Joint Alignment and Distillation for Video Generation via Sample-Guided Distribution Matching

Jiuzhou Lin[1,2†], Junlong Wu[1,2†], Fei Zuo[2,3], Huan Ouyang[2,4], Dewen Fan[2], Boheng Zhang[2], Huaiqing Wang[2], Jia Sun[2], Fan Yang[2*], Houde Liu[1*], Kehai Chen[3], Min Zhang[3], Tingting Gao[2], and Han Li[2]

[1] Tsinghua University
[2] Kuaishou Technology
[3] Harbin Institute of Technology (ShenZhen)
[4] Beijing University of Posts and Telecommunications
[†] Equal contribution. [*] Corresponding authors
✉ {lin-jz24, wu-jl24}@mails.tsinghua.edu.cn

Abstract. Aligning video generative models to human preferences heavily relies on Reinforcement Learning (RL), which suffers from extensive computational overhead. Existing workflows typically treat RL and distillation as disconnected stages: applying RL before distillation incurs prohibitive computational costs, whereas applying RL after distillation frequently leads to model collapse. To overcome these limitations, we propose a unified, single-stage optimization framework grounded in Distribution Matching (DM). In the standard DM framework, distillation updates the model via a gradient direction that minimizes the gap between the real and fake models, guiding generations toward clarity and high fidelity. Building upon this, we introduce DM-Align, which derives a complementary gradient direction to guide the model toward human-preferred samples. Inspired by DPO and GRPO, our method leverages the distributional gap—formulated from either preference pairs or intra-group exploration—to directly construct this preference-guided gradient. By synergizing these two gradient directions, our approach eliminates the need for multi-step reward evaluation and complex ODE-SDE conversions inherent in traditional RL. Comprehensive experiments across multiple foundational video models demonstrate that this sample-guided framework robustly enhances both distillation quality and preference alignment, consistently outperforming both standalone variants and sequential two-stage pipelines.

Keywords: Video Generation · Distribution Matching · Preference Alignment

1 Introduction

Recent advances in visual generative modeling, primarily driven by diffusion models [12, 42] and flow matching models [23, 27], have fundamentally transformed image and video generation. Both open-source models [17, 36, 47] and

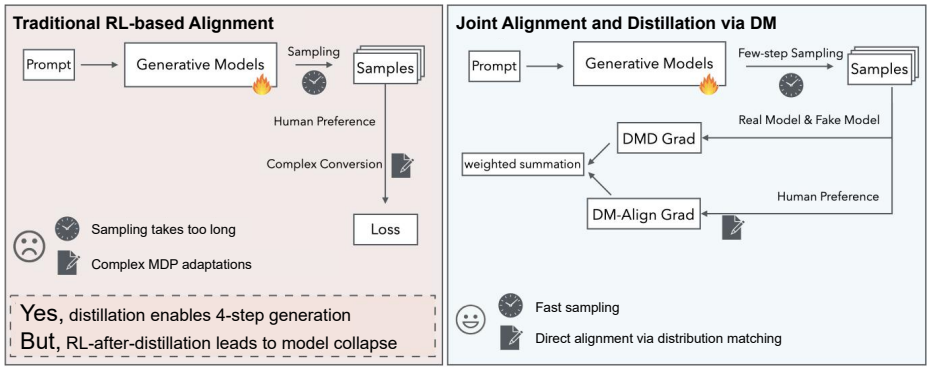

Fig. 1: *Left:* Applying RL before distillation incurs prohibitive multi-step sampling costs and complex MDP adaptations, while applying RL after distillation causes unstable training. *Right:* Our unified Sample-Guided Distribution Matching framework jointly optimizes both objectives in a single stage, enabling fast, high-fidelity preference alignment while entirely bypassing the bottlenecks of traditional RL.

proprietary commercial systems [10, 18] can now generate high-quality images or videos. This progress has established a research paradigm mirroring that of large language models: leveraging open-source foundational models to develop and validate new algorithms. In this work, we focus on two key research directions in diffusion models: 1) *Distillation* [3, 54, 55], which reduces generation time while maintaining output quality, enabling few-step generation comparable to the original multi-step base models; and 2) *Alignment* [19, 25, 51], which fine-tunes models to align with human preferences or aesthetic judgments, often leveraging Reinforcement Learning (RL) techniques like Direct Preference Optimization (DPO) [37] and Group Relative Policy Optimization (GRPO) [41].

As illustrated in Figure 1 (Left), existing workflows typically treat distillation and alignment as disconnected, sequential stages, which inevitably leads to a dilemma. On one hand, applying RL *before* distillation [5] incurs prohibitive computational costs, as obtaining generated samples requires extensive multi-step sampling. More fundamentally, existing RL-based methods adapt the diffusion sampling process into the Markov Decision Process (MDP) formulation of RL algorithms. This paradigm requires calculating trajectory probabilities along the sampling path, necessitating complex conversions into reverse-time Stochastic Differential Equations (SDEs) [1, 44] and reliance on Gaussian distribution assumptions. On the other hand, applying RL *after* distillation to leverage few-step generation [31, 54, 55] is unstable. Fine-tuning a highly compressed distilled model with sparse RL losses frequently destroys the delicate continuous generative mapping, leading to visual artifacts (as illustrated in Figure 5).

To overcome these limitations, as shown in Figure 1 (Right), we shift the paradigm from forcing diffusion into an RL framework to unifying both objectives within a joint **Distribution Matching (DM)** framework. From a theoretical perspective, we observe a gradient compatibility between distillation and

alignment under the DM objective: both objectives can be expressed as score-space distribution-matching gradients and can therefore be optimized within a shared computational pathway. In the standard DM framework, distillation updates the model via a gradient direction that minimizes the gap between the teacher (real) and student (fake) models, guiding generations toward clarity and high fidelity. In a parallel manner, alignment can be formulated by measuring the distributional gap between ordinary generated samples and human-preferred samples. This yields a complementary gradient direction that naturally shifts the model's overall distribution toward optimal structural preferences.

Motivated by this gradient compatibility, we propose a unified, single-stage optimization framework, introducing **DM-Align**. While DM-Align leverages preference samples inspired by RL (such as win-lose pairs from DPO or intra-group exploration from GRPO), it operates fundamentally as a *sample-guided DM loss*. By directly deriving preference-guided gradients based on these sample disparities, our method entirely bypasses the expensive multi-step generation and eliminates the need for complex mathematical adaptations. The distillation gradient and the alignment gradient are synergistically aggregated within the same latent space, guiding the generator toward both high fidelity and human preference without the risk of post-distillation collapse.

To summarize, our key contributions are the following:

- We identify the gradient compatibility between distillation and preference alignment under the Distribution Matching objective. Based on this, we propose a unified, single-stage framework that optimizes both objectives simultaneously, successfully avoiding the prohibitive computational costs of RL-before-distillation and the catastrophic model collapse inherent in RL-after-distillation.
- We introduce **DM-Align**, a sample-guided approach that entirely bypasses the need to adapt the diffusion sampling process into an RL framework. Specifically, we develop **DM-PairLoss** and **DM-GroupLoss** as concrete implementations inspired by DPO and GRPO, directly utilizing preference pairs and group rewards to construct efficient gradient guidance.
- We conduct comprehensive experiments across multiple foundational video models (e.g., Wan 2.1 T2V-1.3B [47] and CogVideoX [53]) and varying reward models (e.g., VideoAlign [25], HPSv2 [50]). Our method demonstrates exceptional generalizability and substantially outperforms strong baselines, achieving significant gains in VBench scores (up to +6.55) and human evaluations (up to +48% net preference) while exhibiting remarkable training efficiency.

