

DUET: A Diversity–Quality Duet of Distillation Experts for Two-Step Video Generation

Zian Li^{1,2*}, Litong Gong⁴, Borui Liao⁴, Pengfei Liu⁴, Xinyu Wang⁵,
Xinyuan Wei¹, Yifan Gao⁴, Tiezheng Ge⁴, Muhan Zhang^{1,3†}

¹Institute for Artificial Intelligence, Peking University, Beijing, China

²School of Intelligence Science and Technology, Peking University, Beijing, China

³State Key Laboratory of General Artificial Intelligence, Peking University, Beijing, China

⁴Alibaba Group, Beijing, China

⁵Shenzhen International Graduate School, Tsinghua University

Abstract

Diffusion models have enabled high-quality video generation in recent years, but the high cost of iterative sampling hinders their practical deployment. Few-step distillation alleviates this cost, yet exposes a quality–diversity trade-off between its two dominant paradigms: trajectory-level distillation (e.g., sCM) favors diversity, whereas distribution-level distillation (e.g., DMD) favors quality. Targeting extreme two-step video generation, we introduce **DUET**, which reconciles the two paradigms through a noise-level duet of experts: an sCM expert takes the high-noise step to lay out diverse structure, and a DMD expert takes the low-noise step to refine appearance detail. Since the two experts are trained independently with their native objectives, DUET sidesteps the optimization difficulties of loss-level combinations and delivers quality and diversity jointly rather than trading one for the other. We further identify the relay interface and the high-noise stage as the remaining bottlenecks, and address them with RL-guided expert adaptation, yielding **DUET+**. With the Wan2.1-T2V-1.3B backbone, DUET lifts the two-step quality of sCM close to the level of DMD while retaining nearly all of its structural diversity—about twice that of DMD—and DUET+ further improves overall quality while preserving this diversity advantage. Together, these results establish noise-level expert specialization as a simple, effective paradigm for reconciling diversity and quality in two-step video generation.

1 Introduction

Since Sora (Brooks et al. 2024) demonstrated the potential of DiT-based (Peebles and Xie 2023) video generation, video diffusion models have advanced at a remarkable pace: a wave of successors (Wan et al. 2025; Team et al. 2025; HaCohen et al. 2024) now generate impressively coherent clips and are seeing rapid adoption in content creation, advertising, and film production.

However, their iterative samplers rely on long trajectories of network evaluations—producing even a single 5-second clip can take tens of minutes on a high-end GPU.

*Work done as an intern at Alibaba Group (zian@stu.pku.edu.cn).

†Correspondence to Muhan Zhang (muhan@pku.edu.cn).



Figure 1: The quality–diversity trade-off between the two distillation paradigms. Each row shows the first frames of videos generated with 5 seeds. **DUET** preserves sCM’s diverse layout while inheriting the high quality of DMD.

Few-step distillation, which has converged on two fundamental paradigms—trajectory-level and distribution-level—addresses this cost by compressing the trajectory into few NFES, but with no free lunch, as illustrated in Figure 1. On the one hand, trajectory-level objectives, exemplified by consistency distillation (Song et al. 2023; Lu and Song 2025), largely preserve the teacher’s sample diversity, yet their mean-seeking nature yields conservative, blurry outputs. On the other hand, distribution-level objectives, exemplified by distribution matching distillation (DMD) (Yin et al. 2024b,a), produce sharp outputs, but their mode-seeking nature concentrates probability mass on substantially fewer modes. The quality–diversity trade-off is not merely an empirical inconvenience of few-step generation, but a symptom of the divergent natures of the two distillation objectives.

A common response to this apparent dilemma is to combine the two distillation losses into a single student training, hoping to inherit diversity from trajectory-level distillation and fidelity from distribution-level distillation (Zheng et al. 2025; Cai et al. 2026; Zou et al. 2026; Ge et al. 2026). Yet *such loss-level combinations typically find a compromise rather than a true escape*—improving one side of the trade-off still comes at the expense of the other—and introduce *hard optimization* due to conflicts between the objectives. For example, rCM (Zheng et al. 2025) augments the sCM objective with a DMD loss down-weighted to 1/100, but its diversity

remains substantially lower than that of its sCM counterpart, as shown in Table 1. Similarly, HiAR (Zou et al. 2026) requires a dedicated design to mitigate conflicts between its trajectory-matching and distribution-matching losses. These methods ask one set of parameters to negotiate between objectives with different preferences; the negotiation can be improved, but it is difficult to remove.

In this paper, we pursue a route that maintains *simplicity*, yet *preserves the advantages of both paradigms as much as possible*. Motivated by expert specialization in Wan2.2’s MoE design and by the distinct roles of different distillation objectives, we propose **DUET** (Diversity–qUality Expert Tandem), built on a principle of *noise-level expert duet*: each interval of the sampling schedule is assigned to the objective whose optimum best matches the role of that interval. Concretely, DUET routes the high-noise interval $[1, \tau]$ to a diversity-preserving sCM expert, where layout, composition, and motion are primarily determined, and the low-noise interval $[\tau, 0]$ to a fidelity-oriented DMD expert, where visual details and appearance are refined. The principle is not tied to a specific step budget; we stress-test it in the extreme two-step regime, where routing is cleanest—each regime is exactly one step—and the trade-off bites hardest. As shown in Figure 1, this simple expert duet preserves much of sCM’s layout diversity while maintaining the visual quality of DMD.

Diverse outputs make downstream preference optimization effective. We therefore further introduce a lightweight RL-based expert adaptation for DUET, yielding **DUET+**. Rather than applying a generic objective to the whole DUET, we optimize each expert according to its own role. For the sCM expert, whose diverse outputs provide the exploration needed for preference learning, we apply GRPO (Shao et al. 2024; Lu, Zuo, and Deng 2026) to steer its high-noise predictions toward higher-reward structures while preserving its coverage behavior. For the DMD expert, we continue DMD training on the sCM expert’s latents, so that it adapts to the actual intermediate distribution produced by the sCM expert at inference. Experiments show that this role-aware adaptation is much more effective than the common init-then-DMD recipe (Yin et al. 2025), which quickly collapses the inherited diversity; DUET+ substantially improves image quality while preserving the diversity advantage.

To summarize, our contributions are fourfold:

- We diagnose the quality–diversity trade-off at the objective level: trajectory- and distribution-level objectives induce distinct noise-to-data mappings, making loss-level combination optimization-hard and compromise-prone.
- We propose **DUET**, a noise-level expert duet that assigns diversity- and fidelity-oriented experts to the noise regimes where they are most effective.
- We introduce a role-aware adaptation that exploits the sCM expert’s diversity for preference optimization and repairs the DMD expert’s relay interface, yielding **DUET+**.
- On Wan2.1-1.3B, we show that quality and diversity can be obtained jointly in the extreme two-step regime: DUET approaches DMD-level quality while preserving much of sCM’s diversity, and DUET+ reaches DMD-level overall quality while preserving the diversity advantage.

2 Related Work

Trajectory distillation. Trajectory-level distillation accelerates sampling by learning shortcut mappings along the teacher’s ODE trajectories. Early methods such as progressive distillation (Salimans and Ho 2022) progressively distill the DDIM (Song, Meng, and Ermon 2020) sampling process, enabling shorter generation trajectories. Consistency models (Song et al. 2023) learn a consistency function that maps different points on the same ODE trajectory to a shared clean-data endpoint. Subsequent works extend this idea to continuous-time formulations sCM (Lu and Song 2025), multi-step, phased, and truncated sampling schemes (Heek, Hoogetboom, and Salimans 2024; Ren et al. 2024; Wang et al. 2024a; Lee et al. 2024). Another line of work directly learns flow maps between arbitrary endpoint pairs (Kim et al. 2024; Geng et al. 2026; Sabour, Fidler, and Kreis 2026; Frans et al. 2025; Chen et al. 2025; Zheng et al. 2025). Despite these differences, these methods share the goal of matching the teacher’s trajectory structure.

Distribution matching distillation. Another family of methods does not attempt to reproduce individual samples along the teacher’s trajectory exactly, but instead aims to match the teacher at the distribution level. Among them, DMD (Yin et al. 2024b,a) is a representative approach, which minimizes an approximate reverse KL divergence between the student and teacher distributions with the aid of an auxiliary fake-score network. Follow-up works replace the reverse KL with alternative distribution distances, including Fisher-divergence-style objectives built on score identities (Zhou et al. 2024; Luo et al. 2024) and general f -divergences (Xu, Nie, and Vahdat 2025). Beyond explicit distribution distances, adversarial distillation instead aligns the student with the teacher or data distribution through a learned discriminator (Sauer et al. 2024b,a; Lin et al. 2025, 2026). Distribution matching has been adopted well beyond text-to-image synthesis, spanning autoregressive and interactive video generation (Yin et al. 2025; Huang et al. 2026; Zhu et al. 2026), molecule generation (Wei et al. 2026), and reward-aware distillation (Jiang et al. 2025; Bai et al. 2026).

Balancing diversity and quality. The mode-seeking behavior of DMD and the mode-covering tendency of trajectory distillation motivate researchers to seek a balance between quality and diversity. One line of work combines the two types of objectives, often viewed as reverse-KL- and forward-KL-style supervision. Some methods use sCM as the dominant objective and add a small amount of DMD regularization (Zheng et al. 2025), while others treat DMD as the primary objective and introduce forward-KL-style supervision (Zou et al. 2026; Cai et al. 2026; Ge et al. 2026). These approaches either require careful tuning of loss coefficients or must mitigate strong gradient conflicts through specialized gradient-routing designs (Zou et al. 2026; Cai et al. 2026). Another line of work stays within distribution matching and mitigates diversity collapse from inside: f -distill (Xu, Nie, and Vahdat 2025) selects divergences with weaker mode-seeking tendencies, AMD (Bai et al. 2026) detects collapsed modes with reward proxies and applies repulsive corrections from the fake-score model, and DP-DMD (Wu et al. 2026a)

reserves the first step for regression and applies DMD only afterward. In contrast, our method avoids directly mixing the two objectives within a single training loss, and preserves diversity structurally through noise-level expert duet.

We provide additional related work about noise-level expert specialization in Appendix A.

