PDMD: PROJECTED DISTRIBUTION MATCHING DISTILLATION FOR VIDEO DIFFUSION MODELS

Anonymous authors
Paper under double-blind review

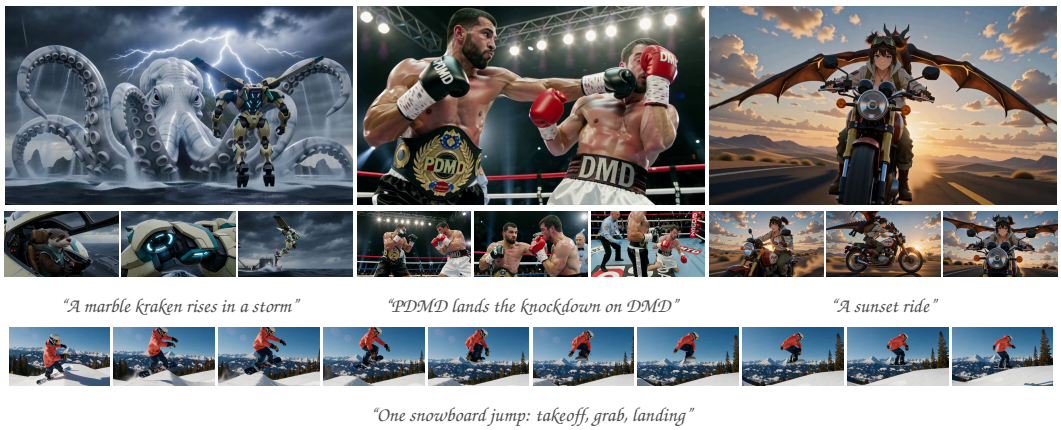

Figure 1: **4-NFE video generation with PDMD.** The 14.4-second, 1366×768 videos generated by PDMD-distilled MiniMax-H3 (MiniMax, 2026a) exhibit natural colors and dynamic motion.

ABSTRACT

Modern video diffusion models require tens of denoising evaluations over long spatiotemporal token sequences. Distribution Matching Distillation (DMD) reduces the number of function evaluations (NFE) to just a few. However, DMD samples can degrade during training, exhibiting progressive oversaturation and artifacts. We trace this instability to critic errors, which enter successive student updates and accumulate over time. We introduce Projected Distribution Matching Distillation (PDMD) to filter critic errors. PDMD projects out the component of the DMD update parallel to the student–critic endpoint residual. At a fixed noisy query, we prove that this residual is an unbiased estimate of the critic's endpoint error. Under high-dimensional assumptions, this projection removes a constant fraction of critic error while discarding only a vanishing fraction of ideal DMD signal. Empirically, the projection stabilizes training and improves sample quality where DMD degrades and develops unnatural textures. PDMD requires only a one-line code change to DMD, with no extra loss, network, data, model pass, or multi-stage training. With Wan2.1, PDMD achieves a VBench total score of 83.73 at 4 NFE, surpassing matched DMD by 1.03 points. On MiniMax-H3 joint video–audio generation, PDMD achieves a VideoGen-Eval visual total score of 83.17, 0.41 points above the strongest distilled baseline. PDMD also achieves the best performance on all six audio metrics among the compared 4-NFE models. Qualitative comparisons and user studies favor PDMD over the distilled baselines in visual quality, motion, and audio quality. Code and models will be publicly available.

1 INTRODUCTION

Video generation has advanced rapidly in visual quality, motion fidelity, and prompt following, as demonstrated by Sora (Brooks et al., 2024), Wan (Team Wan et al., 2025), Seedance (ByteDance Seed, 2025), and MiniMax (MiniMax, 2026a). This progress comes at substantial inference cost: a modern video diffusion transformer (Rombach et al., 2022; Peebles & Xie, 2023) typically requires a large number of function evaluations (NFE), often dozens, each processing a long spatiotemporal token

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



Figure 2: **Qualitative comparisons over matched 4-NFE training.** DMD and DMD2 develop unnatural textures after 500 iterations, whereas PDMD keeps a natural appearance.

sequence. Step distillation (Salimans & Ho, 2022; Song et al., 2023) reduces this cost by compressing a many-step teacher into a few-step student that requires only a few evaluations. Distribution Matching Distillation (DMD) (Yin et al., 2024b) is a widely used approach to step distillation. It trains the student to match the distribution produced by the many-step teacher, using the difference between the student score (learned online by a model called the critic) and the frozen teacher score.

The critic is trained alternately with the student, using student-generated samples to track current student's distribution. However, the online critic has approximation error even for a fixed student and can lag behind a changing one. (Shao et al., 2025; Yin et al., 2024a). Because the critic score enters the DMD update directly, *critic errors* are repeatedly transferred to the student and can accumulate throughout training. The critic error can lead to training instability, with samples becoming progressively oversaturated and eventually losing quality (Shao et al., 2025; Yin et al., 2024a). For example, on MiniMax-H3, DMD becomes visibly oversaturated from 500 iterations onward (Figure 2).

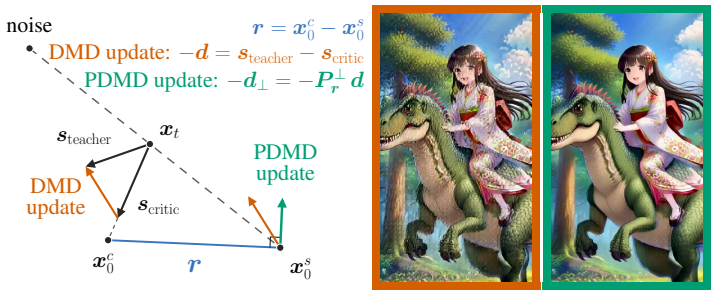

Figure 3: **DMD and PDMD updates.** DMD moves the student along $-d$. The critic score s_{critic}, the online-learning part of d, carries hidden error. PDMD projects out the component of d parallel to the known residual r to suppress critic error, because r is the critic error estimate. The two frames decode one real update target: the PDMD target contains fewer artifacts than the DMD target, while keeping 83% of the update norm.

DMD2 (Yin et al., 2024a) reduces critic errors by taking more critic updates per student update, and also adds a discriminator to improve quality. Other methods combine DMD with additional objectives (Zheng et al., 2026; Gu et al., 2026). We look beyond these additions and ask whether DMD can reduce critic error using what it already computes. Our goal is to use an estimate of the critic error direction that costs no extra data or model pass, to cancel part of the critic error with the estimate, and keep most of the DMD signal. Our key insight is that the direction from the student sample to the critic's one-step denoised prediction, namely the student–critic endpoint residual, provides an estimate of the critic error direction (Section 4.1).

To filter and attenuate critic error, we present Projected Distribution Matching Distillation (PDMD), which projects out the parallel component to the critic error estimate of the DMD update. PDMD is a one-line change to DMD, with no auxiliary objective, network, model pass, data, or training stage. Under high-dimensional concentration and weak-alignment assumptions, our analysis shows PDMD removes a constant fraction of critic error and only a vanishing fraction of useful DMD signal.

As illustrated by Figure 3, PDMD recasts improving DMD as a filtering problem inside the existing update. DMD re-noises a student sample x_0^s to x_t and evaluates the critic score s_{critic} and the frozen teacher score s_{teacher} at the same query. The score difference, $d = s_{\text{critic}} - s_{\text{teacher}}$, defines its student update. However, the critic error within s_{critic} is not directly observable, and cannot simply be subtracted from d. The critic score can be linearly converted to a one-step endpoint x_0^c with no extra model pass. Fortunately, the student–critic endpoint residual $r = x_0^c - x_0^s$ is observable and a critic error estimator. For a fixed query x_t, the ideal critic endpoint is the average student endpoint.

Therefore, the observed student–critic endpoint residual estimates critic error, with a sample-variation term that averages to zero. We prove that, at a fixed query $\boldsymbol{x}_t$, $\boldsymbol{r}$ is an unbiased estimate of the critic endpoint error in Equation (7). Projecting out the component of $\boldsymbol{d}$ parallel to $\boldsymbol{r}$ suppresses critic error. A visual example is provided at right of Figure 3.

The filtering shows up empirically as more stable training, less oversaturation, more motion, and better audio. In 2D analysis, PDMD improves energy distance, on-mode fraction, and collapses to a single mode in $3/20$ runs, against $14/20$ without the projection (Figure 5; Supp. Table B3). On Wan2.1-T2V-1.3B, PDMD reaches **83.73** total score on VBench, 1.03 points above its matched DMD baseline (Table 1). On MiniMax-H3-33B joint video–audio generation, it reaches **83.17** total score on VideoGen-Eval, 0.41 above the strongest distilled baseline, and obtains the highest scores on all six audio metrics among 4-NFE models (Table 2). User studies prefer PDMD to every 4-NFE baseline on visual, motion, and audio quality (Table 2 and Table 1).

Our **contributions** are:

- We connect DMD's training instability and visual degradation to critic error in its score-difference update, and identify the student–critic endpoint residual as an estimate of the critic error.
- We introduce PDMD, a simple and effective algorithm that filters critic error by projecting out the component of the DMD update parallel to the student–critic endpoint residual. Under high-dimensional assumptions, PDMD removes a constant fraction of this error while losing only a vanishing fraction of useful signal.
- Experiments show improved energy distance and on-mode fraction in 2D, better visual quality on Wan2.1, and better video and audio quality on MiniMax-H3. User studies also favor PDMD.

2 Related Work

Distribution matching distillation. DMD approximately minimizes the reverse KL divergence between a few-step student and a diffusion teacher (Yin et al., 2024b). Its online critic is imperfect, and the resulting error can contribute to progressive oversaturation and reduced motion (Yin et al., 2024a; Xu et al., 2025; Zheng et al., 2026). Adversarial distillation, such as DMD2, adds discriminator supervision. DMD2 also uses additional critic updates (Yin et al., 2024a; Xu et al., 2024; Sauer et al., 2024b;a; Lin et al., 2024; 2025). These methods can improve fidelity or critic tracking, but they introduce discriminator training and coupled optimization, which can add memory and training costs and be difficult to stabilize (Ge et al., 2026a). SiD, f-distill, and SGMD instead replace the reverse-KL formulation with semi-implicit Fisher, discriminator-weighted f-divergence, and stop-gradient Fisher objectives, respectively. Each requires additional network evaluations (Zhou et al., 2024; Xu et al., 2025; Wu et al., 2026b). Others keep the DMD objective and add teacher-trajectory, real-data, relational, or reward supervision (Wu et al., 2026a; Chen et al., 2026; Zhang et al., 2026; Bai et al., 2026; Jiang et al., 2026). ADV also identifies oversaturation and temporal collapse in few-step video distillation and counters them with an adaptively weighted regression loss and a temporal regularizer on top of DMD (You et al., 2026). PDMD keeps the DMD score networks and critic training unchanged. It changes only the student update computed from their existing predictions.

Trajectory and hybrid objectives. Consistency models enforce agreement along probability-flow ODE trajectories (Song et al., 2023; Song & Dhariwal, 2024; Kim et al., 2024). rCM extends this approach to video by adding DMD to continuous-time consistency training (Zheng et al., 2026). MeanFlow learns average velocities (Geng et al., 2025), while AnyFlow combines MeanFlow-style flow-map training with on-policy DMD through Flow Map Backward Simulation (Gu et al., 2026). SC-DMD, TMD and FSF-DMD are further hybrids, adding step-size consistency, a MeanFlow-pretrained flow head, or a critic-free update from the flow-map generator's endpoint pseudo-velocity (Ge et al., 2026b; Nie et al., 2026; Kim et al., 2026). These methods need another objective, rollout procedure, or training stage. PDMD requires no additional model, loss, pass, data, or stage beyond DMD.

Error cancellation and directional decomposition. Adaptive noise cancellation and control variates use correlated references to reduce error or variance (Widrow et al., 1975; Glynn & Szechtman, 2002). DDS and FlowEdit likewise take differences between paired predictions for image editing (Hertz et al., 2023; Kulikov et al., 2025), while Perp-Neg and APG project guidance compo-

nents (Armandpour et al., 2023; Sadat et al., 2025). PDMD instead projects the DMD update using the student–critic endpoint residual to suppress critic estimation error.

3 METHODOLOGY

3.1 PRELIMINARIES: DISTRIBUTION MATCHING DISTILLATION

Distribution Matching Distillation (DMD) (Yin et al., 2024b) trains a few-step student to match the teacher distribution using the difference between an online critic score and a frozen teacher score.

Let G_θ denote the few-step student. Given Gaussian input $\boldsymbol{z}$ and condition $\boldsymbol{c}$, the student produces the clean endpoint $\boldsymbol{x}_0^s(\boldsymbol{z}, \boldsymbol{c}) = G_\theta(\boldsymbol{z}, \boldsymbol{c})$, a sample from the student's final distribution. An online critic field $\boldsymbol{s}_{\text{critic}}(\boldsymbol{x}_t, t, \boldsymbol{c})$ is trained to estimate the score at the renoised student endpoints. DMD renoises these samples using the same process and same coefficients α_t and σ_t used to train the teacher:

$$\boldsymbol{x}_t = \alpha_t \boldsymbol{x}_0^s(\boldsymbol{z}, \boldsymbol{c}) + \sigma_t \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}). \tag{1}$$

To update the student under the reverse-KL objective, DMD evaluates both the online critic field $\boldsymbol{s}_{\text{critic}}(\boldsymbol{x}_t, t, \boldsymbol{c})$ and the frozen teacher field $\boldsymbol{s}_{\text{teacher}}(\boldsymbol{x}_t, t, \boldsymbol{c})$, which estimates the score of the teacher distribution. Omitting scalar timestep weights, the DMD update signal is their difference:

$$\boldsymbol{d}(\boldsymbol{x}_t, t, \boldsymbol{c}) = \boldsymbol{s}_{\text{critic}}(\boldsymbol{x}_t, t, \boldsymbol{c}) - \boldsymbol{s}_{\text{teacher}}(\boldsymbol{x}_t, t, \boldsymbol{c}). \tag{2}$$

$\boldsymbol{d}$ is backpropagated as a gradient through $\boldsymbol{x}_0^s = G_\theta(\boldsymbol{z}, \boldsymbol{c})$ to update the student parameters θ. Optimizing the student and critic in alternation trains the few-step student to match the teacher.

3.2 PROJECTED DISTRIBUTION MATCHING DISTILLATION

DMD uses $-\boldsymbol{d}$ as an endpoint-space descent signal. However, because the critic is imperfect, this teacher–critic score difference contains the critic error. To suppress this contamination, we propose Projected Distribution Matching Distillation (PDMD), which projects out the component of the DMD update parallel to the student–critic endpoint residual, an estimate of the critic error.

Specifically, the critic endpoint is $\boldsymbol{x}_0^c = (\boldsymbol{x}_t + \sigma_t^2 \boldsymbol{s}_{\text{critic}})/\alpha_t$. We define $\boldsymbol{r} = \boldsymbol{x}_0^c - \boldsymbol{x}_0^s$ and let

$$\boldsymbol{P}_{\boldsymbol{r}} = \frac{\boldsymbol{r}\boldsymbol{r}^\top}{\|\boldsymbol{r}\|^2}, \qquad \boldsymbol{P}_{\boldsymbol{r}}^\perp = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{r}}. \tag{3}$$

The projected update is

$$\boldsymbol{d}_\perp = \boldsymbol{P}_{\boldsymbol{r}}^\perp \boldsymbol{d} = \boldsymbol{d} - \frac{\langle \boldsymbol{d}, \boldsymbol{r} \rangle}{\|\boldsymbol{r}\|^2} \boldsymbol{r}. \tag{4}$$

When $\boldsymbol{r} = \boldsymbol{0}$, we leave the update unchanged. PDMD uses $\boldsymbol{d}_\perp$ in place of $\boldsymbol{d}$. This one-line change requires no additional loss, network, model pass, data, or training stage.

Algorithm 1: PDMD student step. The highlighted line is the only change from DMD.

```
x_0_s = G(z, cond)
x_t, t = renoise(x_0_s)                    # Eq. (1)
s_t = T.score(x_t, t, cond)
s_c, x_0_c = C.score_end(x_t, t, cond)
d = s_c - s_t                              # Eq. (2)
r = x_0_c - x_0_s
d = d - dot(d, r) / dot(r, r) * r # Eq.  (4)
update(G, dmd_loss(x_0_s, d))
```

4 ANALYSIS

4.1 THE PROJECTION FILTERS THE CRITIC ERROR

Residual as unbiased critic-error estimate. We show that the residual $\boldsymbol{r}$ that PDMD projects along is, at a fixed query, an unbiased estimate of the critic error. Following the re-diffusion process in Equation (1), let Q collect the noisy query $\boldsymbol{x}_t$, its noise level t, and the conditioning $\boldsymbol{c}$, and let $(Q, \boldsymbol{S})$ be a random query–endpoint pair from critic training. Critic training re-diffuses the student's own endpoints, so the endpoint of the pair is the student endpoint, $\boldsymbol{S} = \boldsymbol{x}_0^s$. The critic endpoint is trained with a squared ℓ_2 denoising loss, whose population-risk minimizer at a fixed query $Q = q$ is the conditional mean (Supp. Lemma A1). The population-optimal critic endpoint and the conditional noise are therefore

$$\boldsymbol{c}^\star(q) = \mathbb{E}[\boldsymbol{S} \mid Q = q], \qquad \boldsymbol{\xi} = \boldsymbol{S} - \boldsymbol{c}^\star(q), \qquad \mathbb{E}[\boldsymbol{\xi} \mid Q = q] = \boldsymbol{0}. \tag{5}$$

The learned critic endpoint $\boldsymbol{x}_0^c$ is fixed at q, and its error is

$$\boldsymbol{e} = \boldsymbol{x}_0^c - \boldsymbol{c}^\star(q). \tag{6}$$

The residual at that query is $\boldsymbol{R} = \boldsymbol{x}_0^c - \boldsymbol{S} = \boldsymbol{e} - \boldsymbol{\xi}$. $\boldsymbol{R}$ is random through $\boldsymbol{S}$. The residual $\boldsymbol{r}$ of Equation (3) is one realization of $\boldsymbol{R}$, and $\boldsymbol{P_r}$ is the corresponding realization of the random projector $\boldsymbol{P_R}$. Taking the conditional mean of $\boldsymbol{R} = \boldsymbol{e} - \boldsymbol{\xi}$ and using $\mathbb{E}[\boldsymbol{\xi} \mid Q = q] = \mathbf{0}$ gives

$$\mathbb{E}[\boldsymbol{R} \mid Q = q] = \boldsymbol{e}, \tag{7}$$

so the residual is an unbiased estimate of the critic error. Unlike an arbitrary direction, $\boldsymbol{R}$ contains $\boldsymbol{e}$ by construction.

Error removal guarantee. We now quantify how much critic error the projection removes. Let $\boldsymbol{\Sigma}(q) = \mathrm{Cov}(\boldsymbol{S} \mid Q = q)$, so that $\mathbb{E}[\|\boldsymbol{\xi}\|^2 \mid Q = q] = \operatorname{tr} \boldsymbol{\Sigma}(q)$ and the second moments of the residual are

$$\mathbb{E}[\boldsymbol{e}^\top \boldsymbol{R} \mid Q = q] = \|\boldsymbol{e}\|^2, \qquad \mathbb{E}[\|\boldsymbol{R}\|^2 \mid Q = q] = \|\boldsymbol{e}\|^2 + \operatorname{tr} \boldsymbol{\Sigma}(q). \tag{8}$$

For $\boldsymbol{e} \neq \mathbf{0}$, define the fraction of critic-error energy removed along $\boldsymbol{R}$ as

$$\gamma_e(\boldsymbol{R}) = \frac{\|\boldsymbol{P_R e}\|^2}{\|\boldsymbol{e}\|^2} = \frac{(\boldsymbol{e}^\top \boldsymbol{R})^2}{\|\boldsymbol{e}\|^2 \|\boldsymbol{R}\|^2}. \tag{9}$$

We set $\gamma_e(\mathbf{0}) = 0$ when the residual vanishes. Applying Cauchy–Schwarz to $\mathbb{E}[(\boldsymbol{e}^\top \boldsymbol{R})^2 / \|\boldsymbol{R}\|^2 \mid Q = q]$ with the moments of Equation (8) yields

$$\mathbb{E}[\gamma_e(\boldsymbol{R}) \mid Q = q] \geq \frac{\|\boldsymbol{e}\|^2}{\|\boldsymbol{e}\|^2 + \operatorname{tr} \boldsymbol{\Sigma}(q)}. \tag{10}$$

Moreover, the Bayes risk $L^\star(q) = \mathbb{E}[\|\boldsymbol{S} - \boldsymbol{c}^\star(q)\|^2 \mid Q = q]$ equals $\operatorname{tr} \boldsymbol{\Sigma}(q)$, so the covariance and conditional-risk views are two forms of the same result (Supp. Section A.1). At a fixed noise level, converting endpoint predictions to scores only rescales the corresponding critic error, so γ_e is unchanged in the score space used by DMD.

High-dimensional interpretation. When the conditional noise $\boldsymbol{\xi}$ is concentrated in high dimensions and has no dominant alignment with the fixed error direction $\boldsymbol{e}$, the lower bound in Equation (10) describes a typical update:

$$\gamma_e \approx \frac{\|\boldsymbol{e}\|^2}{\|\boldsymbol{e}\|^2 + \operatorname{tr} \boldsymbol{\Sigma}(q)}. \tag{11}$$

For example, when $\operatorname{tr} \boldsymbol{\Sigma}(q) = \|\boldsymbol{e}\|^2$, the conditional guarantee removes at least 50% of the critic-error energy on average (Supp. Theorem A2 and Corollary A3). Under high-dimensional assumptions, the projection removes a constant fraction of the critic error but only a vanishing fraction of the ideal signal (Supp. Corollaries A5 and A6). On MiniMax-H3, the projection visibly attenuates speckle artifacts and retains more than 80% of the update norm for most of training (Supp. Sections D.2 and E.1).

Critic lag and collapse. DMD2 attributes instability to the critic lagging behind the moving student and adds critic updates (Yin et al., 2024a). In a two-mode model, a strong lag turns the critic's restoring feedback into instability (Supp. Lemma A8). For a critic whose error arises only from this lag, the error is still the conditional mean of $\boldsymbol{r}$, so the bound of Equation (10) applies to it and the projection attenuates the lag error in conditional mean square (Supp. Corollary A9).

4.2 2D Verification

Error filtering. We test the error-filtering performance on a ring of eight Gaussians with an exact teacher score, so the critic is the only learned score. We compare DMD with four variants that each remove one direction from its update (Figure 4; Supp. Section B.1). The other projections remove a similar share of the update norm without targeting the critic error. Only PDMD improves both distributional metrics over DMD. Supp. Figure B1 and Table B2 measure γ_e, the ideal-signal energy removal fraction γ_s, and the net error reduction directly along the run: the projection reduces mean squared error to the ideal update by 28% relative to DMD for $\sigma \in [0.15, 1.1]$.

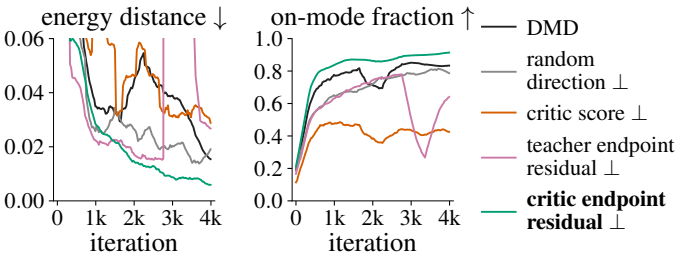

Figure 4: **2D comparison of projection directions** (20-point moving average). Left: energy distance to the target. Right: fraction of samples within 3σ of a mode center. Among the tested directions, only the PDMD projection stabilizes training and improves both metrics over DMD.

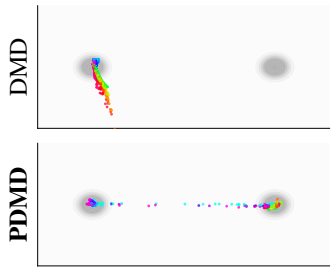

Figure 5: **Two-mode distillation with lagging critic**. This run illustrates DMD collapse (top) and PDMD coverage of both modes (bottom).

