NOT ALL PRIVILEGE IS DISTILLABLE: CALIBRATING TEACHER CORRECTIONS FOR FEW-STEP VIDEO GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Privileged distillation uses clean video continuations to supervise a few-step generator on student-visited denoising states. We find that the largest evaluated evidence budget yields lower mean generation scores and motion activity than an intermediate budget, even when teacher and student share the first frame. To examine reference dependence, we introduce Privilege-Induced Teacher Variance (PITV), an offline measure of normalized variation in teacher corrections across screened references at fixed teacher weights and student states. Higher PITV is associated with poorer full-privilege transfer. For training, Mode-Calibrated Privileged Distillation (MCPD) tests support within a single continuation. It compares corrections from a progressively enlarged evidence subset and the full selected budget, using directional agreement and normalized displacement to weight the full correction at each token. Poorly supported targets move toward the teacher prediction conditioned only on the prompt and first frame. On 4-step Wan2.2 and 8-step LongCat-Video, MCPD improves VBench-I2V Total Score over timestep-dependent scalar shrinkage by 0.6–0.9 points and also outperforms adapted DFD. Controls on Wan2.2 show that matching residual energy or mixing the same teacher queries with fixed weights does not recover MCPD's performance.

1 INTRODUCTION

Distillation has reduced high-quality video generation to a few denoising steps Ding et al. (2025); Nie et al. (2026); Ge et al. (2026); Shaul et al. (2026). Subsequent adaptation must improve the model without disrupting the sampling dynamics learned during distillation Jiang et al. (2026b); Liu et al. (2026). On-policy distillation addresses one source of mismatch by constructing supervision on student-visited states Zhao et al. (2026a); Li et al. (2026b). Privileged variants use paired images or real-video histories as additional teacher context Jiang et al. (2026b); Liu et al. (2026), while hybrid-policy anchoring addresses incompatibilities between teacher conditions and student trajectories Bu et al. (2026). We study how much of the resulting correction should be transferred when the student must generate without that evidence.

We find that additional continuation evidence produces non-monotonic student gains in shared-anchor image-to-video (I2V) post-training. The largest evaluated fixed budget gives lower mean I2V scores and motion activity than a validation-selected intermediate budget. Separately, a frozen snapshot of the training teacher achieves higher reference fidelity when queried with more evidence at the same student states. The distinction illustrated in Figure 1 is between adherence to the paired continuation and improvement of the deployed student. Reference fidelity alone is therefore an insufficient basis for choosing the strength of privileged supervision.

A prompt and first frame need not determine a unique future motion, interaction, or camera path. The paired continuation resolves some of these details for the teacher, while the student generates from its own evolving state without observing that continuation. A teacher can consequently issue reference-specific corrections even when queried at the same noisy latent. Such dependence does not make every correction harmful, but it leaves an information asymmetry that state matching alone does not remove. Related work in reasoning and multimodal distillation also distinguishes learnable

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.

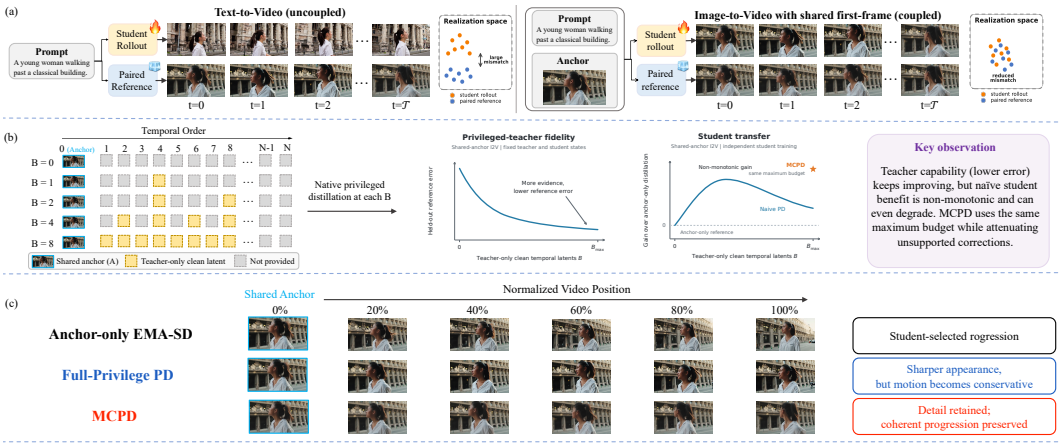

Figure 1: **More privileged evidence need not improve student transfer. (a)** A shared first frame constrains initial content but leaves future continuations ambiguous. **(b)** Teacher reference fidelity and student transfer can diverge as privileged evidence increases. **(c)** Native I2V examples with matched prompts and anchors show differences in appearance and motion progression.

information from view-specific or privilege-dependent supervision Wang et al. (2026); Nguyen et al. (2026); Shen et al. (2026); Jiang et al. (2026a); Wu et al. (2026).

We introduce Privilege-Induced Teacher Variance (PITV) to measure this reference dependence. With teacher weights and student-visible inputs fixed, PITV measures the normalized variation of continuation-induced corrections across screened reference videos under the same condition. Higher PITV is associated with lower condition-level gains from full-privilege distillation, including in a subset whose references are unanimously judged action-compliant. PITV is an offline diagnostic of reference dependence, not a label for harmful supervision. Our training rule instead tests support within a single paired continuation.

We propose **Mode-Calibrated Privileged Distillation (MCPD)** to regulate continuation-induced velocity corrections on the student's native few-step rollout. At each supervised state, an exponential-moving-average (EMA) teacher is queried with the deployment condition alone, a progressive subset of clean continuation latents, and the full selected budget. Each privileged correction is measured relative to the anchor-only prediction. Directional agreement and normalized vector displacement between the two corrections determine a detached token-wise weight on the full residual. Weak support moves the target toward the anchor-only prediction; the progressive query supplies only the support test. Clean evidence never replaces student noisy tokens. Teacher-only inputs and calibration computations are removed after training, retaining the native conditioning interface, sampler, and number of model evaluations.

On 4-step Wan2.2 and 8-step LongCat-Video generators Wan et al. (2025); Team et al. (2025), MCPD improves mean I2V Total Score over timestep-dependent scalar shrinkage by 0.6 and 0.9 points, and over adapted DFD Chen et al. (2026) by 0.7 and 0.8 points, across five post-training runs from one shared initialization per backbone. Energy-matched and fixed-mixture controls on Wan2.2 do not recover MCPD's performance. A short-update probe yields higher held-out learning utility with MCPD than full-privilege targets. Text-only evaluation improves after I2V training, measuring cross-condition transfer into the existing interface. Motion activity increases relative to direct full-privilege transfer, while mean appearance-feature diversity remains close to the initial models.

2 Related Work

Few-step video generation and post-training. DMD and DMD2 established distribution matching for one- and few-step synthesis Yin et al. (2024b;a). Video methods also learn conditional transitions or parallelize denoising predictions Nie et al. (2026); Shaul et al. (2026). Reward feedback provides alternative supervision for distilled outputs and trajectory targets Ding et al. (2025); Li et al. (2026c). Maintaining temporal behavior requires attention to training histories, motivating self-generated rollouts and cache-aware or on-policy objectives Huang et al. (2026); Ge et al. (2026);

Luo et al. (2026). DFD directly addresses post-training after DMD2 initialization Chen et al. (2026). Its real-data teacher-score discrepancy modifies the distribution-matching update. The objective is to recover fidelity and diversity after step distillation, with the sampling budget already fixed.

Privileged on-policy distillation. Privileged learning uses information withheld at deployment Vapnik & Vashist (2009); Lopez-Paz et al. (2016). OPSD combines this asymmetry with student-generated trajectories Zhao et al. (2026a); Penaloza et al. (2026). For step-distilled image models, D-OPSD introduces the paired target through teacher context without replacing the student's noisy state Jiang et al. (2026b). Its teacher and student predict velocities on the same sampled rollout, so paired data enriches the target without providing the supervised denoising state. Extensions accommodate multiple teachers, heterogeneous interfaces, and alternative output or representation targets Li et al. (2026b); Fu et al. (2026); Wei et al. (2026), with CFG branch asymmetry posing a separate matching issue Li et al. (2026a). In video, visual conditions and generated content can disagree. OPSD-V replaces older teacher-cache history with real-video context while retaining the latest student-generated chunk Liu et al. (2026). HPSD instead initializes student sub-trajectories from privileged teacher states Bu et al. (2026). Its trajectory anchoring changes the starting states for student rollout and is not reproduced by sharing a first frame alone. MCPD retains the native shared-anchor rollout and calibrates continuation-induced supervision on its visited states.

Transferability of privileged supervision. Teachability and recoverability analyses assess teacher disagreement through local compatibility and downstream correction Wang et al. (2026); Jiang et al. (2026a). Purification, structured privilege, and matched-information anchoring revise supervision Shen et al. (2026); Zhao et al. (2026b); Wu et al. (2026). In multimodal learning, inference-accessible cues and visual attribution inform target construction Tian et al. (2026); Zhang et al. (2026). More context can also degrade distillation, and apparent gains may reflect context-induced behavior rather than target-specific information Kaur et al. (2026); Ichihara et al. (2026). Clean autoregressive histories Yang et al. (2026) and fixed-budget multi-shot routing Guo et al. (2026) illustrate how video context shapes generation. AVSD provides a close precedent for residual calibration: it retains cross-view consensus and gates additional view-specific support Nguyen et al. (2026). Its geometric and arithmetic teacher aggregates distinguish shared support from information contributed by only a subset of views, with consensus retained in the reconstructed target. MCPD instead takes the deployment-conditioned teacher as its baseline and compares evidence budgets from one continuation, without retaining a privileged consensus floor. PITV separately measures cross-reference variation, holding teacher weights and student-visible denoising inputs fixed.

3 Mode-Calibrated Privileged Distillation

We post-train a step-distilled video generator on pairs (c, x^{gt}). The deployment condition is $C = (c, A)$, where $A = x_1^{\mathrm{gt}}$ in primary I2V and $r = x_{2:T}^{\mathrm{gt}}$ provides teacher-only evidence. An uncoupled T2V diagnostic sets $A = \varnothing$. Sharing the first frame constrains initial appearance but not future dynamics; mode refers to a video realization without explicit assignment.

3.1 Same-State Privileged Post-Training

Let $\sigma_K > \cdots > \sigma_1 > \sigma_0 = 0$ be the native noise schedule. The positive predictor input is $X_k^+ = (z_k^s, \sigma_k, C, \eta_k)$, where η_k contains native coordinates, masks, and any predictor history; X_k^- follows the native negative-conditioning rule. The complete sampler state Σ_k additionally contains solver history and sampling randomness. Starting from $\Sigma_K = \mathrm{Init}_{\mathrm{nat}}(\xi, C)$, with rollout randomness ξ independent of r conditional on C, the student follows

$$\begin{aligned} v_k^{s,\pm} &= F_\theta(X_k^\pm), \\ \Sigma_{k-1} &= \mathrm{sg}\big[\Phi_k\big(\Sigma_k, \mathcal{G}_{\mathrm{nat}}(v_k^{s,+}, v_k^{s,-})\big)\big], \qquad k = K, \dots, 1. \end{aligned} \tag{1}$$

For native CFG, $\mathcal{G}_{\mathrm{nat}}(v^+, v^-) = v^- + w_{\mathrm{cfg}}(v^+ - v^-)$; otherwise, only v^+ is evaluated. Native timestep conversion and anchor handling are retained. Parameters remain fixed within each rollout, and sg stops gradients through the sampled states.

