IB-FLOW: INFORMATION BOTTLENECK-GUIDED CFG DISTILLATION FOR FEW-STEP TEXT-TO-IMAGE GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

While large-scale text-to-image generative models have achieved unprecedented visual performance, their inherent reliance on multi-step iterative solvers incurs severe inference latency. Few-step distillation targeting the Classifier-Free Guidance (CFG) trajectory has emerged as the prevalent dual-dimensional compression paradigm. However, existing frameworks remain subjugated by a coarse-grained blind injection paradigm that perpetually enforces a globally static guidance strength while indiscriminately sampling the supervisor timestep. This state-agnostic design completely disregards the intrinsic nature of image generation as a dynamic evolutionary process characterized by progressive entropy reduction, which not only restricts the performance boundary of few-step compression but also precipitates severe CFG over-conditioning artifacts. To transcend these limitations, we re-examine the distillation procedure through the theoretical lens of Information Theory, formally modeling it as a dynamic mutual information game constrained by the Information Bottleneck (IB) principle. Specifically, we dismantle traditional blind assumptions via a dual-track adaptive framework. To determine the injection target, we propose an instance-aware selection mechanism that transmutes the intractable KL divergence constraint into a closed-form conditional optimum solution predicated on the local vector field norm. To regulate the injection strength, we introduce an entropy-aware schedule that dynamically decays alongside the Signal-to-Noise Ratio (SNR), applying maximal thrust for initial structural anchoring before smoothly reverting to the natural manifold to refine micro-details. Extensive empirical evaluations corroborate that our framework efficiently reduces the over-conditioning artifacts, shattering the performance ceiling to achieve state-of-the-art (SOTA) generative fidelity under extremely stringent 2-step configurations.

1 INTRODUCTION

Recent advances in large-scale generative models, particularly those based on Diffusion Models Ho et al. (2020); Rombach et al. (2022) and Flow Matching Lipman et al. (2023), have achieved unprecedented success in text-to-image generation. By constructing continuous probability paths from simple prior distributions to complex data distributions, they enable highly realistic visual synthesis. However, these models heavily rely on iterative numerical solvers during the sampling phase, typically requiring dozens or even hundreds of Neural Function Evaluations (NFEs). This sequential denoising nature leads to severe inference latency and massive computational overhead, fundamentally limiting their deployment in real-time interactive applications.

To break the inference efficiency bottleneck of diffusion models, the core objective of few-step generation fundamentally necessitates a dual-dimensional compression paradigm, wherein step distillation truncates the multi-step trajectory into one or two steps Yin et al. (2024b); Salimans & Ho (2022); Song et al. (2023), while CFG distillation simultaneously compresses the conditional axis to eliminate the doubled forward-pass overhead imposed by Classifier-Free Guidance (CFG) Ho & Salimans (2022). Recent pioneering studies such as Decoupled DMD Yin et al. (2024a); Liu et al. (2026b) unveil a profound entanglement between these two dimensions by demonstrating that CFG distillation transcends mere computational reduction to act as the quintessential spear propelling the model toward large-stride generation and significantly elevating the quality ceiling of few-step

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.

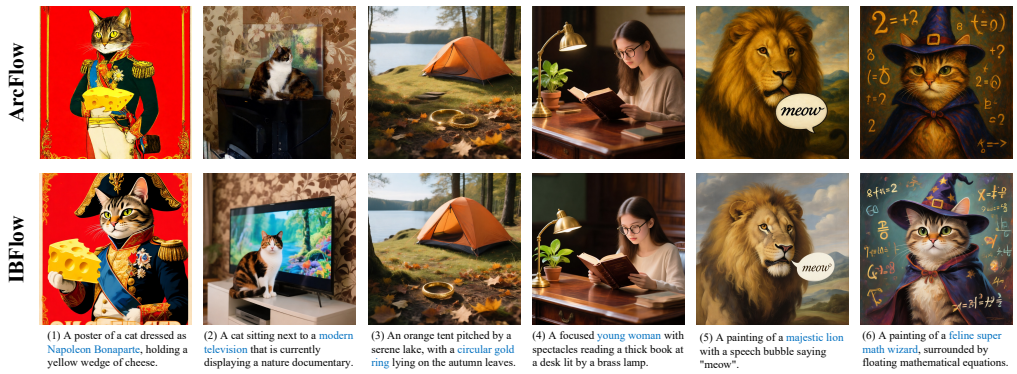


Figure 1: Demonstrating extreme few-step image generation at only 2 NFE, our IBFlow seamlessly improves perceptually orthogonal qualities compared to Arcflow, which improves from: (1,2) structural fidelity, (3,4) textural details, and (5,6) chromatic fidelity.

synthesis. Consequently, the paramount challenge transitions into losslessly condensing the teacher model's burdensome CFG-dependent guidance capability into the student model through a rigorous feature injection constraint co-determined by the **supervisor timestep** and the **guidance strength**.

Despite significant progress in compressing guidance trajectories, existing frameworks rely on a blind injection paradigm that perpetually enforces a globally static guidance strength while indiscriminately sampling the supervisor timestep from empirical priors. This state-agnostic design disregards the intrinsic nature of image generation as a dynamic, progressive entropy-reduction process. In the high-entropy initial stages, the chaotic representation urgently demands conservative structural anchoring; blindly leaping toward distant targets creates a massive semantic gap that precipitates severe noise degradation Liu et al. (2026b). Conversely, during the low-entropy terminal stages when the focus shifts to micro-detail refinement, maintaining rigid guidance disrupts the converged natural manifold, manifesting severe CFG over-conditioning artifacts like color oversaturation and texture sharpening. Ultimately, this rigid adherence to static injection fails to unlock the CFG engine's potential, emerging as the primary impediment to the performance boundary of few-step generative models.

To resolve these limitations, we re-examine the few-step distillation procedure through the theoretical lens of Information Theory to propose a dynamic information injection framework governed by the Information Bottleneck (IB) principle Tishby et al. (2000). Specifically, we dismantle the traditional blind assumptions from two dimensions: **injection target** and **injection strength**. On the one hand, to answer *what target to inject*, we reformulate the selection of the supervisor timestep as a capacity-constrained mutual information maximization problem that transmutes the intractable KL divergence constraint into a zero-overhead closed-form solution predicated on the local vector field norm. This instance-aware mechanism dynamically enforces conservative structural anchoring early on before facilitating advanced detail absorption tailored to specific sample complexities. On the other hand, to answer *how much strength to inject*, we adhere to the minimal sufficient representation tenet of IB to introduce an entropy-aware guidance strength schedule. By dynamically decaying alongside the Signal-to-Noise Ratio (SNR), it applies maximal thrust for initial symmetry breaking before smoothly reverting to the unconditional natural manifold, thereby eradicating over-conditioning artifacts. Extensive empirical evaluations corroborate that our framework achieves superlative generative fidelity under extremely stringent low-step configurations by completely shattering the performance ceiling inherent in traditional static distillation paradigms.