3 Preliminaries

3.1 Flow Matching

Let c denote a text condition, $x_0 \sim p_{\text{data}}(\cdot | c)$ a clean video latent, and $x_1 \sim \mathcal{N}(0, I)$ Gaussian noise; the final video is obtained by decoding a clean latent with the decoder $\text{Dec}(\cdot)$. Flow matching (Lipman et al. 2022) connects data and noise through the linear interpolation

$$x_t = (1 - t)x_0 + tx_1, \quad t \in [0, 1], \quad (1)$$

and trains a teacher velocity field $v_\psi(x_t, t, c)$ by regressing the path velocity, $\min_\psi \mathbb{E}[\|v_\psi(x_t, t, c) - (x_1 - x_0)\|_2^2]$. The teacher defines the probability-flow ODE

$$\frac{dx_t}{dt} = v_\psi(x_t, t, c). \quad (2)$$

Integrating Eq. (2) backward from $t = 1$ to $t = 0$ transports the noise distribution to the data distribution. Throughout this paper, ψ denotes the teacher, and θ and ϕ parameterize the two distilled students introduced below.

3.2 Trajectory-Level Distillation

Trajectory-level distillation learns shortcuts of the teacher ODE (Salimans and Ho 2022; Song et al. 2023; Frans et al. 2025; Lu and Song 2025; Zheng et al. 2025). Among these, consistency distillation is a representative method: a consistency function $f_\theta^{\text{SCM}}(x_t, t, c)$ predicts the clean endpoint of the teacher trajectory passing through x_t , supervised by enforcing consistent predictions at two nearby states on the same trajectory (Song et al. 2023). sCM (Lu and Song 2025) takes the continuous-time limit of this objective; abstracting away parameterization-specific coefficients, its training gradient reduces to

$$\nabla_\theta \mathcal{L}_{\text{CM}}(\theta) = \nabla_\theta \mathbb{E} \left[w(t) f_\theta^{\text{SCM}}(x_t, t, c)^\top \frac{df_{\theta^-}^{\text{SCM}}}{dt} \right], \quad (3)$$

where θ^- denotes a stop-gradient copy, w is a time-dependent weight, and the tangent $\frac{df_{\theta^-}^{\text{SCM}}}{dt} = \partial_t f_{\theta^-}^{\text{SCM}} + \nabla_{x_t} f_{\theta^-}^{\text{SCM}} \cdot v_\psi(x_t, t, c)$ is the total derivative along the teacher ODE, computed efficiently via a Jacobian–vector product. Intuitively, minimizing Eq. (3) drives $\frac{df_{\theta^-}^{\text{SCM}}}{dt} \rightarrow 0$, so the prediction stays constant along each teacher trajectory and hence equals its clean endpoint. Because the supervision is tied to individual teacher trajectories, Eq. (3) constrains the noise-to-data correspondence rather than only the output marginal, which underlies the empirically observed broad sample *coverage* of consistency students.

For multi-step sampling, a clean prediction is re-noised to the next scheduled level via the re-noising operator

$$\mathcal{R}_s(\hat{x}_0, \epsilon) = (1 - s)\hat{x}_0 + s\epsilon, \quad \epsilon \sim \mathcal{N}(0, I), \quad (4)$$

after which f_θ^{SCM} is applied iteratively.

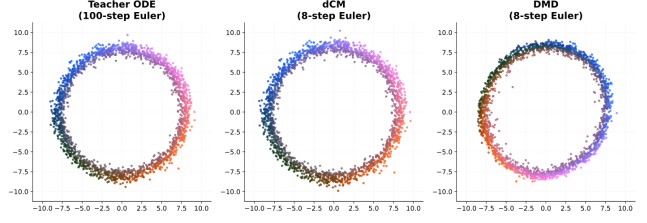


Figure 2: Noise-to-data mappings of different paradigms on a toy ring-shaped Gaussian mixture with a continuum of modes. Color encodes the initial noise.

3.3 Distribution-Level Distillation

Let p_ϕ be the distribution of a few-step generator g_ϕ^{DMD} and p_ψ the teacher distribution. DMD (Yin et al. 2024b,a) optimizes a reverse-KL-style objective

$$\mathcal{L}_{\text{DMD}}(\phi) = D_{\text{KL}}(p_\phi(\cdot | c) \| p_\psi(\cdot | c)), \quad (5)$$

whose practical gradient is estimated from the difference between the teacher score and an online fake score evaluated on student samples. Since samples are drawn from p_ϕ , missing teacher modes receive weak direct pressure; this supplies the familiar *mode-seeking* intuition.

4 Method

4.1 Noise-Data Mapping Conflicts

As two fundamentally different distillation paradigms, DMD and sCM induce distinct optimization preferences: DMD seeks local shortcuts that match the output distribution, whereas sCM enforces path correspondence with the teacher’s trajectory. Starting from the same teacher initialization, the two objectives therefore converge to distinct noise-to-data mappings.

We illustrate with a simple probing experiment. We first train a flow-matching teacher on a ring-shaped 2D Gaussian mixture and distill it separately with sCM and DMD. All students use deterministic 8-step Euler sampling, rather than typical consistency sampling, to preserve mapping consistency. As shown in Figure 2, we pass identically colored initial noise through every model and compare output coordinates, directly visualizing the noise-data correspondence. The sCM student’s outputs remain closely paired with the teacher, whereas DMD realizes a very different mapping.

We hypothesize that, when the two objectives are combined, such a mismatch induces gradient conflicts that lead to hard optimization dynamics and the quality–diversity trade-off, as discussed in Section 2. This suggests that the conflict is intrinsic to forcing one set of parameters to realize both mappings, and can only be sidestepped by keeping the two objectives apart.

4.2 DUET: A Noise-Level Expert Duet

Flow matching models naturally separate semantic formation from detail refinement across noise levels (Wu et al. 2026a; Ren et al. 2026). The high-noise regime determines coarse variables—object identity, layout, camera, and motion pattern—that shape the entire clip, while the low-noise

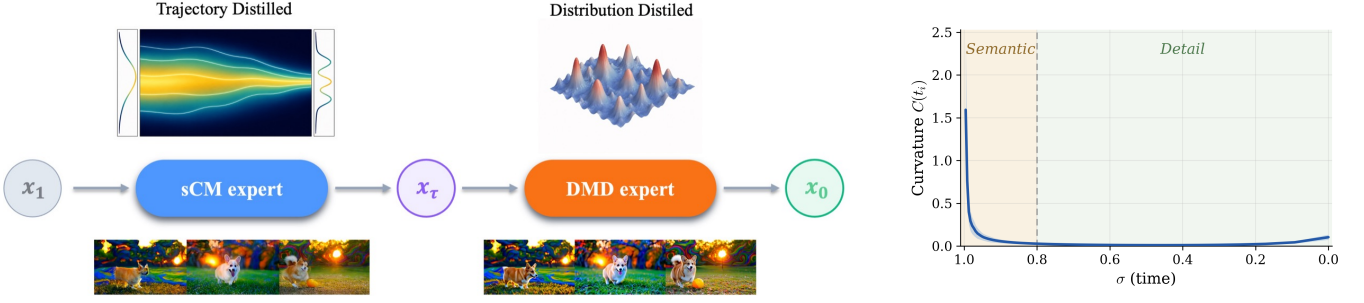


Figure 3: Overview of DUET and its switch-time selection. Left: at inference, the two independently trained experts relay within a single denoising schedule. The sCM expert handles the high-noise interval $t \in [1, \tau]$ to lay out diverse structures; its prediction is then re-noised to $t = \tau$, and the DMD expert completes the low-noise interval $t \in [\tau, 0]$ to refine details. Right: the curvature of the flow-matching teacher is high in the semantic-formation regime and low in the detail-refinement regime; the clear transition motivates setting $\tau = 0.8$.

regime has less freedom to revise this global plan and is better suited for refining edges, texture, color, and other high-frequency appearance. This division mirrors the complementary strengths of the two experts: the sCM expert provides layout diversity and broad coverage, while the DMD expert improves local detail and visual fidelity.

Noise-level division. Before starting, we first probe the two experts with a simple experiment to see the different pattern statistically: we train an sCM expert and a DMD expert separately with their native losses in Eq. (3) and Eq. (5), let each generate multiple videos per prompt in a single step. As shown in Figure 4, the samples of the sCM expert spread over a markedly wider region of the embedding space than those of the DMD expert—an average within-prompt spread about $2.35\times$ larger—confirming its far stronger diversity at the high-noise stage. Moreover, across distinct prompts, the two experts’ sample embeddings land in nearby regions, so the distribution shift is not systematically large, laying the foundation for the relay sampling introduced next.

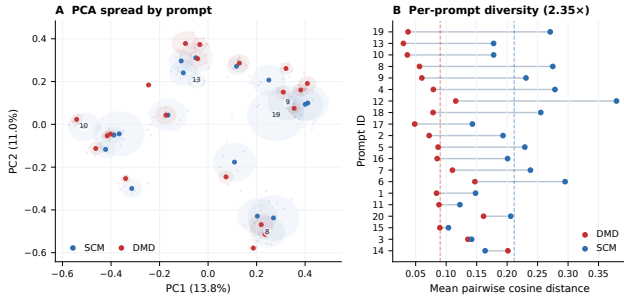


Figure 4: Viclip sample embeddings of the experts.

We therefore preserve both advantages by assigning the high-noise interval $[1, \tau]$ to a coverage-oriented sCM expert f_{θ}^{sCM} and the low-noise interval $[\tau, 0]$ to a quality-oriented DMD expert g_{ϕ}^{DMD} , as shown in Figure 3. In this way, quality and diversity are attained jointly rather than traded off. Since the two experts are trained independently with their native objectives, there is no coefficient to tune between $\mathcal{L}_{\text{CM}}$ and $\mathcal{L}_{\text{DMD}}$, and no gradient conflict to resolve.

Formally, DUET samples in a relay manner: given text c and initial noise x_1 , the sCM expert first predicts a clean endpoint $\hat{x}_0^{\text{sCM}} = f_{\theta}^{\text{sCM}}(x_1, 1, c)$. We then re-noise this prediction to the switch time τ ,

$$x_{\tau}^{\text{sCM}} = \mathcal{R}_{\tau}(\hat{x}_0^{\text{sCM}}, \epsilon_{\tau}), \quad (6)$$

and feed it to the DMD expert for the low-noise step,

$$x_1 \xrightarrow{\mathcal{R}_{\tau} \circ f_{\theta}^{\text{sCM}}} x_{\tau}^{\text{sCM}} \xrightarrow{g_{\phi}^{\text{DMD}}} \hat{x}_0. \quad (7)$$

Algorithm 1 in Appendix B summarizes this. Note that, although two experts are involved, the sampler still uses only two network evaluations, matching the inference cost of a native two-step student while resembling the hard-routing MoE structure of Wan2.2 (Wan et al. 2025).

