SCORE BRIDGE DISTILLATION: STUDENT-ADAPTIVE TARGETS FOR ONE-STEP DIFFUSION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Distribution matching distillation accelerates diffusion model sampling by training a few-step or one-step student with the score difference between the teacher and student distributions, without requiring sample-wise correspondence between their generated outputs. However, the score differences evaluated at different noise levels can induce conflicting update directions for the student, causing them to partially cancel and thereby slowing convergence and limiting the final generation quality. Existing progressive distillation methods rely on a sequence of intermediate distributions fixed before distillation; as the student evolves, this predefined path can become increasingly misaligned with its current distribution. To address this limitation, we propose Score Bridge Distillation (SBD), which periodically reconstructs teacher-guided bridge targets from the current student score estimate, thereby recalibrating the intermediate target as the student evolves while continuously guiding it toward the teacher distribution. SBD alternates between two stages: Bridge Score Adaptation and Bridge Distribution Matching. Bridge Score Adaptation constructs a bridge by adapting the current student score toward the teacher. Bridge Distribution Matching then trains the student to approximate this distribution. As the student distribution evolves, SBD repeatedly reconstructs the bridge distribution so that the training target at each stage remains adapted to the student's current state. This progressive matching process reduces gradient conflicts across noise levels, enabling the student to approach the teacher distribution more effectively. Extensive experiments show that SBD achieves faster convergence and better generation quality than existing distillation methods.

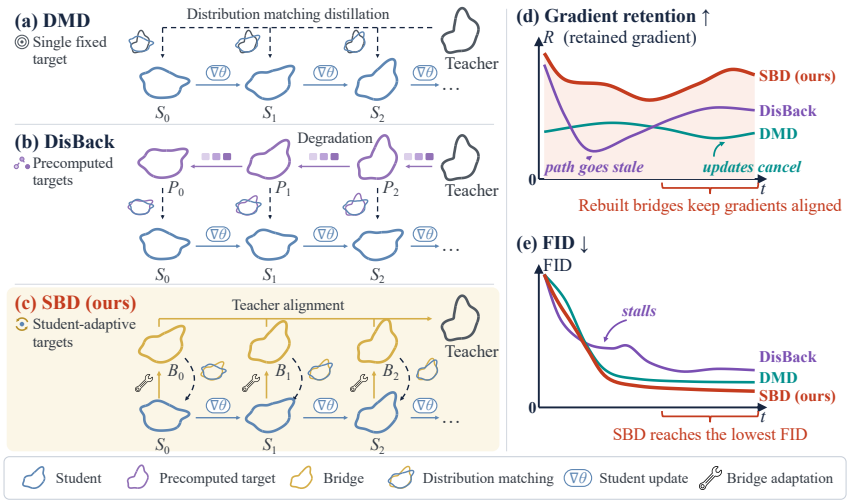


Figure 1: **Target construction in distribution matching distillation.** (a) DMD matches every student state S_i to the same fixed teacher. (b) DisBack precomputes intermediate targets P_i by degrading the teacher and traverses them in reverse. (c) SBD rebuilds a bridge B_i from the current student and adapts it toward the teacher, so each target stays matched to the student it guides. (d, e) Illustration of gradient retention and FID during distillation under the three targets.

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.

1 INTRODUCTION

Diffusion models have become a leading paradigm for high-fidelity image generation (Ho et al., 2020; Song et al., 2021b; Rombach et al., 2022), but their iterative sampling process requires tens of costly network evaluations (Karras et al., 2022). Diffusion distillation accelerates sampling by compressing a multi-step diffusion model into a generator requiring only a few denoising steps (Salimans & Ho, 2022; Song et al., 2023). Among its variants, distribution matching distillation (DMD) enables high-quality one-step generation by aligning the student distribution with the teacher distribution through their score difference, without requiring paired teacher–student samples (Luo et al., 2023; Yin et al., 2024b;a).

To handle the limited overlap between the teacher and student distributions, DMD computes their score differences over noisy marginals at multiple noise levels and aggregates the resulting generator updates. Prior studies have shown that gradients from different timesteps can conflict when training diffusion denoisers (Hang et al., 2023; Go et al., 2023; Park et al., 2024; Ma et al., 2025). In one-step DMD, this conflict directly affects the generated samples: score differences evaluated at different noise levels can prescribe incompatible update directions for the same underlying sample. These directions partially cancel one another in expectation and weaken the effective update toward the teacher distribution, a form of the gradient conflict studied in multi-task learning (Yu et al., 2020). We observe this effect directly in DMD training on CIFAR-10 (Krizhevsky, 2009). Late in DMD training on CIFAR-10, the updates from about one in five pairs of noise levels point in opposing directions, and about one third of the total update is canceled out. This motivates a central question: *How can we construct a training target that moves the student toward the teacher while providing more consistent guidance across noise levels?*

Intermediate distributions turn direct teacher matching into a sequence of stage-wise distribution-matching objectives, echoing intermediate teachers in knowledge distillation (Mirzadeh et al., 2020; Jin et al., 2019; Panigrahi et al., 2025). DisBack instantiates this strategy by first degrading the pretrained teacher toward the initial student and then traversing the recorded distributions in reverse during distillation (Zhang et al., 2025) as illustrated in Figure 1(b). Since this path is precomputed from the initial student, its intermediate targets can become increasingly mismatched with the student as training progresses. As Figure 1(d,e) show, this mismatch appears in the middle of training: the gradient retention of DisBack falls below that of DMD, and its FID stops improving while the other methods continue to improve.

We propose *Score Bridge Distillation* (SBD), an alternating framework for student-adaptive distribution matching, illustrated in Figure 1(c). Our key insight is that the teacher should determine the overall direction of distillation, while the student's current state should determine its immediate training target. Following this principle, *Bridge Score Adaptation* constructs a bridge from the current student and adapts it toward the teacher using teacher alignment and real-data denoising, yielding a teacher-guided yet data-anchored intermediate target. *Bridge Distribution Matching* then freezes the bridge and trains the student to match the resulting distribution. After each matching stage, SBD reconstructs the bridge from the updated student and repeats the procedure, producing a sequence of intermediate targets that evolves with the student rather than following a path fixed before distillation.

