-step policies, which leverage continuous normalizing flows to bypass multi-step sampling while preserving expressiveness. One-Step Flow Q-Learning[0] sits squarely within this cluster, aiming to combine the benefits of flow-based generation with value-guided action selection in a single inference step. This contrasts with nearby approaches like Flow-Based Single-Step Completion[8], which may emphasize trajectory completion over action-level Q-learning, and One-Step Generative MeanFlow[24], which explores mean-field approximations for generative policies. Compared to consistency-based methods such as One-step Diffusion Policy[5] or reward-aware distillation schemes like Reward-Aware Consistency Distillation[1], flow matching offers a distinct mathematical framework that can simplify training dynamics. The central open question across these branches remains how to balance sample quality, computational speed, and the ability to incorporate value functions or safety constraints without sacrificing the multimodal expressiveness that originally motivated diffusion models in offline RL.

## Related Works in Same Category

The following **4 sibling papers** share the same taxonomy leaf node with the original paper:

### 1. Revisiting Diffusion Q-Learning: From Iterative Denoising to One-Step Action Generation

**Authors**: Nguyen, Thanh, Yoo, Chang D., Thanh Nguyen, et al. (6 authors total) | **Year/Venue**: 2025 | **URL**: View paper

#### Abstract

Diffusion Q-Learning (DQL) has established diffusion policies as a high-performing paradigm for offline reinforcement learning, but its reliance on multi-step denoising for action generation renders both training and inference slow and fragile. Existing efforts to accelerate DQL toward one-step denoising typically rely on auxiliary modules or policy distillation, sacrificing either simplicity or performance. It remains unclear whether a one-step policy can be trained directly without such trade-...

#### ⚠ Similarity Notice

These papers share nearly identical titles, abstracts, methodology (One-Step Flow Q-Learning using average velocity fields), experimental setups (D4RL benchmarks), and core technical contributions. The candidate paper appears to be a published or revised version of the original submission, with identical performance claims and approach to accelerating diffusion-based offline RL through one-step action generation.

---

### 2. Flow-Based Single-Step Completion for Efficient and Expressive Policy Learning

**Authors**: Fleming Cody, Prajwal Koirala, Cody H. Fleming | **Year/Venue**: 2025 | **URL**: View paper

#### Abstract

Generative models such as diffusion and flow-matching offer expressive policies for offline reinforcement learning (RL) by capturing rich, multimodal action distributions, but their iterative sampling introduces high inference costs and training instability due to gradient propagation across sampling steps. We propose the \textit{Single-Step Completion Policy} (SSCP), a generative policy trained with an augmented flow-matching objective to predict direct completion vectors from intermediate flow...

#### Relationship Analysis

Both papers belong to the Flow Matching-Based One-Step Policies category, using flow matching or velocity field learning to enable direct one-step action generation in offline RL. They overlap in addressing the computational bottleneck of multi-step diffusion policies by learning flow-based models that support single-step inference, and both integrate their one-step policies into actor-critic frameworks for offline RL. The key difference is that the original paper (OFQL) learns an average velocity field over time intervals to enable accurate one-step generation, while the candidate paper (SSCP) learns completion vectors that predict direct shortcuts from intermediate flow states to final actions, with SSCP additionally extending to goal-conditioned RL and hierarchical policy distillation.

---

### 3. Flow Q-Learning

**Authors**: Park, Seohong, Li, Qiyang, Seohong Park, et al. (9 authors total) | **Year/Venue**: 2025 | **URL**: View paper

### Abstract

We present flow Q-learning (FQL), a simple and performant offline reinforcement learning (RL) method that leverages an expressive flow-matching policy to model arbitrarily complex action distributions in data. Training a flow policy with RL is a tricky problem, due to the iterative nature of the action generation process. We address this challenge by training an expressive one-step policy with RL, rather than directly guiding an iterative flow policy to maximize values. This way, we can complete...