Determining τ . The switch time τ is critical in this extremely low-budget two-step regime. A high τ (close to pure noise) leaves the semantic layout insufficiently denoised and overly blurry—a degradation the DMD expert cannot recover. Conversely, an overly low τ (close to clean data) places excessive demand on the sCM expert while leaving the DMD expert little room for refinement. We choose τ based on the curvature of the flow-matching teacher’s ODE trajectory (Nie et al. 2026; Feng et al. 2026), which is quantified as

$$C(t_i) \propto \left\| \frac{x_{t_i} - x_{t_{i-1}}}{t_i - t_{i-1}} - (x_1 - x_0) \right\|_2^2. \quad (8)$$

Empirically, we find that in flow matching models such as Wan2.1, curvature is high near the noise endpoint, which we interpret as the fast transformation from noise to structured video, and low near the data endpoint, which we interpret as a slower refinement role. As shown in Figure 3, we set the switch time to $\tau = 0.8$: curvature exceeds 1 before this point and drops below 0.1 afterward.

Figure 1 shows that, despite its simplicity, this direct relay already captures the intended division of labor: the diverse layout and motion of the sCM expert are largely maintained, while the DMD expert sharpens appearance details.

4.3 RL-Guided Expert Adaptation

Although the naive expert duet already yields promising results, neither expert operates at its best, resulting in suboptimal samples, as shown in Figure 5.

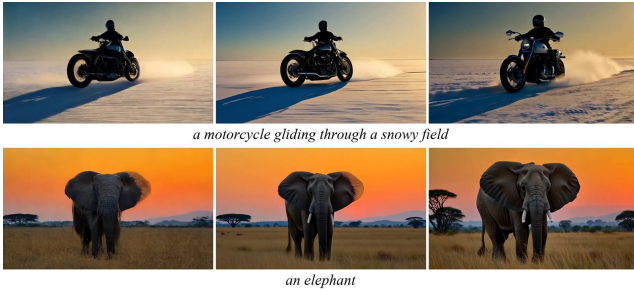


Figure 5: DUET produces high-quality videos, but the direct collaboration is still not optimal. Columns show sCM, DUET, and DUET+ from left to right. DUET+ further improves image quality and visual consistency.

Remaining bottlenecks. We identify two limitations. First, the DMD expert faces a training–inference distribution gap. Let p_{τ}^{sCM} be the distribution of x_{τ}^{sCM} produced by Eq. (6), and let p_{τ}^{DMD} denote that of the intermediate latents produced by the DMD expert’s native backward simulation. Ideal experts, whose predictions from any noise level follow the clean-data distribution, would give $p_{\tau}^{\text{sCM}} \approx p_{\tau}^{\text{DMD}}$; but finite capacity and optimization error generally yield

$$p_{\tau}^{\text{sCM}} \neq p_{\tau}^{\text{DMD}}. \quad (9)$$

The DMD expert is thus evaluated off its training distribution.

Second, the sCM expert’s finite capacity can cap the overall performance. In the layout-first, detail-later paradigm, we identify the layout stage as the bottleneck. This is especially true for video generation, where layout also encodes motion and therefore reflects the model’s understanding of physical dynamics. The performance ceiling therefore rests on the sCM expert, which is also the harder one to train: matching the teacher’s trajectory endpoint requires representing the integral

$$f_{\theta}^{\text{sCM},*}(x_t, t, c) = x_t + \int_t^0 v_{\psi}(x_s, s, c) ds$$

of the teacher velocity field. Precisely in the high-noise interval where the sCM expert operates, the teacher trajectory is highly curved (Figure 3), making this integral hard to fit with limited capacity. The model consequently degrades into a mean-seeking pattern, fitting an intermediate solution between distinct trajectories and sometimes producing physically implausible structures.

Expert-specific adaptation. To address these two limitations, we perform RL-guided expert adaptation for DUET, yielding DUET+. We adapt each expert according to its role.

For the sCM expert, we exploit its diverse samples and use GRPO (Shao et al. 2024; Lu, Zuo, and Deng 2026) to steer this distribution toward the preference distribution defined by selected rewards (Ma et al. 2025; Liu et al. 2026). We treat the sCM expert as a one-step policy: since the relay latent is obtained by re-noising the predicted endpoint, Eq. (6) induces a Gaussian transition kernel $\pi_{\theta}(x_{\tau}^{\text{sCM}} | x_1, c) = \mathcal{N}((1 - \tau)\hat{x}_{0,j}^{\text{sCM}}, \tau^2 I)$, so policy optimization can be applied directly to the sampler used at inference without any

auxiliary SDE. We therefore adopt CM-GRPO (Lu, Zuo, and Deng 2026), which maximizes the advantage-weighted log-likelihood of the sampled relay transitions through the stop-gradient regression objective

$$\begin{aligned} \mathcal{L}_{\text{CM-GRPO}}(\theta) &= \mathbb{E}_{c,j} \left[\left\| \hat{x}_{0,j}^{\text{sCM}} - \text{sg}(\hat{x}_{0,j}^{\text{sCM}} + \Delta_j) \right\|_2^2 \right], \\ \Delta_j &= \frac{(1 - \tau)\hat{A}_j}{2\tau^2} \left(x_{\tau,j}^{\text{sCM}} - (1 - \tau)\hat{x}_{0,j}^{\text{sCM}} \right), \end{aligned} \quad (10)$$

where $\hat{x}_{0,j}^{\text{sCM}} = f_{\theta}^{\text{sCM}}(x_1^j, 1, c)$ is the endpoint prediction of the j -th sampled trajectory and $\text{sg}(\cdot)$ denotes stop-gradient.

The advantage $\hat{A}_j$ is the group-normalized reward. The gradient of Eq. (10) recovers exactly the advantage-weighted score of the Gaussian relay kernel, reinforcing endpoint predictions whose relay latents lead to high-reward videos.

For the DMD expert, we continue training with the DMD loss, but construct the backward simulation with DUET instead of the native rollout. Notably, adaptation proceeds within a single training run: we interleave updates of the sCM expert and DMD expert, so that the DMD expert always adapts to the latest output distribution of the sCM expert.

Direct comparison in Figure 5 demonstrates the effectiveness of this adaptation, with comprehensive quantitative results provided in Section 5.

5 Experiments

In this section, we evaluate DUET through five questions:

Q1: Does DUET jointly deliver DMD quality and sCM diversity? **Q2:** Does RL-guided expert adaptation bring further gains? **Q3:** How does DUET compare with alternatives including rCM (Zheng et al. 2025) and DP-DMD (Wu et al. 2026a)? **Q4:** How do switch time τ and reward choice affect results? **Q5:** Is DUET+ just a better DMD initialization, or do the distinct expert roles matter?

5.1 Experimental Setup

Model and sampling. To demonstrate the effectiveness of DUET, we focus on the rather challenging setting of two-step video generation. We adopt Wan2.1-T2V-1.3B (Wan et al. 2025) as the teacher model, which generates 81-frame text-to-video samples at 832×480 resolution. For sCM training, we use the continuous JVP kernel of rCM (Zheng et al. 2025) for efficiency; for DMD training, we use the distribution-matching objective and basic training setup of DMD2 (Yin et al. 2024a). Detailed configurations are provided in Appendix C. We use the Wan2.1-14B-synthesized data provided by Zheng et al. (2025) for all distillation.

Evaluation. Our evaluation follows the two axes of the quality–diversity trade-off studied in this paper. For *quality*, we use VBench scores as the main benchmark and report the quality-relevant dimensions together with the aggregated quality score. For *diversity*, we measure same-prompt diversity following DP-DMD (Wu et al. 2026a), in the feature spaces of ViCLIP (Wang et al. 2024b), DINO (Caron et al. 2021), and CLIP (Radford et al. 2021).



Figure 6: Qualitative comparisons across DMD, DP-DMD, rCM, sCM, **DUET**, and **DUET+**. Each strip shows first frames of five generations with seeds shared across methods; within-strip variation reflects sample diversity. *Zoom in for best view.*

Method	Diversity $\uparrow$				Quality $\uparrow$							
	ViCLIP	DINO	CLIP	Average	SC	BC	TF	MS	DD	AQ	IQ	Average
sCM	.2085	.2554	.0955	.1865	89.26	91.75	98.19	95.37	55.83	60.69	66.63	81.51
DMD	.0767	.1002	.0412	.0727	94.95	93.30	96.90	96.05	65.79	66.05	68.35	84.38
rCM	.0793	.1105	.0452	.0783 (+7.7%)	93.55	92.86	97.36	93.52	73.89	65.53	67.96	84.26 (+2.75)
DP-DMD	.1183	.1472	.0574	.1076 (+48.0%)	92.04	92.13	97.95	95.20	48.33	62.12	62.47	80.93 (−0.58)
DUET	.1600	.2102	.0834	.1512 (+108.0%)	91.90	93.41	98.38	95.51	67.50	64.94	67.85	83.96 (+2.45)
DUET+	.1620	.2081	.0861	.1521 (+109.2%)	94.13	93.23	98.35	96.19	65.83	65.61	68.14	84.40 (+2.89)

Table 1: Quantitative Results. All methods use exactly two sampling steps (NFE = 2). Parentheses report relative diversity gains over *DMD* and quality point gains over *sCM*. Cells shaded in **red**, **blue**, and **green** mark the best, second- and third-best results in each column.

Baselines. We compare the following methods, *each with just two steps*: 1) **sCM** and **DMD**, which execute two native steps and isolate the two objectives; 2) **rCM**, a loss-level consistency/score-distillation combination (Zheng et al. 2025); 3) **DP-DMD** (Wu et al. 2026a), which assigns ODE regression to the first step and DMD to the second; and 4) our proposed **DUET** and **DUET+**. For fairness, all results are obtained in this work using the same evaluation pipeline.

5.2 Qualitative results

We first present qualitative comparisons to illustrate the effectiveness of **DUET** in Figure 6. Since the outputs are videos, we show only first frames in the main paper and provide more video frames in Appendix E.