The teacher is an EMA copy initialized from the same effective few-step weights. Let $\mathcal{R}(A, r)$ contain eligible continuation slices from the frozen VAE representation. The operator $\mathcal{P}_B^\rho$ exposes

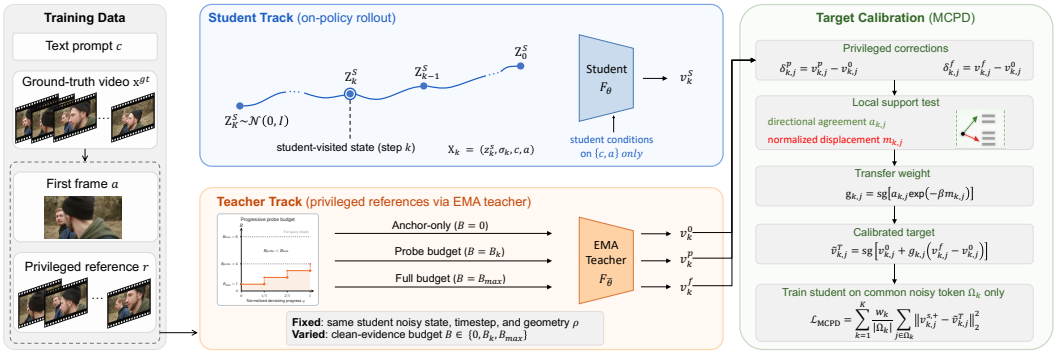

Figure 2: **Mode-Calibrated Privileged Distillation.** At a fixed student state, the EMA teacher provides anchor-only, progressive-budget, and full-budget predictions. The progressive probe gates the full correction with anchor-only fallback. Only common noisy outputs are supervised; teacher-only inputs and computation are removed at deployment.

B clean temporal slices under layout ρ; B excludes the shared anchor and counts neither raw frames nor transformer tokens. The teacher prediction is

$$V_k(B, \rho; r) = F_{\bar{\theta}}\big(\mathrm{sg}[X_k^+]; \mathcal{P}_B^\rho(\mathcal{R}(A, r))\big)\,. \tag{2}$$

As in diffusion and video on-policy distillation Jiang et al. (2026b); Liu et al. (2026), supervision is evaluated at student-visited noisy latents. Teacher-only evidence enters through a clean-conditioning extension with local diffusion level zero. It never overwrites generated slots, drives the student solver, or supplies a ground-truth reconstruction target.

At fixed B, auxiliary-context and condition-aligned layouts differ only in teacher-token coordinates, using a separate temporal range and corresponding video coordinates, respectively. Clean values, modulation, packing, and allowed attention connections remain fixed. Across budgets, student coordinates and the original-token attention-mask submatrix remain unchanged; numerical attention weights may change. Let Ω_k contain valid generated output tokens, excluding clean conditions, privileged slots, and padding. This structural mask is independent of r, and $\|u\|_{\Omega_k}^2 = \sum_{j \in \Omega_k} \|u_j\|_2^2$ sums prediction channels within each token. Appendix A specifies encoding and state constraints.

3.2 Realization Dependence and PITV

Fix the student and teacher snapshots, budget, and layout, and abbreviate $X = X_k^+$. Independent rollout and reference sampling under C need not produce compatible continuations. Define the privileged correction and its conditional mean under a joint law $\pi(X, r)$ as

$$\Delta_{B,\rho}(X, r) = V(B, \rho; X, r) - V(0; X), \qquad \mu_\pi(X) = \mathbb{E}_\pi[\Delta_{B,\rho}(X, r) \mid X]. \tag{3}$$

For a reference-blind predictor h, let $\mathcal{L}_\pi(h) = \mathbb{E}_\pi \|h(X) - \Delta_{B,\rho}(X, r)\|_\Omega^2$.

Proposition 1 (reference-law shift). With finite second moments and an X-measurable mask, μ_π minimizes $\mathcal{L}_\pi$ over unrestricted predictors. If π_{pair} and π_{cmp} have the same X marginal, then

$$\mathcal{L}_{\mathrm{cmp}}(\mu_{\mathrm{pair}}) - \mathcal{L}_{\mathrm{cmp}}(\mu_{\mathrm{cmp}}) = \mathbb{E}_X \|\mu_{\mathrm{pair}}(X) - \mu_{\mathrm{cmp}}(X)\|_\Omega^2. \tag{4}$$

A realization-compatible π_{cmp} is a conceptual comparator, not an observed oracle. Equation equation 4 compares regression targets; it neither identifies that coupling nor guarantees perceptual improvement. The proof and its relation to conditional flow matching are given in Appendix B. This information-set distinction is related to matched-information anchoring and evidence attribution in privileged distillation Wu et al. (2026); Zhang et al. (2026).

We characterize reference dependence with Privilege-Induced Teacher Variance (PITV). Let $\mathcal{Q}(r \mid C)$ be the distribution of accepted sampled continuations under the same condition, and let $\bar{\Delta}_\mathcal{Q}(X) = \mathbb{E}_\mathcal{Q}[\Delta_{B,\rho}(X, r)]$. Holding X fixed, define

$$\mathrm{PITV}_{B,\rho}(X) = \frac{\mathbb{E}_\mathcal{Q} \|\Delta_{B,\rho}(X, r) - \bar{\Delta}_\mathcal{Q}(X)\|_\Omega^2}{\mathbb{E}_\mathcal{Q} \|\Delta_{B,\rho}(X, r)\|_\Omega^2 + \epsilon_P}. \tag{5}$$

PITV measures cross-reference variation relative to correction energy, not recoverability or harmfulness. The unnormalized variance and energy are retained alongside the ratio. Since ordinary paired training supplies one continuation per example, PITV is evaluated offline; its estimator and low-energy conventions are given in Appendix C.

3.3 Calibrated Privilege Transfer

MCPD uses a single-reference support test rather than estimating PITV online. Fix an ordered candidate list $\iota_1, \ldots, \iota_{B_{\max}}$ and nested sets $\mathcal{I}_B = \{\iota_1, \ldots, \iota_B\}$, with $\mathcal{I}_0 = \varnothing$. Increasing B reveals additional clean slices without relocating existing evidence. For $K \geq 2$, the progressive probe uses

$$q_k = \frac{\sigma_K - \sigma_k}{\sigma_K - \sigma_1}, \qquad B_k = B_{\min} + \lfloor (B_{\text{probe}} - B_{\min}) q_k^{\gamma} \rfloor, \tag{6}$$

where $\gamma > 0$ and $1 \leq B_{\min} \leq B_{\text{probe}} < B_{\max}$. This allocation uses less probe evidence at higher noise, without assuming that later student states agree more closely with the paired continuation.

At each supervised state, three detached teacher queries give

$$\begin{aligned} v_k^0 &= V_k(0), & v_k^p &= V_k(B_k, \rho; r), & v_k^f &= V_k(B_{\max}, \rho; r), \\ \delta_{k,j}^p &= v_{k,j}^p - v_{k,j}^0, & \delta_{k,j}^f &= v_{k,j}^f - v_{k,j}^0, & & j \in \Omega_k. \end{aligned} \tag{7}$$

The anchor-only prediction v_k^0 retains all deployment conditions. For each output-token vector, we compute directional agreement and normalized vector displacement:

$$\begin{aligned} a_{k,j} &= \left[\frac{\langle \delta_{k,j}^p, \delta_{k,j}^f \rangle}{\|\delta_{k,j}^p\|_2 \|\delta_{k,j}^f\|_2 + \epsilon_a} \right]_0^1, \\ m_{k,j} &= \frac{2\|\delta_{k,j}^f - \delta_{k,j}^p\|_2}{\|\delta_{k,j}^f\|_2 + \|\delta_{k,j}^p\|_2 + \epsilon_m}, \end{aligned} \tag{8}$$

where $[x]_0^1 = \min(1, \max(0, x))$ and $\epsilon_a, \epsilon_m > 0$. The displacement depends on both length and angle. With $\beta > 0$, the weight and target are

$$g_{k,j} = \text{sg}[a_{k,j} \exp(-\beta m_{k,j})], \tag{9}$$

$$\widetilde{v}_{k,j}^T = \text{sg}\left[v_{k,j}^0 + g_{k,j} \delta_{k,j}^f \right]. \tag{10}$$

The progressive prediction supplies only the support test. A zero gate rejects the entire continuation-induced correction, whereas a unit gate recovers the full teacher. Intermediate gates shrink the full residual locally rather than separating its channel-wise components. Appendix B.2 relates this construction to scalar mixing and states the limits of budget agreement.

The student minimizes

$$\mathcal{L}_{\text{MCPD}} = \mathbb{E}\left[\sum_{k=1}^{K} \frac{w_k}{|\Omega_k|} \sum_{j \in \Omega_k} \|v_{k,j}^{s,+} - \widetilde{v}_{k,j}^T\|_2^2 \right], \qquad w_k \geq 0, \quad \sum_k w_k = 1, \tag{11}$$

where the expectation is over training pairs and native rollout randomness. The uniform all-state formulation uses $w_k = 1/K$. Saved states and targets are treated as constants: gradients are accumulated over the selected states and batch before one student update, followed by $\bar{\theta} \leftarrow \mu\bar{\theta} + (1-\mu)\theta$. Appendix A gives the update procedure and sampled-state estimator.

The loss matches the positive prediction rather than a CFG-composed teacher target, whose branch errors can compensate Li et al. (2026a). Shared parameters can change the negative branch. At deployment, teacher-only evidence and computation are discarded; the native architecture, conditioning paths, solver, guidance rule, and number of function evaluations are retained. Text-only evaluation after I2V post-training is cross-condition transfer, not T2V on-policy optimization.

4 Experiments

We study the effect of continuation evidence on post-training, compare MCPD with fixed-budget and scalar-transfer baselines, and examine the contribution of local target calibration.

4.1 EXPERIMENTAL SETUP

Initialization and optimization. We apply DMD2 Yin et al. (2024a) to the official Wan2.2-TI2V-5B Wan et al. (2025) and LongCat-Video-13.6B Team et al. (2025) models to construct 4-step and 8-step initializations. All methods within a family share the same starting weights and deployment configuration: flow-matching Euler, schedule shift 3.0, and CFG scale 5.0. Shared-anchor I2V is the training task; both branches receive the first frame, while continuation evidence is teacher-only. Wan2.2 updates attention-projection LoRA parameters with rank 128 and $\alpha = 128$; LongCat updates the full denoiser. MCPD supervises all denoising states with uniform weights, accumulates gradients before each optimizer update, and maintains an effective-weight EMA with decay 0.999. Appendix D gives the remaining training details.

Data and evidence. We use 30K clips from WebVid-10M and Panda-70M, removing identifiable source-video overlaps with validation and evaluation data. Wan2.2 post-training uses 256×416 resolution with 53 frames, while LongCat uses 256×256 resolution with 33 frames, preserving their respective latent temporal structures. Evaluation is performed at 480×832/81 frames and 512×512/65 frames, respectively. The default budgets are $(B_{\min}, B_{\text{probe}}, B_{\max}) = (1, 4, 8)$, with $\gamma = 1$ and $\beta = 2$. Budgets count clean temporal latent slices beyond the anchor. At training size, $B_{\max} = 8$ exposes 8 of 13 eligible continuation slices on Wan2.2 and all 8 on LongCat. Full privilege therefore denotes the maximum selected budget, not a common evidence fraction. Candidate selection follows Appendix A.2.

Selection and evaluation. We train for up to 10,000 generator updates and select hyperparameters and checkpoints on 200 validation conditions. Main I2V and T2V results use the complete official VBench++ and VBench suites Huang et al. (2025; 2024). Separate sets support PITV analysis, appearance-diversity measurement, and a 256-condition learning probe split into 128 update and 128 evaluation conditions. I2V/T2V denote official Total Scores. DD measures motion activity; A-Div. measures pairwise cosine distance between temporally averaged DINOv2 features across 8 generations per condition Oquab et al. (2023). We report five post-training runs per method and backbone. Run-level SDs describe individual method scores, whereas paired SDs describe between-method differences. Training-run intervals condition on the shared initialization and evaluation design; Appendix F.2 separates them from condition-bootstrap intervals.

Comparators. BASE is the unmodified initialization. SFT uses data-state flow-matching supervision, and EMA-SD supplies no teacher-only continuation. The deterministic null construction is distinguished from these executed non-privileged runs in Appendix D.1. FULL-PRIV PD directly matches the maximum-budget EMA prediction. BEST-FIXED PD selects $B = 4$ from $\{1, 2, 4, 8\}$ on validation data. BEST-SCALAR and STEP-SCALAR use constant and noise-dependent residual weights. DMD2-CONTINUE and DFD use the shared-anchor distribution-matching adaptations in Appendix D.3 Yin et al. (2024a); Chen et al. (2026). All deployment evaluations omit teacher-only continuation evidence.

