

# Flow Straighter and Faster: Efficient One-Step Generative Modeling via MeanFlow on Rectified Trajectories

Xinxi Zhang<sup>1\*</sup> Shiwei Tan<sup>1\*</sup> Quang Nguyen<sup>1</sup> Quan Dao<sup>1</sup> Ligong Han<sup>2</sup>

Xiaoxiao He<sup>1</sup> Tunyu Zhang<sup>1</sup> Alen Mrdovic<sup>1</sup> Dimitris Metaxas<sup>1</sup>

<sup>1</sup> Rutgers University <sup>2</sup> Red Hat AI Innovation

## Abstract

Flow-based generative models have recently demonstrated strong performance, yet sampling typically relies on expensive numerical integration of ordinary differential equations (ODEs). Rectified Flow enables one-step sampling by learning nearly straight probability paths, but achieving such straightness requires multiple computationally intensive reflow iterations. MeanFlow achieves one-step generation by directly modeling the average velocity over time; however, when trained on highly curved flows, it suffers from slow convergence and noisy supervision. To address these limitations, we propose **Rectified MeanFlow**, a framework that models the *mean velocity field* along the rectified trajectory using only a single reflow step. This eliminates the need for perfectly straightened trajectories while enabling efficient training. Furthermore, we introduce a simple yet effective truncation heuristic that aims to reduce residual curvature and further improve performance. Extensive experiments on ImageNet at  $64^2$ ,  $256^2$ , and  $512^2$  resolutions show that Re-MeanFlow consistently outperforms prior one-step flow distillation and Rectified Flow methods in both sample quality and training efficiency. Code is available at Code is available at <https://github.com/Xinxi-Zhang/Re-MeanFlow>.

## 1 Introduction

Flow models [27, 28] and diffusion models [43, 40] have become a central paradigm in generative modeling, enabling a wide range of applications across various data domains [18, 36, 55, 13, 14, 48]. Compared with earlier paradigms such as GANs [12, 20] and Normalizing Flows [35, 52], these models offer stable training and superior fidelity, but at the cost of expensive sampling. The root cause of this inefficiency is the curvature of the generative trajectories induced by the prior and data distributions. In practice, the velocity fields governing these flows bend sharply, making them difficult to approximate with few discretization steps.

Rectified Flow [28, 24] addresses the multi-step sampling cost of flow models by progressively *straightening* trajectories through an iterative reflow process. In principle, if the trajectory

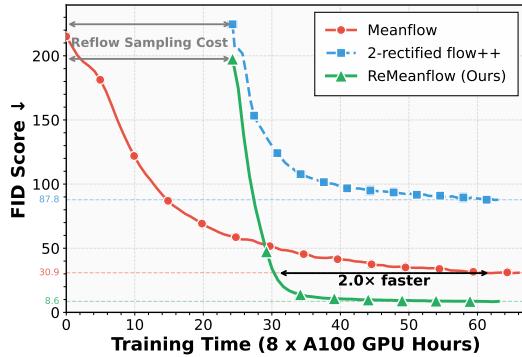


Figure 1: **Re-Meanflow (Ours)** leverages the synergy of trajectory rectification and mean-velocity modeling, achieving the best compute-quality trade-off—reaching strong FID. This synergy yields efficiency and quality that neither rectification nor mean-velocity modeling can achieve alone. All methods are initialized with pretrained EDM2-S [23].

\*Equal contribution

becomes perfectly straight, one-step sampling is achievable. However, Liu et al. [28] theoretically show that achieving a fully straight trajectory requires *infinitely many* reflow iterations. More recently, Lee et al. [24] argue that a single reflow step is sufficient in practice, under the assumption that the intermediate flow mapping is approximately  $L$ -Lipschitz and that edge cases leading to curved trajectories are exceedingly rare. Although this assumption may hold for most data-noise couplings, we find that in practice noticeable edge cases still exist; for these cases, performing Euler one-step sampling on such curved paths will generate invalid samples.

In contrast, MeanFlow [11] completely bypasses ODE integration by directly modeling the mean velocity field over time, eliminating the need for perfectly straight paths and achieving strong single-step performance. However, for highly curved trajectories, MeanFlow suffers from unstable and noisy supervision signals, making the training slow and computationally expensive.

To address these limitations, we propose **Re-Meanflow**, short for **R**ectified **M**ean**F**low, a conceptually simple yet training-effective framework that integrates the MeanFlow objective into the Rectified-flow framework. Specifically, we learn the *mean velocity field* along the rectified trajectory obtained after a *single* reflow step. This bypasses the need for perfectly straight trajectories while making MeanFlow training significantly more efficient and stable, as the rectified trajectory exhibits substantially lower variance than the original highly curved flow. Additionally, we introduce a simple and surprisingly effective *truncation* heuristic that reduces the influence of extreme curvature cases during training. Instead of attempting to explicitly detect curvature, we discard the top 10% couplings ranked by  $\ell_2$  trajectory distance. While this heuristic does not guarantee that all high-curvature trajectories are removed, we hypothesize that large distances often correlate with more curved paths, and applying this filter consistently improves both training stability and sample quality in practice.

Extensive experiments on ImageNet at  $64^2$ ,  $256^2$ , and  $512^2$ , demonstrating that Re-Meanflow **consistently outperforms** both state-of-the-art distillation methods and strong train-from-scratch baselines in terms of *both* generation quality and training efficiency. In particular, compared to its closely related counterpart, 2-rectified flow++ [24], Re-Meanflow reduces FID by **33.4%** while requiring only **10%** of the total compute. Beyond empirical gains, Re-Meanflow introduces a practical shift in how few-step generative models can be trained. Existing distillation pipelines depend heavily on high-end training GPUs, making hyperparameter tuning and repeated runs prohibitively expensive. In contrast, Re-Meanflow offloads the majority of computation to the inference-driven reflow stage, which can be executed on widely available consumer- or inference-grade accelerators, and requires only a lightweight MeanFlow training phase. The training stage of Re-Meanflow accounts for merely **17%** of the total GPU hours used by AYF[37].

## 2 Related Works

**Diffusion and Flow Matching.** Diffusion/flow-based generative models [40, 43, 17, 44, 27, 1, 28] learn to reverse a gradual noising process, where the reverse-time dynamics can be formulated as a deterministic probability-flow ODE [44, 21]. Although these approaches achieve strong performance, they typically require multi-step numerical integration for sampling [41, 21, 30, 53] due to the high curvature of generative paths.