Table 1: **Quantitative comparisons for Wan2.1-T2V-1.3B at 4 NFE on VBench.** † marks our best reimplemented checkpoints; other baselines use official releases. All methods follow the AnyFlow protocol. PDMD leads in quality, dynamic degree, and total score. User-study win–loss shares show substantial preferences for PDMD over all compared methods in visual and motion quality.

		VBench				User study (win − loss, %)		
Method	NFE	**Total**	Quality	Dynamic	Semantic	Text	Visual	Motion
Wan2.1-T2V-1.3B (Team Wan et al., 2025)	50×2	83.06	85.02	68.61	75.23	−0.4	+17.6	+24.3
Wan2.1-T2V-1.3B (Team Wan et al., 2025)	4×2	68.54	73.60	18.61	48.30	–	–	–
rCM (Zheng et al., 2026)	4	83.13	85.14	78.06	75.10	−2.1	+33.8	+17.7
ADV (You et al., 2026)	4	81.47	83.61	63.89	72.91	+2.9	+34.6	+19.2
AnyFlow (Gu et al., 2026)	4	83.54	85.28	59.44	**76.57**	+0.6	+24.0	+28.9
DMD† (Yin et al., 2024b)	4	82.70	85.06	86.39	73.24	+6.5	+53.6	+35.2
DMD2† (Yin et al., 2024a)	4	83.44	85.84	76.67	73.84	+7.5	+39.5	+30.9
PDMD (ours)	4	**83.73**	**85.89**	**89.72**	75.09	–	–	–

Collapse under critic lag. With the critic learning rate at $1/20$ of the student's on a two-mode target, DMD collapses in $14/20$ and $19/20$ of the runs across two learning rates, a random projection in $16/20$ and $19/20$, and PDMD in $3/20$ and $12/20$ (Figure 5; Supp. Section B.2). PDMD typically recovers from an early single-mode state, whereas collapsed DMD runs remain locked.

5 Experiments

5.1 Experimental Setup

Wan2.1-T2V-1.3B. DMD2† is our reimplementation of DMD2 (Yin et al., 2024a) on Wan2.1-T2V-1.3B at 4 NFE. DMD† is our DMD reimplementation with one critic update per student update and no GAN loss. PDMD differs from DMD† only by the projection in Equation (4). All training hyperparameters are unchanged. We fine-tune all weights. For rCM (Zheng et al., 2026), ADV (You et al., 2026), and AnyFlow (Gu et al., 2026), we use their official checkpoints.

MiniMax-H3-33B. We distill the 33B joint video–audio DiT MiniMax-H3 (MiniMax, 2026a) to 4 NFE. Student and critic use rank-128 LoRA with scaling 128 on every block's attention projections and feed-forward layers. We tune learning rates for DMD†, DMD2†, rCM †, and AnyFlow †, reporting each method's best run; Supp. Table C2 details the two-stage rCM † and AnyFlow † recipes. We also evaluate the community's 4-step H3 Turbo LoRA (LarryVrh, 2026) (Supp. Section C.2).

Evaluation. Wan2.1 uses the 944 AnyFlow-augmented VBench prompts (Huang et al., 2023; Gu et al., 2026), at 480p with five seeds per prompt. MiniMax-H3 uses 387 VideoGen-Eval prompts (Yang et al., 2025), at 544p with mixed aspect ratios and a fixed seed. Each H3 prompt is rewritten once with Qwen3.8-27B (Qwen Team, 2026) following official guidance (MiniMax, 2026b), then shared across methods. Video evaluation follows VBench; audio uses six established metrics. In randomized



Figure 6: **Qualitative results on four-step Wan2.1-T2V-1.3B.** Each example shows early and late frames from the same VBench prompt and seed across methods, following the AnyFlow evaluation protocol (Gu et al., 2026). In these examples, PDMD produces natural textures and less oversaturation than the matched DMD and DMD2 models. See Supp. Figure E5 for more visual results.

Table 2: **Quantitative 4-NFE MiniMax-H3-33B comparisons on VideoGen-Eval.** † marks rows reporting the best reimplemented checkpoint. PDMD reaches the highest visual total and performs best across all audio metrics among the distilled models, and the user study prefers it to 4-NFE baselines on visual, motion, and audio quality. Text-alignment judgments are mostly ties (Supp. Table C4).

		Video				Audio						User study (win − loss, %)				
Method	NFE	**Total**	Quality	Dynamic	Semantic	PQ	CE	CU	IS	IB	DeSync ↓	**Overall**	Text	Visual	Motion	Audio
MiniMax-H3-33B (MiniMax, 2026a)	50	82.41	82.22	66.67	83.20	6.567	4.188	6.213	5.15	0.229	0.797	−37.0	−7.5	−35.4	−11.9	−3.9
MiniMax-H3-33B (MiniMax, 2026a)	4	79.48	79.54	44.44	79.23	6.056	3.167	5.580	3.35	0.129	0.932	+70.3	+16.5	+72.6	+44.7	+47.5
H3 Turbo LoRA (LarryVrh, 2026)	4	81.57	81.29	54.78	82.69	6.406	3.917	6.015	4.52	0.183	0.839	+34.6	+2.8	+38.0	+19.4	+9.3
rCM † (Zheng et al., 2026)	4	81.18	80.98	58.40	81.95	6.063	3.711	5.446	3.69	0.194	0.916	+43.7	+3.1	+51.2	+31.5	+17.1
AnyFlow † (Gu et al., 2026)	4	81.97	81.82	64.60	82.60	6.092	3.544	5.591	4.35	0.161	0.905	+59.2	−1.6	+66.4	+27.9	+25.8
DMD † (Yin et al., 2024b)	4	82.76	82.68	61.76	**83.05**	6.381	3.801	5.931	4.79	0.176	0.905	+29.7	+3.6	+25.1	+12.1	+10.9
DMD2 † (Yin et al., 2024a)	4	82.27	82.51	59.95	81.34	6.305	3.434	5.871	4.06	0.153	0.921	+35.1	+9.0	+28.7	+19.4	+22.2
PDMD (ours)	4	**83.17**	**83.25**	**71.83**	82.86	**6.530**	**4.062**	**6.180**	**4.98**	**0.195**	**0.802**	–	–	–	–	–

side-by-side comparisons on matched prompts, 20 annotators choose the better clip or a tie for text alignment, visual and motion quality, plus audio and overall quality for H3. Tables 1 and 2 report win-minus-loss shares for PDMD against each baseline (Supp. Section C).

5.2 Four-Step Text-to-Video Generation on Wan2.1-T2V-1.3B

In Table 1, PDMD scores 83.73 total, above 83.54 for AnyFlow (Gu et al., 2026), 83.06 for the 50-step teacher, and 82.70 for matched DMD †. DMD † peaks at 1,500 iterations, then degrades to 63.04 at 5,000 and 56.09 at 7,500; PDMD scores 83.44 at 5,000 and stays above 83.5 through 10,000. PDMD is also 0.29 points above DMD2† without its additional GAN loss and achieves a higher dynamic-degree score (89.72 versus 76.67). Figure 6 qualitatively shows PDMD produces higher-quality videos with natural textures and less oversaturation. Annotators prefer PDMD to every compared method, the 50-step teacher included, on visual and motion quality; text alignment is mostly ties (Table 1; Supp. Table C3).

5.3 Four-Step Joint Video–Audio Generation on MiniMax-H3-33B

Video. PDMD scores 83.17 total, above every 4-NFE baseline; DMD and AnyFlow reach 82.76 and 81.97, respectively. DMD † peaks at 5,500 iterations with a tenfold lower learning rate; at PDMD's learning rate, it oversaturates from iteration 500 and subsequently collapses. Figures 7 and 8 show more realistic textures, more dynamic motion, and sharper details with PDMD.

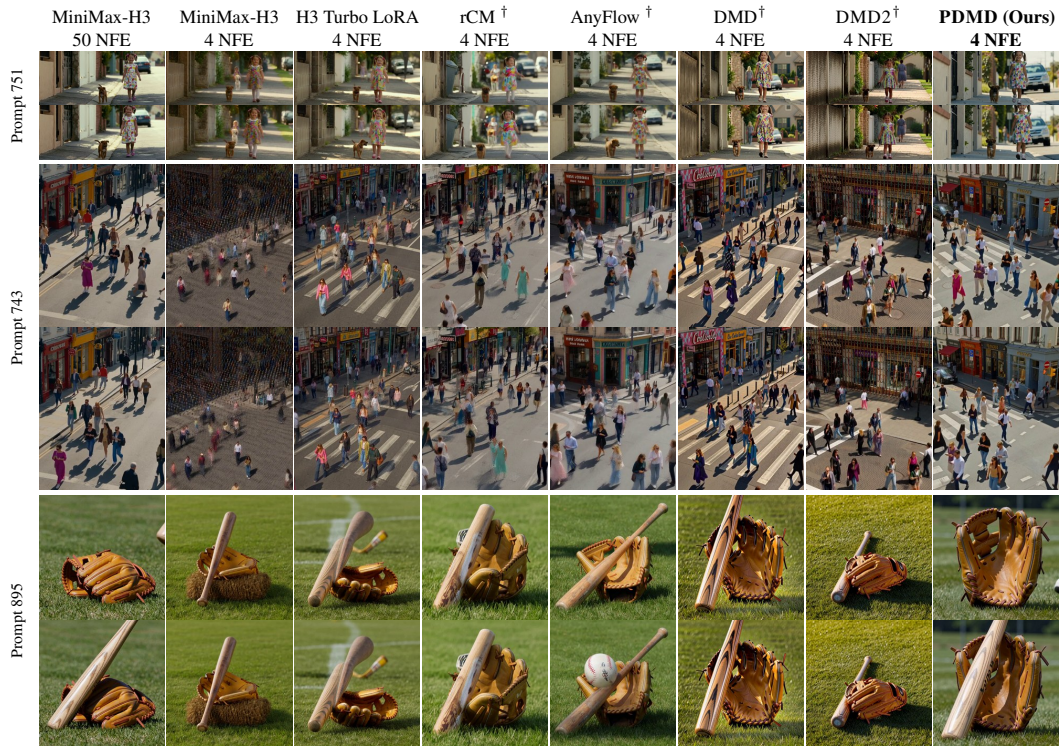

Figure 7: **Qualitative results for 4-NFE MiniMax-H3 on VideoGen-Eval.** PDMD shows fewer artifacts, more realistic textures, and better motion. Each prompt shows an earlier and a later frame of the same clip; Supp. Section E.4 shows more prompts.

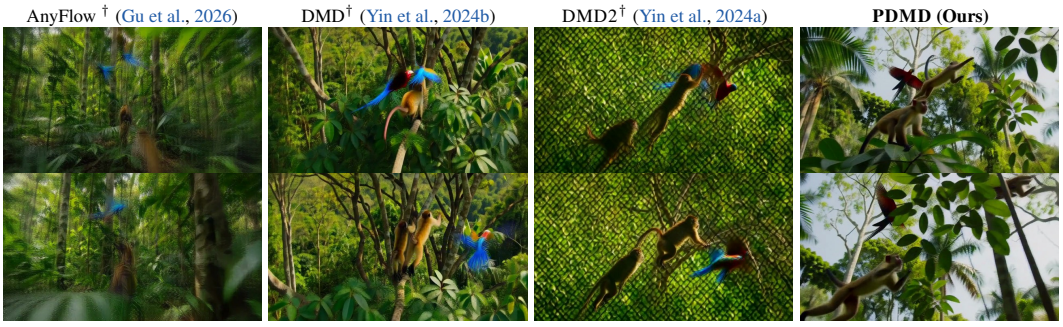

Figure 8: **Motion on 4-NFE MiniMax-H3** (VideoGen-Eval prompt 760). The compared baselines show translucent ghosting artifacts and severely blur the animals, while PDMD keeps them sharp.

Audio. PDMD leads the compared 4-NFE models on all six audio metrics (Table 2), coming within 0.04 of the 50-step teacher on production quality (PQ). Metric definitions are in Supp. Section C.2.

User study. Annotators prefer PDMD to every compared 4-NFE model in overall, visual, motion, and audio quality (Table 2). Text-alignment judgments are mostly ties (64–80%; Supp. Table C4). Unlike on Wan2.1, the 50-step H3 teacher remains preferred by 37.0 percentage points overall and 35.4 in visual quality. This larger residual gap may reflect the frontier teacher's higher quality ceiling, making four-step distillation more challenging than on Wan2.1.

6 Ablations

Stability compared with alternative projections. Six MiniMax-H3 variants differ only in the projection applied to the DMD update. At 1,500 iterations, only PDMD avoids saturation or high-frequency artifacts in Figure 9; Supp. Figure D1 follows three prompts through 3,000 iterations. Figure 10 compares training dynamics quantitatively on all 387 VideoGen-Eval prompts: from 500 to 3,000 iterations, PDMD maintains its total score while the other variants decline.

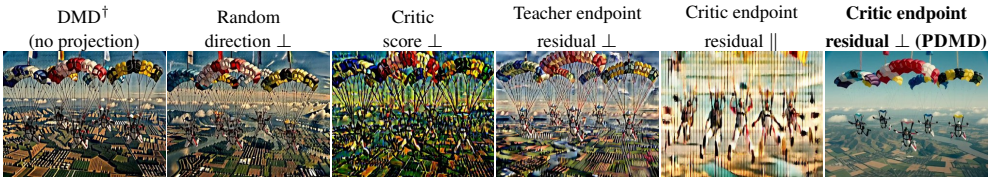

Figure 9: **Qualitative ablation of the projection direction on VideoGen-Eval** (prompt 882). Results at 1,500 training iterations. In this example, projecting out the update component parallel to the student–critic endpoint residual yields less saturation than the other variants.

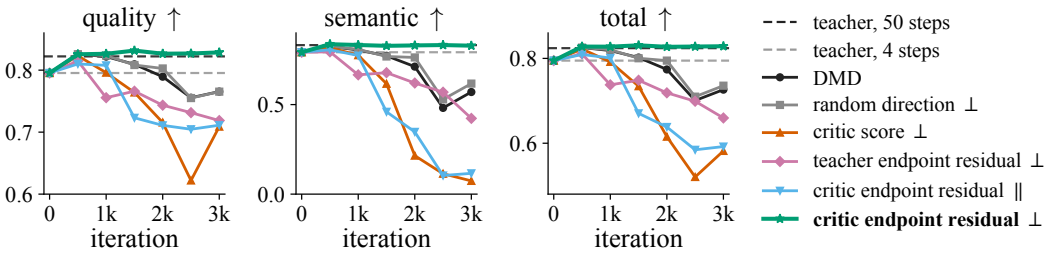

Figure 10: **Quantitative ablation study: training dynamics of MiniMax-H3 on VideoGen-Eval.** PDMD holds its total score; the others decline.

Direction matters more than magnitude. Random projection removes almost no update norm and behaves like DMD. PDMD removes a median 12% and remains stable over the evaluated interval. The residual-parallel variant keeps only 11% and degrades fastest, consistent with the parallel component carrying destabilizing error (Supp. Section D.2). Restoring half of the removed component also leads to degradation, but later than DMD (Supp. Section D.3). Raising the PDMD student learning rate by 1.25× preserves stability through 5,000 iterations, ending at 83.18 total (Supp. Section D.4). DMD at one tenth of the learning rate still oversaturates and peaks at 82.76 total at 5,500 iterations, below PDMD (Supp. Table E1). Together, these controls show that update shrinkage alone does not explain the stability gain. In addition, with one critic update per student update, PDMD scores 82.90 total, versus 82.37 for DMD at the same update ratio, and exceeds every 4-NFE baseline in Table 2 (see Supp. Table D1).

7 Limitations

Single-step quality remains limited (Figure 11) and may require an additional objective, such as a discriminator loss. See Section D.6 and Figure D6 for more 2-NFE results and Section F for further limitations in the supplement.



Figure 11: **Single-step limitation.**

8 Conclusion

We introduced PDMD, a one-line DMD modification that projects out the component of the student update parallel to the student–critic endpoint residual, an unbiased estimate of critic error at a fixed query. Under high-dimensional assumptions, the projection removes a constant fraction of critic error while discarding a vanishing fraction of useful signal. PDMD requires no additional objective, network, model pass, data, or training stage. Across Wan2.1 and MiniMax-H3, PDMD remains stable in settings where DMD degrades, improves video scores, is preferred to the compared 4-NFE baselines in user studies, and performs best on all audio metrics in Table 2. These results highlight PDMD as a promising approach to stable few-step video distillation. PDMD further opens up opportunities to combine critic-error filtering with existing distillation methods as a future work.

Ethics Statement

This work distills two released text-to-video models using existing training prompts and public evaluation benchmarks, without real-video data. The only newly collected data are human preference judgments for the user studies. A distilled student inherits the capabilities and biases of its teacher, and faster generation lowers the cost of producing misleading video. We have not evaluated whether distillation preserves the safety behavior of the base models.

Reproducibility Statement

We will release the training and evaluation code of the video and 2D experiments, the distilled Wan2.1-T2V-1.3B student weights, and the MiniMax-H3 distillation LoRA. Algorithm 1 and Equation (4) give the complete student update; the projection is the only change from DMD, so the method adds one line to an existing DMD implementation. Supp. Section C lists every training setting of the video experiments, including the trainable parameters, data, clip size, batch, optimizer, learning rates, critic update ratio and samplers (Supp. Table C1), and describes the baselines and the evaluation protocols of MiniMax-H3 (Supp. Section C.2) and Wan2.1-T2V-1.3B (Supp. Section C.1), with the prompt sets, seeds, samplers and metric definitions. Supp. Section B specifies the 2D experiments, and Supp. Section A gives the full proofs of the theoretical results.

AI use statement

Generative AI tools assisted with theoretical analysis, implementation, and language polishing. The authors reviewed all AI-assisted content, checked the derivations and proofs, conducted and validated the experiments, and take responsibility for the final content.

References

Mohammadreza Armandpour, Ali Sadeghian, Huangjie Zheng, Amir Sadeghian, and Mingyuan Zhou. Re-imagine the negative prompt algorithm: Transform 2d diffusion into 3d, alleviate janus problem and beyond. *arXiv preprint arXiv:2304.04968*, 2023.

Lichen Bai, Zikai Zhou, Shitong Shao, Wenliang Zhong, Shuo Yang, Shuo Chen, Bojun Chen, and Zeke Xie. Optimizing few-step generation with adaptive matching distillation. In *International Conference on Machine Learning*, 2026.

Tim Brooks, Bill Peebles, Connor Holmes, Will DePue, Yufei Guo, Li Jing, David Schnurr, Joe Taylor, Troy Luhman, Eric Luhman, Clarence Ng, Ricky Wang, and Aditya Ramesh. Video generation models as world simulators. OpenAI technical report, 2024. https://openai.com/index/video-generation-models-as-world-simulators/.

ByteDance Seed. Seedance 1.0: Exploring the boundaries of video generation models. *arXiv preprint arXiv:2506.09113*, 2025.

Siyi Chen, Shaowei Liu, Yixuan Jia, Zian Wang, Huan Ling, Qing Qu, and Jun Gao. Data-forcing distillation: Restoring diversity and fidelity in few-step video generation. *arXiv preprint arXiv:2606.18478*, 2026.

Ho Kei Cheng, Masato Ishii, Akio Hayakawa, Takashi Shibuya, Alexander Schwing, and Yuki Mitsufuji. MMAudio: Taming multimodal joint training for high-quality video-to-audio synthesis. In *CVPR*, 2025.

Xingtong Ge, Xin Zhang, Tongda Xu, Yi Zhang, Xinjie Zhang, Yan Wang, and Jun Zhang. SenseFlow: Scaling distribution matching for flow-based text-to-image distillation. In *International Conference on Learning Representations*, 2026a.

Xingtong Ge, Yi Zhang, Yushi Huang, Dailan He, Xiahong Wang, Bingqi Ma, Guanglu Song, Yu Liu, and Jun Zhang. Salt: Self-consistent distribution matching with cache-aware training for fast video generation. In *European Conference on Computer Vision*, 2026b.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, J. Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. In *Advances in Neural Information Processing Systems*, 2025.

Rohit Girdhar, Alaaeldin El-Nouby, Zhuang Liu, Mannat Singh, Kalyan Vasudev Alwala, Armand Joulin, and Ishan Misra. ImageBind: One embedding space to bind them all. In *CVPR*, 2023.

Peter W. Glynn and Roberto Szechtman. Some new perspectives on the method of control variates. In *Monte Carlo and Quasi-Monte Carlo Methods 2000*, pp. 27–49. Springer, 2002.

Yuchao Gu, Guian Fang, Yuxin Jiang, Weijia Mao, Song Han, Han Cai, and Mike Zheng Shou. Anyflow: Any-step video diffusion model with on-policy flow map distillation. In *European Conference on Computer Vision*, 2026.

N. D. Hayes. Roots of the transcendental equation associated with a certain difference-differential equation. *Journal of the London Mathematical Society*, s1-25(3):226–232, 1950.

Amir Hertz, Kfir Aberman, and Daniel Cohen-Or. Delta denoising score. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023.

Xun Huang, Zhengqi Li, Guande He, Mingyuan Zhou, and Eli Shechtman. Self forcing: Bridging the train-test gap in autoregressive video diffusion. In *Advances in Neural Information Processing Systems*, 2025.

Ziqi Huang, Yinan He, Jiashuo Yu, Fan Zhang, Chenyang Si, Yuming Jiang, Yuanhan Zhang, Tianxing Wu, Qingyang Jin, Nattapol Chanpaisit, Yaohui Wang, Xinyuan Chen, Limin Wang, Dahua Lin, Yu Qiao, and Ziwei Liu. VBench: Comprehensive benchmark suite for video generative models. *arXiv preprint arXiv:2311.17982*, 2023.

Vladimir Iashin, Weidi Xie, Esa Rahtu, and Andrew Zisserman. Synchformer: Efficient synchronization from sparse cues. In *ICASSP*, 2024.

Dengyang Jiang, Dongyang Liu, Zanyi Wang, Qilong Wu, Liuzhuozheng Li, Hengzhuang Li, Xin Jin, Changsheng Lu, Zhen Li, Mengmeng Wang, Steven Hoi, Peng Gao, and Harry Yang. Distribution matching distillation meets reinforcement learning. In *European Conference on Computer Vision*, 2026.

Dongjun Kim, Chieh-Hsin Lai, Wei-Hsiang Liao, Naoki Murata, Yuhta Takida, Toshimitsu Uesaka, Yutong He, Yuki Mitsufuji, and Stefano Ermon. Consistency trajectory models: Learning probability flow ODE trajectory of diffusion. In *International Conference on Learning Representations*, 2024.

Youngjoong Kim, Deokyeong Lee, and Jaesik Park. Distribution matching distillation without fake score network. *arXiv preprint arXiv:2605.19256*, 2026.

Khaled Koutini, Jan Schlüter, Hamid Eghbal-zadeh, and Gerhard Widmer. Efficient training of audio transformers with patchout. In *Interspeech*, 2022.

Vladimir Kulikov, Matan Kleiner, Inbar Huberman-Spiegelglas, and Tomer Michaeli. FlowEdit: Inversion-free text-based editing using pre-trained flow models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025.

LarryVrh. MiniMax-H3 Turbo LoRA: Few-step audio-video generation. https://huggingface.co/larryvrh/MiniMax-H3-Turbo-Lora, 2026. Model card.

Shanchuan Lin, Anran Wang, and Xiao Yang. SDXL-Lightning: Progressive adversarial diffusion distillation. *arXiv preprint arXiv:2402.13929*, 2024.

Shanchuan Lin, Xin Xia, Yuxi Ren, Ceyuan Yang, Xuefeng Xiao, and Lu Jiang. Diffusion adversarial post-training for one-step video generation. In *International Conference on Machine Learning*, 2025.

Yanzuo Lu, Xin Xia, Manlin Zhang, Huafeng Kuang, Jianbin Zheng, Yuxi Ren, and Xuefeng Xiao. Hyper-bagel: A unified acceleration framework for multimodal understanding and generation. *arXiv preprint arXiv:2509.18824*, 2025.