2 Related Work

2.1 Aligning Diffusion Models and Flow Matching Models

Motivated by the success of Reinforcement Learning from Human Feedback (RLHF) [34] in large language models, a significant line of work adapts RL algorithms to align visual generation models with human preferences. These methods

primarily fall into two categories: DPO-based approaches [20, 25, 37, 46, 52, 58] that rely on annotated positive/negative sample pairs, and PPO/GRPO-based methods [2, 9, 19, 24, 39, 51] that utilize reward models for optimization without paired data.

Although RL algorithms for image generation are theoretically transferable to video tasks, practical implementations often encounter severe training instability and convergence difficulties. To address video-specific alignment, methods such as Flow-GRPO [24] and DanceGRPO [51] have emerged. While achieving competitive results, integrating these models into standard RL frameworks introduces inherent complexities. Specifically, adapting continuous ODE-based generative trajectories into the discrete Markov Decision Process (MDP) formulation requires complex reverse-time SDE conversions and explicit trajectory probability estimations [44]. Practically, computing RL losses requires extensive sampling steps (typically 20-40), incurring non-trivial computational overhead. Concurrent work MixGRPO [19] attempts to mitigate this by enhancing sampling efficiency through SDE-ODE alternation. In contrast, we adopt a novel Distribution Matching (DM) perspective, completely circumventing these cumbersome MDP adaptations and multi-step sampling requirements, paving the way to seamlessly unify alignment with highly efficient distillation.

2.2 Diffusion Distillation

Accelerating the multi-step inference of visual generative models [12,23] remains a fundamental challenge. Existing acceleration techniques can be broadly categorized into advanced samplers [16, 26, 59], trajectory distillation [29, 38, 43], and distribution distillation [22, 54, 55]. Among these, Distribution Matching Distillation (DMD) [54, 55] has emerged as a highly effective paradigm, alternately optimizing real and fake models to minimize the KL divergence between generated and target distributions. While initially designed for image synthesis, DMD-based frameworks have been validated by the community as exceptionally stable and effective for video tasks [6, 21].

Inspired by DMD, recent works [11, 31, 40, 45] have proposed various enhancements primarily to improve distillation efficiency. Closely related to our objective are methods like Diff-Instruct++ [30] and concurrent works such as FlashDMD [4] and DMDR [15], which also explore the joint optimization of distillation and RL. However, these approaches typically perform a loss addition, persistently adapting the diffusion sampling process into standard RL algorithms to compute off-the-shelf RL losses. By exploiting the score-space gradient compatibility of distribution matching, we design sample-guided DM losses that formulate preference alignment natively within the DM framework, rather than adding off-the-shelf RL objectives to a distillation loss. Furthermore, unlike these image-centric concurrent efforts, our framework successfully tackles the more complex dynamics of video generation.

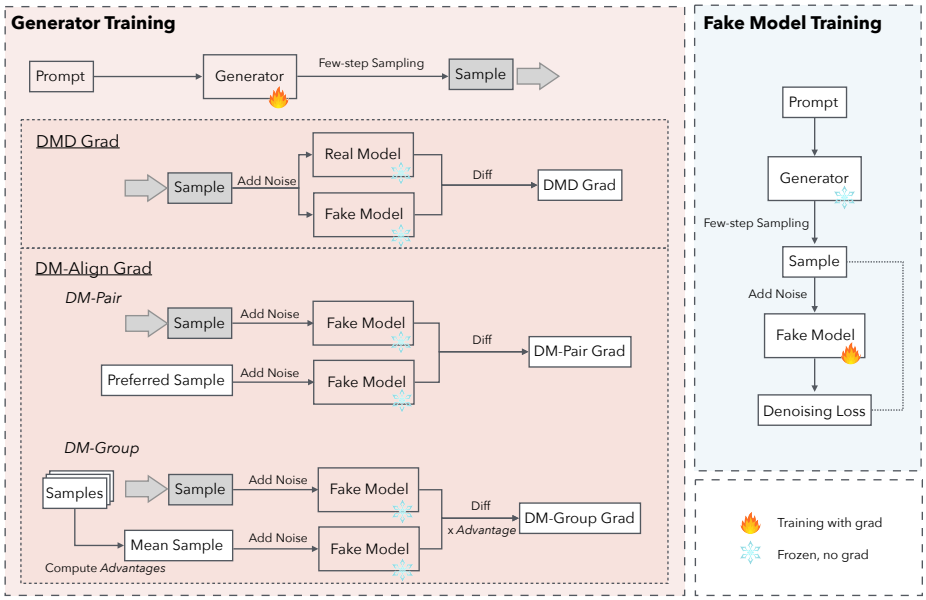

Fig. 2: Overview of our unified framework. *Left:* The generator produces samples via few-step sampling and is updated simultaneously by the distillation gradient and the DM-Align gradient. *Right:* The fake model approximates the generator's current distribution via a standard denoising loss, acting as a robust score estimator for both objectives.

3 Method

In Section 3.1, we briefly review Distribution Matching Distillation (DMD). In Section 3.2, we derive preference alignment as a score-space distribution-matching gradient and show its compatibility with the DMD gradient. Finally, Section 3.3 details our unified, single-stage framework that simultaneously achieves highly efficient distillation and preference alignment.

3.1 Review of Distribution Matching Distillation

Our approach builds upon the foundational Distribution Matching Distillation (DMD) framework [13, 54–56]. The core objective of DMD is to minimize the Kullback-Leibler (KL) divergence between the distilled generator's output distribution (p_{fake}) and the original teacher model's distribution (p_{real}):

$$\mathcal{L}_{\text{DMD}} = D_{\text{KL}}(p_{\text{fake}} \parallel p_{\text{real}}) = \mathbb{E}_{\substack{z\sim\mathcal{N}(0,\mathbf{I}) \\ x=G_\theta(z)}} \left[-(\log p_{\text{real}}(x) - \log p_{\text{fake}}(x))\right]. \quad (1)$$

Since directly estimating probability densities is intractable, DMD optimizes the generator parameters θ by computing the gradient of this divergence. This