The main contributions of this work are summarized as follows:

(1) We expose the fundamental limitations of blind information injection in traditional CFG distillation from first principles, formally modeling the few-step distillation as a dynamic mutual information game constrained by the Information Bottleneck for the first time.

(2) We propose an *instance-aware* dynamic target selection mechanism. Through rigorous mathematical derivation, we prove that the conditional vector field norm can serve as an exact proxy for semantic uncertainty, yielding a zero-overhead closed-form optimal solution.

(3) We introduce an *entropy-aware* dynamic guidance strength scheduling strategy, which perfectly balances the alignment of conditional generation with the fidelity of the natural image manifold, effectively eradicating CFG over-conditioning artifacts.

(4) Extensive quantitative and qualitative experiments demonstrate that our method achieves State-of-the-Art (SOTA) generation quality and text-alignment under the extreme 2-step generation setting.

2 Related Work

2.1 Diffusion and Flow Matching Models

Diffusion models achieve state-of-the-art generative performance by learning to reverse an iterative noise-injection process Sohl-Dickstein et al. (2015); Ho et al. (2020); Song et al. (2021). Building on this, Flow Matching (FM) Lipman et al. (2023) and Rectified Flow Liu et al. (2023a) provide a unified framework based on continuous normalizing flows. By directly regressing the vector field from noise to data, FM yields straighter, optimal transport-guided trajectories that simplify training and elevate large-scale text-to-image synthesis. However, both SDE-based diffusion and ODE-based FM require dozens to hundreds of neural function evaluations (NFEs) during sampling. This heavy computational burden fundamentally hinders their real-time deployment Liu et al. (2026b).

2.2 Few-Step Diffusion Distillation

To address this latency, extensive research focuses on compressing multi-step models into few-step generators Salimans & Ho (2022); Gu et al. (2026); Geng et al. (2025). Key routes include consistency distillation Song et al. (2023); Wang et al. (2024); Zheng et al. (2025), enforcing trajectory self-consistency, and adversarial distillation Sauer et al. (2024); Lin et al. (2024), using GAN objectives to match real data distributions. Among these, score-based methods like Diff-Instruct Luo et al. (2023) and Distribution Matching Distillation (DMD) Yin et al. (2024b;a) are particularly promising Liu et al. (2026b). DMD minimizes an Integral Kullback-Leibler (IKL) divergence between student and teacher distributions, providing an elegant theoretical framework. Building on this, recent DMD variants and trajectory-matching approaches have successfully achieved high-resolution few-step generation Liu et al. (2023b); Zhu et al. (2024); Kim et al. (2025).

2.3 Classifier-Free Guidance and its Distillation

Classifier-Free Guidance (CFG) Ho & Salimans (2022); Saharia et al. (2022); Ho et al. (2022) is pivotal for achieving high-fidelity text-to-image alignment but doubles the per-step computational cost. Integrating CFG into distillation is therefore crucial, as teacher CFG patterns act as the primary driver for few-step conversion Liu et al. (2026b;a); Kim et al. (2025); Luhman & Luhman (2021). Notably, Decoupled DMD Liu et al. (2026b) rigorously isolates this into a Distribution Matching (DM) regularizer and a CFG Augmentation (CA) engine. However, existing implementations universally adopt a globally static CFG scale and randomly sample the CA supervisor timestep. This blind injection paradigm ignores the dynamic entropy-reduction during image generation, inevitably causing severe over-conditioning artifacts Ho & Salimans (2022); Liu et al. (2026b); Sauer et al. (2024); Lin et al. (2024).

3 Method

In this section, we formally delineate our dynamic information injection framework for ultra-efficient few-step generation. We first establish the theoretical bedrock by revisiting Flow Matching and CFG distillation. To accelerate iterative denoising, few-step distillation has been widely used in an attempt to compress tedious multi-step trajectories (e.g., 50 steps) into very low-step trajectories (e.g., 2 steps) without compromising the fidelity of generation.

3.1 Preliminaries

3.1.1 Flow Matching

Flow Matching (FM)Lipman et al. (2023) provides a unified framework based on continuous normalizing flows to learn the probability density path from a simple noise distribution p_0 to a complex data distribution p_1. Compared to traditional diffusion models, FM simplifies the training objective

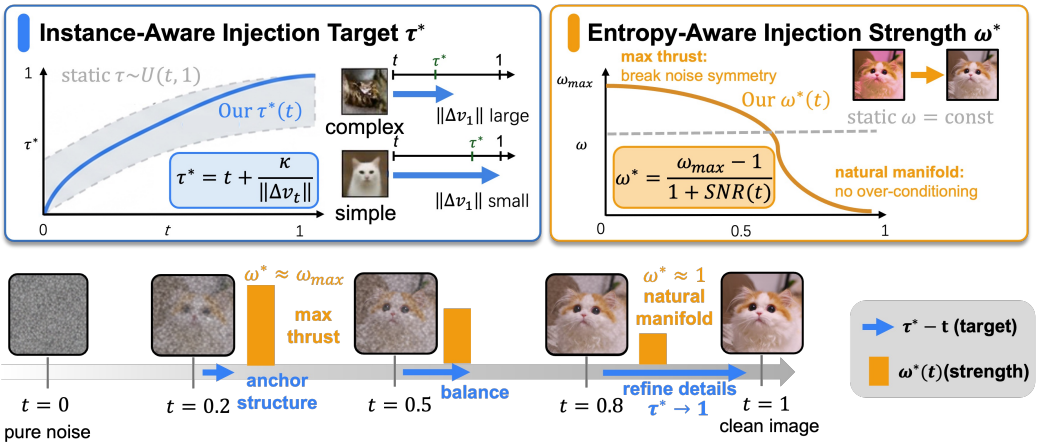


Figure 2: Overview of our dynamic information injection framework. While traditional distillation enforces a static guidance strength ω and blindly samples the supervisor timestep τ, our IB-guided approach dynamically selects the optimal injection target τ^* and strength ω^*.

by directly regressing the vector field, allowing the sampling process to be performed by solving an Ordinary Differential Equation (ODE). Specifically, given data $x_1 \sim p_1$ and noise $x_0 \sim p_0$, we define a linear probability path $x_t = (1-t)x_0 + tx_1$. The conditional vector field corresponding to this path is defined as:

$$u_t(x_t|x_1) = \frac{\mathrm{d}x_t}{\mathrm{d}t} = x_1 - x_0. \tag{1}$$

The training objective for the parameterized neural network $v_\theta(x_t, t)$ is to minimize the expected L_2 error between its prediction and the target vector field:

$$\mathcal{L}_{FM}(\theta) = \mathbb{E}_{t\sim\mathcal{U}(0,1),x_0,x_1} \left\| v_\theta(x_t,t) - (x_1 - x_0) \right\|^2. \tag{2}$$

During inference, generation quality heavily relies on the number of ODE solver steps. Although FM offers smoother trajectories than diffusion models, generating high-quality images with large-scale models typically still requires 40-100 Number of Function Evaluations (NFEs), limiting its application in real-time interactive scenarios.