### Relationship Analysis

Both papers belong to the Flow Matching-Based One-Step Policies category, using flow matching to enable direct one-step action generation in offline RL. They share the core approach of training a one-step policy via flow matching while avoiding multi-step denoising, and both address the computational bottleneck of diffusion-based methods like DQL. The key difference is that the original paper (OFQL) learns an average velocity field directly for one-step generation during both training and inference, whereas the candidate paper (FQL) trains a separate BC flow policy and distills it into a one-step policy that maximizes Q-values, introducing an additional distillation stage.

---

## 4. One-Step Generative Policies with Q-Learning: A Reformulation of MeanFlow

**Authors**: Zeyuan Wang, Da Li, Yulin Chen, Ye Shi, Liang Bai, et al. (7 authors total) | **Year/Venue**: 2025 | **URL**: View paper

### Abstract

We introduce a one-step generative policy for offline reinforcement learning that maps noise directly to actions via a residual reformulation of MeanFlow, making it compatible with Q-learning. While one-step Gaussian policies enable fast inference, they struggle to capture complex, multimodal action distributions. Existing flow-based methods improve expressivity but typically rely on distillation and two-stage training when trained with Q-learning. To overcome these limitations, we propose to re...

### Relationship Analysis

Both papers belong to the Flow Matching-Based One-Step Policies category, using flow matching or velocity field learning to enable direct one-step action generation in offline RL. They share the core approach of reformulating flow-based policies to avoid multi-step denoising, with both leveraging MeanFlow concepts to learn average velocity fields for efficient one-step inference. The key difference is that the original paper (OFQL) focuses on learning the average velocity field directly through a novel loss formulation without auxiliary modules, while the candidate paper reformulates MeanFlow into a residual noise-to-action mapping (g(at,b,t) = at - u(at,b,t)) with value-guided rejection sampling and adaptive behavior cloning coefficients.

# Contributions Analysis

**Overall novelty summary.** The paper proposes One-Step Flow Q-Learning (OFQL), which reformulates diffusion Q-learning within the flow matching paradigm to enable single-step action generation without auxiliary modules or distillation. It resides in the 'Flow Matching-Based One-Step Policies' leaf, which contains five papers including the original work. This leaf is part of the broader 'One-Step Action Generation Methods' branch, indicating a moderately active research direction focused on eliminating iterative denoising. The taxonomy shows twenty-seven total papers across multiple branches, suggesting that one-step generation is a significant but not dominant theme within the field.

The taxonomy reveals that OFQL's leaf sits alongside 'Consistency Distillation for Acceleration' (three papers) and 'Unified Generative Policy Frameworks' (one paper) within the one-step generation category. Neighboring branches include 'Multi-Step Diffusion Policy Methods' with sub-areas for guidance-based optimization and modular training, as well as 'World Model and Latent Space Methods' that integrate diffusion with learned dynamics. The scope note for OFQL's leaf explicitly excludes diffusion-based methods and consistency distillation, positioning flow matching as a distinct mathematical approach. This structural separation suggests the paper targets a specific methodological niche rather than competing directly with the larger multi-step diffusion community.

Among twenty-one candidates examined, seven refutable pairs were identified across three contributions. The core OFQL framework examined nine candidates with three appearing to provide overlapping prior work, while the average velocity field learning contribution examined ten candidates with two potential refutations. The elimination of multi-step denoising examined only two candidates, both flagged as refutable. These statistics indicate that within the limited search scope, each contribution faces at least some prior work overlap, though the majority of examined candidates (fourteen of twenty-one) were non-refutable or unclear. The relatively small candidate pool means the analysis captures top semantic matches rather than exhaustive coverage.

Given the limited search scope of twenty-one candidates, the analysis suggests moderate novelty concerns primarily around the elimination of multi-step denoising, where both examined papers appeared relevant. The flow matching framework and velocity field learning show more mixed signals, with most candidates non-refutable. The taxonomy context indicates OFQL occupies a recognized but not overcrowded research direction, though the sibling papers in the same leaf warrant careful comparison to establish incremental contributions beyond existing flow-based one-step approaches.

