GEOMETRY-GUIDED TEMPORAL ALLOCATION FOR FLOW MATCHING DISTILLATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Flow matching and diffusion models achieve state-of-the-art image synthesis but require slow, iterative inference. While few-step distillation reduces this cost, these methods conventionally distribute training updates natively across time. We propose Complexity-Weighted Rectified Distillation (CWRD), a plug-and-play training wrapper that uses a frozen teacher-geometry profile to reweight timestep sampling. The profile is based on Local Complexity (LC), which measures how the dominant endpoint-sensitivity subspace varies across neighboring latent states. CWRD increases the relative sampling probability of high-LC timesteps, reallocating temporal capacity allocation while strictly preserving the host method's per-example loss, student architecture, and inference procedure. We evaluate CWRD on π-Flow, LADD, TDM, and MFD using high-resolution Stable Diffusion 3.5, SANA and FLux.2 teachers. Across the evaluated configurations, CWRD reduces teacher-referenced FID and improves semantic prompt alignment, with absolute data-referenced FID improvements dependent on the specific host architecture. Complementary trajectory diagnostics show that following a teacher-trajectory prefix before a single suffix update can preserve the teacher output better than native sampling at a matched evaluation budget. These findings suggest that endpoint-sensitivity variation provides a useful empirical signal for temporal allocation in few-step generative models.

1 INTRODUCTION

Diffusion and flow matching models achieve high-quality, high-resolution text-to-image synthesis (Rombach et al., 2022; Esser et al., 2024; Labs et al., 2025). However, their sampling procedures typically require repeated neural network evaluations, making inference computationally expensive. Few-step distillation addresses this cost by transferring a pretrained teacher's generative behavior to a student requiring fewer evaluations. Existing approaches include latent consistency distillation (Luo et al., 2023), trajectory rectification (Liu et al., 2022), and adversarial distillation (Sauer et al., 2024b;a). More recently, π-Flow (Chen et al., 2026) trains a student to predict a network-free policy that produces velocities at multiple integration substeps without additional network evaluations.

Alongside the choice of student parameterization and training objective, distillation requires deciding how to distribute training updates across time. Existing methods employ different timestep distributions and weighting strategies, including transformed time sampling (Chen et al., 2026). Temporal-supervision methods such as SenseFlow (Ge et al., 2025) further transfer teacher behavior within time segments to coarse student anchors. These developments motivate a specific question: can a teacher's spatial variation in endpoint sensitivity provide a useful signal for allocating distillation updates?

We investigate this question by adapting Local Complexity (LC) (Liang et al., 2026; Humayun et al., 2024; 2025) to the teacher's time-indexed suffix map. At a given time t, this map sends a latent state to the endpoint obtained by integrating the frozen teacher to $t = 0$. LC measures how the map's dominant endpoint-sensitivity subspace varies across neighboring latent states at that fixed time. Evaluating LC along teacher trajectories produces a temporal profile of this spatial variation. Our measurements reveal substantial variation across the generation schedule, motivating the use of this teacher-specific profile to guide training emphasis.

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.

Based on this profile, we introduce **Complexity-Weighted Rectified Distillation (CWRD)**, a plug-and-play training wrapper for compatible distillation objectives. CWRD computes a teacher LC profile offline and uses it to reweight the host method's timestep distribution. The profile remains frozen during student training. This changes the relative frequency of training updates across time while preserving the host method's per-example loss, student architecture, and inference procedure. Additional trajectory diagnostics examine its relationship with cross-attention entropy and one-step suffix recovery.

Our contributions are summarized as follows:

- **Temporal Profiling of Endpoint Sensitivity:** We adapt Local Complexity to the teacher's time-indexed suffix map and characterize how spatial variation in endpoint sensitivity changes over the generation schedule. Additional trajectory diagnostics examine its relationship with cross-attention entropy and one-step suffix recovery.
- **Complexity-Weighted Distillation:** We propose CWRD, a training wrapper that uses a frozen teacher profile to reweight timestep sampling without adding student parameters or changing the inference procedure.
- **Efficacy Across Distillation Frameworks:** Across the evaluated configurations of π-Flow, LADD, TDM, and MFD, CWRD reduces data-referenced and teacher-referenced FID relative to the corresponding baseline configurations. Additional comparisons show lower data-referenced FID than the evaluated analytic weighting baseline, with metric-dependent trade-offs in prompt alignment and preference scores.

2 Related Work

The Geometry of Deep Generative Models. Deep generative models are fundamentally viewed as mappings that embed a low-dimensional latent manifold into a high-dimensional observation space. Foundational works across VAEs, Normalizing Flows, and GANs (Shao et al., 2018; Chen et al., 2019; Chadebec & Allassonniere, 2022; Caterini et al., 2021; Dai & Hang, 2021) utilize the Jacobian matrix as a linear approximation of local geometry, demonstrating that manifold curvature renders standard linear interpolation in the latent space suboptimal.

Recently, this geometric framework has been extended to analyze the probability flow ODEs of diffusion models (Park et al., 2023; Tang & Yang, 2024; Kamkari et al., 2024; Farghly et al., 2025). Most relevant to our approach, Humayun et al. (2024; 2025) and Liang et al. (2026) formalized the metrics of *Local Scaling* (LS) and *Local Complexity* (LC) to explicitly quantify the trade-off between generative capacity and manifold spatial variation. However, these prior applications primarily evaluate static generator expressivity or static Out-of-Distribution robustness. They do not analyze how geometric complexity evolves across the iterative, time-dependent integration steps of a generative solver. In this work, we bridge this gap by adapting LC into a temporal profiling tool. By mapping the rapid sensitivity transitions along the ODE trajectory, we transform LC from a passive analytical metric into an active geometric priority signal to weight few-step distillation.

Distillation for Flow-Matched Diffusion Models. To overcome the iterative bottleneck of flow-matched inference, early distillation methods mapped noise to data via straight secants or flow maps (Boffi et al., 2026; Sabour et al., 2026; Tong et al., 2026; Huang et al., 2026). However, forcing highly curved trajectories into rigid linear maps induces severe truncation errors. To circumvent this, methods like Mean Flow Distillation (MFD) (Zhao et al., 2026) and FACM (Peng et al., 2026) optimize time-averaged or localized velocity tracking to prevent shortcut instability.

Because underlying ODE paths are inherently non-homogeneous (Xu et al., 2024), the state-of-the-art has rapidly evolved toward non-linear, sub-path approximations. Most notably, π-Flow (Chen et al., 2026), alongside TMD (Nie et al., 2026), TDM (Luo et al., 2025), and SlimFlow (Zhu et al., 2024), execute dynamic, multi-step velocity imitation rather than static trajectory straightening. Crucially, while these advanced frameworks innovate heavily on *what* the student model predicts, they typically keep their native timestep sampling. CWRD is an orthogonal training wrapper that reweights this sampling using the teacher's spatial variation.

The Temporal Allocation Gap. While state-of-the-art methods innovate on student predictions (*e.g.*, flow maps, curved policies), they almost universally default to native temporal sampling

($t \sim \mathcal{U}[0,1]$). Even rare exceptions like SenseFlow (Ge et al., 2025) address scale gaps rather than the latent space's intrinsic topological curvature. Departing from this uniform-capacity assumption, we identify a highly localized geometric bottleneck in the trajectory. Consequently, we introduce CWRD as an orthogonal, geometry-aware change of measure. By reweighting the training distribution via Local Complexity, CWRD acts as a universal plug-and-play wrapper to seamlessly upgrade the capacity allocation of existing distillation frameworks.

Evaluation Metrics for Text-to-Image Distillation. Evaluating few-step generative models requires a comprehensive suite of metrics to capture both statistical fidelity and perceptual alignment. For distribution-level image quality and diversity, Fréchet Inception Distance (FID) Heusel et al. (2017) and Inception Score (IS) Salimans et al. (2016) remain the standard benchmarks, while Precision and Recall (Sajjadi et al., 2018; Simon et al., 2019; Kynkäänniemi et al., 2019; Liang et al., 2024) further disentangle sample accuracy from manifold coverage. Beyond pure image quality, cross-modal semantic alignment is assessed using CLIP Score Hessel et al. (2021) for global text-image similarity and VQAScore Lin et al. (2024) for fine-grained compositional prompt adherence. Finally, to gauge the visual appeal of the generations relative to human judgments, aesthetic alignment models such as HPSv2.1 Wu et al. (2023); Ma et al. (2025) are utilized, ensuring a holistic evaluation of the distilled student.

3 METHODOLOGY

We study how a frozen temporal geometry profile can improve few-step flow distillation. Our method, *Complexity-Weighted Rectified Distillation* (CWRD), is a plug-and-play training wrapper for compatible time-indexed distillation objectives. It increases the training emphasis assigned to timesteps with higher Local Complexity (LC). LC is used as a geometric priority signal, not as a claim that those timesteps have larger loss or are harder to optimize.