3.1.2 Few-Step CFG Distillation and Decoupled Supervision

CFG is a crucial technique for enhancing text-to-image generation alignment. It strengthens semantic guidance by linearly extrapolating between the conditional vector field $v_t^c(x_t)$ and the unconditional vector field $v_t^u(x_t)$:

$$\tilde{v}_t^c(x_t, \omega) = v_t^u(x_t) + \omega \left(v_t^c(x_t) - v_t^u(x_t) \right), \tag{3}$$

where $\omega > 1$ is the guidance scale. Although CFG significantly improves image quality, it doubles the already high computational cost by requiring two forward passes at each inference step.

Few-step distillation aims to compress the costly multi-step trajectory into extremely few steps. Traditional frameworks Yin et al. (2024a) typically attempt to make the student model implicitly fit the modified trajectory. However, recent insights from Decoupled DMD Liu et al. (2026b) resolves this issue by rigorously decoupled into Distribution Matching (DM) and CFG Augmentation (CA). The gradient direction driving the student model parameter θ updates can be approximated as:

$$\Delta\theta \propto \mathbb{E}_{t,x_t,c} \left[\Big(\underbrace{\Delta_{DM}(\tau_{DM})}_{\text{Shield (Regularizer)}} + \underbrace{\Delta_{CA}(\tau_{CA}, \omega)}_{\text{Spear (Engine)}} \Big) \frac{\partial x_t}{\partial \theta} \right], \tag{4}$$

where the DM term $\Delta_{DM}(\tau_{DM}) = v^{real,c}_{\tau_{DM}}(x_{\tau_{DM}}) - v^{fake,c}_{\tau_{DM}}(x_{\tau_{DM}})$ acts as a regularizing shield to suppress generative artifacts by matching student and teacher distributions at the reference timestep τ_{DM}. More crucially, the CA functions as the core spear for achieving few-step compression:

$$\Delta_{CA}(\tau_{CA}, \omega) = (\omega - 1) \left(v^{real,c}_{\tau_{CA}}(x_{\tau_{CA}}) - v^{real,u}_{\tau_{CA}}(x_{\tau_{CA}}) \right). \tag{5}$$

The CA term distill the deterministic CFG decision pattern of the teacher model at a specific moment into the student model, where the guidance strength ω governing the injection magnitude and the supervisor timestep τ_{CA} determining the injection target are two control knobs of CA.

3.2 Instance-Aware Injection Target via Information Bottleneck

However, existing frameworks remain subjugated by a blind injection paradigm that perpetually enforces a constant guidance strength ω while randomly sampling the supervisor timestep τ_{CA} from a broad distribution. As image generation is a dynamic entropy reduction process, where the demand for both **injection strength** ω and **injection target** τ_{CA} inherently varies across different denoising stages t. This static CA injection severely limits the performance upper bound of few-step generation.

3.2.1 The Information Bottleneck Formulation

In a single forward pass of CFG distillation, the student at time step t absorbs guidance from the teacher by sampling the teacher at time step $\tau_{CA} > t$. A tradeoff is in this process: on the one hand, teachers should provide enough *information supply* for students to get enough guidance. On the other hand, the distribution difference between students and teachers must not be too large, otherwise it will bring high *learning costs*.

Quantifying the Information Supply: The ultimate goal of information injection is to aid the current state X_t in approximating the clean data state X_1. Thus, the effective information of the injected features from τ_{CA} can be rigorously quantified by the conditional mutual information:

$$\mathcal{I}_{supply}^{\tau_{CA}}(X_1, X_t) = I(X_{\tau_{CA}}; X_1 | X_t). \tag{6}$$

Bounding the Learning Cost: We describe the learning cost for students as the difference in distribution between the student and the teacher. In particular, this cost is quantified by the KL divergence between the two marginal distributions Cover & Thomas (2006):

$$D_{KL}(p_t \| p_{\tau_{CA}}) = \int p_t(x) \log \frac{p_t(x)}{p_{\tau_{CA}}(x)} dx. \tag{7}$$

where $p_t(x)$ and $p_{\tau_{CA}}(x)$ denote the marginal probability density functions of the generative process at the current evolutionary state t and the target injection stage τ_{CA}, respectively.

The Information Bottleneck in Distillation: Due to the monotonic diffusion process of flow matching, the predictive information regarding X_1 contained in $X_{\tau_{CA}}$ strictly increases with τ_{CA}. If we merely sought to maximize the supply, we should constantly select $\tau_{CA} = 1$ to acquire the most distinct high-frequency supervision. However, directly learning X_1 at the early stage brings extremely high learning cost, which can lead to distillation divergence Liu et al. (2026b). To solve this problem, we introduce the *Information Bottleneck (IB)* Tishby et al. (2000) and formalize the optimal target time step $\tau_{CA}^*(x_t, t)$ selection problem for the current state x_t at time t as a mutual information maximization problem with cost constraints:

$$\begin{aligned} \tau_{CA}^*(x_t, t) = \arg\max_{\tau \in [t,1]} \quad & I(x_{\tau_{CA}}; X_1 | X_t = x_t) \\ \text{s.t.} \quad & D_{KL}(p_t \| p_{\tau_{CA}}) \leq \delta. \end{aligned} \tag{8}$$

Where hyperparameter δ acts as a restricted instantaneous channel capacity. This formulation compactly reveals the mathematical essence of dynamic injection: during the high-entropy early stages, the divergence constraint forces a lower upper bound on τ_{CA}, ensuring the model absorbs safe, low-frequency structural features. Conversely, in the low-entropy later stages, the constraint relaxes, allowing the model to transition towards $\tau \to 1$ to fully leverage mutual information.

3.2.2 Tractable Approximation via Fisher Information

Based on equation 8, we need to obtain the analytic form of $\tau_{CA}^*(x_t, t)$. However, directly resolving the continuous KL divergence constraint formulated in within high-dimensional pixel space is profoundly intractable. To circumvent this bottleneck, we adopt the continuous vector field perspective inherent in FM and leverage Taylor expansion to transmute constraint into a zero-overhead closed-form solution. Specifically, we have the following lemma:

Lemma 1. *The KL divergence constraint in equation 8 can be approximate as:*

$$D_{KL}(p_t \| p_\tau) \approx \frac{1}{2}(\tau - t)^2 \|\Delta v_t(x_t)\|_2^2. \tag{9}$$

Proof. For an infinitesimal temporal increment $\Delta t = \tau - t$, the local KL divergence can be approximated via Taylor expansion utilizing the Fisher Information Metric (FIM). Under the CFG mechanism, the residual between the conditional and unconditional vector fields, denoted as $\Delta v_t(x_t) = v_t^c(x_t) - v_t^u(x_t)$, is mathematically rigorously proportional to the logarithmic gradient of Pointwise Mutual Information (PMI), where v_t^c and v_t^u are the frozen teacher conditional and unconditional vector fields. Consequently, the divergence penalty precipitated by information injection can be strictly bounded by the squared L_2 norm of the local vector field. The full proof can be found in Appendix A.1. □

Further, we consider the allowable range of $D_{KL}(p_t \| p_\tau)$, which, by Lemma 1, depends on the allowable upper bound of the injection span at the target time step τ. Specifically, we have the following lemma:

Lemma 2. *The injection span at the target time step τ satisfies:*

$$\tau - t \leq \frac{\sqrt{2\delta}}{\|\Delta v_t(x_t)\|_2}. \tag{10}$$

Finally, to maximize the mutual information supply, we adopt a greedy strategy by taking the equality. Combining Lemma 1 and Lemma 2, we have the following theorem:

Theorem 1. *The closed-form optimal solution of τ_{CA}^* is:*

$$\tau_{CA}^*(x_t, t) = \min\left(1, \quad t + \frac{\kappa}{\|v_t^c(x_t) - v_t^u(x_t)\|_2 + \epsilon}\right), \tag{11}$$

where $\tau_{CA}^*(x_t, t) \in [t, 1]$ is the selected teacher-supervision time for the current state x_t. $\kappa = \sqrt{2\delta}$ emerges as a hyperparameter encapsulating the globally divergence budget to explicitly govern the adaptive injection stride, while ϵ serves as a numerical stabilizing constant.

The proof can be found in Appendix A.2 and A.3. Grounded in this formulation, we achieve an Information Bottleneck-guided dynamic optimization for CFG injection target τ across various generative stages t where enforcing conservative macroscopic structural anchoring during the high-entropy initial phases, while facilitating meticulous micro-detail refinement in the solidified latter stages. By supplying adaptive and stabilized supervision for few-step CFG distillation, our framework shatters the performance ceiling and elevates the generative fidelity of synthesized images.

3.3 Entropy-Aware Dynamic Injection Strength

While the instance-aware target τ_{CA}^* ensures the *injection target* resides within a safe semantic divergence bound, it does not dictate the *injection strength*. Existing distillation work generally employs a static CFG scale ω across all time steps, and the consistent guidance strength destroys the local characteristics of the aggregation, leading to CFG overconditioning effects, including color supersaturation and abnormal texture sharpening. This highlights that the system's demand for conditional information C varies dynamically.

We consider the IB to quantify the dynamic demand for conditional information C. Analogous to the optimization of the injection target, we establish a rigorous theoretical equilibrium between the conditioning cost $I(X_t^\omega; C)$ and the resultant generative fidelity $I(X_t^\omega; X_1|C)$:

$$\min_{\omega(t)} \quad \underbrace{I(X_t^\omega; C)}_{\text{Conditioning Cost}} - \beta \underbrace{I(X_t^\omega; X_1|C)}_{\text{Generative Fidelity}}, \tag{12}$$

where X_t^ω denotes the updated internal state driven by the guidance strength ω, and β is a hyperparameter balancing the two terms.

Based on equation 8, we obtain the analytic form of $\omega^*(t)$ as follows:

Theorem 2. *The closed-form optimal solution of $\omega^*(t)$ is:*

$$\omega^*(t) = 1 + (\omega_{max} - 1) \cdot \frac{1}{1 + \gamma \cdot \mathit{SNR}(t)}, \quad \textit{where} \quad \mathit{SNR}(t) = \frac{t^2}{(1-t)^2}, \tag{13}$$

where ω_{max} designates the maximum guidance scale required to shatter the structural symmetry of pure noise in the initial high-entropy stage, and γ is a temperature hyperparameter meticulously calibrating the sensitivity to the SNR transition.

Proof. At any timestep t, the effective information gain achievable by increasing ω is bounded by the system's current residual uncertainty regarding the real data X_1, formally expressed as the conditional entropy $H(X_1|X_t)$. Based on the fundamental I-MMSE relation Guo et al. (2005), this residual uncertainty is analytically proven to be inversely proportional to the system's instantaneous Signal-to-Noise Ratio (SNR). Consequently, the degree of unknownness in the generative process can be explicitly quantified by the precise evolution of the SNR. Furthermore, to maximize generative fidelity while preventing manifold distortion caused by over-fitting the condition C, the supplementary driving force injected by the CFG engine, defined as $\omega(t) - 1$, must be proportional to the current residual uncertainty. By coupling this proportionality with the I-MMSE theorem, we substitute residual uncertainty with the instantaneous SNR. The full proof can be found in Appendix A.4. □

Theorem 2 mathematically guarantees ideal boundary conditions: as $t \to 1$ and residual uncertainty dissipates, the guidance thrust smoothly and inevitably reverts to the unconditional natural manifold $\omega^*(1) \to 1$, thereby eradicating CFG-induced artifacts from first principles.

3.4 Dynamic CFG Augmentation Objective

Finally, by substituting our conclusion in Theorem 1 and Theorem 2 into equation 5, we formulate the ultimate distillation objective. The gradient update for the student parameter θ, driven explicitly by our dynamic CA formulation, is rigorously defined as:

$$\Delta\theta_{CA} \propto \mathbb{E}_{t,x_t,c}\left[\underbrace{(\omega^*(t) - 1)}_{\text{Dynamic Strength}} \underbrace{\left(v^{real,c}_{\tau^*_{CA}}(x_{\tau^*_{CA}}) - v^{real,u}_{\tau^*_{CA}}(x_{\tau^*_{CA}})\right)}_{\text{Dynamic Target}} \nabla_\theta v_\theta(x_t, c)\right]. \tag{14}$$

This theoretically grounded, parameter-free objective elegantly unifies macroscopic structural anchoring and microscopic detail refinement under a strict entropy-reduction trajectory, establishing a foundation for extreme few-step generative synthesis.

4 Experiments

4.1 Experimental setup

We apply the IBFlow distillation framework to three text-to-image teachers spanning different scales and paradigms: FLUX.1-dev Labs et al. (2025), OpenUni-L-512 Wu et al. (2025b), and Qwen-Image-20B Wu et al. (2025a), all of which generate images at 1024×1024 with classifier-free guidance under their default 50×2-NFE schedules. Following ArcFlow Yang et al. (2026), all students are trained on a unified corpus of 2.3M text prompts. We optimize the model under BF16 mixed precision on 32 A100 GPUs. The CA divergence budget is κ=1.5, the SNR temperature is γ=1.0, $\omega_{\max} = 4.0$ is inherited from each teacher's default CFG scale. For evaluation, we follow standard text-to-image protocols and report results on three benchmarks: GenEval Ghosh et al. (2023), DPG-Bench Hu et al. (2024), and OneIG-Bench Chang et al. (2026) for fine-grained alignment along five axes. We deploy ArcFlow with their official released codebases for fair comparison.

4.2 Main Results

Quantitative comparison. Table 1 compares IBFlow with representative 2-step distillation methods on each backbone, including SenseFlow Ge et al. (2025), pi-Flow Chen et al. (2026), ArcFlow Yang et al. (2026), RCGM Sun & Lin (2026), TwinFlow Cheng et al. (2026), and Qwen-Image-Lightning Fan et al. (2025). IBFlow attains the best overall scores across all three backbones at only

Table 1: Results on GenEval, DPG-Bench, and OneIG-Bench. Qwen-Image-20B uses CFG and thus 50×2 NFE.