---

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

## Contribution 1: One-Step Flow Q-Learning (OFQL) framework

**Description**: The authors propose OFQL, a new offline RL framework that reformulates Diffusion Q-Learning within the Flow Matching paradigm. Unlike prior methods, OFQL achieves efficient one-step action generation without requiring auxiliary models, policy distillation, or multi-stage training procedures.

This contribution was assessed against **9 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

### 1. Offline RL Without Off-Policy Evaluation

**URL**: View paper

**Brief Assessment**

Offline RL Without OPE[28] focuses on one-step policy improvement using on-policy Q estimates of the behavior policy, avoiding off-policy evaluation entirely. In contrast, OFQL reformulates Diffusion Q-Learning within the Flow Matching paradigm to achieve one-step action generation. These are fundamentally different approaches to achieving one-step operation in offline RL.

---

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

**URL**: View paper

**Brief Assessment**

Diffusion-dice[2] focuses on distribution correction estimation (DICE) methods with diffusion guidance for offline RL, not on one-step action generation frameworks. The candidate uses multi-step diffusion processes and guide-then-select paradigms, which differs fundamentally from OFQL's flow matching approach for direct one-step generation.

---

### 3. One-Step Generative Policies with Q-Learning: A Reformulation of MeanFlow

**URL**: View paper

**Prior Art Analysis**

One-Step Generative MeanFlow[24] demonstrates that prior work exists for one-step action generation in offline RL without distillation. The candidate paper presents a reformulation of MeanFlow that enables direct noise-to-action mapping through a residual formulation (g(at, b, t) = at - u(at, b, t)), achieving one-step generation without auxiliary models or multi-stage training. Both papers address the same core problem: eliminating multi-step denoising while maintaining expressivity. The candidate's approach uses a different mathematical formulation but achieves the same fundamental goal of one-step action generation without distillation, predating or contemporaneous with the original paper's submission.

**Evidence**

Evidence 1 - **Rationale**: Both papers claim to introduce novel one-step frameworks that avoid distillation. The candidate's reformulation of MeanFlow achieves the same objective as OFQL's average velocity field approach, suggesting prior or concurrent work on this problem. - **Original**: we introduce one-step flow q-learning (ofql), a novel framework that enables effective one-step action generation during both training and inference, without auxiliary modules or distillation. ofql reformulates the dql policy within the flow matching (fm) paradigm but departs from conventional fm by... - **Candidate**: we introduce a one-step generative policy for offline reinforcement learning that mapsnoisedirectly toactionsvia aresidual reformulationof meanflow, making it compatible with q-learning. while one-step gaussian policies enable fast inference, they struggle to capture complex, multimodal action distr...

Evidence 2 - **Rationale**: Both papers address the curved trajectory problem in flow matching and propose solutions for one-step generation. The candidate's residual MeanFlow formulation and the original's average velocity field are different technical approaches to the same fundamental challenge. - **Original**: ofql reformulates the dql policy within the flow matching (lipman et al., 2022) paradigm, we facilitate its efficient action sampling. however, conventional flow matching frequently yields curved trajectories, limiting one-step inference accuracy-an issue rooted in the intrinsic properties of the ma... - **Candidate**: we propose a revisedresidual meanflow formulationwith a carefully constructed mapping: g(at, b, t) =at -u(a t, b, t), where at is a linear interpolation between an offline-dataset action and noise. unlike the naive form, this formulation retains theoretical equivalence to the original meanflow under...

Evidence 3 - **Rationale**: Both papers describe one-step action generation using learned average velocity fields without ODE integration, demonstrating the same core technical contribution. - **Original**: once uθ is learned, actions can be generated in a single step through the approximate endpoint map a=t θ(ϵ, s) =ϵ-u θ(ϵ, r=0, t=1;s), ϵ∼ n(0, i), which eliminates the iterative ode integration required by standard flow matching. - **Candidate**: this formula generates ˆavia one-step velocity estimation using the learnedu θ, without integrating over time. flow policies in rl.recent works have proposed using flow matching to learn policies in rl. in suchflow policies, the policy πθ is implicitly defined by a state-conditional velocity field v...

