HORIZON DISTRIBUTION MATCHING FOR AUTOREGRESSIVE VIDEO DISTILLATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Autoregressive video diffusion offers a scalable route to streaming generation, but maintaining video quality under few-step distillation remains challenging. On-policy learning reduces the discrepancy between training and inference histories, yet leaves open which temporal distributions should be matched along the generated trajectory. For example, full-video matching constrains the sequence distribution, whereas isolated window matching may omit dependencies on preceding history. To study this choice, we introduce *HorizonDMD*, a generalized formulation of on-policy distribution matching over history–horizon views. Each view is specified by its conditioning history and horizon extent, encompassing zero-history global views and local views conditioned on student-generated histories. Full-video matching is a special case. This formulation compares the distributions of noisy horizons conditioned on noisy histories, allowing supervision to be allocated across history depths and temporal extents. Under consistent student and reference marginals, we show that this objective decomposes into a coverage-weighted sum of autoregressive conditional divergences. Consequently, positive coverage preserves the zero-divergence target of full-sequence matching, including identification of the clean sequence law under nondegenerate Gaussian corruption. We further show that prefix matching and conditional horizon matching yield the same gradient when the clean-history distribution is held fixed, providing a basis for horizon-restricted distillation. We also show that prefix and restricted full-video score discrepancies need not agree, so the choice of temporal view can change the distributional guidance.

1 INTRODUCTION

Diffusion models have substantially advanced video generation, enabling high-quality synthesis with coherent appearance and temporal dynamics (Ho et al., 2022; Blattmann et al., 2023). However, bidirectional video generators process the entire clip at once, limiting their suitability for streaming and interactive applications (Yin et al., 2025). Autoregressive video diffusion instead generates frames or blocks sequentially while retaining diffusion modeling within each block (Gao et al., 2025; Yin et al., 2025). It still requires iterative denoising, however, which motivates few-step distillation for low-latency synthesis.

Distribution Matching Distillation (DMD) aligns a few-step student's generated distribution with a pretrained reference distribution (Yin et al., 2024b;a). For autoregressive students, Self Forcing learns from student-generated trajectories, reducing the discrepancy between training and inference histories (Huang et al., 2025). Causal-rCM incorporates this on-policy refinement together with teacher-forcing consistency distillation (Zheng et al., 2026). In both cases, learning from self-generated histories addresses the conditioning context, while the temporal scope of distribution matching remains a separate design choice.

Recent work explores window-based supervision for extended rollouts (Cui et al., 2025; Yang et al., 2025) and history-conditioned or causally aligned distillation (Chen et al., 2026b; Bandyopadhyay et al., 2026). Full-video matching constrains the sequence distribution, including its temporal dependencies, while agreement between isolated window marginals need not identify those dependencies. This raises a common question: *which history-conditioned distributions should be matched, and when does their agreement identify the complete sequence law?*

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.

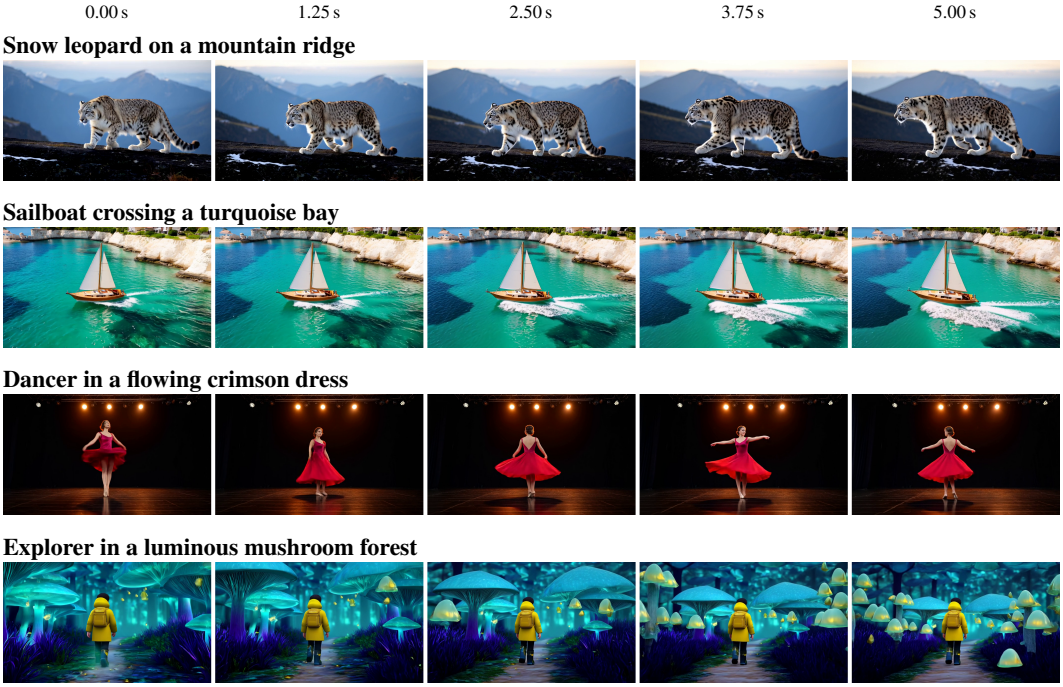

Figure 1: Selected text-to-video results generated autoregressively by HorizonDMD. Each row shows one generated video, with frames sampled at 0, 1.25, 2.50, 3.75, and 5.00 seconds from left to right.

To address this question, we introduce *Horizon Distribution Matching* (HorizonDMD), which formulates on-policy distillation over a family of history–horizon views. Each view specifies both the preceding history and the contiguous horizon whose conditional distribution is matched. An empty history defines a *global view*; a nonempty history defines a *local, history-conditioned view*. Both retain external conditioning, such as a text prompt. This distinction concerns conditioning rather than duration, with full-video matching as the zero-history, full-length special case. A distribution over these views then controls how conditional discrepancies contribute to the matching objective. To connect this objective to DMD training, we analyze the scores of noise-corrupted prefixes containing the history and horizon. Restricting full-video scores to the horizon does not generally recover the corresponding prefix score coordinates. The temporal extent used for scoring can therefore change the horizon update. Autoregressive inference remains unchanged.

Our contributions are threefold. First, we introduce a unified formulation of global and history-conditioned distribution matching for autoregressive distillation. Then, we establish a conditional-coverage characterization of when the idealized objective preserves sequence-level identification in clean and noise-corrupted domains. Finally, we further establish gradient equivalence between prefix matching and conditional horizon matching when the clean-history distribution is held fixed, and characterize the non-equivalence of prefix and restricted full-video scores, providing a conditional foundation for horizon-restricted DMD.

2 Related Work

Diffusion distillation. Progressive distillation compresses sampling trajectories (Salimans & Ho, 2022), while consistency models learn trajectory-consistent mappings (Song et al., 2023). At the distribution level, DMD matches student and reference distributions through their scores (Yin et al., 2024b); DMD2 improves estimation of the evolving student distribution and adds adversarial supervision (Yin et al., 2024a). Phased DMD extends this approach to diffusion-time subintervals (Fan et al., 2025), distinct from the video-time horizons considered here. Further extensions address score estimation (Wu et al., 2026), relational regularization (Zhang et al., 2026), and reward-weighted matching (Lu et al., 2026). These methods develop distillation objectives and estimators; HorizonDMD studies the temporal distributions to which they apply.

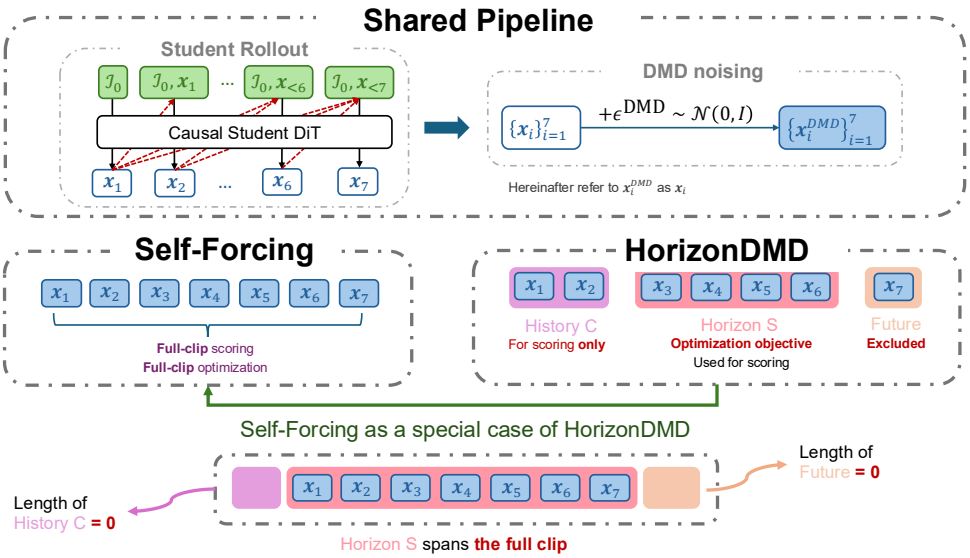

Figure 2: Comparison of full-video DMD and HorizonDMD. Both learn from student-generated rollouts. Full-video DMD uses the entire video for scoring and generator supervision, whereas HorizonDMD scores a history–horizon prefix and restricts the generator loss to the horizon. The full-video setting corresponds to an empty history and a horizon spanning the entire video.

Autoregressive and on-policy video generation. CausVid distills a bidirectional teacher into an autoregressive student (Yin et al., 2025). Flexible sequence noising and history conditioning are explored by Diffusion Forcing and History-Guided Video Diffusion (Chen et al., 2024; Song et al., 2025). For student initialization, Causal Forcing uses causal ODE supervision (Zhu et al., 2026), whereas Causal Forcing++ uses online consistency distillation (Zhao et al., 2026). Both then apply Self Forcing's DMD, which trains on student-generated trajectories (Huang et al., 2025). Causal-rCM combines this on-policy refinement with teacher-forcing consistency distillation (Zheng et al., 2026). For longer rollouts, Self-Forcing++ and LongLive extend on-policy learning using shorter supervision windows (Cui et al., 2025; Yang et al., 2025).