Model	NFE	Geneval	DPG-Bench	OneIG-Bench				
				Alignment	Text	Diversity	Style	Reasoning
FLUX.1-dev	50	0.66	84.16	0.790	0.556	0.238	0.307	0.257
SenseFlow (FLUX)	2	0.60	79.86	0.743	0.230	0.139	0.341	0.212
pi-Flow (GM-FLUX)	2	0.58	82.36	0.764	0.141	0.216	0.332	0.212
ArcFlow (FLUX)	2	0.65	84.29	0.798	0.368	0.210	0.350	0.224
IBFlow-FLUX (Ours)	2	**0.68**	**85.13**	**0.802**	**0.465**	**0.236**	**0.312**	**0.247**
OpenUni-L-512	20×2	0.85	81.54	0.771	0.275	0.132	0.396	0.227
OpenUni-RCGM-512	2	0.85	80.15	0.742	0.005	0.129	0.365	0.196
OpenUni-TwinFlow-512	2	0.85	79.82	0.735	0.006	0.126	0.372	0.219
OpenUni-IBFlow-512 (Ours)	2	**0.86**	**81.15**	**0.756**	**0.089**	**0.231**	**0.385**	**0.231**
Qwen-Image-20B	50×2	0.87	88.32	0.880	0.888	0.194	0.427	0.306
Qwen-Image-Lightning	2	0.85	86.47	0.875	0.879	0.098	0.415	0.292
TwinFlow (Qwen)	2	0.82	87.01	0.862	0.825	0.130	0.364	0.267
pi-Flow (GM-Qwen) Chen et al. (2026)	2	0.83	86.45	0.837	0.634	0.176	0.382	0.259
ArcFlow (Qwen)	2	0.84	87.94	0.868	0.789	**0.182**	0.418	0.283
IBFlow-Qwen (Ours)	2	**0.86**	**88.67**	**0.877**	**0.863**	0.180	**0.425**	**0.301**

Table 2: Fine-grained DPG-Bench breakdown on Qwen-Image-20B at 2 NFE.

Method	NFE	DPG total	L1 axes				L2 axes			
			Global	Entity	Relation	Attribute	Whole	Spatial	Color	Texture
ArcFlow Yang et al. (2026)	2	87.94	83.89	93.09	94.58	90.02	94.86	94.93	94.28	90.08
IBFlow-Qwen (Ours)	2	**88.68**	**85.11**	**93.59**	**95.20**	**90.76**	**95.56**	**95.59**	**94.93**	**91.17**
Improvement (Δ)	–	**+0.74**	**+1.22**	**+0.50**	**+0.62**	**+0.74**	**+0.70**	**+0.66**	**+0.65**	**+1.09**

2 NFE. For challenging Qwen-Image-20B, IBFlow attains **0.86** GenEval and **88.67** DPG-Bench at only 2 NFE, outperforms state-of-the-art ArcFlow on both metrics (0.84 / 87.94), demonstrating that the IB-driven dynamic schedules tighten the injection during high-entropy stages and relax it once the trajectory solidifies, thereby alleviating in one shot the three failure modes that plague every static-CA baseline at 2 NFE. The improvements are more pronounced compared with TwinFlow which have 0.82 / 87.01 on GenEval / DPG-Bench, indicating that its self-adversarial recipe still struggles to preserve fine-grained semantic alignment under such an aggressive step budget.

Table 2 further dissects the DPG-Bench gain over ArcFlow Yang et al. (2026) along the official L1/L2 axes. The two largest improvements pinpoint the role of each dynamic schedule: *Global* rises by **+1.22**, showing that the IB-bounded supervisor τ^*_{CA} keeps the student anchored to safe low-frequency layouts during the high-entropy early stage; *Texture* rises by **+1.09**, mirroring the boundary behavior $\omega^*(t) \to 1$ that retracts late-stage guidance and removes over-sharpening. The remaining axes (Entity, Relation, Color) are also improved, confirming that the two complementary IB-driven schedules jointly lift fidelity across the entire entropy-reduction trajectory rather than along a single dimension.

Qualitative comparison. Fig. 3 compares samples from Qwen-image-Lightning, ArcFlow, TwinFlow and IBFlow, with the two rightmost columns ablating τ^*_{CA} and $\omega^*(t)$. Static-CA baselines fail along four orthogonal dimensions: collapsed *structural fidelity*, broken *semantic alignment*, distorted *chromatic fidelity*, and degraded *textural detail*. IBFlow simultaneously improves all four, matching the multi-step teacher across structure, semantics, color and texture. Removing τ^*_{CA} predominantly damages structural fidelity and semantic alignment, while removing $\omega^*(t)$ predominantly damages chromatic fidelity and textural detail, providing a one-to-one visual confirmation of the complementary roles of our dynamic injection of target and strength.

4.3 Ablation studies

The two dynamic schedules are complementary. Table 3 isolates the effect of each dynamic component on Qwen-Image-20B under identical training budget and hyperparameters. Adding the dynamic injection target τ^*_{CA} alone improves ArcFlow from 0.84 / 87.94 to 0.85 / 88.37; adding the dynamic injection strength $\omega^*(t)$ alone improves it to 0.86 / 88.50; combining both yields the best 0.86 / **88.67**. The two schedules thus address complementary failure modes: τ^*_{CA} prevents the student