We highlight two representative cases. The first prompt is subject-centric, where **DUET** preserves the diverse compositions and poses produced by the sCM expert while substantially improving visual fidelity through the DMD expert. The second prompt tests appearance style. For such a pixel-art prompt, DMD often drifts toward smoother animation-like videos rather than blocky pixel art, suggesting that DMD-dominated distribution matching may prefer high-probability appearance modes while ignoring lower-probability styles. In contrast, **DUET** retains the sCM expert’s coverage of this

lower-probability style while still sharpening local details. Building on the base relay, **DUET+** further enhances the aesthetic quality and repairs residual artifacts left by the direct relay, yielding cleaner and more coherent frames. In parallel, the diversity of rCM remains constrained by the underlying quality–diversity trade-off. DP-DMD loses both fine details and part of the structural diversity: its first stage performs only a simple ODE regression, and the two roles share a single set of parameters rather than explicitly separated experts.

5.3 Quantitative results

Table 1 quantitatively confirms the quality–diversity trade-off and the benefit of our relay design, effectively answering **Q1**, **Q2**, and **Q3**. We make three observations.

First, sCM attains the highest same-prompt diversity but falls far behind DMD on fidelity-oriented VBench dimensions, whereas DMD achieves strong quality at the cost of severe diversity collapse. Compared with rCM and DP-DMD, **DUET** achieves the largest diversity gain over DMD, raising the diversity average from .0727 to .1512 (+108.0%); meanwhile, its quality average rises from sCM’s 81.51 to 83.96 (+2.45 points). Although its diversity score is slightly lower than that of sCM, the visualizations show that **DUET** preserves diverse structures and styles while producing cleaner,

τ	SC	BC	TF	MS	DD	AQ	IQ	Quality
0.3	89.47	92.85	98.52	95.85	59.17	59.89	66.01	81.87
0.8	91.90	93.41	98.38	95.51	67.50	64.94	67.85	83.96
0.934	92.48	93.60	97.92	96.09	58.89	64.07	67.30	83.22

Table 2: Effect of the switch time τ on DUET. The gray cell ($\tau = 0.8$) is our default setting. Red marks the best score in each column.

Reward	SC	BC	TF	MS	DD	AQ*	IQ*	Quality
HPSv3	93.96	94.10	96.76	94.37	61.94	66.86	71.26	84.36
SFS-Nature	91.38	93.15	97.92	94.00	68.61	64.32	69.92	83.84
TA	93.75	93.84	97.98	94.65	57.78	66.41	68.67	83.72

Table 3: Reward-function ablation for CM-GRPO on the sCM expert. $\star$ marks the image-quality-related dimensions we prioritize, and the gray cell (HPSv3) is our default reward. Red marks the best score per column.

more coherent samples, which may lead to a more compact feature-space distribution.

Moreover, **DUET+** keeps this diversity gain (+109.2% over DMD) and further increases the quality improvement over sCM to +2.89 points. Consistent with our qualitative observations, alternatives such as rCM and DP-DMD merely locate a sweet spot along the quality–diversity trade-off, whereas our relay-based methods attain both jointly.

5.4 Ablations

In this section, we answer the remaining questions (**Q4** and **Q5**) by ablating the switch time and the reward choice, and by comparing against init-then-DMD.

Switch Time We compare a small $\tau = 0.3$, the default $\tau = 0.8$, and a large $\tau = 0.934$. For each value of τ , we retrain the corresponding DMD expert with the same switch point in its backward simulation. Table 2 reports the quantitative results, and a qualitative comparison is provided in the Appendix (Figure 9). When τ is small, the sCM expert is overburdened and the DMD expert has little room for refinement, yielding blurry results. When τ is large, the sCM expert contributes too little, leading to under-formed layouts and missing details. The default $\tau = 0.8$, selected from the teacher’s trajectory curvature (Figure 3), balances the two roles and achieves the best scores on most dimensions as well as the highest aggregated quality.

Reward Function We further ablate which preference reward to use for the RL update. To isolate the effect of the reward choice, in this ablation we apply CM-GRPO only to the sCM expert under different rewards, without any DMD-side adaptation, and sample with DUET. Table 3 compares HPSv3 (Ma et al. 2025), SFS-Nature (Wu et al. 2026b), and TA (Liu et al. 2026) on the same VBench dimensions reported in Table 1. HPSv3 is stronger on consistency, aesthetic quality, and imaging quality yet SFS-Nature yields higher Dynamic Degree. Since our primary goal is to improve the fidelity of the sCM expert’s generations, we favor HPSv3 for

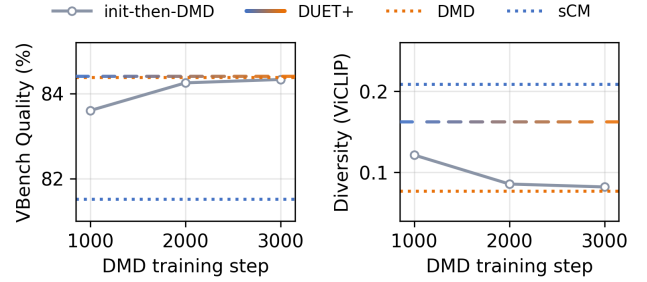


Figure 7: Training dynamics of init-then-DMD, compared with DUET+, DMD, and sCM. Init-then-DMD reaches comparable VBench Quality, but the diversity inherited from the sCM initialization steadily collapses toward the DMD level, remaining well below DUET+.

its improvement on the image-quality deficit caused by the sCM expert’s finite capacity.

Compared to sCM-initialized DMD. **DUET+**’s adaptation resembles a common recipe in DMD-based video distillation: initialize the DMD student from a draft few-step generator—via ODE regression (Yin et al. 2025; Huang et al. 2026) or consistency distillation (Zhu et al. 2026)—and then continue training with the DMD objective, which we call *init-then-DMD*. This mitigates DMD’s local-optimum collapse. We highlight the essential difference and advantage of our method. First, **DUET** is already a strong quality–diversity sampler before adaptation, rather than a warm start. Second, as shown in Figure 7, init-then-DMD reaches comparable VBench Quality but quickly collapses the inherited diversity—about half of **DUET+**. This reveals that the diversity advantage of **DUET+** stems from preserving the distinct roles of the two experts, rather than a better initialization.

6 Conclusion

We propose **DUET**, a noise-level expert duet for two-step video generation: a diversity-preserving sCM expert and a fidelity-oriented DMD expert take the high- and low-noise steps, respectively. Since the two experts are trained independently with their native objectives, **DUET** sidesteps both the quality–diversity trade-off and the optimization difficulty of loss-level combinations. A role-aware, RL-guided expert adaptation further yields **DUET+**, steering the sCM expert toward higher-reward structures and adapting the DMD expert to the actual relay interface. On Wan2.1-T2V-1.3B, **DUET** lifts the two-step quality of sCM to the level of DMD while retaining about twice DMD’s diversity, and **DUET+** reaches DMD-level quality with the diversity advantage largely intact. Ablations validate the switch time τ and the reward choice, and show that the prevalent init-then-DMD recipe instead collapses the inherited diversity. These results establish noise-level distillation expert specialization as a strong few-step paradigm for diverse and high-quality video generation. One limitation is that our evaluation is limited to Wan2.1-T2V-1.3B; however, the properties of sCM and DMD on which **DUET** relies are not backbone-specific, and scaling to larger models like Wan2.1-14B is left for future work.

Acknowledgments

We thank Kaiwen Zheng and Ruibin Li for helpful discussions.

