Novelty Assessment Report

Paper: Diffusion Blend: Inference-Time Multi-Preference Alignment for Diffusion Models
PDF URL: https://openreview.net/pdf?id=M2DXbwO8le
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-30

Abstract

Reinforcement learning (RL) algorithms have been used recently to align diffusion models with downstream objectives such as aesthetic quality and text-image consistency by fine-tuning them to maximize a single reward function under a fixed KL regularization. However, this approach is inherently restrictive in practice, where alignment must balance multiple, often conflicting objectives. Moreover, user preferences vary across prompts, individuals, and deployment contexts, with varying tolerances for deviation from a pre-trained base model. We address the problem of inference-time multi-preference alignment: given a set of basis reward functions and a reference KL regularization strength, can we design a fine-tuning procedure so that, at inference time, it can generate images aligned with any user-specified linear combination of rewards and regularization, without requiring additional fine-tuning? We propose Diffusion Blend, a novel approach to solve inference-time multi-preference alignment by blending backward diffusion processes associated with fine-tuned models, and we instantiate this approach with three algorithms: DB-MPA for multi-reward alignment, DB-KLA for KL regularization control, and DB-MPA-LS for approximating DB-MPA without additional inference cost. Extensive experiments show that Diffusion Blend algorithms consistently outperform relevant baselines and closely match or exceed the performance of individually fine-tuned models, enabling efficient, user-driven alignment at inference-time.

Core Task Landscape

This paper addresses: **Inference-Time Multi-Preference Alignment for Diffusion Models**

A total of **45 papers** were analyzed and organized into a taxonomy with **12 categories**.

Taxonomy Overview

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

- **Inference-Time Alignment Methods**
- **Training-Based Alignment Methods**
- **Multi-Objective Optimization Frameworks**
- **Domain-Specific Applications**

Complete Taxonomy Tree

- Inference-Time Multi-Preference Alignment for Diffusion Models Survey Taxonomy
- Inference-Time Alignment Methods
  - Reward-Guided Inference-Time Alignment ★ (5 papers)
  - [0] Diffusion Blend: Inference-Time Multi-Preference Alignment for Diffusion Models (Anon et al., 2026) View paper
  - [1] Test-time alignment of diffusion models without reward over-optimization (Kim Sun-woo, 2025) View paper
  - [2] Reward-guided controlled generation for inference-time alignment in diffusion models: Tutorial and review (Masatoshi Uehara, 2025) View paper
  - [15] DyMO: Training-Free Diffusion Model Alignment with Dynamic Multi-Objective Scheduling (Xin Xie, 2025) View paper
  - [16] MIRA: Towards Mitigating Reward Hacking in Inference-Time Alignment of T2I Diffusion Models (Singh, 2025) View paper
  - Test-Time Preference Adaptation (3 papers)
  - [4] Steerable adversarial scenario generation through test-time preference alignment (Nie, 2025) View paper
  - [12] Instant Preference Alignment for Text-to-Image Diffusion Models (Li Yang, 2025) View paper
  - [30] Test-Time Preference Optimization for Image Restoration (Bingchen Li, 2025) View paper
- Training-Based Alignment Methods
  - Direct Preference Optimization for Diffusion (6 papers)
  - [11] Calibrated Multi-Preference Optimization for Aligning Diffusion Models (Kyungmin Lee, 2025) View paper
  - [17] PC-Diffusion: Aligning Diffusion Models with Human Preferences via Preference Classifier (Shaomeng Wang, 2025) View paper
  - [21] Diffusion-SDPO: Safeguarded Direct Preference Optimization for Diffusion Models (Fu, 2025) View paper
  - [39] DETONATE: A Benchmark for Text-to-Image Alignment and Kernelized Direct Preference Optimization (Prasad, 2025) View paper
  - [40] Fine-Tuning Diffusion Generative Models via Rich Preference Optimization (Zhao, 2025) View paper
  - [43] A Gradient Guidance Perspective on Stepwise Preference Optimization for Diffusion Models (JTJ Tee, n.d.) View paper
  - Multi-Dimensional Preference Alignment (4 papers)
  - [23] Multi-dimensional Preference Alignment by Conditioning Reward Itself (Jiho Jang, 2025) View paper
  - [24] MapReduce LoRA: Advancing the Pareto Front in Multi-Preference Optimization for Generative Models (Chieh-Yun Chen, 2025) View paper
  - [33] PARM: Multi-Objective Test-Time Alignment via Preference-Aware Autoregressive Reward Model (Lin, 2025) View paper
  - [37] Robust Multi-Objective Preference Alignment with Online DPO (Gupta, 2025) View paper
  - Video and Temporal Preference Alignment (2 papers)