**Few-step Diffusion/Flow Models.** Rectified Flow methods [28, 46, 24] alleviate this by explicitly learning straighter trajectories that enable one-step or few-step sampling, but such straight-path construction itself incurs high computational cost due to heavy training and repeated reflow procedures. Consistency Models [45, 42, 10, 29, 49] train networks to produce invariant outputs across different timesteps, enabling direct few-step or even one-step sampling. Flow Map Models [5, 9, 11] bypass ODE integration by learning the displacement or velocity map over time. Despite their strong empirical results, these few-step paradigms often face stability challenges or high training cost because they must learn mappings along inherently *curved* trajectories, where supervision is noisy, and optimization is difficult. Recent work has sought to mitigate this issue by simplifying the consistency objective [29] or introducing improved loss functions and normalization strategies[6]. Concurrent work CMT [19] aims to improve the efficiency of few-step models by supervising on learned ODE trajectories. Building on these insights, this work aims to combine trajectory simplification with one-step modeling.

**Efficient Training for Diffusion/Flow Models.** Many recent works [51, 26, 50, 47] boost the training efficiency of diffusion models by leveraging external representation learners like DINO

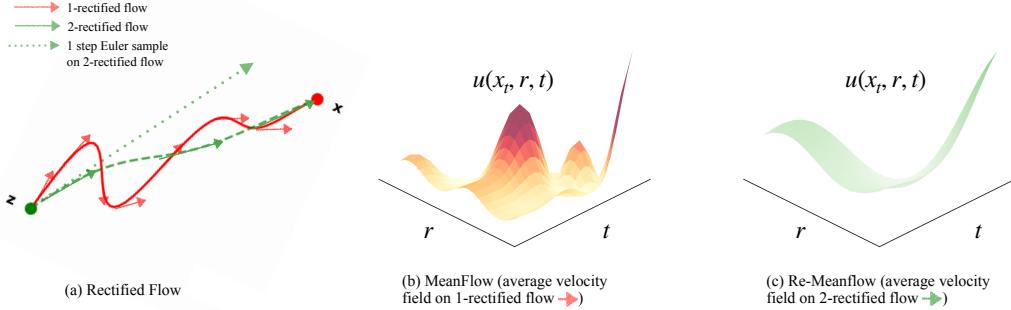


Figure 2: **Why trajectory rectification and mean-velocity modeling reinforce each other.** (a) A 1-rectified flow ( $\rightarrow$ ) still follows highly curved trajectories, requiring many ODE steps. Applying two rounds of rectification ( $\rightarrow$ ) straightens the paths and reduces NFEs, but one-step sampling ( $\cdot \cdot \rightarrow$ ) remains unreliable unless trajectories are nearly straight. (b) MeanFlow estimates the average velocity  $u(\mathbf{z}_t, r, t)$  over all intervals  $(r, t)$ , which in principle avoids the need for straight paths. However, when the underlying velocity field is highly curved, the induced averages become complex and hard to learn. (c) Training MeanFlow on trajectories from a 2-rectified flow yields a significantly smoother average-velocity field, making estimation easier and enabling faster convergence and high-quality one-step generation.

[33]. Orthogonal works [56, 38] adopt masked-transformer designs that exploit spatial redundancy by training on only a subset of tokens. ECT [10] enables efficient consistency models through a progressive training strategy that transitions from diffusion to consistency training. Re-Meanflow contributes in this direction by training a MeanFlow model on a significantly simplified flow path.

### 3 Preliminaries

**Rectified Flow[28].** Given a prior distribution  $p_{\mathbf{z}}$  (taken as  $\mathcal{N}(\mathbf{0}, \mathbf{I})$  in this work) and a data distribution  $p_{\mathbf{x}}$ , Rectified Flow trains a flow model  $v_{\theta}$  that transports  $p_{\mathbf{z}}$  to  $p_{\mathbf{x}}$  via an ordinary differential equation (ODE) over time  $t \in [0, 1]$ . For a data-noise coupling  $(\mathbf{x}, \mathbf{z}) \sim p_{\mathbf{x}\mathbf{z}}$ , where  $p_{\mathbf{x}\mathbf{z}}$  denotes the joint distribution of  $\mathbf{x}$  and  $\mathbf{z}$ , the flow path is defined as a linear interpolation  $\mathbf{z}_t = (1 - t)\mathbf{x} + t\mathbf{z}$ . The flow model  $v_{\theta}$  is trained by regressing the conditional velocity of observed couplings:

$$\mathcal{L}_{\text{RE}}(\theta) = \mathbb{E}_{\mathbf{x}, \mathbf{z} \sim p_{\mathbf{x}\mathbf{z}}, t \sim p_t} \left[ \left\| \frac{d}{dt} \mathbf{z}_t - v_{\theta}(\mathbf{z}_t, t) \right\|_2^2 \right] \quad \text{with } \frac{d}{dt} \mathbf{z}_t = \mathbf{z} - \mathbf{x} \quad (1)$$

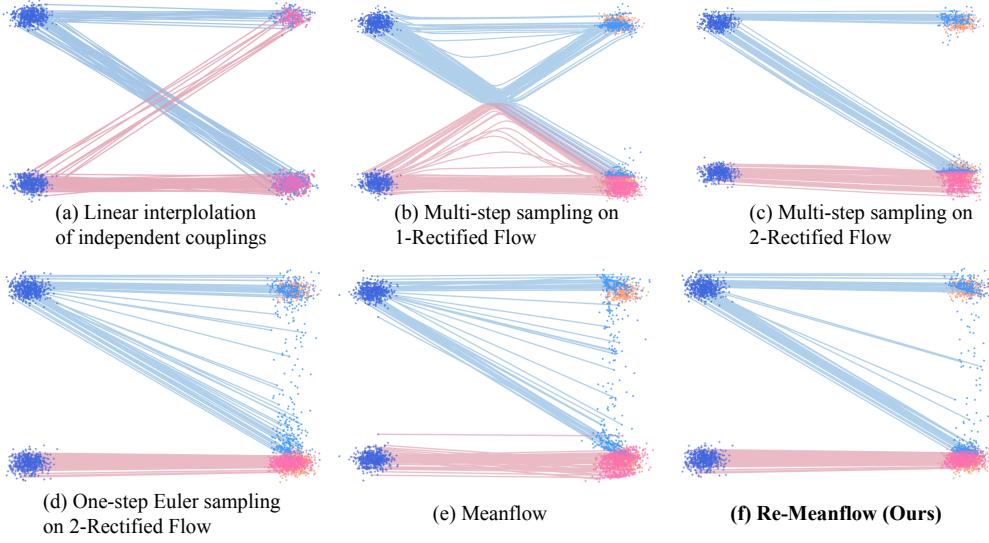
Once  $v_{\theta}$  is learned, a new sample  $\mathbf{x}$  can be generated by solving the ODE for  $\mathbf{z} \sim p_{\mathbf{z}}$ :

$$\mathbf{x}_{\theta}(\mathbf{z}) = \mathbf{z}_0 = \mathbf{z} - \int_0^1 v_{\theta}(\mathbf{z}_{\tau}, \tau) d\tau. \quad (2)$$