Conditional distillation and temporal scope. Context Forcing formulates conditional DMD on student-generated histories, relates it to joint distribution matching through a KL decomposition, and uses a long-context teacher with a growing rollout curriculum (Chen et al., 2026b). Focusing on teacher causality, Context-Matched Distillation aligns teacher scoring with each target's available history and controls through a causal teacher (Bandyopadhyay et al., 2026). HorizonDMD instead considers a distribution of global and history-conditioned views, allowing bidirectional scoring within each selected prefix. Our focus is the conditional coverage and gradient properties of this family of matching objectives. Complementary research directions are discussed in Appendix A.

3 Method

We begin by reviewing diffusion-based distribution matching and on-policy autoregressive generation. We then formulate HorizonDMD over history–horizon views of student-generated videos and analyze its conditional coverage and gradient properties. Finally, we introduce the learning objectives for horizon-restricted distillation.

3.1 Diffusion and distribution matching

Given a video represented in latent space as $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$, the forward diffusion process adds Gaussian noise :

$$\boldsymbol{x}^{(\tau)} = \alpha_\tau \boldsymbol{x} + \sigma_\tau \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}). \tag{1}$$

Here $\boldsymbol{x}_i$ denotes the ith of N consecutive temporal blocks, while $\boldsymbol{x}^{(\tau)}$ indicates the video at diffusion noise level τ. The coefficients $(\alpha_\tau, \sigma_\tau)$ specify the noise schedule, and $\boldsymbol{\epsilon}$ is independent of $\boldsymbol{x}$.

Let q_θ and p denote the student and fixed reference distributions. At noise level τ, their scores are

$$\boldsymbol{s}_r^{\text{full}}(\boldsymbol{x}^{(\tau)}) = \nabla_{\boldsymbol{x}^{(\tau)}} \log r^{(\tau)}(\boldsymbol{x}^{(\tau)}). \tag{2}$$

Here $r^{(\tau)}$ denotes the distribution obtained by applying the Gaussian corruption to $r \in \{q_\theta, p\}$. The student and reference distributions are conditioned on the same text prompt y, which we omit from the notation for brevity.

Distribution Matching Distillation (DMD) is motivated by an ideal reverse-KL objective between these noise-corrupted distributions (Yin et al., 2024b;a):

$$\mathcal{L}_{\text{DMD}}(\theta) = \boldsymbol{E}_{\tau\sim\nu}\left[\text{KL}(q_\theta^{(\tau)} \| p^{(\tau)})\right], \tag{3}$$

where $\boldsymbol{E}$ denotes expectation and ν is a fixed distribution over noise levels with $\alpha_\tau, \sigma_\tau > 0$. For differentiably reparameterized student samples, the gradient is determined by the difference between the two scores:

$$\nabla_\theta \mathcal{L}_{\text{DMD}} = \boldsymbol{E}\left[\alpha_\tau \left(\frac{\partial \boldsymbol{x}}{\partial \theta}\right)^{\mathsf{T}} \left(\boldsymbol{s}_{q_\theta}^{\text{full}} - \boldsymbol{s}_p^{\text{full}}\right)\right]. \tag{4}$$

Both scores are evaluated at $\boldsymbol{x}^{(\tau)}$, and the expectation averages over student sampling, corruption noise, and τ. This identity uses the full sampling derivative and assumes parameter-independent randomness and valid interchange of differentiation and expectation. To estimate the score difference, DMD uses a pretrained teacher for the reference score and trains an auxiliary model ϕ on student samples to estimate the student score.

3.2 On-policy autoregressive generation

For an autoregressive student, the generated sequence distribution factorizes as

$$q_\theta(\boldsymbol{x}) = \prod_{i=1}^{N} q_\theta(\boldsymbol{x}_i \mid \boldsymbol{x}_{<i}). \tag{5}$$

A few-step conditional generator G_θ realizes these factors by recursively sampling

$$\boldsymbol{x}_i = G_\theta(\boldsymbol{\eta}_i; \boldsymbol{x}_{<i}), \qquad i = 1, \ldots, N, \qquad \boldsymbol{x}_{<1} = \varnothing. \tag{6}$$

Here $\boldsymbol{\eta}_i$ collects the parameter-independent sampling randomness for block i, and G_θ generates the block through a few denoising steps conditioned on the preceding outputs.

Teacher forcing supplies ground-truth histories during training. Self Forcing instead trains on student-generated autoregressive rollouts. In its DMD formulation, the resulting video distribution is matched to the reference. On-policy sampling determines the histories encountered during training, but does not prescribe the temporal scope of distribution matching. Within this setting, HorizonDMD extends full-video matching to conditional distributions over history–horizon views.

3.3 Horizon distribution matching

We describe each matching view by its conditioning history and the temporal extent of the horizon being matched. A view (h, k) contains h history blocks followed by k horizon blocks, with $0 \le h < N$ and $1 \le k \le N - h$. We denote its history, horizon, and visible prefix by

$$\boldsymbol{c} = \boldsymbol{x}_{1:h}, \qquad \boldsymbol{z} = \boldsymbol{x}_{h+1:h+k}, \qquad \boldsymbol{v} = (\boldsymbol{c}, \boldsymbol{z}). \tag{7}$$

The endpoint $h+k$ determines the visible prefix, while h determines the history on which the horizon is conditioned. Different history–horizon splits can therefore share the same prefix.

An empty history ($h = 0$) defines a *global view*, whereas a nonempty history ($h > 0$) defines a *local, history-conditioned view*. These terms describe conditioning, not duration: $(0, N)$ gives full-video matching, but shorter zero-history horizons are also global views. Both types retain the external condition y.

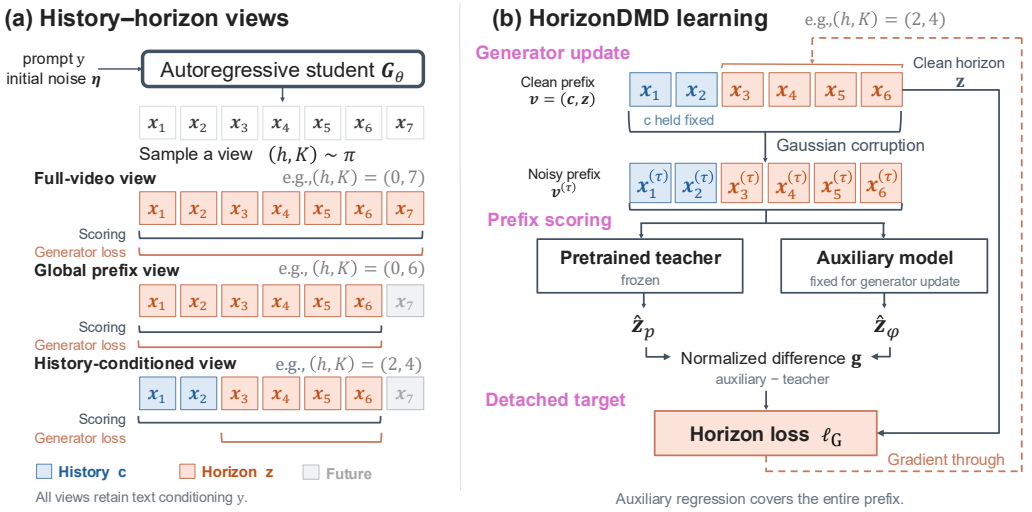

Figure 3: Overview of HorizonDMD. (a) We sample different views of a student-generated video by choosing the history length and horizon extent. A view may cover the full video, a prefix without preceding history, or a horizon conditioned on earlier outputs. (b) For each sampled view, the teacher and auxiliary model evaluate the same noisy prefix. Their denoised predictions are compared over the horizon to train the student, with the clean history held fixed during the update. The auxiliary model is trained on both history and horizon.

Under the Gaussian channel in Eq. (1), the student and reference marginals on $\boldsymbol{v}^{(\tau)} = (\boldsymbol{c}^{(\tau)}, \boldsymbol{z}^{(\tau)})$ induce conditional distributions over the noisy horizon given the noisy history. We compare these conditionals under histories drawn from the student's corrupted distribution:

$$\mathcal{D}^{(\tau)}_{h,k}(\theta) = \boldsymbol{E}_{\boldsymbol{c}^{(\tau)} \sim q^{(\tau)}_\theta} \left[\mathrm{KL}\Big(q^{(\tau)}_\theta(\boldsymbol{z}^{(\tau)} \mid \boldsymbol{c}^{(\tau)}) \,\|\, p^{(\tau)}(\boldsymbol{z}^{(\tau)} \mid \boldsymbol{c}^{(\tau)}) \Big) \right]. \tag{8}$$

Matching therefore conditions on $\boldsymbol{c}^{(\tau)}$, whereas autoregressive generation uses the clean history $\boldsymbol{c}$. For $h = 0$, the history condition is absent.

Let π be a fixed distribution over admissible views. Averaging their divergences over $(h, k) \sim \pi$ and the independently sampled noise level $\tau \sim \nu$ gives the population objective

$$\mathcal{R}_\pi(\theta) = \boldsymbol{E}_{\tau \sim \nu,\, (h,k) \sim \pi} \left[\frac{\omega(k)}{k} \mathcal{D}^{(\tau)}_{h,k}(\theta) \right]. \tag{9}$$

Here $\omega(k) \geq 0$ is a fixed length weight, and division by k measures divergence per block. If π assigns all mass to $(0, N)$, the objective reduces to $\mathcal{R}_\pi = \omega(N)\mathcal{L}_{\mathrm{DMD}}/N$.