- [6] SynCast: Synergizing Contradictions in Precipitation Nowcasting via Diffusion Sequential Preference Optimization (Xu Kaiyi, 2025) View paper
- [25] OnlineVPO: Align Video Diffusion Model with Online Video-Centric Preference Optimization (Zhang Jiacheng, 2024) View paper

- Multi-Objective Optimization Frameworks
  - Pareto-Guided Diffusion Generation (3 papers)
  - [13] Distributional Multi-objective Black-box Optimization for Diffusion-model Inference-time Multi-Target Generation (Tan, 2025) View paper
  - [19] PROUD: PaRetO-gUided diffusion model for multi-objective generation (Yinghua Yao, 2024) View paper
  - [38] Preference-Guided Diffusion for Multi-Objective Offline Optimization (Annadani, 2025) View paper
  - RL-Based Multi-Objective Alignment (3 papers)
  - [3] Uncertainty-aware multi-objective reinforcement learning-guided diffusion models for 3D de novo molecular design (Chen Lianghong, 2025) View paper
  - [7] Evolutionary training-free guidance in diffusion model for 3D multi-objective molecular generation (R Sun, 2025) View paper
  - [8] Expensive Multi-Objective Bayesian Optimization Based on Diffusion Models (Li Bingdong, 2025) View paper
  - Evolutionary Multi-Objective Optimization (2 papers)
  - [35] Open-Ended Evolution of Artistic Styles in Diffusion Models via Island-Based Genetic Algorithms (Marcel Salvenmoser, 2025) View paper
  - [42] EmoDM: A Diffusion Model for Evolutionary Multi-objective Optimization (Yan Xue-ming, 2024) View paper
- Domain-Specific Applications
  - Molecular and Drug Design (4 papers)
  - [5] Training-free multi-objective diffusion model for 3d molecule generation (X Han, 2023) View paper
  - [10] PILOT: equivariant diffusion for pocket-conditioned de novo ligand generation with multi-objective guidance via importance sampling (Julian Cremer, 2024) View paper
  - [20] BoKDiff: Best-of-K Diffusion Alignment for Target-Specific 3D Molecule Generation (Ali Khodabandeh Yalabadi, 2025) View paper
  - [29] PMODiff: Physics-Informed Multi-Objective Optimization Diffusion Model for Protein-Specific 3D Molecule Generation. (Yaoxiang Zhang, 2025) View paper
  - Engineering and Physical Design (3 papers)
  - [9] Airfoil-DDPM: A flexible airfoil generative design method using a multi-objective sampling based diffusion model (Zhe Wen, 2025) View paper
  - [14] Multi-objective evolutionary design of microstructures using diffusion autoencoders (A Suresh, 2023) View paper
  - [28] A generative diffusion model enables multi-objective on-demand inverse design of piezoelectric metamaterials (Chun-Yu Lei, 2025) View paper
  - Sequential Decision-Making Applications (7 papers)
  - [18] Continuous Alignment of Multi-Target Preferences via Instructed Diffusion Model (Y Zhao, 2025) View paper
  - [22] Alignddiff: Aligning diverse human preferences via behavior-customisable diffusion model (Zibin Dong, 2023) View paper
  - [26] Regularized conditional diffusion model for multi-task preference alignment (Yu, 2024) View paper
  - [27] Aligning diffusion behaviors with q-functions for efficient continuous control (Chen Huayu, 2024) View paper
  - [31] MODULI: Unlocking Preference Generalization via Diffusion Models for Offline Multi-Objective Reinforcement Learning (Yuan, 2024) View paper
  - [36] Toward Diffusion-Based Deep Reinforcement Learning for Discrete Decision-Making: Methods and Evaluations (Zhen Chen, 2025) View paper
  - [41] Aligning Machiavellian Agents: Behavior Steering via Test-Time Policy Shaping (Dena Mujtaba, 2025) View paper
  - Specialized Domain Applications (4 papers)
  - [32] Multi-Objective Aerial Collaborative Secure Communication Optimization via Generative Diffusion Model-Enabled Deep Reinforcement Learning (Chuang Zhang, 2024) View paper
  - [34] Aligning Generative Music AI with Human Preferences: Methods and Challenges (Dorien Herremans, 2025) View paper
  - [44] UNLEASHING 2D REWARDS FOR HUMAN PREFER-ENCE ALIGNED TEXT-TO-3D GENERATION VIA PREFERENCE SCORE DISTILLATION (Trellis, n.d.) View paper
  - [45] INFERENCE-TIME ALIGNMENT CONTROL FOR DIFFU (GUIDANCE, n.d.) View paper