---

### 4. Revisiting Diffusion Q-Learning: From Iterative Denoising to One-Step Action Generation

**URL**: View paper

**Prior Art Analysis**

Revisiting Diffusion Q-Learning[7] demonstrates that the same core contribution—a one-step offline RL framework reformulating Diffusion Q-Learning within Flow Matching without auxiliary models or distillation—was already proposed. Both papers introduce OFQL with identical naming, methodology, and claims. The candidate paper presents the same framework architecture, training procedure, and experimental validation, indicating this is the same work rather than independent prior art.

**Evidence**

Evidence 1 - **Rationale**: Both papers introduce the identical framework name (OFQL) with the same core design: one-step action generation without auxiliary modules or distillation, reformulating DQL within flow matching. - **Original**: we introduce one-step flow q-learning (ofql), a novel framework that enables effective one-step action generation during both training and inference, without auxiliary modules or distillation. ofql reformulates the dql policy within the flow matching (fm) paradigm - **Candidate**: we introduce one-step flow q-learning (ofql), a novel framework specifically designed to enable effective one-step action generation during both training and inference, without the need for auxiliary models, policy distillation, or multi-stage training

Evidence 2 - **Rationale**: Both papers claim the same benefits: elimination of multi-step denoising and BPTT, leading to faster and more stable learning. - **Original**: this design removes the need for multi-step denoising and backpropagation-through-time updates, resulting in substantially faster and more robust learning - **Candidate**: as a result, ofql eliminates the necessity of iterative denoising and recursive gradient propagation, providing a faster, more stable, one-step training-inference pipeline

---

### 5. Flow-Based Single-Step Completion for Efficient and Expressive Policy Learning

**URL**: View paper

**Prior Art Analysis**

Flow-Based Single-Step Completion[8] demonstrates that prior work exists on one-step action generation in offline RL without distillation. The candidate paper proposes Single-Step Completion Policy (SSCP) and SSCQL, which achieve one-step action generation through flow-matching with completion vectors, predating the original paper's OFQL framework. Both papers reformulate diffusion-based policies using flow matching paradigms to enable single-step generation, though they employ different technical mechanisms (completion vectors vs. average velocity fields).

**Evidence**

Evidence 1 - **Rationale**: Both papers claim to be the first to enable one-step action generation in offline RL using flow-based methods without distillation. Flow-Based Single-Step Completion[8] proposes SSCP/SSCQL which achieves this through completion vectors, demonstrating prior work exists on this contribution. - **Original**: we introduce one-step flow q-learning (ofql), a novel framework that enables effective one-step action generation during both training and inference, without auxiliary modules or distillation. ofql reformulates the dql policy within the flow matching (fm) paradigm - **Candidate**: we propose the single-step completion policy (sscp), a generative policy trained with an augmented flow-matching objective to predict direct completion vectors from intermediate flow samples, enabling accurate, one-shot action generation. in an off-policy actor-critic framework, sscp combines the ex...

Evidence 2 - **Rationale**: Flow-Based Single-Step Completion[8] presents SSCQL, a complete offline RL framework using single-step flow-based policies with Q-learning, demonstrating prior work on one-step flow Q-learning frameworks. - **Original**: to this end, we introduce one-step flow q-learning (ofql), a novel framework that enables effective one-step action generation during both training and inference, without auxiliary modules or distillation. - **Candidate**: we now instantiate our single-step completion model as an expressive and efficient policy for offline rl in continuous control, termed sscql (single-step completion q-learning). the actor training objective in this method combines a flow-matching loss, a shortcut-based completion loss, and a q-learn...

---

### 6. Gta: Generative trajectory augmentation with guidance for offline reinforcement learning

**URL**: View paper

**Brief Assessment**

GTA[30] focuses on generative trajectory augmentation for offline RL using diffusion models to create high-rewarding trajectories, not on one-step action generation or flow matching frameworks for policy learning.

---

### 7. RecFlow Policy: Fast and Accurate Visuomotor Policy Learning via Rectified Action Flow

**URL**: View paper

**Brief Assessment**

RecFlow Policy[31] focuses on visuomotor imitation learning for robotics using rectified flow, not offline RL with Q-learning. The candidate addresses action generation in a completely different domain (robot manipulation) without Q-function optimization or behavior-regularized actor-critic frameworks.