References

- Bai, L.; Zhou, Z.; Shao, S.; Zhong, W.; Yang, S.; Chen, S.; Chen, B.; and Xie, Z. 2026. Optimizing few-step generation with adaptive matching distillation. *arXiv preprint arXiv:2602.07345*.
- Balaji, Y.; Nah, S.; Huang, X.; Vahdat, A.; Song, J.; Zhang, Q.; Kreis, K.; Aittala, M.; Aila, T.; Laine, S.; et al. 2022. ediff-i: Text-to-image diffusion models with an ensemble of expert denoisers. *arXiv preprint arXiv:2211.01324*.
- Brooks, T.; Peebles, B.; Holmes, C.; DePue, W.; Guo, Y.; Jing, L.; Schnurr, D.; Taylor, J.; Luhman, T.; Luhman, E.; Ng, C.; Wang, R.; and Ramesh, A. 2024. Video generation models as world simulators.
- Cai, S.; Nie, W.; Liu, C.; Berner, J.; Zhang, L.; Ma, N.; Chen, H.; Agrawala, M.; Guibas, L.; Wetzstein, G.; et al. 2026. Mode seeking meets mean seeking for fast long video generation. *arXiv preprint arXiv:2602.24289*.
- Caron, M.; Touvron, H.; Misra, I.; Jégou, H.; Mairal, J.; Bojanowski, P.; and Joulin, A. 2021. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, 9650–9660.
- Chen, J.; Xue, S.; Zhao, Y.; Yu, J.; Paul, S.; Chen, J.; Cai, H.; Han, S.; and Xie, E. 2025. Sana-sprint: One-step diffusion with continuous-time consistency distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 16185–16195.
- Cheng, K.; He, X.; Yu, L.; Tu, Z.; Zhu, M.; Wang, N.; Gao, X.; and Hu, J. 2025. Diff-MoE: Diffusion Transformer with Time-Aware and Space-Adaptive Experts. In *Forty-second International Conference on Machine Learning*.
- Fang, G.; Ma, X.; and Wang, X. 2024. Remix-dit: Mixing diffusion transformers for multi-expert denoising. *Advances in Neural Information Processing Systems*, 37: 107494–107512.
- Feng, J.; Cui, J.; Ban, Y.; and Hsieh, C.-J. 2026. One-Forcing: Towards Stable One-Step Autoregressive Video Generation. *arXiv preprint arXiv:2605.23458*.
- Feng, Z.; Zhang, Z.; Yu, X.; Fang, Y.; Li, L.; Chen, X.; Lu, Y.; Liu, J.; Yin, W.; Feng, S.; et al. 2023. Ernie-vilg 2.0: Improving text-to-image diffusion model with knowledge-enhanced mixture-of-denoising-experts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10135–10145.
- Frans, K.; Hafner, D.; Levine, S.; and Abbeel, P. 2025. One step diffusion via shortcut models. In *International Conference on Learning Representations*, volume 2025, 34668–34684.
- Ge, X.; Zhang, Y.; Huang, Y.; He, D.; Wang, X.; Ma, B.; Song, G.; Liu, Y.; and Zhang, J. 2026. Salt: Self-consistent distribution matching with cache-aware training for fast video generation. *arXiv preprint arXiv:2604.03118*.
- Geng, Z.; Deng, M.; Bai, X.; Kolter, Z.; and He, K. 2026. Mean flows for one-step generative modeling. *Advances in Neural Information Processing Systems*, 38: 75460–75482.
- HaCohen, Y.; Chiprut, N.; Brazowski, B.; Shalem, D.; Moshe, D.; Richardson, E.; Levin, E.; Shiran, G.; Zabari, N.; Gordon, O.; et al. 2024. Ltx-video: Realtime video latent diffusion. *arXiv preprint arXiv:2501.00103*.
- Heek, J.; Hoogeboom, E.; and Salimans, T. 2024. Multistep consistency models. *arXiv preprint arXiv:2403.06807*.
- Huang, X.; Li, Z.; He, G.; Zhou, M.; and Shechtman, E. 2026. Self forcing: Bridging the train-test gap in autoregressive video diffusion. *Advances in Neural Information Processing Systems*, 38: 167283–167308.
- Jiang, D.; Liu, D.; Wang, Z.; Wu, Q.; Li, L.; Li, H.; Jin, X.; Liu, D.; Lu, C.; Li, Z.; et al. 2025. Distribution matching distillation meets reinforcement learning. *arXiv preprint arXiv:2511.13649*.
- Kim, D.; Lai, C.-H.; Liao, W.; Murata, N.; Takida, Y.; Uesaka, T.; He, Y.; Mitsufuji, Y.; and Ermon, S. 2024. Consistency trajectory models: Learning probability flow ode trajectory of diffusion. In *International Conference on Learning Representations*, volume 2024, 44493–44525.
- Lee, S.; Xu, Y.; Geffner, T.; Fanti, G.; Kreis, K.; Vahdat, A.; and Nie, W. 2024. Truncated consistency models. *arXiv preprint arXiv:2410.14895*.
- Lin, S.; Xia, X.; Ren, Y.; Yang, C.; Xiao, X.; and Jiang, L. 2025. Diffusion adversarial post-training for one-step video generation. *arXiv preprint arXiv:2501.08316*.
- Lin, S.; Yang, C.; He, H.; Jiang, J.; Ren, Y.; Xia, X.; Zhao, Y.; Xiao, X.; and Jiang, L. 2026. Autoregressive adversarial post-training for real-time interactive video generation. *Advances in Neural Information Processing Systems*, 38: 41061–41086.
- Lipman, Y.; Chen, R. T.; Ben-Hamu, H.; Nickel, M.; and Le, M. 2022. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*.
- Liu, J.; Liu, G.; Liang, J.; Yuan, Z.; Liu, X.; Zheng, M.; Wu, X.; Wang, Q.; Xia, M.; Wang, X.; et al. 2026. Improving video generation with human feedback. *Advances in Neural Information Processing Systems*, 38: 82155–82192.
- Lu, C.; and Song, Y. 2025. Simplifying, stabilizing and scaling continuous-time consistency models. In *International Conference on Learning Representations*, volume 2025, 50611–50649.
- Lu, Y.; Zuo, R.; and Deng, J. 2026. RAVEN: Real-time Autoregressive Video Extrapolation with Consistency-model GRPO. *arXiv preprint arXiv:2605.15190*.
- Luo, W.; Huang, Z.; Geng, Z.; Kolter, J. Z.; and Qi, G.-j. 2024. One-step diffusion distillation through score implicit matching. *Advances in Neural Information Processing Systems*, 37.
- Lv, Z.; Si, C.; Pan, T.; Chen, Z.; Wong, K.-Y. K.; Qiao, Y.; and Liu, Z. 2025. Dual-expert consistency model for efficient and high-quality video generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 14983–14993.

- Ma, Y.; Wu, X.; Sun, K.; and Li, H. 2025. Hpsv3: Towards wide-spectrum human preference score. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 15086–15095.
- Nie, W.; Berner, J.; Ma, N.; Liu, C.; Xie, S.; and Vahdat, A. 2026. Transition Matching Distillation for Fast Video Generation. *arXiv preprint arXiv:2601.09881*.
- Pan, Z.; Zhuang, B.; Huang, D.-A.; Nie, W.; Yu, Z.; Xiao, C.; Cai, J.; et al. 2025. T-stitch: Accelerating sampling in pre-trained diffusion models with trajectory stitching. In *International Conference on Learning Representations*, volume 2025, 6103–6137.
- Park, B.; Go, H.; Kim, J.-Y.; Woo, S.; Ham, S.; and Kim, C. 2024. Switch diffusion transformer: Synergizing denoising tasks with sparse mixture-of-experts. In *European Conference on Computer Vision*, 461–477. Springer.
- Peebles, W.; and Xie, S. 2023. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, 4195–4205.
- Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, 8748–8763. PmLR.
- Ren, S.; Yu, Q.; He, J.; Shen, X.; and Chen, L.-C. 2026. Frequency-Aware Flow Matching for High-Quality Image Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9074–9083.
- Ren, Y.; Xia, X.; Lu, Y.; Zhang, J.; Wu, J.; Xie, P.; Wang, X.; and Xiao, X. 2024. Hyper-sd: Trajectory segmented consistency model for efficient image synthesis. *Advances in neural information processing systems*, 37: 117340–117362.
- Sabour, A.; Fidler, S.; and Kreis, K. 2026. Align your flow: Scaling continuous-time flow map distillation. *Advances in Neural Information Processing Systems*, 38: 146459–146512.
- Salimans, T.; and Ho, J. 2022. Progressive distillation for fast sampling of diffusion models. *arXiv preprint arXiv:2202.00512*.
- Sauer, A.; Boesel, F.; Dockhorn, T.; Blattmann, A.; Esser, P.; and Rombach, R. 2024a. Fast high-resolution image synthesis with latent adversarial diffusion distillation. *arXiv preprint arXiv:2403.12015*.
- Sauer, A.; Lorenz, D.; Blattmann, A.; and Rombach, R. 2024b. Adversarial diffusion distillation. In *European Conference on Computer Vision*.
- Shao, Z.; Wang, P.; Zhu, Q.; Xu, R.; Song, J.; Bi, X.; Zhang, H.; Zhang, M.; Li, Y.; Wu, Y.; et al. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.
- Song, J.; Meng, C.; and Ermon, S. 2020. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*.
- Song, Y.; Dhariwal, P.; Chen, M.; and Sutskever, I. 2023. Consistency models.
- Team, K.; Chen, J.; Ci, Y.; Du, X.; Feng, Z.; Gai, K.; Guo, S.; Han, F.; He, J.; He, K.; et al. 2025. Kling-Omni Technical Report. *arXiv preprint arXiv:2512.16776*.
- Wan, T.; Wang, A.; Ai, B.; Wen, B.; Mao, C.; Xie, C.-W.; Chen, D.; Yu, F.; Zhao, H.; Yang, J.; et al. 2025. Wan: Open and advanced large-scale video generative models. *arXiv preprint arXiv:2503.20314*.
- Wang, F.-Y.; Huang, Z.; Bergman, A. W.; Shen, D.; Gao, P.; Lingelbach, M.; Sun, K.; Bian, W.; Song, G.; Liu, Y.; et al. 2024a. Phased consistency models. *Advances in neural information processing systems*, 37: 83951–84009.
- Wang, Y.; He, Y.; Li, Y.; Li, K.; Yu, J.; Ma, X.; Li, X.; Chen, G.; Chen, X.; Wang, Y.; et al. 2024b. Internvid: A large-scale video-text dataset for multimodal understanding and generation. In *International Conference on Learning Representations*, volume 2024, 42055–42079.
- Wei, X.; Li, Z.; Yan, S.; Zhou, C.; and Zhang, M. 2026. FlashMol: High-Quality Molecule Generation in as Few as Four Steps. *arXiv preprint arXiv:2605.07020*.
- Wu, T.; Li, R.; Zhang, L.; and Ma, K. 2026a. Diversity-Preserved Distribution Matching Distillation for Fast Visual Synthesis. *arXiv preprint arXiv:2602.03139*.
- Wu, Y.-S.; Luo, R.; Zhu, J.; Tu, T.; Farhadi, A.; Wallingford, M.; Wang, Y.-C. F.; Marschner, S.; and Ma, W.-C. 2026b. Seeing Fast and Slow: Learning the Flow of Time in Videos. *arXiv preprint arXiv:2604.21931*.
- Xu, Y.; Nie, W.; and Vahdat, A. 2025. One-step diffusion models with f-divergence distribution matching. *arXiv preprint arXiv:2502.21148*.
- Yang, S.; Chen, Y.; Wang, L.; Liu, S.; and Chen, Y. 2024. Denoising diffusion step-aware models. In *International Conference on Learning Representations*, volume 2024, 13137–13152.
- Yin, T.; Gharbi, M.; Park, T.; Zhang, R.; Shechtman, E.; Durand, F.; and Freeman, W. T. 2024a. Improved distribution matching distillation for fast image synthesis. *Advances in neural information processing systems*, 37: 47455–47487.
- Yin, T.; Gharbi, M.; Zhang, R.; Shechtman, E.; Durand, F.; Freeman, W. T.; and Park, T. 2024b. One-step diffusion with distribution matching distillation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 6613–6623.
- Yin, T.; Zhang, Q.; Zhang, R.; Freeman, W. T.; Durand, F.; Shechtman, E.; and Huang, X. 2025. From slow bidirectional to fast autoregressive video diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 22963–22974.
- Zheng, K.; Wang, Y.; Ma, Q.; Chen, H.; Zhang, J.; Balaji, Y.; Chen, J.; Liu, M.-Y.; Zhu, J.; and Zhang, Q. 2025. Large scale diffusion distillation via score-regularized continuous-time consistency. *arXiv preprint arXiv:2510.08431*.
- Zhou, M.; Zheng, H.; Wang, Z.; Yin, M.; and Huang, H. 2024. Score identity distillation: Exponentially fast distillation of pretrained diffusion models for one-step generation. In *International Conference on Machine Learning*.

Zhu, H.; Zhao, M.; He, G.; Su, H.; Li, C.; and Zhu, J. 2026. Causal Forcing: Autoregressive Diffusion Distillation Done Right for High-Quality Real-Time Interactive Video Generation. *arXiv preprint arXiv:2602.02214*.

Zou, K.; Zheng, D.; Liu, H.; Hang, T.; Liu, B.; and Yu, N. 2026. Hiar: Efficient autoregressive long video generation via hierarchical denoising. *arXiv preprint arXiv:2603.08703*.