Rectified Flow begins training with independently sampled  $\mathbf{x}$  and  $\mathbf{z}$ , i.e.,  $p_{\mathbf{x}\mathbf{z}}^0(\mathbf{x}, \mathbf{z}) = p_{\mathbf{x}}(\mathbf{x})p_{\mathbf{z}}(\mathbf{z})$ . Training on these independent couplings produces highly curved ODE trajectories, requiring a large number of function evaluations (NFEs) to accurately solve Eq. 2. To mitigate this, Rectified Flow adopts a recursive reflow process to iteratively straighten the ODE trajectories. Given a learned velocity field  $v_{\theta}^k$  trained on  $p_{\mathbf{x}\mathbf{z}}^{k-1}$ , a straighter velocity field  $v_{\theta}^{k+1}$  can be obtained by training on  $p_{\mathbf{x}\mathbf{z}}^k$ , which is sampled by the previous velocity field  $v_{\theta}^k$ . For clarity, we refer to the trained vector field  $v_{\theta}^{k+1}$  as the  $(k+1)$ -rectified flow.

**MeanFlow [11].** Compared to Rectified Flow [28], which models the instantaneous velocity and requires solving the numerical ODE in Eq. 2 during sampling, MeanFlow enables one-step sampling by modeling the average velocity  $u$  between two time points  $t$  and  $r$ :

$$u(\mathbf{z}_t, r, t) \triangleq \frac{1}{t - r} \int_r^t v(\mathbf{z}_{\tau}, \tau) d\tau. \quad (3)$$



**Figure 3: A Class-imbalanced 2D Toy Example.** We consider a controlled 2D setup where a flow model transports a balanced two-component Gaussian mixture on the left to an imbalanced mixture on the right (weights 0.4 for the upper mode and 0.6 for the lower mode). **Panels (a-c):** (a) Linear interpolation of independently sampled couplings  $p_{\mathbf{x}} \times p_{\mathbf{z}}$ , which serve as the training signal for the first velocity model. (b) The resulting 1-rectified flow learned from these independent couplings; the learned velocity field remains noticeably curved. (c) Using the velocity field from (b), we generate a new set of couplings and train a second velocity model on their linear interpolations, yielding the 2-rectified flow. **Panel (d):** Due to imperfect straightening, one-step Euler sampling on the 2-rectified flow still yields noticeable outliers. **Panel (e):** MeanFlow trained directly on independent couplings fails to converge within the training budget because high-variance conditional velocities destabilize learning. **Panel (f):** Re-Meanflow combines trajectory rectification with MeanFlow, eliminating most outliers and achieving more accurate one-step generation.

To train a neural network  $u_{\theta}(\mathbf{x}_t, r, t)$  to approximate this average velocity, MeanFlow derives an implicit training target linking the average velocity, the instantaneous velocity  $v(\mathbf{x}_t, t)$ , and the time derivative of  $u_{\theta}$ :

$$\mathcal{L}_{MF}(\theta) = \mathbb{E}_{\mathbf{x}, \mathbf{z}, r, t} \|u_{\theta}(\mathbf{z}_t, r, t) - \text{sg}(u_{\text{tgt}})\|_2^2 \quad (4)$$

$$\text{with } u_{\text{tgt}} = v(\mathbf{z}_t, t) - (t - r) \frac{d}{dt} u_{\theta}(\mathbf{z}_t, r, t). \quad (5)$$

Here,  $\text{sg}(\cdot)$  denotes the stop-gradient operator, and the derivative  $\frac{d}{dt} u(\mathbf{z}_t, r, t)$  can be computed using the Jacobian-vector product (JVP) interface in frameworks such as PyTorch or JAX. once  $u_{\theta}$  is trained, the entire flow path can be reconstructed from a single evaluation  $u_{\theta}(\mathbf{z}, 0, 1)$  without numerical integration.

## 4 Methods

To motivate Re-Meanflow, we first examine the limitations of existing rectified-flow and MeanFlow approaches and illustrate why combining them yields a substantial synergy with a toy example.

**2D Toy Example Setting.** Figure 3 visualizes a simple transport task from a balanced two-mode Gaussian mixture to an imbalanced two-mode mixture. We compare one-step generation performance across three approaches under a fixed training budget of 20k iterations. We evaluate:

1. **2-rectified flow**, where each reflow iteration is trained for 10k steps;
2. **MeanFlow**, trained for 20k steps directly on independently sampled couplings;
3. **Re-Meanflow (ours)**, which first trains a velocity model for 10k steps to obtain a 1-rectified flow, followed by 10k steps of MeanFlow training on the resulting couplings.

More Detail of the toy example setting are provided in Appendix B.5.

## 4.1 Limitations of Rectified flow and MeanFlow

### Rectified Flow Leaves Residual Curvature Degrading One-Step Sampling.

Rectified Flow seeks to enable one-step sampling by progressively straightening ODE trajectories. However, Liu et al. [28] show that achieving a perfectly straight trajectory theoretically requires applying an infinite number of reflow iterations. Later, Lee et al. [24] argue that a single reflow step already yields trajectories that are “straight enough” for one-step Euler sampling, since couplings in  $p_{xz}^1$  that would induce curved paths are exceedingly rare under the assumption that the 1-rectified flow is  $L$ -Lipschitz:

$$\|\mathbf{x}' - \mathbf{x}''\|_2 \leq L \|\mathbf{z}' - \mathbf{z}''\|_2 \quad \text{where } (\mathbf{x}', \mathbf{z}'), (\mathbf{x}'', \mathbf{z}'') \sim p_{xz}^1. \quad (6)$$

While the  $L$ -Lipschitz condition guarantees that nearby noise samples cannot map to arbitrarily distant data points, it does not prevent  $L$  from being large in practice. When  $L$  is large, even modest differences in noise space can translate into substantial separation in data space, producing early-time bending in the transport paths and degrading the accuracy of one-step Euler sampling.