3.1 TEACHER AND BASE DISTILLATION OBJECTIVE

We use $t = 1$ for noise and $t = 0$ for data. Let $z = x_1 \sim \mathcal{N}(0, I)$ and let $c \sim p(c)$ be a condition. For a fixed classifier-free guidance scale g (Ho & Salimans, 2022), the frozen teacher velocity is

$$v_\phi^g(x, t, c) = v_\phi(x, t, \varnothing) + g\left[v_\phi(x, t, c) - v_\phi(x, t, \varnothing)\right]. \tag{1}$$

The teacher ODE and its reference numerical flow map are

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = v_\phi^g(x_t, t, c), \qquad x_u = \Phi_{s\to u}^\phi(x_s, c). \tag{2}$$

The checkpoint, guidance scale, solver, and time coordinate are fixed during profiling and distillation.

CWRD wraps a base distillation loss rather than defining a new inference architecture. Let $t_i \in \mathcal{T}_{\text{train}}$ denote the training time sampled from the base method's native discrete distribution, where $q_i = \Pr(T = t_i) > 0$ and $\sum_i q_i = 1$. Let

$$\ell_{\text{base}}(\Theta; t_i) = \mathbb{E}_{x_t, c, \omega}\left[\ell(\Theta; x_{t_i}, t_i, c, \omega)\right] \tag{3}$$

denote its conditional loss, where ω collects auxiliary variables and target-construction randomness. Native loss terms and weights remain part of ℓ_{base}. CWRD changes only the distribution or weight of t_i. It leaves the student architecture, target construction, inference schedule, and inference computation unchanged.

3.2 TEMPORAL LOCAL-COMPLEXITY PROFILE

For a teacher state at time $t > 0$, we define the remaining endpoint map $G_{\phi,t,c}(x) := \Phi_{t\to 0}^\phi(x, c)$, where $\Phi_{t\to 0}^\phi$ is evaluated using the frozen guided teacher and a fixed reference solver. The Local Complexity (LC) construction of Liang et al. (2026) measures how a generator's dominant input-sensitivity direction changes across neighboring inputs. We adapt this idea to the time-indexed suffix map for $G_{\phi,t,c}(x)$, use its terminal latent as the output representation, and compare a leading rank-k sensitivity subspace rather than a single direction.

We profile a fixed nonterminal grid $\mathcal{T}_{\text{prof}} = \{t_i\}_{i=1}^{M} \subset (0, 1]$. At the teacher state $x_t^\phi = \Phi_{1\to t}^\phi(z, c)$, let $W = [w_1, \ldots, w_r] \in \mathbb{R}^{d\times r}$ be a random orthonormal probe basis. Using an RMS-scaled finite-difference radius η_t, we estimate the restricted Jacobian of the suffix map along the directions in W:

$$[J_{W,t,c}(x)]_{:,j} = \frac{G_{\phi,t,c}(x + \eta_t w_j) - G_{\phi,t,c}(x)}{\eta_t}, \qquad j = 1, \ldots, \hat{r}. \tag{4}$$

Let $\Pi_k(x; t, c, W)$ denote the projector onto the leading k eigenvectors of the restricted Gram matrix $J_{W,t,c}(x)^\top J_{W,t,c}(x)$.

For $\xi \sim \text{Unif}(\mathbb{S}^{r-1})$ uniform samping over the hypersphere, define the neighboring state $x'_t = x_t^\phi + \rho_t W\xi$, where ρ_t is also scaled by the latent RMS. Using the same W, t, and c at both states, we compute the local subspace variation:

$$\ell_k^{\text{LC}}(t; z, c, W, \xi) = \frac{\left\|\Pi_k(x_t^\phi; t, c, W) - \Pi_k(x'_t; t, c, W)\right\|_F}{\sqrt{2k}\,\rho_t}, \qquad 1 \le k \le \hat{r}/2. \tag{5}$$

To ensure stability, let $R_k(t; z, c, W, \xi) \in \{0, 1\}$ indicate that a predefined normalized boundary eigengap is satisfied at both x_t^ϕ and x'_t. We define the retained coverage $\kappa_k(t)$ and obtain the conditional temporal profile $C(t_i)$:

$$\kappa_k(t_i) = \mathbb{E}_{z,c,W,\xi}[R_k(t_i)], \qquad C(t_i) = \frac{\mathbb{E}_{z,c,W,\xi}[R_k(t_i)\ell_k^{\text{LC}}(t_i)]}{\kappa_k(t_i)} = \mathbb{E}[\ell_k^{\text{LC}}(t_i) \mid R_k(t_i) = 1]. \tag{6}$$

We require $\kappa_k(t_i)$ to exceed a predefined threshold at every training anchor; otherwise, the profiling run is invalid. The profile is computed once on a disjoint split and frozen. The endpoint $t = 0$ is rigorously excluded because the suffix map becomes the identity and its leading sensitivity subspace is not identifiable.

We define the profile maximum: $t_{\text{peak}} = \arg\max_{t\in\mathcal{T}_{\text{prof}}} C(t)$. A high $C(t)$ indicates that the dominant endpoint-sensitivity subspace changes rapidly across neighboring latent states. It does not directly measure the instantaneous velocity-Jacobian norm, trajectory curvature, truncation error, topology, or student learning difficulty.

3.3 LC-Based Temporal Training Weights

We convert the frozen temporal LC profile into a relative training weight for every anchor. Let $q_i = \Pr_q(T = t_i) > 0$, with $\sum_{i=1}^{M} q_i = 1$, denote the native sampling mass of the base distillation method. We write $C_i = C(t_i)$ for the raw LC profile value. Because the magnitude of C_i depends on the teacher and latent coordinate scale, we first normalize the profile within the fixed training grid:

$$\widetilde{C}_i = \frac{C_i - \min_j C_j}{\max_j C_j - \min_j C_j + \epsilon_C}. \tag{7}$$

To prevent min-max normalization from amplifying Monte Carlo noise, we include a predefined flat-profile fallback: if the estimated profile is not stable across independent profiling replicates, we bypass the weighting procedure entirely and set $w_i = 1$ for all i.

We define the shape to explore post-peak capacity allocation, recalling that generation proceeds from $t = 1$ toward $t = 0$, let t_+ denote the largest profiled anchor strictly below the peak t^*. The shapes are defined as, starting a plateau one measured step after the peak and holds at that step's value. This replaces the measured post-peak profile with a constant continuation:

$$\hat{C}_i = \begin{cases} \widetilde{C}_i & t_i \ge t_+ \\ \widetilde{C}_+ & t_i < t_+ \end{cases} \tag{8}$$

We apply flooring and normalization to the plateau profile. Define $\bar{C} = \sum_{j=1}^{M} q_j \hat{C}_j$ under the native sampler q. If $\bar{C} = 0$, we set $w_i = 1$ for all i, recovering native sampling. Otherwise, for a floor parameter $\gamma \in (0, 1]$, define

$$u_i = \max\left(\gamma, \frac{\hat{C}_i}{\bar{C}}\right), \qquad Z = \sum_{j=1}^{M} q_j u_j, \qquad w_i = \frac{u_i}{Z}. \tag{9}$$



Figure 1: Interpolation for sample pairs with LC peaks $t^* \approx 0.6$ and $t^* \approx 0.8$. Test 1 interpolates the source noise at $t = 1$; Test 2 interpolates the teacher states at t^*. Both use the 12-step sampler.

Table 1: Quantitative comparison of interpolation regularity. Peak-LC initialization yields drastically smoother perceptual paths (lower PPL) and higher structural fidelity (PSNR/SSIM) compared to uniform noise interpolation. Both arms use the same 12-step teacher sampler and differ only in where interpolation is applied, *Noise* interpolates the source noise at $t = 1$; *Peak-LC* interpolates the teacher states at the LC-peak time t^* and completes the remaining steps.

LC peak(t)		LPIPS/step ↓	PSNR/step ↑	SSIM/step ↑	CLIP/step↓	PPL ↓
0.84	Noise	0.347 ± 0.019	15.51 ± 0.466	0.644 ± 0.025	0.099 ± 0.010	1540 ± 111
	Peak-LC	$\mathbf{0.174 \pm 0.012}$	$\mathbf{22.208 \pm 0.387}$	$\mathbf{0.795 \pm 0.016}$	$\mathbf{0.080 \pm 0.015}$	$\mathbf{525 \pm 46}$
0.78	Noise	0.317 ± 0.018	16.061 ± 0.533	0.654 ± 0.023	0.0851 ± 0.0113	1421 ± 152
	Peak-LC	$\mathbf{0.155 \pm 0.009}$	$\mathbf{23.434 \pm 0.443}$	$\mathbf{0.801 \pm 0.015}$	$\mathbf{0.058 \pm 0.006}$	$\mathbf{363 \pm 29}$
0.71	Noise	0.296 ± 0.016	16.623 ± 0.435	0.663 ± 0.020	0.080 ± 0.012	1286 ± 123
	Peak-LC	$\mathbf{0.143 \pm 0.009}$	$\mathbf{24.311 \pm 0.464}$	$\mathbf{0.815 \pm 0.017}$	$\mathbf{0.056 \pm 0.009}$	$\mathbf{386 \pm 46}$
0.63	Noise	0.326 ± 0.018	16.457 ± 0.481	0.649 ± 0.018	0.071 ± 0.008	1298 ± 101
	Peak-LC	$\mathbf{0.123 \pm 0.008}$	$\mathbf{25.602 \pm 0.557}$	$\mathbf{0.828 \pm 0.019}$	$\mathbf{0.049 \pm 0.006}$	$\mathbf{265 \pm 34}$
0.53	Noise	0.281 ± 0.018	17.087 ± 0.587	0.709 ± 0.022	0.075 ± 0.008	1226 ± 143
	Peak-LC	$\mathbf{0.078 \pm 0.007}$	$\mathbf{27.765 \pm 0.563}$	$\mathbf{0.902 \pm 0.013}$	$\mathbf{0.040 \pm 0.003}$	$\mathbf{156 \pm 16}$

By construction, the resulting importance weights satisfy $w_i > 0$ and preserve the base measure's total mass, $\sum_{i=1}^{M} q_i w_i = 1$, while γ enforces a non-vanishing lower bound $w_i \geq \frac{\gamma}{Z}$.