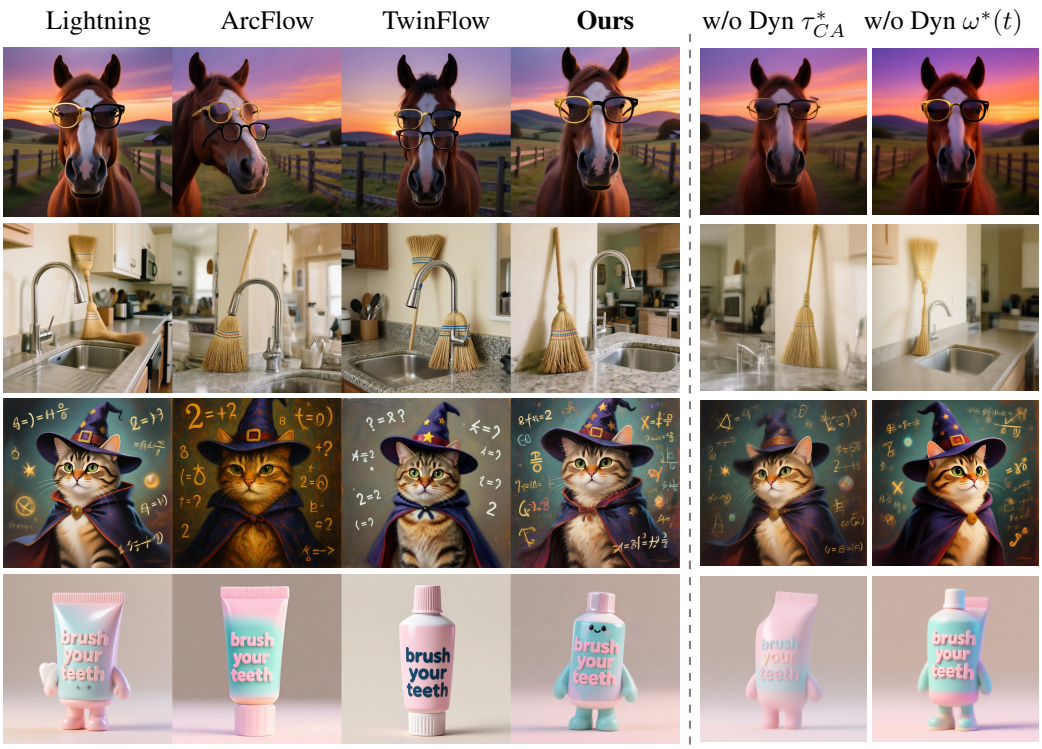


Figure 3: Qualitative comparison on DPG-Bench. IBFlow simultaneously improves *structural fidelity*, *semantic alignment*, *chromatic fidelity*, and *textural detail* over all 2-step baselines, while the rightmost two columns confirm the necessity of the dynamic target τ^*_{CA} and dynamic strength $\omega^*(t)$.

Table 3: Component breakdown on Qwen-Image.

Method	Geneval	DPG
ArcFlow	0.84	87.94
+ Dyn τ^*_{CA}	0.85	88.37
+ Dyn $\omega^*(t)$	0.86	88.50
IBFlow	**0.86**	**88.67**

Table 4: Impact of divergence budget κ.

Budget (κ)	Geneval	DPG
$\kappa = 0.5$	0.85	88.16
$\kappa = 1.0$	0.86	88.32
$\kappa = 1.5$	**0.86**	**88.67**
$\kappa = 3.0$	0.84	86.13

Table 5: Comparison of ω scheduling.

Schedule (ω)	Geneval	DPG
Constant	0.85	88.37
Linear Decay	0.85	87.82
Cosine Decay	0.86	88.16
SNR-Analyt.	**0.86**	**88.67**

from bridging an excessive semantic gap during early stages and recovers high-frequency fidelity, whereas $\omega^*(t)$ removes the over-conditioning artifacts predicted by Theorem 2.

Choice of the divergence budget κ. Table 4 sweeps the IB budget κ, which controls the maximum admissible KL divergence between the current state and the supervisor. A small κ over-restricts the injection target and starves the student of useful supervision, while an excessively large κ violates the local Fisher-information approximation in Lemma 1 and reintroduces distillation instability. The closed-form schedule attains its peak at κ=1.5, where IBFlow recovers the full 0.86 / **88.67**.

5 Conclusion

In this paper, we proposed IB-Flow, a novel Information Bottleneck-guided dynamic CFG distillation framework that significantly elevates the performance boundary of few-step text-to-image generation. Existing frameworks rely on blind injection paradigm that enforces static guidance strength while indiscriminately sampling the supervisor target. To overcome these bottlenecks, we introduced a dual-track adaptive mechanism: an instance-aware target selection strategy that guarantees conservative macroscopic structural anchoring during early high-entropy stages, and an entropy-aware guidance strength schedule that smoothly decays to the natural manifold to eradicate CFG over-conditioning artifacts. Extensive empirical evaluations across multiple large-scale backbones demonstrate that IB-Flow achieves state-of-the-art generative fidelity under extremely 2-step configurations, effectively restoring structural and chromatic fidelity.

Reproducibility Statement

The manuscript specifies the teacher models, training data, optimization settings, compute resources, evaluation benchmarks, and principal hyperparameters used by IB-Flow. The appendix provides complete derivations of the theoretical claims and additional implementation details. We also submit our complete code in the supplementary material to facilitate reproduction of the reported results, subject to the licenses of the underlying models and datasets.

Ethics Statement

The authors have read and adhere to the ICLR Code of Ethics. This work does not involve human subjects or the collection of private personal data. Our experiments use pretrained generative models, synthetic training data, and established evaluation benchmarks subject to their respective licenses. As with other text-to-image generation systems, the resulting models may inherit biases from the underlying models and data and may be misused to create misleading synthetic content. We encourage responsible use together with appropriate provenance, watermarking, and content-safety safeguards.

AI Usage Disclosure

Generative AI tools were used to polish the writing and identify grammatical errors in the manuscript. In addition, FLUX.1-dev, OpenUni, and Qwen-Image-20B was used as a generative teacher to produce the synthetic image data used for zero-shot distillation training. The authors manually reviewed all AI-assisted text and take full responsibility for the generated training data, scientific claims, experiments, and final content of this work.

References

Jingjing Chang, Yixiao Fang, Peng Xing, Shuhan Wu, Wei Cheng, Rui Wang, Xianfang Zeng, Gang Yu, and Hai-Bao Chen. Oneig-bench: Omni-dimensional nuanced evaluation for image generation. *Advances in Neural Information Processing Systems*, 38, 2026.

Hansheng Chen, Kai Zhang, Hao Tan, Leonidas Guibas, Gordon Wetzstein, and Sai Bi. pi-flow: Policy-based few-step generation via imitation distillation. In *International Conference on Learning Representations*, volume 2026, pp. 151521–151547, 2026.

Zhenglin Cheng, Peng Sun, Jianguo Li, and Tao Lin. Twinflow: Realizing one-step generation on large models with self-adversarial flows. In *International Conference on Learning Representations*, volume 2026, pp. 151237–151264, 2026.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2 edition, 2006. doi: 10.1002/047174882X. URL https://doi.org/10.1002/047174882X.

Xiangyu Fan, Zesong Qiu, Fanzhou Wang, Peng Gao, and Lei Yang. Qwen-image-lightning, 2025. URL https://github.com/ModelTC/LightX2V-Qwen-Image-Lightning.

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.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, J Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. *arXiv preprint arXiv:2505.13447*, 2025.

Dhruba Ghosh, Hannaneh Hajishirzi, and Ludwig Schmidt. Geneval: An object-focused framework for evaluating text-to-image alignment. *Advances in Neural Information Processing Systems*, 36: 52132–52152, 2023.

Jiatao Gu, Tianrong Chen, David Berthelot, Huangjie Zheng, Yuyang Wang, Ruixiang Zhang, Laurent Dinh, Miguel Angel Bautista, Joshua Susskind, and Shuangfei Zhai. Starflow: Scaling latent normalizing flows for high-resolution image synthesis. *Advances in Neural Information Processing Systems*, 38:120986–121022, 2026.

Dongning Guo, Shlomo Shamai, and Sergio Verdú. Mutual information and minimum mean-square error in gaussian channels. *IEEE transactions on information theory*, 51(4):1261–1282, 2005.

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

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J Fleet. Video diffusion models. *Advances in neural information processing systems*, 35:8633–8646, 2022.

Xiwei Hu, Rui Wang, Yixiao Fang, Bin Fu, Pei Cheng, and Gang Yu. Ella: Equip diffusion models with llm for enhanced semantic alignment. *arXiv preprint arXiv:2403.05135*, 2024.

