PHYSICS-REINFORCED DISTRIBUTION MATCHING DISTILLATION FOR AUTOREGRESSIVE VIDEO GENERATION

Anonymous authors
Paper under double-blind review



Figure 1: We propose **PhysDMD**, a physics-reinforced distillation framework that improves physical plausibility in few-step autoregressive video generation via direct dynamics alignment and reward-guided optimization. The radar chart summarizes Physics-IQ scores against distilled AR baselines.

ABSTRACT

Autoregressive (AR) video diffusion distillation enables few-step streaming generation, but matching a teacher's distribution does not by itself ensure physically plausible dynamics over extended rollouts. We study how to evaluate and improve physical plausibility in distilled AR models. We first introduce PLV-Bench, an image-to-video benchmark comprising 125 real-world videos, across 16 physical phenomena and spanning at least 10 seconds, with a reference-guided, phase-level evaluation protocol. We then present PhysDMD, which augments Distribution Matching Distillation (DMD) with complementary supervision from a frozen latent world model. To avoid favoring trivially predictable motion, we calibrate the generated transition's prediction energy against its paired real trajectory. The resulting signal directly supervises the final deterministic denoising step and serves as a reward for group-relative optimization of intermediate stochastic transitions. On Physics-IQ-Verified, PhysDMD improves the Overall score from 25.31 to 28.10 over Vanilla-DMD while retaining four-step sampling and requiring no world-model evaluation at inference. Ablations examine the contributions of dynamics alignment, policy optimization, and training-data composition.

1 INTRODUCTION

Large-scale video diffusion models achieve high visual fidelity (Wan et al., 2025; Wu et al., 2025; MiniMax, 2026; Kuaishou, 2026; Google DeepMind, 2026a), enabling applications in game engines (Tang et al., 2025), embodied intelligence (Li et al., 2026c), and world modeling (Jiang et al., 2026; Agarwal et al., 2026; Google DeepMind, 2026b). These applications require low-latency generation with coherent dynamics over long horizons. Autoregressive (AR) video diffusion supports streaming by generating frames or chunks conditioned on previous outputs (Gu et al., 2025; Teng et al., 2025), but iterative denoising at each AR step remains expensive.

Distillation reduces this cost by training few-step students (Yin et al., 2025). Distribution Matching Distillation (DMD) (Yin et al., 2024b;a) aligns the student distribution with a pretrained teacher

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.

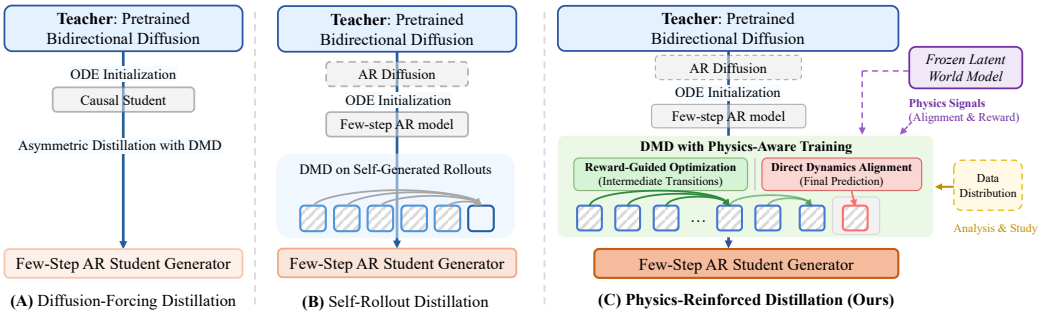

Figure 2: **Training paradigms for AR video diffusion distillation. (A) Diffusion-forcing distillation:** Supervises the student using DMD on data-derived historical inputs (Yin et al., 2025). **(B) Self-rollout distillation:** Applies DMD over self-generated rollouts (Huang et al., 2026a; Zhu et al., 2026). **(C) Physics-reinforced Distillation (Ours):** Augments diffusion distillation with dynamics supervision and reward guidance from a latent world model to improve physical plausibility.

and has been adopted for large-scale generation (Black Forest Labs, 2025; Cai et al., 2025; Zhao et al., 2026a). Recent advances address the quality–diversity trade-off (Wu et al., 2026; Zheng et al., 2026b;a) and incorporate self-rollouts for longer horizons (Huang et al., 2026a). These developments leave an important question:

Visual fidelity alone does not establish whether distilled autoregressive models maintain physically plausible dynamics over extended rollouts.

Evaluation. Standard video evaluations primarily target general-purpose perceptual fidelity (Liu et al., 2024), with emerging benchmarks expanding to assess long-form generation (Zheng et al., 2026c) and specific physical phenomena (Zheng et al., 2025; Bansal et al., 2026). Several physics-oriented benchmarks focus on short text-to-video (T2V) events without paired real-video references. Text-only conditions do not specify an observed trajectory against which to assess the timing and progression of a physical process. To bridge this gap, we introduce PLV-Bench, an image-to-video (I2V) benchmark designed to evaluate (distilled) AR models on extended physical dynamics against real-world trajectories spanning over 10 seconds. We drive this benchmark with a reference-guided, zero-shot vision-language model (VLM) evaluator that performs phase-level alignment, rather than rigid frame-wise matching, to rigorously assess visual and physical integrity. Alongside PLV-Bench, we conduct complementary evaluations under the established Physics-IQ (Motamed et al., 2026).

Our Method. Reward-guided distillation uses reinforcement learning (RL) through sequential post-training (Ren et al., 2024; Miao et al., 2025) or joint optimization (Jiang et al., 2025; Dong et al., 2026; Huang et al., 2026b) to improve target behaviors while retaining distributional supervision. In video generation, motion-oriented rewards (Lu et al., 2026) do not explicitly calibrate transitions against real physical evolution. Latent world models provide a dynamics-aware signal (Garrido et al., 2025; Assran et al., 2025), but using it for inference-time steering (Yuan et al., 2026a) adds search or guidance computation. We instead incorporate this signal during AR distillation. First, we calibrate world-model prediction energy against paired real trajectories to directly supervise the final denoising prediction. Second, we use the same calibrated signal as a reward for Group Relative Policy Optimization (GRPO) (Shao et al., 2024) of intermediate stochastic transitions. Third, we examine sensitivity to training-data composition. Together, the two supervision routes define PhysDMD, which augments DMD without adding inference-time world-model evaluation.

Our contributions are threefold. *First*, we introduce PLV-Bench, an image-conditioned benchmark with 125 real-world reference videos across 16 physical phenomena, paired with a phase-level VLM evaluation to study sustained physical evolution in AR generation. *Second*, we propose PhysDMD, which calibrates latent world-model prediction energy against paired real trajectories and uses the same signal for direct alignment of the final prediction and reward-guided optimization of intermediate stochastic transitions. *Third*, our experiments show improved Physics-IQ-Verified scores at four-step inference and stronger long-horizon dynamics on PLV-Bench, while controlled ablations examine the effects of the two supervision routes and training-data composition.

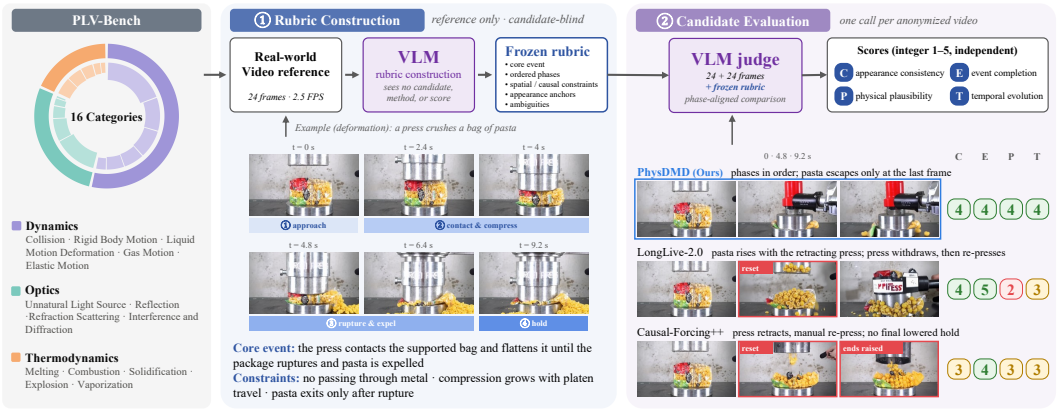

Figure 3: **Overview of PLV-Bench and the reference-guided evaluation pipeline.**

2 PLV-Bench: Benchmarking Physics in Video Distillation

We introduce PLV-Bench to evaluate sustained physical dynamics in distilled AR video generation. Beyond current benchmarks targeting text-only, short-term physical interactions (Meng et al., 2025; Bansal et al., 2026), PLV-Bench provides 125 real-world references spanning 16 phenomena to evaluate the *natural temporal evolution* of physical processes.

2.1 Problem Formulation

While video distillation mitigates heavy inference costs via few-step sampling, standard distribution-matching primarily prioritizes perceptual fidelity. Consequently, its impact on the *physical plausibility* of generated dynamics in distilled AR models remains underexplored, as matching a teacher's distribution does not inherently ensure faithful physical interactions. This motivates our inquiry: *Can we leverage distillation to recover or improve the physical plausibility of pretrained teachers?*

To rigorously investigate this, we first formulate the physical evaluation as an image-conditioned video continuation task. Given a real-world conditioning frame, the distilled autoregressive model sequentially generates subsequent frames based on its self-generated history. The withheld ground-truth continuation serves not as a rigid reconstruction target, but as an empirical reference for the observed timing and natural progression of the physical process.

2.2 Benchmark Curation

We collect candidate videos from WISA-80K (Wang et al., 2025), which contains 80K clips of diverse physical phenomena. Our curation pipeline combines automated filtering with PhD-level expert review. We first apply metadata and quality filters to retain clips of at least 10 seconds and exclude visually degraded videos. This duration supports evaluation of sustained dynamics beyond short physical interactions. We observe that many candidate videos depict physical phenomena in slow motion or at accelerated playback speeds. Accordingly, we use motion diagnostics to assess *temporal naturalness* and remove these manipulated clips to preserve the natural dynamics of the underlying processes. After deduplicating the remaining clips via perceptual hashes and DINOv2 features (Oquab et al., 2024), expert annotators review the final set to correct noisy original labels, assign hierarchical categories, and organize the taxonomy. We detail this curation pipeline in App. B.

2.3 Reference-Guided Evaluation of Long-Horizon Physical Dynamics

Standard visual metrics fail to accurately measure physical plausibility by penalizing benign timing shifts and masking true causal failures. We resolve this via a zero-shot VLM evaluator performing phase-level alignment across four dimensions: Appearance consistency, Event completion, Physical plausibility, and Temporal evolution. Instead of rigid frame-wise matching, the VLM employs reference-specific rubrics to align underlying semantic phases. This strictly isolates benign trajectory deviations from catastrophic structural or temporal collapses, yielding independent, visually grounded scores for each evaluation axis, with detailed descriptions provided in App. B.5.