We also analyze why matching the bridge reduces cross-noise conflict. When the student matches an intermediate distribution instead of the teacher, it receives only part of the teacher–student score difference, and this part varies with the sample and the noise level. Finite adaptation determines how large this part is: noise levels at which the student is far from the teacher produce larger adaptation gradients, so the bridge moves further toward the teacher there. In a two-dimensional example, this reallocation makes the updates from different noise levels point in more consistent directions (Section 5).

The bridge is used only during training, so SBD keeps the original generator architecture and one-step sampling procedure. Compared with existing distribution matching distillation methods, SBD reduces gradient conflicts across noise levels and converges faster (Figure 4), and it also delivers better generation performance (Tables 1–3).

Our contributions are summarized as follows:

- We identify and quantify gradient conflicts across noise levels in one-step distribution matching distillation (Sections 5 and 6.2).

- We propose Score Bridge Distillation, which rebuilds the matching target from the current student and adapts it toward the teacher, achieving faster and better distillation (Section 4).
- Experiments show that SBD reduces gradient conflicts and achieves faster convergence and better generation quality on text-to-image generation with SDXL, unconditional generation, and ImageNet 64×64, without additional inference cost (Section 6).

2 Related Work

Distribution matching distillation. Fast diffusion sampling has been pursued through training-free samplers (Song et al., 2021a; Lu et al., 2022; Karras et al., 2022), trajectory distillation (Salimans & Ho, 2022; Song et al., 2023; Kim et al., 2024; Geng et al., 2025b; Lu & Song, 2025; Zheng et al., 2026; Wang et al., 2024; Ren et al., 2024), direct few-step generation (Frans et al., 2025; Geng et al., 2025a; Zhou et al., 2025a), and adversarial distillation (Goodfellow et al., 2014; Xu et al., 2024; Sauer et al., 2025; 2024; Lin et al., 2024). Distribution matching distillation instead trains a few-step or one-step generator to match the teacher distribution without reproducing its sampling trajectory. Diff-Instruct and variational score distillation optimize the generator using the difference between teacher and generator scores (Luo et al., 2023; Nguyen & Tran, 2024), while DMD and DMD2 make this formulation effective for one-step image generation through online estimation of the student score (Yin et al., 2024b;a). Subsequent work improves the divergence or its estimator (Zhou et al., 2024; 2025b;c; Luo et al., 2024; Xu et al., 2025; Wang et al., 2025b;a), or augments distribution matching with preference rewards (Luo, 2025) and consistency distillation (Chen et al., 2025). These methods primarily determine *which discrepancy* is minimized; SBD instead changes *which distribution* the student is matched to throughout training.

Gradient conflict across noise levels. Different noise levels can induce conflicting optimization signals, a phenomenon closely related to gradient interference in multi-task learning (Yu et al., 2020). Prior diffusion methods mitigate such interference through loss weighting, interval clustering, task routing, or model merging (Hang et al., 2023; Go et al., 2023; Park et al., 2024; Ma et al., 2025). In distillation, Phased DMD assigns different noise subintervals to separate experts (Fan et al., 2026), while the Curriculum Consistency Model balances the teacher–student gap across timesteps (Liu et al., 2025b). In contrast, SBD reduces conflicts among per-noise-level generator gradients by adapting the matching target itself, without partitioning either the noise range or the generator.

Intermediate targets. Intermediate supervision has long been used to bridge large teacher–student gaps, including teacher assistants (Mirzadeh et al., 2020) and routes of teacher checkpoints (Jin et al., 2019), which can induce an implicit curriculum (Panigrahi et al., 2025). In diffusion distillation, DisBack constructs a degradation path from the teacher to the initial student and then traverses this fixed path in reverse (Zhang et al., 2025). Other methods modify the matching target using refined samples (Xie et al., 2024), prescribed real–fake mixtures (Jayashankar et al., 2025), the student's sampling trajectory (Luo et al., 2025), or estimates of teacher reliability (Bai et al., 2026). A related line of work reallocates supervision across guidance terms, timesteps, or trajectory segments (Liu et al., 2025a; Ge et al., 2026; Wu et al., 2026; Liu et al., 2026; Kim et al., 2026). Unlike methods based on a precomputed path, SBD repeatedly rebuilds its intermediate target from the current student as training progresses. Appendix A provides detailed comparisons with the closest approaches.

3 Preliminary

Score distillation trains a one-step generator $G_\theta(z, c)$, with noise $z \sim \mathcal{N}(0, I)$ and condition c, to reproduce the distribution of a pretrained diffusion model. Let q_θ denote the distribution of the generated samples $x = G_\theta(z, c)$. The generated and teacher distributions are compared on noisy samples $y = x + \sigma\epsilon$ with $\epsilon \sim \mathcal{N}(0, I)$ and noise level $\sigma \sim \pi$. For a distribution p with denoiser $D_p(y, \sigma, c)$, the score is $s_p(y, \sigma, c) = (D_p(y, \sigma, c) - y)/\sigma^2$.

Distribution matching distillation (DMD) (Yin et al., 2024b) uses two score functions: the teacher score s_T given by the frozen pretrained denoiser T, and the fake score s_F given by a denoiser F_ψ that tracks the current generated distribution. The generator is updated with the gradient of the reverse KL