MiniMax. MiniMax H3. https://huggingface.co/MiniMaxAI/MiniMax-H3, 2026a. Model card.

MiniMax. Video prompt writing guide. https://huggingface.co/MiniMaxAI/MiniMax-H3/blob/main/docs/VIDEO_PROMPT_WRITING_GUIDE_base_en.md, 2026b.

Weili Nie, Julius Berner, Nanye Ma, Chao Liu, Saining Xie, and Arash Vahdat. Transition matching distillation for fast video generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *IEEE/CVF International Conference on Computer Vision*, 2023.

Qwen Team. Qwen3.8-27B. https://huggingface.co/Qwen/Qwen3.8-27B, 2026. Model card.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

Seyedmorteza Sadat, Otmar Hilliges, and Romann M. Weber. Eliminating oversaturation and artifacts of high guidance scales in diffusion models. In *International Conference on Learning Representations*, 2025.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. In *International Conference on Learning Representations*, 2022.

Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training GANs. In *NeurIPS*, 2016.

Axel Sauer, Frederic Boesel, Tim Dockhorn, Andreas Blattmann, Patrick Esser, and Robin Rombach. Fast high-resolution image synthesis with latent adversarial diffusion distillation. In *ACM SIGGRAPH Asia Conference Papers*, 2024a.

Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial diffusion distillation. In *European Conference on Computer Vision*, 2024b.

Shitong Shao, Hongwei Yi, Hanzhong Guo, Tian Ye, Daquan Zhou, Michael Lingelbach, Zhiqiang Xu, and Zeke Xie. Magicdistillation: Weak-to-strong video distillation for large-scale few-step synthesis. *arXiv preprint arXiv:2503.13319*, 2025.

Yang Song and Prafulla Dhariwal. Improved techniques for training consistency models. In *International Conference on Learning Representations*, 2024.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *Proceedings of the 40th International Conference on Machine Learning*, 2023.

Team Wan, Ang Wang, Baole Ai, Bin Wen, et al. Wan: Open and advanced large-scale video generative models. *arXiv preprint arXiv:2503.20314*, 2025.

Andros Tjandra, Yi-Chiao Wu, Baishan Guo, John Hoffman, Brian Ellis, Apoorv Vyas, Bowen Shi, Sanyuan Chen, Matt Le, Nick Zacharov, Carleigh Wood, Ann Lee, and Wei-Ning Hsu. Meta Audiobox aesthetics: Unified automatic quality assessment for speech, music, and sound. *arXiv preprint arXiv:2502.05139*, 2025.

Wenhao Wang and Yi Yang. VidProM: A million-scale real prompt-gallery dataset for text-to-video diffusion models. In *Advances in Neural Information Processing Systems*, 2024.

Bernard Widrow, John R. Glover, John M. McCool, John Kaunitz, Charles S. Williams, Robert H. Hearn, James R. Zeidler, Eugene Dong, and Robert C. Goodlin. Adaptive noise cancelling: Principles and applications. *Proceedings of the IEEE*, 63(12):1692–1716, 1975.

Tianhe Wu, Ruibin Li, Lei Zhang, and Kede Ma. Diversity-preserved distribution matching distillation for fast visual synthesis. *arXiv preprint arXiv:2602.03139*, 2026a.

Zhuguanyu Wu, Ruihao Gong, Yang Yong, Yushi Huang, Xiangyu Fan, Lei Yang, Dahua Lin, and Xianglong Liu. SGMD: Score gradient matching distillation for few-step video diffusion distillation. In *International Conference on Machine Learning*, 2026b.

Yanwu Xu, Yang Zhao, Zhisheng Xiao, and Tingbo Hou. UFOGen: You forward once large scale text-to-image generation via diffusion GANs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

Yilun Xu, Weili Nie, and Arash Vahdat. One-step diffusion models with f-divergence distribution matching. *arXiv preprint arXiv:2502.15681*, 2025.

Yuhang Yang, Ke Fan, Shangkun Sun, Hongxiang Li, Ailing Zeng, Feilin Han, Wei Zhai, Wei Liu, Yang Cao, and Zheng-Jun Zha. VideoGen-Eval: Agent-based system for video generation evaluation. *arXiv preprint arXiv:2503.23452*, 2025.

Tianwei Yin, Michaël Gharbi, Taesung Park, Richard Zhang, Eli Shechtman, Frédo Durand, and William T. Freeman. Improved distribution matching distillation for fast image synthesis. In *Advances in Neural Information Processing Systems*, 2024a.

Tianwei Yin, Michaël Gharbi, Richard Zhang, Eli Shechtman, Frédo Durand, William T. Freeman, and Taesung Park. One-step diffusion with distribution matching distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024b.

Yuyang You, Yongzhi Li, Jiahui Li, Yadong Mu, Quan Chen, and Peng Jiang. Adaptive video distillation: Mitigating oversaturation and temporal collapse in few-step generation. *arXiv preprint arXiv:2603.21864*, 2026.

Wenhu Zhang, Kun Cheng, Changyuan Wang, Shiyao Li, Yuechen Zhang, Wenbo Li, Jiajun Zha, Jingyi Zhang, Kang Zhao, and Jiaya Jia. CoDMD: Copula-aware distribution matching distillation for fast video generation. *arXiv preprint arXiv:2606.21982*, 2026.

Kaiwen Zheng, Yuji Wang, Qianli Ma, Huayu Chen, Jintao Zhang, Yogesh Balaji, Jianfei Chen, Ming-Yu Liu, Jun Zhu, and Qinsheng Zhang. Large scale diffusion distillation via score-regularized continuous-time consistency. In *International Conference on Learning Representations*, 2026.

Mingyuan Zhou, Huangjie Zheng, Zhendong Wang, Mingzhang Yin, and Hai Huang. Score identity distillation: Exponentially fast distillation of pretrained diffusion models for one-step generation. In *International Conference on Machine Learning*, 2024.

Matthias Zwicker, Wojciech Jarosz, Jaakko Lehtinen, Bochang Moon, Ravi Ramamoorthi, Fabrice Rousselle, Pradeep Sen, Cyril Soler, and Sung-Eui Yoon. Recent advances in adaptive sampling and reconstruction for Monte Carlo rendering. *Computer Graphics Forum*, 34(2):667–681, 2015. doi: 10.1111/cgf.12592.

APPENDIX

A Proofs **14**

A.1 ℓ_2-Optimal Critic Endpoint 14

A.2 High-Dimensional Error Removal and Signal Retention 15

A.3 Critic Lag and Allocation Stability 19

B 2D Experiments **21**

B.1 Experiment 1: Error Filtering on a Multi-Mode Target 21

B.2 Experiment 2: Two-Mode Allocation Under Critic Lag 22

C Experimental Details **24**

C.1 Wan2.1-T2V-1.3B 24

C.2 MiniMax-H3-33B 27

D Additional Controlled Experiments on MiniMax-H3 **29**

D.1 Projection Directions Over Training 29

D.2 Retained Update Norm 29

D.3 Restoring Half of the Removed Component 31

D.4 A Larger Student Step 31

D.5 One Critic Update per Student Update on MiniMax-H3 34

D.6 Two-Step Generation 34

E Additional Visual Comparisons **40**

E.1 The Removed Component in Video Space 40

E.2 Saturation Metrics 40

E.3 Motion Comparisons 40

E.4 Additional Samples 46

F Limitations and Future Work **50**

A PROOFS

A.1 ℓ_2-OPTIMAL CRITIC ENDPOINT

Let $(Q, \boldsymbol{S})$ denote the critic-training query–endpoint pair. We first establish the conditional-moment identities induced by the critic's ℓ_2 endpoint-training objective.

Lemma A1 (ℓ_2-optimal critic endpoint)**.** *Assume* $\mathbb{E}[\|\boldsymbol{S}\|^2 \mid Q = q] < \infty$. *The unique minimizer of the conditional squared-error risk*

$$\mathcal{L}_q(\boldsymbol{v}) := \mathbb{E}[\|\boldsymbol{v} - \boldsymbol{S}\|^2 \mid Q = q] \tag{A1}$$

is

$$\boldsymbol{c}^\star = \arg\min_{\boldsymbol{v}} \mathcal{L}_q(\boldsymbol{v}) = \mathbb{E}[\boldsymbol{S} \mid Q = q]. \tag{A2}$$

Consequently, for

$$\boldsymbol{\xi} = \boldsymbol{S} - \boldsymbol{c}^\star, \qquad \boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{S} \mid Q = q), \tag{A3}$$

we have

$$\mathbb{E}[\boldsymbol{\xi} \mid Q = q] = \mathbf{0}, \qquad \mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^\top \mid Q = q] = \boldsymbol{\Sigma}, \qquad \mathbb{E}[\|\boldsymbol{\xi}\|^2 \mid Q = q] = \operatorname{tr} \boldsymbol{\Sigma}. \tag{A4}$$

Proof. Let $\boldsymbol{m} = \mathbb{E}[\boldsymbol{S} \mid Q = q]$. For any deterministic prediction $\boldsymbol{v}$, write

$$\boldsymbol{v} - \boldsymbol{S} = (\boldsymbol{v} - \boldsymbol{m}) + (\boldsymbol{m} - \boldsymbol{S}). \tag{A5}$$

Expanding its squared norm and taking the conditional expectation gives

$$\begin{aligned} \mathcal{L}_q(\boldsymbol{v}) = \|\boldsymbol{v} - \boldsymbol{m}\|^2 + \mathbb{E}[\|\boldsymbol{S} - \boldsymbol{m}\|^2 \mid Q = q] \\ + 2(\boldsymbol{v} - \boldsymbol{m})^\top \mathbb{E}[\boldsymbol{m} - \boldsymbol{S} \mid Q = q]. \end{aligned} \tag{A6}$$

By the definition of $\boldsymbol{m}$,

$$\mathbb{E}[\boldsymbol{m} - \boldsymbol{S} \mid Q = q] = \boldsymbol{m} - \mathbb{E}[\boldsymbol{S} \mid Q = q] = \mathbf{0}. \tag{A7}$$

Hence

$$\mathcal{L}_q(\boldsymbol{v}) = \mathcal{L}_q(\boldsymbol{m}) + \|\boldsymbol{v} - \boldsymbol{m}\|^2. \tag{A8}$$

The second term is nonnegative and vanishes if and only if $\boldsymbol{v} = \boldsymbol{m}$, which proves both optimality and uniqueness in Equation (A2).

Substituting $\boldsymbol{c}^\star = \boldsymbol{m}$ into the definition of $\boldsymbol{\xi}$ gives

$$\mathbb{E}[\boldsymbol{\xi} \mid Q = q] = \mathbb{E}[\boldsymbol{S} \mid Q = q] - \boldsymbol{c}^\star = \mathbf{0}. \tag{A9}$$

Because $\boldsymbol{c}^\star$ is the conditional mean, the definition of conditional covariance becomes

$$\boldsymbol{\Sigma} = \mathbb{E}\big[(\boldsymbol{S} - \boldsymbol{c}^\star)(\boldsymbol{S} - \boldsymbol{c}^\star)^\top \mid Q = q\big] = \mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^\top \mid Q = q]. \tag{A10}$$

Finally, using $\|\boldsymbol{\xi}\|^2 = \operatorname{tr}(\boldsymbol{\xi}\boldsymbol{\xi}^\top)$ and linearity of conditional expectation and trace,

$$\mathbb{E}[\|\boldsymbol{\xi}\|^2 \mid Q = q] = \mathbb{E}[\operatorname{tr}(\boldsymbol{\xi}\boldsymbol{\xi}^\top) \mid Q = q] \tag{A11}$$

$$= \operatorname{tr} \mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^\top \mid Q = q] = \operatorname{tr} \boldsymbol{\Sigma}. \tag{A12}$$

This proves Equation (A4). □

A.2 High-Dimensional Error Removal and Signal Retention

We make precise the distinction between the structural alignment of critic error with the PDMD residual and the non-structural alignment of an ideal update with a rank-one direction. Throughout this subsection, we condition on a fixed query $Q = q$ and suppress this conditioning from the notation. We use the critic-training pair $(Q, \boldsymbol{S})$ and the corresponding quantities $\boldsymbol{c}^\star$, $\boldsymbol{\xi}$, and $\boldsymbol{\Sigma}$ defined in Lemma A1. The endpoint of the pair is the student endpoint, $\boldsymbol{S} = \boldsymbol{x}_0^s$, so the residual below is the $\boldsymbol{r}$ that PDMD projects along, read at a fixed query (Section 4.1).

For the learned critic endpoint $\boldsymbol{x}_0^c$, define

$$\boldsymbol{e} = \boldsymbol{x}_0^c - \boldsymbol{c}^\star, \qquad \boldsymbol{R} = \boldsymbol{x}_0^c - \boldsymbol{S} = \boldsymbol{e} - \boldsymbol{\xi}. \tag{A13}$$

Thus, unlike a generic direction, $\boldsymbol{R}$ contains the critic error $\boldsymbol{e}$ explicitly.

For $\boldsymbol{e} \neq \mathbf{0}$ and $\boldsymbol{R} \neq \mathbf{0}$, let $\boldsymbol{P}_{\boldsymbol{R}}$ denote the orthogonal projector onto $\boldsymbol{R}$ and define the critic-error removal ratio

$$\gamma_e(\boldsymbol{R}) := \frac{\|\boldsymbol{P}_{\boldsymbol{R}}\boldsymbol{e}\|^2}{\|\boldsymbol{e}\|^2} = \frac{(\boldsymbol{e}^\top \boldsymbol{R})^2}{\|\boldsymbol{e}\|^2\|\boldsymbol{R}\|^2}. \tag{A14}$$

Set $\gamma_e(\mathbf{0}) = 0$ when the residual vanishes.

We next characterize this ratio for a typical realization under explicit high-dimensional assumptions. Define the covariance effective dimension

$$d_{\text{eff}} := \frac{(\operatorname{tr} \boldsymbol{\Sigma})^2}{\operatorname{tr}(\boldsymbol{\Sigma}^2)}. \tag{A15}$$

Theorem A2 (High-dimensional critic-error removal)**.** *Consider a sequence of conditional problems for which* $d_{\text{eff}} \to \infty$*. Write*

$$a = \|\boldsymbol{e}\|^2, \qquad \tau = \operatorname{tr} \boldsymbol{\Sigma}. \tag{A16}$$

Assume $a, \tau > 0$ *and the concentration conditions*

$$\frac{\|\boldsymbol{\xi}\|^2 - \tau}{\tau} = O_{\mathbb{P}}(d_{\text{eff}}^{-1/2}), \qquad \frac{\boldsymbol{e}^{\mathsf{T}}\boldsymbol{\xi}}{\sqrt{a\tau}} = O_{\mathbb{P}}(d_{\text{eff}}^{-1/2}). \tag{A17}$$

Then

$$\gamma_e(\boldsymbol{R}) = \frac{a}{a+\tau} + O_{\mathbb{P}}\left(d_{\text{eff}}^{-1/2}\right). \tag{A18}$$

No balance assumption on a/τ *is required. If additionally* $a/\tau = \Omega(1)$*, then* $\gamma_e(\boldsymbol{R}) = \Theta_{\mathbb{P}}(1)$*.*

Proof. Let

$$v = \frac{\|\boldsymbol{\xi}\|^2 - \tau}{\tau}, \qquad b = \frac{\boldsymbol{e}^{\mathsf{T}}\boldsymbol{\xi}}{\sqrt{a\tau}}, \qquad \kappa = \frac{a}{\tau}. \tag{A19}$$

By assumption, $v, b = O_{\mathbb{P}}(d_{\text{eff}}^{-1/2})$. Substituting $\boldsymbol{R} = \boldsymbol{e} - \boldsymbol{\xi}$ into Equation (A14) gives

$$\gamma_e(\boldsymbol{R}) = \frac{(a - \boldsymbol{e}^{\mathsf{T}}\boldsymbol{\xi})^2}{a(a + \|\boldsymbol{\xi}\|^2 - 2\boldsymbol{e}^{\mathsf{T}}\boldsymbol{\xi})} = \frac{(\sqrt{\kappa} - b)^2}{\kappa + 1 + v - 2\sqrt{\kappa}b}. \tag{A20}$$

Subtracting $\kappa/(\kappa+1)$ exactly yields

$$\gamma_e(\boldsymbol{R}) - \frac{\kappa}{\kappa+1} = \frac{-2\sqrt{\kappa}b + (\kappa+1)b^2 - \kappa v}{(\kappa+1)(\kappa+1+v-2\sqrt{\kappa}b)}. \tag{A21}$$

The perturbation of the second denominator factor relative to $\kappa + 1$ satisfies

$$\frac{|v - 2\sqrt{\kappa}b|}{\kappa+1} \le |v| + |b| = o_{\mathbb{P}}(1), \tag{A22}$$

where $2\sqrt{\kappa}/(\kappa+1) \le 1$. Hence that factor is comparable to $\kappa + 1$ with probability tending to one. On this event, using $\sqrt{\kappa}/(\kappa+1) \le 1/2$ and $\kappa/(\kappa+1)^2 \le 1/4$, the absolute value of Equation (A21) is bounded by a constant multiple of

$$|b| + b^2 + |v| = O_{\mathbb{P}}(d_{\text{eff}}^{-1/2}), \tag{A23}$$

uniformly over every $\kappa > 0$. Since $\kappa/(\kappa+1) = a/(a+\tau)$, this proves Equation (A18). If $a/\tau = \Omega(1)$, the leading term is bounded away from zero, which gives the final statement. □

Corollary A3 (Gaussian sufficient conditions)**.** *Suppose* $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ *and, for a constant* $K < \infty$ *independent of dimension,*

$$\widehat{\boldsymbol{e}}^{\mathsf{T}}\boldsymbol{\Sigma}\widehat{\boldsymbol{e}} \le K\frac{\tau}{d_{\text{eff}}}, \qquad \widehat{\boldsymbol{e}} = \frac{\boldsymbol{e}}{\|\boldsymbol{e}\|}. \tag{A24}$$

Then the concentration conditions in Equation (A17) hold.

Proof. Diagonalizing $\boldsymbol{\Sigma}$ and using the moments of independent standard Gaussian coordinates gives

$$\operatorname{Var}(\|\boldsymbol{\xi}\|^2) = 2\operatorname{tr}(\boldsymbol{\Sigma}^2) = \frac{2\tau^2}{d_{\text{eff}}}. \tag{A25}$$

Chebyshev's inequality proves the first concentration condition. Moreover, $\boldsymbol{e}^{\mathsf{T}}\boldsymbol{\xi}$ is centered Gaussian and

$$\operatorname{Var}\left(\frac{\boldsymbol{e}^{\mathsf{T}}\boldsymbol{\xi}}{\sqrt{a\tau}}\right) = \frac{\widehat{\boldsymbol{e}}^{\mathsf{T}}\boldsymbol{\Sigma}\widehat{\boldsymbol{e}}}{\tau} \le \frac{K}{d_{\text{eff}}}, \tag{A26}$$

which proves the second condition. □

The preceding result is structural: $\boldsymbol{R} = \boldsymbol{e} - \boldsymbol{\xi}$ contains $\boldsymbol{e}$ explicitly. Small signal removal requires a different incoherence assumption on the fixed ideal signal direction (the ideal signal is weakly aligned with the residual).

Proposition A4 (Fixed-signal high-dimensional retention)**.** *At fixed q, let $\boldsymbol{s}^\star_{\mathrm{critic}}(q)$ denote the population-optimal critic score and define the ideal DMD update $\boldsymbol{d}^\star = \boldsymbol{s}^\star_{\mathrm{critic}}(q) - \boldsymbol{s}_{\mathrm{teacher}}(q)$. Assume $\boldsymbol{d}^\star \neq \mathbf{0}$, and let*

$$\boldsymbol{s} = \frac{\boldsymbol{d}^\star}{\|\boldsymbol{d}^\star\|}, \qquad \gamma_s(\boldsymbol{R}) := \frac{\|\boldsymbol{P}_{\boldsymbol{R}}\boldsymbol{d}^\star\|^2}{\|\boldsymbol{d}^\star\|^2} = \frac{(\boldsymbol{s}^\mathsf{T}\boldsymbol{R})^2}{\|\boldsymbol{R}\|^2}, \tag{A27}$$

with $\gamma_s(\mathbf{0}) = 0$. Suppose the concentration conditions in Equation (A17) hold and, for a constant $K_s < \infty$ independent of dimension,

$$(\boldsymbol{s}^\mathsf{T}\boldsymbol{e})^2 \le K_s \frac{a}{d_{\mathrm{eff}}}, \qquad \boldsymbol{s}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{s} \le K_s \frac{\tau}{d_{\mathrm{eff}}}. \tag{A28}$$

Then

$$\gamma_s(\boldsymbol{R}) = O_{\mathbb{P}}(d_{\mathrm{eff}}^{-1}), \qquad 1 - \gamma_s(\boldsymbol{R}) = 1 - O_{\mathbb{P}}(d_{\mathrm{eff}}^{-1}). \tag{A29}$$

Proof. Because $\mathbb{E}[\boldsymbol{\xi}] = \mathbf{0}$ and $\mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T}] = \boldsymbol{\Sigma}$ under the fixed-query conditioning,

$$\mathbb{E}[(\boldsymbol{s}^\mathsf{T}\boldsymbol{\xi})^2] = \boldsymbol{s}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{s} \le K_s \frac{\tau}{d_{\mathrm{eff}}}. \tag{A30}$$

Chebyshev's inequality and Equation (A28) therefore give

$$\boldsymbol{s}^\mathsf{T}\boldsymbol{e} = O\left(\sqrt{\frac{a}{d_{\mathrm{eff}}}}\right), \qquad \boldsymbol{s}^\mathsf{T}\boldsymbol{\xi} = O_{\mathbb{P}}\left(\sqrt{\frac{\tau}{d_{\mathrm{eff}}}}\right). \tag{A31}$$

Since $\boldsymbol{R} = \boldsymbol{e} - \boldsymbol{\xi}$, it follows that

$$(\boldsymbol{s}^\mathsf{T}\boldsymbol{R})^2 = O_{\mathbb{P}}\left(\frac{a+\tau}{d_{\mathrm{eff}}}\right). \tag{A32}$$

Writing v and b as in Equation (A19),

$$\frac{\|\boldsymbol{R}\|^2}{a+\tau} = 1 + \frac{\tau v - 2\sqrt{a\tau}b}{a+\tau} = 1 + o_{\mathbb{P}}(1), \tag{A33}$$

where we used $\tau/(a+\tau) \le 1$ and $2\sqrt{a\tau}/(a+\tau) \le 1$. Dividing Equation (A32) by Equation (A33) proves the result. □

Corollary A5 (High-dimensional error–signal separation)**.** *Under the conditions of Theorem A2 and Proposition A4, if $a/\tau = \Omega(1)$, then*

$$\boxed{\gamma_e = \Theta_{\mathbb{P}}(1), \qquad \gamma_s = O_{\mathbb{P}}(d_{\mathrm{eff}}^{-1})}. \tag{A34}$$

Proof. The conclusion follows immediately from Theorem A2 and Proposition A4 and $a/\tau = \Omega(1)$. □

We next translate the fractional separation into an error guarantee for the projected update. At a fixed noise level, let $\boldsymbol{e}'$ denote the score-space critic error, so that $\boldsymbol{d} = \boldsymbol{d}^\star + \boldsymbol{e}'$, and define $\widetilde{\boldsymbol{d}} = \boldsymbol{P}_{\boldsymbol{R}}^{\perp}\boldsymbol{d}$. The score–endpoint conversion makes $\boldsymbol{e}'$ a scalar multiple of $\boldsymbol{e}$, so its projected energy fraction is also γ_e. Orthogonality gives the exact identity

$$\|\widetilde{\boldsymbol{d}} - \boldsymbol{d}^\star\|^2 = \|\boldsymbol{e}'\|^2(1-\gamma_e) + \|\boldsymbol{d}^\star\|^2\gamma_s, \tag{A35}$$

and therefore

$$\|\boldsymbol{d} - \boldsymbol{d}^\star\|^2 - \|\widetilde{\boldsymbol{d}} - \boldsymbol{d}^\star\|^2 = \|\boldsymbol{e}'\|^2\gamma_e - \|\boldsymbol{d}^\star\|^2\gamma_s. \tag{A36}$$

Corollary A6 (Net improvement over the DMD update)**.** *Under the conditions of Corollary A5, assume $\boldsymbol{e}' \neq \mathbf{0}$ and*

$$\frac{\|\boldsymbol{d}^\star\|^2}{\|\boldsymbol{e}'\|^2} = o(d_{\mathrm{eff}}). \tag{A37}$$

Then

$$\frac{\|\boldsymbol{d} - \boldsymbol{d}^\star\|^2 - \|\widetilde{\boldsymbol{d}} - \boldsymbol{d}^\star\|^2}{\|\boldsymbol{e}'\|^2} = \gamma_e - \frac{\|\boldsymbol{d}^\star\|^2}{\|\boldsymbol{e}'\|^2}\gamma_s = \Theta_{\mathbb{P}}(1). \tag{A38}$$

In particular, $\|\widetilde{\boldsymbol{d}} - \boldsymbol{d}^\star\|^2 < \|\boldsymbol{d} - \boldsymbol{d}^\star\|^2$ with probability tending to one.