3.4 Theoretical Characterization

We first ask when matching history–horizon views is sufficient to identify the full sequence distribution. The answer depends on the weight assigned to each autoregressive conditional. At a fixed noise level τ, write $\mathcal{R}^{(\tau)}_\pi$ for the objective averaged over views, and let $d^{(\tau)}_i = \mathcal{D}^{(\tau)}_{i-1,1}$ denote the single-block conditional divergence at position i. Its effective coverage is

$$\lambda_i = \boldsymbol{E}_{(h,k) \sim \pi} \left[\frac{\omega(k)}{k} \mathbf{1}\{h < i \leq h + k\} \right]. \tag{10}$$

The indicator selects views whose horizon contains position i.

Theorem 1 (Conditional horizon coverage). *Fix τ and assume finite $\mathrm{KL}(q^{(\tau)}_\theta \| p^{(\tau)})$. If every view uses the corresponding marginals and conditionals of these same two sequence laws, then*

$$\mathcal{R}^{(\tau)}_\pi = \sum_{i=1}^{N} \lambda_i d^{(\tau)}_i. \tag{11}$$

With $\lambda_{\min} = \min_i \lambda_i$ *and* $\lambda_{\max} = \max_i \lambda_i$*,*

$$\lambda_{\min} \mathrm{KL}(q_\theta^{(\tau)} \| p^{(\tau)}) \le \mathcal{R}_\pi^{(\tau)} \le \lambda_{\max} \mathrm{KL}(q_\theta^{(\tau)} \| p^{(\tau)}). \tag{12}$$

If $\lambda_{\min} > 0$*, zero horizon risk is equivalent to* $q_\theta^{(\tau)} = p^{(\tau)}$*. Under the Gaussian channel with* $\alpha_\tau \neq 0$*, this is also equivalent to* $q_\theta = p$*. The full proof is provided in Appendix B.*

Positive coverage therefore allows sequence identification from different families of views. Full-video matching covers every conditional at once, while the family of all single-block horizons covers them separately under their complete histories. The first block must still receive positive weight from a global view. Complete-history conditioning is essential: agreement between isolated window marginals need not identify the sequence law (Appendix B.5). Identification also holds after averaging over the prescribed noise levels, and for the clean-channel objective when its KL is finite. These results concern ideal distributional constraints, not convergence of their learned approximations.

The coverage coefficients also determine the relative weights of the conditional divergences. When all λ_i share a positive value, the objective is proportional to full-sequence KL. More generally, distributions over views with identical coverage coefficients induce the same population risk and, wherever differentiable, the same total parameter gradient.

The coverage result characterizes the population objective. We next relate prefix matching to conditional horizon matching while holding the clean-history distribution fixed. For a view (h, k) at noise level τ, the exact prefix score of $r \in \{p, q_\theta\}$ is

$$\boldsymbol{s}_r(\boldsymbol{v}^{(\tau)}) = \nabla_{\boldsymbol{v}^{(\tau)}} \log r^{(\tau)}(\boldsymbol{v}^{(\tau)}), \tag{13}$$

and its horizon coordinates give the conditional score:

$$[\boldsymbol{s}_r(\boldsymbol{v}^{(\tau)})]_z = \nabla_{\boldsymbol{z}^{(\tau)}} \log r^{(\tau)}(\boldsymbol{z}^{(\tau)} \mid \boldsymbol{c}^{(\tau)}). \tag{14}$$

Here $[\cdot]_z$ selects the horizon coordinates.

To isolate changes in the conditional generator, fix a current parameter value θ_0 and keep the clean-history distribution $q_{\theta_0}(\boldsymbol{c})$ fixed while varying $q_\theta(\boldsymbol{z} \mid \boldsymbol{c})$. We denote the resulting derivative at θ_0 by $\nabla_\theta^{\mathrm{fr}}$. Let $\boldsymbol{\delta} = [\boldsymbol{s}_{q_\theta}(\boldsymbol{v}^{(\tau)}) - \boldsymbol{s}_p(\boldsymbol{v}^{(\tau)})]_z$ be the horizon score difference and $\boldsymbol{J}_\theta = \partial \boldsymbol{z} / \partial \theta|_{\boldsymbol{c}}$ the generator Jacobian with the clean history held fixed.

Proposition 1 (Prefix–conditional gradient equivalence)**.** *Fix the reference law and assume that the relevant KL divergences under the frozen-history construction are finite and differentiable in a neighborhood of* θ_0*. At the current parameter value,*

$$\nabla_\theta^{\mathrm{fr}} \mathrm{KL}\Big(q_\theta^{(\tau)}(\boldsymbol{v}^{(\tau)}) \,\|\, p^{(\tau)}(\boldsymbol{v}^{(\tau)})\Big) = \nabla_\theta^{\mathrm{fr}} \mathcal{D}_{h,k}^{(\tau)}. \tag{15}$$

With exact scores, differentiable reparameterized conditional generation, and justified interchange of differentiation and expectation, their common per-block gradient is

$$\nabla_\theta^{\mathrm{fr}} \left[\frac{\mathcal{D}_{h,k}^{(\tau)}}{k} \right] = \frac{\alpha_\tau}{k} \boldsymbol{E}\left[\boldsymbol{J}_\theta^{\mathsf{T}} \boldsymbol{\delta}\right]. \tag{16}$$

The full proof is provided in Appendix B.2.

The history-marginal KL is constant under this construction, so prefix matching and conditional horizon matching have the same derivative. Equation (16) is the corresponding conditional form of the DMD gradient, with expectation over the clean history, conditional rollout, and corruption noise. This equivalence does not generally extend to total on-policy gradients when the history distribution also varies. Across view distributions, equal coverage likewise need not imply equal frozen-history derivatives, because these derivatives depend on the conditioning boundaries (Appendix B.1).

The preceding equivalence concerns a fixed prefix. A shorter prefix requires marginalizing the remaining blocks, which is not generally equivalent to restricting the coordinates of the full-sequence score. Let $\boldsymbol{u} = \boldsymbol{x}_{h+k+1:N}$ be the excluded blocks, and recall the full-sequence score in Eq. (2).

Proposition 2 (Scoring-scope non-equivalence)**.** *For smooth positive densities permitting differentiation under marginalization, the horizon coordinates of the prefix score satisfy*

$$[\boldsymbol{s}_r(\boldsymbol{v}^{(\tau)})]_z = \boldsymbol{E}_{\boldsymbol{u}^{(\tau)} \sim r^{(\tau)}(\cdot \mid \boldsymbol{v}^{(\tau)})} \left[[\boldsymbol{s}_r^{\mathrm{full}}(\boldsymbol{v}^{(\tau)}, \boldsymbol{u}^{(\tau)})]_z \right]. \tag{17}$$

Consequently, the prefix score difference is not generally recovered by restricting full-video score differences to the horizon, even when they are averaged over the student's excluded-block conditional law. The proof and a counterexample are provided in Appendix B.4.

The marginalization law in Eq. (17) is specific to the distribution being scored. Student and reference prefixes therefore require different conditional averages in general. Changing the endpoint $h + k$ can change the score discrepancy itself. By contrast, at a fixed endpoint and the same noised prefix, varying h changes the conditional horizon and its score coordinates, not the prefix score. The proposition establishes distinct supervision, not a universal preference for either scoring view.

3.5 Learning objectives

Following DMD, we use a frozen pretrained teacher to estimate the reference prefix score and train an auxiliary model to estimate the student prefix score. Both models evaluate the same noisy prefix, allowing bidirectional dependence within it while excluding subsequent blocks. The resulting generator supervision is restricted to the horizon.

Under Gaussian corruption, the difference between exact posterior means is proportional to the score discrepancy $\boldsymbol{\delta}$ (Appendix B.3). This identity motivates comparing the teacher and auxiliary model through their denoised horizon predictions, $\widehat{\boldsymbol{z}}_p$ and $\widehat{\boldsymbol{z}}_\phi$, given $\boldsymbol{v}^{(\tau)}$. We use the regression surrogate

$$\ell_G = \frac{\omega(k)}{2n_z} \left\| \boldsymbol{z} - \mathrm{sg}(\boldsymbol{z} - \boldsymbol{g}) \right\|_2^2, \tag{18}$$

where $\boldsymbol{g} = (\widehat{\boldsymbol{z}}_\phi - \widehat{\boldsymbol{z}}_p)/a$ is the normalized prediction difference, $a > 0$ is computed over the horizon as specified in Appendix B.6, and n_z counts the scalar coordinates of the horizon. The operator sg holds the regression target fixed during differentiation, giving

$$\nabla_{\boldsymbol{z}} \ell_G = \frac{\omega(k)}{n_z} \boldsymbol{g}. \tag{19}$$

For equal-sized blocks, choosing $\omega(k) = k/\boldsymbol{E}_\pi[k]$ makes the prefactor $\omega(k)/n_z$ independent of the horizon length.

The auxiliary model estimates the student prefix score, so its denoising objective includes both history and horizon. On the linear flow path $\alpha_\tau = 1 - \sigma_\tau$, we train its velocity predictor with

$$\mathcal{L}_\phi = \boldsymbol{E}\left[\frac{1}{n_v} \left\| F_\phi(\boldsymbol{v}^{(\tau)}, \tau) - (\boldsymbol{\epsilon}_v - \boldsymbol{v}) \right\|_2^2 \right], \tag{20}$$

where F_ϕ predicts velocity with respect to σ_τ, n_v counts the scalar coordinates of the prefix, and $\boldsymbol{\epsilon}_v$ is Gaussian noise. The expectation is over sampled views, student rollouts, noise levels, and corruption noise. The student distribution is held fixed during auxiliary regression. On this path, the clean-prefix prediction before guidance is $\boldsymbol{v}^{(\tau)} - \sigma_\tau F_\phi(\boldsymbol{v}^{(\tau)}, \tau)$; selecting its horizon coordinates gives the corresponding horizon prediction.