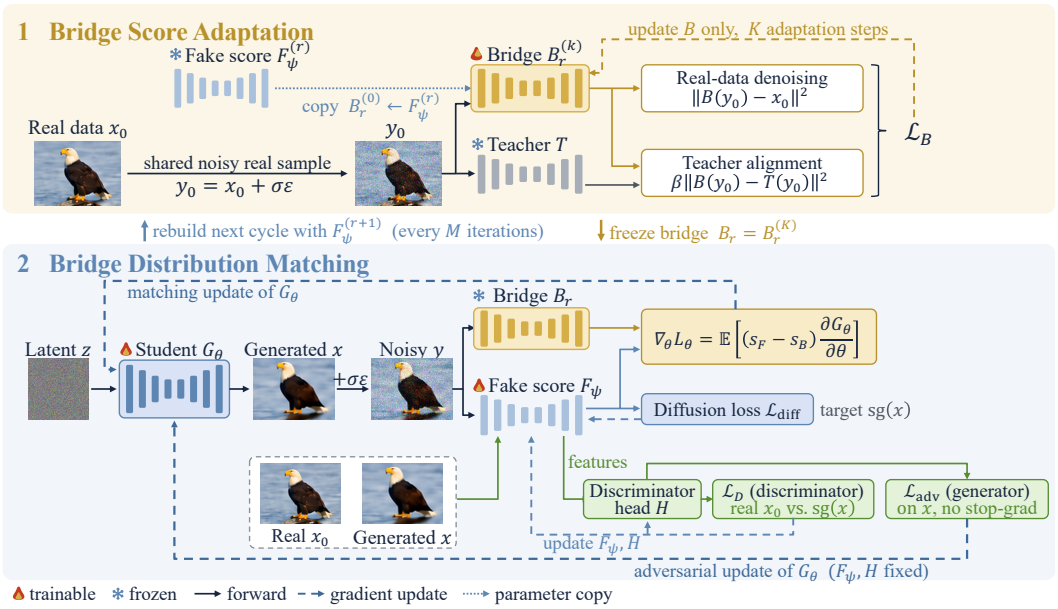


Figure 2: **Overview of SBD.** (1) Bridge Score Adaptation: the bridge is copied from the current fake score F_ψ and updated for K steps with real-data denoising and teacher alignment on shared noisy real samples, while the fake score and teacher remain frozen. (2) Bridge Distribution Matching: the frozen bridge replaces the teacher in the matching field that updates the student G_θ, while the fake score keeps fitting generated samples and the discriminator head provides adversarial supervision. The bridge is rebuilt every M iterations; only the student is used at inference.

divergence between the noisy distributions,

$$\nabla_\theta \mathbb{E}_{\sigma\sim\pi}\mathrm{KL}(q_{\theta,\sigma}\|p_{T,\sigma}) = \mathbb{E}\left[J_\theta G_\theta(z,c)^\top \big(s_F(y,\sigma,c) - s_T(y,\sigma,c)\big)\right], \tag{1}$$

and the fake denoiser is trained alternately on generated samples with the denoising objective

$$\mathcal{L}_F(\psi) = \mathbb{E}\left[\left\|F_\psi(y,\sigma,c) - x\right\|_2^2\right]. \tag{2}$$

Since σ is sampled from π, the generator update in Eq. (1) averages the updates from all noise levels. DMD2 (Yin et al., 2024a) further adds an adversarial loss: a discriminator head on the features of the fake denoiser learns to distinguish real images from generated ones, and the generator is trained to fool it, which provides supervision from real data in addition to the teacher.

4 Score Bridge Distillation

We introduce Score Bridge Distillation (SBD), which constructs the matching target from the student's current state rather than fixing it before distillation. The key idea is to let the teacher determine the overall direction of distillation while the current student determines its immediate target. Accordingly, SBD alternates between two stages: *Bridge Score Adaptation*, which constructs an intermediate target from the current student toward the teacher, and *Bridge Distribution Matching*, which trains the student to match this target. The bridge is then reconstructed from the updated student, allowing the matching target to evolve throughout training (Figure 2).

4.1 Bridge Score Adaptation

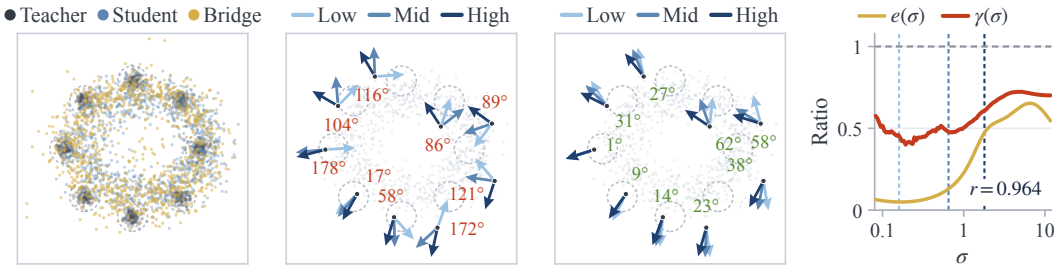
This stage aims to construct an intermediate target that is adapted to the current student. The fake score already tracks the current student distribution, so we initialize a bridge denoiser B_ϕ as a copy of F_ψ and adapt it toward the teacher for K steps with the objective

$$\mathcal{L}_B(\phi) = \mathbb{E}\left[\left\|B_\phi(y_0,\sigma,c) - x_0\right\|_2^2 + \beta\left\|B_\phi(y_0,\sigma,c) - \mathrm{sg}\left(T(y_0,\sigma,c)\right)\right\|_2^2\right], \tag{3}$$

Algorithm 1 Score Bridge Distillation

Require: Generator G_θ, fake denoiser F_ψ, bridge denoiser B_ϕ, teacher T; η, K, M, β

repeat
  $\phi \leftarrow \psi$ ▷ **Stage 1: Bridge Score Adaptation**; initialize B_ϕ from F_ψ
  for $k = 1, \ldots, K$ **do**
    $\phi \leftarrow \phi - \eta \nabla_\phi \mathcal{L}_B(\phi)$ ▷ Update bridge B_ϕ with Eq. (3)
  end for
  for $m = 1, \ldots, M$ **do** ▷ **Stage 2: Bridge Distribution Matching**
    $\theta \leftarrow \theta - \eta \nabla_\theta L_\theta$ ▷ Update generator G_θ with Eq. (4)
    $\psi \leftarrow \psi - \eta \nabla_\psi \mathcal{L}_F(\psi)$ ▷ Update fake denoiser F_ψ with Eq. (2)
  end for
until convergence
return G_θ



(a) Samples (b) Teacher target (c) Bridge target (d) Ratios vs. noise level

Figure 3: **Bridge matching makes cross-noise updates coherent in a 2D problem.** (a) Teacher, student, and bridge samples after 100 adaptation steps. (b,c) Generator updates at the same student samples under three noise levels ($\sigma = 0.16, 0.65, 1.8$, light to dark), with the teacher (b) or the bridge (c) as target; labels give the maximum pairwise angle among the three updates at each sample. (d) Score-error reduction $e(\sigma)$ and gradient-norm ratio $\gamma(\sigma)$ (Eq. 7); vertical lines mark the three noise levels of (b,c).

where (x_0, c) is a real image with its condition, $y_0 = x_0 + \sigma\epsilon$, and sg denotes stop-gradient. The teacher-alignment term moves the bridge toward the teacher, and the real-data term anchors it to the data distribution rather than to the teacher predictions alone. Since the adaptation starts from the student and runs for only a finite number of steps, the bridge score s_B describes an intermediate distribution between the student and the teacher. This stage does not update the generator and requires only a small number of steps ($K = 200$ in our experiments).