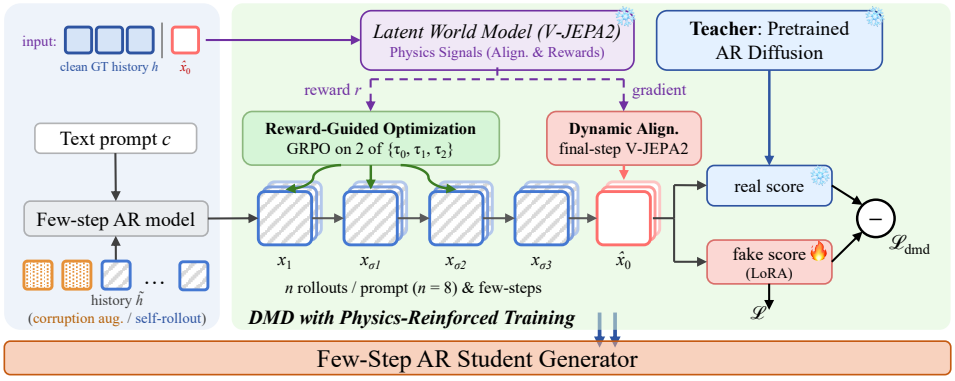

Figure 4: **Overview of PhysDMD.** Latent world-model rewards guide intermediate stochastic transitions, direct alignment supervises the final deterministic prediction, and DMD anchors the student to the teacher distribution.

3 Investigating Physics-Reinforced AR Diffusion Distillation

To address the observed physical inconsistencies, we investigate physics-reinforced training for autoregressive video diffusion distillation. We first review the preliminaries (Sec. 3.1), then explore direct dynamics alignment (Sec. 3.2) and reward-guided policy optimization (Sec. 3.3). Together, these investigations form our physics-reinforced distillation objective (Sec. 3.4).

3.1 Preliminaries: Autoregressive Video Generative Models

An AR video diffusion model factorizes a video into N temporal chunks $\mathbf{x}_0^{1:N}$ conditioned on c (e.g., a text prompt): $p_\theta(\mathbf{x}_0^{1:N} \mid c) = \prod_{i=1}^{N} p_\theta(\mathbf{x}_0^i \mid \mathbf{x}_0^{<i}, c)$. Each chunk may contain one or more frames. We model each conditional factor with diffusion or flow matching, perturbing chunk i as $\mathbf{x}_t^i = \alpha_t \mathbf{x}_0^i + \sigma_t \boldsymbol{\epsilon}^i$, where $\boldsymbol{\epsilon}^i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\alpha_t = 1 - t$, and $\sigma_t = t$. At inference, G_θ denoises each chunk from Gaussian noise conditioned on the generated history, retaining historical keys and values for causal attention.

Teacher Forcing (TF) (Gao et al., 2025) and Diffusion Forcing (DF) (Chen et al., 2024) differ in how they construct training histories. TF conditions a noisy frame $\mathbf{x}_t^i$ on clean ground-truth history $\mathbf{x}_0^{<i}$ to learn $p_{\text{data}}(\mathbf{x}_0^i \mid \mathbf{x}_0^{<i}, c)$. DF independently samples per-frame noise levels t_j and conditions $\mathbf{x}_{t_i}^i$ on $\mathbf{x}_\mathbf{t}^{<i} = (\mathbf{x}_{t_1}^1, \ldots, \mathbf{x}_{t_{i-1}}^{i-1})$, learning $p_{\text{DF}}(\mathbf{x}_0^i \mid \mathbf{x}_\mathbf{t}^{<i}, c)$. Both construct histories from real videos, whereas inference conditions on generated outputs. This discrepancy creates exposure bias: early errors alter later conditioning inputs and can accumulate across the rollout.

Another line of work combines autoregressive generation with diffusion distillation to construct efficient few-step causal video generators. Specifically, CausVid (Yin et al., 2025) initializes an AR student using ODE trajectories from a bidirectional teacher, and applies asymmetric DMD (Yin et al., 2024b). Formally, let $\widetilde{\mathbf{x}}_0 = G_\theta(\boldsymbol{\xi}; c)$ denote a student-generated video, where $\boldsymbol{\xi}$ collects the Gaussian noise sampled across all autoregressive steps. To align the student and teacher distributions, DMD re-noises the generated sample as $\widetilde{\mathbf{x}}_t = \alpha_t \widetilde{\mathbf{x}}_0 + \sigma_t \boldsymbol{\epsilon}'$ with $\boldsymbol{\epsilon}' \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and minimizes the time-averaged reverse KL divergence $\mathcal{L}_{\text{DMD}}(\theta) = \mathbb{E}_t[D_{\text{KL}}(p_{\theta,t} \,\|\, p_{\psi,t})]$ between the noise-perturbed student and teacher distributions. Thus, the generator is updated via the score-difference gradient:

$$\nabla_\theta \mathcal{L}_{\text{DMD}} \approx \mathbb{E}_{\boldsymbol{\xi},t,\boldsymbol{\epsilon}'} \left[w(t)\alpha_t \left(s_\phi(\widetilde{\mathbf{x}}_t, t, c) - s_\psi(\widetilde{\mathbf{x}}_t, t, c) \right)^\top \frac{\partial G_\theta(\boldsymbol{\xi}; c)}{\partial \theta} \right], \tag{1}$$

where the frozen teacher provides the target score s_ψ, and an auxiliary model trained online on student samples estimates the student score s_ϕ. By operating on generated samples rather than ground-truth contexts, DMD naturally supports autoregressive self-rollout. Leveraging this property, Self Forcing (Huang et al., 2026a) unrolls the student during training to condition each frame on prior outputs, applying a video-level DMD objective to the resulting sequence. Causal Forcing (Zhu et al., 2026) adopts a similar self-rollout approach but incorporates a causally aligned initialization; it distills the probability-flow ODE of an AR teacher into the student before DMD post-training.

3.2 Direct Dynamics Alignment via a Latent World Model

Although DMD aligns the student rollout distribution with that of the teacher, its score-based supervision remains agnostic to the physical plausibility of the generated dynamics. Consequently, physically implausible interactions in the teacher distribution can be inherited and further amplified through autoregressive rollout. This motivates our first investigation: whether a pretrained latent world model can provide a direct supervisory signal for physical evolution. We study this using V-JEPA2 (Assran et al., 2025), whose masked latent prediction objective captures predictive spatiotemporal structures beyond low-level appearance cues. We therefore use its latent prediction energy as a dynamics-aware supervision signal for student-generated transitions.

We evaluate this signal across each autoregressive chunk boundary. Given a conditioning history $\mathbf{h}_{\text{cond}}$, obtained from either ground-truth context or self-rollout, the student generates a target chunk $\hat{\mathbf{y}}_\theta$. For V-JEPA2 evaluation, we instead use the clean ground-truth history $\mathbf{h}_{\text{gt}}$ as a stable reference context. This fixes the evaluator context across candidates, separating errors in the conditioning history from the score assigned to the generated transition. Thus, given the corresponding ground-truth target $\mathbf{y}$, we employ the frozen decoder $\mathcal{D}$ to construct boundary-centered clips $\mathbf{v}^{\text{gen}} = [\mathcal{D}(\mathbf{h}_{\text{gt}}); \mathcal{D}(\hat{\mathbf{y}}_\theta)]$ and $\mathbf{v}^{\text{gt}} = [\mathcal{D}(\mathbf{h}_{\text{gt}}); \mathcal{D}(\mathbf{y})]$. We define the latent prediction energy $E_{\text{wm}}(\mathbf{v})$ as the feature-space discrepancy between the target representation and its prediction from the visible context.

A key challenge is that directly minimizing prediction energy can favor trivially predictable dynamics, including motion collapse. We therefore calibrate the generated transition against the intrinsic predictability of its paired ground-truth transition. Specifically, let $e_{\text{gen}} = E_{\text{wm}}(\mathbf{v}^{\text{gen}})$ and $e_{\text{gt}} = \text{sg}[E_{\text{wm}}(\mathbf{v}^{\text{gt}})]$. We define the world-model alignment objective $\mathcal{L}_{\text{WM}}$ via a symmetric Huber penalty applied to the ground-truth-normalized energy gap between e_{gen} and e_{gt}, as detailed in App. C.1. This formulation penalizes generated transitions with prediction energies that are either excessively high or unrealistically low relative to the paired real trajectory. During optimization, the decoder and V-JEPA2 remain frozen, and $\mathcal{L}_{\text{WM}}$ updates the student exclusively through the generated chunk. This provides direct dynamics supervision without backpropagating through the full autoregressive rollout.

3.3 Reward-Guided Policy Optimization with Latent World-Model Rewards

While direct dynamics alignment constrains the deterministic prediction, intermediate stochastic transitions remain susceptible to error accumulation. This motivates our investigation: whether terminal-step alignment alone is sufficient and how to steer intermediate denoising steps toward physically plausible trajectories. To this end, we formulate intermediate transition optimization as an RL problem and adopt GRPO, using the world-model alignment penalty as a physics-aware reward.

Specifically, for each paired training sample $(c, \mathbf{h}_{\text{gt}}, \mathbf{y})$, we generate K target-chunk candidates conditioned on a shared history $\mathbf{h}_{\text{cond}}$. We evaluate each candidate using the alignment objective from Sec. 3.2. For candidate k, we denote its world-model alignment penalty by $\ell_{\text{WM}}^{(k)}$ and define its reward as $r_{\text{WM}}^{(k)} = -\ell_{\text{WM}}^{(k)}$. We then compute the group-relative advantage:

$$A^{(k)} = \frac{r_{\text{WM}}^{(k)} - \bar{r}_{\text{WM}}}{\text{std}\left(\{r_{\text{WM}}^{(j)}\}_{j=1}^{K}\right) + \epsilon_{\text{adv}}}, \qquad \bar{r}_{\text{WM}} = \frac{1}{K}\sum_{j=1}^{K} r_{\text{WM}}^{(j)}. \tag{2}$$

GRPO propagates the terminal world-model reward to intermediate denoising steps without training a value model. We sample candidate trajectories under a reference rollout policy θ_{old}, and replay selected transitions under the current student policy θ to compute the likelihood ratio $q_{k,s}(\theta)$. The chunk-level advantage $A^{(k)}$ is then shared across the selected stochastic transitions $s \in \mathcal{S}$ of candidate k. We optimize these intermediate transitions using the clipped GRPO objective:

$$\mathcal{L}_{\text{GRPO}} = -\frac{1}{K|\mathcal{S}|}\sum_{k=1}^{K}\sum_{s\in\mathcal{S}} \min\Big(q_{k,s}(\theta)A^{(k)}, \text{clip}(q_{k,s}(\theta), 1-\varepsilon_-, 1+\varepsilon_+)\, A^{(k)}\Big), \tag{3}$$

Following few-step generation optimization (Huang et al., 2026b), we update only this selected subset of transitions, and share noise at the remaining steps across candidates to reduce reward variance.