4.2 EVIDENCE BUDGET AND STUDENT TRANSFER

Teacher probes and student outcomes. Figure 3(a) evaluates an independent frozen Wan2.2 teacher with 50 Euler steps; its I2V score increases with evidence. Panel (b) evaluates students trained separately with their own moving few-step EMA teachers. Increasing the fixed budget from 4 to 8 reduces mean I2V from 85.9 to 85.6 and DD from 70.5 to 66.9. This decline is relative to the intermediate budget, not anchor-only training. The panels describe different teacher and student processes; they do not link one teacher's generation quality causally to its transferability. MCPD is evaluated at the single $(1, 4, 8)$ configuration shown.

Training-teacher fidelity. Appendix E.1 freezes the actual FULL-PRIV PD EMA at update 5,000 and queries 256 saved Wan2.2 validation states. On a common region excluding the anchor and exposed-latent neighborhoods, reference error decreases by 14.2%, 27.5%, 43.1%, and 61.8% relative to $B = 0$ at budgets $1, 2, 4, 8$. This is increased reference adherence at fixed weights and states, rather than a measurement of student learning benefit.

Reference dependence. Figure 3(c) relates PITV to condition-level score changes from EMA-SD to FULL-PRIV PD. Eight references per condition pass ViCLIP text–video similarity above 0.25 and intra-video subject consistency above 0.70. Excluding 16 low-energy conditions leaves 184, with Spearman $\rho = -0.83$ (95% CI $[-0.88, -0.76]$). Scores average per-video I2V composites over five training runs and 4 generation seeds (Appendix E.2); they differ from the complete-suite aggregate. PITV measures cross-reference dependence, not MCPD's within-reference support signal.

Local target use. Figure 3(d) replaces the positive prediction in one block of a $2 \times 2 \times 2$ spatiotemporal partition and resumes sampling from the saved state. Utility is the change in a six-dimension per-video composite relative to the corresponding anchor-only intervention. Validation cut points 0.35 and 0.72 define support strata. Full privilege has negative mean utility in the low-support group; MCPD moves this mean toward zero while retaining positive utility in the high-support group. These direct-use effects are consistent with conservative fallback. The learning probe in Section 4.4 separately evaluates parameter updates.

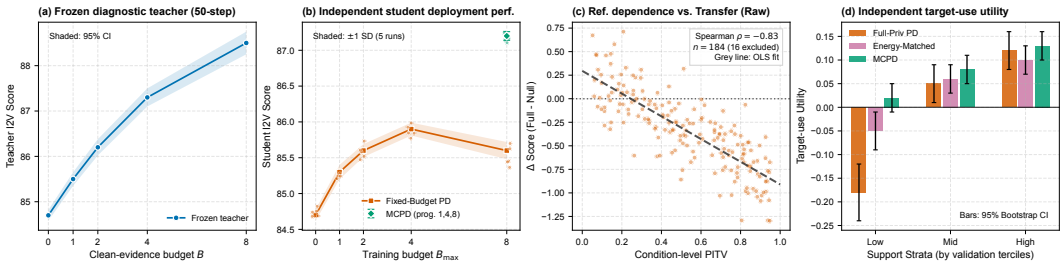

Figure 3: **Evidence and target utility on Wan2.2.** (a) Independent frozen 50-step teacher; shading: 95% condition-level CIs. (b) Fixed-budget students: five-run mean ± SD; MCPD uses $(1, 4, 8)$. (c) PITV versus FULL-PRIV PD−EMA-SD diagnostic score change across 184 retained conditions; grey: OLS fit. (d) Frozen-model block-intervention utility relative to anchor-only, grouped by validation-defined support; 95% paired condition-bootstrap CIs.

4.3 MAIN POST-TRAINING RESULTS

Table 1 reports validation-selected checkpoints. MCPD achieves I2V scores of 87.2 on Wan2.2 and 88.1 on LongCat, improving over BASE by 2.5/2.8 points and FULL-PRIV PD by 1.6/2.0. Against STEP-SCALAR, the gains are 0.6/0.9 points, with paired SDs 0.18/0.21 and 95% CIs $[0.38, 0.82]$/$[0.64, 1.16]$. Gains over adapted DFD are 0.7/0.8 points, with intervals $[0.46, 0.94]$/$[0.53, 1.07]$. The gains thus extend beyond the initialization and uncalibrated target to the evaluated scalar-transfer and distribution-matching alternatives. Appendix F.2 provides the paired statistics and test assumptions.

MCPD also has the highest mean DD, but the evidence varies by comparator. On Wan2.2, its DD advantage is 5.9 points over FULL-PRIV PD and 0.9 over STEP-SCALAR; the 0.4-point difference from DFD has a 95% CI spanning zero. A-Div. is 0.165/0.172 for MCPD, compared with 0.163/0.170 for BASE and 0.168/0.174 for DFD. These appearance-feature means are close to the initialization, with DFD retaining greater diversity. Dimension-level results appear in Appendix F.1.

Native text-only evaluation after I2V post-training improves T2V over BASE by 1.4 points on Wan2.2 (95% CI $[1.05, 1.75]$) and 1.6 on LongCat ($[1.21, 1.99]$). This is cross-condition transfer into the existing T2V interface, rather than direct T2V on-policy training.

Qualitative comparison. Figure 4 shows object-transfer examples under matched native I2V settings. In the last displayed frames, FULL-PRIV PD and STEP-SCALAR leave the block outside the tray. DFD† shows a block inside the tray and another on the table. MCPD shows one block in the tray with the hand withdrawn for both backbones. This highlights task completion and visible object count, which motion activity and time-averaged appearance diversity alone cannot establish.

Human preference. Ten raters assess 25 pairs per comparison. Non-tie overall preference for MCPD is 78.6% against BASE (95% pair-clustered CI $[72.1, 84.2]\%$) and 68.8% against FULL-PRIV PD ($[61.5, 75.8]\%$). Subject-consistency preferences are smaller, at 57.9% and 57.5%, with

Table 1: **Few-step video post-training.** Trained methods report five-run means from validation-selected checkpoints, sharing one initialization per backbone; BASE is fixed. I2V includes run-level SD. I2V/T2V/DD use a 0–100 scale; A-Div. is cosine distance. Best I2V/T2V means are bold. See Table 7 for paired contrasts and Appendix D.3 for † adaptations.

	Wan2.2 (4-step, LoRA)				LongCat (8-step, full parameter)			
Method	I2V	DD	A-Div.	T2V	I2V	DD	A-Div.	T2V
BASE	84.7	71.6	0.163	83.4	85.3	73.4	0.170	84.2
SFT	83.9 ±0.24	67.8	0.151	82.6	84.6 ±0.31	69.2	0.157	83.5
EMA-SD	84.7 ±0.03	71.6	0.163	83.4	85.1 ±0.12	73.0	0.168	84.1
FULL-PRIV PD	85.6 ±0.29	66.9	0.143	83.0	86.1 ±0.35	68.2	0.149	84.0
BEST-FIXED PD	85.9 ±0.27	70.5	0.156	83.8	86.5 ±0.33	72.0	0.162	84.7
BEST-SCALAR	86.2 ±0.23	71.2	0.159	84.0	86.7 ±0.28	73.0	0.166	84.9
STEP-SCALAR	86.6 ±0.21	71.9	0.162	84.4	87.2 ±0.27	73.8	0.169	85.2
DMD2-CONTINUE†	85.8 ±0.32	70.8	0.157	84.1	86.2 ±0.38	72.5	0.164	84.8
DFD†	86.5 ±0.26	72.4	0.168	84.5	87.3 ±0.30	74.3	0.174	85.3
MCPD	**87.2** ±0.24	72.8	0.165	**84.8**	**88.1** ±0.29	75.1	0.172	**85.8**

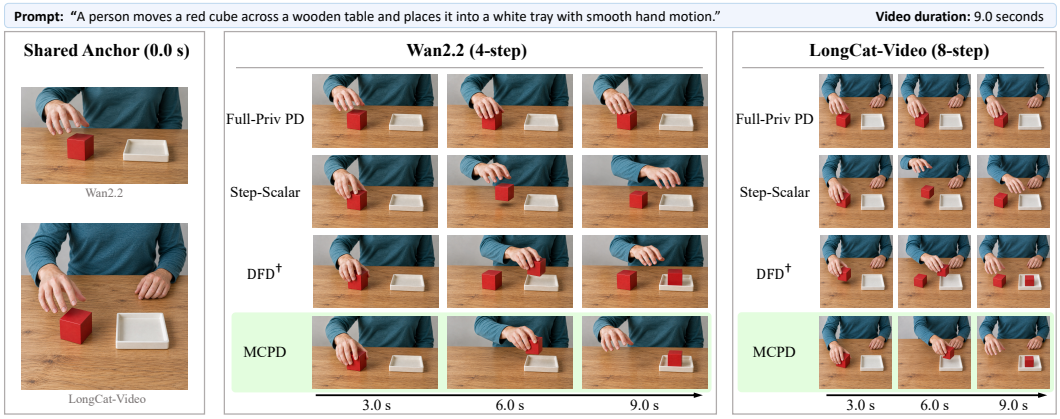

Figure 4: **Native I2V object-transfer comparison.** Within each backbone—Wan2.2 (4 steps) and LongCat-Video (8 steps)—methods share the prompt, input frame, sampling settings, and displayed frame indices, with paired random inputs. Each input frame appears once at left; aspect ratios are preserved. Inference uses no teacher-only evidence or calibration. † denotes the shared-anchor adaptation in Table 1.

interval lower limits close to 50%. Appendix F.3 gives all criterion-level counts and intervals. The study evaluates these two comparators, not DFD or STEP-SCALAR.

4.4 LOCAL CALIBRATION BEYOND GLOBAL SHRINKAGE

Table 2 compares target constructions on Wan2.2. ENERGY-MATCHED replaces local gates by a sample-and-state scalar with the same privileged residual energy, evaluated on its own current states. THREE-QUERY MIX uses the same queries with fixed mixing coefficients. MCPD improves their I2V scores by 1.1 and 0.9 points, with paired 95% CIs $[0.85, 1.35]$ and $[0.66, 1.14]$. Uniform residual-energy scaling and fixed mixing therefore do not recover the performance of the local rule in these runs. Energy matching does not equate complete gradients or optimizer updates.

A constant weak probe at $B = 4$ gives 86.8 I2V, 0.4 points below MCPD (95% CI $[0.19, 0.61]$). This comparison changes both evidence allocation across noise levels and total probe evidence. COSINE-ONLY removes the displacement penalty, changing support information and attenuation strength. SHIFTED-GATE changes spatial gate–residual correspondence and residual energy. CR-VEL evaluates a continuous consensus–residual target. MCPD exceeds their mean I2V scores by 1.2, 1.4, and 0.7 points, respectively. These comparisons support the combined target construction rather than separately identifying schedule order, displacement information, or fallback choice. Appendix D.4 defines the controls, and Appendix F.2 reports their paired statistics.

Table 2: **Target-construction comparisons on Wan2.2.** I2V entries are five-run means ± run-level SD; other metrics are means, on Table 1's scales. Paired-difference statistics appear in Appendix F.2. These controls can change more than one aspect of the target.

Target construction	I2V	DD	A-Div.	T2V
FULL-PRIV PD	85.6 ± 0.29	66.9	0.143	83.0
ENERGY-MATCHED	86.1 ± 0.27	71.1	0.159	84.1
FIXED-PROBE ($B = 4$)	86.8 ± 0.25	72.1	0.163	84.5
COSINE-ONLY	86.0 ± 0.28	70.8	0.156	83.8
THREE-QUERY MIX	86.3 ± 0.26	71.5	0.161	84.3
SHIFTED-GATE	85.8 ± 0.30	70.8	0.155	83.9
CR-VEL	86.5 ± 0.25	72.1	0.162	84.5
MCPD	$\mathbf{87.2} \pm 0.24$	72.8	0.165	**84.8**

Short-update learning utility. We apply five parameter updates on 128 conditions and evaluate 128 disjoint conditions against matched anchor-only updates. MCPD's mean diagnostic-score gain is $+0.15$ (95% CI $[0.08, 0.22]$), versus -0.08 ($[-0.15, -0.01]$) for FULL-PRIV PD. Their paired contrast is $+0.23$ (condition-bootstrap interval $[0.12, 0.34]$). This short-horizon probe measures held-out learning, separately from direct target use and full post-training (Appendix E.3).