Narrative

Core task: inference-time multi-preference alignment for diffusion models. The field addresses how to steer diffusion models toward multiple, potentially conflicting objectives without retraining. The taxonomy organizes work into four main branches. Inference-Time Alignment Methods focus on guiding generation during sampling, often using reward signals or gradient-based steering to balance competing preferences on the fly. Training-Based Alignment Methods instead modify model weights or learn auxiliary modules to encode preferences, trading flexibility for potentially stronger alignment. Multi-Objective Optimization Frameworks provide algorithmic tools—such as Pareto front approximation or scalarization strategies—that can be applied at either training or inference time. Domain-Specific Applications demonstrate these techniques in specialized contexts like molecular design, materials engineering, and image restoration, where multiple design criteria must be satisfied simultaneously.

A particularly active line of work explores reward-guided inference-time steering, where methods like Test-time Alignment[1] and Reward-guided Tutorial[2] adjust sampling trajectories using differentiable reward models. These approaches contrast with training-free multi-objective frameworks such as Training-free Multi-objective[5], which combine multiple objectives without gradient-based guidance. Diffusion Blend[0] sits squarely within the reward-guided inference-time cluster, emphasizing dynamic blending of multiple reward signals during generation. Compared to DyMO[15] and MIRA[16]—neighbors that also tackle multi-preference scenarios—Diffusion Blend[0] focuses on flexible, on-the-fly composition rather than pre-trained preference encodings. A central open question across these branches is how to efficiently navigate trade-offs among conflicting objectives while maintaining sample quality, especially when the number of preferences scales or when domain-specific constraints arise.

Related Works in Same Category

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

1. Test-time alignment of diffusion models without reward over-optimization

Authors: Kim Sun-woo, Kim Min-Kyu, Park Dongmin | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Diffusion models excel in generative tasks, but aligning them with specific objectives while maintaining their versatility remains challenging. Existing fine-tuning methods often suffer from reward over-optimization, while approximate guidance approaches fail to optimize target rewards effectively. Addressing these limitations, we propose a training-free, test-time method based on Sequential Monte Carlo (SMC) to sample from the reward-aligned target distribution. Our approach, tailored for diffu...

Relationship Analysis

Both papers belong to the reward-guided inference-time alignment category, using reward models to steer diffusion generation at inference time without additional fine-tuning. They overlap in addressing multi-preference alignment and avoiding reward over-optimization through inference-time methods. However, the original paper (Diffusion Blend) blends backward diffusion processes from multiple fine-tuned models using linear combinations of rewards and KL regularization control, while the candidate paper uses Sequential Monte Carlo (SMC) sampling with tempering to sample from reward-aligned target distributions, focusing on particle-based guidance rather than model blending.