A Additional Related Work

Noise-level expert specialization. A series of works observe that different noise levels of the generative process demand different capabilities, and therefore specialize model capacity along the denoising schedule. eDiff-I (Balaji et al. 2022) trains an ensemble of expert denoisers for different noise intervals, ERNIE-ViLG 2.0 (Feng et al. 2023) adopts a mixture of denoising experts across uniformly divided stages, and follow-up timestep-aware MoE architectures further extend this design (Park et al. 2024; Fang, Ma, and Wang 2024; Cheng et al. 2025); Wan2.2 (Wan et al. 2025) instantiates this idea at scale with a hard-routed high-noise/low-noise expert pair. Another line targets efficiency by assigning models of different sizes to different timesteps (Pan et al. 2025; Yang et al. 2024). Nevertheless, all of these specialize capacity under a *single diffusion training paradigm*, primarily to improve *fidelity or efficiency in many-step sampling*. In the distillation scenario, DCM (Lv et al. 2025) decouples video consistency distillation into a semantic expert and a detail expert trained on high- and low-noise samples, respectively; however, it uses consistency losses for both experts and does not fully exploit the respective advantages of DMD and sCM. **DUET** is novel in that we assign *two different distillation objectives* with complementary behaviors to the two noise regimes—a coverage-seeking sCM expert for high-noise structure and a mode-seeking DMD expert for low-noise refinement—and relay between independently trained experts in the extreme *two-step generation* regime, targeting the win-win in *quality-diversity*.

B Method Details

Algorithm 1 summarizes the two-step **DUET** sampler and the RL-guided expert adaptation schedule described in the main paper.

Algorithm 1 **DUET** Inference and RL-Guided Expert Adaptation

Require: text c , switch time τ , sCM expert f_{θ}^{sCM} , DMD expert g_{ϕ}^{DMD} , fake-score critic, sCM updates per round N_{sCM} , critic updates per round N_{critic}

- 1: **Two-step inference**
- 2: Sample $x_1, \epsilon_{\tau} \sim \mathcal{N}(0, I)$
- 3: $x_{\tau}^{\text{sCM}} \leftarrow \mathcal{R}_{\tau}(f_{\theta}^{\text{sCM}}(x_1, 1, c), \epsilon_{\tau})$
- 4: $\hat{x}_0 \leftarrow g_{\phi}^{\text{DMD}}(x_{\tau}^{\text{sCM}}, \tau, c)$
- 5: Output decoded video $\text{Dec}(\hat{x}_0)$
- 6: **RL-guided expert adaptation**
- 7: **for** each adaptation round **do**
- 8: Update g_{ϕ}^{DMD} with $\mathcal{L}_{\text{DMD}}$ using relay latents from the current f_{θ}^{sCM}
- 9: **for** $i = 1, \dots, N_{\text{sCM}}$ **do**
- 10: Update f_{θ}^{sCM} by CM-GRPO (Eq. (10))
- 11: **end for**
- 12: **for** $i = 1, \dots, N_{\text{critic}}$ **do**
- 13: Update the critic with the denoising score-matching loss on samples generated by the current relay
- 14: **end for**
- 15: **end for**
- 16: **return** adapted experts $f_{\theta}^{\text{sCM}}, g_{\phi}^{\text{DMD}}$ as **DUET+**

C Training and Evaluation Configurations

C.1 Training Configurations

Training devices. All training runs are conducted on 16–32 NVIDIA H800 GPUs, with gradient accumulation adjusted accordingly for the global batch size of 64. Training the sCM expert, the DMD expert, and **DUET+** takes approximately 1,024, 360, and 128 GPU hours, respectively.

Common settings. All four methods use classifier-free guidance 5.0, timestep shift 5.0, bfloat16 precision, FusedAdamW, global batch size 64, and generator learning rate 2×10^{-6} . The base students, including the sCM expert, DMD expert, and DP-DMD, are initialized from the teacher; **DUET+** is initialized from the two trained experts. For the methods using the DMD-side fake-score critic, we set the critic learning rate to 4×10^{-7} , critic weight decay to 0.01, critic updates to $N_{\text{critic}} = 5$, and switch time to $\tau = 0.8$. Table 4 summarizes the remaining method-specific configurations, which are further detailed as follows.

- **sCM expert configuration.** The sCM expert is trained with the sCM objective in Eq. (3), LogNormal time sampling with mean 0.7 and standard deviation 1.6, and the tangent warm-up schedule of sCM (Lu and Song 2025) for 1,000 iterations.
- **DP-DMD configuration.** DP-DMD follows the official recipe (Wu et al. 2026a) with diversity weight 0.05 and $K = 5$ anchor steps, keeping the same critic setting for comparability.

- **rCM.** The rCM baseline is not trained by us. We use the publicly released Wan2.1-T2V-1.3B 480p checkpoint from Hugging Face, linked from the official rCM repository (Zheng et al. 2025). Note that, as stated in its model card, the released checkpoint is not trained under the same setting as the rCM paper: it is reproduced with limited synthetic data, whereas the original rCM uses high-quality internal data, and may therefore perform worse than the officially reported results. Nevertheless, the comparison remains fair, since all methods in this work are likewise trained on the same synthetic dataset.
- **DUET+ configuration.** DUET+ is initialized from the two trained experts above and optimized jointly for 500 iterations, where each cycle includes one DMD step, $N_{\text{sCM}} = 5$ CM-GRPO updates, and $N_{\text{critic}} = 5$ critic updates, following Algorithm 1. CM-GRPO uses HPSv3 as the reward, a group size of 16 with 8 groups per policy update, and a 4-step sCM expert rollout on the shifted grid, where the last transition skips to the clean endpoint and a random single transition receives the policy update; we do not use a KL loss, and advantages are clipped at 5.0. For the sCM expert optimizer, we set the weight decay to 0 and Adam’s ϵ to 10^{-10} .

	sCM expert	DMD expert	DP-DMD	DUET+
Training iterations	5,000	3,000	3,000	500
Adam betas	(0.9, 0.999)	(0.0, 0.999)	(0.0, 0.999)	(0.0, 0.999)
EMA	Yes	Yes	Yes	No

Table 4: Training configurations of the sCM expert, the DMD expert, DP-DMD, and DUET+.

C.2 Evaluation Configurations

Evaluation details. All methods are evaluated with exactly two sampling steps ($\text{NFE} = 2$). For VBench, we generate videos with the prompt suites of the dimensions reported in Table 1, using the augmented prompts provided by rCM (Zheng et al. 2025); each prompt is sampled 5 times, except for the Temporal Flickering prompts, which are sampled 25 times. For diversity, we follow DP-DMD (Wu et al. 2026a) and compute the same-prompt diversity

$$D = 1 - \frac{2}{R(R-1)} \sum_{i < j} \cos(x^{(i)}, x^{(j)}), \quad (11)$$

where $x^{(1)}, \dots, x^{(R)}$ are the ℓ_2 -normalized embeddings of the R videos generated from the same prompt. For the DINO and CLIP feature spaces, we uniformly sample 8 frames per video, extract frame-level features, and average-pool them into a video embedding; for ViCLIP, the 8 uniformly sampled frames are directly encoded into a video-level embedding. Diversity is computed for every prompt of the full VBench prompt suite ($R = 5$, or $R = 25$ for the Temporal Flickering prompts) and then averaged over prompts.

Inference cost. Since the two experts of DUET each execute exactly one of the two sampling steps, DUET doubles the parameter count kept in memory, but its FLOPs and latency are identical to those of a native two-step student.

Explicit Trade-off of rCM. rCM exhibits an explicit quality–diversity trade-off, which can be controlled through the σ_{max} parameter of its EDM-style initial sampling step: a large σ_{max} substantially reduces diversity while improving quality, and vice versa. Figure 8 visualizes the generations under the two settings: with $\sigma_{\text{max}} = 80$, the default of the official inference script, rCM trades quality for diversity, yet both its diversity and quality remain below those of DUET. To preserve the generation quality of rCM, we report its results under $\sigma_{\text{max}} = 1600$. For our DUET, we also take $\sigma_{\text{max}} = 1600$ for fairness.

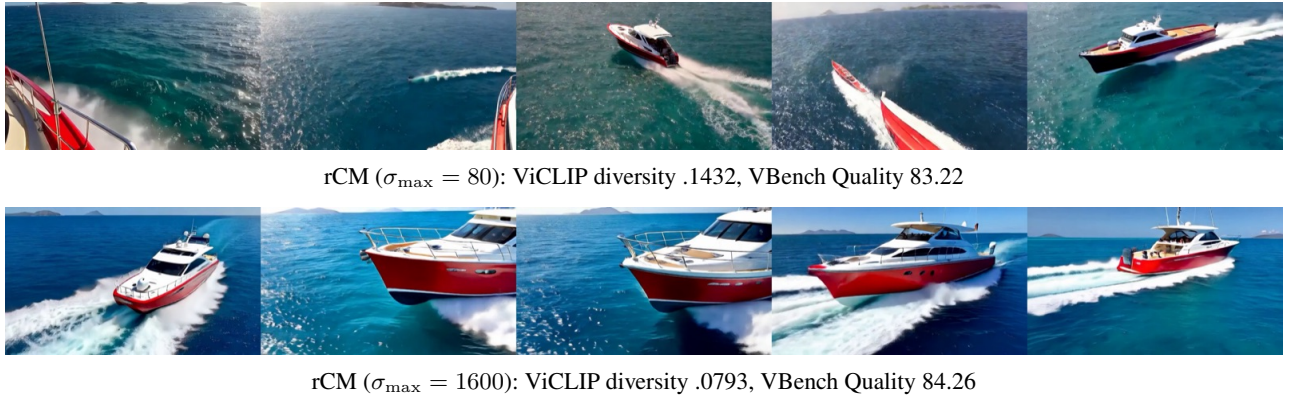


Figure 8: Visualization of the σ_{max} trade-off of rCM. Each row shows the first frames of five same-prompt generations (“A boat accelerating to gain speed”) with the corresponding scores below. $\sigma_{\text{max}} = 80$ yields diverse compositions with degraded quality, whereas $\sigma_{\text{max}} = 1600$ improves quality but collapses the generations to nearly identical layouts.

D More Experimental Results

Qualitative comparison across switch times. Figure 9 shows samples of **DUET** at different switch times τ on two prompts, using the same initial noise within each prompt block. The default $\tau = 0.8$ preserves rich structural details while inheriting the high fidelity of the DMD expert; $\tau = 0.934$ loses part of the fine details, and $\tau = 0.3$ can hardly inherit the DMD expert’s fidelity at all.