4.5 LORA-RANK COMPARISONS AND TRAINING COST

MCPD improves over STEP-SCALAR at Wan2.2 LoRA ranks 64, 128, and 256 by 0.4, 0.6, and 0.7 I2V points; all three paired 95% intervals lie above zero (Appendix F.4). LongCat's full-parameter configuration gives the 0.9-point gain in Table 1. These results cover the tested Wan2.2 ranks and a second backbone, conditional on the shared initialization within each family.

Table 3: **Rank comparisons and post-training cost.** Score differences are MCPD minus STEP-SCALAR, averaged over five paired runs; parentheses give I2V paired SD. GPU-hours show single allocations on eight H800 GPUs, in MCPD / STEP-SCALAR order. Validation search and checkpoint serialization are excluded.

Backbone	Adaptation	ΔI2V (SD)	ΔDD	ΔT2V	GPU-hours
Wan2.2	LoRA 64	$+0.4$ (0.16)	$+0.7$	$+0.3$	172 / 150
Wan2.2	LoRA 128	$+0.6$ (0.18)	$+0.9$	$+0.4$	175 / 152
Wan2.2	LoRA 256	$+0.7$ (0.20)	$+0.9$	$+0.4$	179 / 155
LongCat	Full parameter	$+0.9$ (0.21)	$+1.3$	$+0.6$	263 / 223

At the main settings, measured GPU-hours are 175 versus 152 on Wan2.2 and 263 versus 223 on LongCat, approximately 15% and 18% above the two-query STEP-SCALAR. A repeat of the Wan2.2 rank-128 timing differed by 2.1%. Costs are reported separately from validation-selected checkpoint quality. Deployment removes the teacher and calibration paths; merging Wan2.2's LoRA adds no inference module or model evaluation.

5 CONCLUSION

We find that increasing teacher-only continuation evidence yields non-monotonic mean gains in shared-anchor I2V post-training of few-step video generators. PITV characterizes reference dependence at fixed teacher weights and student states, while MCPD uses a progressive evidence probe to calibrate full-privilege corrections relative to the anchor-only prediction. On Wan2.2 and LongCat-Video, MCPD improves I2V scores over direct privileged transfer, timestep-dependent scalar shrinkage, and adapted DFD. Energy-matched and fixed-mixture controls on Wan2.2 support the value of local calibration beyond uniform attenuation. The native inference interface and sampler remain unchanged. Budget agreement can still accept reference-specific corrections that remain stable as evidence grows. These findings motivate evaluating privileged supervision through student benefit rather than reference fidelity alone.

6 AI USE STATEMENT

We used Gemini (Google) only for grammar correction and language polishing during the writing of this paper. No AI tools were involved in the research ideation, methodology, experiments, or analysis. The authors take full responsibility for all content in this manuscript.

REFERENCES

Jiazi Bu, Pengyang Ling, Yujie Zhou, Yibin Wang, Yuhang Zang, Xuanlang Dai, Shengyuan Ding, Tianyi Wei, Xiaohang Zhan, Jiaqi Wang, et al. Hpsd: Hybrid-policy self-distillation for text-image-to-video diffusion models. *arXiv preprint arXiv:2608.13205*, 2026.

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

Zihan Ding, Chi Jin, Difan Liu, Haitian Zheng, Krishna Kumar Singh, Qiang Zhang, Yan Kang, Zhe Lin, and Yuchen Liu. Dollar: Few-step video generation via distillation and latent reward optimization. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 17961–17971. IEEE, 2025.

Siming Fu, Zheming Fu, Ruizhe He, Hualiang Wang, Jie Huang, Xiaoxiao Ma, Mingchen Zhong, Weihu Huang, Xiaoxuan He, and Haojun Xu. Any-opd: Heterogeneous on-policy distillation for flow-matching models via representation-space bridging. *arXiv preprint arXiv:2608.03316*, 2026.

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

Xu Guo, Zhengxuan Wei, Xinghui Li, Hanzhuo Huang, Xinyu Liu, Xiangyang Luo, Min Wei, Yiran Zhu, Qiulin Wang, Yulong Xu, et al. Contextmaster: Interactive multi-shot video creation via fixed-budget sparse context routing. *arXiv preprint arXiv:2608.04956*, 2026.

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, 2026.

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.

Yuki Ichihara, Naoto Iwase, Mohammad Atif Quamar, and Junpei Komiyama. Privileged solutions or context-induced teacher behavior? dissecting on-policy self-distillation. *arXiv preprint arXiv:2608.09228*, 2026.

De Jiang, Zhengyang Zhang, Kehong Yuan, and Shaohua Ma. Not every divergence should be suppressed: Counterfactual recoverability in on-policy distillation. *arXiv preprint arXiv:2608.04408*, 2026a.

Dengyang Jiang, Xin Jin, Dongyang Liu, Zanyi Wang, Mingzhe Zheng, Ruoyi Du, Xiangpeng Yang, Qilong Wu, Zhen Li, Peng Gao, et al. D-opsd: On-policy self-distillation for continuously tuning step-distilled diffusion models. *arXiv preprint arXiv:2605.05204*, 2026b.

Simran Kaur, Narutatsu Ri, Yinghui He, Liam Fowl, and Sanjeev Arora. Rethinking on-policy self-distillation for thinking models. *arXiv preprint arXiv:2607.05184*, 2026.

Bingnan Li, Haozhe Wang, Haozhong Xiong, Fangtai Wu, Jinpeng Yu, Yang Shi, Jiaming Liu, and Ruihua Huang. Rethinking classifier-free guidance in on-policy diffusion distillation. *arXiv preprint arXiv:2607.24731*, 2026a.

Quanhao Li, Junqiu Yu, Kaixun Jiang, Yujie Wei, Zhen Xing, Pandeng Li, Ruihang Chu, Shiwei Zhang, Yu Liu, and Zuxuan Wu. Diffusionopd: A unified perspective of on-policy distillation in diffusion models. *arXiv preprint arXiv:2605.15055*, 2026b.

Rui Li, Bingyu Li, Yuanzhi Liang, Haibin Huang, Chi Zhang, and XueLong Li. Reward-aware trajectory shaping for few-step visual generation. *arXiv preprint arXiv:2604.14910*, 2026c.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

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.

David Lopez-Paz, Léon Bottou, Bernhard Schölkopf, and Vladimir Vapnik. Unifying distillation and privileged information. In *International Conference on Learning Representations*, 2016.

Yang Luo, Shengju Qian, Xiaohang Tang, Zirui Zhu, Yong Liu, Xin Wang, and Yang You. On-policy adversarial flow distillation for autoregressive video generation. *arXiv preprint arXiv:2605.26105*, 2026.

Duy Nguyen, Hanqi Xiao, Archiki Prasad, Zaid Khan, Anirban Das, Austin Zhang, Sambit Sahu, Hyunji Lee, Elias Stengel-Eskin, and Mohit Bansal. Avsd: Adaptive-view self-distillation by balancing consensus and teacher-specific privileged signals. *arXiv preprint arXiv:2605.20643*, 2026.

Weili Nie, Julius Berner, Nanye Ma, Chao Liu, Saining Xie, and Arash Vahdat. Transition matching distillation for fast video generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4645–4655, 2026.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

Emiliano Penaloza, Dheeraj Vattikonda, Nicolas Gontier, Alexandre Lacoste, Laurent Charlin, and Massimo Caccia. Privileged information distillation for language models. *arXiv preprint arXiv:2602.04942*, 2026.

Neta Shaul, Chao Liu, Arash Vahdat, and Julius Berner. Parallel decoding distillation for fast image and video generation. *arXiv preprint arXiv:2607.26004*, 2026.

Zhanming Shen, Jintao Tong, Shaotian Yan, Chen Shen, Hao Chen, Wentao Ye, Xiaomeng Hu, Rui Miao, Haobo Wang, Junbo Zhao, et al. Purified opsd: On-policy self-distillation without losing how to think. *arXiv preprint arXiv:2607.02234*, 2026.

Meituan LongCat Team, Xunliang Cai, Qilong Huang, Zhuoliang Kang, Hongyu Li, Shijun Liang, Liya Ma, Siyu Ren, Xiaoming Wei, Rixu Xie, and Tong Zhang. Longcat-video technical report, 2025. URL `https://arxiv.org/abs/2510.22200`.

Kanghui Tian, Siyuan Liu, Ziang Yan, Sheng Xia, Shuai Dong, and Yi Wang. Vicur: Visual cues as recoverable privilege for multimodal on-policy distillation. *arXiv preprint arXiv:2606.05718*, 2026.

Vladimir Vapnik and Akshay Vashist. A new learning paradigm: Learning using privileged information. *Neural networks*, 22(5-6):544–557, 2009.

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.

Yuanyi Wang, Su Lu, Yanggan Gu, Pengkai Wang, Yifan Yang, Zhaoyi Yan, Congkai Xie, Jianmin Wu, and Hongxia Yang. Not all disagreement is learnable: Token teachability in on-policy distillation. *arXiv preprint arXiv:2605.26844*, 2026.

Qingyan Wei, Guangzhao Li, Xiaobing Tu, Yinggui Wang, Xiantao Zhang, Jinkui Ren, Xiaohong Liu, and Linfeng Zhang. Step-opd: Rethinking output targets and internal dynamics in on-policy distillation for diffusion models. *arXiv preprint arXiv:2608.04887*, 2026.

Jianyu Wu, Yizhou Wang, Encheng Su, Chen Tang, and Shixiang Tang. Dapd: Dual-anchored policy distillation. *arXiv preprint arXiv:2608.01735*, 2026.

Lingxiao Yang, Liu Liu, Moran Li, Han Feng, Wenjian Cao, Jiangning Zhang, and Ye Shi. In-context forcing: Uncovering context effects in autoregressive video diffusion. *arXiv preprint arXiv:2608.05237*, 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.

Kangning Zhang, Yixing Li, Shuai Shao, Qingyao Li, Zhengxi Lu, Zhiyuan Yao, Jianghao Lin, Wenxiang Jiao, Yuan Lu, Weiwen Liu, et al. Vad: Attributing visual evidence for target reconstruction in multimodal on-policy distillation. *arXiv preprint arXiv:2607.28590*, 2026.

Siyan Zhao, Zhihui Xie, Mengchen Liu, Jing Huang, Guan Pang, Feiyu Chen, and Aditya Grover. Self-distilled reasoner: On-policy self-distillation for large language models. In *International Conference on Machine Learning*, 2026a.

Xuyang Zhao, Liting Zhang, Zichen Xu, Zhihu Wang, Xu Caiyue, Shiwan Zhao, and Qicheng Li. Is more privileged information better? from solution traces to problem-solving structure in self-distilled reasoning. *arXiv preprint arXiv:2608.01589*, 2026b.

Supplementary Material

CONTENTS

A Training and Conditioning Details **14**

A.1 Native states and evidence encoding 14

A.2 Nested budgets and supervision weights 14

A.3 Optimization procedure 15

B Reference-Blind Regression and Calibration **15**

B.1 Conditional projection and reference-law shift 15

B.2 Properties and limits of the support rule 16

C Estimating Privilege-Induced Teacher Variance **17**

D Experimental Protocol and Comparator Implementations **17**

D.1 Initialization, tensor shapes, and sampling 17

D.2 Simple targets and validation selection 18

D.3 Distribution-matching comparisons 19

D.4 Target-construction controls 19

E Teacher Diagnostics and Independent Target Utility **19**

E.1 Teacher generation and fixed-state reference fidelity 19

E.2 PITV and condition-level transfer 20

E.3 Target-use, local fitting, and learning utility 21

E.4 Within-condition re-pairing protocol 22

F Evaluation, Statistics, and Supplementary Results **22**

F.1 Dataset roles and score definitions 22

F.2 Training variation and paired differences 24

F.3 Blinded human evaluation 25

F.4 LoRA-rank comparisons and post-training cost 25

A Training and Conditioning Details

A.1 Native states and evidence encoding

We use the deployment sampler of each DMD2-initialized few-step generator throughout post-training. The predictor input X_k contains the noisy latent, noise level, deployment condition, coordinates, structural masks, and predictor history. The complete state Σ_k additionally contains solver history and sampling randomness. Teacher queries share the saved student-visible input; a continuation after intervention starts from the complete saved state. Inputs are read-only, and caches affected by evidence or position coordinates are kept separate across queries.