---

### 8. Diffusion Policies creating a Trust Region for Offline Reinforcement Learning

**URL**: View paper

**Brief Assessment**

Diffusion Trust Region[12] focuses on a dual-policy approach with diffusion trust region loss for behavior regularization, not on reformulating DQL within the Flow Matching paradigm for one-step generation without distillation as OFQL does.

---

### 9. SAC Flow: Sample-Efficient Reinforcement Learning of Flow-Based Policies via Velocity-Reparameterized Sequential Modeling

**URL**: View paper

**Brief Assessment**

SAC Flow[29] focuses on stabilizing flow-based policy training through velocity reparameterization (Flow-G and Flow-T architectures) to address gradient pathologies, rather than proposing a one-step action generation framework. The candidate addresses training stability via architectural innovations, not the elimination of multi-step denoising that OFQL targets.

---

## Contribution 2: Average velocity field learning for one-step generation

**Description**: The authors introduce a novel approach that learns an average velocity field instead of the conventional marginal velocity field used in Flow Matching. This design enables accurate direct action prediction from a single step, eliminating the need for iterative denoising and curved trajectory approximations.

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. Flow map matching

**URL**: View paper

**Brief Assessment**

Flow Map Matching[38] focuses on learning two-time flow maps for few-step generation in image synthesis, not on average velocity fields for direct action prediction in reinforcement learning contexts.

---

### 2. Consistency Flow Matching: Defining Straight Flows with Velocity Consistency

**URL**: View paper

**Brief Assessment**

Consistency Flow Matching[39] focuses on enforcing self-consistency in velocity fields for flow matching in generative modeling (image generation), not on reinforcement learning or action prediction tasks as in the original paper.

---

### 3. Flowmp: Learning motion fields for robot planning with conditional flow matching

**URL**: View paper

**Brief Assessment**

FlowMP[32] focuses on learning motion fields for robot trajectory planning with second-order dynamics (acceleration), not on one-step action generation in offline RL. The technical domains and objectives differ fundamentally.

---

### 4. Splitmeanflow: Interval splitting consistency in few-step generative modeling

**URL**: View paper

**Prior Art Analysis**

SplitMeanFlow[35] demonstrates that average velocity field learning for one-step generation was previously introduced by MeanFlow[9]. The candidate paper explicitly states that MeanFlow offered the insight of modeling average velocity rather than instantaneous velocity for large-step generation, and that SplitMeanFlow builds upon this foundation. The original paper's claim to introduce a 'novel approach that learns an average velocity field' is refuted by the candidate's acknowledgment that MeanFlow already established this paradigm, with SplitMeanFlow providing an alternative algebraic formulation rather than introducing the concept itself.

**Evidence**

Evidence 1 - **Rationale**: This pair shows that MeanFlow[9] already introduced the concept of modeling average velocity for large-step generation before the original paper. The candidate explicitly credits MeanFlow with this 'profound insight,' indicating prior work exists. - **Original**: we introduce one-step flow q-learning (ofql), a novel framework that enables effective one-step action generation during both training and inference, without auxiliary modules or distillation. ofql reformulates the dql policy within the flow matching (fm) paradigm but departs from conventional fm by... - **Candidate**: Building on this momentum, meanflow [9] offered a profound and physically intuitive insight: for large-step generation, it is more effective to directly model the average velocityalong the entire path connecting noise to data, rather than theinstantaneous velocity at each point. This conceptual shif...

Evidence 2 - **Rationale**: While the original paper claims to address curved trajectories by learning average velocity, the candidate shows this is built upon MeanFlow's existing average velocity framework, with SplitMeanFlow contributing a different mathematical formulation (algebraic vs differential) rather than the core concept. - **Original**: ofql reformulates the dql policy within the flow matching (lipman et al., 2022) paradigm, we facilitate its efficient action sampling. however, conventional flow matching frequently yields curved trajectories, limiting one-step inference accuracy-an issue rooted in the intrinsic properties of the ma... - **Candidate**: the cornerstone of our approach is the additivity property of definite integrals: for any intermediate times ∈ [r, t], the integral over [r, t] is the sum of the integrals over[r, s] and [s, t] (i.e., rt r = rs r + rt s ). by substituting the definition of displacement,(t - r)u(zt, r, t) = rt r vdτ,...