This construction guarantees a positive lower bound of $w_i \geq \gamma/Z$ and ensures $\mathbb{E}_{T\sim q}[w(T)] = \sum_i q_i w_i = 1$, preserving the expected scalar weight and normalizing the sampling distribution. The CWRD objective applies this scalar multiplier to the base loss:

$$\mathcal{L}_{\text{CWRD}}(\Theta) = \sum_{i=1}^{M} q_i w_i \, \ell_{\text{base}}(\Theta; t_i). \tag{10}$$

Our motivation is that high-LC anchors expose the shared student to more heterogeneous endpoint-sensitivity orientations. Increasing their relative training frequency may therefore improve the learned approximation. This is an empirical allocation hypothesis, not a claim that high-LC anchors necessarily have larger losses or are harder to optimize.

Equivalent sampling view. We apply w as a multiplier on the sampled loss, Eq. 10, leaving the host's timestep sampling unchanged. In expectation, this is equivalent to sampling from the reweighted measure $p_w(t_i) = q_i w_i$:

$$\mathbb{E}_{T\sim p_w}[\ell_{\text{base}}(\Theta; T)] = \mathbb{E}_{T\sim q}[w(T)\ell_{\text{base}}(\Theta; T)]. \tag{11}$$

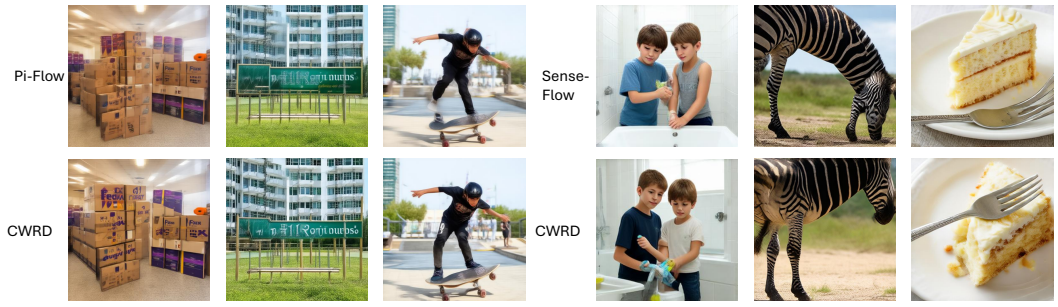

Figure 2: Qualitative distillation comparisons across baseline methods and our proposed *CWRD* integration. The inclusion of *CWRD* mitigates structural artifacts and improves fine-grained detail synthesis. For more examples please check Appendix. E.

4 EXPERIMENTS

In this section, we empirically express the geometric properties of the Local Complexity (LC) profile and evaluate the efficacy of Complexity-Weighted Rectified Distillation (CWRD). We first establish the temporal compressibility of the post-peak ODE trajectory, showing that it constitutes a highly regular, one-step-recoverable suffix. We then demonstrate that weighting distillation by this LC profile yields an advantage with few-step generation fidelity compared to native baselines. The experiment setup is provided in Appendix A.

4.1 TRAJECTORY DIAGNOSTICS AROUND THE LC PEAK

In this part, we empirically examine interpolation regularity and one-step suffix recovery around the Local Complexity (LC) peak. Specifically, we analyze the latent dynamics around the LC peak through two controlled diagnostic tests: representation interpolation (comparing initial noise space vs. LC-peak states) and one-step suffix recovery (generating directly from the LC peak). These experiments aim to validate that the ODE trajectory transitions into a temporally compressible, highly regular regime immediately following the LC peak. Note that Sec.4.1.1 for interpolation uses a 12-step teacher schedule, and Sec.4.2 for suffix-recovery comparison uses 20 steps.

4.1.1 INTERPOLATION AT LC-PEAK STATES

We test whether states at the LC peak support smoother interpolation than the source noise. Given two samples A and B generated from noises $x_{1,A}$ and $x_{1,B}$, we compare two strategies:

Test 1 (Noise interpolation): We spherically interpolate (Slerp) between $x_{1,A}$ and $x_{1,B}$ and generate each intermediate noise with the 12-step teacher sampler.

Test 2 (LC-peak interpolation): We run the same 12-step sampler on both noises up to the LC-peak time t^*, obtaining $z_A^* = \Phi_{1\to t^*}^{\phi}(x_{1,A}, c)$ and $z_B^* = \Phi_{1\to t^*}^{\phi}(x_{1,B}, c)$. We linearly interpolate $z_\lambda^* = (1-\lambda)z_A^* + \lambda z_B^*$ for $\lambda \in [0,1]$ and complete sampling from t^* to $t = 0$ with the remaining steps of the same schedule. Both tests use the same sampler and step count; they differ only in the time at which interpolation is applied.

As detailed in Table 1, interpolating at the LC peak yields smoother paths than interpolating the source noise: per-step LPIPS falls from 0.28–0.35 to 0.08–0.17 and PPL from above 1200 to 156–525. Peak-LC initialization yields drastically smoother perceptual paths (lower PPL) and higher structural fidelity (PSNR/SSIM) compared to uniform noise interpolation. Smoothness generally increases as t^* decreases, as expected, since later states already determine more of the image, measuring path smoothness. We complement these quantitative metrics with visual interpolation sequences in Figure 1, anchored the LC peak at $t \simeq 0.8$ and $t \simeq 0.6$, which visually confirm the absence of artifacting when interpolating safely within the compressible suffix. $t^* \approx 0.8$ and $t^* \approx 0.6$.

Table 2: Distillation results across teachers and host methods. Following the protocol of Chen et al. (2026), we train on the COCO-10k prompt set and evaluate on a separate set of COCO prompts not used during training. Lower is better for FID and FID versus the teacher; higher is better for all other metrics. Bold values indicate the better result within each matched `original`/`CWRD` comparison for a fixed host and Neural Function Evaluation (NFE) budget. Italicized rows denote teacher or released-checkpoint references and are not included in the matched comparisons. A dash indicates that the metric was not evaluated. "ft" denotes fine-tuning. $\text{FID}_t \downarrow$ denotes FID with teacher.

	NFE	FID↓	$\text{FID}_t \downarrow$	IS↑	Precision↑	Recall↑	CLIP ↑	HPSv2.1 ↑	VQAScore ↑
Teacher: SD3.5	–	20.368	–	42.501	0.701	0.369	26.400	0.285	0.899
π-Flow	4	21.580	11.615	37.821	0.635	0.319	26.153	0.262	0.859
π-Flow `CWRD`	4	**18.724**	**7.976**	**39.515**	**0.696**	**0.348**	**26.227**	**0.270**	**0.875**
π-Flow	8	20.122	4.209	38.870	0.710	0.343	25.735	0.284	0.868
π-Flow `CWRD`	8	**19.305**	**3.502**	**39.505**	**0.724**	**0.349**	**25.792**	**0.288**	**0.880**
LADD	4	22.813	4.157	38.409	0.742	0.348	**26.253**	0.270	0.604
LADD `CWRD`	4	**18.138**	**4.077**	**40.654**	**0.756**	**0.424**	26.235	**0.272**	**0.605**
TDM ft	4	24.04	9.16	37.10	**0.581**	0.235	26.39	0.258	0.606
TDM ft `CWRD`	4	**21.95**	**9.09**	**37.11**	0.580	**0.258**	**26.64**	**0.268**	0.606
Teacher: SANA	–	24.385	–	38.105	0.725	0.420	26.128	0.297	0.911
MFD	4	33.353	13.398	**33.317**	0.609	0.193	25.389	**0.289**	0.831
MFD `CWRD`	4	**30.584**	**13.286**	33.164	**0.615**	**0.209**	**25.795**	0.281	**0.847**
MFD	8	32.044	8.245	**32.944**	0.623	0.208	25.251	**0.296**	0.817
MFD `CWRD`	8	**28.973**	**7.977**	32.715	**0.630**	**0.216**	**25.617**	0.289	**0.855**
Teacher: Flux.2	–	14.300	–	38.337	0.695	0.472	26.892	0.272	0.933
π-Flow	4	12.650	2.835	38.051	0.695	0.488	26.675	0.255	0.899
π-Flow `CWRD`	4	**12.600**	**2.734**	**38.291**	**0.698**	**0.493**	**26.768**	**0.258**	**0.899**

4.1.2 One-Step Suffix Recovery

To test whether the empirically observed compressibility boundary (t_{rec}) securely aligns with our calculated t_{peak}, we execute a controlled step-budget ablation. We compare a standard uniform step allocation against an *LC-Suffix One-Step* strategy. In the LC-Suffix strategy, we integrate uniformly strictly up to the designated LC peak, and then substitute the entire remaining trajectory with a single Euler jump to $t = 0$. The competing baseline (*Peak-Step Matched Sampling*) utilizes the exact same total Neural Function Evaluation (NFE) budget but distributes the steps uniformly across the entire $t \in [1, 0]$ domain.