We encode the shared anchor through the native first-frame-only path. Teacher evidence comes from the frozen VAE encoding of the complete paired video, followed by selection of continuation slices. Encoding the full clip preserves temporal context at the continuation boundary, while the student-visible anchor remains independent of future frames. One temporal latent slice contains multiple spatial tokens and may represent several raw frames. The privilege budget counts these slices after excluding the anchor.

Aligned placement assigns privileged tokens their corresponding video coordinates; auxiliary placement uses a separate temporal range. The two layouts retain the same clean values, modulation, physical packing, and allowed attention connections at a fixed budget. Increasing the budget leaves generated-token coordinates and the original-token attention-mask submatrix unchanged. Additional keys can still change the numerical attention weights.

Privileged slices receive local diffusion level zero and remain separate from generated slots. At $B = 0$, the extension reduces to the native positive-branch input. Padding used for absent evidence is masked. Before forming residuals, we map teacher outputs to the canonical generated-token order. The supervision set $\Omega_{i,k}$ contains valid generated outputs only, excluding anchor tokens, privileged slots, and padding. Reference-dependent loss masks are treated separately in Appendix B.

A.2 Nested budgets and supervision weights

An evidence path fixes both a candidate pool and a reveal order. For $L \geq B_{\max}$ eligible continuation slices, a midpoint construction places $B_{\max}$ candidates at zero-based indices

$$t_m = \left\lfloor \frac{(m - \frac{1}{2})L}{B_{\max}} \right\rfloor, \qquad m = 1, \ldots, B_{\max}. \tag{12}$$

For an eight-candidate pool, the permutation $(4, 8, 2, 6, 1, 3, 5, 7)$ indexes candidate identities. Revealing its first B entries yields nested sets without relocating previously exposed evidence. A leading-prefix path is a different construction.

Our main configuration is $(B_{\min}, B_{\text{probe}}, B_{\max}, \gamma) = (1, 4, 8, 1)$. We evaluate Eq. (6) at the native solver noise levels. The resulting integer sequence, rather than uniformly spaced step indices, determines the probe at each state. Reversing this sequence preserves its budget multiset; replacing q_k by $1 - q_k$ need not do so after rounding. A distinct nonempty probe requires $B_{\max} \geq 2$. When $B_{\max} = 2$, only $B_k = 1$ is available; $B_k = 0$ makes the progressive residual vanish, whereas $B_k = B_{\max}$ removes the weak/full comparison.

We sum prediction channels, average valid tokens, and then average examples. For a saved state,

$$\ell_{i,k} = \frac{1}{|\Omega_{i,k}|} \sum_{j \in \Omega_{i,k}} \|v^{s,+}_{i,k,j} - \widetilde{v}^{T}_{i,k,j}\|_2^2. \tag{13}$$

The main experiments supervise every denoising state with $w_k = 1/K$. More generally, for nonnegative weights summing to one, sampling $k \sim p$ with $p_k > 0$ wherever $w_k > 0$ gives

$$\widehat{\mathcal{L}}_i = \frac{w_k}{p_k} \ell_{i,k}, \qquad \mathbb{E}_{k \sim p}[\widehat{\mathcal{L}}_i] = \sum_k w_k \ell_{i,k}. \tag{14}$$

At fixed model snapshots and detached states and targets, the same identity holds for the semi-gradient. It does not differentiate through the state distribution. Training at one fixed state instead changes the objective unless all supervision weight is assigned to that state.

Algorithm 1 One MCPD optimization update

Require: Student coordinates θ, teacher weights $\bar{W}$, batch $\mathcal{D}$ of size N_{eff}, native sampler, state weights w_k, evidence order, budgets, layout ρ, and EMA decay μ.

Clear gradients; hold $W(\theta)$ and $\bar{W}$ fixed until the optimizer step.
for each $(c, x^{\text{gt}}) \in \mathcal{D}$ **do**
  Construct the native condition and detached continuation evidence separately.
  Collect a native student rollout without gradients and save the supervised inputs.
  for each state k with $w_k > 0$ **do**
    Query detached v_k^0, v_k^p, and v_k^f at the same saved input.
    Align generated outputs and compute residuals and g_k on Ω_k.
    Form $\widetilde{v}_k^T = \text{sg}[v_k^0 + g_k \odot \delta_k^f]$.
    Re-evaluate the native student positive prediction with gradients.
    Accumulate the gradient of $w_k \ell_k / N_{\text{eff}}$.
  end for
end for
Apply gradient clipping and take one optimizer step on θ.
Update $\bar{W} \leftarrow \mu \bar{W} + (1-\mu) W(\theta)$.

A.3 Optimization procedure

Let $W(\theta)$ denote effective student weights and $\bar{W}$ the teacher weights. For full-parameter training, $W(\theta) = \theta$; for LoRA, $W(\theta)$ includes the frozen initialization and adapter products. Algorithm 1 keeps both snapshots fixed while collecting rollouts, constructing targets, and accumulating gradients over a global batch. We re-evaluate saved student inputs with gradients and update the teacher only after the student optimizer step.

The algorithm expresses a global-batch average. With G equal-sized data-parallel shards containing $n = N_{\text{eff}}/G$ examples each, averaging gradients of locally $1/n$-normalized losses gives the same gradient. Summed gradient reduction instead requires local normalization by N_{eff}.

For a newly adapted matrix at update u,

$$W_u = W_{\text{init}} + s_{\text{L}} B_{\text{L},u} A_{\text{L},u}, \qquad s_{\text{L}} = \alpha_{\text{L}}/r_{\text{L}}, \tag{15}$$

where W_{init} includes the preceding step-distillation adaptation. The new adapter starts with zero effective update. We maintain its teacher in dense effective-weight space:

$$\begin{aligned} \bar{W}_u &= W_{\text{init}} + \bar{D}_u, \\ \bar{D}_u &= \mu \bar{D}_{u-1} + (1-\mu) s_{\text{L}} B_{\text{L},u} A_{\text{L},u}, \qquad \bar{D}_0 = 0. \end{aligned} \tag{16}$$

EMA values use float32 without rank truncation. Averaging effective updates differs from averaging adapter factors, since $\text{EMA}(B_{\text{L}})\,\text{EMA}(A_{\text{L}}) \neq \text{EMA}(B_{\text{L}} A_{\text{L}})$ in general. The accumulated update $\bar{D}_u$ can also exceed rank r_{L}.

Residual subtraction, norm reductions, and support statistics use float32. We set $\epsilon_a = 10^{-6}$ and $\epsilon_m = 10^{-4}$ on both backbones; their units are squared prediction and prediction units, respectively. Zero progressive residuals yield $a = 0$, and zero full residuals yield no privileged update. Manually setting $g = 0$ or $g = 1$ recovers the anchor-only or full teacher target. Positive-only matching updates shared denoiser parameters, so it can also change negative predictions Li et al. (2026a). The strict null setting and the non-privileged training control are distinguished in Appendix D.1.

B Reference-Blind Regression and Calibration

B.1 Conditional projection and reference-law shift

Fix both model snapshots, the budget, layout, and an X-measurable structural mask. Let $\Delta = \Delta(X, r)$ have finite second moments, and set $\mu_\pi(X) = \mathbb{E}_\pi[\Delta \mid X]$. For any square-integrable reference-blind predictor,

$$\begin{aligned} \mathcal{L}_\pi(h) &= \mathbb{E}_\pi \|h - \mu_\pi + \mu_\pi - \Delta\|_\Omega^2 \\ &= \mathbb{E}_X \|h - \mu_\pi\|_\Omega^2 + \mathbb{E}_\pi \|\Delta - \mu_\pi\|_\Omega^2. \end{aligned} \tag{17}$$

The cross term vanishes because $h - \mu_\pi$ and the mask are X-measurable and $\mathbb{E}_\pi[\Delta - \mu_\pi \mid X] = 0$. Hence μ_π minimizes risk on supervised coordinates. Applying the identity under π_{cmp} with $h = \mu_{\mathrm{pair}}$ proves Eq. (4) when the laws share the same X marginal and the risks are finite. An X-measurable token-count normalization preserves the conditional minimizer.

At a fixed student snapshot, independent rollout and reference sampling conditional on the complete deployment condition gives

$$\pi_{\mathrm{pair}}(X, r \mid C) = p_\theta(X \mid C)\, p_{\mathrm{data}}(r \mid C). \tag{18}$$

This concerns sampling randomness conditional on the trained model, not independence of learned weights from their training data. A comparison that changes only the coupling also preserves the reference marginal conditional on C.

The projection result characterizes a regression target under a specified law. Conditional averaging is appropriate in, for example, the marginal-vector-field construction of flow matching Lipman et al. (2022). The training law π_{pair}, a conceptual comparison law π_{cmp}, and the screened diagnostic law $\mathcal{Q}$ need not coincide. In a finite dataset with one reference per exact condition, memorization can remove empirical conditional variance. PITV under $\mathcal{Q}$ therefore describes reference dependence, rather than establishing irreducible error for the actual finite-capacity student.

Reference-dependent weights. For a nonnegative coordinate weight $\omega_\ell(X, r)$, assume $0 < \mathbb{E}[\omega_\ell \mid X] < \infty$ and $\mathbb{E}[\omega_\ell \Delta_\ell^2 \mid X] < \infty$ almost surely, with finite integrated weighted risk. Completing the weighted square gives

$$h_\ell^\star(X) = \frac{\mathbb{E}[\omega_\ell(X, r)\Delta_\ell(X, r) \mid X]}{\mathbb{E}[\omega_\ell(X, r) \mid X]}. \tag{19}$$

Coordinates with zero expected weight are unconstrained. The weight includes reference-dependent normalization; thus a GT-derived ignore mask can alter the optimum. Fixing one such mask during a diagnostic does not change the weighting rule used in training.

B.2 Properties and limits of the support rule

For nonzero residuals, write $r_p = \|\delta^p\|_2$, $r_f = \|\delta^f\|_2$, and $c = \cos(\delta^p, \delta^f)$. Ignoring the displacement stabilizer,

$$m^2 = \frac{4[(r_f - r_p)^2 + 2r_f r_p(1 - c)]}{(r_f + r_p)^2}. \tag{20}$$

The statistic combines length mismatch and angular disagreement. Removing the cosine factor from the gate therefore produces a displacement-only rule, not a magnitude-only rule. Since $g_j \in [0, 1]$,

$$\|g \odot \delta^f\|_\Omega^2 \le \|\delta^f\|_\Omega^2. \tag{21}$$

For detached teachers and gate, the target also satisfies

$$\begin{aligned}\|s - [(1 - g)v^0 + gv^f]\|_2^2 = (1 - g)\|s - v^0\|_2^2 + g\|s - v^f\|_2^2 \\ - g(1 - g)\|v^f - v^0\|_2^2.\end{aligned} \tag{22}$$

The last term contributes no student gradient. Thus, at a fixed query, MCPD is gradient-equivalent to an adaptive mixture of anchor-only and full-privilege regression losses. This local identity holds with the teacher predictions and sampled states treated as constants.

Under a fixed joint law π, the unrestricted calibrated regression target is

$$v^\star(X) = V(0; X) + \mathbb{E}_\pi[g(X, r) \odot \Delta(X, r) \mid X]. \tag{23}$$

Neither contraction nor this conditional mean specifies a perceptually optimal target. A scalar gate attenuates all channels of a token together. Moreover, if $\delta^p(X, r) = \delta^f(X, r) = \zeta(r)u$ for nonzero u and equiprobable $\zeta \in \{-1, 1\}$, the gate and PITV both approach one as their stabilizers vanish. Conversely, a useful correction revealed only at full budget is rejected when $\delta^p = 0$. These two cases separate within-reference support from cross-reference agreement and explain why the support rule requires empirical evaluation.