The gradient identity in Eq. (19) is exact for the detached regression target. It does not, however, equate the practical parameter update with the ideal conditional KL gradient: learned or guided predictions, normalization, and gradient truncation can alter that update (Appendix B.6). The estimated prefix scores also need not correspond to consistent marginals of a common sequence distribution, as required by the coverage analysis. Algorithm 1 summarizes the resulting alternating training procedure.

4 Experiments

4.1 Quantitative comparison

Comparison with SF DMD. Both SF DMD and HorizonDMD learn from student-generated autoregressive trajectories. SF DMD performs distribution matching on the full generated clip, corresponding to the view $(h, k) = (0, N)$. HorizonDMD generalizes this formulation to sampled history–horizon views. For each view, the teacher and auxiliary model evaluate the jointly corrupted

Algorithm 1 HorizonDMD: single-view alternating training

Require: Student G_θ, auxiliary flow predictor F_ϕ, and frozen pretrained teacher D_p
Require: View and noise-level distributions π, ν; length weights $\omega(k)$ and normalization floor $\varepsilon_{\mathrm{n}} > 0$
Require: Warmup length $W \geq 0$, generator update interval $r \geq 1$, and post-warmup iterations T

for $t = -W, \ldots, T-1$ **do**
  Sample $(h, k) \sim \pi$ and conditioning y_ϕ.
  Roll out $\boldsymbol{v} \sim q_\theta(\boldsymbol{v} \mid y_\phi)$ for $h + k$ blocks; set $\boldsymbol{v} \leftarrow \mathrm{sg}(\boldsymbol{v})$.
  Sample $\tau \sim \nu$ and $\boldsymbol{\epsilon}_v \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$.
  $\boldsymbol{v}^{(\tau)} \leftarrow (1 - \sigma_\tau)\boldsymbol{v} + \sigma_\tau \boldsymbol{\epsilon}_v$.
  $\ell_\phi \leftarrow \frac{1}{n_v} \| F_\phi(\boldsymbol{v}^{(\tau)}, \tau; y_\phi) - (\boldsymbol{\epsilon}_v - \boldsymbol{v}) \|_2^2$.
  Update ϕ using $\nabla_\phi \ell_\phi$; keep θ fixed.
  if $t \geq 0$ **and** $t \bmod r = 0$ **then**
    Sample fresh conditioning y_G and rollout randomness.
    Roll out h history blocks $\boldsymbol{c} \sim q_\theta(\boldsymbol{c} \mid y_G)$; set $\boldsymbol{c} \leftarrow \mathrm{sg}(\boldsymbol{c})$.
    Generate k blocks $\boldsymbol{z} \sim q_\theta(\boldsymbol{z} \mid \boldsymbol{c}, y_G)$ with retained gradients.
    $\boldsymbol{v} \leftarrow (\boldsymbol{c}, \boldsymbol{z})$.
    Independently sample $\tau \sim \nu$ and $\boldsymbol{\epsilon}_v \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$.
    $\boldsymbol{v}^{(\tau)} \leftarrow (1 - \sigma_\tau)\,\mathrm{sg}(\boldsymbol{v}) + \sigma_\tau \boldsymbol{\epsilon}_v$.
    $\widehat{\boldsymbol{z}}_p \leftarrow [D_p(\boldsymbol{v}^{(\tau)}, \tau; y_G)]_z$.
    $\widehat{\boldsymbol{z}}_\phi \leftarrow [D_\phi(\boldsymbol{v}^{(\tau)}, \tau; y_G)]_z$.
    $a \leftarrow \max\{\varepsilon_{\mathrm{n}}, \|\boldsymbol{z} - \widehat{\boldsymbol{z}}_p\|_1 / n_z\}$.
    $\boldsymbol{g} \leftarrow (\widehat{\boldsymbol{z}}_\phi - \widehat{\boldsymbol{z}}_p)/a$.
    $\ell_G \leftarrow \frac{\omega(k)}{2 n_z} \| \boldsymbol{z} - \mathrm{sg}(\boldsymbol{z} - \boldsymbol{g}) \|_2^2$.
    Update θ using $\nabla_\theta^{\mathrm{ret}} \ell_G$; keep ϕ fixed.
  end if
end for
return G_θ

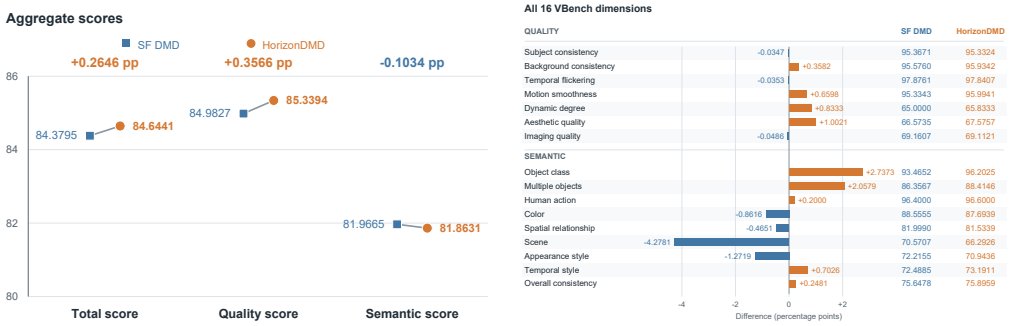

Figure 4: VBench comparison between SF DMD and HorizonDMD. Left: aggregate total, quality, and semantic scores. Right: differences across all 16 evaluation dimensions, computed as HorizonDMD minus SF DMD, with the corresponding absolute scores shown alongside. Scores are normalized and scaled by 100; differences are reported in percentage points. This figure compares the historical SF DMD evaluation with configuration H1; the two results are not from a common-noise paired evaluation.

prefix $\boldsymbol{x}^{(\tau)}_{1:h+k}$, while the generator surrogate is restricted to $\boldsymbol{x}_{h+1:h+k}$ with the generated history held fixed. Auxiliary denoising covers the entire selected prefix. Thus, the distinction lies in the temporal scope of training supervision and the coordinates receiving the generator loss; the autoregressive inference procedure is unchanged.

Quantitative results. Table 1 reports the aggregate VBench results. Among the evaluated configurations, single-view HorizonDMD with $p_{\mathrm{full}} = 1/3$ achieves the highest observed total and quality scores, 84.6441 and 85.3394, respectively. Relative to SF DMD in the current evaluation, the total and quality scores increase by 0.1268 and 0.2857 points, while the semantic score decreases by 0.5089 points. The corresponding total-score difference against the historical SF DMD evaluation is 0.2646 points. These results indicate a quality–semantic trade-off rather than uniform improvement across all evaluation criteria.

Table 1: VBench comparison of SF DMD and HorizonDMD. Scores are scaled by 100; higher is better. Boldface indicates the highest value within the current-evaluation block. H1–H4 denote the configurations in Table 4. The historical SF DMD evaluation is reported separately.

Method / configuration	p_{full}	Step	Total ↑	Quality ↑	Semantic ↑
Historical evaluation					
SF DMD	—	600	84.3795	84.9827	81.9665
Current evaluation					
SF DMD	—	600	84.5174	85.0537	**82.3720**
H1: HorizonDMD (single-view)	$1/3$	800	**84.6441**	**85.3394**	81.8631
H2: HorizonDMD (single-view)	$3/8$	700	84.4631	85.2796	81.1972
H3: HorizonDMD (two-prefix)	$3/8$	600	84.4434	85.1436	81.6429
H4: H3 + split RNG	$3/8$	600	84.4445	85.0032	82.2099

SF DMD uses a fixed checkpoint from our own training run. For each HorizonDMD configuration, we select the highest-total-score checkpoint among its completed evaluations, without a separate validation set. Checkpoint search ranges may differ across configurations. Here $p_{\text{full}} = \Pr(k = N)$ denotes the full-video-view probability in HorizonDMD; SF DMD directly uses full-video matching.

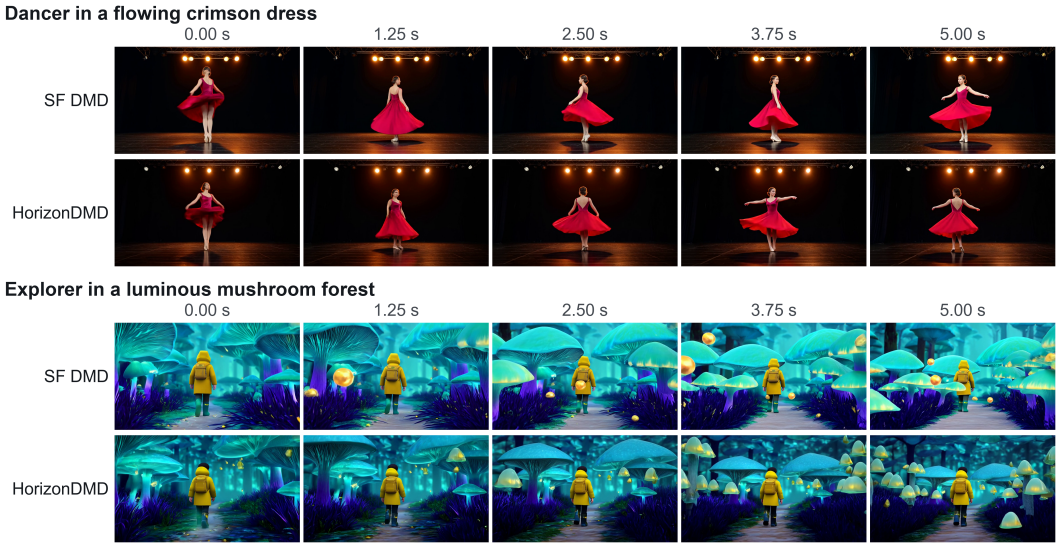

Figure 5: Qualitative comparison between SF DMD and HorizonDMD under the same text prompts and matched sampling noise. For each prompt, the upper row shows SF DMD and the lower row shows HorizonDMD. Columns correspond to 0, 1.25, 2.50, 3.75, and 5.00 seconds.