Beomsu Kim, Yu-Guan Hsieh, Michal Klein, Jong Chul Ye, Bahjat Kawar, James Thornton, et al. Simple reflow: Improved techniques for fast flow models. In *International Conference on Learning Representations*, volume 2025, pp. 67740–67766, 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.

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

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

Deyuan Liu, Peng Sun, Yansen Han, Zhenglin Cheng, Chuyan Chen, and Tao Lin. Self-adversarial one step generation via condition shifting. *arXiv preprint arXiv:2604.12322*, 2026a.

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

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023a.

Xingchao Liu, Xiwen Zhang, Jianzhu Ma, Jian Peng, et al. Instaflow: One step is enough for high-quality diffusion-based text-to-image generation. In *The Twelfth International Conference on Learning Representations*, 2023b.

Eric Luhman and Troy Luhman. Knowledge distillation in iterative generative models for improved sampling speed. *arXiv preprint arXiv:2101.02388*, 2021.

Weijian Luo, Tianyang Hu, Shifeng Zhang, Jiacheng Sun, Zhenguo Li, and Zhihua Zhang. Diff-instruct: A universal approach for transferring knowledge from pre-trained diffusion models. *Advances in Neural Information Processing Systems*, 36:76525–76546, 2023.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in neural information processing systems*, 35:36479–36494, 2022.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. *arXiv preprint arXiv:2202.00512*, 2022.

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

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. pmlr, 2015.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

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

Peng Sun and Tao Lin. Any-step generation via n-th order recursive consistent velocity field estimation. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=GnawtLKGkP.

Naftali Tishby, Fernando C Pereira, and William Bialek. The information bottleneck method, 2000. URL https://arxiv.org/abs/physics/0004057.

Fu-Yun Wang, Zhaoyang Huang, Alexander Bergman, Dazhong Shen, Peng Gao, Michael Lingelbach, Keqiang Sun, Weikang Bian, Guanglu Song, Yu Liu, et al. Phased consistency models. In *Advances in Neural Information Processing Systems*, 2024.

Chenfei Wu, Jiahao Li, Jingren Zhou, Junyang Lin, Kaiyuan Gao, Kun Yan, Sheng-ming Yin, Shuai Bai, Xiao Xu, Yilei Chen, et al. Qwen-image technical report. *arXiv preprint arXiv:2508.02324*, 2025a.

Size Wu, Zhonghua Wu, Zerui Gong, Qingyi Tao, Sheng Jin, Qinyue Li, Wei Li, and Chen Change Loy. Openuni: A simple baseline for unified multimodal understanding and generation. *arXiv preprint arXiv:2505.23661*, 2025b.

Zihan Yang, Shuyuan Tu, Licheng Zhang, Qi Dai, Yu-Gang Jiang, and Zuxuan Wu. Arcflow: Unleashing 2-step text-to-image generation via high-precision non-linear flow distillation. *arXiv preprint arXiv:2602.09014*, 2026.

Tianwei Yin, Michaël Gharbi, Taesung Park, Richard Zhang, Eli Shechtman, Fredo Durand, and Bill 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.

Kaiwen Zheng, Yuji Wang, Qianli Ma, Huayu Chen, Jintao Zhang, Yogesh Balaji, Jianfei Chen, Ming-Yu Liu, Jun Zhu, and Qinsheng Zhang. Large scale diffusion distillation via score-regularized continuous-time consistency. *arXiv preprint arXiv:2510.08431*, 2025.

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 PROOFS OF THE INFORMATION-BOTTLENECK FORMULATION

This appendix provides the full proofs of Lemma 1, Lemma 2, Theorem 1, and Theorem 2 stated in the main text. The presentation strictly follows the order used in Section 3: each subsection corresponds to exactly one statement, and the labels match those referenced from the main paper.

A.1 PROOF OF LEMMA 1 (LOCAL KL BOUND VIA FISHER INFORMATION)

We prove that, under the Flow Matching framework, the local KL divergence between marginals at t and $\tau = t + \Delta t$ admits the approximation

$$D_{KL}(p_t \| p_\tau) \approx \tfrac{1}{2} (\tau - t)^2 \, \|\Delta v_t(x_t)\|_2^2.$$

For continuous normalizing flows, the marginal density $p_t(x_t)$ satisfies the continuity equation

$$\partial_t p_t + \nabla \cdot (p_t v_t) = 0. \tag{15}$$

For Gaussian probability paths used in FM (e.g., OT-Flow or VP-Flow), the predicted vector field v_t is intrinsically coupled with the Stein score $\nabla_{x_t} \log p_t$ via a time-dependent scalar weighting σ_t:

$$v_t(x_t) \propto \mathbb{E}[X_1 \mid X_t = x_t] - x_t \propto \sigma_t^2 \, \nabla_{x_t} \log p_t(x_t). \tag{16}$$

The CFG residual is $\Delta v_t = v_t(x_t, c) - v_t(x_t, \emptyset)$. Substituting equation 16,

$$\begin{aligned} \Delta v_t(x_t) &\propto \nabla_{x_t} \log p_t(x_t \mid c) - \nabla_{x_t} \log p_t(x_t) \\ &= \nabla_{x_t} \log \frac{p_t(x_t, c)}{p_t(x_t)\, p(c)} = \nabla_{x_t} \, \mathrm{PMI}_t(x_t; c), \end{aligned} \tag{17}$$

so $\|\Delta v_t\|_2$ is proportional to the spatial gradient of the pointwise mutual information injected by the condition c.

By a second-order Taylor expansion in Δt, the KL divergence between two infinitesimally close distributions on the same parametric path satisfies

$$D_{KL}(p_t \| p_{t+\Delta t}) \approx \tfrac{1}{2} (\Delta t)^2 \, \mathcal{I}_F(t), \quad \mathcal{I}_F(t) = \mathbb{E}_{p_t}\left[\left\| \partial_t \log p_t \right\|^2 \right]. \tag{18}$$

For the CA-driven transition, the dominant contribution to the temporal score variation is the conditional shift induced by Δv_t. Translating this temporal variation into the spatial PMI gradient gives

$$D_{KL}^{CA}(p_t \| p_\tau) \approx \tfrac{1}{2} (\tau - t)^2 \left\| \nabla_{x_t} \mathrm{PMI}_t(x_t; c) \right\|_2^2 \propto \tfrac{1}{2} (\tau - t)^2 \, \|\Delta v_t(x_t)\|_2^2, \tag{19}$$

which establishes Lemma 1. □

A.2 PROOF OF LEMMA 2 (ADMISSIBLE INJECTION SPAN)

Given the IB budget $D_{KL}(p_t \| p_\tau) \le \delta$, we plug the approximation from Lemma 1 (proved in Appendix A.1) into the constraint:

$$\tfrac{1}{2} (\tau - t)^2 \, \|\Delta v_t(x_t)\|_2^2 \;\le\; \delta. \tag{20}$$

Since $(\tau - t) \ge 0$ and $\|\Delta v_t(x_t)\|_2 \ge 0$, taking square roots and rearranging yields

$$\tau - t \;\le\; \frac{\sqrt{2\delta}}{\|\Delta v_t(x_t)\|_2}, \tag{21}$$

which is the admissible injection span stated in Lemma 2. The bound is instance-aware: a large CFG residual (sharp local conditional gradient) tightens the admissible reach, while a small residual permits looking further ahead. □

A.3 Proof of Theorem 1 (Closed-Form Optimal Injection Target)

We prove the closed-form solution

$$\tau^*_{CA}(x_t, t) = \min\left(1,\ t + \frac{\kappa}{\|v^c_t(x_t) - v^u_t(x_t)\|_2 + \epsilon}\right), \quad \kappa = \sqrt{2\delta}.$$

Under flow matching, the forward process $\{X_\tau\}$ is a monotone information channel from X_t towards X_1, so

$$\tau_1 \le \tau_2 \ \Rightarrow\ I(X_{\tau_1}; X_1 \mid X_t) \le I(X_{\tau_2}; X_1 \mid X_t). \tag{22}$$

The IB objective equation 8 is therefore non-decreasing in τ, and its optimum is attained at the largest τ permitted by the constraint.