---

2. Reward-guided controlled generation for inference-time alignment in diffusion models: Tutorial and review

Authors: Masatoshi Uehara, Yulai Zhao, Chenyu Wang, Xiner Li, Aviv Regev, et al. (7 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

This tutorial provides an in-depth guide on inference-time guidance and alignment methods for optimizing downstream reward functions in diffusion models. While diffusion models are renowned for their generative modeling capabilities, practical applications in fields such as biology often require sample generation that maximizes specific metrics (e.g., stability, affinity in proteins, closeness to target structures). In these scenarios, diffusion models can be adapted not only to generate realist...

Relationship Analysis

Both papers belong to the reward-guided inference-time alignment category, focusing on techniques that use reward models or guidance signals to steer diffusion model generation toward desired objectives without fine-tuning. The candidate paper provides a comprehensive tutorial and review of inference-time guidance methods including SMC-based guidance, value-based sampling, and classifier guidance, while the original paper (Diffusion Blend) proposes a specific novel algorithm that blends backward diffusion processes from multiple fine-tuned models to achieve multi-preference alignment at inference time. The key difference is that the candidate is a survey/tutorial covering the broader landscape of inference-time techniques, whereas the original presents a new method (DB-MPA, DB-KLA, DB-MPA-LS) for combining multiple reward objectives through diffusion process blending.

---

3. DyMO: Training-Free Diffusion Model Alignment with Dynamic Multi-Objective Scheduling

Authors: Xin Xie, Dong Gong | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Text-to-image diffusion model alignment is critical for improving the alignment between the generated images and human preferences. While training-based methods are constrained by high computational costs and dataset requirements, training-free alignment methods remain underexplored and are often limited by inaccurate guidance. We propose a plug-and-play training-free alignment method, DyMO, for aligning the generated images and human preferences during inference. Apart from text-aware human pre...

Relationship Analysis

Both papers belong to the Reward-Guided Inference-Time Alignment category, using reward models or guidance signals to steer diffusion model generation toward desired objectives at inference time. They overlap in addressing multi-preference alignment through inference-time guidance mechanisms that avoid additional fine-tuning. However, the original paper (Diffusion Blend) blends backward diffusion processes from multiple fine-tuned models to achieve linear combinations of rewards, while the candidate paper (DyMO) uses dynamic scheduling of multiple objectives (semantic alignment and preference scores) with attention map manipulation and does not require pre-fine-tuned models for each reward.

---

4. MIRA: Towards Mitigating Reward Hacking in Inference-Time Alignment of T2I Diffusion Models

Authors: Singh, Utsav, Kevin Zhai, Thatipelli, Anirudh, et al. (20 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Diffusion models excel at generating images conditioned on text prompts, but the resulting images often do not satisfy user-specific criteria measured by scalar rewards such as Aesthetic Scores. This alignment typically requires fine-tuning, which is computationally demanding. Recently, inference-time alignment via noise optimization has emerged as an efficient alternative, modifying initial input noise to steer the diffusion denoising process towards generating high-reward images. However, this...

Relationship Analysis

Both papers belong to the Reward-Guided Inference-Time Alignment category, using reward models or guidance signals to steer diffusion model generation at inference time. They overlap in addressing inference-time alignment for diffusion models with reward functions, but differ fundamentally in their approach: the original paper (Diffusion Blend) blends backward diffusion processes from multiple fine-tuned models to enable multi-preference alignment without additional fine-tuning, while the candidate paper (MIRA) focuses on mitigating reward hacking during noise optimization by introducing an image-space KL regularization constraint to prevent distributional drift from the base model.

Contributions Analysis

Overall novelty summary. The paper proposes Diffusion Blend, a framework for inference-time multi-preference alignment that enables dynamic composition of multiple reward functions and KL regularization strengths without additional fine-tuning. It resides in the 'Reward-Guided Inference-Time Alignment' leaf, which contains five papers total, including the original work. This leaf sits within the broader 'Inference-Time Alignment Methods' branch, indicating a moderately populated research direction focused on steering generation during sampling rather than through model retraining. The taxonomy reveals this is an active but not overcrowded area, with sibling work exploring related reward-guided and test-time adaptation strategies.