3.4 Unified Physics-Reinforced Distillation

Synthesizing the preceding investigations, we formulate a unified objective with complementary supervision across the few-step generation trajectory.

Joint Objective. During training, we reuse the same historical context for DMD and GRPO. Since the final denoising transition is deterministic, we exclude it from policy optimization and directly apply the differentiable world-model alignment objective. We denote this terminal term by $\mathcal{L}_{\mathrm{WM}}^{\mathrm{term}}$ and optimize the student with:

$$\mathcal{L}_{\mathrm{total}} = \lambda_{\mathrm{DMD}}\mathcal{L}_{\mathrm{DMD}} + \lambda_{\mathrm{RL}}\mathcal{L}_{\mathrm{GRPO}} + \lambda_{\mathrm{WM}}\mathcal{L}_{\mathrm{WM}}^{\mathrm{term}}. \tag{4}$$

Crucially, gradients from $\mathcal{L}_{\mathrm{WM}}^{\mathrm{term}}$ propagate exclusively through the final denoising step, with preceding states and history detached. To reduce computational cost, we evaluate this term on a sampled subset of M out of K candidates (e.g., $M = 2, K = 8$).

Context Decoupling. To improve robustness to imperfect histories, we condition the student model on perturbed histories, while evaluating the world-model reward against the clean ground-truth history. This separates imperfect generation contexts from the reference used for physics-aware supervision.

4 Experiments and Analysis

4.1 Implementation Details

Training. We causally adapt a DiT-based latent flow model (Peebles & Xie, 2023; Wan et al., 2025) for chunk-wise AR generation and fine-tune it on curated high-quality data to obtain the *AR baseline*. Following CausVid (Yin et al., 2025) and Self Forcing (Huang et al., 2026a), we initialize a four-step student using 22K ODE solution trajectories from this baseline, with prompts drawn from LLM-refined OpenVid captions (Nan et al., 2025) for initialization and subsequent distillation. We train *Vanilla-DMD* with AdamW and a global batch size of 128, using learning rates of 2.0×10^{-6} for the generator and 4.0×10^{-7} for the fake-score model, with one generator update per five fake-score updates. We initialize PhysDMD from Vanilla-DMD and jointly optimize DMD, direct dynamics alignment, and reward-guided policy optimization. We sample $K = 8$ continuations per context, set both learning rates to 1.0×10^{-6}, and train in `bfloat16` with rank-256 LoRA on linear layers.

Evaluation. We evaluate on Physics-IQ-Verified (Motamed et al., 2026), which comprises 66 controlled real-world experiments with multi-view captures. Under its I2V protocol, models generate 5-second continuations at 24 FPS from a switch frame and a best-practice prompt. The benchmark calibrates alignment with a real continuation against variation between independent recordings (averaged across four runs). We also evaluate longer rollouts on PLV-Bench (Sec. 2); Table 1 reports the available scores. These include VBench (Huang et al., 2024) metrics (subject/background consistency, motion smoothness) alongside our reference-guided VLM evaluations via GPT-5.6-Sol (appearance consistency, event completion, physical plausibility, and temporal evolution).

4.2 Comparisons

We compare PhysDMD with seven open-source models: the bidirectional Wan-Video 2.1 (Wan et al., 2025); the AR models SANA-Video (Chen et al., 2026c) and Helios-Base (Yuan et al., 2026b);

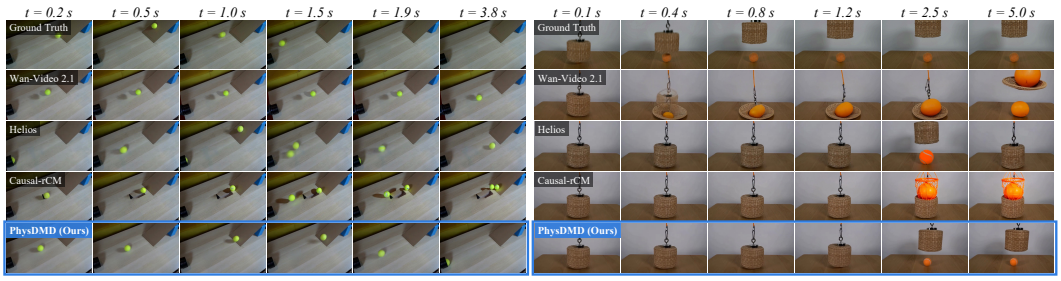

Figure 5: **Qualitative comparisons on Physics-IQ** against representative video generation models.

Table 1: **Quantitative comparisons on Physics-IQ-Verified and PLV-Bench.** Physics-IQ results are reported across four independent runs. Higher is better for all reported scores; **bold** and underline denote the best and second-best results, respectively. NFE denotes the number of function evaluations.

(a) Physics-IQ-Verified

Method	NFE	ST-IoU↑	Sp-IoU↑	wSp-IoU↑	MSE-based↑	Overall↑
Bidirectional Video Diffusion Models						
Wan-Video 2.1	50×2	22.88 ±1.64	**34.30** ±0.27	19.17 ±0.37	20.27 ±0.66	24.15 ±0.57
(Distilled) Autoregressive Video Diffusion Models						
SANA-Video	50×2	9.47 ±0.69	16.02 ±0.69	8.38 ±0.41	3.46 ±0.40	9.33 ±0.45
Helios	50×2	27.67 ±3.94	32.68 ±3.94	**20.68** ±2.41	28.36 ±1.00	27.35 ±1.01
CausVid	3	17.31 ±0.69	16.65 ±0.15	12.65 ±0.17	3.44 ±0.03	12.51 ±0.10
LongLive-2.0	4	14.40 ±0.49	33.48 ±1.08	18.14 ±0.54	19.82 ±0.42	21.46 ±0.28
Causal-Forcing++	2	23.19 ±1.61	32.81 ±0.68	18.48 ±0.30	16.58 ±0.31	22.76 ±0.35
Causal-rCM	4	22.62 ±1.04	32.22 ±1.13	19.95 ±0.51	18.66 ±0.97	23.36 ±0.55
PhysDMD	4	**29.42** ±0.99	31.13 ±0.47	19.61 ±1.24	**32.24** ±0.76	**28.10** ±0.34

(b) PLV-Bench

Method	NFE	Sub. Cons.	Bg. Cons.	Mot. Sm.	Event Comp.	Phys. Plaus.	Temp. Evol.	App. Cons.
SANA-Video	50×2	80.41	88.65	98.08	3.51	2.58	2.86	2.18
Helios	50×2	89.35	93.91	99.02	**3.92**	3.76	**3.56**	4.18
CausVid	3	79.23	87.09	98.39	3.02	2.70	2.63	2.02
LongLive-2.0	4	89.38	94.07	99.13	3.69	3.37	3.50	4.14
Causal-Forcing++	2	89.41	92.56	98.92	3.30	3.12	3.03	3.95
Causal-rCM	4	88.32	92.05	98.65	3.57	2.98	3.13	3.35
PhysDMD	4	**92.05**	**95.65**	**99.47**	3.76	**3.77**	3.42	**4.34**

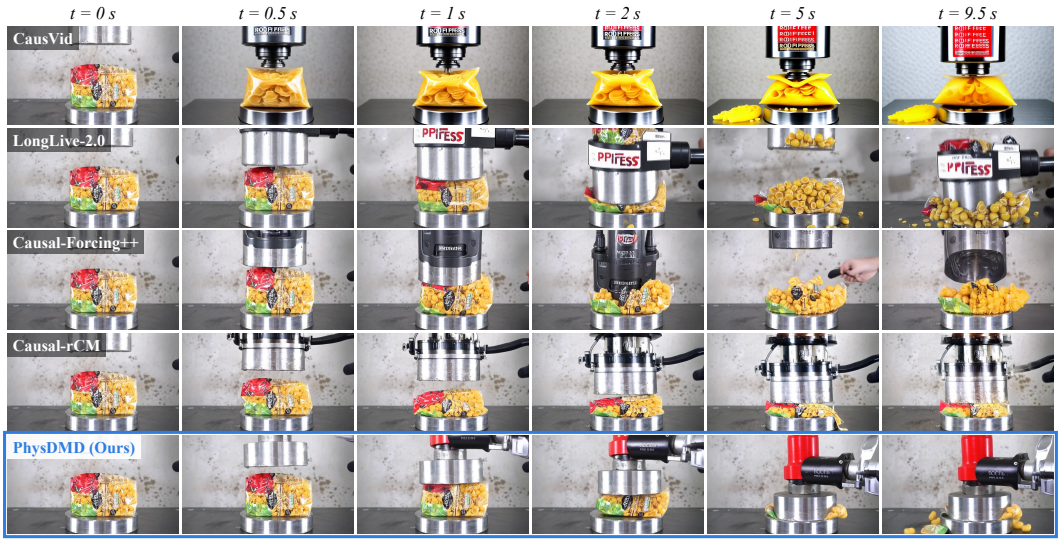

Figure 6: **Qualitative comparisons on PLV-Bench** against representative distilled AR models.

and the distilled AR models CausVid (Yin et al., 2025), LongLive-2.0 (Chen et al., 2026e), Causal-Forcing++ (Zhao et al., 2026b), and Causal-rCM (Zheng et al., 2026a).

Image-Conditioned Generation on Physics-IQ. In Table 1, PhysDMD achieves the highest Overall score among the compared models (28.10), improving on Causal-rCM (23.36) by 4.74 points. It also leads in ST-IoU and MSE-based scores, while other models retain higher Sp-IoU and wSp-IoU scores. These results indicate improved spatiotemporal alignment at four-step inference, with qualitative examples in Fig. 5. More qualitative results are provided in App. E.1.

Table 2: **Ablations on Physics-IQ-Verified.** The multi-step AR baseline is shown as a reference. Bold denotes the best mean among four-step variants. The final block uses relative-Huber alignment and GRPO; only the training-data composition changes. Higher is better.