5 Conclusion

We introduced HorizonDMD, which generalizes full-video distribution matching to a family of global and history-conditioned views for autoregressive video distillation. Under consistent sequence marginals and positive conditional coverage, the ideal horizon objective preserves the zero-divergence target of full-sequence matching. Our analysis further connects prefix matching to conditional horizon updates when the clean-history distribution is held fixed, and shows why restricting full-video scores does not generally recover the same guidance. In the reported VBench comparison, the selected single-view configuration achieves higher total and quality scores than the re-evaluated SF DMD baseline, alongside a lower semantic score. These observations highlight the importance of examining how temporal supervision affects different aspects of generation quality, rather than relying on aggregate scores alone. Understanding how the distribution over views shapes this trade-off remains a direction for further study.

AI USE STATEMENT

We used generative AI tools to assist with manuscript editing, LaTeX preparation, figure and table preparation, code-related assistance, and discussions of experimental design and result interpretation. The reported numerical measurements were obtained from model training and evaluation runs rather than generated by these tools. The authors take responsibility for the accuracy and integrity of the final manuscript, code, and reported results.

ETHICS STATEMENT

This work studies the distillation of autoregressive video generators. As with other video-generation systems, the resulting models could be misused to create misleading or impersonating content and may inherit biases from their pretrained components. Responsible use should respect applicable model and data licenses, privacy, and consent, and should disclose synthetic content where appropriate.

REPRODUCIBILITY STATEMENT

Section 3 specifies the matching objective and learning losses, and Algorithm 1 summarizes single-view training. Section 4 reports the evaluated training configurations, checkpoint-selection procedure, and aggregate scores. Appendix C provides the common-checkpoint comparison and raw per-dimension scores. The theoretical assumptions are stated in Section 3.4, with complete proofs in Appendix B.

REFERENCES

Hmrishav Bandyopadhyay, Xuanchi Ren, Zijian Huang, Jay Zhangjie Wu, Tianshi Cao, Ruilong Li, Bryan Chu, Sanja Fidler, Yi-Zhe Song, and Zian Wang. Context-Matched Distillation: Teacher causality for autoregressive video distillation. *arXiv preprint arXiv:2608.13391*, 2026. doi: 10.48550/arXiv.2608.13391. URL https://arxiv.org/abs/2608.13391v1.

Andreas Blattmann, Tim Dockhorn, Sumith Kulal, Daniel Mendelevitch, Maciej Kilian, Dominik Lorenz, Yam Levi, Zion English, Vikram Voleti, Adam Letts, Varun Jampani, and Robin Rombach. Stable Video Diffusion: Scaling latent video diffusion models to large datasets. *arXiv preprint arXiv:2311.15127*, 2023. doi: 10.48550/arXiv.2311.15127. URL https://arxiv.org/abs/2311.15127.

Shengqu Cai, Weili Nie, Chao Liu, Julius Berner, Lvmin Zhang, Nanye Ma, Hansheng Chen, Maneesh Agrawala, Leonidas Guibas, Gordon Wetzstein, and Arash Vahdat. Mode seeking meets mean seeking for fast long video generation. *arXiv preprint arXiv:2602.24289*, 2026. doi: 10.48550/arXiv.2602.24289. URL https://arxiv.org/abs/2602.24289v3.

Boyuan Chen, Diego Martí Monsó, Yilun Du, Max Simchowitz, Russ Tedrake, and Vincent Sitzmann. Diffusion Forcing: Next-token prediction meets full-sequence diffusion. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-0759. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/2aee1c4159e48407d68fe16ae8e6e49e-Abstract-Conference.html.

Hanmo Chen, Chenghao Xu, Xu Yang, Xuan Chen, and Cheng Deng. Past- and future-informed KV cache policy with salience estimation in autoregressive video diffusion. *arXiv preprint arXiv:2601.21896*, 2026a. doi: 10.48550/arXiv.2601.21896. URL https://arxiv.org/abs/2601.21896v3.

Shuo Chen, Cong Wei, Sun Sun, Ping Nie, Kai Zou, Ge Zhang, Ming-Hsuan Yang, and Wenhu Chen. Context Forcing: Consistent autoregressive video generation with long context. In *International Conference on Machine Learning*, 2026b. URL https://chenshuo20.github.io/Context_Forcing.

Justin Cui, Jie Wu, Ming Li, Tao Yang, Xiaojie Li, Rui Wang, Andrew Bai, Yuanhao Ban, and Cho-Jui Hsieh. Self-Forcing++: Towards minute-scale high-quality video generation. *arXiv*

preprint arXiv:2510.02283, 2025. doi: 10.48550/arXiv.2510.02283. URL https://arxiv.org/abs/2510.02283.

Xiangyu Fan, Zesong Qiu, Zhuguanyu Wu, Fanzhou Wang, Zhiqian Lin, Tianxiang Ren, Dahua Lin, Ruihao Gong, and Lei Yang. Phased DMD: Few-step distribution matching distillation via score matching within subintervals. *arXiv preprint arXiv:2510.27684*, 2025. doi: 10.48550/arXiv.2510.27684. URL https://arxiv.org/abs/2510.27684v3.

Kaifeng Gao, Jiaxin Shi, Hanwang Zhang, Chunping Wang, Jun Xiao, and Long Chen. Ca2-VDM: Efficient autoregressive video diffusion model with causal generation and cache sharing. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 18550–18565. PMLR, 2025. URL https://proceedings.mlr.press/v267/gao25m.html.

Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J. Fleet. Video diffusion models. In *Advances in Neural Information Processing Systems*, volume 35, 2022. doi: 10.52202/068431-0628. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/39235c56aef13fb05a6adc95eb9d8d66-Abstract-Conference.html.

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*, volume 38, 2025. doi: 10.52202/085713-5576. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/f4823f831af67a3ef15e41a85434422a-Abstract-Conference.html.

Zhennan Jiang, Shangqing Zhou, Yutong Jiang, Zefang Huang, Mingjie Wei, Yuhui Chen, Tianxing Zhou, Zhen Guo, Hao Lin, Quanlu Zhang, Yu Wang, Haoran Li, Chao Yu, and Dongbin Zhao. WoVR: World models as reliable simulators for post-training VLA policies with RL. *arXiv preprint arXiv:2602.13977*, 2026. doi: 10.48550/arXiv.2602.13977. URL https://arxiv.org/abs/2602.13977v2.

Dimitrios Karageorgiou, Symeon Papadopoulos, Ioannis Kompatsiaris, and Efstratios Gavves. Test-time noise guided adaptation for realistic autoregressive video generation. In *European Conference on Computer Vision*, 2026. URL https://arxiv.org/abs/2607.15849v1.

Hongyu Liu, Chun Wang, Feng Gao, Xuanhua He, Yue Ma, Ziyu Wan, Yong Zhang, Xiaoming Wei, and Qifeng Chen. OPSD-V: On-policy self-distillation for post-training few-step autoregressive video generators. *arXiv preprint arXiv:2607.08766*, 2026. doi: 10.48550/arXiv.2607.08766. URL https://arxiv.org/abs/2607.08766v1.

Yunhong Lu, Yanhong Zeng, Haobo Li, Hao Ouyang, Qiuyu Wang, Ka Leong Cheng, Jiapeng Zhu, Hengyuan Cao, Zhipeng Zhang, Xing Zhu, Yujun Shen, and Min Zhang. Reward Forcing: Efficient streaming video generation with rewarded distribution matching distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 34385–34397, 2026. URL https://openaccess.thecvf.com/content/CVPR2026/html/Lu_Reward_Forcing_Efficient_Streaming_Video_Generation_with_Rewarded_Distribution_Matching_CVPR_2026_paper.html.

Jiayi Luo, Qiyan Liu, Tengyang Wang, JunHao Liu, Jiayu Chen, Cong Wang, Hanxin Zhu, Chen Gao, Xiaobin Hu, Qingyun Sun, and Zhibo Chen. Future Forcing: Future-aware training-free KV cache policy for autoregressive video generation. *arXiv preprint arXiv:2605.30083*, 2026a. doi: 10.48550/arXiv.2605.30083. URL https://arxiv.org/abs/2605.30083v1.

Yawen Luo, Xiaoyu Shi, Junhao Zhuang, Yutian Chen, Quande Liu, Xintao Wang, Pengfei Wan, and Tianfan Xue. ShotStream: Streaming multi-shot video generation for interactive storytelling. *arXiv preprint arXiv:2603.25746*, 2026b. doi: 10.48550/arXiv.2603.25746. URL https://arxiv.org/abs/2603.25746v1.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. *arXiv preprint arXiv:2202.00512*, 2022. doi: 10.48550/arXiv.2202.00512. URL https://arxiv.org/abs/2202.00512.

Kiwhan Song, Boyuan Chen, Max Simchowitz, Yilun Du, Russ Tedrake, and Vincent Sitzmann. History-guided video diffusion. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 56242–56280. PMLR, 2025. URL https://proceedings.mlr.press/v267/song25b.html.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 32211–32252. PMLR, 2023. URL https://proceedings.mlr.press/v202/song23a.html.

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. *arXiv preprint arXiv:2605.30116*, 2026. doi: 10.48550/arXiv.2605.30116. URL https://arxiv.org/abs/2605.30116v2.

Shuai Yang, Wei Huang, Ruihang Chu, Yicheng Xiao, Yuyang Zhao, Xianbang Wang, Muyang Li, Enze Xie, Yingcong Chen, Yao Lu, Song Han, and Yukang Chen. LongLive: Real-time interactive long video generation. *arXiv preprint arXiv:2509.22622*, 2025. doi: 10.48550/arXiv.2509.22622. URL https://arxiv.org/abs/2509.22622.

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*, volume 37, 2024a. doi: 10.52202/079017-1505. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/54dcf25318f9de5a7a01f0a4125c541e-Abstract-Conference.html.