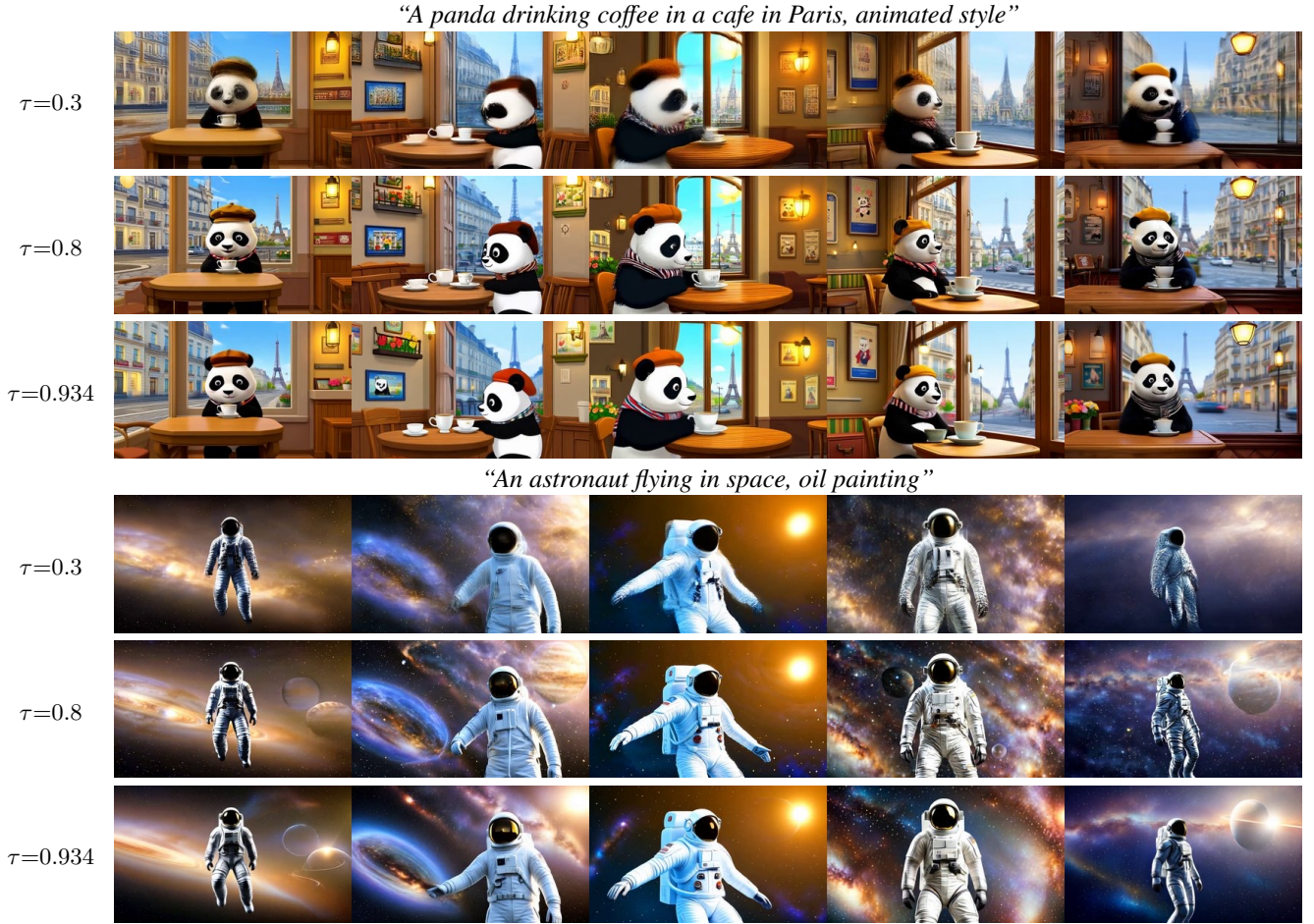


Figure 9: Qualitative comparison of **DUET** under different switch times τ on two prompts. The default $\tau = 0.8$ preserves rich structural details while inheriting the high fidelity of the DMD expert; $\tau = 0.934$ loses part of the fine details, and $\tau = 0.3$ can hardly inherit the DMD expert’s fidelity at all.

Reward curves of CM-GRPO. Figure 10 plots the training reward curve of CM-GRPO applied to the sCM expert with the HPSv3 reward, the setting adopted by **DUET+**. The reward mean rises steadily over the roughly 300 optimization iterations, indicating a stable preference-optimization process.

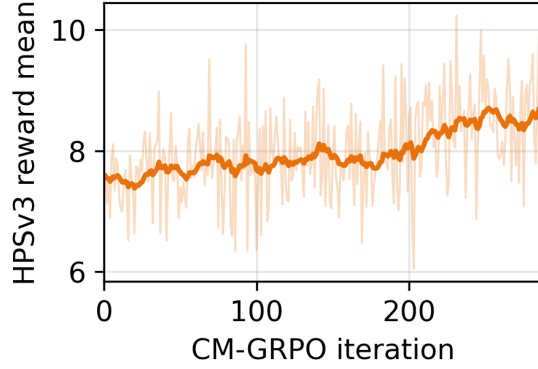


Figure 10: Training reward curve of CM-GRPO on the sCM expert with the HPSv3 reward. The light curve shows the per-iteration reward mean and the dark curve shows an EMA-smoothed version.

E Additional Qualitative Results

We provide additional qualitative results on five new prompts, covering all six models in Table 1. For each prompt, every model generates five videos with the same random seed, so the i -th generation of every model shares the same initial noise. Figures 11–15 show the first frames of all five generations and visualize sample diversity, while Figures 16–20 show five uniformly spaced frames of one randomly selected generation per prompt, with the generation index kept identical across models, visualizing temporal consistency.



Figure 11: First frames of five generations on the prompt “A cute happy Corgi playing in park, sunset, pixel art”; each column is one shared seed.



Figure 12: First frames of five generations on the prompt “A person playing guitar”; each column is one shared seed.



Figure 13: First frames of five generations on the prompt “A raccoon that looks like a turtle, digital art”; each column is one shared seed.



Figure 14: First frames of five generations on the prompt “A storm trooper vacuuming the beach”; each column is one shared seed.

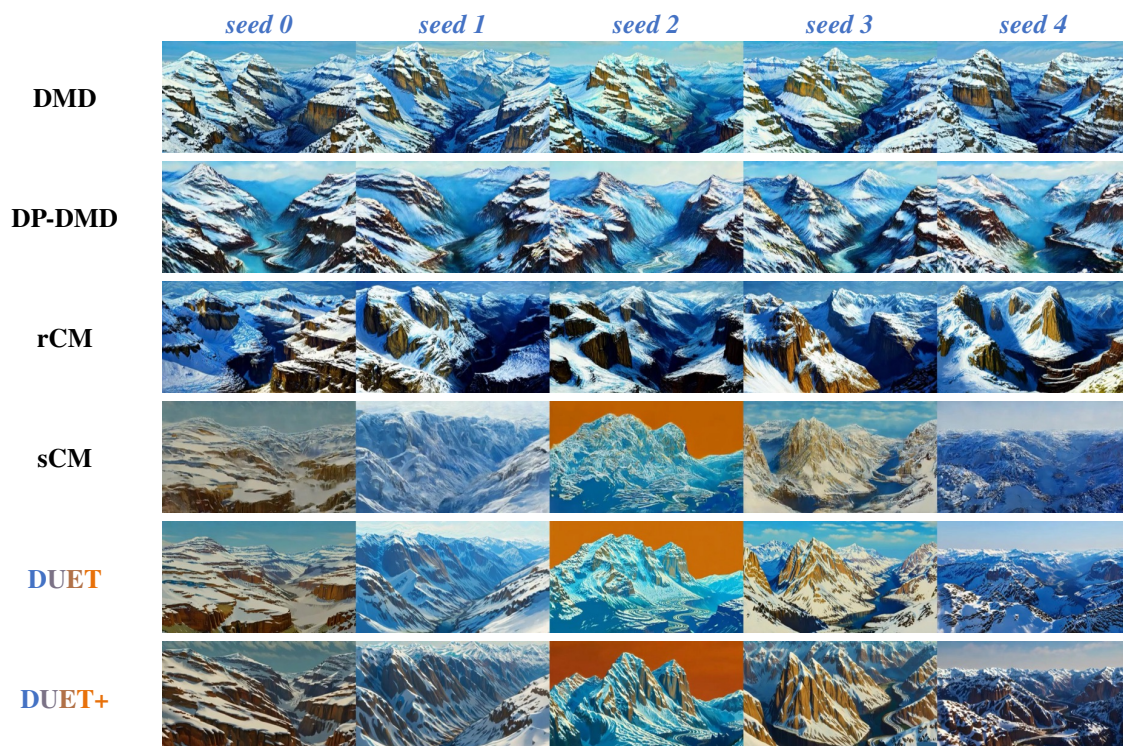


Figure 15: First frames of five generations on the prompt “Snow rocky mountains peaks canyon. snow blanketed rocky mountains surround and shadow deep canyons. the canyons twist and bend through the high elevated mountain peaks”; each column is one shared seed.



Figure 16: Five frames of generation #4 on the prompt “A cute happy Corgi playing in park, sunset, pixel art” (shared across models); each column is one time step.