As detailed in Appendix B.1 Table 5, the LC-Suffix methodology drastically outperforms matched-budget uniform sampling across all tested peak locations. For instance, at Peak 8, the LC-Suffix achieves an SSIM of 0.8723, whereas the Peak-Step baseline collapses to 0.5821. To further isolate the LC peak as the exact boundary of this compressibility, in Appendix B.2, Table 6 presents a launch-offset evaluation. When the single-step Euler jump is launched prematurely (offset $o > 0$, prior to the LC peak), structural fidelity degrades precipitously, with LPIPS rising from 0.185 at the peak to 0.604 at an offset of 4.

Together, these results show that, in the evaluated settings, following the teacher trajectory to the LC peak before a single suffix update improves teacher-referenced reconstruction relative to the tested earlier-launch and matched-budget sampling baselines. Once the generator traverses the rapid spatial variation up to the LC peak, the remainder of the generative schedule transitions into a temporally compressible suffix that can be bypassed in a single step generation without severe degradation. Figure 3 visually corroborates this in Appendix B.3.

4.2 Complexity Weighted Rectified Distillation(CWRD) Efficacy

We empirically evaluate the efficacy of CWRD as a plug-and-play training wrapper. To demonstrate its broad compatibility, we apply the offline LC profile to four distinct state-of-the-art distillation baselines encompassing both regression and adversarial objectives: π-Flow, Latent Adversarial Diffusion Distillation (LADD), Trajectory Distillation Models (TDM), and Mean Flow Distillation (MFD). The evaluations span three distinct high-resolution teacher architectures (Stable Diffusion 3.5(Esser et al., 2024), SANA (Xie et al., 2024) and Flux.2 (Labs, 2025)).

Table 3: Comparison of CWRD and σ^2 across teachers and sampling budgets. Bold values indicate the better result within each method and NFE pair. "ft" denotes fine-tuning.

Method	NFE	FID ↓	IS ↑	Precision ↑	Recall ↑	CLIP ↑	HPSv2.1 ↑	VQAScore ↑
SD3.5								
π-Flow CWRD	4	**18.724**	**39.515**	**0.696**	**0.348**	26.227	**0.270**	**0.875**
π-Flow σ^2	4	23.468	36.894	0.621	0.287	**26.304**	0.260	0.865
π-Flow CWRD	8	**19.305**	**39.505**	**0.724**	**0.349**	25.792	**0.288**	**0.880**
π-Flow σ^2	8	21.888	37.928	0.693	0.317	**25.885**	0.283	0.861
SenseFlow ft	–	24.690	41.350	0.789	0.426	25.930	0.280	0.867
SenseFlow ft CWRD	–	**24.540**	**42.130**	**0.797**	**0.437**	**26.150**	**0.281**	**0.869**
SANA								
MFD CWRD	4	**30.584**	**33.164**	**0.615**	**0.209**	**25.795**	0.281	0.847
MFD σ^2	4	32.680	32.402	0.605	0.194	25.454	**0.290**	**0.851**
MFD CWRD	8	**28.973**	**32.715**	**0.630**	**0.216**	**25.617**	0.289	**0.855**
MFD σ^2	8	31.253	32.069	0.612	0.203	25.477	**0.299**	0.845

To rigorously isolate the effect of our geometric priority signal, every paired comparison utilizes identically matched student architectures, optimizer steps, batch sizes, and target-generation budgets. For the baseline arms, the training timesteps are drawn from the native distribution $q(t)$. For the CWRD arms, we apply the exact same base loss but alter the training measure to $p_w(t)$ using the pre-computed LC profile, reallocating expected training emphasis toward the rapid-sensitivity transitions (the LC peak).

As detailed in Table 2, CWRD consistently improves distributional fidelity across all evaluated architectures and solver budgets. Most notably, integrating CWRD reduces both data-referenced FID and teacher-referenced distillation FID (FID_t) across all seven matched comparisons. For instance, when applying MFD to SANA, data-referenced FID decreases substantially from 33.353 to 30.584 at 4 NFE, and from 32.044 to 28.973 at 8 NFE. These distributional gains are broadly accompanied by enhancements in structural coherence and semantic alignment: CLIP scores improve in six of the seven configurations (with a marginal decrease for LADD), while VQAScore improves in five and remains stable for TDM and π-Flow on Flux.2. We do observe minor metric-dependent tradeoffs in human preference scores, as HPSv2.1 improves for the SD3.5 and Flux.2 configurations but decreases slightly for MFD on SANA. Overall, these quantitative results support our hypothesis that prioritizing high-LC timesteps improves allocation over native sampling of gradient capacity compared to native sampling. Finally, as illustrated in Figure 2, this optimized temporal allocation directly translates to enhanced qualitative generation. Across various distillation methodologies, CWRD yields notable improvements in fine-grained details and substantially mitigates structural artifacts, such as supernumerary limbs or anomalous leg placements.

4.3 Comparison with the SenseFlow strategy

Although SenseFlow (Ge et al., 2025) applies an analytic temporal weighting heuristic, $w(t) = (1 - \sigma_t)^2$, it ignores the underlying spatial geometry of the vector field. We hypothesize that our data-driven LC profile, which tracks intrinsic representational variation, provides a more effective allocation signal. To evaluate this, we benchmark CWRD against the σ^2 weighting within π-Flow and MFD. Crucially, we also replace the native σ^2 in SenseFlow with CWRD to compare directly against its original formulation.

As shown in Table 3, CWRD outperforms the analytic σ^2 heuristic across virtually all configurations. In the π-Flow (4NFE) setting, CWRD improves FID from 23.468 to 18.724. Most compellingly, wrapping the native SenseFlow algorithm with CWRD yields strict improvements across all evaluated metrics compared to its original formulation. These results suggest LC-based allocation is more effective than the σ^2 heuristic in our experiments.

4.4 Local Complexity Alignment

The efficacy of our proposed Complexity-Weighted Rectified Distillation (CWRD) relies heavily on precise synchronization between the concentrated sampling weights and the underlying trajectory dynamics. In this section, we validate the necessity and physical meaning of this geometric

Table 4: Quantitative comparison of standard CWRD and a misaligned control (Shift $\sigma^\star \pm \mathcal{R}$). In the shifted condition, the LC peak hinge is displaced by a random signed offset, redrawn at every optimizer step, to disrupt temporal alignment. Fixed offsets are reported in Appendix C.1.

Method	FID↓	IS↑	Precision↑	Recall↑	CLIP ↑	HPSv2.1 ↑	VQAScore ↑
CWRD	**18.724**	**39.515**	**0.696**	**0.348**	26.227	**0.270**	**0.875**
shift $\sigma^\star \pm \mathcal{R}$	20.317	38.945	0.688	0.328	26.251	0.269	0.702

alignment from two complementary perspectives. We examine temporal allocation through a shift ablation and an analysis of the relationship between LC and cross-attention concentration.

4.4.1 Local Complexity Shift Ablation

To assess the effect of temporal placement, we compare CWRD with a matched Shifted Allocation control. We perturb the plateau hinge along the noise-level axis σ while keeping the measured LC profile and the native π-Flow sampling procedure fixed. At each optimizer step, we sample a signed offset r with $|r| \in [0.15, 0.50]$. Negative offsets move the hinge toward $\sigma = 0$, the clean-latent endpoint x_0, whereas positive offsets move it toward $\sigma = 1$, the pure-noise endpoint. The resulting profile weights the training loss as in standard CWRD. Student initialization, training prompts, optimizer settings, batch size, training budget, and the underlying LC profile are matched across conditions. We compare CWRD with randomized and fixed hinge perturbations, as detailed in Appendix C.

As reported in Table 4, CWRD achieves lower FID than the shifted control (18.724 versus 20.317) and higher VQAScore (0.875 versus 0.702), while the control achieves a marginally higher CLIP score (26.251 versus 26.227). These results support the effectiveness of the LC-aligned allocation relative to the tested randomized alternative. Because the shift changes across steps, the control is trained on a mixture of allocations, so this comparison does not isolate temporal alignment at a fixed effective allocation concentration. Results for more fixed hinge offsets are reported in Appendix C.1.