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*, pp. 6613–6623, 2024b. URL https://openaccess.thecvf.com/content/CVPR2024/html/Yin_One-step_Diffusion_with_Distribution_Matching_Distillation_CVPR_2024_paper.html.

Tianwei Yin, Qiang Zhang, Richard Zhang, William T. Freeman, Fredo Durand, Eli Shechtman, and Xun Huang. From slow bidirectional to fast autoregressive video diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22963–22974, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Yin_From_Slow_Bidirectional_to_Fast_Autoregressive_Video_Diffusion_Models_CVPR_2025_paper.html.

Yonghao Yu, Lang Huang, Runyi Li, Zerun Wang, and Toshihiko Yamasaki. Video-Mirai: Autoregressive video diffusion models need foresight. *arXiv preprint arXiv:2606.03971*, 2026. doi: 10.48550/arXiv.2606.03971. URL https://arxiv.org/abs/2606.03971v1.

Shenghai Yuan, Yuanyang Yin, Zongjian Li, Xinwei Huang, Xiao Yang, and Li Yuan. Helios: Real real-time long video generation model. *arXiv preprint arXiv:2603.04379*, 2026. doi: 10.48550/arXiv.2603.04379. URL https://arxiv.org/abs/2603.04379v1.

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. doi: 10.48550/arXiv.2606.21982. URL https://arxiv.org/abs/2606.21982v1.

Min Zhao, Hongzhou Zhu, Kaiwen Zheng, Zihan Zhou, Bokai Yan, Xinyuan Li, Xiao Yang, Chongxuan Li, and Jun Zhu. Causal Forcing++: Scalable few-step autoregressive diffusion distillation for real-time interactive video generation. *arXiv preprint arXiv:2605.15141*, 2026. doi: 10.48550/arXiv.2605.15141. URL https://arxiv.org/abs/2605.15141v3.

Kaiwen Zheng, Guande He, Min Zhao, Jintao Zhang, Huayu Chen, Jianfei Chen, Chen-Hsuan Lin, Ming-Yu Liu, Jun Zhu, and Qianli Ma. Causal-rCM: A unified teacher-forcing and self-forcing open recipe for autoregressive diffusion distillation in streaming video generation and interactive

world models. *arXiv preprint arXiv:2606.25473*, 2026. doi: 10.48550/arXiv.2606.25473. URL https://arxiv.org/abs/2606.25473v1.

Hongzhou Zhu, Min Zhao, Guande He, Hang Su, Chongxuan Li, and Jun Zhu. Causal Forcing: Autoregressive diffusion distillation done right for high-quality real-time interactive video generation. In *International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2602.02214v5.

Junhao Zhuang, Shiyi Zhang, Yuxuan Bian, Yaowei Li, Yawen Luo, Yijun Liu, Weiyang Jin, Songchun Zhang, Xianglong He, Xuying Zhang, Haoran Li, Haoyang Huang, Zeyue Xue, and Nan Duan. Self Gradient Forcing: Native long video extrapolation. *arXiv preprint arXiv:2607.20368*, 2026. doi: 10.48550/arXiv.2607.20368. URL https://arxiv.org/abs/2607.20368v2.

Kai Zou, Dian Zheng, Hongbo Liu, Tiankai Hang, Bin Liu, and Nenghai Yu. HiAR: Efficient autoregressive long video generation via hierarchical denoising. *arXiv preprint arXiv:2603.08703*, 2026. doi: 10.48550/arXiv.2603.08703. URL https://arxiv.org/abs/2603.08703v1.

APPENDIX

A ADDITIONAL RELATED WORK AND SCOPE

Related approaches address temporal learning through the source of supervision, gradient propagation, denoising dynamics, and historical memory. These directions complement the question of which history-conditioned distributions are matched.

Supervision and generated histories. The reference supervision can vary both in its statistical target and in the histories under which it is evaluated. OPSD-V uses an exponential-moving-average teacher for on-policy velocity regression, replacing older generated context with real-video history while retaining the most recent generated block (Liu et al., 2026). Mode Seeking meets Mean Seeking combines global flow matching on real long videos with local sliding-window DMD (Cai et al., 2026). Reward Forcing introduces motion-reward-weighted DMD alongside compressed historical memory (Lu et al., 2026). These methods modify the supervision source or the preferred output distribution; our coverage analysis instead assumes conditionals of a common pair of sequence laws.

Gradient propagation and representation supervision. Temporal learning also depends on how historical representations receive supervision. Self Gradient Forcing allows later DMD losses to train historical key–value representations while holding the generated history latents fixed (Zhuang et al., 2026). This propagates gradients through the re-encoded memory, not through the full autoregressive rollout. Video-Mirai transfers foresight from a frozen bidirectional encoder through representation alignment (Yu et al., 2026). These approaches concern gradient propagation and representation targets, which are distinct from the conditional distributions chosen for matching.

Denoising dynamics and test-time adaptation. Temporal robustness can also be addressed through the generation process itself. HiAR organizes denoising hierarchically across video blocks, aligning their noise levels within each denoising stage (Zou et al., 2026). This changes the generation schedule, rather than the Gaussian corruption used for prefix matching. TANGO adapts generation at test time using look-ahead noise statistics motivated by a noise-consistency hypothesis (Karageorgiou et al., 2026). Here the focus is inference-time trajectory adaptation, rather than the training distribution over matching views.

Historical memory and long-rollout robustness. Memory selection determines which past information remains available during generation. PaFu-KV learns salience estimates for retaining historical key–value entries (Chen et al., 2026a), whereas Future Forcing estimates future relevance from historical query statistics for training-free cache management (Luo et al., 2026a). Its future-aware selection does not require realized future frames. Beyond memory selection, Helios combines compressed representations, anti-drift context construction, and hierarchical distillation (Yuan et al., 2026). These approaches address long-rollout robustness through memory or training design; HorizonDMD does not prescribe a new inference-time memory policy.

Multi-shot generation and world-model applications. Interactive settings introduce additional conditioning structures and learning objectives. ShotStream distills a next-shot teacher conditioned on sparse historical context and trains intra-shot and inter-shot generation for interactive storytelling (Luo et al., 2026b). This selected cross-shot context differs from the complete-history conditioning assumed by our coverage theorem. WoVR uses action-conditioned world models as simulators for reinforcement learning of vision–language–action policies (Jiang et al., 2026), extending the role of generated trajectories from video synthesis to closed-loop policy learning.

B PROOFS AND INTERPRETATION

We use the notation of Section 3. External conditioning y is fixed and suppressed. Statements about scores assume smooth positive densities and the integrability needed to interchange differentiation and integration. Conditional identities hold almost everywhere under the stated laws. The arguments apply classical KL and score identities to the history–horizon objective.

B.1 CONDITIONAL HORIZON COVERAGE

Fix τ and abbreviate $q = q_\theta^{(\tau)}$ and $p = p^{(\tau)}$ in this subsection. Write $\boldsymbol{x}_i^{(\tau)}$ for block i of the noised sequence. The single-block discrepancy is

$$d_i^{(\tau)} = \boldsymbol{E}_{\boldsymbol{x}_{<i}^{(\tau)} \sim q} \left[\mathrm{KL}\Big(q(\boldsymbol{x}_i^{(\tau)} \mid \boldsymbol{x}_{<i}^{(\tau)}) \,\|\, p(\boldsymbol{x}_i^{(\tau)} \mid \boldsymbol{x}_{<i}^{(\tau)}) \Big) \right]. \tag{21}$$

Applying the conditional KL chain rule to a horizon gives

$$\mathcal{D}_{h,k}^{(\tau)} = \sum_{i=h+1}^{h+k} d_i^{(\tau)}. \tag{22}$$

In this expansion, averaging the horizon's earlier blocks together with its history produces the student marginal on $\boldsymbol{x}_{<i}^{(\tau)}$. Because all views use the same two sequence laws, $d_i^{(\tau)}$ is independent of which horizon contains position i. Consequently,

$$\mathcal{R}_\pi^{(\tau)} = \sum_{h,k} \pi(h,k) \frac{\omega(k)}{k} \sum_{i=h+1}^{h+k} d_i^{(\tau)} = \sum_{i=1}^{N} \left[\sum_{h,k} \pi(h,k) \frac{\omega(k)}{k} \mathbf{1}\{h < i \le h+k\} \right] d_i^{(\tau)}. \tag{23}$$

This proves Eq. (11). Since each $d_i^{(\tau)}$ is nonnegative and $\sum_i d_i^{(\tau)} = \mathrm{KL}(q\|p)$, Eq. (12) follows. If $\lambda_{\min} > 0$, zero horizon risk is equivalent to equality of the noised sequence laws.

To recover the clean laws, consider their characteristic functions. For any Fourier variable $\boldsymbol{\xi}$,

$$\varphi_{q_\theta^{(\tau)}}(\boldsymbol{\xi}) = \varphi_{q_\theta}(\alpha_\tau \boldsymbol{\xi}) \exp\left(-\frac{\sigma_\tau^2 \|\boldsymbol{\xi}\|_2^2}{2} \right), \tag{24}$$

and the analogous identity holds for $p^{(\tau)}$. The exponential factor never vanishes. For $\alpha_\tau \neq 0$, equality of the corrupted laws therefore implies equality of the clean characteristic functions everywhere, hence $q_\theta = p$. The converse follows by applying the same corruption channel. At $\alpha_\tau = 0$, the channel discards the clean signal and this identification is impossible.

Nonnegativity implies that $\mathcal{R}_\pi = 0$ only if $\mathcal{R}_\pi^{(\tau)} = 0$ for ν-almost every τ. Since ν is supported on nondegenerate noise levels, the same zero-risk identification holds after averaging over τ. This is an exact-equality result; Gaussian smoothing does not in general provide a stable bound on clean KL from a small noisy KL. The chain-rule proof also applies directly to clean laws whenever their KL is finite.