Proof. By Corollary A5 and Equation (A37),

$$\frac{\|\boldsymbol{d}^\star\|^2}{\|\boldsymbol{e}'\|^2}\gamma_s = o_{\mathbb{P}}(1), \tag{A39}$$

whereas γ_e is bounded away from zero with probability tending to one. Substitution into Equation (A36) proves both claims. □

Thus, under the stated assumptions, error alignment is structural, whereas signal alignment is asymptotically weak: PDMD removes a constant fraction of critic-error energy, while the fraction of ideal-signal energy removed along the same rank-one direction vanishes with effective dimension.

Evidence at scale. Although $\boldsymbol{d}^\star$ and $\boldsymbol{e}$ are not observable on MiniMax-H3, the removed component itself can be decoded and seen. The same update target, decoded with and without the projection, loses its speckle and keeps its content (Figure 3; Figure E1; Section E.1). Three further measurements point the same way. The kept component carries more than 80% of the update norm for most of training (Section D.2). Removing any of four other directions instead does not stabilize training (Section 6, Section D.2). Putting half of the removed component back brings the degradation back (Section D.3). The removed direction carries the destabilizing part of the update, and training with the projected update stays stable where DMD degrades (Figure 2).

Projecting versus subtracting an error estimate. Since $\mathbb{E}[\boldsymbol{R}] = \boldsymbol{e}$, subtracting an estimate of the error is an alternative to deleting a direction. Write $\boldsymbol{e}' = \kappa\boldsymbol{e}$ with $\kappa > 0$ and consider the family

$$\widetilde{\boldsymbol{d}}_\lambda = \boldsymbol{d} - \lambda\kappa\boldsymbol{R}, \qquad \lambda \geq 0, \tag{A40}$$

which is the DMD update at $\lambda = 0$ and subtracts the unbiased estimate at $\lambda = 1$. Since $\boldsymbol{d} - \lambda\kappa\boldsymbol{R} = \boldsymbol{d}^\star + (1-\lambda)\boldsymbol{e}' + \lambda\kappa\boldsymbol{\xi}$ and $\mathbb{E}[\boldsymbol{\xi}] = \boldsymbol{0}$,

$$\mathbb{E}\|\widetilde{\boldsymbol{d}}_\lambda - \boldsymbol{d}^\star\|^2 = \kappa^2\big[(1-\lambda)^2 a + \lambda^2\tau\big], \tag{A41}$$

minimized at $\lambda^\star = a/(a+\tau)$ with value $\kappa^2 a\tau/(a+\tau)$.

The unbiased choice $\lambda = 1$ has expected squared error $\kappa^2\tau$: it removes the error in mean but injects the full conditional noise. The optimal fixed coefficient $\lambda^\star$ achieves $\kappa^2 a\tau/(a+\tau)$ but depends on the unobservable a and τ. The projection needs neither; its realized error is given exactly by Equation (A35). Its coefficient $\langle \boldsymbol{d}, \boldsymbol{R}\rangle/\|\boldsymbol{R}\|^2$ equals $\kappa\lambda^\star(1 + o_{\mathbb{P}}(1))$ under the conditions of Corollary A6, so it asymptotically estimates the optimal fixed subtraction coefficient from the two available vectors. This convergence in probability does not by itself establish equality of the expected risks.

Three further properties favor the projection. The projection never reverses the DMD direction: $\langle \boldsymbol{P}_{\boldsymbol{R}}^{\perp}\boldsymbol{d}, \boldsymbol{d}\rangle = \|\boldsymbol{P}_{\boldsymbol{R}}^{\perp}\boldsymbol{d}\|^2 \geq 0$ for every realization, so the angle between $\boldsymbol{P}_{\boldsymbol{R}}^{\perp}\boldsymbol{d}$ and $\boldsymbol{d}$ never exceeds 90°, whereas $\langle \boldsymbol{d} - \lambda\kappa\boldsymbol{R}, \boldsymbol{d}\rangle = \|\boldsymbol{d}\|^2 - \lambda\kappa\langle \boldsymbol{R}, \boldsymbol{d}\rangle$ turns negative once the subtracted term is large and aligned with $\boldsymbol{d}$. The projection never lengthens the update, $\|\boldsymbol{P}_{\boldsymbol{R}}^{\perp}\boldsymbol{d}\| \leq \|\boldsymbol{d}\|$, whereas $\boldsymbol{d} - \lambda\kappa\boldsymbol{R}$ can be longer than $\boldsymbol{d}$. The projection is invariant to any nonzero rescaling of $\boldsymbol{R}$, whereas subtraction needs κ at every noise level, and an overestimate flips the remaining error. Section D.3 evaluates this family at half the projection's own coefficient, and that run degrades as DMD does.

Biased critic-error estimates with smaller deviation. In graphics, a biased estimator with smaller variance often replaces an unbiased one to lower the mean squared error (Zwicker et al., 2015). A biased critic-error estimate with smaller mean-squared deviation offers a related tradeoff in our analysis. The proofs use $\boldsymbol{\xi}$ only through $\|\boldsymbol{\xi}\|^2$ and $\boldsymbol{e}^{\mathsf{T}}\boldsymbol{\xi}$. Let the subtracted endpoint deviate from $\boldsymbol{c}^\star$ by $\boldsymbol{\xi}' = \boldsymbol{m} + \boldsymbol{\zeta}$ with a fixed $\boldsymbol{m}$ and a centered $\boldsymbol{\zeta}$, so $\boldsymbol{R}' = \boldsymbol{e} - \boldsymbol{\xi}'$ has bias $-\boldsymbol{m}$, and write $\tau' = \|\boldsymbol{m}\|^2 + \operatorname{tr}\operatorname{Cov}(\boldsymbol{\zeta})$. If $\boldsymbol{\xi}'$ satisfies Equation (A17) with τ' in place of τ, Theorem A2 gives $\gamma_e = a/(a+\tau') + O_{\mathbb{P}}(d_{\text{eff}}^{-1/2})$: a bias costs the same as an equal amount of variance, and a deviation with smaller mean square removes more error. The fixed $\boldsymbol{m}$ may align with the signal, but it adds at most $\|\boldsymbol{m}\|^2/(a+\tau')$ to γ_s up to a constant: the smaller the bias, the less its direction matters.

A.3 Critic Lag and Allocation Stability

Distribution matching leaves the allocation of mass between separated modes weakly constrained, and the online critic sees that allocation only as it was some updates earlier. This section models the resulting feedback loop and shows that the student–critic endpoint residual observes the delayed error itself. Lag is one component of the critic error $\boldsymbol{e}$ of Equation (6); Equation (10) and Section A.2 bound the removal of the total error, and this section models the lag component alone.

We keep the notation of Equation (5): $(Q, \boldsymbol{S})$ is the critic-training query–endpoint pair, $\boldsymbol{c}^\star(q) = \mathbb{E}[\boldsymbol{S} \mid Q = q]$, $\boldsymbol{\xi} = \boldsymbol{S} - \boldsymbol{c}^\star(q)$, and $\boldsymbol{\Sigma}(q) = \mathrm{Cov}(\boldsymbol{S} \mid Q = q)$. Critic training re-diffuses the student's own endpoints (Equation (1)), so the critic-training endpoint is the student endpoint, $\boldsymbol{S} = \boldsymbol{x}_0^s$. The residual that PDMD projects along, $\boldsymbol{r} = \boldsymbol{x}_0^c - \boldsymbol{x}_0^s$ of Equation (3), is therefore the conditional endpoint residual $\boldsymbol{R} = \boldsymbol{e} - \boldsymbol{\xi}$ of Section 4.1, and $\boldsymbol{P}_{\boldsymbol{r}} = \boldsymbol{P}_{\boldsymbol{R}}$.

Allocation slice. At a fixed external condition, let the target have two separated mode regions with masses $\pi^\star$ and $1 - \pi^\star$ for $\pi^\star \in (0, 1)$, and otherwise arbitrary within-mode distributions. Let $\mathcal{M}(\boldsymbol{x}_0) \in \{+, -\}$ denote the mode assignment and let p_+ and p_- be the student's mode-conditioned endpoint distributions. We isolate the local slice in which these distributions are fixed while their allocation varies,

$$p_\theta(\boldsymbol{x}_0) = \pi(\theta) p_+(\boldsymbol{x}_0) + (1 - \pi(\theta)) p_-(\boldsymbol{x}_0), \tag{A42}$$

where the student induces the latent basins $A_\pm(\theta) = \{\boldsymbol{z} : \mathcal{M}(G_\theta(\boldsymbol{z})) = \pm\}$ and hence the allocation $\pi(\theta) = \mathbb{P}_{\boldsymbol{z}}[\boldsymbol{z} \in A_+(\theta)]$. Write u for training time, $\pi(u)$ for the allocation of the student at time u, and

$$\Delta(u) = \pi(u) - \pi^\star \tag{A43}$$

for its deviation from the matched allocation.

Fix the noise level t and the external condition. The query is then the noisy point itself, $Q_t = \alpha_t \boldsymbol{S} + \sigma_t \boldsymbol{\epsilon}$ of Equation (1), so conditioning on $Q = q$ and on $Q_t = q$ coincide and we use whichever is clearer. With $M = \mathcal{M}(\boldsymbol{S})$, define the mode posterior, the mode-conditioned posterior means, and their contrast,

$$\begin{gathered} w_\pi(q) = \mathbb{P}(M = + \mid Q_t = q), \qquad \boldsymbol{m}_\pm(q) = \mathbb{E}[\boldsymbol{S} \mid Q_t = q, M = \pm], \\ \Delta\boldsymbol{m}(q) = \boldsymbol{m}_+(q) - \boldsymbol{m}_-(q). \end{gathered} \tag{A44}$$

Let $\boldsymbol{c}_\pi^\star(q) = \mathbb{E}[\boldsymbol{S} \mid Q_t = q]$ denote the population-optimal critic endpoint when the student allocation is π, so that $\boldsymbol{c}^\star(q) = \boldsymbol{c}_{\pi(u)}^\star(q)$ at training time u.

Lemma A7 (Allocation moves the critic target along the mode contrast)**.** *For every allocation π, $\boldsymbol{c}_\pi^\star(q) = \boldsymbol{m}_-(q) + w_\pi(q)\Delta\boldsymbol{m}(q)$, and for any two allocations π and π',*

$$\boldsymbol{c}_{\pi'}^\star(q) - \boldsymbol{c}_\pi^\star(q) = \big(w_{\pi'}(q) - w_\pi(q)\big)\Delta\boldsymbol{m}(q). \tag{A45}$$

Proof. Conditioning on M gives $\boldsymbol{c}_\pi^\star(q) = w_\pi(q)\boldsymbol{m}_+(q) + (1 - w_\pi(q))\boldsymbol{m}_-(q)$, which is the first identity. The mode-conditioned distributions are held fixed in Equation (A42), so $\boldsymbol{m}_\pm(q)$ does not depend on π and only $w_\pi(q)$ moves with the allocation; subtracting the first identity at π' and at π proves Equation (A45). □

Delayed allocation feedback. Near the matched allocation we treat Δ as a scalar channel. Against a critic that is population-optimal for the current student, we model the population update as restoring the match at a rate $\lambda > 0$, $\dot\Delta(u) = -\lambda\Delta(u)$. The online critic is instead trained against the student it saw earlier: writing $\tau > 0$ for that delay and keeping the same channel, the critic responds to $\pi(u - \tau)$ rather than $\pi(u)$, and the allocation obeys the delay equation

$$\dot\Delta(u) = -\lambda\Delta(u - \tau). \tag{A46}$$

Lemma A8 (Critical delay)**.** *Let $\lambda, \tau > 0$. Every root of the characteristic equation*

$$z + \lambda e^{-z\tau} = 0 \tag{A47}$$

has negative real part if and only if $\lambda\tau < \pi/2$; at $\lambda\tau = \pi/2$ the pair $z = \pm i\lambda$ lies on the imaginary axis. The zero solution of Equation (A46) is therefore asymptotically stable for $\lambda\tau < \pi/2$ and unstable for $\lambda\tau > \pi/2$.

Proof. Substituting $\Delta(u) = e^{zu}$ in Equation (A46) gives $ze^{zu} = -\lambda e^{z(u-\tau)}$, which is Equation (A47). Let $z = x + iy$ be a root with $x \geq 0$; roots occur in conjugate pairs, so we may take $y \geq 0$. From $z = -\lambda e^{-z\tau}$,

$$|z| = \lambda e^{-x\tau} \leq \lambda, \qquad x = -\lambda e^{-x\tau}\cos(y\tau), \qquad y = \lambda e^{-x\tau}\sin(y\tau). \tag{A48}$$

Since $\lambda e^{-x\tau} > 0$, the middle identity and $x \geq 0$ force $\cos(y\tau) \leq 0$, and the smallest $y\tau \geq 0$ with $\cos(y\tau) \leq 0$ is $\pi/2$, so $y\tau \geq \pi/2$. The first identity gives $y \leq |z| \leq \lambda$ and hence $y\tau \leq \lambda\tau$. A root with nonnegative real part therefore requires $\lambda\tau \geq \pi/2$, so $\lambda\tau < \pi/2$ leaves every root in the open left half plane. At $\lambda\tau = \pi/2$, substituting $z = i\lambda$ in Equation (A47) gives $i\lambda + \lambda e^{-i\pi/2} = i\lambda - i\lambda = 0$, so $\pm i\lambda$ are roots; for $\lambda\tau > \pi/2$ this pair has crossed into the right half plane (Hayes, 1950). □

The residual observes the lag. Model a lagging critic as one whose only error is staleness: at training time u its endpoint is population-optimal for the student of time $u - \tau$, $\boldsymbol{x}_0^c = \boldsymbol{c}^\star_{\pi(u-\tau)}(q)$. Its endpoint-space error Equation (6) is then the lag error, and Lemma A7 places it on the mode contrast,

$$\boldsymbol{e} = \boldsymbol{e}^{x_0}_{\mathrm{lag}}(q) := \boldsymbol{c}^\star_{\pi(u-\tau)}(q) - \boldsymbol{c}^\star_{\pi(u)}(q) = \big(w_{\pi(u-\tau)}(q) - w_{\pi(u)}(q)\big)\Delta\boldsymbol{m}(q). \tag{A49}$$

At a fixed noise level, the mixture-level Tweedie identity $\boldsymbol{s}_\pi(q) = (\alpha_t \boldsymbol{c}^\star_\pi(q) - q)/\sigma_t^2$, where $\boldsymbol{s}_\pi = \nabla_q \log p_{\pi,t}$ is the score of the density $p_{\pi,t}$ of Q_t at allocation π, converts this into the score-space lag error

$$\boldsymbol{e}^s_{\mathrm{lag}}(q) = \boldsymbol{s}_{\pi(u-\tau)}(q) - \boldsymbol{s}_{\pi(u)}(q) = \frac{\alpha_t}{\sigma_t^2}\boldsymbol{e}^{x_0}_{\mathrm{lag}}(q), \tag{A50}$$

which is the error this critic contributes to the DMD update Equation (2). In the notation of Section A.2, the pure-lag model is the case $\boldsymbol{e}' = \boldsymbol{e}^s_{\mathrm{lag}}(q)$ of $\boldsymbol{d} = \boldsymbol{d}^\star + \boldsymbol{e}'$. Because the critic-training endpoint is the student's own endpoint, the residual carries this error explicitly,

$$\boldsymbol{r} = \boldsymbol{x}_0^c - \boldsymbol{x}_0^s = \boldsymbol{e}^{x_0}_{\mathrm{lag}}(q) - \boldsymbol{\xi}, \qquad \mathbb{E}[\boldsymbol{r} \mid Q = q] = \boldsymbol{e}^{x_0}_{\mathrm{lag}}(q), \tag{A51}$$

where the second identity uses $\mathbb{E}[\boldsymbol{\xi} \mid Q = q] = \mathbf{0}$ from Lemma A1. The conditional mean of the observable residual is the stale-score error itself, up to the scalar σ_t^2/α_t.

Corollary A9 (Conditional attenuation of the lag error)**.** *Let* $\boldsymbol{e}^{x_0}_{\mathrm{lag}}(q) \neq \mathbf{0}$. *Then*

$$\mathbb{E}\left[\frac{\|\boldsymbol{P}_{\boldsymbol{r}}\boldsymbol{e}^s_{\mathrm{lag}}\|^2}{\|\boldsymbol{e}^s_{\mathrm{lag}}\|^2}\,\middle|\, Q = q\right] \geq \frac{\|\boldsymbol{e}^{x_0}_{\mathrm{lag}}(q)\|^2}{\|\boldsymbol{e}^{x_0}_{\mathrm{lag}}(q)\|^2 + \operatorname{tr}\boldsymbol{\Sigma}(q)}, \tag{A52}$$

and the part of the lag error that survives the projection Equation (4) obeys

$$\mathbb{E}\big[\|\boldsymbol{P}_{\boldsymbol{r}}^\perp \boldsymbol{e}^s_{\mathrm{lag}}\|^2 \mid Q = q\big] \leq \frac{\operatorname{tr}\boldsymbol{\Sigma}(q)}{\|\boldsymbol{e}^{x_0}_{\mathrm{lag}}(q)\|^2 + \operatorname{tr}\boldsymbol{\Sigma}(q)}\,\|\boldsymbol{e}^s_{\mathrm{lag}}(q)\|^2 < \|\boldsymbol{e}^s_{\mathrm{lag}}(q)\|^2. \tag{A53}$$

Proof. By Equation (A50) the two lag errors differ by the positive scalar α_t/σ_t^2, which cancels in the ratio, so the left-hand side of Equation (A52) is $\mathbb{E}[\gamma_e(\boldsymbol{r}) \mid Q = q]$ of Equation (9) with $\boldsymbol{e} = \boldsymbol{e}^{x_0}_{\mathrm{lag}}(q)$. By Equation (A51), $\boldsymbol{r} = \boldsymbol{e} - \boldsymbol{\xi}$ with $\mathbb{E}[\boldsymbol{\xi} \mid Q = q] = \mathbf{0}$ and $\mathbb{E}[\|\boldsymbol{\xi}\|^2 \mid Q = q] = \operatorname{tr}\boldsymbol{\Sigma}(q)$ (Lemma A1), which are exactly the conditional moments Equation (8) behind Equation (10); that bound is Equation (A52). Since $\boldsymbol{P}_{\boldsymbol{r}}$ and $\boldsymbol{P}_{\boldsymbol{r}}^\perp$ are complementary orthogonal projectors, $\|\boldsymbol{P}_{\boldsymbol{r}}^\perp \boldsymbol{e}^s_{\mathrm{lag}}\|^2 = \|\boldsymbol{e}^s_{\mathrm{lag}}\|^2 - \|\boldsymbol{P}_{\boldsymbol{r}}\boldsymbol{e}^s_{\mathrm{lag}}\|^2$, and taking conditional expectations turns Equation (A52) into Equation (A53); the last inequality is strict because $\|\boldsymbol{e}^{x_0}_{\mathrm{lag}}(q)\| > 0$. □

The delayed feedback of Equation (A46) reaches the student through $\boldsymbol{e}^s_{\mathrm{lag}}$. DMD applies it in full. By Corollary A9, PDMD passes on, in conditional mean square, at most the fraction

$$\rho(q) = \frac{\operatorname{tr}\boldsymbol{\Sigma}(q)}{\|\boldsymbol{e}^{x_0}_{\mathrm{lag}}(q)\|^2 + \operatorname{tr}\boldsymbol{\Sigma}(q)} < 1. \tag{A54}$$

Treating this fraction as a constant gain on the scalar channel gives a one-parameter model of the projected feedback, $\dot{\Delta}(u) = -\rho\lambda\Delta(u - \tau)$, in which the stability condition of Lemma A8 becomes $\rho\lambda\tau < \pi/2$. The model is a heuristic: ρ bounds an energy fraction in expectation, the constant-gain step stands in for a quantity that varies with the query and with the critic, and the projection also acts on the ideal update, so ρ scales the lag error rather than the whole channel. The two-mode

results in Table B3 are consistent with this qualitative interpretation. The gain ρ tracks the state of the critic. When the critic is stale, the lag error dominates the conditional spread of the student's endpoints, $\rho \to 0$, and the delayed feedback is suppressed. When the critic has caught up, $\boldsymbol{e}_{\text{lag}} = \mathbf{0}$, the residual is pure conditional noise, and the update passes through up to the signal-removal term of Equation (A35). The projection damps the allocation feedback when the critic is stale and releases it when the critic is current. A fixed reduction of the step, such as the random direction in Table B3, has no such dependence.

B 2D EXPERIMENTS

This section tests two theoretical predictions in a minimal, fully observable setting. Experiment 1 measures whether the projection preferentially removes critic error while retaining the distribution-matching signal (Equation (10) and Corollary A5). Experiment 2 measures mode allocation under critic lag and tests the attenuation of Corollary A9.

Shared protocol. Both experiments use a variance-exploding (VE) diffusion toy in $\mathbb{R}^2$ in which the critic error and ideal update can be estimated from student samples. The teacher score is *analytic* (the exact posterior-mean denoiser of the Gaussian-mixture target), so the online critic is the only learned score in the system. Experiment 1 measures its deviation from the population-optimal student score. The student is a one-step generator $\boldsymbol{x}_0^s = G_\theta(\boldsymbol{z})$, $\boldsymbol{z} \sim \sigma_{\max}\mathcal{N}(\mathbf{0}, \boldsymbol{I})$, and the critic is an EDM-parameterized denoiser trained by denoising score matching on the student's own re-diffused samples (Equation (1) with $\alpha_t = 1$); both are 3-layer MLPs of width 128. Training uses Adam (β_1=0, β_2=0.999), a learning rate of 2×10^{-3} for both networks unless stated otherwise, a batch size of 1,024, 4,000 iterations in Experiment 1 and 6,000 iterations in Experiment 2, and noise levels $\sigma \sim \text{LogUniform}[0.02, 5]$ with the standard DMD weighting. PDMD differs from the DMD variant *only* by the per-sample projection of Equation (4); seeds, data, and schedules are shared. We report the *energy distance* between 2,048 generated and target samples (a rotation-invariant two-sample distance, zero iff the distributions match); *modes covered*, the number of mixture components whose share of on-mode samples exceeds 25% of its target share; the *on-mode fraction*, the share of samples within $3\sigma_{\text{mode}}$ of a component center (a precision proxy); and the *removed fraction*, the relative reduction in update norm, $1 - \|\boldsymbol{d}_{\text{kept}}\|/\|\boldsymbol{d}\|$.