Combining this monotonicity with the admissible span of Lemma 2, the unconstrained-by-terminal optimum is obtained:

$$\tau^\diamond = t + \frac{\sqrt{2\delta}}{\|\Delta v_t(x_t)\|_2}. \tag{23}$$

The supervisor timestep is bounded above by the terminal time $\tau \le 1$, since X_1 is the clean data state. Clipping $\tau^\diamond$ from above and adding a small $\epsilon > 0$ to the denominator to prevent division by zero, we obtain

$$\tau^*_{CA}(x_t, t) = \min\left(1,\ t + \frac{\kappa}{\|v^c_t(x_t) - v^u_t(x_t)\|_2 + \epsilon}\right), \tag{24}$$

with $\kappa = \sqrt{2\delta}$, which proves Theorem 1. □

A.4 Proof of Theorem 2 (Closed-Form Optimal Injection Strength)

We prove the SNR-driven schedule

$$\omega^*(t) = 1 + (\omega_{\max} - 1) \cdot \frac{1}{1 + \gamma\,\mathrm{SNR}(t)}, \quad \mathrm{SNR}(t) = \frac{t^2}{(1-t)^2}.$$

The IB condition-injection objective equation 12 contains the generative-fidelity term $I(X^\omega_t; X_1 \mid C)$. By the chain rule of conditional mutual information,

$$I(X^\omega_t; X_1 \mid C) \ =\ H(X_1 \mid C) \ -\ H(X_1 \mid X^\omega_t, C). \tag{25}$$

The first term $H(X_1 \mid C)$ is an intrinsic property of the data–condition joint distribution and is therefore independent of the guidance schedule ω. Consequently, the only ω-dependent component of the fidelity term is the negative residual entropy $-H(X_1 \mid X^\omega_t, C)$, which means the achievable fidelity gain obtained by raising ω is upper-bounded by the residual uncertainty about X_1 left after observing X^ω_t and C:

$$\Delta I(X^\omega_t; X_1 \mid C) \ \le\ H(X_1 \mid X^\omega_t, C) \ \le\ H(X_1 \mid X_t), \tag{26}$$

where the second inequality follows because conditioning on additional information (the condition C on top of X_t) cannot increase entropy. Equation equation 26 shows that the marginal fidelity gain from ω is governed by $H(X_1 \mid X_t)$, the residual uncertainty about the clean data given the current state.

By the I-MMSE relation Guo et al. (2005), when the forward trajectory is locally modelled as an additive Gaussian noise channel, the posterior variance of estimating X_1 from X_t scales inversely with the instantaneous signal-to-noise ratio. Up to a logarithmic factor,

$$H(X_1 \mid X_t) \ \propto\ \frac{1}{1 + \mathrm{SNR}(t)}. \tag{27}$$

Under the Rectified Flow path $X_t = tX_1 + (1-t)X_0$ with $X_0 \sim \mathcal{N}(0, I)$, the signal amplitude is exactly t and the noise amplitude is $1 - t$, giving

$$\mathrm{SNR}(t) = \frac{\|tX_1\|_2^2}{\|(1-t)X_0\|_2^2} = \frac{t^2}{(1-t)^2}. \tag{28}$$

To minimize the IB cost without inducing manifold distortion via over-conditioning, the supplementary CFG drive $\Delta\omega(t) = \omega(t) - 1$ should scale with the residual uncertainty in equation 27. Anchoring the maximum drive at the high-entropy initial stage with $\omega(0) = \omega_{\max}$ and substituting equation 28, we obtain

$$\omega^*(t) = 1 + (\omega_{\max} - 1) \cdot \frac{1}{1 + \gamma\,\mathrm{SNR}(t)}, \tag{29}$$

with $\gamma > 0$ controlling the SNR sensitivity. As $t \to 1$, $\mathrm{SNR}(t) \to \infty$ and $\omega^*(t) \to 1$, so the guidance smoothly retracts to the unconditional manifold and CFG over-conditioning artifacts are eliminated at the source. This proves Theorem 2. □

B Compute Resources

Each main training in the manuscript is conducted on 32 NVIDIA A100 GPUs with 80GB memory and takes approximately 1.5 days. We report the compute configuration for the experiments in the paper, while additional exploratory runs may require extra compute beyond the final reported settings.

C Limitations

Despite the superior efficiency and fidelity achieved by IB-Flow, several limitations remain to be addressed in future work:

Upper Bound of Teacher Distribution. As a distillation-based framework, IB-Flow primarily focuses on compressing the iterative sampling trajectory of a pre-trained teacher model. While our information-theoretic dynamic schedules significantly reduce over-conditioning artifacts and bridge the semantic gap, the fundamental generative capacity remains upper-bounded by the teacher's distribution and the training corpus. Since we maintain the same dataset as the teacher to ensure a fair comparison, the model does not exhibit a radical shift in the diversity or the underlying data distribution beyond what the teacher has already mastered.

Specific to Flow-based Trajectories. Our current mathematical derivation for the dynamic injection target τ^*_{CA} and strength ω^* is optimized for Flow Matching and Rectified Flow frameworks. While the Information Bottleneck principle is universal, extending these closed-form solutions to traditional SDE-based diffusion models might require additional approximations for the non-linear variance schedules.

D Broader Impacts

The development of IB-Flow carries both positive societal benefits and potential ethical risks:

Positive Impacts. By reducing the computational cost of high-quality image generation by $25\times$ to $50\times$, our framework democratizes access to state-of-the-art AI. This significantly lowers the hardware barrier for creative individuals and researchers, reducing the carbon footprint associated with large-scale inference and enabling real-time interactive applications on edge devices.

Negative Impacts. As with any advanced generative technology, the enhanced fidelity and extreme speed of IB-Flow could be misused for the rapid creation of deceptive synthetic media or "Deepfakes." The ability to generate near-photorealistic images in only two steps complicates the task of real-time detection and could potentially be exploited for disinformation campaigns. We encourage the community to integrate robust digital watermarking and provenance standards (e.g., C2PA) to mitigate these risks.