Necessity and sharpness of coverage. If $\lambda_j = 0$, choose independent binary blocks with reference success probability $1/2$ at every position and student success probability $3/4$ only at j. In the clean-channel case, $d_j^{(\tau)} > 0$ and all other conditional divergences vanish. The risk is zero although the sequence laws differ. Independent Gaussian coordinates with a mean shift only at j give the same counterexample after nondegenerate Gaussian corruption. Concentrating the discrepancy at a minimum- or maximum-coverage position also shows that the constants in Eq. (12) are sharp.

Equal coverage and frozen-history gradients. For $\rho \in [0,1]$, assign probability ρ to the full-video view $(0, N)$ and probability $1-\rho$ to the uniform distribution over $(i-1, 1)$, $i = 1, \dots, N$. With $\omega(k) = k/\boldsymbol{E}_\pi[k]$,

$$\boldsymbol{E}_\pi[k] = \rho N + 1 - \rho, \qquad \lambda_i = \frac{\rho + (1-\rho)/N}{\rho N + 1 - \rho} = \frac{1}{N}. \tag{25}$$

All these mixtures induce full-sequence KL divided by N and thus have identical total parameter gradients wherever differentiable. Nevertheless, freezing different histories can give different semi-gradients. Approximate scores, normalization, and further gradient truncation can also produce different practical updates.

B.2 Proof of prefix–conditional gradient equivalence

Fix a pair (h, k) and a current parameter value θ_0. For a varying parameter ϑ, define the clean prefix law

$$\bar{q}_\vartheta(\boldsymbol{c}, \boldsymbol{z}) = q_{\theta_0}(\boldsymbol{c}) q_\vartheta(\boldsymbol{z} \mid \boldsymbol{c}). \tag{26}$$

Apply the same coordinate-wise Gaussian channel to obtain $\bar{q}_\vartheta^{(\tau)}$. Its noisy-history marginal is $q_{\theta_0}^{(\tau)}(\boldsymbol{c}^{(\tau)})$, independent of ϑ. Let $\bar{\mathcal{D}}_{h,k}^{(\tau)}(\vartheta)$ denote the conditional divergence in Eq. (8) computed using this student law. Define

$$\nabla_\theta^{\mathrm{fr}} \mathcal{D}_{h,k}^{(\tau)}\Big|_{\theta=\theta_0} := \nabla_\vartheta \bar{\mathcal{D}}_{h,k}^{(\tau)}(\vartheta)\Big|_{\vartheta=\theta_0}. \tag{27}$$

The same frozen-law construction defines the derivative of the KL between the noisy prefix distributions. The clean-history law is frozen before corruption; generation itself remains conditioned on clean history.

Proof of Proposition 1. For every ϑ in the stated neighborhood, the KL chain rule gives

$$\mathrm{KL}\Big(\bar{q}_\vartheta^{(\tau)}(\boldsymbol{v}^{(\tau)}) \,\|\, p^{(\tau)}(\boldsymbol{v}^{(\tau)})\Big) = \mathrm{KL}\Big(q_{\theta_0}^{(\tau)}(\boldsymbol{c}^{(\tau)}) \,\|\, p^{(\tau)}(\boldsymbol{c}^{(\tau)})\Big) + \bar{\mathcal{D}}_{h,k}^{(\tau)}(\vartheta). \tag{28}$$

The history-marginal term is finite and independent of ϑ. Differentiating and evaluating at $\vartheta = \theta_0$ proves Eq. (15). No score approximation is involved in this equality of population derivatives. Differentiating the corresponding log-density factorization with respect to $\boldsymbol{z}^{(\tau)}$ also proves Eq. (14).

For the score representation, consider a reparameterized variable $\boldsymbol{w}_\vartheta = T_\vartheta(\boldsymbol{\xi})$ with density r_ϑ and a fixed reference density p. Under the stated regularity assumptions,

$$\nabla_\vartheta \mathrm{KL}(r_\vartheta \| p) = \boldsymbol{E}\big[(\partial_\vartheta \boldsymbol{w}_\vartheta)^{\mathsf{T}} \left(\nabla_{\boldsymbol{w}} \log r_\vartheta(\boldsymbol{w}_\vartheta) - \nabla_{\boldsymbol{w}} \log p(\boldsymbol{w}_\vartheta)\right)\big]. \tag{29}$$

The explicit density-derivative term has zero expectation because its integral is the derivative of the normalization $\int r_\vartheta(\boldsymbol{w})\, d\boldsymbol{w} = 1$.

In the frozen-history construction, the noisy-history coordinates have zero parameter derivative, while the horizon derivative is $\alpha_\tau \partial_\vartheta \boldsymbol{z}|_{\boldsymbol{c}}$. Only horizon score coordinates contribute. At $\vartheta = \theta_0$, the frozen law coincides with the current student's visible marginal. Applying Eq. (29) to Eq. (28) and dividing by k gives Eq. (16). □

Without freezing the history law, the total derivatives of the two KL objectives can differ by the derivative of the history-marginal KL. The proposition therefore does not establish equality of their unrestricted on-policy gradients.

B.3 Exact denoiser and score identities

For a law $r \in \{p, q_\theta\}$, let $\widehat{\boldsymbol{v}}_r(\boldsymbol{v}^{(\tau)})$ denote the conditional mean of the clean prefix given its corrupted value under r. The Gaussian score identity is

$$\boldsymbol{s}_r(\boldsymbol{v}^{(\tau)}) = \frac{\alpha_\tau \widehat{\boldsymbol{v}}_r(\boldsymbol{v}^{(\tau)}) - \boldsymbol{v}^{(\tau)}}{\sigma_\tau^2}. \tag{30}$$

Subtracting the teacher and student expressions and restricting to the horizon gives

$$\boldsymbol{\delta} = \frac{\alpha_\tau}{\sigma_\tau^2}(\widehat{\boldsymbol{z}}_{q_\theta} - \widehat{\boldsymbol{z}}_p). \tag{31}$$

Here $\widehat{\boldsymbol{z}}_{q_\theta}$ is the exact student posterior mean, not the learned auxiliary estimate $\widehat{\boldsymbol{z}}_\phi$. Both posterior means in Eq. (31) condition on the entire noisy prefix. Combining this equation with Eq. (16) gives

$$\nabla_\theta^{\mathrm{fr}} \left[\frac{\mathcal{D}_{h,k}^{(\tau)}}{k}\right] = \frac{\alpha_\tau^2}{k\sigma_\tau^2} \boldsymbol{E}\big[\boldsymbol{J}_\theta^{\mathsf{T}}(\widehat{\boldsymbol{z}}_{q_\theta} - \widehat{\boldsymbol{z}}_p)\big]. \tag{32}$$

The horizon weight multiplies this expression by $\omega(k)$. The practical normalized update does not in general preserve the ideal factor $\alpha_\tau^2/(k\sigma_\tau^2)$ or its exact denoisers.

B.4 Proof of scoring-scope non-equivalence

Fix (h, k) and τ. The full score and the excluded blocks $\boldsymbol{u}$ are defined in Section 3.4. For comparison with the prefix discrepancy, define

$$\boldsymbol{\delta}^{\mathrm{full}}(\boldsymbol{v}^{(\tau)}, \boldsymbol{u}^{(\tau)}) = \left[\boldsymbol{s}^{\mathrm{full}}_{q_\theta}(\boldsymbol{v}^{(\tau)}, \boldsymbol{u}^{(\tau)}) - \boldsymbol{s}^{\mathrm{full}}_{p}(\boldsymbol{v}^{(\tau)}, \boldsymbol{u}^{(\tau)}) \right]_z . \tag{33}$$

Here we write $\boldsymbol{\delta}(\boldsymbol{v}^{(\tau)})$ to make the prefix discrepancy's input explicit. The exact difference is

$$\begin{aligned} &\boldsymbol{E}_{q_\theta^{(\tau)}(\boldsymbol{u}^{(\tau)}|\boldsymbol{v}^{(\tau)})}[\boldsymbol{\delta}^{\mathrm{full}}(\boldsymbol{v}^{(\tau)}, \boldsymbol{u}^{(\tau)})] - \boldsymbol{\delta}(\boldsymbol{v}^{(\tau)}) \\ &\quad = \boldsymbol{E}_{p^{(\tau)}(\boldsymbol{u}^{(\tau)}|\boldsymbol{v}^{(\tau)})}[[\boldsymbol{s}^{\mathrm{full}}_{p}(\boldsymbol{v}^{(\tau)}, \boldsymbol{u}^{(\tau)})]_z] \\ &\qquad - \boldsymbol{E}_{q_\theta^{(\tau)}(\boldsymbol{u}^{(\tau)}|\boldsymbol{v}^{(\tau)})}[[\boldsymbol{s}^{\mathrm{full}}_{p}(\boldsymbol{v}^{(\tau)}, \boldsymbol{u}^{(\tau)})]_z]. \end{aligned} \tag{34}$$

Proof of Proposition 2. For either law, differentiation under the marginalization integral gives

$$\begin{aligned} \nabla_{\boldsymbol{v}^{(\tau)}} \log r^{(\tau)}(\boldsymbol{v}^{(\tau)}) &= \frac{\int \nabla_{\boldsymbol{v}^{(\tau)}} r^{(\tau)}(\boldsymbol{v}^{(\tau)}, \boldsymbol{u}^{(\tau)})\, d\boldsymbol{u}^{(\tau)}}{r^{(\tau)}(\boldsymbol{v}^{(\tau)})} \\ &= \boldsymbol{E}_{r^{(\tau)}(\boldsymbol{u}^{(\tau)}|\boldsymbol{v}^{(\tau)})}[[\boldsymbol{s}^{\mathrm{full}}_{r}(\boldsymbol{v}^{(\tau)}, \boldsymbol{u}^{(\tau)})]_v]. \end{aligned} \tag{35}$$

Here $[\cdot]_v$ selects visible-prefix coordinates. Restricting further to the horizon proves Eq. (17). Apply that identity separately to student and teacher. In the difference in Eq. (34), the student terms cancel, leaving the teacher score averaged under two different conditional laws. □

For a concrete example, take scalar visible and excluded coordinates v and u, respectively, with no preceding history, and let $q_\theta^{(\tau)} = \mathcal{N}(\mathbf{0}, \boldsymbol{I}_2)$ and $p^{(\tau)} = \mathcal{N}(\mathbf{0}, \left[\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right])$ for $0 < |\rho| < 1$. Choose a noise level with $\alpha_\tau \neq 0$ and $\sigma_\tau^2 < 1 - |\rho|$, so both laws arise from positive-definite Gaussian clean laws under the common channel. The visible marginals agree, so their score discrepancy is zero. However, the student-conditional average of the restricted full-score discrepancy is

$$\boldsymbol{E}_{q_\theta^{(\tau)}(u|v)}[\delta^{\mathrm{full}}(v, u)] = \frac{\rho^2 v}{1 - \rho^2}. \tag{36}$$

The scalar notation reflects the one-dimensional target in this example. The full score detects a dependence mismatch invisible to this single prefix. Non-equivalence therefore does not establish superiority of prefix matching; it establishes that the two views can provide different guidance. This comparison is between score vectors, not their products with sample-dependent generator Jacobians.