Matched-strength targets. Constant and timestep-dependent shrinkage use $v^0 + \alpha\delta^f$ and $v_k^0 + \alpha_k\delta_k^f$. For a fixed sample and state, the energy-matched target is

$$\lambda_E = \begin{cases} \sqrt{\dfrac{\sum_{j\in\Omega} \|g_j\delta_j^f\|_2^2}{\sum_{j\in\Omega} \|\delta_j^f\|_2^2}}, & \sum_{j\in\Omega} \|\delta_j^f\|_2^2 > 0, \\ 0, & \text{otherwise}, \end{cases} \qquad v_j^E = v_j^0 + \lambda_E\delta_j^f. \tag{24}$$

It matches privileged residual energy on the specified outputs, rather than the complete gradient or optimizer update. Replacing anchor-only fallback by a progressive target floor gives

$$[v^p + g \odot (v^f - v^p)] - [v^0 + g \odot (v^f - v^0)] = (1 - g) \odot \delta^p. \tag{25}$$

This alternative retains weak privilege at low-support locations and changes target energy as well as the fallback prediction.

C Estimating Privilege-Induced Teacher Variance

PITV is computed with fixed model snapshots, student-visible inputs, budget, layout, and structural mask. References are sampled independently of student rollout randomness conditional on C and accepted by the screening rule in Appendix E.2. This procedure defines the diagnostic law $\mathcal{Q}(r \mid C)$.

For $M \geq 2$ corrections at one state,

$$\begin{aligned} \widehat{\mu} &= \frac{1}{M}\sum_{m=1}^{M} \Delta_m, \\ \widehat{V} &= \frac{1}{M}\sum_{m=1}^{M} \|\Delta_m - \widehat{\mu}\|_\Omega^2, \qquad \widehat{E} = \frac{1}{M}\sum_{m=1}^{M} \|\Delta_m\|_\Omega^2, \\ \widehat{\mathrm{PITV}} &= \frac{\widehat{V}}{\widehat{E} + \epsilon_P}. \end{aligned} \tag{26}$$

The $1/M$ convention yields $\widehat{E} = \widehat{V} + \|\widehat{\mu}\|_\Omega^2$ and a ratio in $[0, 1]$. For i.i.d. accepted references, $\mathbb{E}[\widehat{V}] = (M-1)V/M$. This factor does not in general describe the bias of the ratio, so we retain the empirical ratio without multiplying it by $M/(M-1)$.

Let $D_{i,k}$ count supervised scalar coordinates and define $\widehat{e}_{i,k} = \widehat{E}_{i,k}/D_{i,k}$ and $\widehat{\nu}_{i,k} = \widehat{V}_{i,k}/D_{i,k}$. With $\epsilon_P = D_{i,k}\epsilon_{\mathrm{coord}}$ and nonnegative state weights summing to one,

$$e_i = \sum_k w_k\widehat{e}_{i,k}, \qquad \nu_i = \sum_k w_k\widehat{\nu}_{i,k}, \qquad \widehat{P}_i = \frac{\nu_i}{e_i + \epsilon_{\mathrm{coord}}}. \tag{27}$$

Without the stabilizer, this is an energy-weighted combination of state-level ratios. We set $\epsilon_{\mathrm{coord}} = 10^{-12}$ and exclude conditions with $e_i < \tau = 10^{-10}$ in native squared-prediction units after state aggregation. For retained conditions,

$$\widehat{P}_i = \frac{\nu_i}{e_i}\frac{e_i}{e_i + \epsilon_{\mathrm{coord}}}, \qquad \frac{100}{101} \leq \frac{e_i}{e_i + \epsilon_{\mathrm{coord}}} < 1. \tag{28}$$

The stabilizer therefore reduces a nonzero unstabilized ratio by at most approximately 0.99%. This bounds its denominator effect, not finite-reference or forward-prediction error. At $B = 0$, all privileged corrections vanish. Energy, variance, and their ratio remain separate quantities: large relative variation can occur at small absolute energy, while low variation alone does not establish a useful correction.

D Experimental Protocol and Comparator Implementations

D.1 Initialization, tensor shapes, and sampling

We construct four-step Wan2.2 and eight-step LongCat initializations with DMD2 Yin et al. (2024a). The base models are `Wan-AI/Wan2.2-TI2V-5B` and LongCat-13.6B. Preceding distillation

adapters are merged before post-training, and each family uses one shared effective initialization. Wan2.2 learns a new adapter on attention projections `q`, `k`, `v`, `o`; LongCat updates the full denoiser. Table 4 lists the main settings. Wan2.2 uses a per-device microbatch of one and 16 accumulation steps.

Table 4: Main post-training configuration. Sizes are height × width. Euler denotes the flow-matching Euler solver used by the constructed few-step initialization.

Setting	Wan2.2	LongCat-13.6B
Sampling steps	4	8
Solver / shift / CFG	Euler / 3.0 / 5.0	Euler / 3.0 / 5.0
Prediction	$v = \epsilon - y$	$v = \epsilon - y$
Adaptation	LoRA 128, $\alpha = 128$	Full parameter
Training size / frames	256×416 / 53	256×256 / 33
Training temporal slices	14	9
Token grid per slice	8×13	16×16
Evaluation size / frames	480×832 / 81	512×512 / 65
Effective batch size	128	128
Maximum generator updates	10,000	10,000
Base learning rate	10^{-5}	5×10^{-6}
AdamW (β_1, β_2)	$(0.9, 0.999)$	$(0.9, 0.999)$
Initial moment states	$m_0 = v_0 = 0$	$m_0 = v_0 = 0$
Weight decay	0.01	0.01
Teacher EMA decay	0.999	0.999
$(B_{\min}, B_{\text{probe}}, B_{\max})$	$(1, 4, 8)$	$(1, 4, 8)$
Schedule γ / gate β	1.0 / 2.0	1.0 / 2.0
(ϵ_a, ϵ_m)	$(10^{-6}, 10^{-4})$	$(10^{-6}, 10^{-4})$

Spatial and temporal accounting. Wan2.2 training pixels have shape $[N, 3, 53, 256, 416]$. Temporal stride four and spatial stride sixteen produce latents $[N, C_z, 14, 16, 26]$; 2×2 spatial patches yield an 8×13 token grid per slice. LongCat processes 33 frames at 256×256 into nine temporal slices with a 16×16 token grid. Excluding the anchor leaves 13 and 8 continuation slices. Thus $B_{\max} = 8$ covers $8/13$ and $8/8$ of the eligible slices at the respective training sizes.

Flow convention. Both implementations use $z_\sigma = (1 - \sigma)y + \sigma\epsilon$ and $v = \epsilon - y$. The implied clean prediction is $\widehat{y} = z_\sigma - \sigma v$. CFG gives $v_{\text{guided}} = v^- + w_{\text{cfg}}(v^+ - v^-)$, followed by the Euler update $z_{\sigma'} = z_\sigma + (\sigma' - \sigma)v_{\text{guided}}$ and native anchor handling. The shift is 3.0. These are the deployment settings of the few-step initializations; continuous solver noise and model timestep are distinct quantities.

EMA-SD and the null setting. With identical deterministic predictions and zero optimizer moments, the initial matching gradient is zero. A Wan2.2 adapter initialized with $B_{\text{L}} = 0$ has zero effective update even if weight decay rescales A_{L}. This describes the mathematical null setting. The executed Wan2.2 EMA-SD runs have mean I2V 84.7 and run-level SD 0.03; their score summary alone is not an exact null test. For full-parameter LongCat, weight decay can first produce $W_1 = (1 - \eta\lambda_{\text{wd}})W_0$ at zero matching gradient. Subsequent student–EMA differences can generate nonzero matching gradients. LongCat EMA-SD gives 85.1 ± 0.12, compared with the fixed BASE score of 85.3. We treat the executed runs as non-privileged controls and keep them distinct from the deterministic null construction.

D.2 SIMPLE TARGETS AND VALIDATION SELECTION

SFT predicts $\epsilon - y^{\text{gt}}$ on data-noised states with a shared first-frame condition and no anchor-output loss. FULL-PRIV PD matches v^f without clipping or gating. BEST-FIXED PD selects $B = 4$ from $\{1, 2, 4, 8\}$ on validation data. BEST-SCALAR uses $v^0 + \alpha\delta^f$. STEP-SCALAR uses early/late coefficients selected from $\{0.25, 0.5, 0.75, 1.0\}^2$, with $(\alpha_{\text{e}}, \alpha_{\text{l}}) = (0.75, 0.50)$. All methods are trained for at most 10,000 generator updates. Checkpoints at $\{100, 300, 1000, 3000, 6000, 10000\}$ are evaluated on the 200 validation conditions, and main-table results use the selected checkpoints. Test scores do not enter selection.

D.3 DISTRIBUTION-MATCHING COMPARISONS

DMD2-CONTINUE Yin et al. (2024a) and DFD Chen et al. (2026) share the common initialization and are adapted to shared-anchor I2V, denoted by †. Both train a fake-score network with AdamW at learning rate 10^{-5}, using five fake-network updates per generator update. Distribution-matching losses exclude anchor coordinates.

For the affine Gaussian path, a conditional mean velocity and its score satisfy, for $\sigma > 0$,

$$\begin{aligned} s_v(z,\sigma,C) &= -\frac{z-(1-\sigma)\widehat{y}_v(z,\sigma,C)}{\sigma^2} = -\frac{z+(1-\sigma)v(z,\sigma,C)}{\sigma}, \\ \widehat{y}_v(z,\sigma,C) &= z-\sigma v(z,\sigma,C). \end{aligned} \tag{29}$$

Applying this expression to learned velocities gives the corresponding score estimates. Let x^f, x^r be generated and real clean samples, with $x^a_\sigma = (1-\sigma)x^a + \sigma\epsilon^a$ for $a \in \{f, r\}$. Writing $J_\theta = \partial x^f/\partial\theta$ and absorbing the noising-chain factor into $w(\sigma)$, the DFD generator direction is

$$\mathbb{E}\big[J_\theta^\top w(\sigma)\left\{s_\phi(x^f_\sigma, C) - (1-p)s_T(x^f_\sigma, C) - p\, s_T(x^r_\sigma, C)\right\}\big]. \tag{30}$$

Here s_T and s_ϕ are teacher and fake-score estimates. We use $p = 0.5$, omit the DFD adversarial discriminator, and add no separate real-data flow-regression term.

HPSD Bu et al. (2026) initializes student sub-trajectories from privileged teacher states; OPSD-V Liu et al. (2026) constructs teacher caches from real-video history for autoregressive generation. FULL-PRIV PD is an uncalibrated control for our state construction, not a reproduction of these methods. Our empirical comparisons are restricted to the implementations evaluated here.

D.4 TARGET-CONSTRUCTION CONTROLS

ENERGY-MATCHED applies Eq. (24) on its own current states. FIXED-PROBE retains all three queries and fixes the weak budget at four. COSINE-ONLY uses $g = a$, removing the displacement penalty and changing attenuation strength. THREE-QUERY MIX uses

$$\widetilde{v}_{\text{mix}} = v^0 + \alpha[(1-\lambda)\delta^p + \lambda\delta^f], \tag{31}$$

with validation-selected $(\alpha, \lambda) = (1, 0.5)$.

SHIFTED-GATE cyclically translates gates by one transformer-grid position along each spatial axis without changing time. It preserves gate values but generally changes residual energy. CR-VEL uses a consensus–residual target inspired by AVSD Nguyen et al. (2026): $c_j = (\delta^p_j + \delta^f_j)/2$, $d_j = \delta^f_j - c_j$, and $t^{\text{CR}}_j = v^0_j + \alpha(c_j + h_j d_j)$, with cosine-based h_j and $\alpha = 1$. This is a velocity-space analogue rather than a probability-space reproduction. The measured comparisons concern these target constructions as a whole.

Local calibration versus uniform attenuation. Figure 5 complements Table 2 with native I2V outputs from students post-trained using different target constructions. In the displayed frames, ENERGY-MATCHED keeps the cat near its starting position. Both FULL-PRIV PD and MCPD show a leap, while MCPD's last displayed frame shows the cat resting on the opposite slab.

E TEACHER DIAGNOSTICS AND INDEPENDENT TARGET UTILITY

E.1 TEACHER GENERATION AND FIXED-STATE REFERENCE FIDELITY

Figure 3(a) evaluates an independent frozen diagnostic teacher using its own 50-step Euler rollout, fixed continuation evidence, and the I2V Total Score Q_B. The students in panel (b) instead train with their own evolving few-step EMA teachers.