Variant	NFE	ST-IoU↑	Sp-IoU↑	wSp-IoU↑	MSE-based↑	Overall↑
Reference and Distillation Baselines						
AR baseline	50×2	31.43 ±3.40	30.09 ±0.97	19.81 ±1.21	31.54 ±1.94	28.22 ±1.69
Vanilla-DMD	4	23.01 ±1.09	32.57 ±1.45	19.06 ±1.19	26.60 ±0.87	25.31 ±0.63
Matched-step DMD	4	24.80 ±0.86	32.62 ±2.82	19.43 ±2.08	27.12 ±0.87	25.99 ±1.39
Dynamics Alignment Only (Original Data)						
One-sided hinge	4	24.35 ±1.28	33.78 ±1.82	20.47 ±1.52	27.74 ±0.83	26.58 ±1.06
Relative Huber	4	24.24 ±0.79	**34.63** ±1.31	**20.93** ±0.97	27.73 ±0.70	26.88 ±0.74
Dynamics Alignment + GRPO: Training-Data Composition						
Original DMD data	4	**29.83** ±0.63	30.83 ±0.78	19.37 ±0.38	32.06 ±0.65	28.02 ±0.35
Filtered + hard (1:1)	4	27.47 ±1.31	31.27 ±0.32	19.46 ±0.49	30.57 ±0.59	27.19 ±0.42
PhysDMD (filtered)	4	29.42 ±0.99	31.13 ±0.47	19.61 ±1.24	**32.24** ±0.76	**28.10** ±0.34

Extended-Horizon Generation on PLV-Bench. PLV-Bench extends evaluation to longer real-world processes and diverse physical phenomena. In Table 1, PhysDMD achieves the best subject/background consistency and motion smoothness, while attaining the best physical plausibility (3.77) and event completion (3.76) among distilled AR models. Fig. 6 illustrates representative rollouts; additional results are provided in App. E.2.

4.3 Ablation Studies

Table 2 examines the training objectives and data composition on Physics-IQ-Verified; additional qualitative results appear in App. E.

Can direct dynamics alignment improve physical plausibility? Matched-step DMD, which simply continues DMD training, increases the Overall score from 25.31 to 25.99. With direct alignment, the one-sided hinge and relative-Huber variants reach 26.58 and 26.88, respectively. Relative Huber improves the spatial scores over the hinge variant, while their ST-IoU and MSE-based scores are similar. The strongest gains from relative-Huber alignment are therefore in spatial overlap.

Does policy optimization complement terminal alignment? With the original data composition, adding GRPO to relative-Huber alignment raises the Overall score from 26.88 to 28.02. The full PhysDMD model reaches 28.10, close to the multi-step AR baseline's 28.22, while retaining four-step sampling. The largest changes are in ST-IoU and MSE-based scores, consistent with complementary supervision of intermediate transitions.

4.4 Analysis: Impact of Training-Data Composition

We compare data compositions under the same joint objective (App. D.1). The original and filtered data yield similar Overall scores (28.02 and 28.10), whereas mixing filtered and hard samples at a 1:1 ratio yields 27.19. Thus, the tested hard-sample mixture provides no improvement, and filtering has a modest effect on the reported mean. Within these configurations, the hard-sample mixture has a larger effect than the difference between the original and filtered data.

5 Related Work

5.1 Autoregressive Long Video Models

Current bidirectional video diffusion models (Wan et al., 2025; Wu et al., 2025) yield impressive capabilities in generating detailed short clips. To scale them to longer generation, recent studies have shifted toward autoregressive rollouts. Early AR methods typically rely on autoregressive Transformers to generate discrete spatiotemporal tokens sequentially at inference time (Yan et al., 2021; Kondratyuk et al., 2024). Subsequent methods combine AR factorization with continuous

denoising diffusion or flow matching (Jin et al., 2025; Gu et al., 2025), where frames or chunks are generated sequentially in an outer loop and iteratively denoised in an inner loop. To mitigate the exposure bias in rollouts, Diffusion Forcing (Chen et al., 2024) trains on sequences with independent per-frame noise levels, a formulation further augmented via variable-length history guidance (Song et al., 2025). To scale this paradigm, recent architectures (Teng et al., 2025; Chen et al., 2025; Kodaira et al., 2026; Yuan et al., 2026b) utilize chunk-wise AR denoising to support minute-scale and streaming generation. Furthermore, parallel efforts preserve coherent historical context via explicit context-window management (Guo et al., 2025; Zhang et al., 2026a) or training-free memory and positional adaptation (Li et al., 2026a;b; Yesiltepe et al., 2026; Kim et al., 2026).

5.2 Diffusion Distillation

To reduce the latency of iterative sampling, extensive studies distill multi-step diffusion models into efficient few-step generators via trajectory- or distribution-based formulations. Trajectory-based methods either approximate teacher transitions (Salimans & Ho, 2022) or enforce consistency along probability-flow ODEs (Luo et al., 2023; Lu & Song, 2025). Advancing this paradigm to large-scale image and video models, rCM (Zheng et al., 2026b) augments continuous-time consistency with score regularization, while AnyFlow (Gu et al., 2026b) combines flow-map learning with on-policy distribution matching for flexible any-step generation. Conversely, distribution-based methods align the student and target distributions at the generative level. DMD (Yin et al., 2024b) successfully scales score distillation to large-scale text-to-image generation. DMD2 (Yin et al., 2024a) removes costly regression supervision and improves distribution matching with stabilized score estimation and adversarial training. Subsequent variants further expand distribution matching through improved training objectives (Xu et al., 2025; Chen et al., 2026a; Liu et al., 2026a), phase-wise multi-step distillation (Fan et al., 2026), and hybrid trajectory–distribution matching (Luo et al., 2025).

5.3 Autoregressive Video Diffusion Distillation

An emerging line of research integrates autoregressive video generation with diffusion distillation for efficient, few-step causal synthesis. CausVid (Yin et al., 2025) establishes this paradigm by distilling a pretrained bidirectional diffusion model into a causal student via Distribution Matching Distillation. Within this family, Self Forcing (Huang et al., 2026a) reduces the train–inference gap by replacing ground-truth conditioning with autoregressive self-rollouts, enabling distillation on self-generated trajectories. To push temporal boundaries, LongLive (Yang et al., 2026) applies streaming DMD to newly generated clips during long autoregressive rollouts, while Rolling Forcing (Liu et al., 2026b) optimizes extended denoising windows conditioned on self-generated histories. Concurrently, Self-Forcing++ (Cui et al., 2026) supervises sampled segments from self-generated long videos, extrapolating generation beyond the teacher's original temporal horizon. Beyond horizon extension, more recent work refines the underlying distillation mechanism: Causal Forcing and its extension (Zhu et al., 2026; Zhao et al., 2026b) resolve teacher–student architectural mismatches during causal student initialization, Context Forcing (Chen et al., 2026d) aligns long-context students with long-context teacher guidance, and Reward Forcing (Lu et al., 2026) introduces reward-weighted distribution matching to improve motion dynamics. WMReward (Yuan et al., 2026a) uses latent world-model rewards for inference-time search and steering. PhysDMD instead incorporates reference-calibrated world-model supervision during distillation, combining direct terminal alignment with policy updates to intermediate transitions; the world model is not used at inference.

6 Conclusion

We presented PLV-Bench, an image-to-video benchmark with real-world references for studying sustained physical dynamics, and PhysDMD, a physics-reinforced distillation framework for few-step AR video generation. By calibrating latent world-model prediction energy against paired real trajectories, PhysDMD provides direct supervision for the final deterministic prediction and reward guidance for intermediate stochastic transitions. On Physics-IQ-Verified, it improves the Overall score from 25.31 to 28.10 over Vanilla-DMD, approaching the multi-step AR baseline while retaining four-step sampling and requiring no inference-time world model. These results support incorporating reference-calibrated dynamics supervision into video distillation; the learned energy remains a proxy for physical plausibility rather than a guarantee of physical correctness.

AI USE STATEMENT

In this work, we used generative AI tools for grammar checking and language polishing to improve the readability of the manuscript. We also used GPT-5.6 Sol as an evaluation instrument for generated videos, as described in the evaluation protocol and experiments. We did not use generative AI tools to originate the research ideas, design the proposed methodology, or determine the experimental setup. All AI-assisted writing was reviewed by the authors to ensure that the original scientific meaning was preserved and that no unintended claims or factual inaccuracies were introduced. The AI-generated evaluation outputs were reviewed for consistency with the predefined scoring protocol. We take responsibility for the final content of this work, including all text, experimental claims, evaluation results, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

Our benchmark construction involves human annotation to verify physical phenomena, temporal naturalness, and category labels in real-world videos. Annotators follow predefined review criteria, and detailed annotation protocols. The annotation process focuses exclusively on physical content and does not require annotators to infer sensitive personal attributes. The videos are sourced from existing public video datasets, and we use them solely for research and video models evaluation.

REPRODUCIBILITY STATEMENT

We provide detailed descriptions of our method, training procedure, and evaluation protocol in the main paper and Appendix. The Appendix further documents implementation settings, benchmark construction, data processing, and annotation procedures. These details are intended to facilitate reproduction of our experiments and future research on physics-reinforced video distillation.

REFERENCES

Niket Agarwal, Arslan Ali, Jon Allen, Martin Antolini, Adeline Aubame, Alisson Azzolini, Junjie Bai, Maciej Bala, Yogesh Balaji, Josh Bapst, et al. Cosmos 3: Omnimodal world models for physical ai. *arXiv preprint arXiv:2606.02800*, 2026.

Mido Assran, Adrien Bardes, David Fan, Quentin Garrido, Russell Howes, Matthew Muckley, Ammar Rizvi, Claire Roberts, Koustuv Sinha, Artem Zholus, et al. V-jepa 2: Self-supervised video models enable understanding, prediction and planning. *arXiv preprint arXiv:2506.09985*, 2025.

Hritik Bansal, Clark Peng, Yonatan Bitton, Roman Goldenberg, Aditya Grover, and Kai-Wei Chang. Videophy-2: A challenging action-centric physical commonsense evaluation in video generation. In *International Conference on Learning Representations*, volume 2026, pp. 118456–118470, 2026.

Black Forest Labs. FLUX.2: Next generation image generation. `https://bfl.ai/models/flux-2`, 2025.

Huanqia Cai, Sihan Cao, Ruoyi Du, Peng Gao, Aiming Hao, Steven Hoi, Zhaohui Hou, Shijie Huang, Dengyang Jiang, Yuming Jiang, et al. Z-image: An efficient image generation foundation model with single-stream diffusion transformer. *arXiv preprint arXiv:2511.22699*, 2025.

Boyuan Chen, Diego Martí Monsó, Yilun Du, Max Simchowitz, Russ Tedrake, and Vincent Sitzmann. Diffusion forcing: Next-token prediction meets full-sequence diffusion. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 24081–24125. Curran Associates, Inc., 2024. doi: 10.52202/079017-0759. URL `https://proceedings.neurips.cc/paper_files/paper/2024/file/2aee1c4159e48407d68fe16ae8e6e49e-Paper-Conference.pdf`.

Guanjie Chen, Shirui Huang, Yifu Sun, Kai Liu, Jianchen Zhu, Xiaoye Qu, Yu Cheng, and Peng Chen. Flash-dmd: Towards high-fidelity few-step image generation with efficient distillation and joint reinforcement learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6010–6020, 2026a.