**Toy Evidence (Panel d).** Due to the effect of class imbalance, samples associated with the lighter target mode must traverse *disproportionately long distances* toward the denser component. Consequently, even when their noise samples are close, the corresponding data points can lie far apart, yielding large effective  $L$  in Eq. 6 and causing trajectories to bend near  $t \approx 0$ . The severe one-step deviations in Fig. 3(d) reflect precisely this early-time curvature. Similar geometric imbalances arise naturally in real datasets, underscoring the need for methods that remain stable and accurate even in the presence of residual curvature after rectification.

**MeanFlow Struggles on Highly Curved Independent Couplings.** Training MeanFlow directly on curved flows suffers from noisy and unstable supervision signals, which is a well-known cause of slow convergence in machine learning. Concretely, in the training objective of Eq. 4, the marginal velocity field is not directly available and must be approximated from conditional velocities. Because the couplings are sampled independently, these conditional velocities exhibit high variance. Moreover, the derivative term  $(t - r) \frac{d}{dt} u(\mathbf{z}_t, r, t)$  relies on Jacobian-vector products of the model output without explicit supervision, introducing additional noise that compounds the variance. Finally, when trajectories are highly curved, the corresponding average velocity field becomes irregular and difficult to approximate, as illustrated in Fig. 2b.

**Toy Evidence (Panel e).** MeanFlow trained directly on independent couplings struggles to learn a coherent transport map under the same training budget, resulting in poor overall performance.

## 4.2 Re-Meanflow: Efficient One-Step Generative Modeling via MeanFlow on Rectified Trajectories

Motivated by these observations, we propose **Re-Meanflow**, which models the mean velocity field along rectified trajectories, avoiding relying on perfectly straight ODE paths. Concretely, given a pretrained flow model  $v_\theta$  trained on the independent couplings  $p_{xz} = p_x p_z$ , we obtain a straighter coupling distribution  $p_{xz}^1$  by transporting samples using the 1-rectified flow  $v_\theta^1$ . We then train a MeanFlow model  $u_\theta$  on these rectified couplings under the MeanFlow objective (Eq. 5).

This combination substantially accelerates MeanFlow training compared to using independent couplings. Rectified trajectories contain far fewer path intersections, allowing conditional velocities to better approximate the marginal velocity field and also keeping instantaneous velocities closer to their temporal averages. As a result, the discrepancy between  $u(\mathbf{z}_t, r, t)$  and  $v(\mathbf{z}_t, t)$  in Eq. 5 is significantly reduced.

**Toy Evidence (Panel f).** Re-Meanflow (ours) achieves more accurate generation and yields many fewer invalid one-step samplings, demonstrating the complementary roles of the two components.

**Takeaway.** Trajectory rectification sufficiently reduces curvature to enable efficient MeanFlow training, while MeanFlow removes the need for fully straight trajectories. This synergistic advantage of Re-Meanflow is further evident on real ImageNet data in Fig. 1, where it outperforms using MeanFlow or Rectified Flow alone. A visual illustration of this synergy is provided in Fig. 2.

### 4.3 Distance Truncation

As discussed in Sec. 4.1, a small but consequential subset of couplings exhibits *large data-noise distances*, often arising from low-density regions where samples must travel substantially farther than typical cases. Such pairs tend to produce higher curvature even after rectification, contributing disproportionately to instability during training.

This observation motivates a simple filtering heuristic: after generating candidate couplings, we discard the top  $k\%$  (e.g., 10%) with the largest data-noise  $\ell_2$  distance. This **Distance Truncation** selectively removes the pairs most associated with excessive transport length, and in practice leads to noticeably straighter flows and more stable learning.

We also examine this behavior on real datasets. As shown in Fig. 4, couplings with large data-noise distances consistently populate the high-loss tail of the flow regression objective, suggesting that they coincide with high-curvature, high-variance regions. As shown in our empirical results in Sec. 5.2, filtering these pairs yields a significant performance gain in practice.

## 5 Experiments

**Datasets.** We conduct experiments on ImageNet [7] at  $64^2$  resolution in pixel space and at  $256^2$  and  $512^2$  resolutions in latent space [36].

**Settings.** We initialize Re-Meanflow from a pretrained flow or diffusion model. For ImageNet- $64^2$  and ImageNet- $512^2$ , we initialize Re-Meanflow from the pretrained EDM2-S model [23]. For ImageNet- $256^2$ , we initialize from the pretrained SiT-XL model [31]. During the reflow process, we generate 5M data-noise couplings using the default procedures described in the corresponding papers. During sampling for better synthetic image quality, we apply classifier-free guidance (CFG)[16] for ImageNet- $256^2$  and Autoguidance[22] for ImageNet- $64^2$  and  $512^2$ . More details on the implementation and the hyperparameter settings are provided in Appendix B.

**Baselines.** We compare Re-Meanflow against recent state-of-the-art one-step flow-based methods, selecting baselines with comparable architecture or computational cost to ensure fair comparisons.

### 5.1 One Step Generation Quality

For all experiments, we evaluate image quality using Fréchet Inception Distance (FID) [15]. When comparing with prior methods, we select those with comparable architectures and parameter counts. Across all evaluated resolutions, **Re-Meanflow consistently outperforms state-of-the-art one-step flow-based generation methods.**

On the EDM2-S backbone [23], Re-Meanflow achieves superior performance at both  $64^2$  and  $512^2$  resolutions. On ImageNet- $64^2$ , Re-Meanflow outperforms the closely related 2-rectified flow++ [24] by **33.4%** in FID, and slightly improves over recent state-of-the-art one-step baselines[29, 25, 37]. On ImageNet- $512^2$ , Re-Meanflow also delivers strong one-step image quality, achieving a **9%** FID gain over AYF[37] and outperforming strong consistency distillation methods[37, 19].

On the SiT backbone [31] for ImageNet- $256^2$ , Re-Meanflow slightly surpasses MeanFlow [11] despite being trained *without* real-image supervision and relying solely on synthesized reflow samples. This is noteworthy because training exclusively on self-generated data is known to degrade performance due to self-conditioning effects [2]. We attribute this improvement to the more favorable optimization landscape created by rectified mean-velocity learning: although the supervision is limited to synthetic couplings, rectification produces a smoother and lower-variance trajectory family, effectively narrowing the search space. As a result, the model converges more reliably toward a high-quality solution.