4.2 Bridge Distribution Matching

Given the bridge, this stage trains the student to match the bridge distribution instead of the teacher distribution. The bridge is frozen and replaces the teacher score in the generator update,

$$\nabla_\theta L_\theta = \mathbb{E}\big[J_\theta G_\theta(z, c)^\top \big(s_F(y, \sigma, c) - s_B(y, \sigma, c)\big)\big], \tag{4}$$

while the fake denoiser is still trained with Eq. (2). After M iterations, the student has moved toward the bridge, and SBD constructs a new bridge from the updated fake score. Repeating the two stages produces a sequence of intermediate targets that follows the student throughout training. Compared with DMD, the final target of SBD is still the teacher, while the immediate target at each stage is adapted to the current student. Algorithm 1 summarizes SBD, and implementation details are provided in Appendix B.1.

5 Why Does the Bridge Reduce Cross-Noise Conflict?

In DMD, let $g_\sigma^T(x)$ denote the effective update induced on a student sample x at noise level σ, obtained by averaging the score-difference guidance $s_F - s_T$ over its noisy copies $y = x + \sigma\epsilon$. The DMD update then aggregates these per-level updates across noise levels. When the updates of the

same sample point in different directions, part of them cancels. The bridge changes the strength and the direction of each per-level update before they are averaged.

To see how this can happen, consider an idealized bridge defined directly in distribution space as $b_\sigma = (1-\lambda_\sigma)q_\sigma + \lambda_\sigma p_\sigma$, where q_σ and p_σ denote the noisy student and teacher marginals and $0 < \lambda_\sigma < 1$. Its score difference satisfies

$$s_{q_\sigma}(y) - s_{b_\sigma}(y) = \rho_\sigma(y)\big(s_{q_\sigma}(y) - s_{p_\sigma}(y)\big), \qquad \rho_\sigma(y) = \frac{\lambda_\sigma p_\sigma(y)}{(1-\lambda_\sigma)q_\sigma(y) + \lambda_\sigma p_\sigma(y)}. \quad (5)$$

Under exact student-score estimation, replacing the teacher with this bridge scales the score difference at each noisy copy y by $\rho_\sigma(y) \in (0,1)$. Since $\rho_\sigma(y)$ varies across noisy realizations of the same sample x, bridge matching reweights their contributions non-uniformly. The resulting per-level update is therefore not generally a scalar rescaling of the teacher-matching update, and can change in both magnitude and direction.

The learned bridge is more general than the mixture. It starts from the current student score, where the matching update vanishes, and is adapted toward the teacher for a finite number of steps. The teacher-alignment gradient at each noise level is proportional to the bridge–teacher residual, so different noise levels move toward the teacher by different amounts before the bridge is frozen. We check both effects in the two-dimensional example of Figure 3, where only the bridge is adapted, for 100 steps with Eq. (3).

Direction. Under direct teacher matching, the same samples receive very different update directions from different noise levels (Figure 3b); with the adapted bridge, they are markedly more aligned (Figure 3c). Let $\{g_\ell\}_{\ell=1}^{L}$ be the updates of a sample at L noise levels, or the corresponding generator gradients in parameter space. Besides the *conflict rate*, the fraction of pairs with $\cos(g_i, g_j) < 0$, we measure the *retention rate*, the fraction of update magnitude kept after aggregation:

$$R = \Big\| \sum_{\ell=1}^{L} g_\ell \Big\|_2 \Big/ \sum_{\ell=1}^{L} \|g_\ell\|_2. \quad (6)$$

$R = 1$ when all per-level updates are aligned and decreases as they cancel. In this example, the bridge lowers the conflict rate from 10.2% to 2.9% and raises R from 0.899 to 0.956 in sample space, and raises R from 0.941 to 0.957 in parameter space.

Magnitude. For each noise level, we measure

$$e(\sigma) = 1 - \frac{\mathrm{RMS}(s_B - s_T)}{\mathrm{RMS}(s_F - s_T)}, \qquad \gamma(\sigma) = \frac{\|g_\sigma^B\|_2}{\|g_\sigma^T\|_2}, \quad (7)$$

where F, B, and T denote the fake, bridge, and teacher scores. The score-error reduction $e(\sigma)$ measures how far the bridge has moved from the student toward the teacher at noise level σ, while the gradient-norm ratio $\gamma(\sigma)$ measures the resulting update strength relative to direct teacher matching. Their profiles closely follow each other across noise levels (Pearson $r = 0.964$; Figure 3d): noise levels that adapt further toward the teacher also induce stronger student updates. Together with the directional changes in Figures 3b and 3c, this shows that finite bridge adaptation reshapes each per-level update before aggregation, changing both its direction and its relative strength and thereby reducing cross-noise cancellation.

6 Experiments

Experiments are conducted on different models across various datasets. We first compare the one-step generation performance of SBD with other distillation methods (Section 6.1). Secondly, we analyze how SBD changes the cross-noise gradient conflict and the convergence during training (Section 6.2). Finally, we conduct ablation studies on the design choices of SBD (Section 6.3). The experiment setup is provided in Appendix B.3.

6.1 One-step generation

SBD achieves better one-step generation performance than existing distribution matching distillation methods. Experiments are conducted on (1) text-to-image generation with SDXL (Podell et al., 2024),

Table 1: **Text-to-image generation with SDXL.** Evaluated on 10K COCO-2014 prompts. Among one-step generators, the best result is in **bold** and the second best is underlined; "–" denotes a metric not reported. *Computed with 100 NFE. †Generates 512×512 images.

		Distribution ↓		Alignment & preference ↑			
Model	NFE	FID	Patch-FID	CLIP	PickScore	HPSv2	MPS
SDXL (Podell et al., 2024)	50	19.36*	21.38*	35.03	22.25	27.37	10.70
SDXL-Turbo† (Sauer et al., 2025)	1	24.57	**23.94**	35.01	22.15	27.10	10.43
SDXL-Lightning (Lin et al., 2024)	1	23.92	31.65	33.50	21.92	27.16	10.23
DMD2 (Yin et al., 2024a)	1	19.01	26.98	35.22	22.10	27.45	10.69
DisBack (Zhang et al., 2025)	1	18.96	26.89	–	–	–	–
DMDX (Lu et al., 2025)	1	–	–	35.26	**22.27**	27.70	**11.20**
SBD (ours)	1	**18.52**	26.86	**35.43**	22.17	**27.86**	10.97

Table 2: **Unconditional generation.** FID ↓. Among one-step generators, the best result is in **bold** and the second best is underlined.