4.4.2 Mechanism Probing: Cross-Attention Crystallization

To ground our macroscopic analysis, we examine the relationship between Local Complexity (LC) and cross-attention concentration in the SD3.5 teacher. By computing the Shannon entropy of spatial attention maps for subject tokens, where lower entropy denotes highly concentrated attention, we observe that the maximum of the averaged LC profile and the minimum of the averaged attention-entropy profile occur at the same sampled time, $t = 0.448$. As detailed in Appendix D, per-trajectory analyses consistently reveal strong negative LC–entropy correlations prior to the LC peak, followed by weaker, heterogeneous associations afterward.

These findings associate spatial variation in endpoint sensitivity with changes in cross-attention concentration, providing complementary empirical motivation for LC-guided temporal allocation. CWRD uses the frozen LC profile to reweight the host method's native timestep distribution, adjusting the relative frequency of training updates across the generation schedule.

5 Conclusion

In this work, we investigate the geometric properties of trajectory distillation using the Local Complexity (LC) metric, observing a distinct transition from an early phase of high spatial variation to a temporally compressible suffix. To leverage this structural property, we propose Complexity-Weighted Rectified Distillation (CWRD), an orthogonal training wrapper that reweights native timestep sampling to prioritize regions of rapid geometric transition. Evaluations across diverse architectures indicate that CWRD consistently improves generative fidelity and semantic alignment, performing favorably against both native sampling and recent analytic schedule heuristics. Overall, our findings suggest that few-step distillation benefits from allocating gradient capacity according to intrinsic temporal geometry. Limitations and future directions are discussed in Appendix F.

AI USE STATEMENT

We used generative AI tools primarily to revise an initial manuscript written by the authors, improving language, clarity, organization, and LaTeX presentation. The authors reviewed the AI-assisted revisions for consistency with the experimental procedures and reported evidence and made the final decisions on their inclusion. We take full responsibility for the final manuscript, including all text, scientific claims, analyses, and conclusions developed with AI assistance.

ETHICS STATEMENT

This work studies temporal training allocation for distilling pretrained text-to-image models. Accelerating image generation can improve accessibility while also lowering the cost of producing misleading or harmful content. Distilled models may retain biases and harmful behaviors present in their teachers and training data. Our evaluation focuses on generation quality and teacher-referenced fidelity; these metrics do not establish fairness, privacy, or safety guarantees. Responsible use should respect the licenses and usage conditions of the underlying models and datasets.

REPRODUCIBILITY STATEMENT

The main paper describes the LC profiling procedure, temporal weight construction, and timestep resampling used by CWRD. Appendix A describes the experimental setup, while the remaining appendices provide details of the trajectory diagnostics, attention analysis, and shifted-allocation control. We include our implementation in the supplementary material to support examination and reproduction of the method during peer review. We will publicly release the code upon acceptance.

REFERENCES

Nicholas Boffi, Michael Albergo, and Eric Vanden-Eijnden. How to build a consistency model: Learning flow maps via self-distillation. *Advances in Neural Information Processing Systems*, 38: 33346–33382, 2026.

Anthony L Caterini, Gabriel Loaiza-Ganem, Geoff Pleiss, and John P Cunningham. Rectangular flows for manifold learning. *Advances in neural information processing systems*, 34:30228–30241, 2021.

Clément Chadebec and Stephanie Allassonniere. A geometric perspective on variational autoencoders. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022.

Hansheng Chen, Kai Zhang, Hao Tan, Leonidas Guibas, Gordon Wetzstein, and Sai Bi. pi-flow: Policy-based few-step generation via imitation distillation. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=1vAte8dsap.

Nutan Chen, Francesco Ferroni, Alexej Klushyn, Alexandros Paraschos, Justin Bayer, and Patrick van der Smagt. Fast approximate geodesics for deep generative models. In *International Conference on Artificial Neural Networks*, pp. 554–566. Springer, 2019.

Mengyu Dai and Haibin Hang. Manifold matching via deep metric learning for generative modeling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6587–6597, October 2021.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first international conference on machine learning*, 2024.

Tyler Farghly, Peter Potaptchik, Samuel Howard, George Deligiannidis, and Jakiw Pidstrigach. Diffusion models and the manifold hypothesis: Log-domain smoothing is geometry adaptive. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=4JihzQXNJn.

Xingtong Ge, Xin Zhang, Tongda Xu, Yi Zhang, Xinjie Zhang, Yan Wang, and Jun Zhang. Senseflow: Scaling distribution matching for flow-based text-to-image distillation. *arXiv preprint arXiv:2506.00523*, 2025.

Jack Hessel, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, and Yejin Choi. Clipscore: A reference-free evaluation metric for image captioning. In *Proceedings of the 2021 conference on empirical methods in natural language processing*, pp. 7514–7528, 2021.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

Zemin Huang, Weijian Luo, Zhengyang Geng, and Guojun Qi. One-step diffusion and flow distillation through implicit generator matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026.

Ahmed Imtiaz Humayun, Ibtihel Amara, Candice Schumann, Golnoosh Farnadi, Negar Rostamzadeh, and Mohammad Havaei. On the local geometry of deep generative manifolds. In *ICML 2024 Workshop on Geometry-grounded Representation Learning and Generative Modeling*, 2024.

Ahmed Imtiaz Humayun, Ibtihel Amara, Cristina Nader Vasconcelos, Deepak Ramachandran, Candice Schumann, Junfeng He, Katherine A Heller, Golnoosh Farnadi, Negar Rostamzadeh, and Mohammad Havaei. What secrets do your manifolds hold? understanding the local geometry of generative models. In *The Thirteenth International Conference on Learning Representations*, 2025.

Hamidreza Kamkari, Brendan L Ross, Rasa Hosseinzadeh, Jesse C Cresswell, and Gabriel Loaiza-Ganem. A geometric view of data complexity: Efficient local intrinsic dimension estimation with diffusion models. *Advances in Neural Information Processing Systems*, 37:38307–38354, 2024.

Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved precision and recall metric for assessing generative models. *Advances in Neural Information Processing Systems*, 32, 2019.

Black Forest Labs. Flux. 2: Frontier visual intelligence. *https://bfl. ai/blog/flux-2*, 2025.

Black Forest Labs, Stephen Batifol, Andreas Blattmann, Frederic Boesel, Saksham Consul, Cyril Diagne, Tim Dockhorn, Jack English, Zion English, Patrick Esser, Sumith Kulal, Kyle Lacey, Yam Levi, Cheng Li, Dominik Lorenz, Jonas Müller, Dustin Podell, Robin Rombach, Harry Saini, Axel Sauer, and Luke Smith. Flux.1 kontext: Flow matching for in-context image generation and editing in latent space, 2025. URL https://arxiv.org/abs/2506.15742.

Yuanbang Liang, Jing Wu, Yu-Kun Lai, and Yipeng Qin. Efficient precision and recall metrics for assessing generative models using hubness-aware sampling. In *Forty-first International Conference on Machine Learning*, 2024.

Yuanbang Liang, Zhengwen Chen, and Yu-Kun Lai. Geometric decoupling: Diagnosing the structural instability of latent. In *Forty-third International Conference on Machine Learning*, 2026.

Zhiqiu Lin, Deepak Pathak, Baiqi Li, Jiayao Li, Xide Xia, Graham Neubig, Pengchuan Zhang, and Deva Ramanan. Evaluating text-to-visual generation with image-to-text generation. In *European Conference on Computer Vision*, pp. 366–384. Springer, 2024.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.

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 *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 17719–17728. IEEE, 2025.

Yuhang Ma, Xiaoshi Wu, Keqiang Sun, and Hongsheng Li. Hpsv3: Towards wide-spectrum human preference score. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 15086–15095. IEEE, 2025.

Weili Nie, Julius Berner, Nanye Ma, Chao Liu, Saining Xie, and Arash Vahdat. Transition matching distillation for fast video generation. *arXiv preprint arXiv:2601.09881*, 2026.

Yong-Hyun Park, Mingi Kwon, Jaewoong Choi, Junghyo Jo, and Youngjung Uh. Understanding the latent space of diffusion models through the lens of riemannian geometry. *Advances in Neural Information Processing Systems*, 36:24129–24142, 2023.

Yansong Peng, Kai Zhu, Yu Liu, Pingyu Wu, Hebei Li, Xiaoyan Sun, and Feng Wu. Facm: Flow-anchored consistency models. In *International Conference on Learning Representations*, volume 2026, pp. 7520–7543, 2026.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 10674–10685. ieee, 2022.

Amirmojtaba Sabour, Sanja Fidler, and Karsten Kreis. Align your flow: Scaling continuous-time flow map distillation. *Advances in Neural Information Processing Systems*, 38:146459–146512, 2026.

Mehdi SM Sajjadi, Olivier Bachem, Mario Lucic, Olivier Bousquet, and Sylvain Gelly. Assessing generative models via precision and recall. *Advances in neural information processing systems*, 31, 2018.

Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *Advances in neural information processing systems*, 29, 2016.

Axel Sauer, Frederic Boesel, Tim Dockhorn, Andreas Blattmann, Patrick Esser, and Robin Rombach. Fast high-resolution image synthesis with latent adversarial diffusion distillation. In *SIGGRAPH Asia 2024 Conference Papers*, pp. 1–11, 2024a.

Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial diffusion distillation. In *European Conference on Computer Vision*, pp. 87–103. Springer, 2024b.

Hang Shao, Abhishek Kumar, and P Thomas Fletcher. The riemannian geometry of deep generative models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 315–323, 2018.

Loïc Simon, Ryan Webster, and Julien Rabin. Revisiting precision and recall definition for generative model evaluation. *arXiv preprint arXiv:1905.05441*, 2019.

Rong Tang and Yun Yang. Adaptivity of diffusion models to manifold structures. In *International Conference on Artificial Intelligence and Statistics*, pp. 1648–1656. PMLR, 2024.

Shangyuan Tong, Nanye Ma, Saining Xie, and Tommi Jaakkola. Flow map distillation without data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 33973–33984, 2026.

Xiaoshi Wu, Yiming Hao, Keqiang Sun, Yixiong Chen, Feng Zhu, Rui Zhao, and Hongsheng Li. Human preference score v2: A solid benchmark for evaluating human preferences of text-to-image synthesis. *arXiv preprint arXiv:2306.09341*, 2023.

Enze Xie, Junsong Chen, Junyu Chen, Han Cai, Haotian Tang, Yujun Lin, Zhekai Zhang, Muyang Li, Ligeng Zhu, Yao Lu, et al. Sana: Efficient high-resolution image synthesis with linear diffusion transformers. *arXiv preprint arXiv:2410.10629*, 2024.

Chen Xu, Tianhui Song, Weixin Feng, Xubin Li, Tiezheng Ge, Bo Zheng, and Limin Wang. Accelerating image generation with sub-path linear approximation model. In *European Conference on Computer Vision*, pp. 323–339. Springer, 2024.

An Zhao, Shengyuan Zhang, Zhongjian Sun, Yixiang Zhou, Zejian Li, Ling Yang, Tianrun Chen, and Lingyun Sun. Mean flow distillation: Robust and stable distillation for flow matching models. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=kOMsEkgMuz.

Yuanzhi Zhu, Xingchao Liu, and Qiang Liu. Slimflow: Training smaller one-step diffusion models with rectified flow. In *European Conference on Computer Vision*, pp. 342–359. Springer, 2024.

A Experimental Setup

All distillation experiments are executed on a single NVIDIA RTX 5090 GPU. Following the protocol of Chen et al. (2026), we utilize the COCO-10k prompt dataset[1] for training and evaluate on a separate set of COCO prompts not used during training. To rigorously isolate the effect of our geometric priority signal, every paired comparison utilizes identically matched student architectures, optimizer steps, batch sizes, and target-generation budgets. Because CWRD is an orthogonal, plug-and-play wrapper, we compare directly against the native baselines, where training timesteps are drawn from the distribution $q(t)$. For the CWRD arms, we apply the exact same base loss but alter the training measure to $p_w(t)$ using the pre-computed LC profile, reallocating expected training emphasis toward the rapid-sensitivity transitions (the LC peak).

For each host method, we follow its official implementation and default training configuration, retaining the optimizer, learning rate, initialization, and parameter-update settings. CWRD modifies only the per-timestep loss weight. The baseline retains each host method's native timestep distribution q, corresponding to constant additional weights $w_i = 1$. CWRD introduces an LC-derived change of measure, $p_w(t_i) = q_i w_i$, while preserving the native per-example loss.

B More experiment for LC prediction distillation difficult

B.1 LC-Suffix methodology details

As detailed in Table 5, the LC-Suffix methodology drastically outperforms matched-budget native sampling across all tested peak locations. For instance, at Peak 8, the LC-Suffix achieves an SSIM of 0.8723, whereas the uniform Peak-Step baseline collapses to 0.5821.

B.2 Sensitivity to the launch step.

For each run, let k denote the step at which Local Complexity (LC) reaches its peak. We launch a single Euler update to $t = 0$ from step $k - o$, where $o \in \{0, 1, 2, 3, 4\}$, and evaluate the resulting image against the corresponding run's 12-step reference output.

The results at Table 6 show a monotonic deterioration in reconstruction quality as the launch step moves earlier. Across the tested offsets, each additional step earlier decreases PSNR by approximately $1.9\,\mathrm{dB}$ and increases LPIPS by approximately 0.10 on average. These observations support further investigation of the LC peak as a launch location. However, four runs provide limited evidence, and this comparison alone does not distinguish an LC-specific effect from the benefit of a shorter remaining integration horizon. A full evaluation is needed to assess the robustness of this trend.

B.3 Example for one-step LC peak

Figure 3 provides a visual comparison of the one-step recovery diagnostic. The *LC-Suffix* strategy preserves the precise macroscopic layout of the 20-step Ground Truth (GT) by expending integra-

[1] https://huggingface.co/datasets/Lakonik/t2i-prompts-coco-10k

Table 5: Generation quality at different LC-peak locations with the 20-step Stable Diffusion 3.5 teacher reference at different LC-peak locations. *Peak Step Matched Sampling* uses a sampling budget matched to the number of steps associated with the LC peak, together with the final generation step. This baseline matches the step count but does not follow or reuse the same flow trajectory to the LC-peak state. *LC Suffix One-Step* follows the original generation trajectory to the LC-peak state and then completes the remaining suffix in a single step. Results are reported as mean ± standard deviation.

LC Peak (t)	Method	PSNR ↑	SSIM ↑	LPIPS ↓
0.84	Peak Step	11.800 ± 0.349	0.4924 ± 0.0326	0.6823 ± 0.0203
	LC-Suffix	$\mathbf{17.878} \pm 0.438$	$\mathbf{0.6293} \pm 0.0312$	$\mathbf{0.3381} \pm 0.0244$
0.78	Peak Step	12.276 ± 0.316	0.4796 ± 0.0278	0.6228 ± 0.0200
	LC Suffix	$\mathbf{19.687} \pm 0.284$	$\mathbf{0.6748} \pm 0.0157$	$\mathbf{0.2617} \pm 0.0127$
0.71	Peak Step	12.546 ± 0.312	0.4872 ± 0.0237	0.6093 ± 0.0178
	LC Suffix	$\mathbf{22.202} \pm 0.339$	$\mathbf{0.7318} \pm 0.0151$	$\mathbf{0.2150} \pm 0.0114$
0.63	Peak Step	13.485 ± 0.272	0.5187 ± 0.0198	0.5422 ± 0.0159
	LC Suffix	$\mathbf{24.399} \pm 0.329$	$\mathbf{0.7996} \pm 0.0110$	$\mathbf{0.1434} \pm 0.0081$
0.53	Peak Step	14.673 ± 0.347	0.5821 ± 0.0213	0.4504 ± 0.0182
	LC Suffix	$\mathbf{27.663} \pm 0.485$	$\mathbf{0.8723} \pm 0.0088$	$\mathbf{0.0836} \pm 0.0068$
0.40	Peak Step	16.877 ± 0.675	0.6667 ± 0.0256	0.3528 ± 0.0273
	LC Suffix	$\mathbf{32.090} \pm 0.773$	$\mathbf{0.9310} \pm 0.0060$	$\mathbf{0.0341} \pm 0.0035$

Table 6: Preliminary launch-offset evaluation over four runs with a 12-step SD3.5 teacher schedule. Offset $o = 0$ launches the Euler update at the LC peak; larger offsets launch it o steps earlier along the trajectory.

Offset o	PSNR ↑	SSIM ↑	LPIPS ↓
0 (at peak)	23.66	0.800	0.185
1	21.09	0.730	0.269
2	19.13	0.671	0.384
3	17.79	0.632	0.474
4	16.10	0.573	0.604

tion capacity strictly prior to the LC peak and executing a single jump afterward. In contrast, the matched-budget *Peak-Step* baseline suffers from severe structural degradation.

Specifically, *Peak-Step Matched Sampling* uses a sampling budget identical to the number of steps associated with the LC peak, plus the final generation step. This baseline matches the total step count but does not follow or reuse the identical flow trajectory to the LC-peak state. Conversely, the *LC-Suffix One-Step* methodology follows the native generation trajectory exactly up to the LC-peak state, and then completes the remaining suffix in a single generative step.

C LC-Peak Hinge Shift Formulation

This section describes the plateau-hinge ablation presented in Section 4.4.1. We evaluate alternative hinge placements in noise-level space while retaining the native π-Flow sampling procedure. Sec.3.3 defines the main CWRD method. This appendix describes the separate plateau-hinge perturbation controls, including their weight construction, normalization, and training implementation.