B.5 Why unconditioned windows are insufficient in general

The coverage theorem uses complete-history conditionals, not window marginals alone. For $\boldsymbol{x} = (x_1, x_2, x_3) \in \{0, 1\}^3$ and $0 < \eta < 1$, consider

$$p(\boldsymbol{x}) = \frac{1 + \eta(-1)^{x_1 + x_2 + x_3}}{8}, \qquad q(\boldsymbol{x}) = \frac{1 - \eta(-1)^{x_1 + x_2 + x_3}}{8}. \tag{37}$$

Marginalizing any coordinate cancels the parity term, so all proper window marginals agree. Yet $\mathrm{KL}(q\|p) = \eta \log((1 + \eta)/(1 - \eta)) > 0$. Independent coordinate-wise corruption preserves agreement of these marginals, while nondegenerate Gaussian corruption still distinguishes the complete laws. This is an identification counterexample, not a claim that window-based training cannot be effective.

B.6 Normalization and the detached-surrogate gradient

The positive scale in the normalized update is

$$a = \max\left\{ \varepsilon_{\mathrm{n}}, \frac{\|\boldsymbol{z} - \widehat{\boldsymbol{z}}_p\|_1}{n_z} \right\}, \qquad \boldsymbol{g} = \frac{\widehat{\boldsymbol{z}}_\phi - \widehat{\boldsymbol{z}}_p}{a}, \tag{38}$$

Table 2: HorizonDMD configurations evaluated at checkpoint index 700. All scores use the same scale as Table 1. Boldface indicates the highest value in each column. Matching checkpoint indices does not imply equal training compute.

Configuration	Total ↑	Quality ↑	Semantic ↑
H1: Single-view, $p_{\text{full}} = 1/3$	**84.4634**	85.0896	81.9583
H2: Single-view, $p_{\text{full}} = 3/8$	84.4631	**85.2796**	81.1972
H3: Two-prefix	84.3607	85.0106	81.7613
H4: Two-prefix + split RNG	84.2649	84.7916	**82.1583**

where $\varepsilon_{\text{n}} > 0$ is a scalar numerical floor. Both the discrepancy and normalization use horizon coordinates only. The regression target is recomputed at the current parameter value and held fixed only during differentiation. For Eq. (18), this gives

$$\nabla_{\boldsymbol{z}} \ell_G = \frac{\omega(k)}{n_z} \left(\boldsymbol{z} - \text{sg}(\boldsymbol{z} - \boldsymbol{g})\right) = \frac{\omega(k)}{n_z} \boldsymbol{g}. \tag{39}$$

The second equality substitutes the target's forward value only after differentiation. It does not differentiate through $\boldsymbol{g}$ or its sample-dependent normalizer.

Let $\boldsymbol{J}_\theta^{\text{ret}}$ denote the Jacobian retained by the training computation graph. The chain rule gives

$$\nabla_\theta^{\text{ret}} \ell_G = \frac{\omega(k)}{n_z} (\boldsymbol{J}_\theta^{\text{ret}})^{\mathsf{T}} \boldsymbol{g}. \tag{40}$$

If the graph stops gradients through intermediate generated states, this retained Jacobian need not equal the full conditional-generation Jacobian in Eq. (16). With equal-sized blocks, n_z is proportional to k; the chosen length weight makes $\omega(k)/n_z$ independent of k, without guaranteeing equal gradient norms across views.

The surrogate identity is exact for its detached target. It does not identify learned or guided predictions with exact posterior means, remove the effect of normalization, or equate a truncated training gradient with the full conditional KL derivative.

C Supplementary quantitative results

C.1 Comparison at a common checkpoint index

At checkpoint index 700, H1 and H2 attain nearly identical total scores, with a difference of approximately 0.00021 points, but exhibit different quality–semantic trade-offs. Thus, H1's higher best-observed total score cannot be attributed solely to the change in p_{full}. At the same index, H3 and H4 have lower total scores than H2, despite higher semantic scores. These observations do not establish a total-score benefit from two-prefix grouping or the split-RNG estimator.

C.2 Raw per-dimension scores

Score conventions. Table 3 reports raw metric values, whereas Figure 4 uses benchmark-normalized dimension scores multiplied by 100. These are distinct scales: for example, the historical SF DMD scene score is 0.580233 before normalization and 70.5707 after normalization and scaling. Aggregate quality and semantic scores use the benchmark normalization and weights, rather than a simple mean of the raw values.

Dimension-level results. HorizonDMD scores higher in nine dimensions and lower in seven. On the normalized scale, the largest increases are in object class (2.7373 points) and multiple objects (2.0579 points), while the largest decrease is in scene (4.2781 points). Motion smoothness, dynamic degree, and aesthetic quality also increase; subject consistency, temporal flickering, and imaging quality decrease slightly. This pattern is consistent with the aggregate quality–semantic trade-off in Table 1.

Table 3: Raw VBench dimension scores for historical SF DMD and HorizonDMD (H1). Values are reported before benchmark normalization, without multiplication by 100, and rounded to six decimal places. Boldface indicates the higher value in each row. These two evaluation records are not a common-noise paired evaluation.

Dimension	SF DMD	HorizonDMD
Quality dimensions		
Subject consistency	**0.960444**	0.960148
Background consistency	0.967329	**0.969974**
Temporal flickering	**0.992127**	0.991996
Motion smoothness	0.983899	**0.985823**
Dynamic degree	0.650000	**0.658333**
Aesthetic quality	0.665735	**0.675757**
Imaging quality	**0.691607**	0.691121
Semantic dimensions		
Object class	0.934652	**0.962025**
Multiple objects	0.863567	**0.884146**
Human action	0.964000	**0.966000**
Color	**0.885555**	0.876939
Spatial relationship	**0.819990**	0.815339
Scene	**0.580233**	0.545058
Appearance style	**0.206425**	0.202805
Temporal style	0.263858	**0.266416**
Overall consistency	0.275358	**0.276261**

Table 4: HorizonDMD training configurations. The configurations differ in the full-video-view probability, prefix grouping, or training-randomness policy. RNG denotes a random-number stream.

ID	p_{full}	View group	Training randomness
H1	$1/3$	One (h, k) view	Random-state replay across accumulation microsteps
H2	$3/8$	One (h, k) view	Same as H1
H3	$3/8$	Two histories with a shared k; one view if $k = N$	Same as H1
H4	$3/8$	Same as H3	Rollout replay with a separate advancing score RNG stream

C.3 Training configurations

View distribution. For $N = 7$ blocks, we parameterize the horizon distribution by the full-video-view probability p_{full}:

$$\pi_k(k) = \begin{cases} (1 - p_{\text{full}})/6, & k = 1, \dots, 6, \\ p_{\text{full}}, & k = 7. \end{cases} \tag{41}$$

For a given horizon length, the history length is uniform over the admissible positions:

$$\pi(h \mid k) = \frac{1}{N - k + 1}, \qquad h \in \{0, \dots, N - k\}. \tag{42}$$

The configurations with $p_{\text{full}} = 1/3$ and $3/8$ therefore assign probability $1/9$ and $5/48$, respectively, to each shorter horizon. Here p_{full} refers specifically to the full-video view, rather than all zero-history views: a global view with $h = 0$ may also have $k < N$.

Evaluation records. The two SF DMD rows correspond to the same checkpoint evaluated in different runs, not to different training methods. The historical SF DMD result and H1 do not form a common-noise paired evaluation. We report observed differences without a claim of statistical

significance. Aggregate scores follow

$$S_{\text{total}} = \frac{4S_{\text{quality}} + S_{\text{semantic}}}{5}, \tag{43}$$

with a dynamic-degree weight of 0.5 in the quality aggregate.

Single-view and two-prefix estimators. H1 and H2 use one sampled history–horizon view. H3 samples two distinct legal history lengths without replacement for the same horizon length and averages their losses. The two prefixes share a student rollout. At $k = N$, only $h = 0$ is admissible, yielding one view. This configuration changes the sampling estimator over history positions, rather than introducing two different horizon lengths.

Randomness across accumulation microsteps. H1–H3 replay a rank-local random state across accumulation microsteps. H4 retains rollout-randomness replay but uses a separate rank-local score stream that advances across microsteps and updates. Within each generator DMD loss, the teacher and auxiliary predictions still use the same score timestep and noisy prefix. This modification concerns the training estimator and does not alter inference-time sampling.

Shared training settings. H1–H4 use a global batch size of 64, eight accumulation microsteps, stop-gradient history handling, horizon-length weighting $\omega(k) = k/\mathbb{E}_\pi[k]$, 50 auxiliary-model warmup iterations, and auxiliary-model updates preceding generator updates. The selected view or view group is shared across accumulation microsteps and distributed ranks. On generator-update iterations, the generator and auxiliary model use the same view or view group.