Guibin Chen, Dixuan Lin, Jiangping Yang, Chunze Lin, Junchen Zhu, Mingyuan Fan, Hao Zhang, Sheng Chen, Zheng Chen, Chengcheng Ma, et al. Skyreels-v2: Infinite-length film generative model. *arXiv preprint arXiv:2504.13074*, 2025.

Haoyu Chen, Kaichen Zhou, Hang Hua, Kaile Zhang, Jingwen Qian, Wufei Ma, Haonan Chen, Chunjiang Liu, Yizhou Zhao, Xiaoyuan Wang, et al. Memobench: Benchmarking world modeling in dynamically changing environments. In *European Conference on Computer Vision*, pp. 658–676. Springer, 2026b.

Junsong Chen, Yuyang Zhao, Jincheng Yu, Ruihang Chu, Junyu Chen, Shuai Yang, Xianbang Wang, Yicheng Pan, Zhou Daquan, Huan Ling, et al. Sana-video: Efficient video generation with block linear diffusion transformer. In *International Conference on Learning Representations*, volume 2026, pp. 38973–38997, 2026c.

Shuo Chen, Cong Wei, Sun Sun, Tiancheng SHEN, Ping Nie, Kai Zou, Ge Zhang, Ming-Hsuan Yang, and Wenhu Chen. Context forcing: Consistent autoregressive video generation with long context. In *Forty-third International Conference on Machine Learning*, 2026d. URL https://openreview.net/forum?id=lv86f7Qr9V.

Yukang Chen, Luozhou Wang, Wei Huang, Shuai Yang, Bohan Zhang, Yicheng Xiao, Ruihang Chu, Weian Mao, Qixin Hu, Shaoteng Liu, et al. Longlive-2.0: An nvfp4 parallel infrastructure for long video generation. *arXiv preprint arXiv:2605.18739*, 2026e.

Jiaxing 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. In *International Conference on Learning Representations*, volume 2026, pp. 85802–85822, 2026.

Linwei Dong, Ruoyu Guo, Ge Bai, Zehuan Yuan, Yawei Luo, and Changqing Zou. Guiding distribution matching distillation with gradient-based reinforcement learning. *arXiv preprint arXiv:2604.19009*, 2026.

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. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 41667–41676, 2026.

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 *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=YbtH1aoE1V.

Quentin Garrido, Nicolas Ballas, Mahmoud Assran, Adrien Bardes, Laurent Najman, Michael Rabbat, Emmanuel Dupoux, and Yann LeCun. Intuitive physics understanding emerges from self-supervised pretraining on natural videos. *arXiv preprint arXiv:2502.11831*, 2025.

Google DeepMind. Introducing gemini omni. https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-omni/, 2026a.

Google DeepMind. Genie 3: A new frontier for world models, 2026b. URL https://deepmind.google/models/genie/.

Jing Gu, Xian Liu, Yu Zeng, Ashwin Nagarajan, Fangrui Zhu, Daniel Hong, Yue Fan, Qianqi Yan, Kaiwen Zhou, Ming-Yu Liu, et al. $phyworldbench$: A comprehensive evaluation of physical realism in text-to-video models. In *International Conference on Learning Representations*, volume 2026, pp. 75130–75164, 2026a.

Yuchao Gu, Weijia Mao, and Mike Zheng Shou. Long-context autoregressive video modeling with next-frame prediction. *arXiv preprint arXiv:2503.19325*, 2025.

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. *arXiv preprint arXiv:2605.13724*, 2026b.

Yuwei Guo, Ceyuan Yang, Ziyan Yang, Zhibei Ma, Zhijie Lin, Zhenheng Yang, Dahua Lin, and Lu Jiang. Long context tuning for video generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 17281–17291, 2025.

Adam W Harley, Yang You, Xinglong Sun, Yang Zheng, Nikhil Raghuraman, Yunqi Gu, Sheldon Liang, Wen-Hsuan Chu, Achal Dave, Suya You, et al. Alltracker: Efficient dense point tracking at high resolution. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5253–5262. IEEE, 2025.

Xun Huang, Zhengqi Li, Guande He, Mingyuan Zhou, and Eli Shechtman. Self forcing: Bridging the train-test gap in autoregressive video diffusion. *Advances in Neural Information Processing Systems*, 38:167283–167308, 2026a.

Yushi Huang, Xiangxin Zhou, Ruoyu Wang, Chi Zhang, Jun Zhang, and Tianyu Pang. Reinforcing few-step generators via reward-tilted distribution matching. *arXiv preprint arXiv:2605.26108*, 2026b.

Ziqi Huang, Yinan He, Jiashuo Yu, Fan Zhang, Chenyang Si, Yuming Jiang, Yuanhan Zhang, Tianxing Wu, Qingyang Jin, Nattapol Chanpaisit, et al. Vbench: Comprehensive benchmark suite for video generative models. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21807–21818. IEEE, 2024.

Ziqi Huang, Fan Zhang, Xiaojie Xu, Yinan He, Jiashuo Yu, Ziyue Dong, Qianli Ma, Nattapol Chanpaisit, Chenyang Si, Yuming Jiang, et al. Vbench++: Comprehensive and versatile benchmark suite for video generative models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Dengyang Jiang, Dongyang Liu, Zanyi Wang, Qilong Wu, Liuzhuozheng Li, Hengzhuang Li, Xin Jin, David Liu, Changsheng Lu, Zhen Li, et al. Distribution matching distillation meets reinforcement learning. *arXiv preprint arXiv:2511.13649*, 2025.

Yuxin Jiang, Yuchao Gu, Ivor Tsang, and Mike Zheng Shou. Olaf-world: Orienting latent actions for video world modeling. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=5TiuerrwR8.

Yang Jin, Zhicheng Sun, Ningyuan Li, Kun Xu, Hao Jiang, Nan Zhuang, Quzhe Huang, Yang Song, Yadong Mu, and Zhouchen Lin. Pyramidal flow matching for efficient video generative modeling. In *International Conference on Learning Representations*, volume 2025, pp. 23378–23402, 2025.

Youngrae Kim, Qixin Hu, C-C Jay Kuo, and Peter A Beerel. Memrope: Training-free infinite video generation via evolving memory tokens. *arXiv preprint arXiv:2603.12513*, 2026.

Akio Kodaira, Tingbo Hou, Ji Hou, Markos Georgopoulos, Felix Juefei-Xu, Masayoshi Tomizuka, and Yue Zhao. Streamdit: Real-time streaming text-to-video generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 29200–29210, 2026.

Dan Kondratyuk, Lijun Yu, Xiuye Gu, Jose Lezama, Jonathan Huang, Grant Schindler, Rachel Hornung, Vighnesh Birodkar, Jimmy Yan, Ming-Chang Chiu, Krishna Somandepalli, Hassan Akbari, Yair Alon, Yong Cheng, Joshua V. Dillon, Agrim Gupta, Meera Hahn, Anja Hauth, David Hendon, Alonso Martinez, David Minnen, Mikhail Sirotenko, Kihyuk Sohn, Xuan Yang, Hartwig Adam, Ming-Hsuan Yang, Irfan Essa, Huisheng Wang, David A Ross, Bryan Seybold, and Lu Jiang. Videopoet: A large language model for zero-shot video generation. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=LRkJwPIDuE.

Kuaishou. Kling ai. https://kling.ai/, 2026. Official website.

Haodong Li, Shaoteng Liu, Zhe Lin, and Manmohan Chandraker. Rolling sink: Bridging limited-horizon training and open-ended testing in autoregressive video diffusion. *arXiv preprint arXiv:2602.07775*, 2026a.

Jia Li, Xiaomeng Fu, Xurui Peng, Weifeng Chen, Youwei Zheng, Tianyu Zhao, Jiexi Wang, Fangmin Chen, Xing Wang, and Hayden Kwok-Hay So. Train short, inference long: Training-free horizon extension for autoregressive video generation. *arXiv preprint arXiv:2602.14027*, 2026b.

Lin Li, Qihang Zhang, Yiming Luo, Shuai Yang, Ruilin Wang, Luyao Zhang, Mingrui Yu, Zelin Gao, Nan Xue, Boyu Zhou, Xing Zhu, Mingyu Ding, Yujun Shen, and Yinghao Xu. Causal World Modeling for Robot Control. In *Proceedings of Robotics: Science and Systems*, Sydney, Australia, July 2026c. doi: 10.15607/RSS.2026.XXII.016.

Dongyang Liu, Gao Peng, David Liu, DU DU, Zhen Li, Qilong Wu, Xin Jin, Sihan Cao, Shifeng Zhang, Steven HOI, et al. Decoupled dmd: Cfg augmentation as the spear, distribution matching as the shield. In *International Conference on Learning Representations*, volume 2026, pp. 140643–140666, 2026a.

Kunhao Liu, Wenbo Hu, Jiale Xu, Ying Shan, and Shijian Lu. Rolling forcing: Autoregressive long video diffusion in real time. In *International Conference on Learning Representations*, volume 2026, pp. 91177–91196, 2026b.

Yaofang Liu, Xiaodong Cun, Xuebo Liu, Xintao Wang, Yong Zhang, Haoxin Chen, Yang Liu, Tieyong Zeng, Raymond Chan, and Ying Shan. Evalcrafter: Benchmarking and evaluating large video generation models. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22139–22149. IEEE, 2024.

Cheng Lu and Yang Song. Simplifying, stabilizing and scaling continuous-time consistency models. In *International Conference on Learning Representations*, volume 2025, pp. 50611–50649, 2025.

Yunhong Lu, Yanhong Zeng, Haobo Li, Hao Ouyang, Qiuyu Wang, Ka Leong Cheng, Jiapeng Zhu, Hengyuan Cao, Zhipeng Zhang, Xing Zhu, et al. 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.

Simian Luo, Yiqin Tan, Longbo Huang, Jian Li, and Hang Zhao. Latent consistency models: Synthesizing high-resolution images with few-step inference. *arXiv preprint arXiv:2310.04378*, 2023.

Yihong Luo, Tianyang Hu, Jiacheng Sun, Yujun Cai, and Jing Tang. Learning few-step diffusion models by trajectory distribution matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025.

Fanqing Meng, Jiaqi Liao, Xinyu Tan, Quanfeng Lu, Wenqi Shao, Kaipeng Zhang, Yu Cheng, Dianqi Li, and Ping Luo. Towards world simulator: Crafting physical commonsense-based benchmark for video generation. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=dIjMswSzgF.

Zichen Miao, Zhengyuan Yang, Kevin Lin, Ze Wang, Zicheng Liu, Lijuan Wang, and Qiang Qiu. Tuning timestep-distilled diffusion model using pairwise sample optimization. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 13809–13830, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/243697ace81f57daef8737ff2c5cffd3-Paper-Conference.pdf.

MiniMax. Minimax h3: An open model breaking the boundaries between tasks and modalities. https://www.minimax.io/blog/minimax-h3, 2026. Official model release.