---

### 5. High-dimensional Mean-Field Games by Particle-based Flow Matching

**URL**: View paper

**Brief Assessment**

High-dimensional Mean-Field Games[33] focuses on mean-field games using flow matching for particle-based optimization in game-theoretic settings, not on offline reinforcement learning policy learning. The average velocity field concept in [33] serves a different purpose (disentangling particle trajectories in MFG dynamics) compared to the original paper's use for direct action prediction in RL.

---

### 6. Revisiting Diffusion Q-Learning: From Iterative Denoising to One-Step Action Generation

**URL**: View paper

**Prior Art Analysis**

Revisiting Diffusion Q-Learning[7] presents the identical technical innovation of learning an average velocity field instead of marginal velocity for direct one-step action prediction. Both papers use the same mathematical formulation, the same mean flow identity from Geng et al. (2025), and the same justification for why this enables accurate one-step generation without curved trajectory approximations.

**Evidence**

Evidence 1 - **Rationale**: Both papers use the identical mean flow identity formulation from Geng et al. (2025) to compute the average velocity, with identical mathematical notation and implementation. - **Original**: to address this, we adopt an equivalent reformulation based on themeanflow identity(geng et al., 2025): u(at, r, t;s) =v(at, t;s)-(t-r) d dtu(at, r, t;s) - **Candidate**: to address this, we adopt an equivalent reformulation based on themeanflow identity(geng et al., 2025): u(at, r, t;s) =v(at, t;s)-(t-r) d dtu(at, r, t;s)

Evidence 2 - **Rationale**: Both papers describe the identical one-step endpoint mapping using the learned average velocity field, with the same mathematical formulation for eliminating ODE integration. - **Original**: once uθ is learned, actions can be generated in a single step through the approximate endpoint map a=t θ(ϵ, s) =ϵ-u θ(ϵ, r=0, t=1;s), ϵ~ n(0, i), which eliminates the iterative ode integration required by standard flow matching - **Candidate**: by approximating u with a neural network, actions can be computed in a single step (i.e., a= ϵ-u θ(ϵ, r= 0, t= 1;s) ), thereby eliminating the need for iterative ode integration in fm

---

### 7. Parametric model reduction of mean-field and stochastic systems via higher-order action matching

**URL**: View paper

**Brief Assessment**

Higher-order Action Matching[37] focuses on learning average velocity fields for parametric model reduction of mean-field and stochastic systems across physics parameters, not for direct action prediction in offline RL. The candidate addresses optimal transport and population dynamics in physical systems, whereas the original paper addresses reinforcement learning policy optimization.

---

### 8. Streaming Flow Policy: Simplifying diffusion/flow-matching policies by treating action trajectories as flow trajectories

**URL**: View paper

**Brief Assessment**

Streaming Flow Policy[34] focuses on streaming action trajectories by treating them as flow trajectories with incremental velocity field integration, rather than learning an average velocity field for direct one-step action prediction as in the original paper.

---

### 9. Flow matching with semidiscrete couplings

**URL**: View paper

**Brief Assessment**

Semidiscrete Flow Matching[36] focuses on optimal transport couplings for noise-data pairing in flow matching, not on average velocity field parameterization for direct action prediction in reinforcement learning contexts.

---

### 10. FlowPolicy: Enabling Fast and Robust 3D Flow-based Policy via Consistency Flow Matching for Robot Manipulation

**URL**: View paper

**Brief Assessment**

FlowPolicy[40] focuses on consistency flow matching for robotic manipulation with 3D visual conditions, not on average velocity field learning as proposed in the original paper. The candidate uses consistency constraints on velocity fields rather than learning average velocity fields directly.