B.1 EXPERIMENT 1: ERROR FILTERING ON A MULTI-MODE TARGET

Setup. The target distribution is a ring of eight Gaussians (radius 2, standard deviation $\sigma_{\text{mode}} = 0.12$). We train five variants. Four remove one direction from the DMD update $\boldsymbol{d}$, and the unprojected baseline provides a reference: (a) **DMD**: no direction is removed; (b) **random direction** $\perp$: remove a fresh random direction for each sample. In 2D, this projection removes $1 - 2/\pi \approx 36\%$ of the update norm on average, close to the share PDMD removes. Variant (b) therefore tests whether reducing the update norm alone explains the gain; (c) **critic score** $\perp$: remove the direction of $\boldsymbol{s}_{\text{critic}}(\boldsymbol{x}_t)$. The critic-score direction is the closest alternative to the PDMD direction; the two differ only by the query displacement $\sigma_t\boldsymbol{\epsilon}$; (d) **student–teacher endpoint residual** $\perp$: remove $\boldsymbol{r}_{\text{teacher}} = \boldsymbol{x}_0^{\text{teacher}} - \boldsymbol{x}_0^s$. The teacher endpoint $\boldsymbol{x}_0^{\text{teacher}}$ comes from $\boldsymbol{s}_{\text{teacher}}$ through the standard time-dependent rescaling and costs no extra forward pass. The teacher is exact, so $\boldsymbol{r}_{\text{teacher}}$ contains no critic error; (e) **student–critic endpoint residual** $\perp$: the PDMD update, which removes $\boldsymbol{r} = \boldsymbol{x}_0^c - \boldsymbol{x}_0^s$. The error-removal analysis predicts an ordering based on the information carried by each direction. The student–critic endpoint residual has the structural form $\boldsymbol{R} = \boldsymbol{e} - \boldsymbol{\xi}$ targeted by the removal bound. The critic-score direction also contains critic error, but mixes it with the full marginal-score geometry rather than isolating $\boldsymbol{R}$; variants (b) and (d) do not contain critic error, and variant (c) can also remove useful signal. All variants share the seed, data, and schedule. The figures show one run per variant with 200 recorded snapshots.

Distribution metrics. Figure 4 in the main text shows the training curves. Figure B2 shows the generated samples of every variant at iterations $0, 1{,}000, \ldots, 4{,}000$. We average each metric over the last quarter of training (iterations 3,000 to 4,000). PDMD reaches energy distance 0.0080 and on-mode fraction 0.903. DMD reaches 0.0273 and 0.838. The random direction reaches 0.0180 and 0.795. The critic-score projection reaches 0.0332 and 0.423. The teacher-residual projection reaches 2.04 and 0.413; the large error spike near iteration 3,000, visible in both figures, dominates

the average for this variant. Averaged over the shown $4{,}000$ iterations, the random direction removes 36% of the update norm, the critic-score projection 39%, the teacher-residual projection 43%, and PDMD 42%; the four deletions shrink the step by a similar amount. The ordering is consistent with the predicted error geometry. Among the tested directions, removing the student–critic endpoint residual is the only deletion that improves both metrics over DMD. Removing a random direction or the error-free teacher residual does not reproduce the gain. Since a rank-one deletion in $d = 2$ removes one of only two available directions, this experiment is a stringent signal-retention test; PDMD still improves both metrics over DMD.

Direct measurement of the removal ratios. The metrics above read the projection through its consequences. To look at the mechanism itself, Figure B1 measures γ_e and γ_s of Section A.2 directly along the PDMD run. In 2D we estimate the population-optimal critic endpoint by a kernel average over a bank of 2^{17} student samples, $\boldsymbol{c}^\star(q) = \sum_i \boldsymbol{x}_i \mathcal{N}(q; \boldsymbol{x}_i, \sigma^2 \boldsymbol{I}) / \sum_i \mathcal{N}(q; \boldsymbol{x}_i, \sigma^2 \boldsymbol{I})$; the same weights give $\operatorname{tr} \boldsymbol{\Sigma}(q)$. With the analytic teacher, this gives a Monte Carlo estimate of $\boldsymbol{d}^\star = (\boldsymbol{c}^\star(q) - \boldsymbol{x}_0^{\text{teacher}}(q))/\sigma^2$; the critic error is $\boldsymbol{e} = \boldsymbol{x}_0^c(q) - \boldsymbol{c}^\star(q)$ and the residual is $\boldsymbol{r} = \boldsymbol{x}_0^c(q) - \boldsymbol{x}_0^s$ for the probe's own endpoint. Every 100 iterations, 512 probes at each of 12 log-spaced noise levels in $[0.02, 5]$ give the per-sample $\gamma_e(\boldsymbol{r})$, $\gamma_s(\boldsymbol{r})$, and the lower bound $\|\boldsymbol{e}\|^2/(\|\boldsymbol{e}\|^2 + \operatorname{tr} \boldsymbol{\Sigma}(q))$ of Equation (10); the diagnostics use a separate random stream, so the measured run reproduces the stored PDMD run bit for bit. In $d = 2$ a uniformly random rank-one deletion removes half of the energy of any fixed vector on average, and $d_{\text{eff}} \leq 2$, so the high-dimensional limit $\gamma_s \to 0$ is not available in the plane. What the plane can test is preferential error removal: γ_e exceeds γ_s on average in the evaluated noise band (Table B1). Figure B1 shows exactly that split: γ_e peaks at 0.75 where the critic carries error, both ratios return to $1/2$ once the critic is accurate, and the lower bound of Equation (10) holds at every snapshot and noise level. Table B1 repeats the measurement for the other three deleted directions on the same probes: the random direction sits at $1/2$ for both ratios, the student–teacher endpoint residual removes more signal than error, and only the student–critic endpoint residual removes clearly more error than signal, 0.18 ± 0.01 more error energy than the random direction and above the theoretical bound in both blocks.

Table B2 reports the net effect of each deletion on the same probes, in the two forms of Equation (A36) and Corollary A6. With $\widetilde{\boldsymbol{d}} = \boldsymbol{P}_{\boldsymbol{b}}^{\perp} \boldsymbol{d}$ the update after deleting the direction $\boldsymbol{b}$, the error ratio $\nu = \mathbb{E}\|\widetilde{\boldsymbol{d}} - \boldsymbol{d}^\star\|^2 / \mathbb{E}\|\boldsymbol{d} - \boldsymbol{d}^\star\|^2$ is the distance of the projected update to the ideal update relative to that of the DMD update, and the improved fraction φ is the share of probes on which the projected update is closer to $\boldsymbol{d}^\star$, the empirical form of the probability in Corollary A6.

In this 2D setting, deleting the student–critic endpoint residual reduces the average error in the evaluated noise band: in the band $\sigma \in [0.15, 1.1]$ our projection makes the error fall by 28%. Deleting a random direction lowers it by 16%, the critic score by 20%, and the student–teacher endpoint residual by 8%; over all twelve noise levels the student–teacher endpoint residual raises the error by 8%. Paired over the same snapshot and noise level, the student–critic endpoint residual has a lower error ratio than the random direction in 82% of the cells in the band, than the critic score in 68%, and than the teacher endpoint residual in 99%. The same holds on a run trained without the projection: probes along the DMD run, where each deletion is computed at every snapshot but never used to update the student, give the same ordering, $\nu = 0.90$ for the student–critic endpoint residual against 0.98, 0.94 and 1.21.

B.2 Experiment 2: Two-Mode Allocation Under Critic Lag

Setup. The target is the symmetric two-mode mixture $\frac{1}{2}\mathcal{N}(+\boldsymbol{\mu}, \sigma_{\text{mode}}^2 \boldsymbol{I}) + \frac{1}{2}\mathcal{N}(-\boldsymbol{\mu}, \sigma_{\text{mode}}^2 \boldsymbol{I})$ with $\boldsymbol{\mu} = (2, 0)$ and $\sigma_{\text{mode}} = 0.2$, a Gaussian instance of the allocation slice in Equation (A42). The observable of interest is the antisymmetric mode imbalance $a_k = \text{share}(\text{samples in mode } +) - \frac{1}{2}$; the symmetric solution has $a = 0$ and total collapse has $|a| = \frac{1}{2}$. A critic that lags behind the student motivates the delayed-feedback model of Equation (A46), and the observable residual has conditional mean equal to the resulting tracking error (Equation (A51)). We stress this interaction by running the critic at a learning rate $20\times$ smaller than the student's and repeat the entire comparison at two step sizes to test whether the effect is specific to one setting: student 2×10^{-3} with critic 1×10^{-4}, and student 1×10^{-3} with critic 5×10^{-5}. Three variants use the same 20 consecutive seeds (0–19) at each step size: plain DMD, the random-direction control (b) of Section B.1, and PDMD; nothing

Table B1: **Removal ratios of the four deleted directions on the ring of eight.** Mean γ_e (critic error removed) and γ_s (ideal signal removed) over the same probes, noise draws, and snapshots as Figure B1, along the PDMD run. Left: the noise band $\sigma \in [0.15, 1.1]$ in which the critic carries error. Right: all 12 noise levels. In the plane a uniformly random rank-one deletion removes $1/2$ of any fixed vector's energy on average, so $1/2$ is the reference for both ratios. The last column of each block is the mean lower bound $\|\boldsymbol{e}\|^2/(\|\boldsymbol{e}\|^2 + \operatorname{tr} \boldsymbol{\Sigma}(q))$ of Equation (10) on the same probes; it is a bound on γ_e for the student–critic endpoint residual and is listed in that row.

	$\sigma \in [0.15, 1.1]$				all levels			
Deleted direction	γ_e	γ_s	$\gamma_e - \gamma_s$	γ_e bound	γ_e	γ_s	$\gamma_e - \gamma_s$	γ_e bound
Random direction	0.50	0.50	0.00	–	0.50	0.50	0.00	–
Critic score	0.52	0.49	0.03	–	0.52	0.49	0.03	–
Student–teacher residual	0.56	0.61	−0.05	–	0.55	0.59	−0.04	–
Student–critic residual (PDMD)	**0.68**	0.57	**0.11**	0.25	**0.61**	0.56	**0.04**	0.21

Table B2: **Net effect of the four deletions on the ring of eight.** Same probes, noise draws and snapshots as Table B1, along the PDMD run. ν (↓) is the error ratio $\mathbb{E}\|\widetilde{\boldsymbol{d}} - \boldsymbol{d}^\star\|^2/\mathbb{E}\|\boldsymbol{d} - \boldsymbol{d}^\star\|^2$: below 1 the deletion moves the update closer to the ideal update, above 1 farther away. φ (↑) is the improved fraction, the share of probes on which the projected update is closer to $\boldsymbol{d}^\star$ than the DMD update. Left: the noise band $\sigma \in [0.15, 1.1]$ in which the critic carries error. Right: all 12 noise levels. Best value per column in bold.

	$\sigma \in [0.15, 1.1]$		all levels	
Deleted direction	$\nu \downarrow$	$\varphi \uparrow$	$\nu \downarrow$	$\varphi \uparrow$
Random direction	0.84	0.68	0.95	0.63
Critic score	0.81	0.71	**0.88**	0.64
Student–teacher residual	0.92	0.67	1.08	0.62
Student–critic residual (PDMD)	**0.72**	**0.74**	0.89	**0.66**

else changes. This slow-tracking regime mirrors the practical motivation for DMD2, which adds more critic updates (TTUR) to improve tracking of the moving student (Yin et al., 2024a).

Results. In the lag-stressed regime, PDMD yields the predicted reduction in allocation collapse (Table B3): plain DMD collapses to a single mode in $14/20$ seeds (mean final $|a| = 0.355$), while PDMD collapses in $3/20$ (mean 0.138) at equal precision (mean on-mode fraction 0.973 vs. 0.970) and reaches a mean energy distance of 0.386 against DMD's 1.274. The random-direction control remains close to DMD, collapsing in $16/20$ seeds, with a mean final $|a| = 0.396$ and mean energy distance 1.449, while removing a mean 0.363 of the update norm against PDMD's 0.514: the reduction follows the direction that is removed rather than the amount. The two remaining directions of Section B.1 also reduce collapse in this setting: the critic-score projection yields $0/20$ and the student–teacher-endpoint-residual projection yields $5/20$, but with lower on-mode fractions (0.730 and 0.871 versus 0.973); Experiment 1 rules out both alternatives on accuracy grounds. The smaller step size reproduces the comparison: DMD collapses in $19/20$ seeds and the random direction in $19/20$, while PDMD collapses in $12/20$. The trajectories are more informative than the counts (Figure B3). DMD and PDMD both pass through a strongly imbalanced transient early in training (the one-step student initially covers a single mode, $|a| \approx 0.5$). Plain DMD then *locks*: every collapsed seed remains at $|a| = 0.500$ for the rest of training. PDMD *recovers*: a typical seed follows $|a|$: $0.50 \to 0.40 \to 0.15 \to 0.03$ and then fluctuates near balance. Corollary A9 provides a model for the first part of the transition: the projection attenuates the delayed component that drives the imbalance. Empirically, the remaining distribution-matching field transports mass back toward balance. Seventeen of the twenty PDMD seeds finish with both modes covered.

Connection to critic tracking in practice. We realize delayed tracking by lowering the critic learning rate. DMD2 addresses the same critic-tracking challenge in image distillation and increases the critic update frequency so that it follows the moving student more closely (Yin et al., 2024a). Our

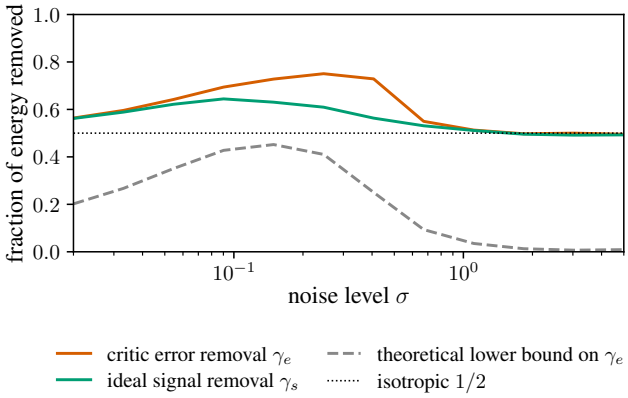

Figure B1: **Direct measurement of the removal ratios on the ring of eight.** Along the PDMD run of Section B.1: the fraction of critic-error energy γ_e and of ideal-signal energy γ_s that the projection removes, against the noise level. The grey dashed line is the theoretical lower bound $\|\boldsymbol{e}\|^2/(\|\boldsymbol{e}\|^2 + \operatorname{tr}\boldsymbol{\Sigma}(q))$ of Equation (10) on γ_e. Each curve is a mean over the snapshots from iteration 100 to 4,000. The dotted line marks the mean energy fraction $1/2$ removed by a uniformly random rank-one deletion from a fixed vector in the plane.

Table B3: **Two-mode collapse counts over 20 consecutive seeds at two step sizes.** Collapse is modes covered < 2 at the end of training; $|a|$ is the final imbalance; $|a|$, on-mode fraction, and energy distance are means $\pm$ standard deviations over seeds. The critic learning rate is $20\times$ smaller than the student's in both blocks. The random direction is control (b) from Section B.1: the same rank-one deletion along an unstructured direction.

Student lr	Variant	Collapsed ↓	Mean $\lvert a\rvert$ ↓	On-mode fraction	Mean energy ↓
2×10^{-3}	DMD	$14/20$	0.355 ± 0.219	0.970 ± 0.045	1.274 ± 0.856
	Random direction $\perp$	$16/20$	0.396 ± 0.197	0.975 ± 0.065	1.449 ± 0.756
	PDMD (ours)	$\mathbf{3/20}$	$\mathbf{0.138 \pm 0.162}$	0.973 ± 0.039	$\mathbf{0.386 \pm 0.689}$
1×10^{-3}	DMD	$19/20$	0.477 ± 0.104	0.913 ± 0.231	1.899 ± 0.618
	Random direction $\perp$	$19/20$	0.476 ± 0.108	0.991 ± 0.018	1.791 ± 0.440
	PDMD (ours)	$\mathbf{12/20}$	$\mathbf{0.331 \pm 0.215}$	0.925 ± 0.224	$\mathbf{1.317 \pm 1.046}$

two-mode experiment makes the associated allocation response directly observable, while the video results show its large-scale signature as degradation that grows with training (Figure 2).

C Experimental Details

This section describes the training and evaluation settings for the paper's three experimental domains and their baselines.

C.1 Wan2.1-T2V-1.3B

Training. We build training on the rCM codebase (Zheng et al., 2026) with its consistency loss disabled and the DMD implementation of Hyper-Bagel (Lu et al., 2025), so DMD is the only objective, and train every parameter of the student from the Wan2.1-T2V-1.3B weights. Training uses 42K captions without real-video latents. The frozen Wan2.1 teacher uses guidance scale 5 throughout distribution matching. Table C1 lists the optimizer, learning rates, batch and sampler. One critic step follows every student step.

Baselines. DMD2† is our reimplementation of DMD2 on Wan2.1-T2V-1.3B, with the DMD2 discriminator head and five critic steps per student step. Its discriminator positives are clean samples generated by rolling out MiniMax-H3, chosen for their higher visual quality than Wan2.1-T2V-1.3B

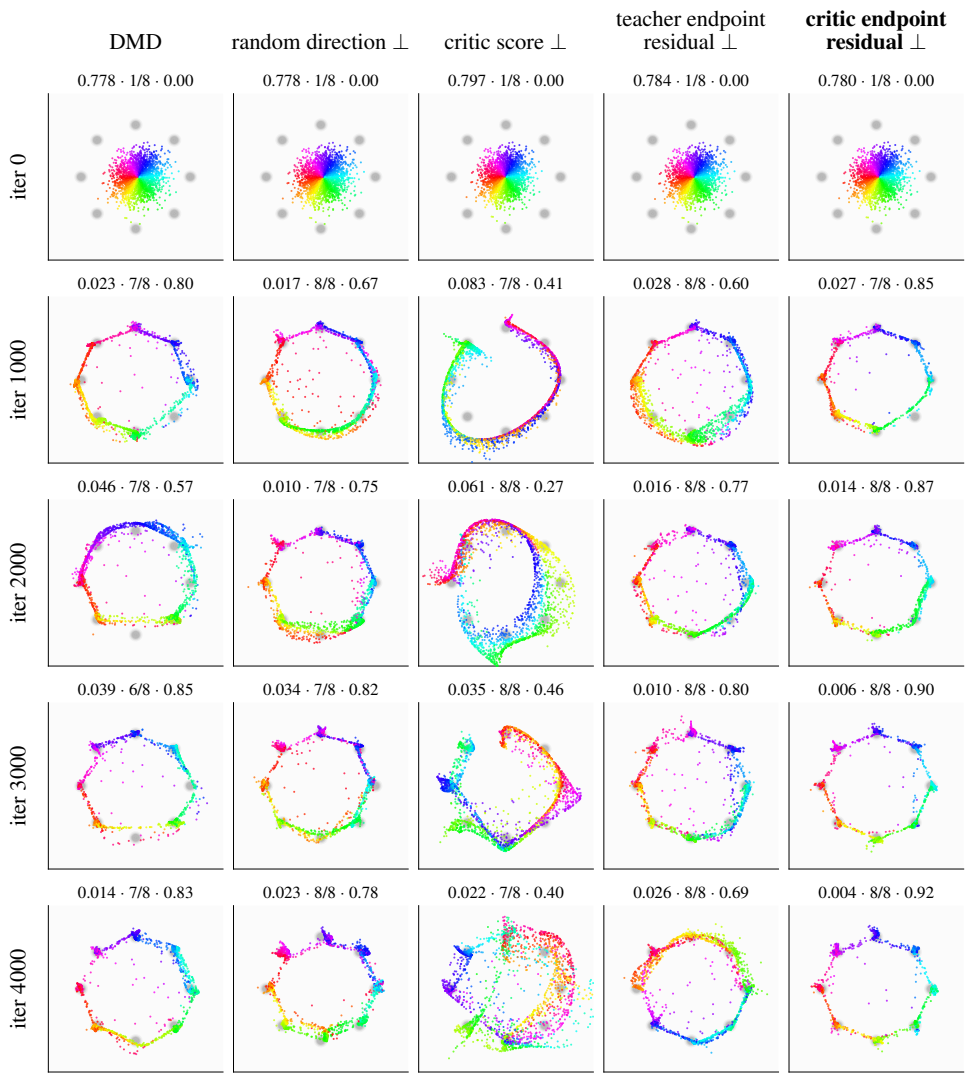

Figure B2: **Generated samples of the five 2D variants over training.** Columns are the five variants of Section B.1. Rows are iterations $0, 1{,}000, \ldots, 4{,}000$ of the same runs as Figure 4. Point color encodes the angle of the latent $\boldsymbol{z}$, visualizing how latent directions map to target modes. Each panel lists the energy distance (↓, lower is better), the number of covered modes (↑), and the on-mode fraction (↑, higher is better) at the shown iteration. The student–critic-endpoint-residual column (PDMD) covers all eight modes from iteration $2{,}000$ onward and stays tight. The critic-score column covers the modes but does not concentrate. The teacher-residual column shows the instability near iteration $3{,}000$.

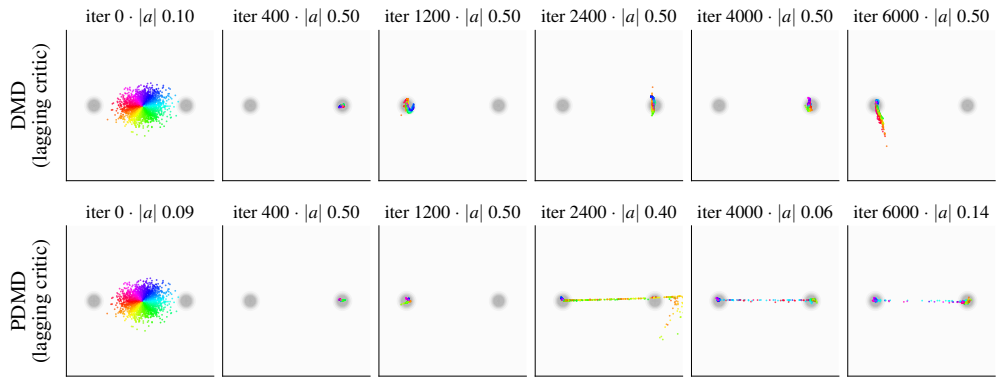

Figure B3: **Two-mode collapse process under a lagging critic (seed 0).** Rows are plain DMD and PDMD; columns are training iterations. Each panel lists the iteration and the imbalance $|a|$; $|a| = 0$ is the balanced solution and $|a| = 0.5$ is full collapse. Both variants start from a collapsed transient at iteration 400. Plain DMD stays locked on one mode for all 6,000 iterations. PDMD transports mass back between the modes (iteration 2,400 shows samples in transit) and settles near balance. The main text shows the collapsed DMD endpoint in Figure 5.

Table C1: **Training configurations of the video experiments.** Our DMD, DMD2, and PDMD runs in Table 1 share the Wan2.1 column and the DMD †, DMD2† and PDMD rows of Table 2 share the MiniMax-H3 column, except where a baseline paragraph says otherwise.

	Wan2.1-T2V-1.3B	MiniMax-H3-33B
Trainable parameters	all	LoRA, rank and scaling 128
Training data	42K captions, no real video	248K prompts, no real video
DMD2 discriminator positives	H3 teacher rollouts	H3 teacher rollouts
Clip	81 frames, 480×832	5 s, 544p, mixed aspect ratios
GPUs / global batch	64 H100 / 64	16 H100 / 16
Optimizer	AdamW $(0, 0.999)$, wd 0.01	AdamW $(0, 0.9)$, no wd
Student / critic learning rate	4×10^{-7} / 8×10^{-8}	5×10^{-5} / 1×10^{-5}
Critic updates per student update	1	5
Teacher guidance in distillation	5	none
Student sampler (4 NFE)	$\sigma_{\max} = 320$, $t = 0.934, 0.76, 0.11$	released scheduler, shift 12 / 3

rollouts. These H3-generated samples provide the adversarial supervision; the distribution-matching teacher remains Wan2.1-T2V-1.3B. DMD † removes the discriminator head and the extra critic steps, so the critic and the student update once each. PDMD adds only the projection to the DMD † run. Table 1 reports DMD † at 1,500 iterations, DMD2† at 6,500 and PDMD at 5,500, the best checkpoint of each run under the evaluation below. rCM, ADV (You et al., 2026) and AnyFlow are the publicly released Wan2.1-T2V-1.3B checkpoints: rCM samples with $\sigma_{\max} = 1600$ and intermediate times 0.783, 0.609 and 0.406, ADV with intermediate times 0.900, 0.757 and 0.522, and AnyFlow with its official 4-step bidirectional setting.