Algorithm 1 Unified Single-Stage Optimization for Distillation and Alignment

Require: Pretrained model μ_{real}, conditional input $\mathcal{D} = \{\mathbf{c}\}$, preference data $\mathcal{D}_{\text{pref}}$.
Require: Few-step timesteps $\mathcal{T} = \{\tau_1, \ldots, \tau_Q\}$, update ratio N, group size K.
$G_\theta \leftarrow \texttt{copyWeights}(\mu_{\text{real}})$; $\mu_{\text{fake}} \leftarrow \texttt{copyWeights}(\mu_{\text{real}})$
$iter \leftarrow 1$
while training **do**
  Sample batch $\mathbf{z} \sim \mathcal{N}(0, I)^B$ and $\mathbf{c} \sim \mathcal{D}$
  Sample distillation timestep $\tau \sim \text{Uniform}(\mathcal{T})$
  $\mathbf{x}, \mathbf{x}_{\text{final}} \leftarrow \texttt{generateSamples}(G_\theta, \mathbf{z}, \mathbf{c}, \tau)$ ▷ Shared generation step
  if $iter \pmod{N+1} \neq 0$ **then** ▷ Phase 1: Update Fake Model (N steps)
    Sample denoising timestep $t \sim \mathcal{U}(0, 1)$ and noise $\epsilon \sim \mathcal{N}(0, I)$
    $\mathbf{x}_t \leftarrow \alpha_t \cdot \texttt{stopgrad}(\mathbf{x}) + \sigma_t \epsilon$
    $\mathcal{L}_{\text{denoise}} \leftarrow \texttt{denoisingLoss}(\mu_{\text{fake}}(\mathbf{x}_t, t), \texttt{stopgrad}(\mathbf{x}))$
    $\mu_{\text{fake}} \leftarrow \texttt{update}(\mu_{\text{fake}}, \nabla \mathcal{L}_{\text{denoise}})$
  else ▷ Phase 2: Update Generator (1 step)
    Sample denoising timestep $t \sim \mathcal{U}(0, 1)$
    $\nabla \mathcal{L}_{\text{DMD}} \leftarrow \texttt{DMDGradient}(\mu_{\text{real}}, \mu_{\text{fake}}, \mathbf{x}, t)$ ▷ Eq. 2
    if using DM-PairLoss **then**
      $\mathbf{x}_+ \leftarrow \texttt{sample}(\mathcal{D}_{\text{pref}}, \mathbf{c})$
      $\nabla \mathcal{L}_{\text{align}} \leftarrow \texttt{DM-PairGradient}(\mu_{\text{fake}}, \mathbf{x}_{\text{final}}, \mathbf{x}_+, t)$ ▷ Eq. 6
    else if using DM-GroupLoss **then**
      $\mathbf{x}_{\text{group}} \leftarrow \texttt{groupSamples}(\mathbf{x}_{\text{final}}, K)$
      $\nabla \mathcal{L}_{\text{align}} \leftarrow \texttt{DM-GroupGradient}(\mu_{\text{fake}}, \mathbf{x}_{\text{final}}, \mathbf{x}_{\text{group}}, t)$ ▷ Eq. 7
    end if
    $\nabla \mathcal{L}_G \leftarrow \lambda_{\text{DMD}} \nabla \mathcal{L}_{\text{DMD}} + \lambda_{\text{align}} \nabla \mathcal{L}_{\text{align}}$ ▷ Synergistic gradient aggregation
    $G_\theta \leftarrow \texttt{update}(G_\theta, \nabla \mathcal{L}_G)$
  end if
  $iter \leftarrow iter + 1$
end while

is achieved by perturbing data distributions with Gaussian noise to create overlapping manifolds:

$$\nabla_\theta \mathcal{L}_{\text{DMD}} \approx -\mathbb{E}_{t,z}\left[\Big(\mathbf{s}_{\text{real}}\big(F(G_\theta(z), t), t\big) - \mathbf{s}_{\text{fake}}\big(F(G_\theta(z), t), t\big)\Big)\frac{dG_\theta(z)}{d\theta}\right], \quad (2)$$

where $F(\cdot, t)$ denotes the forward diffusion process at noise level t, and $\mathbf{s}_p(\mathbf{x}_t, t) = \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$ denotes the score function of the noisy distribution p_t. As illustrated in Figure 2, the real model serves as a fixed teacher, while the fake denoising model μ_ϕ^{fake} is continuously updated to approximate the generator's output distribution and provides the score estimator $\mathbf{s}_{\text{fake}}(\cdot, t)$:

$$\mathcal{L}_\phi^{\text{denoise}} = \|\mu_\phi^{\text{fake}}(\mathbf{x}_t, t) - \mathbf{x}_0\|_2^2, \quad (3)$$

where $\mathbf{x}_0 = G_\theta(z)$ is the generated sample and $\mathbf{x}_t = F(\mathbf{x}_0, t)$ is its noisy counterpart. This framework provides a highly stable mechanism for few-step trajectory compression.

3.2 Alignment via Sample-Guided Distribution Matching

While DMD successfully minimizes the gap to the teacher model, aligning the model with human preferences requires a fundamental paradigm shift. Traditional RL approaches attempt to solve this by adapting diffusion sampling into a Markov Decision Process (MDP), necessitating complex trajectory probability estimations and extensive multi-step sampling. In contrast, we reveal a score-space gradient compatibility: preference alignment can be natively formulated within the DM framework by matching the generator distribution to a preference-optimal target distribution.

Theoretical Formulation. Following the standard RLHF framework under the Bradley-Terry model, the optimal policy p^* that maximizes expected reward while remaining anchored to a reference distribution p_{ref} has the closed-form solution [35]:

$$p^*(x) = \frac{1}{Z} p_{\text{ref}}(x) \exp\left(\frac{r(x)}{\beta}\right). \tag{4}$$

We formulate preference alignment as minimizing $D_{\text{KL}}(p_\theta \parallel p^*)$. Utilizing the reparameterization trick [28, 49], the gradient to update the generator G_θ operates entirely in the score domain:

$$\nabla_\theta \mathcal{L}_{\text{align}} \approx -\mathbb{E}_{t,z}\left[\Big(\nabla_{x_t} \log p_t^*(x_t) - \nabla_{x_t} \log p_{\theta,t}(x_t)\Big) \frac{dG_\theta(z)}{d\theta}\right]. \tag{5}$$

Approximating the Target Score. The exact target score $\nabla_{x_t} \log p_t^*(x_t)$ is mathematically intractable. However, we observe that the fake model ($\mathbf{s}_{\text{fake}}$), continuously trained by Eq. 3, acts as an exceptionally robust denoiser. Based on Tweedie's equivalence [7, 33], the score can be directly estimated from the denoising prediction. Thus, given a human-preferred sample x_+, feeding its noisy variant $F(x_+, t)$ into the fake model yields an implicit projection onto the learned high-reward manifold. This allows us to approximate the intractable target score with $\mathbf{s}_{\text{fake}}(F(\mathbf{x}_+, t), t)$, yielding our core DM-Align gradient:

$$\nabla_\theta \mathcal{L}_{\text{align}} \approx -\mathbb{E}_{t,z}\left[\Big(\mathbf{s}_{\text{fake}}\big(F(x_+, t), t\big) - \mathbf{s}_{\text{fake}}\big(F(G_\theta(z), t), t\big)\Big) \frac{dG_\theta}{d\theta}\right]. \tag{6}$$

By operating strictly in the score space, Eq. 6 entirely bypasses the need to adapt the diffusion sampling process into a traditional RL framework. Based on this, we design two concrete sample-guided implementations, which are integrated into the main training loop (Algorithm 1):

DM-PairLoss. Inspired by DPO, this variant utilizes pre-annotated positive samples (e.g., SFT data) as x_+, directly substituting them into Eq. 6. The current generator's output serves as the negative sample, establishing a straightforward contrastive gradient that pulls the generation toward the preferred anchor.