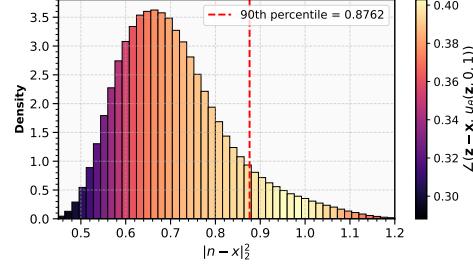


Figure 4: **Distance-Error Correlation.** Histogram of data-noise  $\ell_2$  distances on ImageNet  $512^2$ , colored by angular error (computed as  $\arccos(\frac{\langle \mathbf{z}-\mathbf{x}, u_\theta(\mathbf{z}, 0.1) \rangle}{\|\mathbf{z}-\mathbf{x}\|_2 \|u_\theta(\mathbf{z}, 0.1)\|_2})$ ), the angle between the predicted and ground-truth velocities). Errors are computed using a Mean-Flow model trained on *untruncated* 1-rectified couplings (config (d) in Tab. 3). A clear high-distance high-error tail (90th percentile marked) motivates our Distance Truncation.

Table 1: **Class-conditional generation on ImageNet.** All results use classifier-free guidance (CFG) when supported; “ $\times 2$ ” indicates that CFG doubles the effective NFE.  $\dagger$  denotes methods initialized from, or directly comparable to, the diffusion backbones also marked with  $\dagger$ . (iCT[42] result at  $256^2$  is reported by IMM[57], and ECT/ECD[10] results at  $512^2$  are reported by CMT[19]).

(a) ImageNet $64^2$			(b) ImageNet $256^2$			(c) ImageNet $512^2$		
Method	NFE	FID	Method	NFE	FID	Method	NFE	FID
<b>Diffusion models</b>								
EDM2-S [23] $\dagger$	$63 \times 2$	1.58	ADM [8]	$250 \times 2$	10.94	EDM2-S [23] $\dagger$	$63 \times 2$	2.23
+ Autoguidance [22]	$63 \times 2$	1.01	DiT-XL [34]	$250 \times 2$	2.27	+ Autoguidance [22]	$63 \times 2$	1.34
EDM2-XL [23]	$63 \times 2$	1.33	SIT-XL [31] $\dagger$	$250 \times 2$	2.06	EDM2-XXL [23]	$63 \times 2$	1.81
<b>Few-step models</b>								
2-rectified flow++ [24]	1	4.31	Re-Meanflow (ours) $\dagger$	1	3.41	Re-Meanflow (ours) $\dagger$	1	3.03
iCT [42]	1	4.02	<b>Few-step models</b>			<b>Few-step models</b>		
ECD-S [10] $\dagger$	1	3.30	iCT [42]	1	34.6	ECT [10] $\dagger$	1	9.98
sCD-S [29] $\dagger$	1	2.97	SM [9]	1	10.6	ECD [10] $\dagger$	1	8.47
TCM [25] $\dagger$	1	2.88	iSM [32]	1	5.27	CMT [19] $\dagger$	1	3.38
AYF [37] $\dagger$	1	2.98	IMM [57]	$1 \times 2$	7.77	SCT-S [29]	1	10.13
MeanFlow [11]			MeanFlow [11]	1	3.43	sCD-S [29] $\dagger$	1	3.07
Re-Meanflow (ours)	1	2.87	<b>Re-Meanflow (ours)<math>\dagger</math></b>			AYF [37] $\dagger$	1	3.32



Figure 5: **Qualitative results for Re-Meanflow (NFE=1) on ImageNet  $64^2$  (Left),  $256^2$  (Middle), and  $512^2$  (Right).**

## 5.2 Training Efficiency

As noted by Lee et al. [24], distillation methods involving reflow can remain computationally competitive, even though they require generating additional data-noise couplings.

### Synergistic Benefits of Re-Meanflow.

We compare Re-Meanflow against MeanFlow [11] and 2-rectified flow++ [24] on ImageNet-512 $^2$  without using CFG. All methods are initialized from EDM2-S [23] and follow the recommended training configurations. Re-Meanflow and 2-rectified flow++ are trained on the same set of 5M couplings generated during the reflow process, whereas MeanFlow is trained on independent couplings.

For fairness, we follow the training settings reported in the original papers with one exception: we *do not* include the LPIPS [54] loss for 2-rectified flow++. LPIPS operates in pixel space and cannot be used directly in the latent space setting of our experiments; moreover, LPIPS may leak perceptual information correlated with FID, potentially exaggerating improvements. Instead, we adopt the Huber loss, the strongest configuration reported for 2-rectified flow++ without LPIPS.

To ensure a fair comparison, we define the total training budget as the GPU hours required to (1) generate all 5M couplings and (2) train Re-Meanflow for 100k iterations. We allocate the *same*

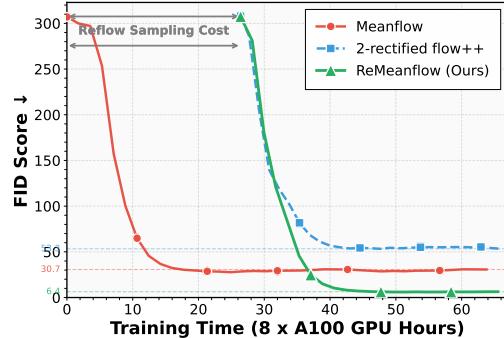


Figure 6: **Comparison of convergence on ImageNet-256 $^2$ .** Re-Meanflow achieves the best FID under the same training budget.

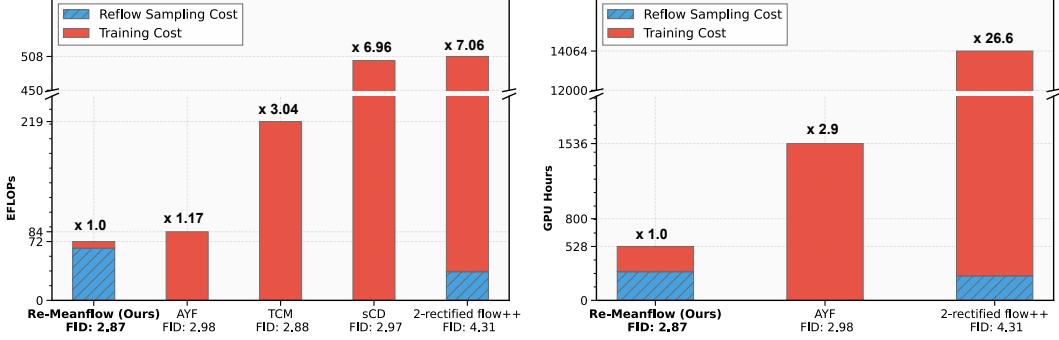


Figure 7: **Comparison of total computational cost on ImageNet-64<sup>2</sup> in EFLOPs (left) and GPU hours (right).** Each bar is decomposed into the training cost (solid red) and the reflow sampling cost (blue hatched). Re-Meanflow (ours) is used as the computational baseline. Numbers above the bars indicate a multiplicative factor relative to Re-Meanflow (1.0 $\times$ ).