For the training-teacher probe, we freeze the FULL-PRIV PD EMA at update 5,000 and query 256 saved Wan2.2 validation states. The implied clean prediction is $\widehat{y}^B_{i,k} = z^s_{i,k} - \sigma_k V_k(B, \rho; r_i)$. A common region $\mathcal{H}_i$ excludes the anchor and the union of exposed-latent neighborhoods across budgets.

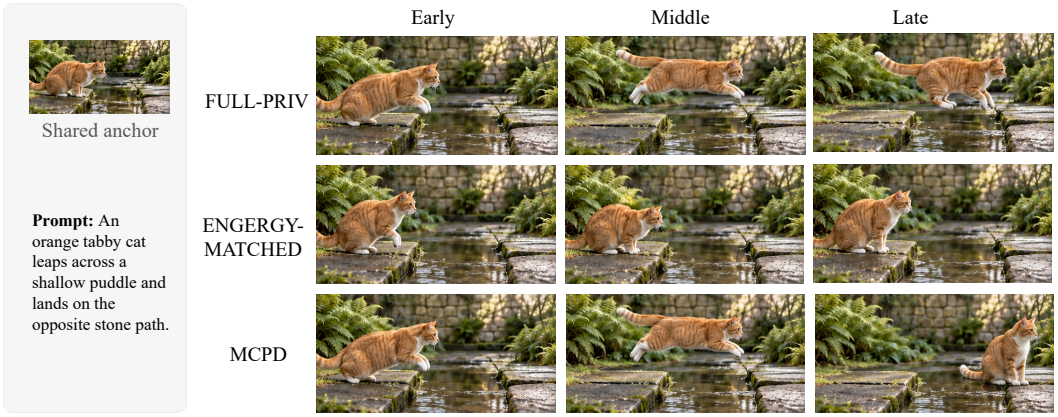

Figure 5: **Qualitative ablation of privileged-target calibration.** Rows show native I2V outputs from students post-trained with FULL-PRIV PD (FULL), ENERGY-MATCHED, or MCPD. Methods share the prompt, input frame, generation settings, and displayed frame indices. Teacher-only continuation evidence is absent at inference.

With $\mathcal{I}_H = \{i : D_{\mathcal{H}_i} > 0\}$ and $N_H = |\mathcal{I}_H|$,

$$E_B = \frac{1}{N_H} \sum_{i \in \mathcal{I}_H} \sum_k w_k \frac{\|\widehat{y}^B_{i,k} - y^{\mathrm{gt}}_i\|^2_{\mathcal{H}_i}}{D_{\mathcal{H}_i}}. \tag{32}$$

Relative reductions $100(1 - E_B/E_0)$ are 14.2%, 27.5%, 43.1%, and 61.8% at $B = 1, 2, 4, 8$. The common mask makes the budgets comparable on the same outputs. This evaluates aggregate reference adherence at a frozen EMA snapshot, separately from complete teacher generation and student learning.

E.2 PITV AND CONDITION-LEVEL TRANSFER

We use 200 conditions with eight continuations sampled from the base multi-step generator. References pass VBench `overall_consistency`, based on ViCLIP text–video similarity, above 0.25 and intra-video `subject_consistency` above 0.70. I2V references share the input anchor. The condition-level energy threshold in Appendix C removes 16 conditions, leaving 184.

For method m, the condition score averages a per-video I2V composite over five training runs and four generation seeds:

$$\begin{aligned} q_i^m &= \frac{1}{20} \sum_{b=1}^{5} \sum_{s=1}^{4} q_{\mathrm{I2V}}(Y^m_{i,b,s}, C_i), \\ \Delta_i^{\mathrm{tr}} &= q_i^{\text{Full-Priv PD}} - q_i^{\text{EMA-SD}}. \end{aligned} \tag{33}$$

We use VBench v1.0 min–max bounds and equal weights for Quality Score and I2V Score, on a 0–100 scale. This differs from the 4:1 quality/semantic weighting for T2V and from a complete-suite aggregate. Figure 3(c) gives $\rho = -0.83$ (95% CI $[-0.88, -0.76]$) between $\widehat{P}_i$ and Δ_i^{tr}.

Reference audit. Three blinded raters assess 800 references from 100 conditions, with a video-level action-compliance rate of 86%. All eight references are unanimously action-compliant for 45 conditions. The restricted correlation is $\rho = -0.79$ (95% CI $[-0.87, -0.65]$). The association therefore persists in this action-compliant subset, without treating the audit as a test of every aspect of reference quality or compatibility.

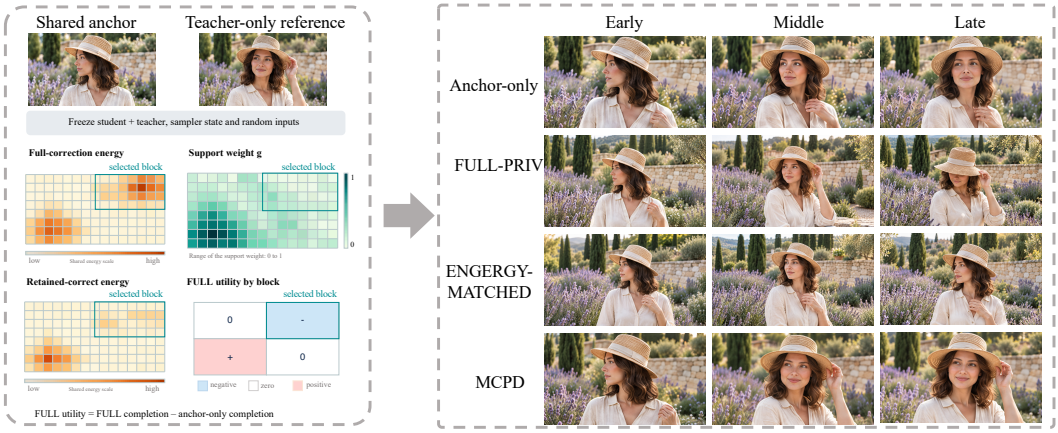

Figure 6: **Schematic illustration of fixed-state target-use diagnostics. Left:** The shared anchor and teacher-only reference contextualize the full-correction energy, support weight g, retained-correction energy, and blockwise FULL utility. The energy maps share a schematic color scale. The utility map depicts one temporal half of the $2 \times 2 \times 2$ partition; each sign represents a separate hypothetical block intervention relative to anchor-only. **Right:** Conceptual early, middle, and late stages for ANCHOR-ONLY, FULL, ENERGY-MATCHED, and MCPD. At a single saved state, the illustrated protocol replaces only the positive-branch prediction within the selected block and then resumes sampling. ENERGY-MATCHED matches MCPD's privileged residual energy within that block. All portraits and maps are synthetic illustrations, not outputs or measurements from intervention rollouts.

E.3 TARGET-USE, LOCAL FITTING, AND LEARNING UTILITY

Block interventions. We save complete student states and partition generated outputs into a $2 \times 2 \times 2$ temporal–height–width grid. For a valid block $\mathcal{B} \subseteq \Omega_{i,k}$ and target $t \in \{0, f, E, \mathrm{MCPD}\}$,

$$v_{i,k,j}^{t,+,\mathcal{B}} = v_{i,k,j}^{s,+} + \mathbf{1}[j \in \mathcal{B}]\big(v_{i,k,j}^{t,+} - v_{i,k,j}^{s,+}\big). \tag{34}$$

The negative prediction at this state is unchanged. Native CFG is applied, and subsequent predictions are recomputed on the intervened trajectory. Branches share subsequent random increments where the sampler uses them.

Direct-use utility. The diagnostic combines Subject Consistency, Background Consistency, Motion Smoothness, Dynamic Degree, Aesthetic Quality, and Imaging Quality:

$$q_{\mathrm{diag}}(Y, C) = s_{\mathrm{diag}} \sum_{\ell \in \mathcal{D}_6} \lambda_\ell^{\mathrm{diag}} \mathcal{N}_\ell\big(d_\ell(Y, C)\big). \tag{35}$$

Here d_ℓ, $\mathcal{N}_\ell$, $\lambda_\ell^{\mathrm{diag}}$, and s_{diag} denote dimension outputs, normalizations, weights, and output scale. Utility is measured relative to the corresponding anchor-only intervention:

$$U_{\mathrm{priv}}^{t,\mathcal{B}} = q_{\mathrm{diag}}(Y^{t,\mathcal{B}}, C) - q_{\mathrm{diag}}(Y^{0,\mathcal{B}}, C). \tag{36}$$

For ENERGY-MATCHED, λ_E is computed within the intervened block. Support cut points 0.35 and 0.72 are set on validation data. As g approaches zero, the MCPD target approaches anchor-only, making near-zero incremental utility consistent with conservative fallback. Figure 3(d) evaluates direct use; no parameters are updated.

Visual interpretation of fixed-state diagnostics. Figure 6 schematically illustrates the direct-use protocol above. The FULL utility map corresponds to $U_{\mathrm{priv}}^{f,\mathcal{B}}$, whose baseline is the anchor-only teacher intervention $Y^{0,\mathcal{B}}$. Positive utility therefore indicates improvement over this intervention baseline, not necessarily over the unmodified student. At a fixed sample and saved state, we suppress

the sample and denoising-step indices. For a d-dimensional output-token correction δ_j^f, the full and retained energies are

$$e_j^f = \frac{\|\delta_j^f\|_2^2}{d}, \qquad e_j^{\mathrm{MCPD}} = g_j^2 e_j^f.$$

Displaying correction energy alongside g_j distinguishes support weight from the amount of correction retained. Energy attenuation follows from the target construction and does not itself establish beneficial target use. Within-block ENERGY-MATCHED preserves MCPD's total privileged residual energy in the intervened block while replacing tokenwise scaling by the common coefficient λ_E. If g_j is constant across tokens with nonzero full corrections, the two targets coincide. The schematic thus distinguishes correction magnitude, budget support, and downstream utility; it does not provide a measured comparison of the intervention outcomes.

Local fitting statistic. For fixed states, masks, and detached targets, let $E_{t,\mathrm{pre}}$ and $E_{t,\mathrm{post}}$ be regression errors before and after n updates. The local fit ratio is

$$\mathrm{LFR}_{t,n} = \frac{E_{t,\mathrm{pre}} - E_{t,\mathrm{post}}}{\max(E_{t,\mathrm{pre}}, \epsilon_{\mathrm{fit}})}. \tag{37}$$

It measures fitting on saved queries, whereas the learning probe evaluates generation outside the update set.

Learning utility. We take five updates at learning rate 10^{-5} on 128 conditions and evaluate 128 disjoint conditions. Let $\theta_t^{(n)}$ and $\theta_0^{(n)}$ denote target-specific and anchor-only updated students. The learning composite uses the VBench-I2V dimensions applicable to each condition:

$$q_{\mathrm{Total}}(Y, C) = s_{\mathrm{Total}} \sum_{\ell \in \mathcal{D}(C)} \lambda_{\ell,C}^{\mathrm{Total}} \mathcal{N}_\ell\big(d_\ell(Y, C)\big), \tag{38}$$

where coefficients include dimension weighting and output scale. With S_Y seeds per evaluation condition,

$$U_{\mathrm{learn}}^{t,n} = \frac{1}{N_{\mathrm{test}} S_Y} \sum_{i,s} \left[q_{\mathrm{Total}}(Y_{i,s}^{\theta_t^{(n)}}, C_i) - q_{\mathrm{Total}}(Y_{i,s}^{\theta_0^{(n)}}, C_i) \right]. \tag{39}$$

The means are $+0.15$ (95% CI $[0.08, 0.22]$) for MCPD and -0.08 ($[-0.15, -0.01]$) for FULL-PRIV PD. We also compute their contrast within each evaluation condition and resample complete conditions 5,000 times. The mean contrast is $+0.23$ with a paired bootstrap interval $[0.12, 0.34]$. It measures the difference between the two short updates, not either update's improvement over the unmodified model. This per-video score and the block-intervention composite are separate from dataset-level official Total Scores.