Evaluation. We reuse AnyFlow's evaluation end to end: its 944 augmented VBench prompts with five samples each (4,720 clips), its per-sample seeds and video writer, and the same 16 VBench dimensions and weights for the quality, semantic and total scores. All distilled rows use 4 NFEs without classifier-free guidance. The two teacher rows use UniPC with 50 and 4 steps, a shift of 8, and guidance scale 5, which doubles their NFE. The dynamic-degree column is the VBench dimension of that name.

User study. The preference columns of Table 1 follow the MiniMax-H3 protocol of Section C.2 without the audio or overall-quality questions, with a pool of 20 annotators. The prompts are a uniform subset of VBench, every tenth prompt, 95 in all, and every clip uses seed 0. PDMD is paired with each of six baselines of Table 1: the 50-step teacher, rCM, ADV, AnyFlow, DMD† and DMD2†, 570 pairs in all, each judged on text alignment, visual quality and motion quality. Table C3 gives the

Table C2: **Two-stage recipes of the rCM † and AnyFlow † ports on MiniMax-H3.** Both follow their reference: a trajectory stage first, then a stage that adds DMD on top of it. Every other setting follows Table C1. Iterations count every update, as in Section C.2, so the student update count of a stage with a critic is a fraction of it. Stage 1 is listed at the checkpoint that initializes stage 2; training it further did not reach the reported model.

	Stage 1	Stage 2 (the reported model)
rCM †: dCM, then dCM + DMD		
Learning rate	5×10^{-6}	5×10^{-5} (critic 2×10^{-5})
Iterations	3,000	6,000, saved every 500
Initialization	LoRA from a full-parameter run	stage 1
Student:critic	no critic	1:4
Loss	L_{dCM}	$100\, L_{\mathrm{dCM}} + L_{\mathrm{DMD}}$
Reported	–	2,000 iterations, 400 student updates
AnyFlow †: flow map, then flow map + DMD		
Learning rate	5×10^{-5}	5×10^{-6} (critic 5×10^{-6})
Iterations	500	3,000, saved every 500
Initialization	base model	stage 1
Student:critic	no critic	1:1
Loss	L_{flowmap}	$L_{\mathrm{flowmap}} + 1.0\, L_{\mathrm{DMD}}$
Reported	–	3,000 iterations, 1,500 student updates

Table C3: **Vote distribution behind the Wan2.1 user study of Table 1.** Each row is PDMD against one baseline, on the three questions of that study. Win, tie and loss are shares of the judgments, in percent, and net is the win share minus the loss share. Text alignment draws ties on roughly 82–88% of judgments throughout; the visual and motion questions separate them. The grey row is the 50-step teacher.

	Text alignment				Visual quality				Motion quality			
Baseline	Win	Tie	Loss	Net	Win	Tie	Loss	Net	Win	Tie	Loss	Net
Wan2.1-T2V-1.3B, 50×2 NFE	8.5	82.5	8.9	−0.4	36.7	44.3	19.1	+17.6	39.9	44.5	15.6	+24.3
DMD †	9.3	87.8	2.8	+6.5	58.1	37.4	4.5	+53.6	37.7	59.8	2.5	+35.2
DMD2†	10.1	87.2	2.7	+7.5	49.9	39.7	10.4	+39.5	35.9	59.0	5.1	+30.9
ADV (You et al., 2026)	10.7	81.5	7.8	+2.9	46.6	41.4	12.0	+34.6	33.3	52.7	14.1	+19.2
rCM (Zheng et al., 2026)	5.7	86.6	7.8	−2.1	46.7	40.4	12.9	+33.8	29.6	58.4	11.9	+17.7
AnyFlow (Gu et al., 2026)	8.7	83.3	8.1	+0.6	39.5	45.0	15.5	+24.0	39.7	49.5	10.8	+28.9

win, tie and loss shares. The undistilled 4-step model is left out because many of its clips are black frames.

C.2 MiniMax-H3-33B

Training. The student and critic use LoRA adapters with rank and scaling parameter 128 on the attention projections and both feed-forward layers of every DiT block; the teacher is frozen and, being guidance-distilled, is sampled without classifier-free guidance. Training uses the 248K prompt-only VidProM prompts (Wang & Yang, 2024) released with Self-Forcing (Huang et al., 2025), without rewriting and without real-video data. Each GPU holds one 5-second 544p clip; training batches contain mixed aspect ratios. Table C1 lists the optimizer, learning rates, and batch size. The projection of Equation (4) is applied per sample and separately to the video and audio latents, each with its own residual $\boldsymbol{r}$. The remaining training choices follow Hyper-Bagel (Lu et al., 2025), a multi-step DMD distillation framework. We also adopt the two-timescale update rule (TTUR, several critic updates per student update) of DMD2 (Yin et al., 2024a): five critic updates follow every student update, and we count every update as one iteration.

Baselines. All DMD-family rows of Table 2 share the setup above. PDMD is reported at 3,500 iterations. DMD † is reported at 5,500 iterations of the run with both learning rates divided by

ten, the best DMD run of our learning-rate search; at the PDMD learning rate DMD peaks at 500 iterations and degrades afterwards (Figure 10). DMD2† adds the DMD2 discriminator head and is reported at 500 iterations at the PDMD learning rate. The discriminator's positive examples are clean samples generated by rolling out the frozen MiniMax-H3 teacher; no real-video data are used. The same search was run for DMD2†, with both learning rates divided by ten and by one hundred over 0–6,000 iterations. No checkpoint of either run exceeded the DMD2 500-iteration score at the original learning rate. rCM † and AnyFlow † each port a two-stage reference recipe to MiniMax-H3; Table C2 gives both stages and Section C.2 the two places where our port departs from the reference. rCM † is reported at 2,000 iterations of its second stage and AnyFlow † at 3,000, which at their update ratios are 400 and 1,500 student updates against PDMD's 3,500 iterations. For both rCM † and AnyFlow † we trained three learning rates, the one in Table C2 and that rate multiplied and divided by five, and report the best of the three. We also varied the weight of the DMD term of rCM † and saw no clear difference in the final scores. H3 Turbo LoRA (LarryVrh, 2026) is the released v4-600 EMA adapter.

Where the ports depart from the reference

Two settings could not be carried over unchanged. First, both references obtain the time derivative of the student with a Jacobian–vector product, which MiniMax-H3 does not support, so we use the finite-difference tangent that the reference implementation also provides (its `fd_type=2`) and approximate $\mathrm{d}u/\mathrm{d}t$ by finite differences in both ports. Second, the AnyFlow † flow-map stage uses $\epsilon = 0.005$, a fresh coupling per step, flow-map ratios 0.5 and 0.25, and fp32 embedders; its second stage applies DMD to the rollout endpoint with $N \in \{2, 4, 8, 16, 50\}$, and keeps the reference's rule that the critic and student learning rates are equal.

Evaluation. Every MiniMax-H3 score in the paper, including the per-checkpoint tracking of Section 6, is computed on all 387 prompts of VideoGen-Eval (Yang et al., 2025) with one harness; we do not reimplement any scoring rule. Each prompt is rewritten once with Qwen3.8-27B (Qwen Team, 2026), using the official MiniMax-H3 video-prompt writing guide (MiniMax, 2026b) as the system prompt, and the same rewritten prompt is given to every method. One clip per prompt is rendered at 4 NFEs with the released scheduler (shift 12 for video and 3 for audio), 124 frames at 24 fps, seed 42, and 544p with a prompt-dependent aspect ratio: across the 387 prompts we cycle through 21:9, 9:21, 16:9, 9:16, 1:1, 4:3 and 3:4. Quality is computed from the seven prompt-independent VBench (Huang et al., 2023) dimensions (subject consistency, background consistency, temporal flickering, motion smoothness, dynamic degree, aesthetic quality, imaging quality), normalized and weighted as in VBench. The nine semantic dimensions (object class, multiple objects, human action, color, spatial relationship, scene, appearance style, temporal style, overall consistency) follow each VBench dimension's own rule, with Qwen3.8-27B as the vision-language judge; the semantic score is their VBench-weighted average, and the total score is $(4 \cdot \mathrm{quality} + \mathrm{semantic})/5$. Semantic dimensions differ in how many of the 387 prompts they apply to (spatial relationship 25, appearance style 30, overall consistency all 387), so we read them at the aggregate level.

Audio metrics. The first four audio columns of Table 2 evaluate the generated audio alone: PQ, CE and CU are the production quality, content enjoyment and content usefulness of Audiobox Aesthetics (Tjandra et al., 2025), and IS is the Inception score (Salimans et al., 2016) over PaSST (Koutini et al., 2022) audio tags. Two additional metrics compare the audio with the video: IB is the ImageBind (Girdhar et al., 2023) cosine between the audio and the video of the same clip, and DeSync is the audio–video offset in seconds predicted by Synchformer (Iashin et al., 2024); both come from the av-benchmark suite of MMAudio (Cheng et al., 2025). PDMD leads both among the 4-NFE rows.

User study. The preference columns of Table 2 come from a pairwise study on the same 387 VideoGen-Eval clips as the table, judged by 20 annotators. Each pair shows the PDMD clip and the clip of one baseline for the same prompt and seed, side by side in random left–right order, and the annotator picks the better clip or a tie on five questions: an overall judgment, text alignment, visual quality, motion quality, and audio quality. The seven baselines are the rows of Table 2, including the 50-step teacher and the undistilled 4-step model. For each baseline and question the table reports the share of votes for PDMD minus the share of votes for the baseline, in percent, so a positive number favors PDMD and ties count for neither side. Table C4 gives the underlying win, tie and loss shares. All 387 pairs were judged for each baseline.

Table C4: **Vote distribution behind the MiniMax-H3 user study of Table 2.** Each row is PDMD against one baseline, on the five questions of that study. Win, tie and loss are shares of the judgments, in percent, and net is the win share minus the loss share, computed from the displayed percentages. Text alignment draws ties on roughly 64–80% of judgments throughout. Grey rows are the undistilled base model at 50 and 4 sampling steps.

	Overall quality				Text alignment				Visual quality				Motion quality				Audio quality			
Baseline	Win	Tie	Loss	Net	Win	Tie	Loss	Net	Win	Tie	Loss	Net	Win	Tie	Loss	Net	Win	Tie	Loss	Net
MiniMax-H3-33B, 50 NFE	14.2	34.6	51.2	−37.0	10.6	71.3	18.1	−7.5	8.0	48.6	43.4	−35.4	8.0	72.1	19.9	−11.9	7.5	81.1	11.4	−3.9
MiniMax-H3-33B, 4 NFE	79.3	11.6	9.0	+70.3	26.4	63.8	9.8	+16.6	78.3	16.0	5.7	+72.6	50.1	44.4	5.4	+44.7	49.6	48.3	2.1	+47.5
H3 Turbo LoRA	51.2	32.3	16.5	+34.7	13.4	76.0	10.6	+2.8	49.9	38.2	11.9	+38.0	26.9	65.6	7.5	+19.4	13.7	81.9	4.4	+9.3
DMD †	45.2	39.3	15.5	+29.7	15.2	73.1	11.6	+3.6	34.9	55.3	9.8	+25.1	19.9	72.4	7.8	+12.1	15.5	79.8	4.7	+10.8
DMD2†	51.7	31.8	16.5	+35.2	19.1	70.8	10.1	+9.0	39.0	50.6	10.3	+28.7	27.6	64.1	8.3	+19.3	26.6	69.0	4.4	+22.2
rCM †	60.2	23.3	16.5	+43.7	16.3	70.5	13.2	+3.1	62.8	25.6	11.6	+51.2	39.8	51.9	8.3	+31.5	27.9	61.2	10.9	+17.0
AnyFlow †	69.5	20.2	10.3	+59.2	9.0	80.4	10.6	−1.6	72.4	21.7	5.9	+66.5	33.9	60.2	5.9	+28.0	30.0	65.9	4.1	+25.9

D Additional Controlled Experiments on MiniMax-H3

The ablations of Section 6 change the projection direction only and keep every other setting fixed. This section provides additional measurements and controls for those runs. Section D.1 follows the six projection variants on three prompts to 3,000 iterations. Section D.2 compares retained update norms, and Section E.1 decodes one training batch to show what the projection removes. Sections D.3 and D.4 address two questions left open by the ablations: whether PDMD stays stable only because the projection shrinks the update, and whether a partial projection still stabilizes training. Section D.6 repeats the comparison at 2 NFE, and Section D.5 trains PDMD with one critic update per student update. Every run uses the 4-NFE MiniMax-H3 recipe of Section 6: the same initialization, data, LoRA rank, TTUR schedule, and sampler, except where the subsection says otherwise. We save a checkpoint every 500 iterations and score each one on the 387 VideoGen-Eval prompts with the protocol of Section C.2. Scores are in percent.

D.1 Projection Directions Over Training

Figure 9 shows one prompt at 1,500 iterations. Figure D1 follows the same six variants on three SoRA prompts to 3,000 iterations, with the frames at 500, 1,500 and 3,000 iterations stacked under each prompt. In these examples, all five control variants degrade in different ways, while PDMD continues to follow its prompt at 3,000 iterations.

D.2 Retained Update Norm

The four projection variants shown in Figure D2, trained on MiniMax-H3 in the 4-NFE setup of Section 6, each retain one component of the DMD update $\boldsymbol{d}$ and discard its complement. The error-removal ratio γ_e of Table B1 is not available here: it is defined against the population-optimal critic endpoint $\boldsymbol{c}^\star(q)$, which we estimate from a large student-sample bank in the 2D setting of Section B.1. In large-scale video generation we therefore measure what is observable, how much of the update each variant retains, and read the error question from three proxies: this norm ratio, the decoded removed component of Section E.1, and the partial-projection control of Section D.3. We log, at every student step, the ratio $\|\boldsymbol{d}_{\text{kept}}\|/\|\boldsymbol{d}\|$ between the retained component and the full update, separately for the video and the audio latents. Figure D2 shows the ratio over training. The DMD variant applies no projection, so its ratio is 1 by definition and is not drawn.

Three observations follow. First, the random-direction variant keeps the full norm: a random direction in a high-dimensional latent space is nearly orthogonal to $\boldsymbol{d}$, so removing that component changes almost nothing, and the variant behaves like DMD. Second, PDMD removes between 8% and 17% of the video update norm for over 90% of training (median 12%). The critic-score variant removes a median of 6% and still degrades; it removes more only late in training, after its scores have already collapsed (Figure 10). These comparisons suggest that a smaller update alone does not explain the stability of PDMD. Third, the residual-parallel variant keeps only 5–30% of the video update (median 11%). This variant degenerates fastest, consistent with the parallel component containing a destabilizing part of the update. Taken together, Figure D2 shows that stability does not follow the amount removed: the random direction removes almost none of the update and degrades, the

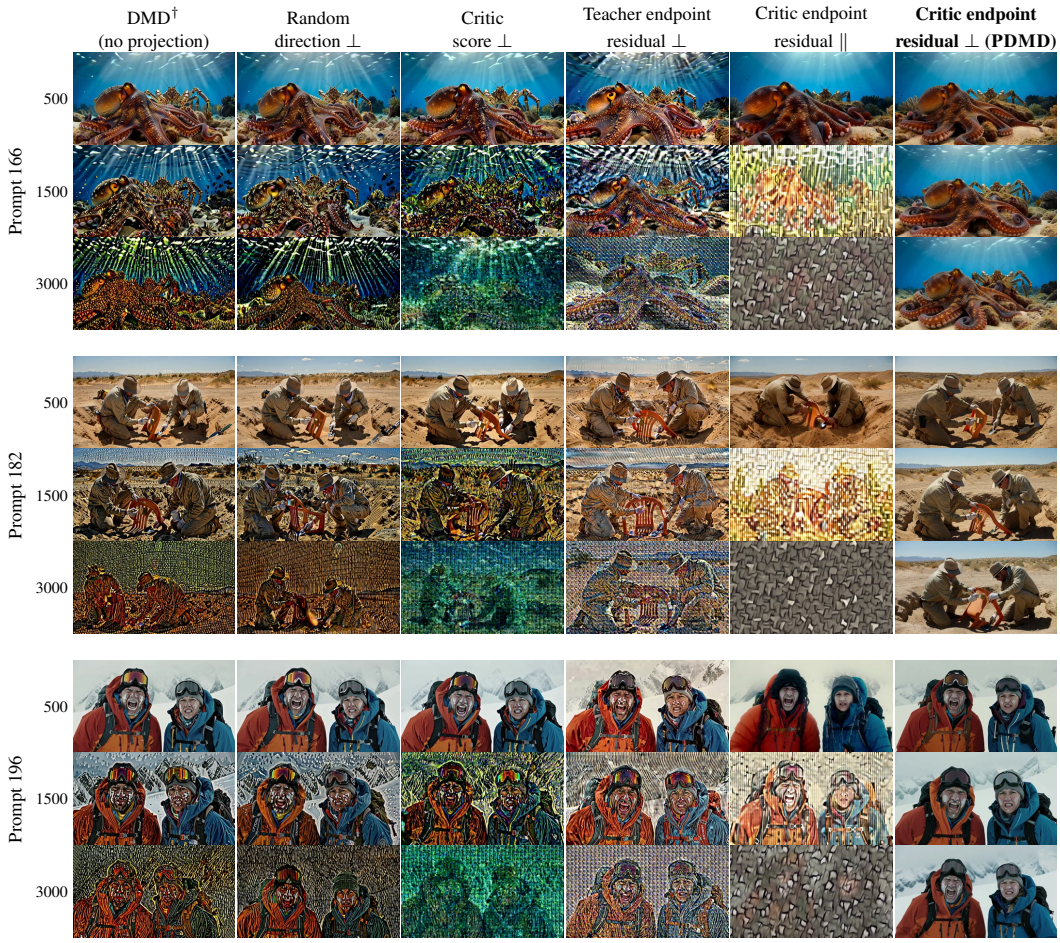

Figure D1: **Effect of projection direction during 4-NFE MiniMax-H3 distillation.** Six variants that differ only in which direction is removed from the DMD update $\boldsymbol{d}$, one per column, in the order of Figure 9. Each row is one of three SoRA prompts, given by its prompt id, and every row is a tight vertical triple: the same run at 500, 1,500 and 3,000 iterations, top to bottom, so reading downwards follows one training trajectory. Every frame is taken at 30% of the generated 1344×768 clip; prompt and seed are the same across variants. In these examples, all five control variants degrade in different ways. Retaining the student-critic endpoint-parallel component instead of removing it is the fastest failure: by 3,000 iterations it produces the same prompt-independent pebble texture for every prompt. PDMD continues to follow its prompt at 3,000 iterations. The five control variants show severe degradation on all three illustrated prompts.

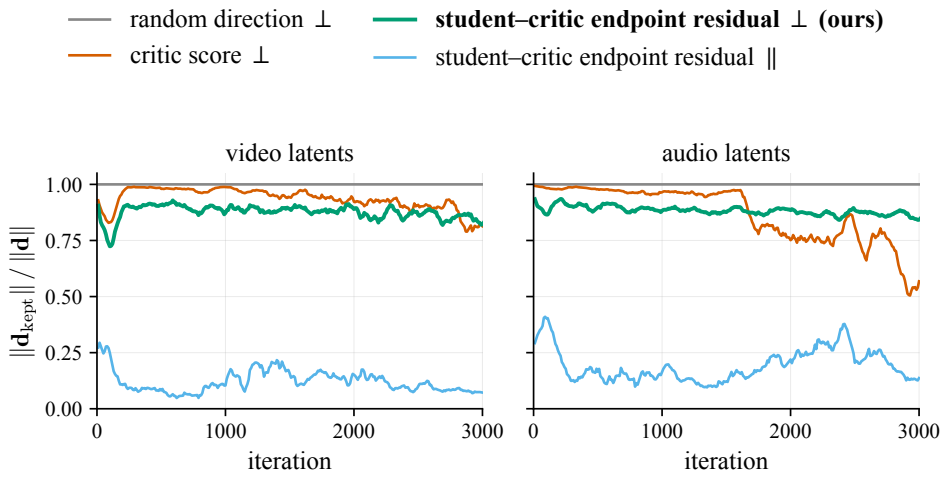

Figure D2: **Retained fraction of the DMD update norm on MiniMax-H3 at 4 NFE.** d_{kept} is the component the variant keeps, as a 100-iteration running mean. Left: video latents. Right: audio latents.

residual-parallel variant retains a median 11% and degrades fastest, whereas PDMD retains 88% and remains stable over the evaluated interval. What separates them is which component is removed, which is the ordering γ_e measures directly in 2D.

D.3 Restoring Half of the Removed Component

PDMD removes the component of $\boldsymbol{d}$ along the student–critic endpoint residual $\boldsymbol{r}$ entirely. To test whether a partial projection is enough, we interpolate between DMD and PDMD with

$$\boldsymbol{d}_\beta = \boldsymbol{d}_\perp + \beta \ \text{proj}_{\boldsymbol{r}}(\boldsymbol{d}), \tag{D1}$$

where $\beta = 0$ recovers PDMD and $\beta = 1$ recovers plain DMD. We train the midpoint $\beta = 0.5$ for 5,000 iterations with the learning rates of PDMD. At $\beta = 0.5$ half of the removed component is put back. The residual-parallel component scales linearly with β. If it drives degradation, restoring it should reduce the stability benefit; the analysis does not predict a quantitative degradation rate.

Figure D3 is consistent with this prediction. The $\beta = 0.5$ run stays at the level of the stable PDMD run for the first 2,000 iterations, with a total score between 82.64 and 82.98. The run crosses the 4-step base-model score at twice the iteration count of DMD. DMD drops below the 4-step base model (79.48 total) at 2,000 iterations and reaches 70.07 at 2,500; the $\beta = 0.5$ run drops below the base model at 4,000 iterations and reaches 74.90 at 5,000. In this setting, restoring half of the residual-parallel component is enough to reproduce the degradation, while removing the full component remains stable over the evaluated interval.

D.4 A Larger Student Step

Section D.2 shows that PDMD removes a median 12% of the video update norm. To test whether the projection helps only because its update is smaller than the DMD update, we train PDMD with the student learning rate raised from 5×10^{-5} to 6.25×10^{-5}, a factor of 1.25, and keep the critic learning rate at 1×10^{-5}. The projected update keeps between 83% and 92% of the DMD norm for most of training (Section D.2), motivating this scale factor: $1.25 \times 0.83 > 1$. This is a learning-rate stress test, not an exact match of parameter-update norms, because the generator Jacobian and AdamW also transform the update.

Figure D3 plots the run next to DMD at the base learning rate. The $1.25\times$ run does not exhibit a sustained decline over this interval. The total score of every one of the ten checkpoints lies between 82.82 and 83.48, above the 50-step teacher (82.41), and the last checkpoint at 5,000 iterations scores 83.18. DMD at the base learning rate, in the same figure, has fallen to 70.07 by 2,500 iterations. A

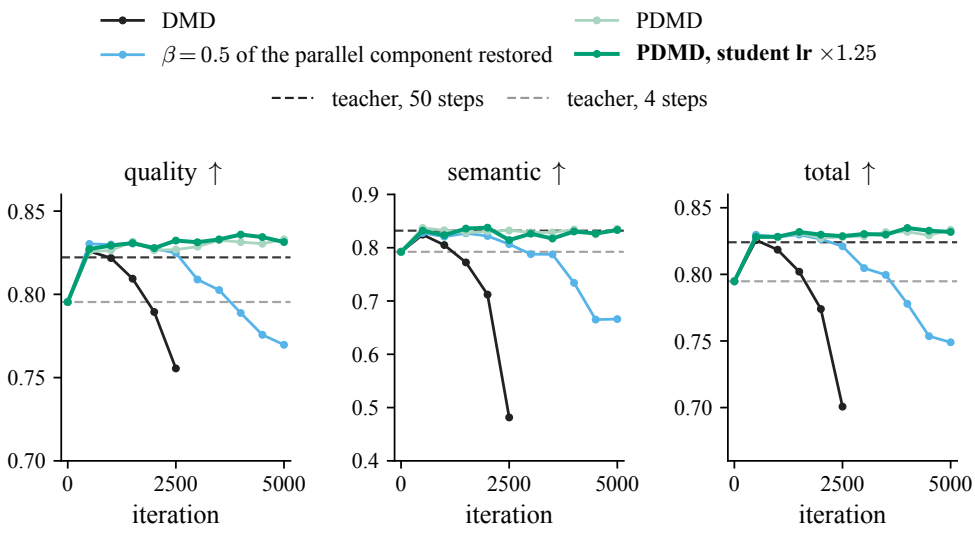

Figure D3: **A larger student step and a partial projection on MiniMax-H3 at 4 NFE.** Quality, semantic, and total scores of every 500-iteration checkpoint up to 5,000 iterations on the 387 VideoGen-Eval prompts. DMD is the corresponding variant in Figure 10. The light green curve shows PDMD at the base learning rate, as far as it was scored, and the dark green curve shows PDMD trained with a $1.25\times$ larger student learning rate (Section D.4). The blue curve shows the run that restores half of the residual-parallel component removed by PDMD, using Equation (D1) with $\beta = 0.5$ (Section D.3). Iteration 0 is the undistilled base model at 4 NFE. Horizontal lines show the teacher scores at 50 and 4 sampling steps.

larger step does not reproduce the degradation, suggesting that update direction, rather than magnitude alone, drives the stability gain. Figure D4 shows frames from four checkpoints of the same run, and the frames remain visually consistent over this interval.