Profile Construction. We linearly interpolate the normalized LC profile from 12 measured anchors onto a grid of $M = 1000$ points, $\sigma_i = i/(M-1)$, for $i = 0, \ldots, M-1$. Here, σ denotes the

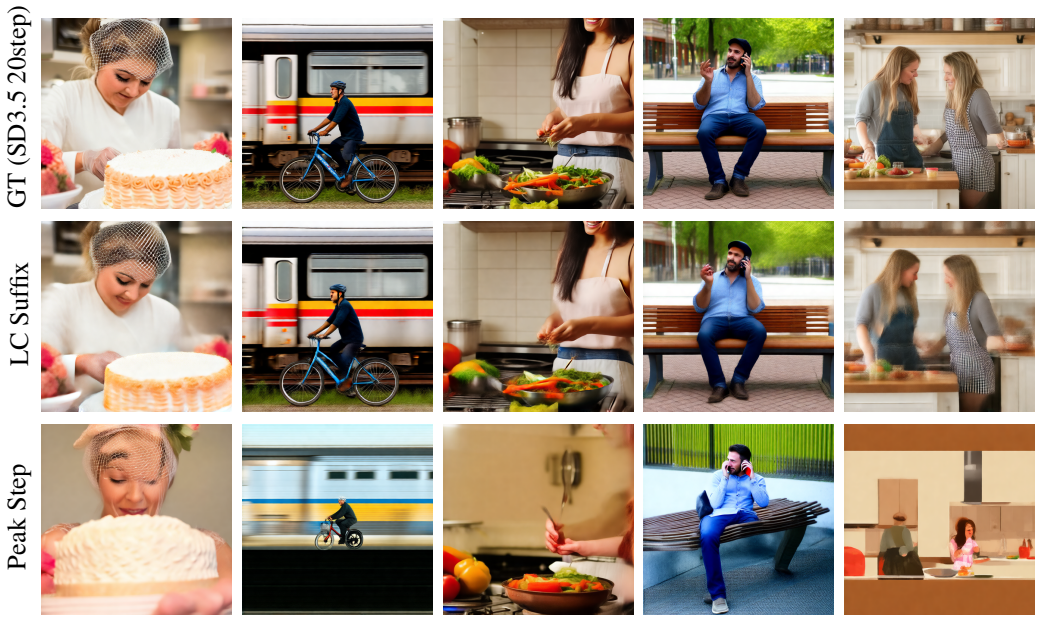

Figure 3: Visual comparison of one-step recovery. The *LC Suffix* preserves the precise macroscopic layout of the 20-step Ground Truth (GT) by expending capacity prior to the LC peak and executing a single jump afterward. The matched-budget *Peak Step* baseline suffers from severe structural degradation. *Peak Step Matched Sampling* uses a sampling budget matched to the number of steps associated with the LC peak, together with the final generation step. This baseline matches the step count but does not follow or reuse the same flow trajectory to the LC-peak state. *LC Suffix One-Step* follows the original generation trajectory to the LC-peak state and then completes the remaining suffix in a single generation step.

normalized time coordinate t used in the main text, with $\sigma = 0$ corresponding to the clean-latent endpoint and $\sigma = 1$ to pure noise. Let C_i denote the interpolated profile, and define $i^\star = \arg\max_i C_i$ and $\sigma^\star = i^\star/(M-1)$.

Let R denote the signed random offset and r a sampled realization. For an offset r, the perturbed hinge is

$$s_r = \mathrm{clip}(\sigma^\star + r, 0.05, 0.95), \qquad h_r = \mathrm{round}\big((M-1)s_r\big), \tag{12}$$

where rounding uses the nearest integer, with ties rounded to the nearest even integer. Negative offsets move the hinge toward $\sigma = 0$, later in generation, whereas positive offsets move it toward $\sigma = 1$, earlier in generation. The perturbed profile is

$$b_i^{(r)} = \begin{cases} C_{h_r}, & i < h_r, \\ C_i, & i \geq h_r. \end{cases} \tag{13}$$

Setting $r = 0$ places the hinge at the interpolated LC maximum within this perturbation construction. This construction changes the plateau length and height; it does not cyclically permute the weights.

Randomized and Fixed Offsets. The randomized control draws one offset r_k per optimizer step. Both directions are eligible in the reported configuration. We sample a sign $S \in \{-1, +1\}$ with equal probability and set $R = SA$, where the magnitude A follows

$$A \mid S = -1 \sim \mathrm{Uniform}[0.15, 0.50), \qquad A \mid S = +1 \sim \mathrm{Uniform}[0.15, 0.95 - \sigma^\star). \tag{14}$$

The positive magnitude is therefore bounded by approximately 0.3174, whereas the negative magnitude can approach 0.50. These asymmetric bounds prevent clipping in the randomized control. At each training step $k = 0, \ldots, 3999$, we initialize a fresh NumPy random generator with seed $42 \cdot 100003 + k$, sampling the sign before the magnitude.

The fixed controls use $r \in \{-0.50, -0.25, -0.15, +0.15, +0.25, +0.50\}$. Only the nominal $+0.50$ offset activates clipping: its hinge is bounded at $s_r = 0.95$, and an effective displacement of approximately $+0.3174$.

Table 7: Changes in generation metrics under fixed shifts $r \in \{\pm 0.15, \pm 0.25, \pm 0.50\}$ of the LC-derived allocation profile. Each row reports the difference relative to the aligned allocation ($r = 0$).

r	FID ↓	IS ↑	Precision ↑	Recall ↑
-0.50	$+0.265$	-0.282	-0.001	-0.005
-0.25	$+0.149$	-0.156	-0.001	-0.002
-0.15	$+0.089$	-0.193	-0.001	0.000
$+0.15$	$+0.371$	-0.267	0.000	-0.003
$+0.25$	$+1.721$	-0.863	-0.017	-0.030
$+0.50$	$+2.509$	-1.415	-0.044	-0.043

Weight Normalization. Every profile undergoes the same flooring and normalization. Defining the grid mean as $\langle a \rangle = M^{-1} \sum_{i=0}^{M-1} a_i$, we compute

$$a_i = \max(b_i^{(r)}, 10^{-8}), \qquad u_i = \max\left(0.2, \frac{a_i}{\langle a \rangle}\right), \qquad \omega_i^{(r)} = \frac{u_i}{\langle u \rangle}. \tag{15}$$

This normalization ensures $M^{-1} \sum_{i=0}^{M-1} \omega_i^{(r)} = 1$. It does not guarantee equal expected weights under the native intermediate-time distribution or equal weight variance across arms.

Application During Training. Anchor sampling, intermediate-state construction, and target generation follow native π-Flow. We multiply each intermediate-state mean squared error (MSE) term by the weight at its noise level. Define

$$I(\sigma) = \mathrm{clip}(\mathrm{round}((M-1)\sigma), 0, M-1)\,. \tag{16}$$

The weighted loss at optimizer step k is

$$\widehat{\mathcal{L}}_k(\Theta) = \frac{1}{BJ} \sum_{b=1}^{B} \sum_{j=1}^{J} \omega_{I(\sigma_{b,j})}^{(r_k)}\, \ell_{b,j}(\Theta), \tag{17}$$

where $\ell_{b,j}$ denotes the native intermediate-state MSE. With batch size $B = 4$ and $J = 2$ intermediate states per example, each optimizer step contains eight weighted terms. The randomized control shares one offset r_k and its resulting profile across all eight terms. Static controls construct their profiles once and reuse them throughout training. The timestep sampling distribution remains unchanged.

This ablation compares the chosen LC-based plateau with alternative hinge placements. Because the perturbations also change the weight distribution, it does not isolate temporal alignment at a fixed allocation concentration.

C.1 The different peak shift results

As detailed in Table 7, the LC-aligned allocation achieves the lowest reported FID among the tested shift settings. The shifted allocations increase FID by between 0.089 and 2.509 relative to the aligned reference. They also yield lower Inception Scores, while precision and recall are lower or unchanged at the reported precision. These results support the effectiveness of the LC-derived temporal placement relative to the evaluated fixed-shift controls.

D Cross-Attention Entropy and Local Complexity

This section reports aggregate profiles and per-trajectory analyses of the relationship between Local Complexity (LC) and spatial cross-attention entropy, complementing Section 4.4.2. We extract cross-attention maps associated with the selected subject text tokens from the teacher's Multimodal Diffusion Transformer (MM-DiT) blocks and compute Shannon entropy over visual positions. For a normalized spatial attention distribution, lower entropy indicates greater concentration of attention mass across visual positions. Entropy does not encode the geometric arrangement of those positions or establish that a semantic layout has become fixed.

D.1 Aggregated Trajectory Dynamics

Table 8 and Figure 4 summarize the measured profiles along a 20-step generation schedule, with generation proceeding from $t = 1$ toward $t = 0$. Across the reported states, mean attention entropy generally decreases before the LC peak and increases afterward. The mean LC profile reaches its maximum and the mean entropy profile reaches its minimum at the same sampled time, $t = 0.448$. We interpret this coincidence as an aggregate association on the evaluated time grid.

Separately, averaging correlations computed for individual teacher trajectories yields a mean Pearson correlation of -0.699 and a mean Spearman correlation of -0.694. These values summarize trajectory-level correlations; they are not correlations computed from the averaged profiles in Table 8.