---

## Contribution 3: Elimination of multi-step denoising and BPTT in policy learning

**Description**: By adopting the average velocity field formulation, OFQL removes the computational bottleneck of multi-step denoising chains and recursive gradient propagation (BPTT) that plague diffusion-based policies. This results in faster training, more stable optimization, and improved inference efficiency.

This contribution was assessed against **2 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

### 1. Revisiting Diffusion Q-Learning: From Iterative Denoising to One-Step Action Generation

**URL**: View paper

**Prior Art Analysis**

Revisiting Diffusion Q-Learning[7] demonstrates the same computational benefits: removal of multi-step denoising chains and backpropagation-through-time (BPTT) in policy learning. Both papers present identical experimental evidence showing faster training, improved stability, and higher inference efficiency compared to diffusion-based policies, with the same benchmark results on D4RL.

**Evidence**

Evidence 1 - **Rationale**: Both papers report identical experimental measurements demonstrating the computational efficiency gains from eliminating multi-step denoising, with the same numerical results. - **Original**: dql's training time scales nearly linearly with the number of denoising steps-from 11.7 hours at 5 steps to 49.5 hours at 50 steps-while ofql completes training in only 6.3 hours. at inference, ofql reaches 846.5 hz, compared to 238.7 hz for 5-step dql - **Candidate**: dql's training time scales nearly linearly with the number of denoising steps-from 11.7 hours at 5 steps to 49.5 hours at 50 steps-while ofql completes training in only 6.3 hours. at inference, ofql reaches 846.5 hz, compared to 238.7 hz for 5-step dql and just 35.5 hz for 50-step dql

Evidence 2 - **Rationale**: Both papers present the same experimental validation on D4RL showing that elimination of multi-step denoising leads to both computational efficiency and performance improvements. - **Original**: extensive experiments on the d4rl benchmark show that ofql, despite generating actions in a single step, not only significantly reduces computation during both training and inference but also outperforms multistep dql by a large margin - **Candidate**: extensive empirical evaluations on the d4rl benchmark demonstrate that ofql not only surpasses dql in performance but also significantly improves both training and inference efficiency, all while maintaining a simple learning pipeline

---

### 2. Flow-Based Single-Step Completion for Efficient and Expressive Policy Learning

**URL**: View paper

**Prior Art Analysis**

Flow-Based Single-Step Completion[8] demonstrates prior work on eliminating multi-step denoising and backpropagation-through-time (BPTT) in policy learning. The candidate explicitly addresses both computational bottlenecks by enabling single-step action generation that avoids iterative denoising chains and recursive gradient propagation. The paper shows that their single-step completion approach eliminates the need for BPTT across training, critic updates, and inference, achieving the same goals as the original paper's contribution.

**Evidence**

Evidence 1 - **Rationale**: Both papers achieve the same goal of removing multi-step denoising through single-step generation mechanisms, demonstrating Flow-Based Single-Step Completion[8] addresses this bottleneck. - **Original**: this design removes the need for multi-step denoising and backpropagation-through-time updates, resulting in substantially faster and more robust learning. - **Candidate**: by enabling single-step inference, our method resolves both issues efficiently. third, test-time action generation is similarly reduced to a single forward pass: $\pi\theta(s) = z + h\theta(z, s,0, 1)$, where $z \sim p0$.

Evidence 2 - **Rationale**: Flow-Based Single-Step Completion[8] achieves dramatic speedups by eliminating the BPTT and multi-step sampling bottlenecks identified in the original paper, demonstrating prior work on this contribution. - **Original**: the training speed is doubly affected: beyond the diffusion loss, dql requires two rounds of policy sampling per iteration-one for the current action and another for the next-to compute all loss components. in addition, dql leverages the reparameterization trick to backpropagate through the entire d... - **Candidate**: sscql demonstrates a significant advantage in training and inference efficiency, being up to 64x faster in training than dql, a strong-performing diffusion actor baseline, and offering over an order-of-magnitude speedup in inference.

---

## Appendix: Text Similarity Detection

Textual similarity detection checked 19 papers and found 3 similarity segment(s) across 1 paper(s).