Figure D4: **PDMD with a** $1.25\times$ **larger student step also does not degrade over 5,000 iterations.** The first four columns are checkpoints of the single PDMD run of Section D.4. The last column, set slightly apart, is the best DMD checkpoint over a learning-rate search, the DMD † row of Table 2. Rows are labelled with the VideoGen-Eval prompt id; every panel is the frame at 80% of the clip, with the same prompt, seed and sampler throughout.

D.5 One Critic Update per Student Update on MiniMax-H3

The DMD-family MiniMax-H3 rows of Table 2 take five critic updates per student update, the two-timescale update rule (TTUR) of DMD2 described in Section C.2. On Wan2.1, our DMD and PDMD runs take one critic update per student update, while DMD2† takes five. We repeat PDMD on MiniMax-H3 with one critic update per student update and every other setting of Section C.2 unchanged. Table D1 compares the 1,000-iteration checkpoint of this run with the DMD † and PDMD rows of Table 2, which keep the five critic updates, and with DMD trained with one critic update per student update, reported at 1,500 iterations, on the video and audio metrics. With one critic update per student update, PDMD scores 82.90 total, above DMD † with five and above every 4-NFE baseline of Table 2. Our projection does not depend on the extra critic updates. It is compatible with them.

Table D1: **PDMD with one critic update per student update on MiniMax-H3.** VideoGen-Eval scores on the 387 prompts (percentages) and audio metrics of the same clips, as in Table D2. Student : critic is the ratio of student to critic updates; the best value per column is bold.

			Video			Audio quality				Audio–video	
Method	NFE	Student : critic	**Total**	Quality	Semantic	PQ	CE	CU	IS	IB	DeSync ↓
DMD †	4	1 : 1	82.37	82.24	82.89	6.321	3.674	5.900	4.32	0.148	0.891
DMD †	4	1 : 5	82.76	82.68	**83.05**	6.381	3.801	5.931	4.79	0.176	0.905
PDMD	4	1 : 1	82.90	83.14	81.92	**6.538**	**4.123**	**6.209**	**5.08**	0.191	0.808
PDMD (Table 2)	4	1 : 5	**83.17**	**83.25**	82.86	6.530	4.062	6.180	4.98	**0.195**	**0.802**

D.6 Two-Step Generation

We repeat the comparison at 2 NFE. The two runs share the setup of Section C.2 with the student sampled in two steps; the projection is the only difference between them. Over the 6,000 iterations, every checkpoint of PDMD scores 80 total or above on the 387 VideoGen-Eval prompts, while the run without the projection falls below 60. Table D2 reports the best DMD checkpoint at 1,000 iterations and the PDMD checkpoint at 2,000, with the audio metrics of the same two checkpoints: PDMD leads on all six. Figure D5 shows frames of these two checkpoints next to the four-step models. Figures D6 to D9 provide eight additional prompts comparing DMD at 2 NFE with PDMD at 2 and 4 NFE, with three uniformly sampled frames per clip.

Table D2: **Best checkpoint of DMD and PDMD at 2 NFE on MiniMax-H3.** VideoGen-Eval scores on the 387 prompts (percentages) and audio metrics of the same clips. PQ, CE, CU and IS are the audio-quality metrics of Table 2; IB and DeSync are the audio–video agreement metrics of Table 2. Arrows give the preferred direction; the best value per column is bold.

		Video			Audio quality				Audio–video	
Method	NFE	**Total**	Quality	Semantic	PQ	CE	CU	IS	IB	DeSync ↓
DMD †	2	81.58	81.87	80.39	4.529	2.380	3.781	1.88	0.085	1.022
PDMD (ours)	2	**82.88**	**82.83**	**83.05**	**6.061**	**3.719**	**5.597**	**4.16**	**0.177**	**0.838**

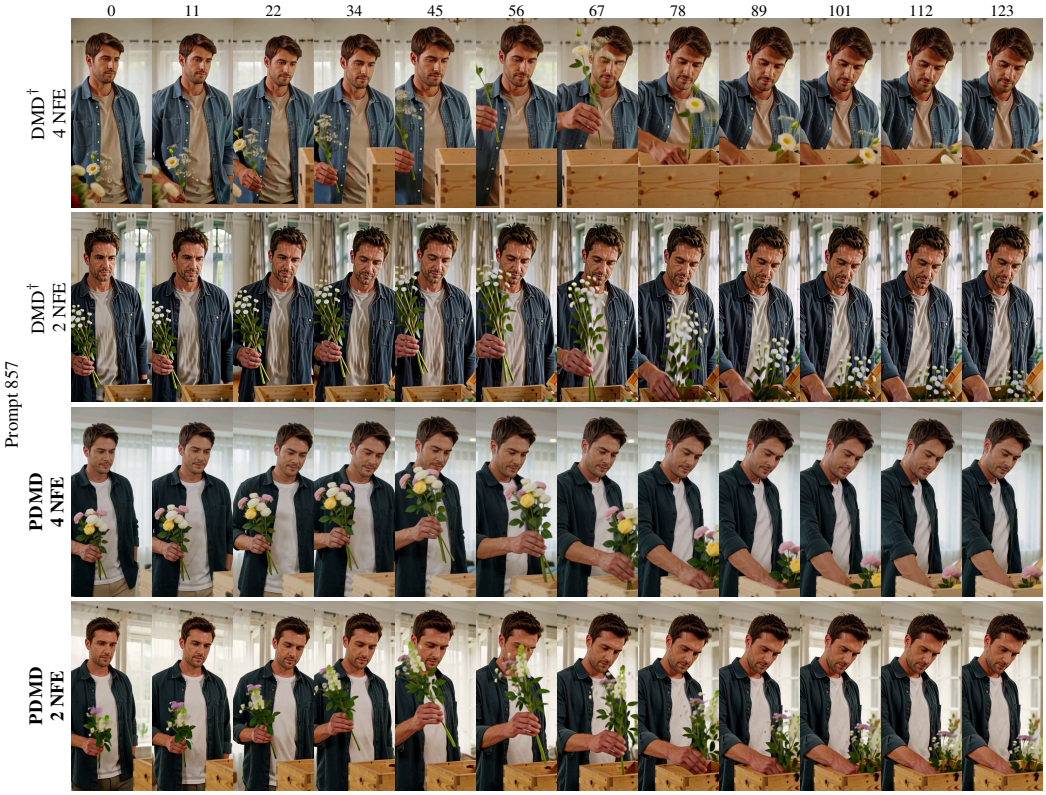

Figure D5: **Motion comparisons of DMD and PDMD at 2 NFE and 4 NFE.** DMD† at 4 and 2 NFE and PDMD at 4 and 2 NFE, using the checkpoints reported in Tables 2 and D2. VideoGen-Eval prompt 857 on MiniMax-H3, one row per model; twelve uniformly spaced frames of the 124-frame clip, left to right, labelled by frame index, with the same prompt and seed across rows. Against PDMD at 4 NFE, PDMD at 2 NFE shows slightly lower motion quality and less realistic object texture, but remains clearly better than the unprojected DMD† at 2 NFE.



Figure D6: **Additional two-step generation results (part 1 of 4).** DMD† at 2 NFE, PDMD at 2 NFE, and PDMD at 4 NFE on MiniMax-H3 for VideoGen-Eval prompts 736 and 735. Each row shows three uniformly sampled frames (0, 62, 123; zero-based) from a 124-frame clip, with the original aspect ratio preserved. † denotes our reimplementation.

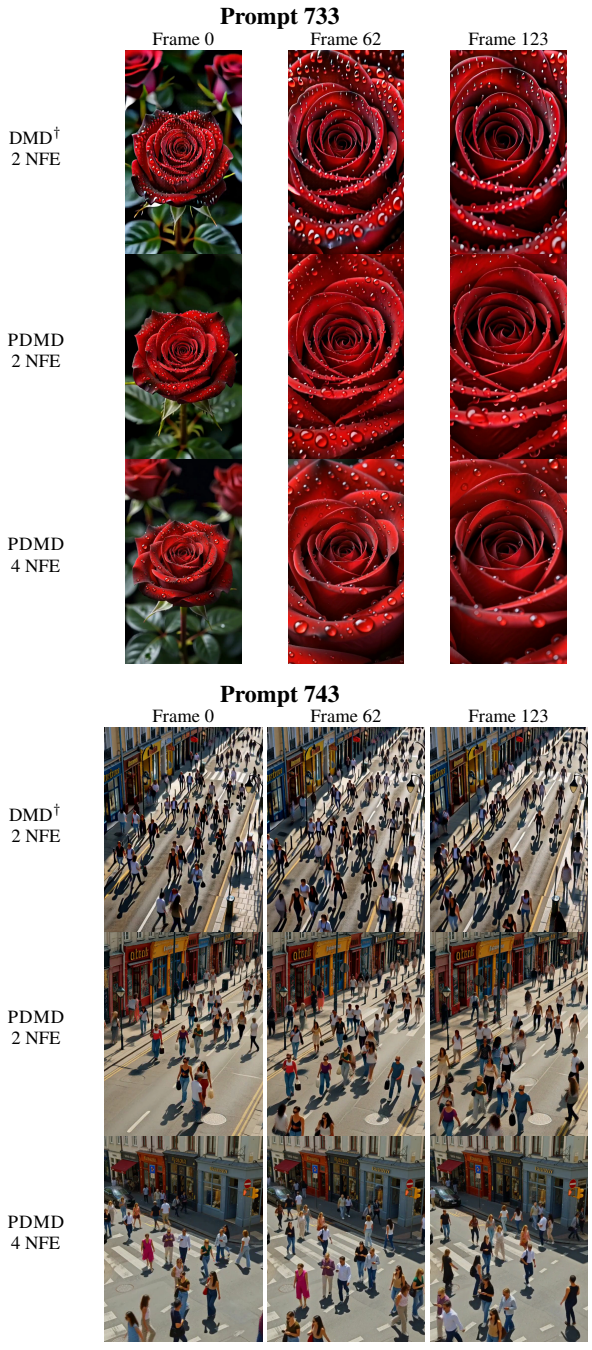

Figure D7: **Additional two-step generation results (part 2 of 4).** DMD† at 2 NFE, PDMD at 2 NFE, and PDMD at 4 NFE on MiniMax-H3 for VideoGen-Eval prompts 733 and 743. Each row shows three uniformly sampled frames (0, 62, 123; zero-based) from a 124-frame clip, with the original aspect ratio preserved. † denotes our reimplementation.

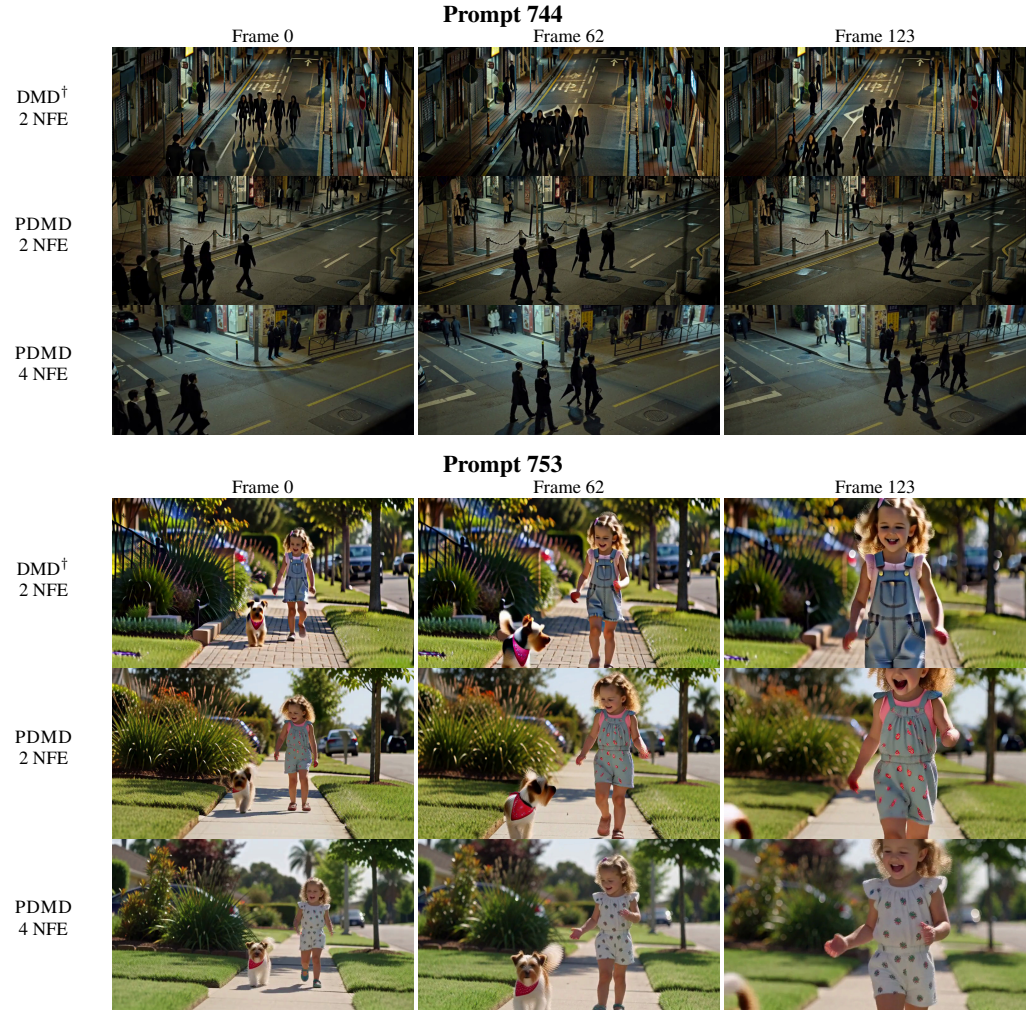

Figure D8: **Additional two-step generation results (part 3 of 4).** DMD† at 2 NFE, PDMD at 2 NFE, and PDMD at 4 NFE on MiniMax-H3 for VideoGen-Eval prompts 744 and 753. Each row shows three uniformly sampled frames (0, 62, 123; zero-based) from a 124-frame clip, with the original aspect ratio preserved. † denotes our reimplementation.

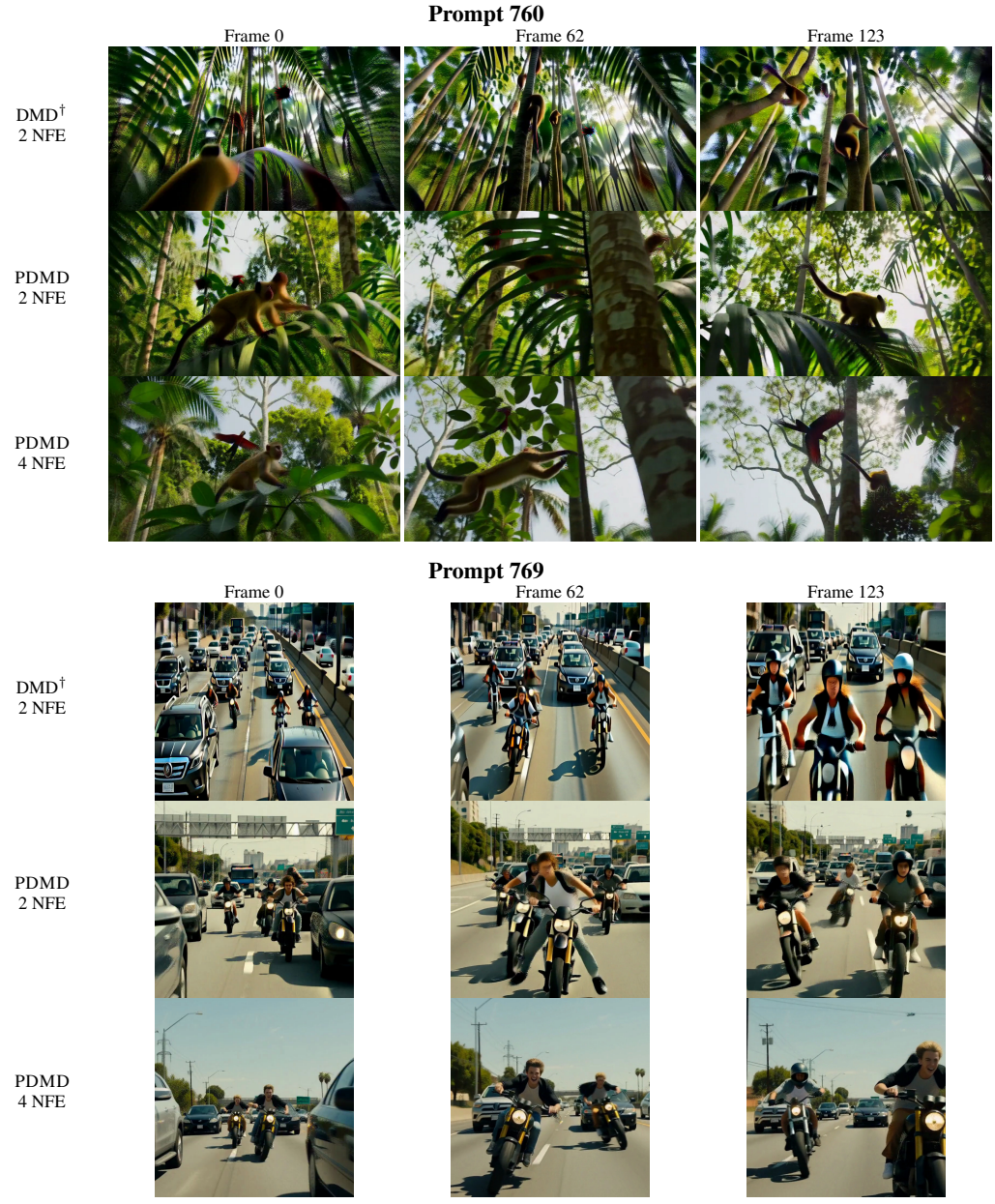

Figure D9: **Additional two-step generation results (part 4 of 4).** DMD† at 2 NFE, PDMD at 2 NFE, and PDMD at 4 NFE on MiniMax-H3 for VideoGen-Eval prompts 760 and 769. Each row shows three uniformly sampled frames (0, 62, 123; zero-based) from a 124-frame clip, with the original aspect ratio preserved. † denotes our reimplementation.

E Additional Visual Comparisons

E.1 The Removed Component in Video Space

The norm ratio measures how much of the DMD update the projection removes. Figure E1 shows what the removed part looks like after decoding. The images come from the PDMD run of Section 6 at iteration 500. At each checkpoint, every rank decodes the sample from its last student step three ways: the student endpoint $\boldsymbol{x}_s$, the DMD target $\boldsymbol{x}_s - \boldsymbol{d}/Z$ built from the raw numerator, and the PDMD target $\boldsymbol{x}_s - \boldsymbol{d}^{\perp}/Z$ built from the projected one. The two targets share the normalizer Z and the same student step; the projection is their only difference. Under the 125k-token packing each rank holds one sample, and Figure E1 shows twelve of the sixteen ranks of that step in rank order.

The DMD target carries high-frequency speckle that the student endpoint does not have. The PDMD target keeps the same subject, colors, and frame layout, and carries visibly less of that speckle. The projection removes between 5% and 30% of the video update norm across the 16 samples at this step (median 16%); the visual comparison associates the removed component with speckle artifacts.

E.2 Saturation Metrics

Table E1 lists two color statistics of the models in Tables 1 and 2. Chroma is the mean CIELAB $C^* = \sqrt{a^{*2} + b^{*2}}$ over pixels. Saturation is the mean C^* divided by the mean lightness L^*. Both are averaged over frames and clips.

PDMD has lower mean saturation than DMD on both backbones. On Wan2.1, DMD† sits at 0.709 saturation against 0.587 for the 50-step teacher, and PDMD at 0.476; on MiniMax-H3, DMD† and DMD2† sit at 0.460 and 0.520 against 0.405, and PDMD at 0.357. rCM and AnyFlow sit above the teacher on both color statistics for both backbones. These aggregate statistics complement the visual comparisons and user studies; lower saturation alone does not establish better perceptual quality.

Table E1: **Color statistics of the models in Tables 1 and 2.** Wan2.1 values are means over the 4,720 clips of the VBench protocol in Section C.1; MiniMax-H3 values are means over the 387 VideoGen-Eval clips of Section C.2.

Method	Saturation (C^*/L^*)	Chroma C^*
Wan2.1-T2V-1.3B		
Wan2.1-T2V-1.3B, 50 steps	0.587	21.5
Wan2.1-T2V-1.3B, 4 steps	1.233	27.6
rCM	0.633	23.6
AnyFlow	0.607	25.9
DMD†	0.709	23.5
DMD2†	0.437	17.9
PDMD (ours)	0.476	17.9
MiniMax-H3-33B		
MiniMax-H3-33B, 50 steps	0.405	15.0
MiniMax-H3-33B, 4 steps	0.669	17.7
H3 Turbo LoRA	0.548	17.2
rCM†	0.440	15.2
AnyFlow†	0.437	15.3
DMD†	0.460	16.6
DMD2†	0.520	18.3
PDMD (ours)	0.357	14.0

E.3 Motion Comparisons

Beyond the dynamic-degree scores in Tables 1 and 2, this section shows more examples of the motion in our distilled results. Figure E2 shows Wan2.1 prompt 119 with eighteen uniformly spaced frames per method, one column per method. Figures E3 and E4 show MiniMax-H3 prompts 857 and 895 with twelve and ten uniformly spaced frames per method, one row per method. Figure D5 adds the two-step runs of Section D.6 on prompt 857, below the four-step rows of DMD † and PDMD.