total GPU-hour budget to MeanFlow and to 2-rectified flow++, and compare the resulting FID as a function of GPU hours. As shown in Fig. 1, Re-Meanflow achieves the best FID 8.6, reaching the final MeanFlow FID 2 $\times$  faster, while 2-rectified flow++ converges extremely slowly, achieving only 87.8 FID within the same budget. We repeat this experiment on ImageNet-256<sup>2</sup> and observe consistent results, as shown in Fig.6.

This demonstrates that combining MeanFlow with rectified flow is not merely an additive combination of two components: rectification reduces trajectory variance and stabilizes the MeanFlow objective, while mean-velocity learning enables efficient one-step sampling. The two components reinforce each other, yielding a synergistic effect that neither method achieves independently.

#### Overall Training Cost Comparison.

We further evaluate Re-Meanflow by estimating the total training cost in FLOPs and GPU hours on ImageNet-64<sup>2</sup>. The protocol for computing these metrics follows Lee et al. [24] and is detailed in the Appendix C.1. We compare against AYF [37], the strongest existing distillation baseline in both quality and efficiency, and against the closely related 2-rectified flow++ [24].

As shown in Fig. 7, Re-Meanflow achieves the *lowest overall compute cost* among recent distillation approaches. In terms of GPU hours, Re-Meanflow is **26.6 $\times$  faster** than 2-rectified flow++, reinforcing the results observed in our controlled efficiency experiments. Even compared to AYF, which does not require coupling generation, Re-Meanflow remains **2.9 $\times$  faster**.

We also compare estimated FLOPs. Although Re-Meanflow shows a smaller advantage under this metric, FLOPs alone do not fully reflect practical runtime. A substantial portion of our compute lies in the inference-only reflow stage, which can be executed on widely accessible inference-grade GPUs and runs significantly faster than training workloads with the same FLOP count. Training steps incur considerable overhead, such as data loading and batching, and rely on computationally expensive backward and JVP operations that typically require high-end training GPUs.

**Takeaway.** This highlights a key practical benefit of Re-Meanflow: by shifting most of the computational burden away from scarce high-end training GPUs and onto widely available inference-grade accelerators, Re-Meanflow substantially improves the accessibility and real-world feasibility of one-step flow distillation.

### 5.3 Ablation Study

**Distance Truncation.** We evaluate different truncation thresholds for Re-Meanflow, ranging from 5% to 15%, and report the resulting FID scores in Tab. 2. All truncation levels outperform the no-truncation baseline. The best result is achieved with a 10% threshold, which reduces FID from 3.5 to 3.03.

Table 2: Ablation on truncation strength on ImageNet-512<sup>2</sup>. Percentage improvement is computed relative to the no-truncation baseline.

Truncation Strength	FID $\downarrow$	Improvement (%)
No Truncation [8]	3.50	-
Top 5%	3.10	11.4%
Top 10%	3.03	13.4%
Top 15%	3.19	8.9%

As the threshold increases beyond 10%, performance begins to degrade (e.g., 3.19 at 15%). This pattern reflects the trade-off inherent in truncation: while discarding extremely long-distance pairs helps suppress the high-curvature tail, removing too many couplings reduces data coverage and weakens the supervision signal. Overall, the empirical trend supports our analysis in Sec. 4.3: moderate truncation effectively filters the most problematic cases while preserving sufficient coupling diversity for training.

#### Important Implementation Details.

To identify the components most critical for stable and high-quality training for Re-Meanflow, we conduct ablations on ImageNet-512<sup>2</sup> and showcase their impact on FID in Table 3.

- (a) *Hyperparameter adjustments.* Because rectified trajectories are significantly straighter, we reduce the normalization strength in the MeanFlow adaptive loss from 1.0 to 0.5 (similar to a Pseudo-Huber loss). We also find that sampling the classifier-free guidance scale from a uniform distribution improves generation quality.
- (b) Following Sabour et al. [37], Lu and Song [29], we replace the EDM2 time embedding  $\text{emb}(\log \sigma_t)$  with  $\text{emb}(t)$  to obtain more stable Jacobian–vector product computations. We perform a brief alignment stage that trains the new embedding to match the original outputs at the corresponding noise levels. Details of this alignment are provided in Appendix B.1.
- (c) Since rectified trajectories are smoother and require less mid-range emphasis, we adopt the U-shaped  $t$  distribution from Lee et al. [24] instead of using the standard lognormal schedule.
- (d) Re-Meanflow exhibits unusually high variance when  $t > 0.95$  and  $r < 0.4$ , a phenomenon we analyze in more detail in the Appendix. Inspired by the truncation strategy used in TCM [25], we mitigate this issue by avoiding this high-variance time region. This modification not only improves FID, but also accelerates convergence. More detail is presented in Appendix B.3.
- (e) Applying our trajectory-distance truncation heuristic (Sec. 4.3) further improves efficiency and generation quality. Additional analysis of distance–error correlations across resolutions, which motivates our truncation threshold, is provided in Appendix C.2.

## 6 Conclusion

We presented **Re-Meanflow**, a simple yet efficient framework that combines the complementary strengths of Rectified Flow [28] and MeanFlow [11]. By leveraging the synergy between trajectory straightening and mean-velocity learning, Re-Meanflow outperforms both components when used independently. Empirically, on ImageNet 64<sup>2</sup>, 256<sup>2</sup>, and 512<sup>2</sup>, Re-Meanflow delivers substantial improvements in both efficiency and generation quality, surpassing current state-of-the-art one-step flow-based models.

**Limitation.** Since our method does not access real data during training and relies entirely on synthetic samples generated by pretrained diffusion or flow models, its performance naturally depends on the quality of these generated couplings. A promising direction is to incorporate real data into the training process, as explored in Lee et al. [24], Seong et al. [39]. Another complementary line of work investigates how to improve generative models using only their own outputs [2, 3, 4], which suggests that self-generated data can still be leveraged effectively with appropriate regularization. Alternatively, one could improve the synthetic supervision directly by using stronger backbone models to generate the couplings.