Model	NFE	FFHQ	AFHQv2	LSUN-Bedroom	LSUN-Cat
EDM (Karras et al., 2022)	79	2.39	1.96	3.57	6.69
CD (Song et al., 2023)	1	12.58	10.75	7.80	11.00
DMD	1	11.23	9.78	7.69	10.76
DisBack (Zhang et al., 2025)	1	10.88	9.97	6.99	10.30
SBD (ours)	1	**6.14**	**7.07**	**6.56**	**9.60**

(2) unconditional generation on FFHQ 64×64 (Karras et al., 2019), AFHQv2 64×64 (Choi et al., 2020; Karras et al., 2021), LSUN-Bedroom 256×256, and LSUN-Cat 256×256 (Yu et al., 2015), and (3) class-conditional generation on ImageNet 64×64 (Deng et al., 2009). All SBD models follow the training settings of DisBack (Zhang et al., 2025).

Text-to-image generation. We use SBD to distill SDXL and evaluate FID (Heusel et al., 2017) and Patch-FID (Chai et al., 2022; Lin et al., 2024) on COCO 2014 (Lin et al., 2014), together with CLIP score (Radford et al., 2021), PickScore (Kirstain et al., 2023), HPSv2 (Wu et al., 2023), and MPS (Zhang et al., 2024). The results are presented in Table 1. Among one-step generators, SBD achieves the best FID, CLIP score, and HPSv2. Some generation samples are shown in Appendix D.

Unconditional generation. The results are shown in Table 2. SBD achieves the lowest FID on all four datasets. Compared with DMD, it reduces the FID from 11.23 to 6.14 on FFHQ and from 9.78 to 7.07 on AFHQv2.

Class-conditional generation. SBD achieves an FID of 1.20 on class-conditional ImageNet 64×64 with a single network evaluation, outperforming DMD2 and DisBack (Table 3). The ImageNet model is distilled from the pre-trained EDM model (Karras et al., 2022).

6.2 Gradient conflict and convergence

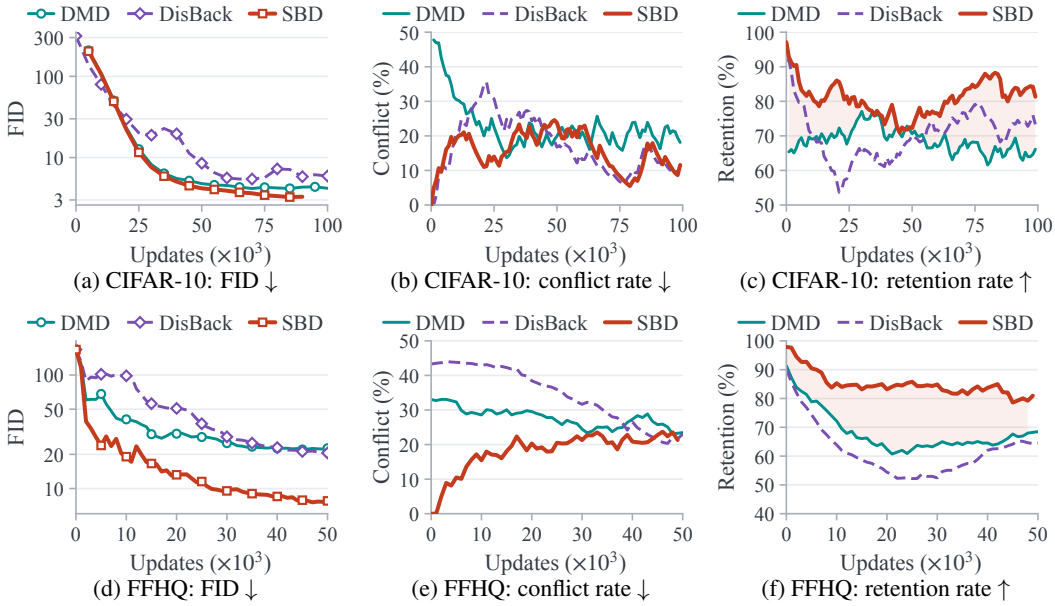
Figure 4 tracks FID and the gradient statistics during training on CIFAR-10 (a–c) and FFHQ 64×64 (d–f). On CIFAR-10, SBD and DMD follow similar FID trajectories during the first 20K updates, after which SBD stays consistently lower: at 90K updates, SBD reaches FID 3.29, compared with 4.32 for DMD, and DisBack ends at 5.92 after 100K updates. The gradient statistics account for this gap. Averaged over the last 40K updates, SBD reduces the cross-noise conflict rate from 20.0% under DMD to 10.1% and raises gradient retention from 65.6% to 84.1%, so a larger share of every update moves the student toward the teacher. DisBack reaches a comparable late-stage conflict rate (10.7%), but between 20K and 40K updates its conflict rate peaks above those of the other methods, its retention falls to its minimum, and its FID stalls near 20, consistent with a precomputed path that is mismatched with the evolving student. The FID improvement on CIFAR-10 persists with a larger batch of 2,048, from 2.13 to 1.67 at 70K iterations. On FFHQ, the same pattern holds: SBD has the

Table 3: **Class-conditional ImageNet** 64×64**.** Gray rows list other one-step methods for reference. Among the remaining one-step methods, the best result is in **bold** and the second best is underlined.

Model	NFE	FID ↓
EDM (teacher) (Karras et al., 2022)	511	1.36
CTM (Kim et al., 2024)	1	1.92
SiD (Zhou et al., 2024)	1	1.71
SiDA (Zhou et al., 2025c)	1	1.35
Uni-Instruct (Wang et al., 2025b)	1	1.28
DMD2 (Yin et al., 2024a)	1	1.51
DisBack (Zhang et al., 2025)	1	1.38
DMD2, longer training (Yin et al., 2024a)	1	1.28
SBD (ours)	1	**1.20**

Table 4: **Ablation study.** FID ↓; each row changes one factor from the default SBD.