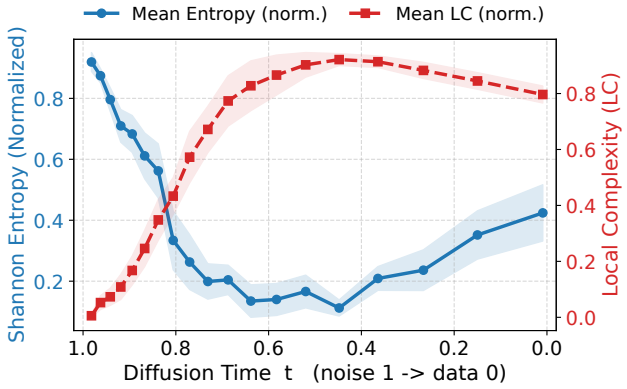

Figure 4: Cross-attention entropy versus Local Complexity across a 20-step schedule. The LC peak coincides with the entropy minimum, separating a broadly decreasing entropy trend from a subsequent stabilization.

Table 8: Cross-attention entropy and Local Complexity by latent state (20-step schedule) with full results (mean $\pm$ standard error). The data reveals a strong negative correlation (Spearman -0.6942, Pearson -0.6990) between the macroscopic LC metric and the semantic attention entropy.

Time t	Entropy (Norm)	LC (Norm)	Entropy	LC
0.982	0.919 ± 0.031	0.006 ± 0.006	6.342 ± 0.053	1.750 ± 0.186
0.962	0.874 ± 0.023	0.053 ± 0.025	6.309 ± 0.050	2.671 ± 0.502
0.941	0.796 ± 0.039	0.073 ± 0.031	6.254 ± 0.059	3.010 ± 0.587
0.919	0.710 ± 0.054	0.109 ± 0.045	6.197 ± 0.071	3.667 ± 0.820
0.894	0.683 ± 0.061	0.168 ± 0.052	6.180 ± 0.073	4.778 ± 0.920
0.867	0.611 ± 0.075	0.246 ± 0.065	6.138 ± 0.078	6.488 ± 1.232
0.838	0.562 ± 0.089	0.348 ± 0.073	6.101 ± 0.086	8.578 ± 1.453
0.806	0.334 ± 0.097	0.433 ± 0.066	5.925 ± 0.092	10.394 ± 1.517
0.770	0.263 ± 0.090	0.573 ± 0.092	5.863 ± 0.091	13.234 ± 2.105
0.731	0.199 ± 0.058	0.672 ± 0.083	5.799 ± 0.067	15.598 ± 2.477
0.687	0.204 ± 0.048	0.774 ± 0.090	5.794 ± 0.058	17.783 ± 2.857
0.638	0.135 ± 0.053	0.828 ± 0.089	5.737 ± 0.065	18.799 ± 2.878
0.583	0.140 ± 0.053	0.866 ± 0.072	5.739 ± 0.067	19.456 ± 2.658
0.520	0.166 ± 0.054	0.903 ± 0.044	5.764 ± 0.067	20.130 ± 2.374
0.448	0.112 ± 0.027	0.921 ± 0.021	5.705 ± 0.067	20.490 ± 2.162
0.364	0.209 ± 0.039	0.914 ± 0.021	5.773 ± 0.065	20.352 ± 2.094
0.266	0.236 ± 0.067	0.883 ± 0.029	5.809 ± 0.065	19.805 ± 2.104
0.150	0.352 ± 0.079	0.845 ± 0.030	5.868 ± 0.078	19.060 ± 2.047
0.009	0.424 ± 0.092	0.797 ± 0.030	5.923 ± 0.078	18.131 ± 2.004

Table 9: Per-trajectory correlations between cross-attention entropy and LC, grouped by LC peak time t^*. Results are reported as mean $\pm$ SEM across runs. Generation proceeds from $t = 1$ toward $t = 0$. The peak state is included in both segments.

	Pre-peak		Post-peak	
t^*	Pearson r	Spearman ρ	Pearson r	Spearman ρ
0.841	-0.763 ± 0.044	-0.855 ± 0.044	$+0.218 \pm 0.079$	$+0.239 \pm 0.082$
0.784	-0.759 ± 0.026	-0.837 ± 0.029	$+0.223 \pm 0.057$	$+0.253 \pm 0.059$
0.716	-0.678 ± 0.033	-0.746 ± 0.034	-0.165 ± 0.050	-0.203 ± 0.050
0.634	-0.682 ± 0.030	-0.732 ± 0.030	-0.085 ± 0.053	-0.128 ± 0.052
0.532	-0.670 ± 0.036	-0.708 ± 0.038	-0.076 ± 0.068	-0.033 ± 0.063
0.404	-0.634 ± 0.045	-0.676 ± 0.046	-0.165 ± 0.084	-0.092 ± 0.081

D.2 Instance-Level Dynamics for Different LC Peaks

To examine variation across trajectories, we compute Pearson and Spearman correlations separately for each teacher trajectory before and after its own LC peak. Measurements are aligned by latent state and diffusion time: LC recorded at step s and entropy recorded at step $s + 1$ both correspond to state z_s at $t = \sigma[s]$.

For a trajectory whose LC peak occurs at state k, the pre-peak and post-peak segments contain states $1, \ldots, k$ and $k, \ldots, N - 1$, respectively, so the peak state is included in both segments. We exclude segments containing fewer than three states. Per-trajectory correlations are averaged within each peak-time group, giving equal weight to every included trajectory, and are reported with their standard errors.

As shown in Table 9, mean pre-peak correlations are negative across all evaluated peak-time groups: Pearson correlations range from -0.763 to -0.634, and Spearman correlations range from -0.855 to -0.676. Thus, greater LC is associated with lower attention entropy during the pre-peak segment. After the peak, associations are weaker and heterogeneous, with mean Pearson correlations ranging from -0.165 to $+0.223$ and mean Spearman correlations ranging from -0.203 to $+0.253$. These results characterize an empirical relationship between endpoint-sensitivity variation and attention concentration across visual positions. The aggregate coincidence of the extrema and the trajectory-level correlations provide complementary diagnostics; they do not establish a universal boundary for layout completion or a causal explanation of distillation performance.

E Further Examples for different distillation method with CWRD

We present additional qualitative comparisons between baseline distillation methods and their counterparts incorporating CWRD. Figure 5 provides representative generation results across the evaluated models. These comparisons complement the quantitative evaluation by illustrating differences in structural coherence, the occurrence of visual artifacts, and the fidelity of fine-grained details.

F Limitation and Future works

our approach is subject to two notable limitations. First, the current Local Complexity metric is computed as a global, population-level expectation; while effective on aggregate, this macroscopic approximation does not accurately capture the exact, instance-specific geometric curvature of individual generative trajectories. Second, formulating the temporal weighting schedule $w(t)$ inherently requires an offline pre-computation phase to map the geometric profile, rendering CWRD a two-stage pipeline. To address these constraints, future research will explore dynamic, online LC estimation techniques. Bypassing the pre-computation bottleneck in favor of an instance-adaptive weighting scheme would allow the distillation process to react natively to sample-specific trajectory curvature, further optimizing capacity allocation during active training.

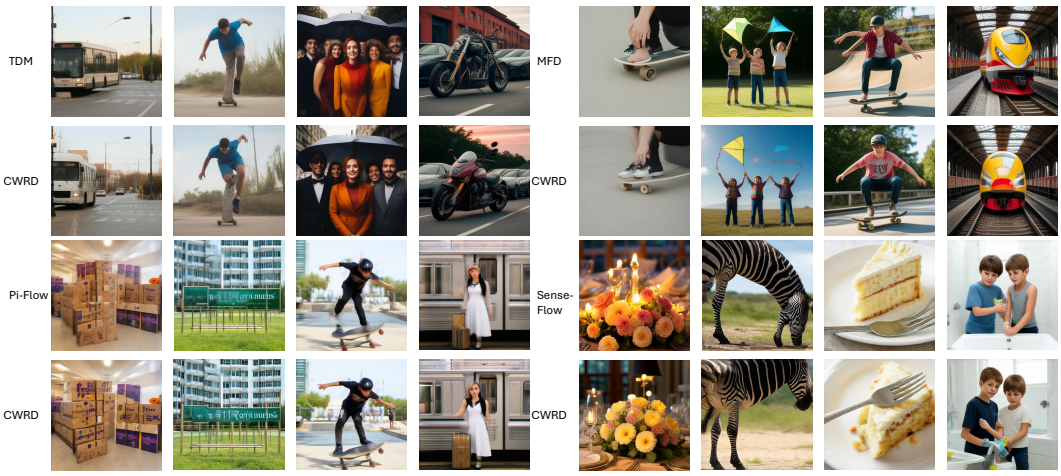

Figure 5: Additional qualitative comparisons between baseline distillation methods and their CWRD-augmented counterparts. The selected examples illustrate reductions in structural artifacts and improvements in fine-grained visual detail following the integration of CWRD.