**Future Work.** Looking ahead, the core idea behind Re-Meanflow is architecture-agnostic. We believe that leveraging straighter trajectories is a broadly applicable principle that may benefit other few-step generative paradigms. Re-Meanflow also offers practical advantages by shifting most of the training burden from scarce high-end training GPUs to widely available inference-grade accelerators capable of performing the reflow process. As a proof of concept, we view Re-Meanflow as a promising direction to scaling one-step paradigms to larger domains, such as text-to-image generation.

Table 3: Ablation study on ImageNet-512<sup>2</sup> for key implementation details of Re-Meanflow.

Training configurations	FID ↓
Base (Best Setting reported in [11])	7.81
(a) + Hyperparameter adjustments	7.22
(b) + Time embedding change	4.60
(c) + U-shaped $t$ distribution	3.71
(d) + Avoid high-variance ( $t, r$ ) region	3.50
(e) + Distance truncation	3.03

## References

- [1] Michael S Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. *arXiv preprint arXiv:2209.15571*, 2022.
- [2] Sina Alemohammad, Josue Casco-Rodriguez, Lorenzo Luzi, Ahmed Imtiaz Humayun, Hossein Babaei, Daniel LeJeune, Ali Siahkoohi, and Richard Baraniuk. Self-consuming generative models go mad. In *The Twelfth International Conference on Learning Representations*, 2023.
- [3] Sina Alemohammad, Ahmed Imtiaz Humayun, Shruti Agarwal, John Collomosse, and Richard Baraniuk. Self-improving diffusion models with synthetic data. *arXiv preprint arXiv:2408.16333*, 2024.
- [4] Sina Alemohammad, Zhangyang Wang, and Richard G Baraniuk. Neon: Negative extrapolation from self-training improves image generation. *arXiv preprint arXiv:2510.03597*, 2025.
- [5] Nicholas M Boffi, Michael S Albergo, and Eric Vanden-Eijnden. Flow map matching. *arXiv preprint arXiv:2406.07507*, 2, 2024.
- [6] Quan Dao, Khanh Doan, Di Liu, Trung Le, and Dimitris Metaxas. Improved training technique for latent consistency models. *arXiv preprint arXiv:2502.01441*, 2025.
- [7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [8] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34, 2021.
- [9] Kevin Frans, Danijar Hafner, Sergey Levine, and Pieter Abbeel. One step diffusion via shortcut models. *arXiv preprint arXiv:2410.12557*, 2024.
- [10] Zhengyang Geng, Ashwini Pokle, William Luo, Justin Lin, and J Zico Kolter. Consistency models made easy. *arXiv preprint arXiv:2406.14548*, 2024.
- [11] 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.
- [12] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.
- [13] Xiaoxiao He, Chaowei Tan, Ligong Han, Bo Liu, Leon Axel, Kang Li, and Dimitris N Metaxas. Dmcvr: Morphology-guided diffusion model for 3d cardiac volume reconstruction. In *International conference on medical image computing and computer-assisted intervention*, pages 132–142. Springer Nature Switzerland Cham, 2023.
- [14] Xiaoxiao He, Quan Dao, Ligong Han, Song Wen, Minhao Bai, Di Liu, Han Zhang, Martin Renqiang Min, Felix Juefei-Xu, Chaowei Tan, et al. Dice: Discrete inversion enabling controllable editing for multinomial diffusion and masked generative models. *arXiv preprint arXiv:2410.08207*, 2024.
- [15] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.
- [16] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.
- [17] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.
- [18] 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.

- [19] Zheyuan Hu, Chieh-Hsin Lai, Yuki Mitsufuji, and Stefano Ermon. Cmt: Mid-training for efficient learning of consistency, mean flow, and flow map models. *arXiv preprint arXiv:2509.24526*, 2025.
- [20] 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*, pages 4401–4410, 2019.
- [21] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35: 26565–26577, 2022.
- [22] Tero Karras, Miika Aittala, Tuomas Kynkänniemi, Jaakko Lehtinen, Timo Aila, and Samuli Laine. Guiding a diffusion model with a bad version of itself. *Advances in Neural Information Processing Systems*, 37:52996–53021, 2024.
- [23] Tero Karras, Miika Aittala, Jaakko Lehtinen, Janne Hellsten, Timo Aila, and Samuli Laine. Analyzing and improving the training dynamics of diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24174–24184, 2024.
- [24] Sangyun Lee, Zinan Lin, and Giulia Fanti. Improving the training of rectified flows. *Advances in neural information processing systems*, 37:63082–63109, 2024.
- [25] Sangyun Lee, Yilun Xu, Tomas Geffner, Giulia Fanti, Karsten Kreis, Arash Vahdat, and Weili Nie. Truncated consistency models. *arXiv preprint arXiv:2410.14895*, 2024.
- [26] Xingjian Leng, Jaskirat Singh, Yunzhong Hou, Zhenchang Xing, Saining Xie, and Liang Zheng. Repa-e: Unlocking vae for end-to-end tuning with latent diffusion transformers. *arXiv preprint arXiv:2504.10483*, 2025.
- [27] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.
- [28] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.
- [29] Cheng Lu and Yang Song. Simplifying, stabilizing and scaling continuous-time consistency models. *arXiv preprint arXiv:2410.11081*, 2024.
- [30] 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. *Advances in neural information processing systems*, 35:5775–5787, 2022.
- [31] Nanye Ma, Mark Goldstein, Michael S Albergo, Nicholas M Boffi, Eric Vanden-Eijnden, and Saining Xie. Sit: Exploring flow and diffusion-based generative models with scalable interpolant transformers. In *European Conference on Computer Vision*, pages 23–40. Springer, 2024.
- [32] Anh Nguyen, Viet Nguyen, Duc Vu, Trung Dao, Chi Tran, Toan Tran, and Anh Tran. Improved training technique for shortcut models. *arXiv preprint arXiv:2510.21250*, 2025.
- [33] Maxime Oquab, Timothée Darzet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.
- [34] William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4195–4205, 2023.
- [35] Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR, 2015.
- [36] 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*, pages 10684–10695, 2022.