Variant	FFHQ	CIFAR-10
DMD (teacher target)	11.23	4.32
Initialization		
Copy of the teacher	52.74	3.92
Rebuilding		
Built once ($M = \infty$)	20.38	3.96
Adaptation objective		
Real-data term only	10.55	3.36
Teacher term only	7.57	3.54
Controls		
DMD, norm-matched	12.35	4.88
DMD, compute-matched	11.03	4.27
SBD (default)	6.14	3.29



(a) CIFAR-10: FID ↓ (b) CIFAR-10: conflict rate ↓ (c) CIFAR-10: retention rate ↑

(d) FFHQ: FID ↓ (e) FFHQ: conflict rate ↓ (f) FFHQ: retention rate ↑

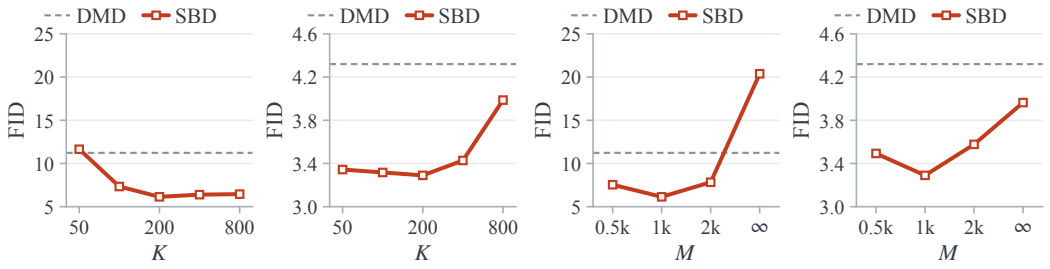
Figure 4: **Training dynamics of DMD, DisBack, and SBD on CIFAR-10 (a–c) and FFHQ** 64×64 **(d–f).** The horizontal axis counts generator and fake-network updates on CIFAR-10 and generator updates on FFHQ; FID uses a log scale, and the retention rate follows Eq. 6.

lowest FID at every evaluated checkpoint, keeps gradient retention near 80% throughout training (81.5% on average, compared with 63.6% for DMD and 56.7% for DisBack), and has the lowest average conflict rate (22.4%, compared with 25.9% and 30.5%).

6.3 Ablation Study

We ablate SBD on FFHQ 64×64 and CIFAR-10 in the settings of Sections 6.1 and 6.2, changing one factor at a time from the default configuration: the bridge is copied from the current fake score, adapted for $K = 200$ steps with $\beta = 1$, and rebuilt every $M = 1000$ iterations. Table 4 reports FID, and Figure 5 varies K and M.

Where the bridge starts. Initializing the bridge from the teacher instead of the current fake score removes its dependence on the student and turns it into a periodically fine-tuned teacher. This is the most damaging change: FID rises from 6.14 to 52.74 on FFHQ and from 3.29 to 3.92 on



(a) K, FFHQ-64 (b) K, CIFAR-10 (c) M, FFHQ-64 (d) M, CIFAR-10

Figure 5: **Adaptation horizon and rebuilding interval.** FID of SBD when varying (a,b) the adaptation horizon K with $M = 1000$ and (c,d) the rebuilding interval M with $K = 200$; $M = \infty$ builds the bridge once at the start of training. The dashed line is DMD.

CIFAR-10 (Table 4). These results show that retaining dependence on the current student is critical for constructing an effective bridge.

Rebuilding the bridge. A bridge built once and kept fixed ($M = \infty$) is conditioned on the initial student only, as the DisBack path is, while it is still constructed by the adaptation procedure of SBD. Figure 5c,d further varies the rebuilding interval. Building the bridge once raises FID to 20.38 on FFHQ and 3.96 on CIFAR-10, so a target fixed at the start of training gives up the gain; on FFHQ it is even worse than DMD (11.23). Rebuilding every $M = 1000$ iterations is best on both datasets: rebuilding more frequently ($M = 500$: 7.54 and 3.49) gives the student too few updates to sufficiently exploit each bridge, whereas rebuilding less frequently ($M = 2000$: 7.83 and 3.58) causes the bridge to lag behind the evolving student.

Adaptation objective. Adapting with real-data denoising alone ($\beta = 0$) removes the teacher guidance, and adapting with teacher alignment alone removes the data anchor. Both terms are needed. Without teacher alignment, FID rises to 10.55 on FFHQ and 3.36 on CIFAR-10; without real-data denoising, it rises to 7.57 and 3.54.

Adaptation horizon. At $K = 0$ the bridge equals the fake score and the update in Eq. (4) vanishes, while a large K moves the bridge away from its student initialization toward a fine-tuned teacher. FFHQ and CIFAR-10 share the optimum at $K = 200$ (Figure 5a,b). A short horizon leaves the bridge close to the fake score and weakens the update ($K = 50$: 11.65 on FFHQ, 3.34 on CIFAR-10), while a long horizon moves the bridge closer to the teacher and degrades performance ($K = 800$: 6.45 and 3.99), consistent with losing the benefit of an intermediate, student-adaptive target.

Step size and compute. Bridge matching can change the norm of the teacher-induced update (Section 5), so its benefit could in principle come from a different effective step size. A norm-matched DMD keeps the teacher as the target but rescales the teacher field at each noise level to the norm of the bridge field on the same samples. It reaches an FID of 12.35 on FFHQ and 4.88 on CIFAR-10, no better than DMD, so the gain of SBD cannot be explained by update magnitude alone. A compute-matched DMD trains for additional iterations so that its GPU time equals that of SBD, whose adaptation adds K bridge updates every M iterations. The extra training changes the FID of DMD only slightly, to 11.03 on FFHQ and 4.27 on CIFAR-10, which remains far from the 6.14 and 3.29 of SBD. The improvement therefore does not come from the additional computation either.

7 Conclusion

This paper introduces Score Bridge Distillation (SBD), which repeatedly reconstructs the matching target from the current student and adapts it toward the teacher. By making the immediate target student-adaptive, SBD reshapes the updates induced at different noise levels before aggregation, reducing cross-noise gradient conflicts and improving optimization. Across diverse one-step generation settings, including high-resolution text-to-image generation, SBD achieves faster convergence and better generation quality than existing distribution matching distillation methods, without changing the generator or its sampling procedure.