Saman Motamed, Laura Culp, Kevin Swersky, Priyank Jaini, and Robert Geirhos. Do generative video models understand physical principles? In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2026.

Kepan Nan, Rui Xie, Penghao Zhou, Tiehan Fan, Zhenheng Yang, Zhijie Chen, Xiang Li, Jian Yang, and Ying Tai. Openvid-1m: A large-scale high-quality dataset for text-to-video generation. In *International Conference on Learning Representations*, volume 2025, pp. 1045–1064, 2025.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel HAZIZA, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Herve Jegou, Julien Mairal, Patrick

Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=a68SUt6zFt. Featured Certification.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4172–4182. IEEE, 2023.

Yuxi Ren, Xin Xia, Yanzuo Lu, Jiacheng Zhang, Jie Wu, Pan Xie, Xing Wang, and Xuefeng Xiao. Hyper-sd: Trajectory segmented consistency model for efficient image synthesis. *Advances in neural information processing systems*, 37:117340–117362, 2024.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=TIdIXIpzhoI.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Kiwhan Song, Boyuan Chen, Max Simchowitz, Yilun Du, Russ Tedrake, and Vincent Sitzmann. History-guided video diffusion. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=j8Vr3E3vhy.

Kaiyue Sun, Kaiyi Huang, Xian Liu, Yue Wu, Zihan Xu, Zhenguo Li, and Xihui Liu. T2v-compbench: A comprehensive benchmark for compositional text-to-video generation. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8406–8416. IEEE, 2025.

Junshu Tang, Jiacheng Liu, Jiaqi Li, Longhuang Wu, Haoyu Yang, Penghao Zhao, Siruis Gong, Xiang Yuan, Shuai Shao, Linfeng Zhang, et al. Hunyuan-gamecraft-2: Instruction-following interactive game world model. *arXiv preprint arXiv:2511.23429*, 2025.

Hansi Teng, Hongyu Jia, Lei Sun, Lingzhi Li, Maolin Li, Mingqiu Tang, Shuai Han, Tianning Zhang, WQ Zhang, Weifeng Luo, et al. Magi-1: Autoregressive video generation at scale. *arXiv preprint arXiv:2505.13211*, 2025.

Team Wan, Ang Wang, Baole Ai, Bin Wen, Chaojie Mao, Chen-Wei Xie, Di Chen, Feiwu Yu, Haiming Zhao, Jianxiao Yang, et al. Wan: Open and advanced large-scale video generative models. *arXiv preprint arXiv:2503.20314*, 2025.

Jing Wang, Ao Ma, Ke Cao, Jun Zheng, Jiasong Feng, Zhanjie Zhang, Wanyuan Pang, and Xiaodan Liang. WISA: World simulator assistant for physics-aware text-to-video generation. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=4jWuS5hye1.

Bing Wu, Chang Zou, Changlin Li, Duojun Huang, Fang Yang, Hao Tan, Jack Peng, Jianbing Wu, Jiangfeng Xiong, Jie Jiang, et al. Hunyuanvideo 1.5 technical report. *arXiv preprint arXiv:2511.18870*, 2025.

Tianhe Wu, Ruibin Li, Lei Zhang, and Kede Ma. Diversity-preserved distribution matching distillation for fast visual synthesis. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=R3JzLU2qZj.

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

Wilson Yan, Yunzhi Zhang, Pieter Abbeel, and Aravind Srinivas. Videogpt: Video generation using vq-vae and transformers. *arXiv preprint arXiv:2104.10157*, 2021.

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. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 90037–90059, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/91a1610c6ed9e02d33f826b46f472b92-Paper-Conference.pdf.

Hidir Yesiltepe, Tuna Meral, Adil Kaan Akan, Kaan Oktay, and Pinar Yanardag. Infinity-rope: Action-controllable infinite video generation emerges from autoregressive self-rollout. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 40256–40265, 2026.

Tianwei Yin, Michaël Gharbi, Taesung Park, Richard Zhang, Eli Shechtman, Fredo Durand, and William T Freeman. Improved distribution matching distillation for fast image synthesis. *Advances in neural information processing systems*, 37:47455–47487, 2024a.

Tianwei Yin, Michaël Gharbi, Richard Zhang, Eli Shechtman, Fredo Durand, William T Freeman, and Taesung Park. One-step diffusion with distribution matching distillation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6613–6623. IEEE, 2024b.

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 *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22963–22974. IEEE, 2025.

Jianhao Yuan, Xiaofeng Zhang, Felix Friedrich, Nicolas Beltran-Velez, Melissa Hall, Reyhane Askari-Hemmat, Xiaochuang Han, Nicolas Ballas, Michal Drozdzal, and Adriana Romero-Soriano. Inference-time physics alignment of video generative models with latent world models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026a.

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*, 2026b.

Yu Yuan, Xijun Wang, Tharindu Wickremasinghe, Zeeshan Nadir, Bole Ma, and Stanley Chan. Newtongen: Physics-consistent and controllable text-to-video generation via neural newtonian dynamics. In *International Conference on Learning Representations*, volume 2026, pp. 12847–12878, 2026c.

Lvmin Zhang, Shengqu Cai, Muyang Li, Gordon Wetzstein, and Maneesh Agrawala. Frame context packing and drift prevention in next-frame-prediction video diffusion models. *Advances in Neural Information Processing Systems*, 38:30546–30566, 2026a.

Xiangdong Zhang, Jiaqi Liao, Shaofeng Zhang, Fanqing Meng, Xiangpeng Wan, Junchi Yan, and Yu Cheng. Videorepa: Learning physics for video generation through relational alignment with foundation models. *Advances in Neural Information Processing Systems*, 38:122647–122676, 2026b.

Bing Zhao, Chenfei Wu, Deqing Li, Hao Meng, Jiahao Li, Jie Zhang, Jingren Zhou, Junyang Lin, Kaiyuan Gao, Kuan Cao, et al. Qwen-image-2.0 technical report. *arXiv preprint arXiv:2605.10730*, 2026a.

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*, 2026b.

Dian Zheng, Ziqi Huang, Hongbo Liu, Kai Zou, Yinan He, Fan Zhang, Lulu Gu, Yuanhan Zhang, Jingwen He, Wei-Shi Zheng, et al. Vbench-2.0: Advancing video generation benchmark suite for intrinsic faithfulness. *arXiv preprint arXiv:2503.21755*, 2025.

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*, 2026a.

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*, volume 2026, pp. 2582–2603, 2026b.

Xiangqing Zheng, Chengyue Wu, Kehai Chen, and Min Zhang. Locot2v-bench: Benchmarking long-form and complex text-to-video generation. In *Forty-third International Conference on Machine Learning*, 2026c. URL https://openreview.net/forum?id=F22sU0Amh1.

Fengzhe Zhou, Jiannan Huang, Jialuo Li, Deva Ramanan, and Humphrey Shi. Pai-bench: A comprehensive benchmark for physical ai. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21522–21536, 2026.

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 *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=BYInOck3gr.

A Extended Related Works

A.1 Video Generation Benchmark

Although proprietary video models (Google DeepMind, 2026a; Kuaishou, 2026) exhibit striking quality, their opacity makes open-source alternatives essential for reproducible evaluation. Accelerated by this shift, extensive studies have established benchmarks to evaluate text-to-video (T2V) generation. Early efforts decompose perceptual quality into fine-grained dimensions, evaluating visual fidelity, temporal coherence, and compositional alignment (Huang et al., 2024; Liu et al., 2024; Sun et al., 2025). As capabilities scale to longer sequences, recent evaluations shift toward complex temporal structures, targeting multi-scene narrative coherence (Huang et al., 2025; Zheng et al., 2026c) and long-term memory consistency (Chen et al., 2026b). Beyond perceptual evaluation, recent works increasingly emphasize physical plausibility, prompting the development of physics-aware generation models (Zhang et al., 2026b; Yuan et al., 2026c) and targeted benchmarks. Specifically, existing benchmarks assess models' adherence to physical commonsense and dynamic realism in short, text-driven events (Meng et al., 2025; Bansal et al., 2026; Gu et al., 2026a), or ground evaluations in real-world experiments testing fundamental principles (Motamed et al., 2026; Zhou et al., 2026).

B Additional Details of PLV-Bench

B.1 Source Pool and Automated Screening

We initialize the candidate pool with 8,124 clips from WISA-80K (Wang et al., 2025), retaining only videos with a duration of at least 10 seconds. We decode sampled frames and apply the automated screening criteria listed in Table 3. Clips with missing metrics or failing any criterion are removed, leaving 3,777 clips.

Table 3: Automated screening criteria for the initial candidate pool.

Criterion	**Acceptance Rule**
Duration	≥ 10 s
Decoded frames	≥ 81
Bits per pixel per frame	≥ 0.03
Bit rate	≥ 0.5 Mb/s
Mean noise estimate	≤ 10
Mean blocking score	≤ 8
Mean sharpness score	≥ 80
Brightness / blur	Normal / sharp
Motion	Optimal, medium, static, or camera-driven

B.2 Two-Stage Deduplication

We perform two-stage deduplication to remove redundant and near-duplicate videos.

Perceptual Hashing. We compute dHash for eight uniformly sampled frames per video. Clips are linked if their mean frame-wise similarity is ≥ 0.92 and at least four temporally aligned frames exhibit a Hamming distance ≤ 10. We construct a graph over matched pairs, identify connected components, and retain the highest-quality clip from each component. This stage reduces the pool from 3,777 to 2,764 clips.

Semantic Deduplication. We extract DINOv2 (ViT-L/14) features across 32 sampled frames. Candidates are coarsely retrieved using a 0.65 video-level cosine similarity threshold, then verified as near-duplicates if they satisfy both video-level (≥ 0.70) and aligned frame-level (≥ 0.60) similarity criteria. Applying identical graph-based pruning yields the final 1,526 high-quality clips.

B.3 Motion Diagnostics and Human Verification

Fine-Grained Motion Profiling. To capture detailed temporal dynamics independent of the initial Farneback screening, we compute RAFT optical flow statistics across one- and five-frame offsets for the remaining 1,526 clips. This categorizes the videos into `optimal_motion` (307), `medium_quality` (380), `static` (385), `shaky` (395), and `camera_driven` (59).

Diverse Sampling & Human Verification. To minimize redundancy while preserving physical and kinematic diversity, we sample 400 representative clips based on decoupled video attributes and feature-space coverage. Finally, annotators review these clips via a custom web interface (Fig. 7) to verify event validity and taxonomic accuracy. Annotators also document primary and secondary physical phenomena with textual justifications, and rigorously filter temporal artifacts such as abnormal speeds, reversals, loops, or discontinuous cuts.