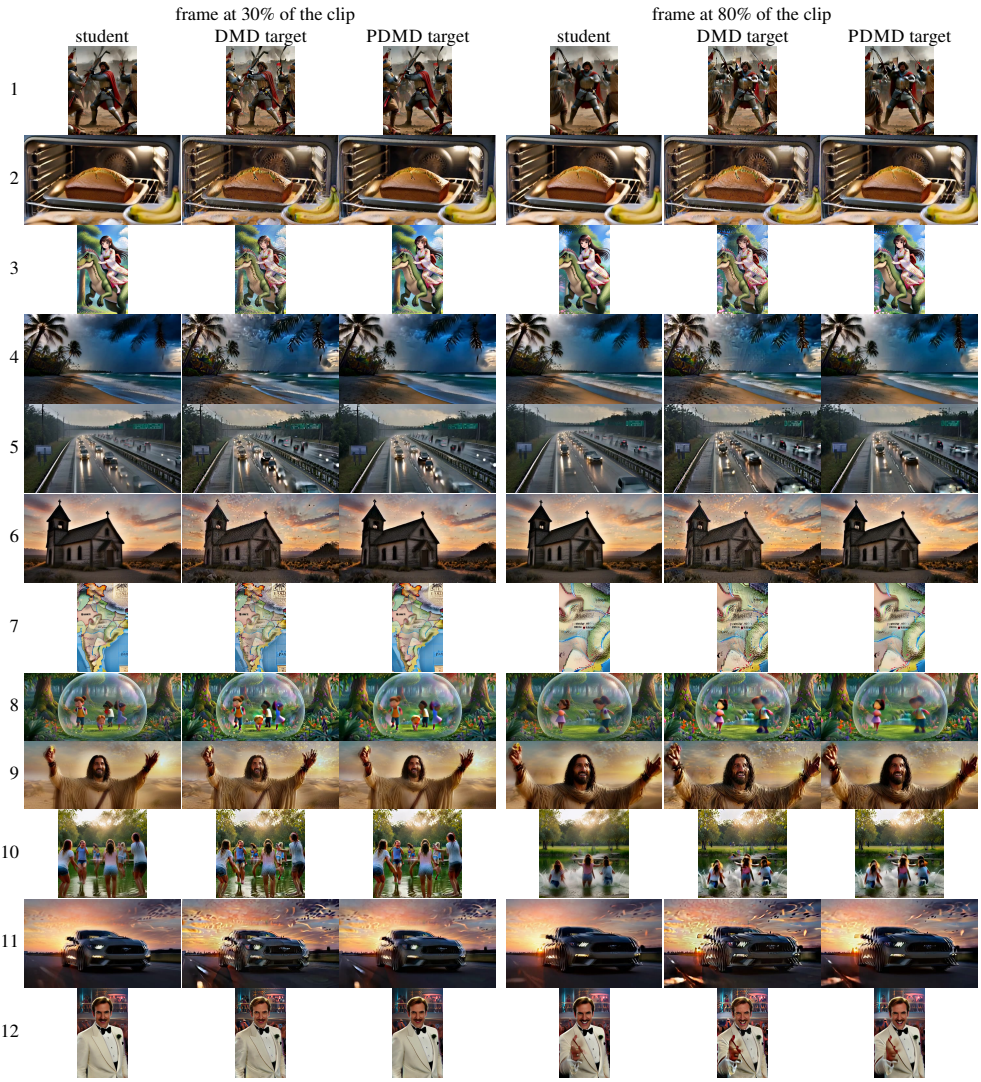

Figure E1: **One student step, before and after the projection.** Results are from the training iteration-500 of the 4-NFE MiniMax-H3 PDMD run in Section 6. Each row is one of the data-parallel ranks of the same student step, decoded three ways: the student endpoint $\boldsymbol{x}_s$, the DMD target $\boldsymbol{x}_s - \boldsymbol{d}/Z$ built from the raw numerator, and the PDMD target $\boldsymbol{x}_s - \boldsymbol{d}^{\perp}/Z$ built from the projected one. Under the 125k-token packing each rank holds one sample; the rows are twelve of the sixteen ranks of that step, in rank order. Left: the frame at 30% of the clip. Right: the frame at 80%. The DMD target carries high-frequency speckle, but the PDMD target retains the same subject, colors, and layout while exhibiting less speckle. This comparison associates the removed component with visible artifacts.

The three scenes show three recurring failures of the baselines. Motion is lost: in 895 the distilled baselines except AnyFlow† are almost static while PDMD follows the teacher. Motion is misread: in 119 every baseline except ADV turns the rider's upper body around. Motion is blurred: Turbo LoRA, rCM† and AnyFlow† blur the slowly moving bouquet in 857 into translucent ghosting. Several baselines also carry a color cast (119) or a smeared, shifted tone (DMD† and DMD2† in 857). PDMD avoids these visible failures in the three scenes.

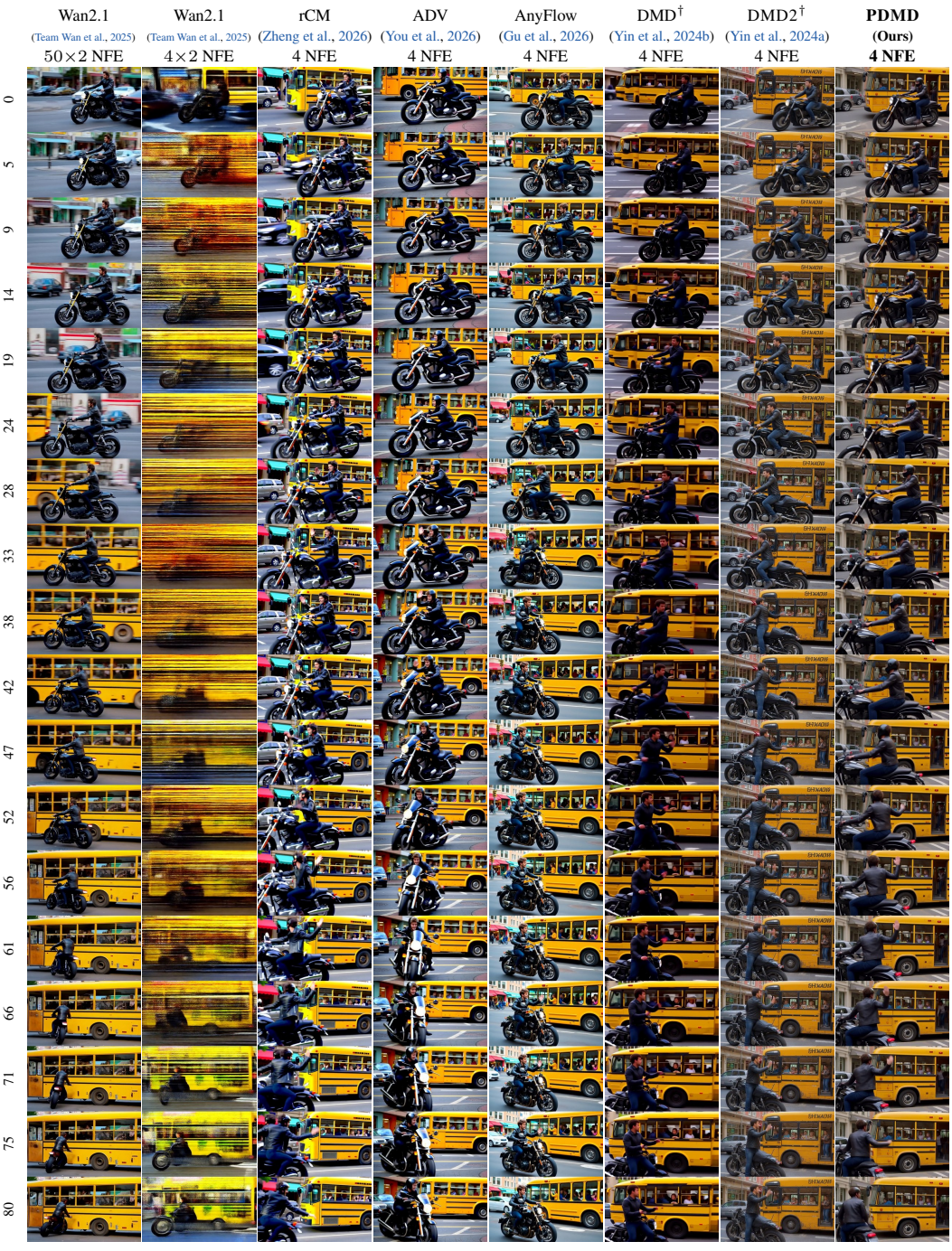

Figure E2: **Motion on four-step Wan2.1-T2V-1.3B, VBench prompt 119.** Each column is one method from Table 1; the eighteen rows are uniformly spaced frames of the 81-frame clip, top to bottom, labelled by frame index, with the same prompt and seed across methods. Besides the visible color cast of the other methods, every baseline except ADV misreads the rider's motion and turns his upper body by 180 degrees.



Figure E3: **Motion on four-step MiniMax-H3, VideoGen-Eval prompt 857.** Each row is one method from Table 2; the twelve frames are uniformly spaced over the 124-frame clip, left to right, labelled by frame index, with the same prompt and seed across methods. Turbo LoRA, rCM† and AnyFlow† blur the slowly moving bouquet into translucent ghosting. DMD† and DMD2† exhibit shifted colors and smeared textures.

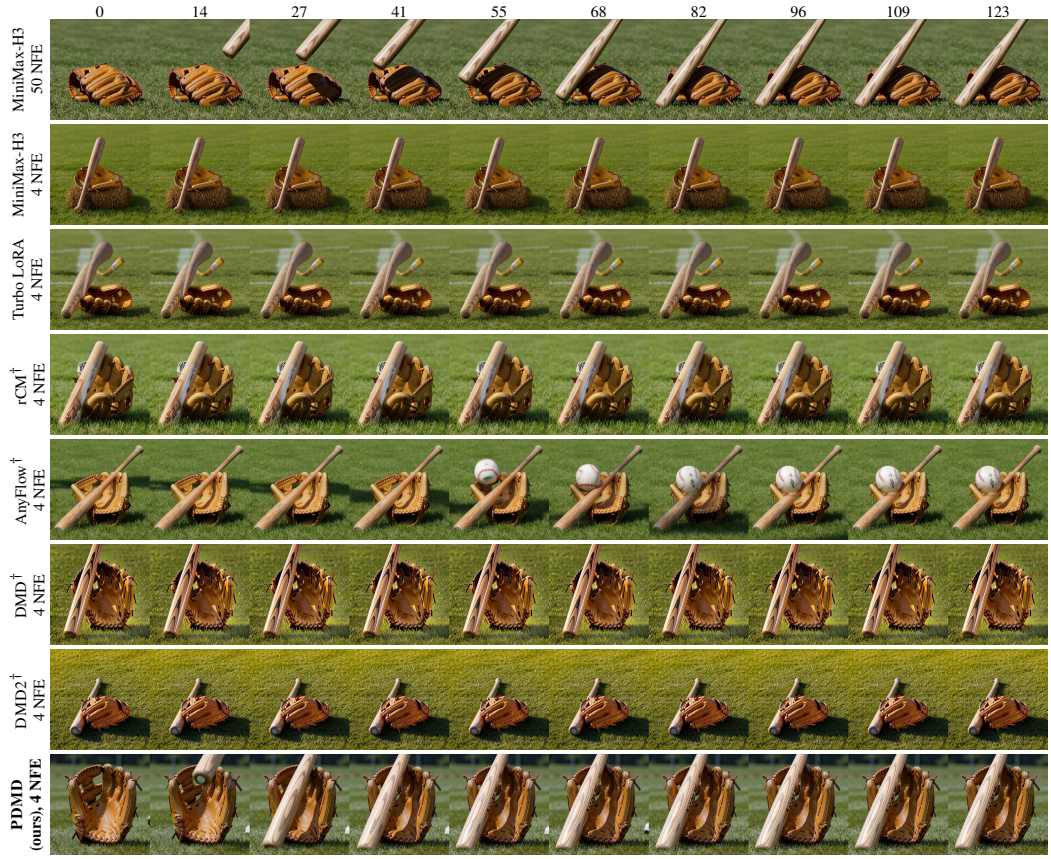

Figure E4: **Motion on four-step MiniMax-H3, VideoGen-Eval prompt 895.** Each row is one method from Table 2; the ten frames are uniformly spaced over the 124-frame clip, left to right, labelled by frame index, with the same prompt and seed across methods. Except for AnyFlow†, the distilled baselines render an almost completely static video. PDMD shows the baseball bat entering the frame, similar to the teacher.

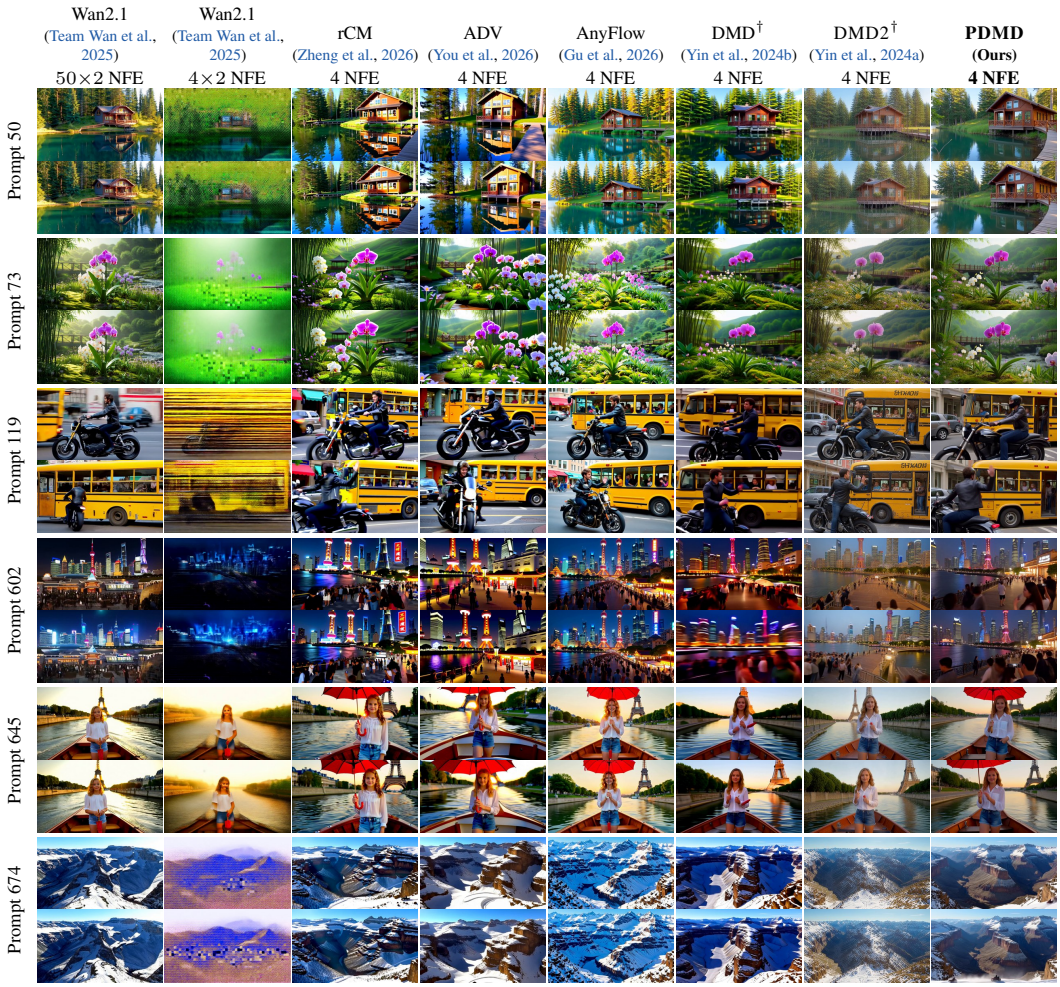

Figure E5: **Additional qualitative results for four-step Wan2.1-T2V-1.3B.** Each block shows an early and a later frame from one video under the same layout, methods, protocol, and seed as Figure 6; † marks our implementations (Section C.1).

E.4 Additional Samples

Figure E5 shows more prompts for the methods of Table 1 on Wan2.1-T2V-1.3B. Figure E6 compares the eight models in Table 2 on eight of the 387 VideoGen-Eval prompts, with two frames from every clip over two pages; Figure 2 instead compares the methods over training.

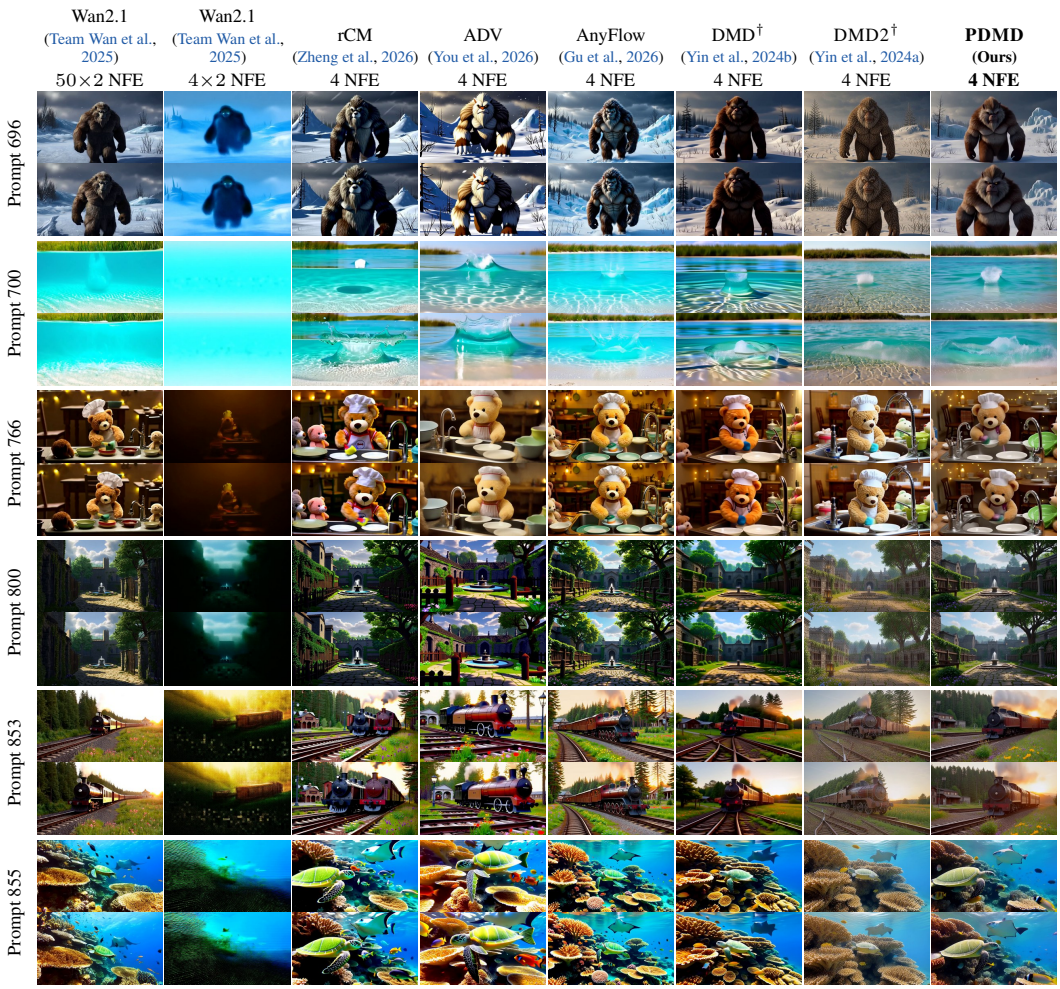

Figure E5: **Additional qualitative results for four-step Wan2.1-T2V-1.3B (continued).** The remaining prompts use the same methods, frame positions, protocol, and seed as the preceding page.

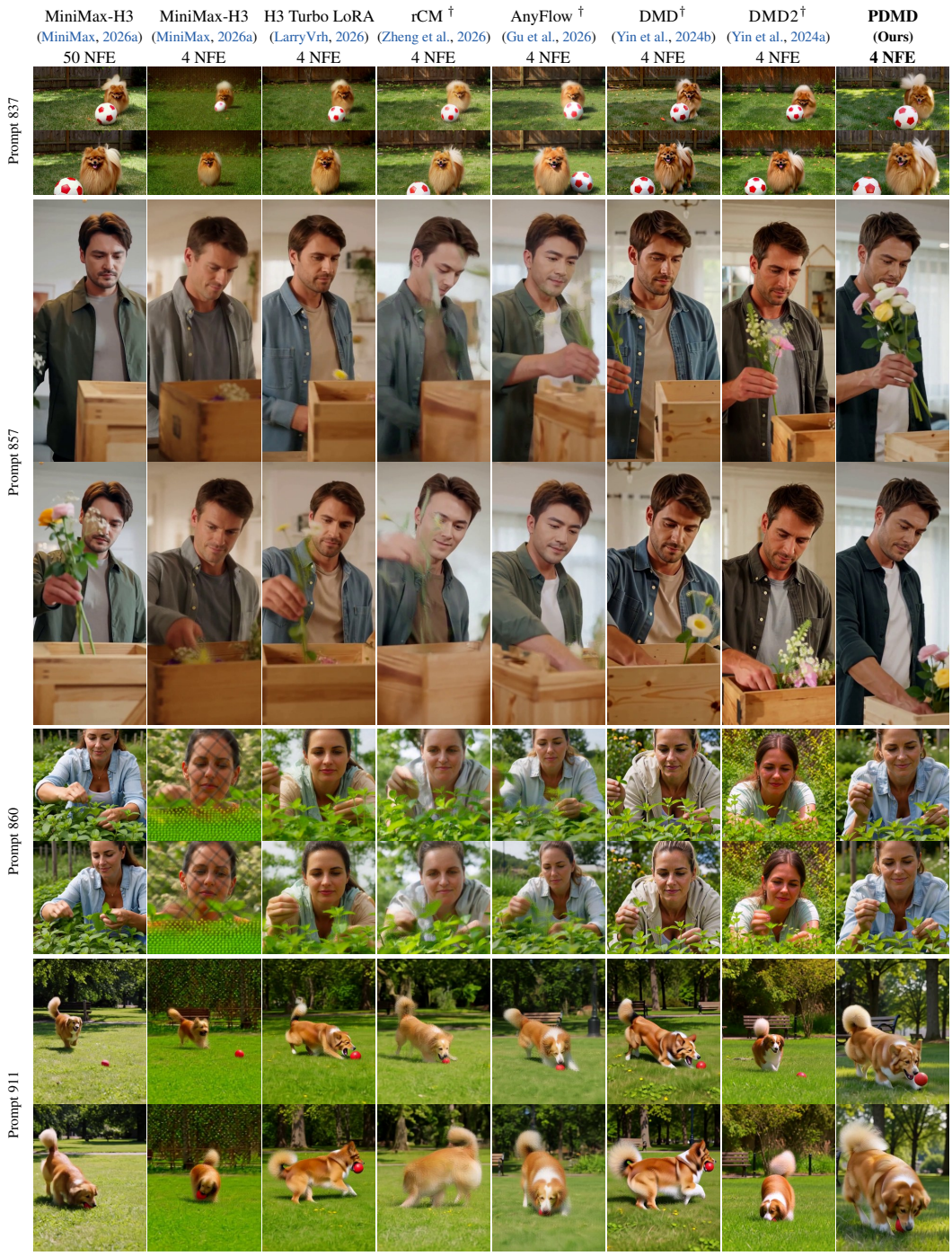

Figure E6: **Qualitative results on four-step MiniMax-H3 (part 1 of 2).** The eight models of Table 2 on the first four of the 8 prompts, two frames of every clip per prompt with the earlier frame above. Every clip is 124 frames, and rows keep the benchmark's own aspect ratio, so their heights differ. † marks our reimplementations.



Figure E7: **Qualitative results on four-step MiniMax-H3 (part 2 of 2).** Four more prompts, same models, same layout and the same two-frame rule as Figure E6.

F Limitations and Future Work

PDMD does not yet yield satisfactory single-step video generation in our experiments. At 1 NFE, samples remain smeared and fall below the quality of the four-step results (Figure 11). Improving single-step quality may require an additional objective, such as a discriminator loss; we leave this investigation to future work. As in graphics, where a biased estimator with smaller variance often replaces an unbiased one, using a biased critic error estimate with smaller deviation is another possible improvement (Section A.2). Whether the residual direction remains the right one to be removed when a single evaluation must cover the whole video generation is left to future work.

The theory guarantees conditional error removal, not convergence of the coupled student–critic optimization. A nonvanishing error-removal fraction requires the critic error to remain appreciable relative to conditional endpoint variance, and signal retention requires weak alignment with the residual. If the useful signal aligns with the residual, the projection can remove it as well; when the critic is exact, it can still discard useful signal (Equation (A35)). Although these assumptions cannot be verified directly on the video models, the decoded PDMD target carries visibly less high-frequency speckle than the DMD target (Figure E1).

Our video experiments cover two backbones, primarily at 4 NFE, and do not establish preservation of diversity across prompts and seeds. On MiniMax-H3, the 50-step teacher remains preferred in the user study despite PDMD's higher aggregate video score (Table 2), indicating a remaining perceptual-quality gap and a limitation of the automatic metrics.