DM-GroupLoss. For reward-model-guided alignment (inspired by GRPO), we dynamically synthesize the alignment target. For a given prompt, we sample

a group of K outputs $\{x_1, \ldots, x_K\}$ from the generator and compute their rewards and advantages A_i. We designate the average of the top-k samples as the preferred anchor $\bar{x}_+$. The gradient adaptively handles the group:

$$\nabla_\theta \mathcal{L}_{\text{DM-Group}} \approx -\mathbb{E}_{t,z}\left[\sum_{i=1}^{K} A_i \Big(\mathbf{s}_{\text{fake}}\big(F(x_i,t),t\big) - \mathbf{s}_{\text{fake}}\big(F(\bar{x}_+,t),t\big)\Big)\frac{dG_\theta}{d\theta}\right]. \tag{7}$$

This formulation creates a dynamic distribution shift: poor generations ($A_i < 0$) are pulled toward the high-quality group mean $\bar{x}_+$, while superior generations ($A_i > 0$) push the distribution further toward the peak of the preference manifold.

3.3 Unified Single-Stage Optimization

As rigorously outlined in Algorithm 1, our framework completely eliminates the traditional barrier between distillation and RL. The generator is optimized via a synergistic combination of both objectives:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{DMD}}\mathcal{L}_{\text{DMD}} + \lambda_{\text{align}}\mathcal{L}_{\text{align}}. \tag{8}$$

This unification delivers immense computational benefits. Specifically, the generator performs the few-step denoising trajectory according to the schedule $\mathcal{T}$ and reuses the generated samples for both objectives. For score estimation, the DMD and DM-Align gradients are evaluated after applying forward noise at an independently sampled level $t \sim \mathcal{U}(0,1)$. Simultaneously, the generator completes the full few-step denoising process to produce the final output $\mathbf{x}_{\text{final}}$. These fully generated samples are then directly utilized as the high-quality negative inputs for DM-PairLoss or the reliable evaluation inputs for DM-GroupLoss. By seamlessly sharing the latent space and the score estimator ($\mathbf{s}_{\text{fake}}$), DM-Align leverages the acceleration of DMD to generate samples, while DMD benefits from the structural guidance of DM-Align to align with human preferences, achieving mutually enhanced video generation.

4 Experiments

4.1 Experiment Setup

We primarily employ Wan 2.1-T2V-1.3B [47] as our base model, generating 5-second videos at 16 FPS with a resolution of 832×480. We term the pair-based and group-based variants of our method (Section 3) as **DM-Align (Pair)** and **DM-Align (Group)**, respectively.

Datasets. The choice of distinct datasets is driven by the algorithmic requirements of the two variants. For DM-Align (Pair), we adopt the filtered ConsistID dataset [57] (roughly 6K samples). Because DM-Align (Pair) requires high-quality real video pairs to construct contrastive DPO gradients, this human-centric dataset serves as an ideal choice. Furthermore, real-world video data

inherently contains much more natural, fluid human motions and realistic physical dynamics compared to generated content, providing a superior upper bound for preference alignment. Conversely, because DM-Align (Group) utilizes an external reward model to score and guide the optimization process, it naturally supports the use of unannotated textual prompts. Thus, we utilize 20K highly diverse prompts from the expansive VidProM dataset [48] to broadly cover different generation scenarios and thoroughly explore the reward landscape. This strategic dataset selection aligns naturally with the paradigms established in previous RL alignment works [24, 51].

Baselines and Reward. Baselines include the raw model, standalone DMD2 [54], standalone RL (Flow-DPO [25]/DanceGRPO [51]), and a two-stage sequential pipeline (RL + DMD2). For DM-Align (Group), we observe that the VideoAlign reward model [25] tends to assign erroneously high motion/visual scores to degenerate content (e.g., pure white noise or flickering frames). To prevent training collapse for both DanceGRPO and our method, we primarily focus on Textual Alignment (TA) for the VidProM dataset. Additionally, to validate the versatility of our framework across diverse reward signals, we employ the image-based HPSv2 model [50] as an alternative reward in our sensitivity analysis (Section 4.3). Following standard conventions, the overall video reward is calculated by averaging the per-frame image scores.

Training and Evaluation. We conduct all experiments on 32 NVIDIA H100 GPUs using the Wan2.1-T2V-1.3B base model. Both the generator and the fake model are optimized with AdamW, using learning rates of 2×10^{-6} and 4×10^{-7}, respectively. The fake model $\mathbf{s}_{\text{fake}}$ is updated continuously to provide accurate score estimation, while the generator G_θ is updated every 5 steps for training stability. We use a per-GPU batch size of 1, start EMA from step 500 with a decay rate of 0.99, and perform distillation over timesteps $[1000, 750, 500, 250]$ with a teacher CFG scale of 6.0. For DM-Group, we set the group size to $K = 8$ and average the top-4 samples to construct the preference-guided target. In our combined loss function, λ_{align} and λ_{DMD} are set to 0.5. We provide a comprehensive sensitivity analysis of these weights, including various static ratios and dynamic scheduling, on the CogVideoX model in Section 4.3. While a DMD-only warm-up might seem intuitive to prevent early-stage instability, we find it unnecessary. First, even during the initial stages where generations are blurry, the coarse color blocks and basic motion patterns already provide relative advantages to establish a roughly correct direction for preference gradient guidance. Second, the DMD gradient inherently dominates this early phase to rapidly establish distribution consistency, natively averting instability without explicit tuning. Additionally, we note that standard RL baselines like DanceGRPO and Flow-DPO heavily rely on Exponential Moving Average (EMA) to maintain training stability, without which the visual quality of their generated samples tends to degrade. In contrast, our unified framework exhibits robust convergence natively. While DM-Align and DMD2 converge stably within 500 steps, we train them for 1,000 steps to ensure complete convergence. Conversely, due to the substantially higher per-step time cost of DanceGRPO, we train it for 100 steps. For evaluation, we curate a test

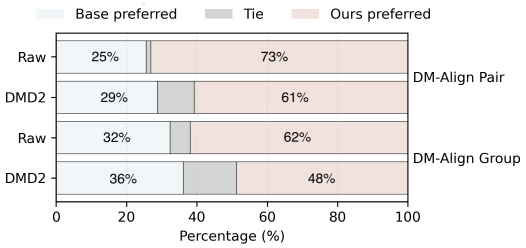

Fig. 3: Human evaluation (GSB protocol).

Table 2: TA scores on VidProM.

Method	TA Score
Wan-T2V-1.3B (raw)	0.69
DMD2	0.75
DanceGRPO	1.37
DanceGRPO + DMD2	1.40
DM-Align (Group)	**1.65**

set of 640 samples and adopt VBench [14] for automated metrics alongside the GSB (Good, Same, Bad) protocol for human evaluation.

4.2 Main Experiments

The quantitative VBench results are summarized in Table 1, where NFE denotes the Number of Function Evaluations. Overall, standalone DMD2 effectively distills the 100-NFE generation process into a more efficient 4-NFE few-step generation, demonstrating inherent baseline stability. Our single-stage approach successfully integrates alignment into this framework without compromising these distillation capabilities.