The taxonomy structure shows that Diffusion Blend's leaf is adjacent to 'Test-Time Preference Adaptation' (three papers) within the same parent branch, and neighbors the 'Training-Based Alignment Methods' branch, which includes 'Direct Preference Optimization for Diffusion' (six papers) and 'Multi-Dimensional Preference Alignment' (four papers). The scope notes clarify that inference-time methods like Diffusion Blend differ from training-based approaches by avoiding model weight updates, and from multi-objective optimization frameworks by focusing on reward-guided steering rather than Pareto-optimal solution generation. This positioning suggests the work bridges inference-time flexibility with multi-preference handling, a boundary less densely explored than single-objective training methods.

Among nine candidates examined, the 'Inference-time multi-preference alignment problem formulation' contribution shows two refutable candidates out of eight examined, indicating some prior work addresses similar problem settings within the limited search scope. The 'Diffusion Blend framework and algorithms' contribution examined one candidate with no refutations, suggesting the specific blending mechanism may be more distinctive. The 'Theoretical approximation for control term' contribution examined zero candidates, leaving its novelty unassessed by this analysis. These statistics reflect a focused search rather than exhaustive coverage, so the presence of two refutable candidates for the problem formulation does not definitively establish lack of novelty but signals overlapping prior work exists among top semantic matches.

Based on the limited search scope of nine candidates, the work appears to occupy a moderately explored niche within inference-time alignment, with the problem formulation showing some overlap with existing methods but the algorithmic approach potentially more distinctive. The taxonomy context reveals this sits in an active but not saturated research direction, with clear boundaries separating it from training-based and Pareto-optimization approaches. The analysis covers top semantic matches and does not claim exhaustive field coverage.

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

Contribution 1: Inference-time multi-preference alignment problem formulation

Description: The authors formalize a new problem where diffusion models must align with arbitrary user-specified linear combinations of multiple reward functions and varying KL regularization strengths at inference time, without additional fine-tuning. This extends beyond standard single-reward alignment to accommodate diverse and dynamic user preferences.

This contribution was assessed against **8 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Rewards-in-context: Multi-objective alignment of foundation models with dynamic preference adjustment

URL: View paper

Prior Art Analysis

Rewards-in-context[48] demonstrates that the problem of inference-time alignment with multiple reward functions was formulated and solved prior to the original paper. The candidate explicitly addresses the same core challenge: enabling models to align with arbitrary user-specified combinations of multiple rewards at inference time without additional fine-tuning. Both papers formulate this as a multi-objective alignment problem where users specify preference weights over multiple reward functions, and the model must generate outputs aligned with these preferences dynamically. The candidate's formulation in Section 2 and problem statement in Section 3 directly parallel the original paper's contribution, showing this was not a novel formulation by the original authors.

Evidence

Evidence 1 - **Rationale**: Both papers identify the same fundamental problem: multi-objective alignment with heterogeneous preferences that must be handled efficiently. - **Original**: we address the problem of inference-time multi-preference alignment: given a set of basis reward functions and a reference kl regularization strength, can we design a finetuning procedure so that, at inference time, it can generate images aligned with any user-specified linear combination of rewards... - **Candidate**: we consider the problem of multi-objective alignment of foundation models with human preferences, which is a critical step towards helpful and harmless ai systems. however, it is generally costly and unstable to fine-tune large foundation models using reinforcement learning (rl), and the multi-dimen...

Evidence 2 - **Rationale**: Both papers explicitly formulate the problem as requiring inference-time adaptation to user-specified preference weights without additional fine-tuning, demonstrating this problem formulation existed prior to the original work. - **Original**: these limitations motivate the need for a more flexible approach:inference-time multi-preference alignment, where the user specifies their preference vector, i.e., weights over a set of basis reward functions such as alignment, aesthetics, or human preference, along with a desired regularization str... - **Candidate**: in this paper, we introduce rewards-in-context (ric), which conditions the response of a foundation model on multiple rewards in its prompt context and applies supervised fine-tuning for alignment. the salient features of ric are simplicity and adaptivity, as it only requires supervised fine-tuning ...