E.4 WITHIN-CONDITION RE-PAIRING PROTOCOL

A complementary pairing construction holds fixed eight student rollouts and eight accepted references under the same prompt and anchor. Define the corresponding-frame DINOv2 cost

$$D_{i,sm} = \frac{1}{L_D} \sum_{\ell=1}^{L_D} d_D\big(\phi_D(Y_{i,s}^{\mathrm{native}}(\tau_\ell)), \phi_D(r_{i,m}(\tau_\ell))\big). \tag{40}$$

Minimum- and maximum-distance assignments minimize and maximize $\sum_s D_{i,s,\pi(s)}$ over bijections π, while random assignment uses a uniform permutation. Each construction preserves both empirical marginals and changes only their pairing. The distance is an appearance-and-structure proxy computed without utility outcomes. This protocol is distinct from Figure 3(d)'s support strata; no separate re-pairing result is reported.

F EVALUATION, STATISTICS, AND SUPPLEMENTARY RESULTS

F.1 DATASET ROLES AND SCORE DEFINITIONS

We curate 30K clips from WebVid-10M and Panda-70M. Source-video identifiers or canonical URLs, where available, identify training overlaps with validation and evaluation media for removal.

Table 5: Dataset roles. Counts distinguish clips from conditions; A-Div. uses a subset of the official evaluation suite.

Set	Size	Role
Post-training	30K clips	Filtered WebVid-10M and Panda-70M mixture.
Main evaluation	Full suites	Official VBench++ I2V and VBench T2V conditions.
Validation	200 conditions	Hyperparameter and checkpoint selection.
PITV diagnostic	200 conditions	Eight references each; 184 conditions after energy filtering.
A-Div.	200 conditions	Official-suite subset with eight generations per condition.
Learning probe	256 conditions	128 update and 128 disjoint evaluation conditions.

This procedure covers identifiable media rather than a complete deduplication guarantee for all prompt conditions. A-Div. intentionally overlaps its parent suite. Clips used at the maximum budget have at least eight eligible continuation slices after excluding the anchor.

Main results use the complete official VBench++ I2V and VBench T2V suites Huang et al. (2025; 2024). The PITV set does not replace those suites. The learning probe's 128 update and 128 evaluation conditions are disjoint; each condition can generate multiple videos and denoising states.

Appearance diversity. We extract DINOv2 `vitb14` features from $S = 8$ matched generations for $N_A = 200$ conditions and average over video time before normalization:

$$\bar{\phi}_{i,s} = \frac{1}{L_A}\sum_{\ell=1}^{L_A}\phi(Y_{i,s}(\tau_\ell)), \qquad e_{i,s} = \frac{\bar{\phi}_{i,s}}{\|\bar{\phi}_{i,s}\|_2 + \epsilon_A},$$
$$A_i = \frac{2}{S(S-1)}\sum_{s<s'}(1 - e_{i,s}^\top e_{i,s'}), \qquad \text{A-Div.} = \frac{1}{N_A}\sum_i A_i. \tag{41}$$

Temporal averaging discards ordering. A-Div. thus describes appearance-representation variation, not temporal modes. Run-level SDs range from 0.003 to 0.005. The reported MCPD–BASE contrast is $+0.002$ with SD 0.004 and $p = 0.33$. A failure to reject a zero difference does not establish equivalence; we describe the observed means as close rather than infer statistical preservation.

Table 6: Wan2.2 motion and intra-video consistency. Trained methods report means ± run-level SD across five runs; BASE is a fixed reference. Scores are raw dimensions multiplied by 100. Tests against BASE, where used, concern mean post-training performance relative to that fixed score.

Method	Dynamic Degree	Motion Smoothness	Subject Consistency
BASE	71.6	94.2	66.8
FULL-PRIV PD	66.9 ±0.72	94.6 ±0.16	67.5 ±0.21
STEP-SCALAR	71.9 ±0.64	94.3 ±0.13	67.2 ±0.18
DFD†	72.4 ±0.58	94.1 ±0.17	66.7 ±0.19
MCPD	72.8 ±0.55	94.7 ±0.14	67.6 ±0.17

Motion and consistency. On Wan2.2, MCPD's DD gain over FULL-PRIV PD is $+5.9$ (paired SD 0.70; 95% CI $[5.03, 6.77]$; $p < 0.001$). The gain over STEP-SCALAR is $+0.9$ (SD 0.62; $p = 0.031$), whereas the $+0.4$ difference from DFD has SD 0.58 and CI $[-0.32, 1.12]$ ($p = 0.20$). On LongCat, DD improves by $+1.3$ over STEP-SCALAR (SD 0.71; $p = 0.015$). The results show a larger activity gain over uncalibrated privilege than over the strongest alternatives. Wan2.2 smoothness and subject-consistency means differ from FULL-PRIV PD by 0.1 point each; their individual method SDs do not determine a paired equivalence test.

Additional score variation. Excluding EMA-SD, run-level SDs range from 0.21–0.32 for I2V and 0.22–0.34 for T2V on Wan2.2, and 0.27–0.38 and 0.26–0.36 on LongCat. The EMA-SD I2V SDs are 0.03 and 0.12, respectively. MCPD's T2V SDs are 0.28/0.31; relative to fixed BASE scores, the gains are 1.4/1.6 with CIs $[1.05, 1.75]$/$[1.21, 1.99]$. The rank comparisons have paired SD ranges 0.55–0.71 for DD and 0.24–0.31 for T2V. These quantities refer to different contrasts and are not interchangeable.

F.2 TRAINING VARIATION AND PAIRED DIFFERENCES

For a fixed backbone, initialization, and comparator, let S_b^m be the benchmark score at method m's validation-selected checkpoint in replicate b. We compute

$$d_b = S_b^{\mathrm{MCPD}} - S_b^{\mathrm{control}}, \qquad \bar{d} = \frac{1}{5}\sum_{b=1}^{5} d_b, \qquad s_d = \sqrt{\frac{1}{4}\sum_{b=1}^{5}(d_b - \bar{d})^2}. \tag{42}$$

Paired runs use matched data order and evaluation design. The paired variance obeys

$$s_d^2 = s_{\mathrm{MCPD}}^2 + s_{\mathrm{control}}^2 - 2\widehat{\mathrm{Cov}}(S^{\mathrm{MCPD}}, S^{\mathrm{control}}). \tag{43}$$

Positive covariance can reduce difference variance, but pairing does not guarantee an SD smaller than either marginal SD. The bound $|s_{\mathrm{MCPD}} - s_{\mathrm{control}}| \le s_d \le s_{\mathrm{MCPD}} + s_{\mathrm{control}}$ always holds. For fixed BASE, subtracting its score leaves $s_d = s_{\mathrm{MCPD}}$; inference is a one-sample comparison against that fixed reference rather than a second set of training replicates.

We compute two-sided t-tests and pointwise 95% intervals from unrounded differences:

$$t = \frac{\bar{d}}{s_d/\sqrt{5}}, \qquad \mathrm{CI}_{95\%} = \bar{d} \pm t_{0.975,4}\frac{s_d}{\sqrt{5}}, \qquad t_{0.975,4} = 2.776. \tag{44}$$

The tests assume independent run pairs and approximately normal differences. With five pairs, tail behavior is weakly assessed. The reported p-values and intervals are nominal, comparison-wise summaries, not multiplicity-adjusted familywise claims. Tables 7 and 8 collect the principal and target-construction comparisons.

Table 7: Principal I2V contrasts: MCPD minus comparator. Mean differences and paired SDs are score points. CIs and two-sided p-values use five run differences with df $= 4$. For fixed BASE, the difference SD equals MCPD's score SD.

Backbone	Comparator	Gain (SD)	95% CI	p
Wan2.2	BASE	+2.5 (0.24)	[2.20, 2.80]	< 0.001
	FULL-PRIV PD	+1.6 (0.24)	[1.30, 1.90]	< 0.001
	STEP-SCALAR	+0.6 (0.18)	[0.38, 0.82]	0.002
	DFD†	+0.7 (0.19)	[0.46, 0.94]	0.001
LongCat	BASE	+2.8 (0.29)	[2.44, 3.16]	< 0.001
	FULL-PRIV PD	+2.0 (0.27)	[1.66, 2.34]	< 0.001
	STEP-SCALAR	+0.9 (0.21)	[0.64, 1.16]	< 0.001
	DFD†	+0.8 (0.22)	[0.53, 1.07]	0.001

Table 8: Paired Wan2.2 I2V differences between MCPD and the target-construction controls. Results summarize the same five-run comparisons as Table 2; p-values are two-sided and nominal.

Comparator	Mean gain	Paired SD	p
FULL-PRIV PD	+1.6	0.24	< 0.001
ENERGY-MATCHED	+1.1	0.20	< 0.001
FIXED-PROBE	+0.4	0.17	0.006
COSINE-ONLY	+1.2	0.22	< 0.001
THREE-QUERY MIX	+0.9	0.19	< 0.001
SHIFTED-GATE	+1.4	0.24	< 0.001
CR-VEL	+0.7	0.19	0.001

All five runs share one DMD2 initialization per backbone. The intervals quantify variability present in the recorded score pairs; independent reconstruction of the initialization and separate benchmark or evaluation-seed resampling are not included. Common implementation or evaluator effects are likewise not estimated by run dispersion. Condition-bootstrap intervals use 5,000 resamples of complete conditions with paired observations retained, conditional on the evaluated checkpoints and other quantities held fixed. They measure a different uncertainty from training-run intervals.

F.3 BLINDED HUMAN EVALUATION

Ten raters evaluate 50 video pairs, 25 against BASE and 25 against FULL-PRIV PD. Each pair is rated for overall preference, motion naturalness, and subject consistency, yielding 1,500 criterion-level judgments. Method identities are hidden and display order randomized. Overall-preference agreement is Krippendorff's $\alpha = 0.44$. Table 9 reports counts from MCPD's perspective and pair-clustered intervals for each criterion.

Table 9: Human preference for MCPD. Each row contains ten judgments per video pair for 25 pairs. Non-tie preference is $W/(W + L)$. The 95% intervals resample complete pairs and are pointwise across criteria.

Comparator	Criterion	W	T	L	Non-tie (%)	95% CI (%)
BASE	Overall	165	40	45	78.6	[72.1,84.2]
	Motion	152	40	58	72.4	[65.8,79.0]
	Subject	110	60	80	57.9	[50.6,65.2]
FULL-PRIV PD	Overall	150	32	68	68.8	[61.5,75.8]
	Motion	145	30	75	65.9	[58.9,72.9]
	Subject	115	50	85	57.5	[50.2,64.8]

Overall tie rates are 16.0% against BASE and 12.8% against FULL-PRIV PD, or 14.4% pooled. Overall and motion preference intervals lie above 50%. Subject-consistency preferences are smaller: their intervals also exclude 50%, but only narrowly at the lower endpoints of 50.6% and 50.2%. Pair resampling retains all ten judgments together and supports inference over pairs conditional on this rater panel. The study covers these comparators, not DFD or STEP-SCALAR.

F.4 LORA-RANK COMPARISONS AND POST-TRAINING COST

At Wan2.2 ranks 64, 128, and 256, MCPD's I2V gains over STEP-SCALAR are 0.4, 0.6, and 0.7 with paired SDs 0.16, 0.18, and 0.20. The 95% intervals are $[0.20, 0.60]$, $[0.38, 0.82]$, and $[0.45, 0.95]$, with nominal $p = 0.005, 0.002, 0.001$. LongCat's full-parameter result is $+0.9$ (SD 0.21; CI $[0.64, 1.16]$; $p < 0.001$). These estimate method differences within the tested settings, not the difference of gains between ranks or a controlled adaptation-type effect.

For method m, GPU-hours and relative cost are

$$H_m = \sum_{d \in \mathcal{D}_m} \frac{T_{m,d}}{3600}, \qquad \mathrm{CostRatio}_m = \frac{H_m}{H_{\text{Step-Scalar}}}, \tag{45}$$

where $T_{m,d}$ is allocated device time in seconds. On eight H800 GPUs, measured MCPD / STEP-SCALAR GPU-hours are 172/150, 175/152, and 179/155 for Wan2.2 ranks 64, 128, and 256, and 263/223 for LongCat. The main rank-128 and LongCat ratios are approximately 1.15 and 1.18. A repeat timing of Wan2.2 rank 128 differed by 2.1%. Timing includes data loading, forward/backward passes, and optimizer updates, excluding validation search and checkpoint serialization. These allocation costs are separate from validation-selected checkpoint quality.