Specifically, **DM-Align (Pair)** outperforms all baselines, achieving the highest average score (82.78) with only 4 NFE. When analyzing the metrics, both our method and Flow-DPO significantly improve Motion Quality. However, the standalone Flow-DPO shows limited gains in Visual Quality (e.g., its Aesthetic Quality drops to 56.32). While the two-stage Flow-DPO + DMD2 pipeline can partially recover this Visual Quality (58.26), it drastically diminishes the Motion Quality benefits (Dynamic Degree drops from 67.13 to 53.43). In contrast, our single-stage optimization stably improves both aspects simultaneously, achieving a robust Dynamic Degree of 66.25 and an Aesthetic Quality of 60.89.

For **DM-Align (Group)**, our method similarly achieves the highest average score (84.40), showing substantial improvements over other baselines. However, similar to the pair variant, the sequential two-stage pipeline suffers from severe performance degradation after distillation. For instance, while DanceGRPO initially achieves a high Dynamic Degree of 75.78, applying DMD2 afterward causes a precipitous drop to 60.85 (a 14.93 reduction). Our approach completely circumvents this post-distillation collapse, pushing the Dynamic Degree to an impressive 80.19.

Human evaluation results (Figure 3) demonstrate that both variants of our method significantly outperform the raw model and DMD2 in overall human preference. DM-Align (Pair) shows particularly pronounced improvements (+48% and +32% net preference) as it leverages real human preference data, providing highly accurate gradient directions for alignment. Although DM-Align (Group) relies on multiple sampling and reward model estimations for preference-guided gradients, it still delivers substantial performance gains (+30% and +12% net

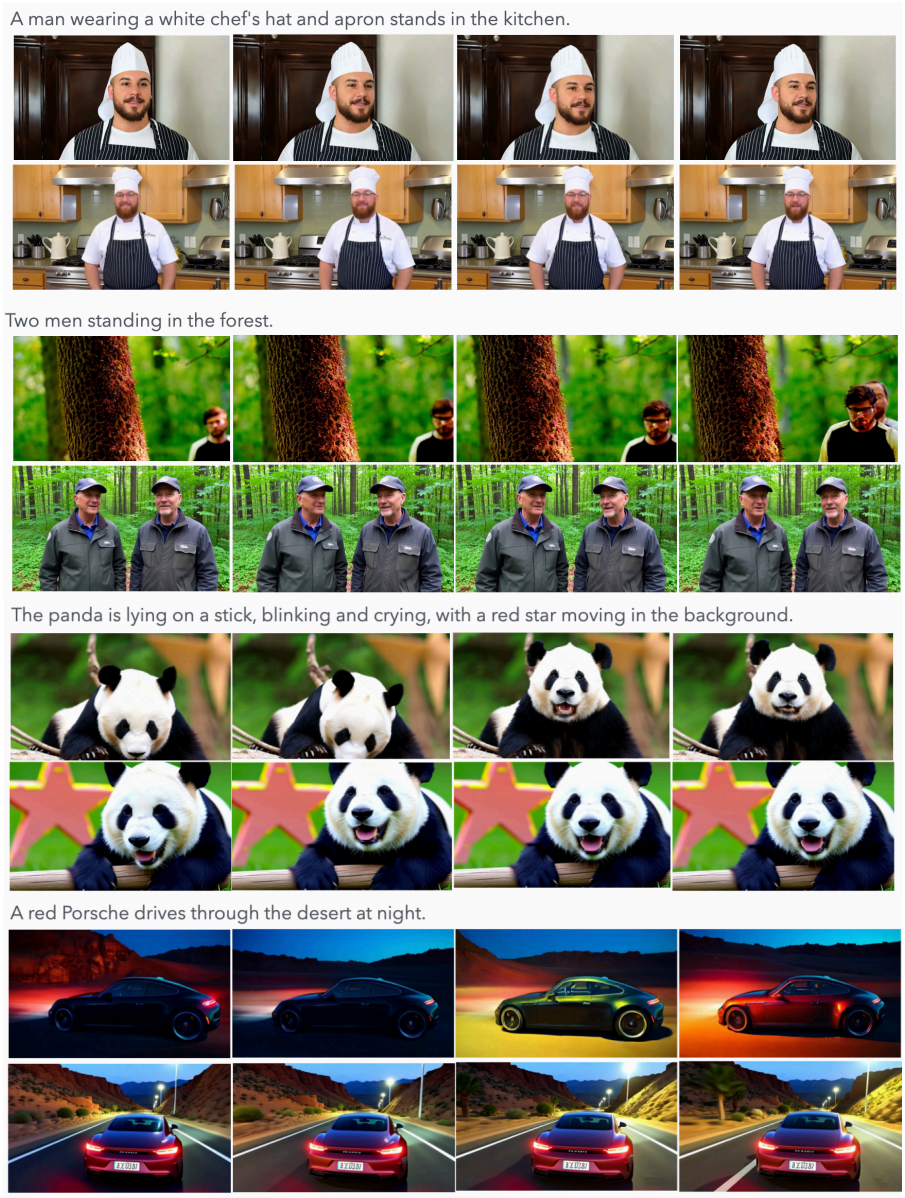

Fig. 4: Qualitative comparison between the raw base model and our DM-Align. For each prompt, the upper row presents the raw base model outputs, and the lower row displays our DM-Align results. The *top two prompts* showcase DM-Align (Pair): whereas the base model produces a frozen, single-person output, our method correctly generates the requested kitchen background and depicts two persons with more natural movements. The *bottom two prompts* showcase DM-Align (Group): our method accurately generates specific textual details, such as a red star and a red Porsche, demonstrating superior prompt consistency.

Table 1: Main results on VBench evaluation. The best results are in **bold**. The upper section presents results on the ConsistID dataset using DM-Align (Pair), while the lower section shows results on the VidProM dataset using DM-Align (Group).

Method	NFE	Average Score	Temporal Consistency		Motion Quality		Visual Quality	
			Subject Consistency	Background Consistency	Motion Smoothness	Dynamic Degree	Aesthetic Quality	Imaging Quality
DM-Align (Pair) on ConsistID Dataset								
Wan-T2V-1.3B (raw model)	100	78.20	96.57	94.46	99.01	50.78	56.09	72.29
DMD2	4	79.68	98.03	95.78	99.10	52.81	57.64	74.73
Flow-DPO	100	81.54	98.10	95.56	99.25	**67.13**	56.32	72.91
Flow-DPO + DMD2	4	79.89	98.14	**96.01**	99.18	53.43	58.26	74.35
DM-Align (Pair) (Ours)	4	**82.78**	**98.80**	95.76	**99.35**	66.25	**60.89**	**75.61**
DM-Align (Group) on VidProM Dataset								
Wan-T2V-1.3B (raw model)	100	77.85	93.93	94.63	98.63	56.40	57.33	66.20
DMD2	4	78.88	95.53	95.22	98.60	51.64	**63.02**	69.30
DanceGRPO	100	82.76	96.04	95.86	98.33	75.78	60.44	70.15
DanceGRPO + DMD2	4	80.54	96.60	95.89	98.43	60.85	60.71	70.81
DM-Align (Group) (Ours)	4	**84.40**	**97.13**	**96.49**	**98.99**	**80.19**	62.17	**71.45**

preference). Furthermore, regarding Textual Alignment (TA)—the primary reward metric focused on during DM-Align (Group) training—Table 2 confirms that our framework provides superior guidance, drastically outperforming the next best baseline (1.65 vs. 1.40). As illustrated in Figure 4, our approach exhibits clear visual improvements in both prompt consistency and motion smoothness compared to the base model outputs.