Evidence 3 - **Rationale**: Both papers formulate the exact same mathematical problem: given basis reward functions and user preference weights w, generate outputs aligned with linear combinations of rewards at inference time without retraining. - **Original**: given a set of basis reward functions(ri)m i=1 and a basis kl regularization weightα, can we design a fine-tuning procedure such that when the user specifies their reward or regularization preferences through parameters w and λ at inference time, the model generates images aligned with the linear re... - **Candidate**: during the inference stage, users assign preferences w = [w1, . . . , wn] for different reward dimensions r1, . . . , rn . to adjust the llm policy according to the user preferences w, it is necessary to map these preferences w to the desired rewards that will be used as conditionings in prompts.

---

2. A general framework for inference-time scaling and steering of diffusion models

URL: View paper

Brief Assessment

Inference-Time Scaling[47] focuses on steering diffusion models with reward functions at inference time using particle-based methods, but does not address the specific problem of aligning with arbitrary linear combinations of multiple reward functions and varying KL regularization strengths without fine-tuning, which is the core novelty claim of the original paper.

---

3. INFERENCE-TIME DIFFUSION MODEL ALIGNMENT VIA RANDOM ORDINARY EQUATIONS

URL: View paper

Brief Assessment

Random Ordinary Equations[51] focuses on inference-time alignment via trajectory search using random ODEs and MCTS for single reward functions, not on multi-preference alignment with arbitrary linear combinations of multiple rewards and varying KL regularization strengths.

---

4. Test-time alignment of diffusion models without reward over-optimization

URL: View paper

Brief Assessment

Test-time Alignment[1] focuses on single-reward alignment at test-time using Sequential Monte Carlo sampling, not multi-preference alignment with arbitrary linear combinations of multiple rewards as formulated in the original paper.

---

5. PARM: Multi-Objective Test-Time Alignment via Preference-Aware Autoregressive Reward Model

URL: View paper

Prior Art Analysis

PARM[33] demonstrates that the problem of inference-time alignment with multiple reward functions was previously addressed in the context of language models. The candidate paper explicitly formulates multi-objective test-time alignment where models adapt to diverse multi-dimensional user preferences during inference while keeping the base model frozen, using preference vectors to specify weights

over multiple reward dimensions. This directly parallels the original paper's formulation of inference-time multi-preference alignment for diffusion models, showing that the core concept of dynamic preference-based alignment at inference time without retraining existed prior to the original work.

Evidence

Evidence 1 - **Rationale**: Both papers formulate the same core problem: enabling models to align with user-specified combinations of multiple objectives at inference time without additional fine-tuning. The candidate demonstrates this concept was already established for LLMs. - **Original**: we address the problem of inference-time multi-preference alignment: given a set of basis reward functions and a reference kl regularization strength, can we design a finetuning procedure so that, at inference time, it can generate images aligned with any user-specified linear combination of rewards... - **Candidate**: multi-objective test-time alignment aims to adapt large language models (llms) to diverse multidimensional user preferences during inference while keeping llms frozen.

Evidence 2 - **Rationale**: Both papers use preference vectors to represent user-specified linear combinations of multiple reward functions. The candidate's formulation with preference vectors α predates the original's formulation with weights w, demonstrating prior work on this problem structure. - **Original**: this motivates us to address the following questions: given a set of basis reward functions(ri)m i=1 and a basis kl regularization weightα, can we design a fine-tuning procedure such that when the user specifies their reward or regularization preferences through parameters w and λ at inference time,... - **Candidate**: in multi-objective alignment, users expect the llm's outputs to align with their multi-dimensional needs, which can be represented as a preference vector, α = (α1, · ··, αk) ∈ Δk-1, where αi denotes the weight for the i-th preference dimension