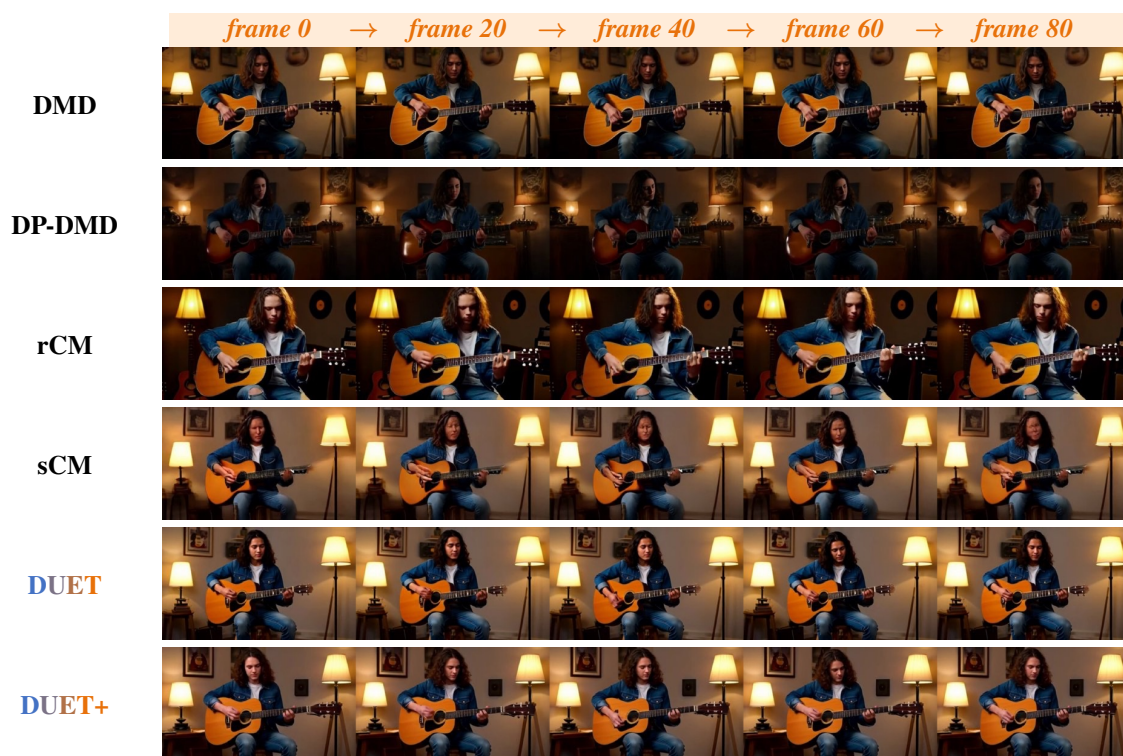


Figure 17: Five frames of generation #1 on the prompt “A person playing guitar” (shared across models); each column is one time step.

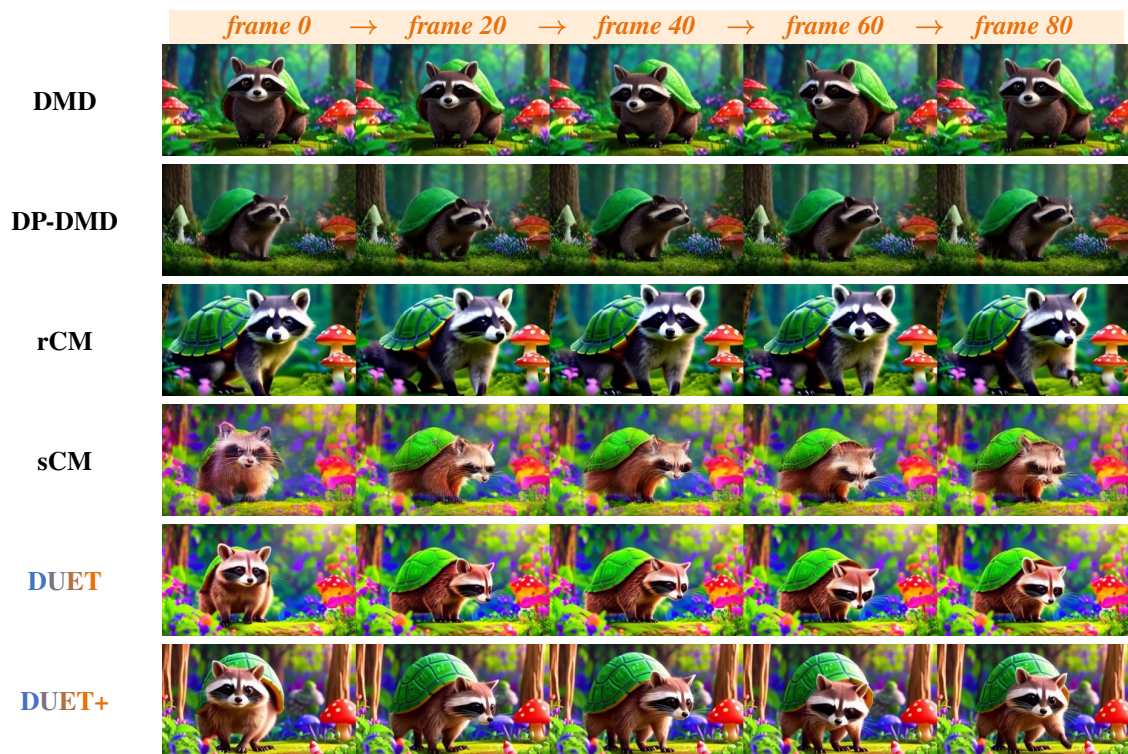


Figure 18: Five frames of generation #4 on the prompt “A raccoon that looks like a turtle, digital art” (shared across models); each column is one time step.

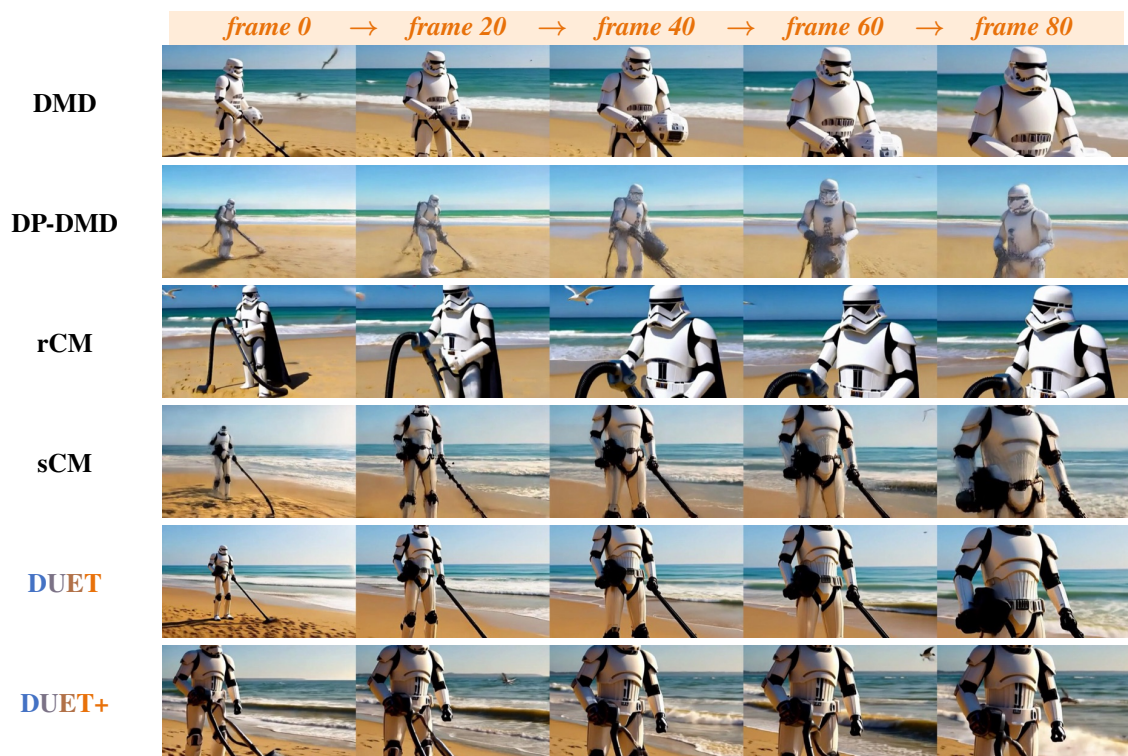


Figure 19: Five frames of generation #4 on the prompt “A storm trooper vacuuming the beach” (shared across models); each column is one time step.

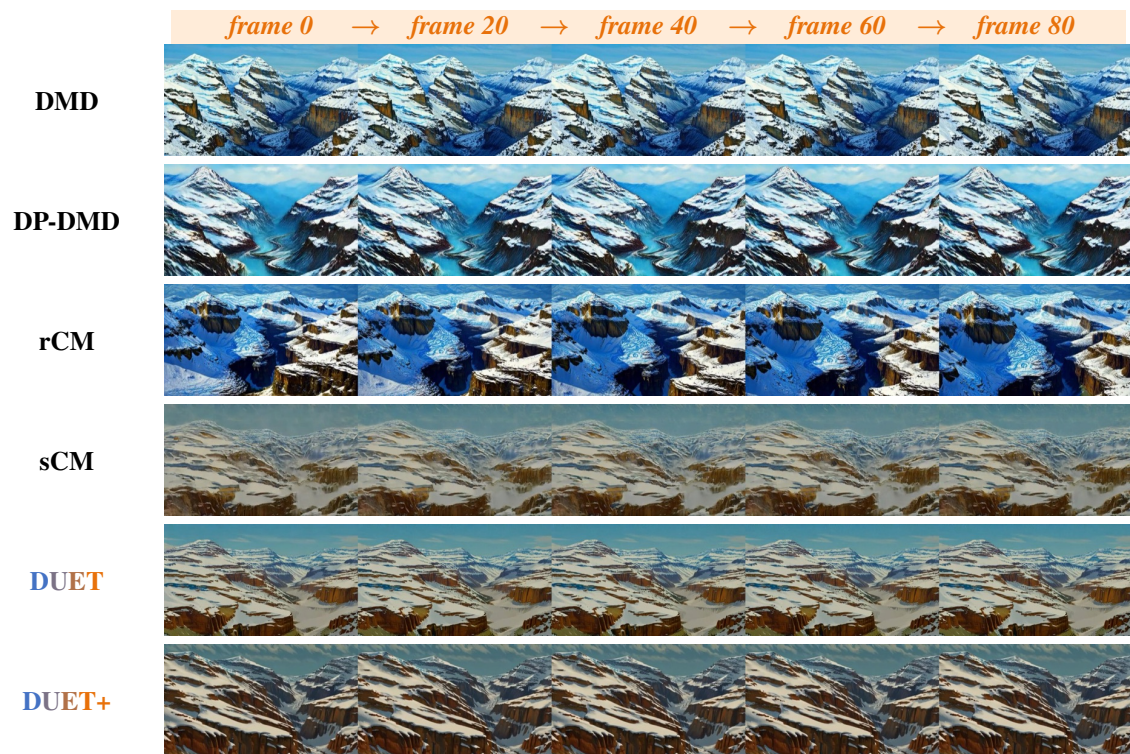


Figure 20: Five frames of generation #1 on the prompt “*Snow rocky mountains peaks canyon. snow blanketed rocky mountains surround and shadow deep canyons. the canyons twist and bend through the high elevated mountain peaks*” (shared across models); each column is one time step.