4.3 Efficiency, Scalability, and Generality

Beyond the primary VBench metrics, we extend our evaluation to validate the core architectural advantages of DM-Align.

Training Efficiency. As shown in Table 3, traditional RL methods like DanceGRPO incur prohibitive multi-step sampling costs, requiring over 1500 NFE and 400 seconds per optimization step. In stark contrast, our single-stage DM-Align completely eliminates multi-step sampling bottlenecks, reducing the training time by nearly 30× while natively aligning the model.

Scalability to MoE Architectures. To validate scalability, we conducted experiments on the latest large-scale Mixture-of-Experts (MoE) architecture, **Wan2.2-T2V-A14B** [47]. The Wan2.2 architecture consists of two distinct 14B models for high-noise and low-noise regimes. Empirically, we observe that applying DM-Align solely to the high-noise expert is sufficient to achieve substantial gains (Table 4). Since high-level semantic consistency is predominantly determined in the high-noise regime, aligning only this expert is both remarkably effective and computationally efficient.

Generality and Sensitivity. We extended DM-Align to **CogVideoX-2B** [53] and evaluated it using an alternative reward model (**HPSv2**). As presented in Table 5, DM-Align successfully boosts performance over the standalone DMD baseline on both models and metrics without specific hyper-parameter

Table 3: Training efficiency comparison. 'Time (s)' denotes the time per optimization step in seconds.

Method	NFE	Time (s)
DMD2	46	11
DanceGRPO	1536	423
DM-Align (Ours)	**51**	**15**

Table 4: VBench Scores on the latest MoE model (Wan2.2-A14B).

Method	NFE	Score
Raw Base	100	80.00
DMD	4	78.35
DM-Align	4	**83.45**

Table 5: Generality & Sensitivity analysis ($\lambda_{\text{DMD}} : \lambda_{\text{align}}$) across different backbones and reward models. "Varying" denotes a linear schedule where the weight ratio $\lambda_{\text{DMD}} : \lambda_{\text{align}}$ is linearly scheduled from $1:0$ to $1:3$ over training.

Base Model	Reward	1:0 (DMD)	3:1	**1:1 (Ours)**	1:3	Varying
Wan-1.3B	TA (↑)	0.75	1.34	**1.65**	<u>1.56</u>	1.45
	HPSv2 (↑)	20.50	24.01	**28.89**	12.37	<u>27.56</u>
CogVideoX	TA (↑)	0.22	**0.58**	<u>0.55</u>	0.08	0.49
	HPSv2 (↑)	18.71	20.92	**25.01**	13.78	<u>22.11</u>

tuning. Furthermore, we conducted a sensitivity analysis on the weight ratio $\lambda_{\text{DMD}} : \lambda_{\text{align}}$, keeping their sum constrained to 1. Alongside various static ratios, we also test a dynamic "Varying" strategy, where the ratio $\lambda_{\text{DMD}} : \lambda_{\text{align}}$ is linearly scheduled from 1:0 to 1:3 over the training steps, while normalizing the two weights to maintain $\lambda_{\text{DMD}} + \lambda_{\text{align}} = 1$. This schedule starts from pure distillation to establish structural consistency and gradually increases the alignment strength to refine human preferences. The results reveal that performance remains highly robust across a wide range of configurations. Severe instability or degradation only occurs when the DMD weight is excessively low (e.g., 1:3), where the insufficient distillation signal fails to sustain the base few-step generation capability.

4.4 Analysis of Alternative Approaches

Why Not Sequential Pipelines? An intuitive approach is the two-stage pipeline: either "Distillation-after-RL" or "RL-after-Distillation". As demonstrated in Table 1, optimizing alignment and distillation in two isolated stages often results in objective misalignment. Specifically, applying distillation to a previously RL-aligned model tends to degrade the very metrics (e.g., motion quality and dynamic degree) that the initial RL stage sought to improve. Conversely, performing RL training *after* the distillation process to leverage fast sampling is equally catastrophic. As shown in Figure 5, fine-tuning a distilled model leads to rapid degradation, manifesting as noticeable blurring artifacts and flickering motions, eventually progressing to complete model collapse. We hypothesize that a distilled model's parameters are hypersensitive; even minor gradient adjustments

Fig. 5: Visual examples of catastrophic model collapse when attempting to fine-tune a distilled model (the sequential "RL-after-Distillation" pipeline).

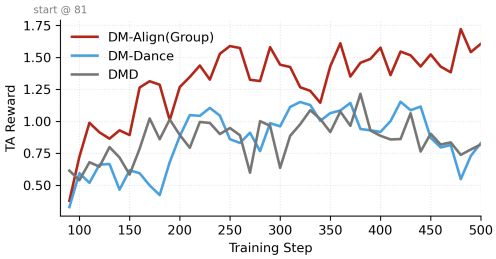

Fig. 6: Reward progression during training. Unlike direct RL loss integration, DM-Align provides superior and stable reward guidance.

from sparse RL signals shatter its few-step generation capability. This reinforces the necessity of our unified, single-stage distribution matching framework.

Why Not Direct RL Loss Integration? Another straightforward alternative to our method would be directly incorporating off-the-shelf RL losses alongside the DMD optimization process. To investigate this, we developed a variant that combines the DanceGRPO loss with the standard DMD loss. As illustrated in Figure 6, this direct RL loss variant performs nearly identically to the standalone DMD, struggling to provide meaningful reward progression. We attribute this to the overly conservative gradient updates characteristic of standard RL losses designed for visual generation. Because applying standard RL directly to continuous generative models is notoriously unstable, their loss formulations are heavily constrained. In contrast, our DM-Align operates natively in the distribution matching space, leveraging the fake model's unified denoising vector field to provide dense, stable, and highly compatible guidance.

Does the Gain Come from the Dataset? A potential concern is that the improvement of DM-Align may primarily stem from the high-quality ConsistID data rather than the proposed unified formulation. To disentangle this dataset factor, we conduct a prompt-only cross-evaluation on ConsistID, where DM-Group uses only the textual prompts without accessing the paired real videos. As shown in Table 6, DM-Group still substantially outperforms DMD2, the sequential RL baseline DanceGRPO, and the concurrent joint-optimization method DMDR. This result suggests that the performance gain is not merely a byproduct of high-quality paired data, but arises from formulating preference alignment as a native score-space distribution-matching gradient compatible with DMD.

4.5 Limitations and Discussion

While our framework successfully bridges distillation and preference alignment, several limitations remain. First, although we provide a theoretical derivation

Table 6: Cross-evaluation on ConsistID using prompts only. DM-Group uses only textual prompts without paired real videos.