Evidence 3 - **Rationale**: The candidate paper addresses the same core challenge of enabling inference-time preference alignment without retraining, demonstrating that this problem formulation existed in prior work on language models. - **Original**: these limitations motivate the need for a more flexible approach:inference-time multi-preference alignment, where the user specifies their preference vector, i.e., weights over a set of basis reward functions such as alignment, aesthetics, or human preference, along with a desired regularization str... - **Candidate**: to address these limitations, we aim to jointly train a single arm across all preferences by optimizing the following multi-objective optimization problem, min θ (ℓ(πθ, d1), · ··, ℓ(πθ, dk))⊤

---

6. Effective Test-Time Scaling of Discrete Diffusion through Iterative Refinement

URL: View paper

Brief Assessment

Iterative Refinement[49] focuses on test-time scaling for discrete diffusion models through iterative noising-denoising refinement, not on multi-preference alignment with arbitrary linear combinations of reward functions at inference time without fine-tuning.

---

7. Steerable adversarial scenario generation through test-time preference alignment

URL: View paper

Brief Assessment

Steerable Adversarial[4] addresses test-time preference alignment for adversarial scenario generation in autonomous driving, not general diffusion model alignment with multiple reward functions. The technical domains and applications are fundamentally different.

---

8. GenARM: Reward Guided Generation with Autoregressive Reward Model for Test-time Alignment

URL: View paper

Brief Assessment

GenARM[50] focuses on test-time alignment for LLMs using autoregressive reward models for next-token generation, while the original paper addresses inference-time multi-preference alignment for diffusion models with linear combinations of multiple reward functions. These are fundamentally different model architectures and generation paradigms.

Contribution 2: Diffusion Blend framework and algorithms

Description: The authors introduce Diffusion Blend, a principled method that blends backward diffusion trajectories from reward-specific fine-tuned models. They propose three concrete algorithms: DB-MPA enables multi-reward alignment, DB-KLA provides KL regularization control, and DB-MPA-LS achieves similar performance without extra inference overhead.

This contribution was assessed against **1 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Aligning Text-to-Image Diffusion Models with Reward Backpropagation

URL: View paper

Brief Assessment

Reward Backpropagation[46] focuses on end-to-end backpropagation through the denoising process for single reward optimization, not on blending backward diffusion trajectories from multiple reward-specific models for inference-time multi-preference alignment.

Contribution 3: Theoretical approximation for control term in backward diffusion

Description: The authors derive a theoretical result showing that the backward diffusion for any reward combination can be expressed via a control term, and they propose an approximation that decomposes this term into contributions from basis reward models. This enables blending of fine-tuned models to achieve arbitrary preference alignment without retraining.

This contribution was assessed against **0 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

Appendix: Text Similarity Detection

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

References

- [0] Diffusion Blend: Inference-Time Multi-Preference Alignment for Diffusion Models View paper
- [1] Test-time alignment of diffusion models without reward over-optimization View paper
- [2] Reward-guided controlled generation for inference-time alignment in diffusion models: Tutorial and review View paper
- [3] Uncertainty-aware multi-objective reinforcement learning-guided diffusion models for 3D de novo molecular design View paper
- [4] Steerable adversarial scenario generation through test-time preference alignment View paper
- [5] Training-free multi-objective diffusion model for 3d molecule generation View paper
- [6] SynCast: Synergizing Contradictions in Precipitation Nowcasting via Diffusion Sequential Preference Optimization View paper
- [7] Evolutionary training-free guidance in diffusion model for 3D multi-objective molecular generation View paper
- [8] Expensive Multi-Objective Bayesian Optimization Based on Diffusion Models View paper