AI USE STATEMENT

In this work, we used generative AI tools to assist with the mathematical derivations in Section 5. We have not used generative AI tools to generate data, implement the method, or interpret the results, and the remaining tasks with required disclosure are not applicable to this work. Additionally, we used generative AI tools to polish the writing and to search for related work. We have reviewed all AI-assisted work: the authors drafted all sections, checked every derivation, and take full responsibility for the content of this paper, including text and claims produced with the aid of generative AI.

REFERENCES

Lichen Bai, Zikai Zhou, Shitong Shao, Wenliang Zhong, Shuo Yang, Shuo Chen, Bojun Cheng, and Zeke Xie. Optimizing few-step generation with adaptive matching distillation. In *International Conference on Machine Learning (ICML)*, 2026.

Lucy Chai, Michaël Gharbi, Eli Shechtman, Phillip Isola, and Richard Zhang. Any-resolution training for high-resolution image synthesis. In *European Conference on Computer Vision (ECCV)*, Lecture Notes in Computer Science, pp. 170–188. Springer, 2022. doi: 10.1007/978-3-031-19787-1_10.

Junsong Chen, Shuchen Xue, Yuyang Zhao, Jincheng Yu, Sayak Paul, Junyu Chen, Han Cai, Song Han, and Enze Xie. SANA-Sprint: One-step diffusion with continuous-time consistency distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025.

Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. StarGAN v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8188–8197, 2020.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.

Xiangyu Fan, Zesong Qiu, Zhuguanyu Wu, Fanzhou Wang, Zhiqian Lin, Tianxiang Ren, Dahua Lin, Ruihao Gong, and Lei Yang. Phased dmd: Few-step distribution matching distillation via score matching within subintervals. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

Kevin Frans, Danijar Hafner, Sergey Levine, and Pieter Abbeel. One step diffusion via shortcut models. In *International Conference on Learning Representations*, 2025.

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. In *International Conference on Learning Representations*, 2026. URL https://arxiv.org/abs/2506.00523.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, J. Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. In *Advances in Neural Information Processing Systems*, 2025a. URL https://arxiv.org/abs/2505.13447.

Zhengyang Geng, Ashwini Pokle, William Luo, Justin Lin, and J. Zico Kolter. Consistency models made easy. In *International Conference on Learning Representations*, 2025b.

Hyojun Go, JinYoung Kim, Yunsung Lee, Seunghyun Lee, Shinhyeok Oh, Hyeongdon Moon, and Seungtaek Choi. Addressing negative transfer in diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, pp. 27199–27222, 2023.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 27, 2014.

Tiankai Hang, Shuyang Gu, Chen Li, Jianmin Bao, Dong Chen, Han Hu, Xin Geng, and Baining Guo. Efficient diffusion training via Min-SNR weighting strategy. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7441–7451, 2023.

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. In *Advances in Neural Information Processing Systems*, 2017. URL https://arxiv.org/abs/1706.08500.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

Tejas Jayashankar, Jongha Jon Ryu, and Gregory W. Wornell. Score-of-mixture training: One-step generative model training made simple via score estimation of mixture distributions. In *International Conference on Machine Learning (ICML)*, 2025.

Xiao Jin, Baoyun Peng, Yichao Wu, Yu Liu, Jiaheng Liu, Ding Liang, Junjie Yan, and Xiaolin Hu. Knowledge distillation via route constrained optimization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1345–1354, 2019.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4401–4410, 2019.

Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pp. 12104–12114, 2020.

Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 852–863, 2021.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, 2022. URL https://arxiv.org/abs/2206.00364.

Dongjun Kim, Chieh-Hsin Lai, Wei-Hsiang Liao, Naoki Murata, Yuhta Takida, Toshimitsu Uesaka, Yutong He, Yuki Mitsufuji, and Stefano Ermon. Consistency trajectory models: Learning probability flow ODE trajectory of diffusion. In *International Conference on Learning Representations*, 2024. URL https://arxiv.org/abs/2310.02279.

Youngjoong Kim, Deokyeong Lee, and Jaesik Park. Distribution matching distillation without fake score network. *arXiv preprint arXiv:2605.19256*, 2026. URL https://arxiv.org/abs/2605.19256.

Yuval Kirstain, Adam Polyak, Uriel Singer, Shahbuland Matiana, Joe Penna, and Omer Levy. Pick-a-pic: An open dataset of user preferences for text-to-image generation. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pp. 36652–36663, 2023.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

Shanchuan Lin, Anran Wang, and Xiao Yang. SDXL-Lightning: Progressive Adversarial Diffusion Distillation. *arXiv preprint arXiv:2402.13929*, 2024.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *European Conference on Computer Vision (ECCV)*, volume 8693 of *Lecture Notes in Computer Science*, pp. 740–755. Springer, 2014. doi: 10.1007/978-3-319-10602-1_48.

Dongyang Liu, Peng Gao, David Liu, Ruoyi Du, Zhen Li, Qilong Wu, Xin Jin, Sihan Cao, Shifeng Zhang, Hongsheng Li, and Steven Hoi. Decoupled DMD: CFG augmentation as the spear, distribution matching as the shield. *arXiv preprint arXiv:2511.22677*, 2025a. URL https://arxiv.org/abs/2511.22677.

Yunpeng Liu, Boxiao Liu, Yi Zhang, Xingzhong Hou, Guanglu Song, Yu Liu, and Haihang You. See further when clear: Curriculum consistency model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025b.

Yuxi Liu, Haoyu Li, Yixiang Cai, Tengxu Sun, Zekun Zhang, Baole Ai, Ang Wang, Jiamang Wang, Lin Qu, Kun Yuan, and Kai Zhang. CrossDistill: Balancing quality and diversity via trajectory-level hybrid few-step distillation. *arXiv preprint arXiv:2609.14725*, 2026. URL https://arxiv.org/abs/2609.14725.

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

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. In *Advances in Neural Information Processing Systems*, 2022. URL https://arxiv.org/abs/2206.00927.

Yanzuo Lu, Yuxi Ren, Xin Xia, Shanchuan Lin, Xing Wang, Xuefeng Xiao, Andy J. Ma, Xiaohua Xie, and Jian-Huang Lai. Adversarial distribution matching for diffusion distillation towards efficient image and video synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025.

Weijian Luo. Diff-instruct++: Training one-step text-to-image generator model to align with human preferences. *Transactions on Machine Learning Research*, 2025.