Method	Dynamic	Aesthetic	Temporal
DMD2	74.90	64.19	95.51
DanceGRPO	78.42	66.85	96.12
DMDR [15]	77.15	67.40	95.88
DM-Group	**81.35**	**68.12**	**97.15**

for the alignment gradient, the joint optimization dynamics of the weighted loss summation lack strict theoretical convergence guarantees. Second, our implementation is built upon the foundational DMD framework. This restricts the distillation mechanism to DMD-based techniques, excluding other recent GAN-based alternatives [6,22,32] that have shown promising distillation performance.

However, our primary algorithm design intentionally prioritizes simplicity and foundational validation. Recently, advanced DMD-based enhancements have emerged [8, 31, 40]. Because our DM-Align operates orthogonally as a native, score-based gradient provider, it can be seamlessly integrated with these advancements to yield further improvements. We leave these exciting integrations for future work. More broadly, we hope this work invites a higher-level perspective on distillation and alignment—transitioning them from traditionally separated stages into a unified paradigm.

5 Conclusion

In this paper, we introduced DM-Align, a novel framework that unifies video generation model distillation and human preference alignment into a single cohesive stage through distribution matching. By operating entirely in the score domain, our approach eliminates the need to adapt the diffusion sampling process into an MDP formulation, simultaneously bypassing time-intensive multi-step generation bottlenecks that plague conventional diffusion RL methods. Extensive evaluations across diverse base models and reward signals demonstrate that our method preserves the efficiency of few-step distillation while securing robust alignment with human preferences and exhibiting strong generalization capabilities. Both automated metrics and human evaluations confirm that DM-Align significantly outperforms standalone baselines and conventional two-stage pipelines, paving the way for more efficient and unified generative model optimization.

Acknowledgment

This work was supported in part by the Shenzhen Science and Technology Program (Grant No. RCJC20210706091946001), and in part by the Shenzhen Science and Technology Program (Grant No. ZDCY20250901104207008).

References

1. Albergo, M.S., Boffi, N.M., Vanden-Eijnden, E.: Stochastic interpolants: A unifying framework for flows and diffusions. arXiv preprint arXiv:2303.08797 (2023)
2. Black, K., Janner, M., Du, Y., Kostrikov, I., Levine, S.: Training diffusion models with reinforcement learning. arXiv preprint arXiv:2305.13301 (2023)
3. Chadebec, C., Tasar, O., Benaroche, E., Aubin, B.: Flash diffusion: Accelerating any conditional diffusion model for few steps image generation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 15686–15695 (2025)
4. Chen, G., Huang, S., Liu, K., Zhu, J., Qu, X., Chen, P., Cheng, Y., Sun, Y.: Flash-dmd: Towards high-fidelity few-step image generation with efficient distillation and joint reinforcement learning (2025), https://arxiv.org/abs/2511.20549
5. Chen, G., Lin, D., Yang, J., Lin, C., Zhu, J., Fan, M., Zhang, H., Chen, S., Chen, Z., Ma, C., et al.: Skyreels-v2: Infinite-length film generative model. arXiv preprint arXiv:2504.13074 (2025)
6. Cheng, J., Ma, B., Ren, X., Jin, H., Yu, K., Zhang, P., Li, W., Zhou, Y., Zheng, T., Lu, Q.: Pose: Phased one-step adversarial equilibrium for video diffusion models. arXiv preprint arXiv:2508.21019 (2025)
7. Efron, B.: Tweedie's formula and selection bias. Journal of the American Statistical Association **106**(496), 1602–1614 (2011)
8. Fan, X., Qiu, Z., Wu, Z., Wang, F., Lin, Z., Ren, T., Lin, D., Gong, R., Yang, L.: Phased dmd: Few-step distribution matching distillation via score matching within subintervals (2025), https://arxiv.org/abs/2510.27684
9. Fan, Y., Watkins, O., Du, Y., Liu, H., Ryu, M., Boutilier, C., Abbeel, P., Ghavamzadeh, M., Lee, K., Lee, K.: Dpok: Reinforcement learning for fine-tuning text-to-image diffusion models. Advances in Neural Information Processing Systems **36**, 79858–79885 (2023)
10. Google DeepMind: Veo 2. https://deepmind.google/technologies/veo/veo-2/ (Dec 2024), accessed: 2025-09-15
11. Gu, Y., Li, X., Hu, Y., Zhuang, B.: Video-blade: Block-sparse attention meets step distillation for efficient video generation. arXiv preprint arXiv:2508.10774 (2025)
12. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
13. Huang, X., Li, Z., He, G., Zhou, M., Shechtman, E.: Self forcing: Bridging the train-test gap in autoregressive video diffusion. arXiv preprint arXiv:2506.08009 (2025)
14. Huang, Z., He, Y., Yu, J., Zhang, F., Si, C., Jiang, Y., Zhang, Y., Wu, T., Jin, Q., Chanpaisit, N., et al.: Vbench: Comprehensive benchmark suite for video generative models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21807–21818 (2024)
15. Jiang, D., Liu, D., Wang, Z., Wu, Q., Li, L., Li, H., Jin, X., Liu, D., Li, Z., Zhang, B., Wang, M., Hoi, S., Gao, P., Yang, H.: Distribution matching distillation meets reinforcement learning (2025), https://arxiv.org/abs/2511.13649
16. Karras, T., Aittala, M., Aila, T., Laine, S.: Elucidating the design space of diffusion-based generative models. Advances in neural information processing systems **35**, 26565–26577 (2022)
17. Kong, W., Tian, Q., Zhang, Z., Min, R., Dai, Z., Zhou, J., Xiong, J., Li, X., Wu, B., Zhang, J., et al.: Hunyuanvideo: A systematic framework for large video generative models. arXiv preprint arXiv:2412.03603 (2024)
18. Kuaishou: Kling ai. https://klingai.kuaishou.com/ (2024), accessed: 2025-09-15