The following **1 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. Revisiting Diffusion Q-Learning: From Iterative Denoising to One-Step Action Generation

**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.

## References


- [0] One-Step Flow Q-Learning: Addressing the Diffusion Policy Bottleneck in Offline Reinforcement Learning View paper
- [1] Accelerating Diffusion Models in Offline RL via Reward-Aware Consistency Trajectory Distillation View paper
- [2] Diffusion-dice: In-sample diffusion guidance for offline reinforcement learning View paper
- [3] OM2P: Offline multi-agent mean-flow policy View paper
- [4] Reasoning with Latent Diffusion in Offline Reinforcement Learning View paper
- [5] One-step diffusion policy: Fast visuomotor policies via diffusion distillation View paper
- [6] Safe offline reinforcement learning with feasibility-guided diffusion model View paper
- [7] Revisiting Diffusion Q-Learning: From Iterative Denoising to One-Step Action Generation View paper
- [8] Flow-Based Single-Step Completion for Efficient and Expressive Policy Learning View paper
- [9] Offline Reinforcement Learning With Reverse Diffusion Guide Policy View paper
- [10] Flow Q-Learning View paper
- [11] Contrastive Energy Prediction for Exact Energy-Guided Diffusion Sampling in Offline Reinforcement Learning View paper
- [12] Diffusion Policies creating a Trust Region for Offline Reinforcement Learning View paper
- [13] DAWM: Diffusion Action World Models for Offline Reinforcement Learning via Action-Inferred Transitions View paper
- [14] Robust Policy Learning via Offline Skill Diffusion View paper
- [15] Offline Reinforcement Learning with Generative Trajectory Policies View paper
- [16] Accelerating Diffusion Planners in Offline RL via Reward-Aware Consistency Trajectory Distillation View paper
- [17] Efficient Diffusion Policies for Offline Reinforcement Learning View paper
- [18] Toward Fast and Generalizable Decision-Making with Diffusion Models View paper
- [19] Streaming Diffusion Policy: Fast Policy Synthesis with Variable Noise Diffusion Models View paper
- [20] Modular Diffusion Policy Training: Decoupling and Recombining Guidance and Diffusion for Offline RL View paper
- [21] Diffusion Policies with Value-Conditional Optimization for Offline Reinforcement Learning View paper
- [22] Entropy-regularized Diffusion Policy with Q-Ensembles for Offline Reinforcement Learning View paper
- [23] Multi-agent Coordination via Flow Matching View paper
- [24] One-Step Generative Policies with Q-Learning: A Reformulation of MeanFlow View paper
- [25] Diffusion Policies for Out-of-Distribution Generalization in Offline Reinforcement Learning View paper
- [26] Integrating Diffusion Models into Model-Based Reinforcement Learning for Real-Time Robotic Control A Theoretical Review View paper
- [27] DyDiff: Long-Horizon Rollout via Dynamics Diffusion for Offline Reinforcement Learning View paper
- [28] Offline RL Without Off-Policy Evaluation View paper
- [29] SAC Flow: Sample-Efficient Reinforcement Learning of Flow-Based Policies via Velocity-Reparameterized Sequential Modeling View paper
- [30] Gta: Generative trajectory augmentation with guidance for offline reinforcement learning View paper
- [31] RecFlow Policy: Fast and Accurate Visuomotor Policy Learning via Rectified Action Flow View paper
- [32] Flowmp: Learning motion fields for robot planning with conditional flow matching View paper
- [33] High-dimensional Mean-Field Games by Particle-based Flow Matching View paper
- [34] Streaming Flow Policy: Simplifying diffusion/flow-matching policies by treating action trajectories as flow trajectories View paper
- [35] Splitmeanflow: Interval splitting consistency in few-step generative modeling View paper
- [36] Flow matching with semidiscrete couplings View paper
- [37] Parametric model reduction of mean-field and stochastic systems via higher-order action matching View paper
- [38] Flow map matching View paper
- [39] Consistency Flow Matching: Defining Straight Flows with Velocity Consistency View paper
- [40] FlowPolicy: Enabling Fast and Robust 3D Flow-based Policy via Consistency Flow Matching for Robot Manipulation View paper