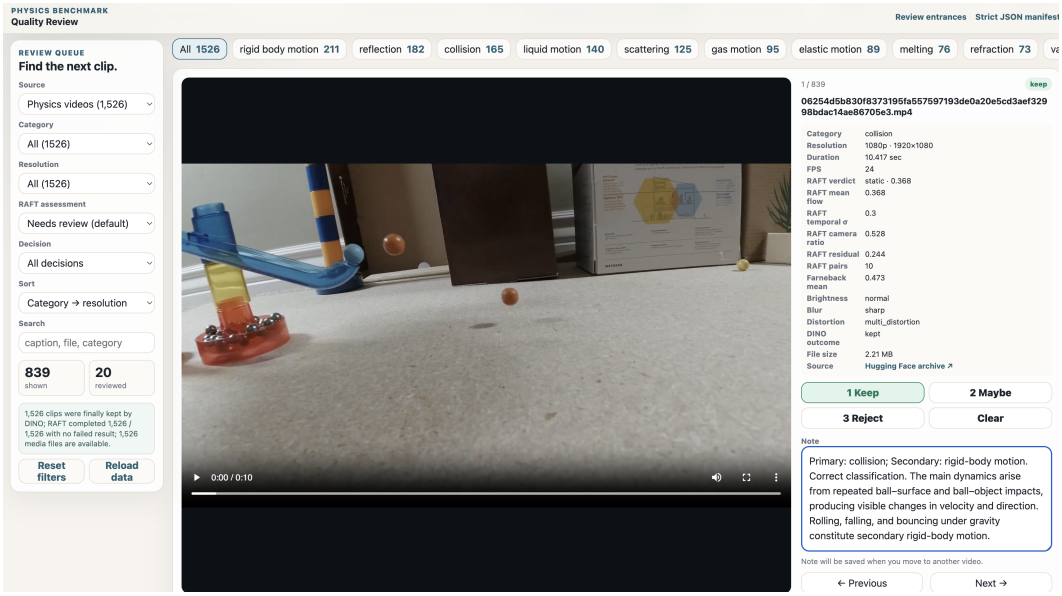

Figure 7: Our web-based interface for human verification. Annotators verify the physics category of each video, identify the primary and secondary physical phenomena, and provide a short justification.

B.4 Final Benchmark Taxonomy and Distribution

Following rigorous human verification, we construct the final PLV-Bench using a curated set of 125 high-quality videos. Fig. 8 illustrates the benchmark's hierarchical taxonomy.

B.5 Reference-Guided VLM Evaluation

We detail the zero-shot VLM evaluation protocol introduced in Sec. 2.3, which assesses long-horizon physical dynamics without requiring frame-wise temporal synchronization.

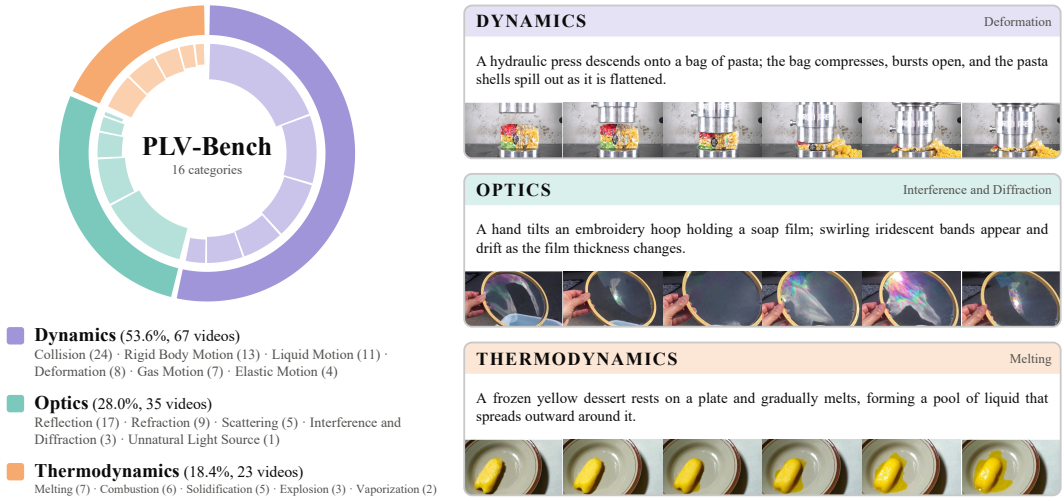

Figure 8: **Overview of PLV-Bench.** *Left:* A chart maps the 125 curated videos into 16 physical subcategories across Dynamics, Optics, and Thermodynamics. *Right:* Representative video sequences and textual annotations illustrate select physical phenomena.

B.5.1 Reference-Specific Rubric Construction

Before evaluating candidates, the VLM analyzes each reference video to construct a case-specific rubric specifying: (1) the core physical event or target state; (2) necessary phases; (3) spatial, causal, and state-transition constraints; (4) appearance anchors; and (5) potential visual ambiguities. The rubric is then frozen and shared across all candidates for the same reference.

B.5.2 Phase-Level Candidate Evaluation

Each anonymized candidate is evaluated independently against the frozen rubric using phase-level rather than timestamp-level alignment. This allows valid variations in speed, trajectory, and magnitude while identifying missing phases, implausible interactions, and unexplained state resets. To ensure full-trajectory coverage, the evaluator inspects structured temporal windows and verifies state transitions across the sequence.

B.5.3 Evaluation Dimensions

We report four independent scores on a discrete 1–5 scale:

- **Appearance Consistency (C):** Measures consistency of the subject, background, materials, and colors, independently of physical correctness.
- **Event Completion (E):** Measures whether the core physical event and its required terminal state are realized.
- **Physical Plausibility (P):** Evaluates spatial, causal, and state-transition consistency without requiring exact geometric replication of the reference.
- **Temporal Evolution (T):** Measures the ordering, continuity, and persistence of the required physical phases, including failures such as unexplained reversals or resets.

B.5.4 Evidence-Grounded Scoring

Method identities, existing scores, and expected rankings are withheld from the evaluator. Each score must be supported by explicit visual evidence, including relevant frame intervals, state transitions, and reference–candidate phase correspondences. The four dimensions are scored independently, such that appearance consistency cannot compensate for incomplete or physically implausible dynamics.

The evaluator outputs structured visual evidence and integer scores, which are programmatically validated before aggregation.

B.5.5 Full Evaluation Prompt

Due to its length and the inclusion of programmatic formatting tokens, the complete zero-shot VLM evaluation prompt is provided as a separate text file (`eval_prompt.txt`) in our supplementary material.

C Additional Details of PhysDMD

C.1 Additional Details of Direct Dynamics Alignment

This section provides the formal formulation of the direct dynamics alignment objective introduced in Sec. 3.2. Let $e_{\text{gen}} = E_{\text{wm}}(\mathbf{v}^{\text{gen}})$ denote the latent world-model prediction energy of the generated transition, and $e_{\text{gt}} = \text{sg}[E_{\text{wm}}(\mathbf{v}^{\text{gt}})]$ represent the fixed reference energy from the paired ground-truth trajectory.

One-sided hinge objective. We initially consider a one-sided formulation that penalizes the generated transition strictly when its prediction energy exceeds the ground-truth reference:

$$\mathcal{L}_{\text{WM}}^{\text{hinge}} = \mathbb{E}_{(\mathbf{h}_{\text{gt}},\mathbf{y}),\,\mathbf{z}}\left[\max(0,\, e_{\text{gen}} - e_{\text{gt}} - m)\right], \tag{5}$$

where m is a non-negative margin (set to $m = 0$ in our experiments). The objective penalizes energy above the reference plus the margin, but yields zero loss when $e_{\text{gen}} \leq e_{\text{gt}} + m$. It therefore does not discourage anomalously low prediction energies, which can arise from attenuated or frozen motion.

Symmetric relative-Huber objective. To avoid favoring transitions that are either less predictable or unrealistically more predictable than their paired real trajectories, we measure their deviation using a ground-truth-normalized relative energy gap:

$$d_{\text{rel}} = \frac{e_{\text{gen}} - e_{\text{gt}}}{\max(e_{\text{gt}}, \epsilon)}, \tag{6}$$

where $\epsilon > 0$ ensures numerical stability. We then impose a symmetric Huber penalty on this normalized gap:

$$\mathcal{L}_{\text{WM}} = \mathbb{E}_{(\mathbf{h}_{\text{gt}},\mathbf{y}),\,\mathbf{z}}\left[\rho_\delta(d_{\text{rel}})\right], \tag{7}$$

where $\mathbf{z}$ denotes the generative noise for the target chunk, and ρ_δ is defined as:

$$\rho_\delta(d) = \begin{cases} \frac{1}{2}d^2, & |d| \leq \delta, \\ \delta\left(|d| - \frac{1}{2}\delta\right), & |d| > \delta. \end{cases} \tag{8}$$

Unlike the one-sided hinge, the relative-Huber objective penalizes both positive and negative deviations from the reference energy. Furthermore, the ground-truth normalization calibrates the penalty against the natural predictability of diverse physical events, while the linear tails of the Huber loss gracefully mitigate the impact of severe relative-energy outliers.

D Additional Experiments

D.1 Additional Details of Training-Data Composition

This section details the training-data configurations evaluated in Table 2. To rigorously isolate the impact of data composition, all evaluated variants share identical model initializations and optimization hyperparameters.

Motion-Driven Candidate Pool Construction. We derive the candidate pool from the OpenVidHD dataset (Nan et al., 2025). To ensure kinematic reliability, we filter out sequences with visual or temporal artifacts (e.g., severe blur, exposure anomalies, short duration), retaining only clips classified as `optimal_motion` or `medium_quality`. From this refined subset, we restrict the pool to samples possessing the requisite GT history latents and text annotations for subsequent selection.

Difficulty and Motion Characterization. To profile sample difficulty, we extract dense trajectories via AllTracker (Harley et al., 2025) to derive a latent spatiotemporal motion mask. We then formulate two complementary difficulty metrics: (1) a motion-weighted data distribution distillation discrepancy measuring correction energy within dynamic regions, and (2) the relative-Huber latent world-model discrepancy (Sec. C.1) capturing the energy deviation between predicted and ground-truth transitions. These metrics serve strictly as heuristic proxies for data curation, rather than absolute indicators of physical correctness.

Reference-Calibrated Selection. To calibrate these metrics, we establish a reference distribution from previously optimized clips. We define high-discrepancy thresholds via the reference median and truncate extreme outliers at an upper-percentile bound. Discarding candidates that score low on both metrics yields three distinct strata: DMD-high, WM-high, and dual-high. To preserve kinematic diversity, we stratify these candidates across motion-magnitude bins and sample proportionally. The resulting novel clips are subsequently merged with the reference set to create the final training corpus.