19. Li, J., Cui, Y., Huang, T., Ma, Y., Fan, C., Yang, M., Zhong, Z.: Mixgrpo: Unlocking flow-based grpo efficiency with mixed ode-sde. arXiv preprint arXiv:2507.21802 (2025)
20. Liang, Z., Yuan, Y., Gu, S., Chen, B., Hang, T., Li, J., Zheng, L.: Step-aware preference optimization: Aligning preference with denoising performance at each step. arXiv preprint arXiv:2406.04314 **2**(5), 7 (2024)
21. LightX2V: Lightx2v: Light video generation inference framework. `https://github.com/ModelTC/lightx2v` (2025)
22. Lin, S., Xia, X., Ren, Y., Yang, C., Xiao, X., Jiang, L.: Diffusion adversarial post-training for one-step video generation. arXiv preprint arXiv:2501.08316 (2025)
23. Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. arXiv preprint arXiv:2210.02747 (2022)
24. Liu, J., Liu, G., Liang, J., Li, Y., Liu, J., Wang, X., Wan, P., Zhang, D., Ouyang, W.: Flow-grpo: Training flow matching models via online rl. arXiv preprint arXiv:2505.05470 (2025)
25. Liu, J., Liu, G., Liang, J., Yuan, Z., Liu, X., Zheng, M., Wu, X., Wang, Q., Qin, W., Xia, M., et al.: Improving video generation with human feedback. arXiv preprint arXiv:2501.13918 (2025)
26. Liu, L., Ren, Y., Lin, Z., Zhao, Z.: Pseudo numerical methods for diffusion models on manifolds. arXiv preprint arXiv:2202.09778 (2022)
27. Liu, X., Gong, C., Liu, Q.: Flow straight and fast: Learning to generate and transfer data with rectified flow. arXiv preprint arXiv:2209.03003 (2022)
28. Lu, C., Song, Y.: Simplifying, stabilizing and scaling continuous-time consistency models. arXiv preprint arXiv:2410.11081 (2024)
29. Luo, S., Tan, Y., Huang, L., Li, J., Zhao, H.: Latent consistency models: Synthesizing high-resolution images with few-step inference. arXiv preprint arXiv:2310.04378 (2023)
30. Luo, W.: Diff-instruct++: Training one-step text-to-image generator model to align with human preferences. arXiv preprint arXiv:2410.18881 (2024)
31. Luo, Y., Hu, T., Sun, J., Cai, Y., Tang, J.: Learning few-step diffusion models by trajectory distribution matching. arXiv preprint arXiv:2503.06674 (2025)
32. Mao, X., Jiang, Z., Wang, F.Y., Zhang, J., Chen, H., Chi, M., Wang, Y., Luo, W.: Osv: One step is enough for high-quality image to video generation. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 12585–12594 (2025)
33. Meng, C., Song, Y., Li, W., Ermon, S.: Estimating high order gradients of the data distribution by denoising. Advances in Neural Information Processing Systems **34**, 25359–25369 (2021)
34. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al.: Training language models to follow instructions with human feedback. Advances in neural information processing systems **35**, 27730–27744 (2022)
35. Peters, J., Mulling, K., Altun, Y.: Relative entropy policy search. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 24, pp. 1607–1612 (2010)
36. Polyak, A., Zohar, A., Brown, A., Tjandra, A., Sinha, A., Lee, A., Vyas, A., Shi, B., Ma, C.Y., Chuang, C.Y., et al.: Movie gen: A cast of media foundation models. arXiv preprint arXiv:2410.13720 (2024)
37. Rafailov, R., Sharma, A., Mitchell, E., Manning, C.D., Ermon, S., Finn, C.: Direct preference optimization: Your language model is secretly a reward model. Advances in neural information processing systems **36**, 53728–53741 (2023)

38. Salimans, T., Ho, J.: Progressive distillation for fast sampling of diffusion models. arXiv preprint arXiv:2202.00512 (2022)
39. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347 (2017)
40. Shao, S., Yi, H., Guo, H., Ye, T., Zhou, D., Lingelbach, M., Xu, Z., Xie, Z.: Magicdistillation: Weak-to-strong video distillation for large-scale few-step synthesis. arXiv preprint arXiv:2503.13319 (2025)
41. Shao, Z., Wang, P., Zhu, Q., Xu, R., Song, J., Bi, X., Zhang, H., Zhang, M., Li, Y., Wu, Y., et al.: Deepseekmath: Pushing the limits of mathematical reasoning in open language models. arXiv preprint arXiv:2402.03300 (2024)
42. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
43. Song, Y., Dhariwal, P., Chen, M., Sutskever, I.: Consistency models (2023)
44. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456 (2020)
45. Sun, Y., Wu, J., Cao, Y., Xu, C., Wang, Y., Cao, W., Luo, D., Wang, C., Fu, Y.: Swiftvideo: A unified framework for few-step video generation through trajectory-distribution alignment. arXiv preprint arXiv:2508.06082 (2025)
46. Wallace, B., Dang, M., Rafailov, R., Zhou, L., Lou, A., Purushwalkam, S., Ermon, S., Xiong, C., Joty, S., Naik, N.: Diffusion model alignment using direct preference optimization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8228–8238 (2024)
47. Wan, T., Wang, A., Ai, B., Wen, B., Mao, C., Xie, C.W., Chen, D., Yu, F., Zhao, H., Yang, J., et al.: Wan: Open and advanced large-scale video generative models. arXiv preprint arXiv:2503.20314 (2025)
48. Wang, W., Yang, Y.: Vidprom: A million-scale real prompt-gallery dataset for text-to-video diffusion models. Advances in Neural Information Processing Systems **37**, 65618–65642 (2024)
49. Wang, Z., Lu, C., Wang, Y., Bao, F., Li, C., Su, H., Zhu, J.: Prolificdreamer: High-fidelity and diverse text-to-3d generation with variational score distillation. Advances in neural information processing systems **36**, 8406–8441 (2023)
50. Wu, X., Hao, Y., Sun, K., Chen, Y., Zhu, F., Zhao, R., Li, H.: Human preference score v2: A solid benchmark for evaluating human preferences of text-to-image synthesis. arXiv preprint arXiv:2306.09341 (2023)
51. Xue, Z., Wu, J., Gao, Y., Kong, F., Zhu, L., Chen, M., Liu, Z., Liu, W., Guo, Q., Huang, W., et al.: Dancegrpo: Unleashing grpo on visual generation. arXiv preprint arXiv:2505.07818 (2025)
52. Yang, K., Tao, J., Lyu, J., Ge, C., Chen, J., Li, Q., Shen, W., Zhu, X., Li, X.: Using human feedback to fine-tune diffusion models without any reward model (2024). URL https://arxiv. org/abs/2311.13231 (2024)
53. Yang, Z., Teng, J., Zheng, W., Ding, M., Huang, S., Xu, J., Yang, Y., Hong, W., Zhang, X., Feng, G., et al.: Cogvideox: Text-to-video diffusion models with an expert transformer. arXiv preprint arXiv:2408.06072 (2024)
54. Yin, T., Gharbi, M., Park, T., Zhang, R., Shechtman, E., Durand, F., Freeman, B.: Improved distribution matching distillation for fast image synthesis. Advances in neural information processing systems **37**, 47455–47487 (2024)
55. Yin, T., Gharbi, M., Zhang, R., Shechtman, E., Durand, F., Freeman, W.T., Park, T.: One-step diffusion with distribution matching distillation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6613–6623 (2024)

56. Yin, T., Zhang, Q., Zhang, R., Freeman, W.T., Durand, F., Shechtman, E., Huang, X.: From slow bidirectional to fast autoregressive video diffusion models. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 22963–22974 (2025)
57. Yuan, S., Huang, J., He, X., Ge, Y., Shi, Y., Chen, L., Luo, J., Yuan, L.: Identity-preserving text-to-video generation by frequency decomposition. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 12978–12988 (2025)
58. Zhang, T., Da, C., Ding, K., Yang, H., Jin, K., Li, Y., Gao, T., Zhang, D., Xiang, S., Pan, C.: Diffusion model as a noise-aware latent reward model for step-level preference optimization. arXiv preprint arXiv:2502.01051 (2025)
59. Zhao, W., Bai, L., Rao, Y., Zhou, J., Lu, J.: Unipc: A unified predictor-corrector framework for fast sampling of diffusion models. Advances in Neural Information Processing Systems **36**, 49842–49869 (2023)