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. In *Advances in Neural Information Processing Systems*, 2023. URL https://arxiv.org/abs/2305.18455.

Weijian Luo, Zemin Huang, Zhengyang Geng, J. Zico Kolter, and Guo-jun Qi. One-step diffusion distillation through score implicit matching. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

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

Qianli Ma, Xuefei Ning, Dongrui Liu, Li Niu, and Linfeng Zhang. Decouple-then-merge: Finetune diffusion models as multi-task learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

Seyed-Iman Mirzadeh, Mehrdad Farajtabar, Ang Li, Nir Levine, Akihiro Matsukawa, and Hassan Ghasemzadeh. Improved knowledge distillation via teacher assistant. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020.

Thuan Hoang Nguyen and Anh Tran. SwiftBrush: One-step text-to-image diffusion model with variational score distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. URL https://arxiv.org/abs/2312.05239.

Abhishek Panigrahi, Bingbin Liu, Sadhika Malladi, Andrej Risteski, and Surbhi Goel. Progressive distillation induces an implicit curriculum. In *International Conference on Learning Representations (ICLR)*, 2025.

Byeongjun Park, Sangmin Woo, Hyojun Go, Jin-Young Kim, and Changick Kim. Denoising task routing for diffusion models. In *International Conference on Learning Representations (ICLR)*, 2024.

Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. SDXL: Improving latent diffusion models for high-resolution image synthesis. In *International Conference on Learning Representations*, 2024.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763, 2021.

Yuxi Ren, Xin Xia, Yanzuo Lu, Jiacheng Zhang, Jie Wu, Pan Xie, Xing Wang, and Xuefeng Xiao. Hyper-SD: Trajectory segmented consistency model for efficient image synthesis. In *Advances in Neural Information Processing Systems*, 2024.

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 (CVPR)*, pp. 10684–10695, 2022.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. In *International Conference on Learning Representations*, 2022. URL https://arxiv.org/abs/2202.00512.

Axel Sauer, Frederic Boesel, Tim Dockhorn, Andreas Blattmann, Patrick Esser, and Robin Rombach. Fast high-resolution image synthesis with latent adversarial diffusion distillation. *arXiv preprint arXiv:2403.12015*, 2024. URL https://arxiv.org/abs/2403.12015.

Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial Diffusion Distillation. In *European Conference on Computer Vision*, pp. 87–103. Springer, 2025.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a. URL https://arxiv.org/abs/2010.02502.

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*, 2021b.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *International Conference on Machine Learning*, 2023. URL https://arxiv.org/abs/2303.01469.

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

Leyang Wang, Mingtian Zhang, Zijing Ou, and David Barber. Vardiu: A variational diffusive upper bound for one-step diffusion distillation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025a.

Yifei Wang, Weimin Bai, Colin Zhang, Debing Zhang, Weijian Luo, and He Sun. Uni-Instruct: One-step diffusion model through unified diffusion divergence instruction. In *Advances in Neural Information Processing Systems*, 2025b. URL https://arxiv.org/abs/2505.20755.

Tianhe Wu, Ruibin Li, Lei Zhang, and Kede Ma. Diversity-preserved distribution matching distillation for fast visual synthesis. *arXiv preprint arXiv:2602.03139*, 2026. URL https://arxiv.org/abs/2602.03139.

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.

Sirui Xie, Zhisheng Xiao, Diederik P. Kingma, Tingbo Hou, Ying Nian Wu, Kevin Patrick Murphy, Tim Salimans, Ben Poole, and Ruiqi Gao. Em distillation for one-step diffusion models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Yanwu Xu, Yang Zhao, Zhisheng Xiao, and Tingbo Hou. UFOGen: You forward once large scale text-to-image generation via diffusion GANs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

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

Tianwei Yin, Michaël Gharbi, Taesung Park, Richard Zhang, Eli Shechtman, Frédo Durand, and William T. Freeman. Improved distribution matching distillation for fast image synthesis. In *Advances in Neural Information Processing Systems*, 2024a. URL https://arxiv.org/abs/2405.14867.

Tianwei Yin, Michaël Gharbi, Richard Zhang, Eli Shechtman, Frédo Durand, William T. Freeman, and Taesung Park. One-step diffusion with distribution matching distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024b. URL https://arxiv.org/abs/2311.18828.

Fisher Yu, Ari Seff, Yinda Zhang, Shuran Song, Thomas Funkhouser, and Jianxiong Xiao. LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*, 2015.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Shengyuan Zhang, Ling Yang, Zejian Li, An Zhao, Chenye Meng, Changyuan Yang, Guang Yang, Zhiyuan Yang, and Lingyun Sun. Distribution backtracking builds a faster convergence trajectory for diffusion distillation. In *International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2408.15991.

Sixian Zhang, Bohan Wang, Junqiang Wu, Yan Li, Tingting Gao, Di Zhang, and Zhongyuan Wang. Learning multi-dimensional human preference for text-to-image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8018–8027, 2024.

Kaiwen Zheng, Yuji Wang, Qianli Ma, Huayu Chen, Jintao Zhang, Yogesh Balaji, Jianfei Chen, Ming-Yu Liu, Jun Zhu, and Qinsheng Zhang. Large scale diffusion distillation via score-regularized continuous-time consistency. In *International Conference on Learning Representations (ICLR)*, 2026.

Linqi Zhou, Stefano Ermon, and Jiaming Song. Inductive moment matching. In *International Conference on Machine Learning*, pp. 78651–78686, 2025a.

Mingyuan Zhou, Huangjie Zheng, Zhendong Wang, Mingzhang Yin, and Hai Huang. Score identity distillation: Exponentially fast distillation of pretrained diffusion models for one-step generation. In *International Conference on Machine Learning*, 2024. URL https://arxiv.org/abs/2404.04057.

Mingyuan Zhou, Zhendong Wang, Huangjie Zheng, and Hai Huang. Guided score identity distillation for data-free one-step text-to-image generation. In *International Conference on Learning Representations (ICLR)*, 2025b.

Mingyuan Zhou, Huangjie Zheng, Yi Gu, Zhendong Wang, and Hai Huang. Adversarial score identity distillation: Rapidly surpassing the teacher in one step. In *International Conference on Learning Representations*, 2025c. URL https://arxiv.org/abs/2410.14919.