Training-Data Configurations. We evaluate three data compositions (Table 2): *Original* inherits the uncurated Vanilla-DMD distribution; *Mixed* blends motion-filtered data with extreme-difficulty samples (1:1 ratio); and *Filtered* (adopted by PhysDMD) utilizes the motion-refined training corpus. The mixed configuration yields a lower Overall score (27.19) than the original (28.02) and filtered (28.10) configurations. These results favor the latter two selection recipes in this setting, with only a small difference between them.

D.2 Ablation on History Conditioning Strategies

Synthetic Corruption vs. Self-Rollout. To robustify the autoregressive student against imperfect histories, we evaluate *synthetic history corruption*, which introduces heuristic perturbations (e.g., stochastic noise and saturation jitter) to the ground-truth prefixes, against *online self-rollout*, which dynamically synthesizes the complete conditioning history via the active student policy. As reported in Table 4, both recipes achieve nearly identical Overall Physics-IQ scores. We default to synthetic corruption, which achieves a similar Overall score without generating the complete history during training.

Table 4: **Ablation of history conditioning strategies on Physics-IQ-Verified.**

Variant	Self-Rollout	History Corruption	Saturation Jitter	NFE	ST-IoU↑	Sp-IoU↑	wSp-IoU↑	MSE-based↑	Overall↑
Synthetic history corruption	–	✓	✓	4	29.83 ±0.63	30.83 ±0.78	19.37 ±0.38	32.06 ±0.65	28.02 ±0.35
Full-history self-rollout	✓	–	✓	4	31.61 ±1.00	28.88 ±1.50	18.60 ±1.05	32.93 ±0.82	28.01 ±0.50

E Additional Qualitative Results

E.1 Additional Qualitative Results on Physics-IQ-Verified

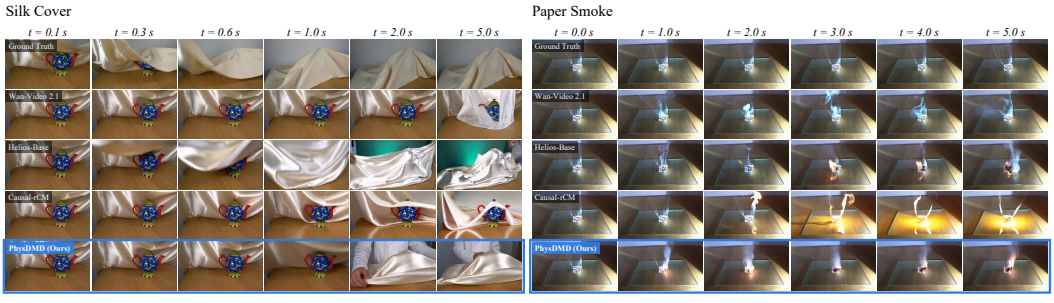

Figure 9: **Additional qualitative comparisons on Physics-IQ**.

E.2 Additional Qualitative Results on PLV-Bench

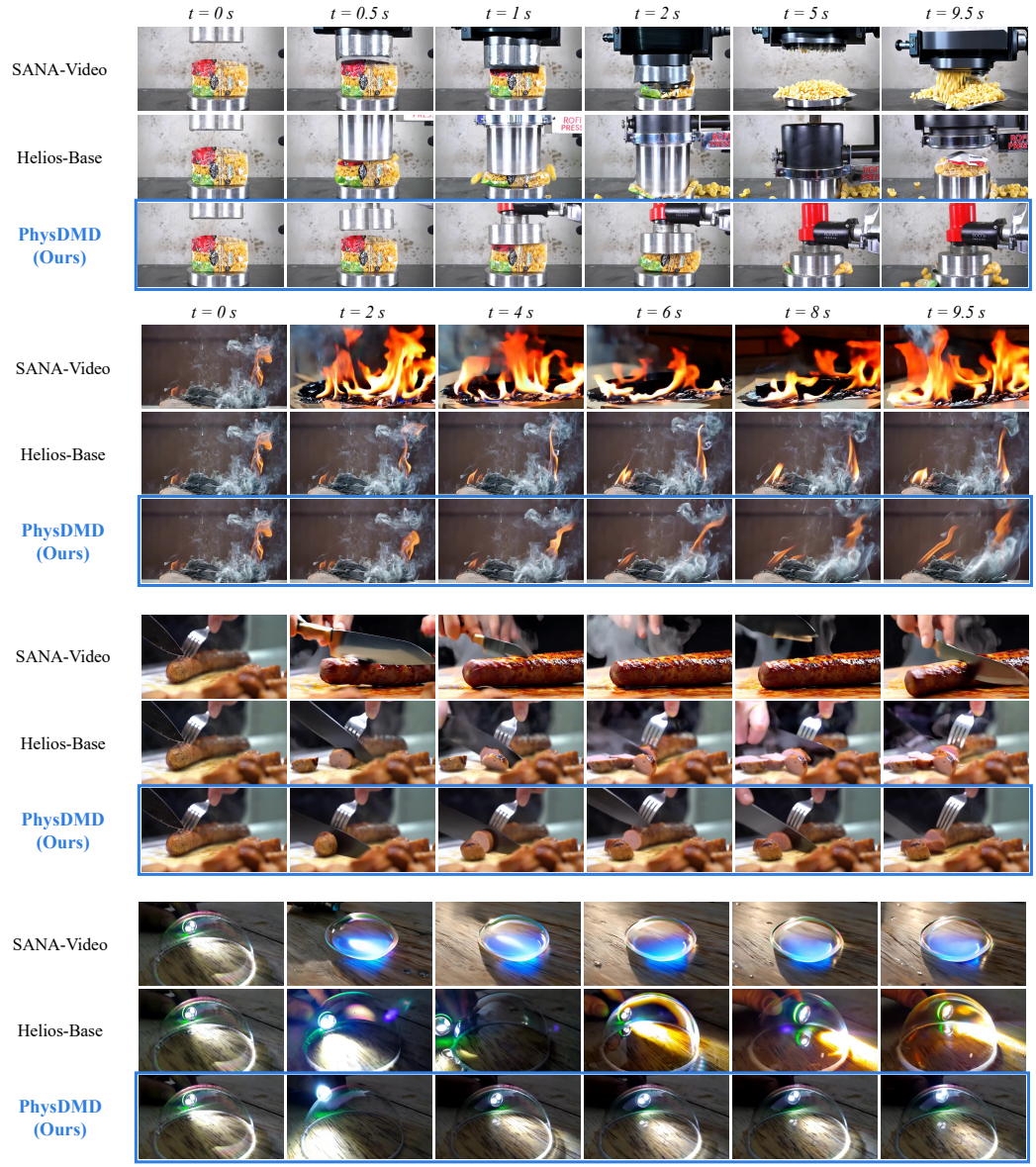

Figure 10: **Additional qualitative comparisons on PLV-Bench** against standard AR models.

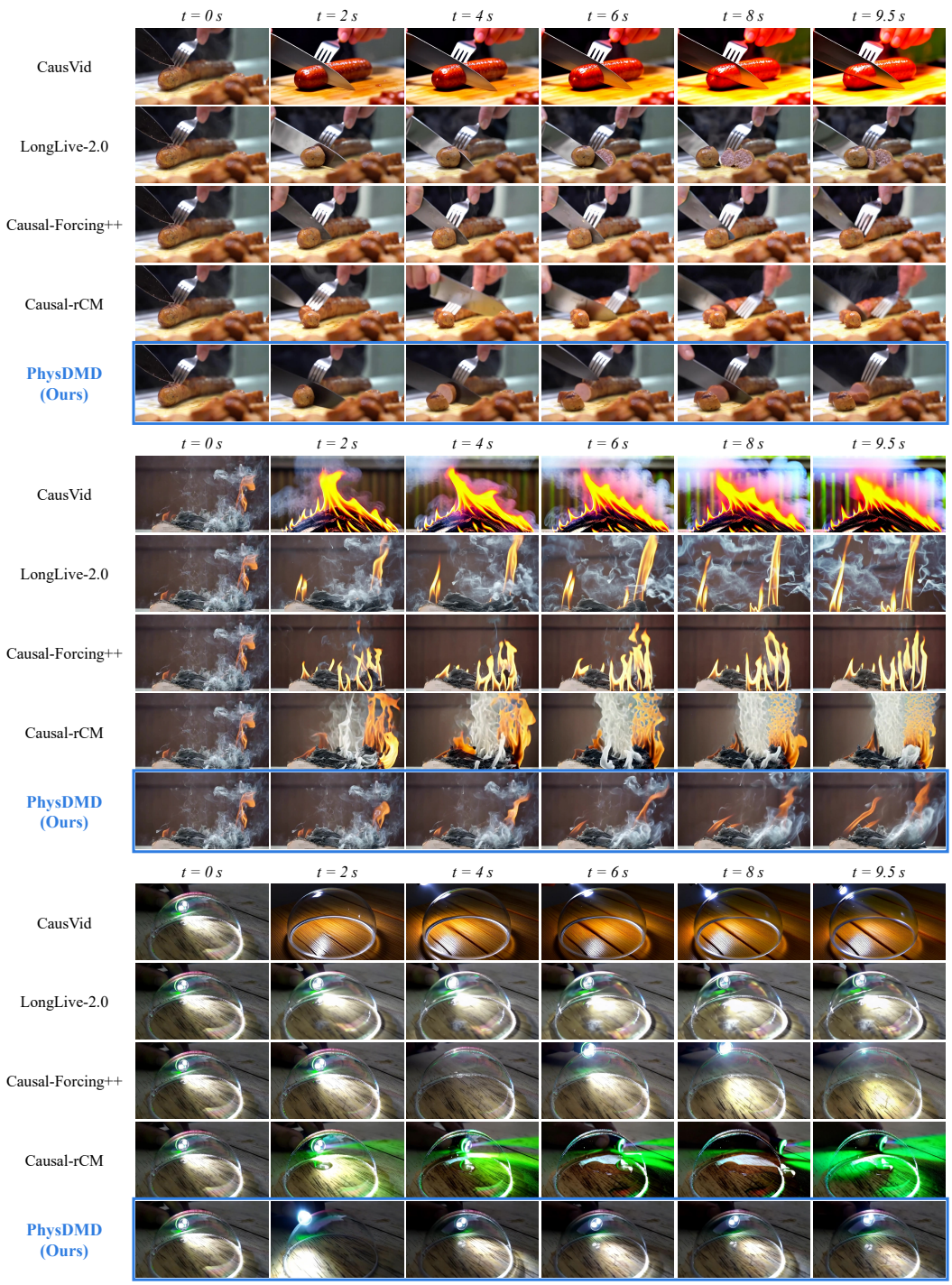

Figure 11: **Additional qualitative comparisons on PLV-Bench** against distilled AR models.

E.3 Qualitative Ablation Results

Qualitative Ablation. Figure 12 provides a qualitative comparison of the ablated variants on Physics-IQ-Verified.



Figure 12: **Qualitative ablation on Physics-IQ-Verified.**