- [9] Airfoil-DDPM: A flexible airfoil generative design method using a multi-objective sampling based diffusion model View paper
- [10] PILOT: equivariant diffusion for pocket-conditioned de novo ligand generation with multi-objective guidance via importance sampling View paper
- [11] Calibrated Multi-Preference Optimization for Aligning Diffusion Models View paper
- [12] Instant Preference Alignment for Text-to-Image Diffusion Models View paper
- [13] Distributional Multi-objective Black-box Optimization for Diffusion-model Inference-time Multi-Target Generation View paper
- [14] Multi-objective evolutionary design of microstructures using diffusion autoencoders View paper
- [15] DyMO: Training-Free Diffusion Model Alignment with Dynamic Multi-Objective Scheduling View paper
- [16] MIRA: Towards Mitigating Reward Hacking in Inference-Time Alignment of T2I Diffusion Models View paper
- [17] PC-Diffusion: Aligning Diffusion Models with Human Preferences via Preference Classifier View paper
- [18] Continuous Alignment of Multi-Target Preferences via Instructed Diffusion Model View paper
- [19] PROUD: PaRetO-gUided diffusion model for multi-objective generation View paper
- [20] BoKDiff: Best-of-K Diffusion Alignment for Target-Specific 3D Molecule Generation View paper
- [21] Diffusion-SDPO: Safeguarded Direct Preference Optimization for Diffusion Models View paper
- [22] Aligndiff: Aligning diverse human preferences via behavior-customisable diffusion model View paper
- [23] Multi-dimensional Preference Alignment by Conditioning Reward Itself View paper
- [24] MapReduce LoRA: Advancing the Pareto Front in Multi-Preference Optimization for Generative Models View paper
- [25] OnlineVPO: Align Video Diffusion Model with Online Video-Centric Preference Optimization View paper
- [26] Regularized conditional diffusion model for multi-task preference alignment View paper
- [27] Aligning diffusion behaviors with q-functions for efficient continuous control View paper
- [28] A generative diffusion model enables multi-objective on-demand inverse design of piezoelectric metamaterials View paper
- [29] PMODiff: Physics-Informed Multi-Objective Optimization Diffusion Model for Protein-Specific 3D Molecule Generation. View paper
- [30] Test-Time Preference Optimization for Image Restoration View paper
- [31] MODULI: Unlocking Preference Generalization via Diffusion Models for Offline Multi-Objective Reinforcement Learning View paper
- [32] Multi-Objective Aerial Collaborative Secure Communication Optimization via Generative Diffusion Model-Enabled Deep Reinforcement Learning View paper
- [33] PARM: Multi-Objective Test-Time Alignment via Preference-Aware Autoregressive Reward Model View paper
- [34] Aligning Generative Music AI with Human Preferences: Methods and Challenges View paper
- [35] Open-Ended Evolution of Artistic Styles in Diffusion Models via Island-Based Genetic Algorithms View paper
- [36] Toward Diffusion-Based Deep Reinforcement Learning for Discrete Decision-Making: Methods and Evaluations View paper
- [37] Robust Multi-Objective Preference Alignment with Online DPO View paper
- [38] Preference-Guided Diffusion for Multi-Objective Offline Optimization View paper
- [39] DETONATE: A Benchmark for Text-to-Image Alignment and Kernelized Direct Preference Optimization View paper
- [40] Fine-Tuning Diffusion Generative Models via Rich Preference Optimization View paper
- [41] Aligning Machiavellian Agents: Behavior Steering via Test-Time Policy Shaping View paper
- [42] EmoDM: A Diffusion Model for Evolutionary Multi-objective Optimization View paper
- [43] A Gradient Guidance Perspective on Stepwise Preference Optimization for Diffusion Models View paper
- [44] UNLEASHING 2D REWARDS FOR HUMAN PREFER-ENCE ALIGNED TEXT-TO-3D GENERATION VIA PREFERENCE SCORE DISTILLATION View paper
- [45] INFERENCE-TIME ALIGNMENT CONTROL FOR DIFFU View paper
- [46] Aligning Text-to-Image Diffusion Models with Reward Backpropagation View paper
- [47] A general framework for inference-time scaling and steering of diffusion models View paper
- [48] Rewards-in-context: Multi-objective alignment of foundation models with dynamic preference adjustment View paper
- [49] Effective Test-Time Scaling of Discrete Diffusion through Iterative Refinement View paper
- [50] GenARM: Reward Guided Generation with Autoregressive Reward Model for Test-time Alignment View paper
- [51] INFERENCE-TIME DIFFUSION MODEL ALIGNMENT VIA RANDOM ORDINARY EQUATIONS View paper