- [37] Amirmojtaba Sabour, Sanja Fidler, and Karsten Kreis. Align your flow: Scaling continuous-time flow map distillation. *arXiv preprint arXiv:2506.14603*, 2025.
- [38] Vikash Sehwag, Xianghao Kong, Jingtao Li, Michael Spranger, and Lingjuan Lyu. Stretching each dollar: Diffusion training from scratch on a micro-budget. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 28596–28608, 2025.
- [39] Kim Shin Seong, Mingi Kwon, Jaeseok Jeong, and Youngjung Uh. Balanced conic rectified flow. *arXiv preprint arXiv:2510.25229*, 2025.
- [40] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. pmlr, 2015.
- [41] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.
- [42] Yang Song and Prafulla Dhariwal. Improved techniques for training consistency models. *arXiv preprint arXiv:2310.14189*, 2023.
- [43] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.
- [44] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.
- [45] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. 2023.
- [46] Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *arXiv preprint arXiv:2302.00482*, 2023.
- [47] Ge Wu, Shen Zhang, Ruijing Shi, Shanghua Gao, Zhenyuan Chen, Lei Wang, Zhaowei Chen, Hongcheng Gao, Yao Tang, Jian Yang, et al. Representation entanglement for generation: Training diffusion transformers is much easier than you think. *arXiv preprint arXiv:2507.01467*, 2025.
- [48] Qipan Xu, Zhenting Wang, Xiaoxiao He, Ligong Han, and Ruixiang Tang. Can large vision-language models detect images copyright infringement from genai? *arXiv preprint arXiv:2502.16618*, 2025.
- [49] Ling Yang, Zixiang Zhang, Zhilong Zhang, Xingchao Liu, Minkai Xu, Wentao Zhang, Chenlin Meng, Stefano Ermon, and Bin Cui. Consistency flow matching: Defining straight flows with velocity consistency. *arXiv preprint arXiv:2407.02398*, 2024.
- [50] Jingfeng Yao, Bin Yang, and Xinggang Wang. Reconstruction vs. generation: Taming optimization dilemma in latent diffusion models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 15703–15712, 2025.
- [51] Sihyun Yu, Sangkyung Kwak, Huiwon Jang, Jongheon Jeong, Jonathan Huang, Jinwoo Shin, and Saining Xie. Representation alignment for generation: Training diffusion transformers is easier than you think. *arXiv preprint arXiv:2410.06940*, 2024.
- [52] Shuangfei Zhai, Ruixiang Zhang, Preetum Nakkiran, David Berthelot, Jiatao Gu, Huangjie Zheng, Tianrong Chen, Miguel Angel Bautista, Navdeep Jaitly, and Josh Susskind. Normalizing flows are capable generative models. *arXiv preprint arXiv:2412.06329*, 2024.
- [53] Qinsheng Zhang and Yongxin Chen. Fast sampling of diffusion models with exponential integrator. *arXiv preprint arXiv:2204.13902*, 2022.
- [54] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

- [55] Xinxi Zhang, Song Wen, Ligong Han, Felix Juefei-Xu, Akash Srivastava, Junzhou Huang, Vladimir Pavlovic, Hao Wang, Molei Tao, and Dimitris Metaxas. Soda: Spectral orthogonal decomposition adaptation for diffusion models. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 4665–4682. IEEE, 2025.
- [56] Hongkai Zheng, Weili Nie, Arash Vahdat, and Anima Anandkumar. Fast training of diffusion models with masked transformers. *arXiv preprint arXiv:2306.09305*, 2023.
- [57] Linqi Zhou, Stefano Ermon, and Jiaming Song. Inductive moment matching. *arXiv preprint arXiv:2503.07565*, 2025.

## Appendix

<b>A More Discussion</b>	<b>1</b>
A.1 Broader Impact . . . . .	1
A.2 Additional Details for Sec. 4.1 . . . . .	2
A.3 Comparison with CMT [19] . . . . .	2
<b>B Implementation Details</b>	<b>2</b>
B.1 Pretrained Model Conditioning . . . . .	3
B.2 Time Distribution . . . . .	3
B.3 Avoiding High-Variance Time Regions . . . . .	4
B.4 Training with Guidance . . . . .	4
B.5 2D Toy Example Setting . . . . .	5
<b>C More Experiment Details and Results</b>	<b>5</b>
C.1 Computation Estimation of Each Method . . . . .	5
C.2 Distance-Error Correlation at Other Resolutions . . . . .	6
<b>D Algorithm for Re-Meanflow</b>	<b>7</b>
<b>E More Qualitative Results</b>	<b>8</b>
E.1 ImageNet-64 <sup>2</sup> . . . . .	8
E.2 ImageNet-256 <sup>2</sup> . . . . .	9
E.3 ImageNet-512 <sup>2</sup> . . . . .	10

## A More Discussion

In this section, we provide additional analysis and context for our method. We begin by discussing the broader impact of our approach. We then revisit the theoretical motivation for using a single reflow iteration and explain how our perspective complements existing analyses. Finally, we compare our method with the concurrent CMT method [19].

### A.1 Broader Impact

This work introduces a simple yet effective framework for efficient one-step generative modeling by training a MeanFlow model on rectified trajectories. Through extensive evaluations across multiple ImageNet settings, varying in resolution, latent representation, and model architecture, we demonstrate that Re-Meanflow achieves strong performance and high efficiency.

Beyond empirical gains, our method highlights a practical and socially relevant shift in how large generative models may be trained. Traditional few-step or one-step distillation pipelines often require expensive, high-end training hardware (e.g., A100-class GPUs), limiting accessibility to well-resourced institutions. In contrast, Re-Meanflow moves many of the computational burden to the inference reflow stage, followed by a lightweight MeanFlow training phase. Because inference workloads can be executed efficiently on widely available consumer- or inference-grade accelerators, our framework reduces reliance on scarce training GPUs and lowers the barrier to experimentation.

## A.2 Additional Details for Sec. 4.1

Lee et al. [24] analyze when multiple Reflow iterations are truly necessary for achieving straight trajectories in rectified flows. Their argument centers on how trajectory *intersections* affect the learned velocity field.

Consider two 1-rectified couplings  $(\mathbf{x}', \mathbf{z}')$  and  $(\mathbf{x}'', \mathbf{z}'')$ . A trajectory intersection occurs if there exists  $t \in [0, 1]$  such that

$$(1 - t)\mathbf{x}' + t\mathbf{z}' = (1 - t)\mathbf{x}'' + t\mathbf{z}''.$$

If such an intersection happens, both trajectories pass through the same intermediate point. Because rectified-flow training regresses the conditional expectation  $E[\mathbf{x}|\mathbf{x}_t]$ , the model must assign a *single* velocity to this shared point. As a result, the learned velocity field cannot simultaneously point toward both  $\mathbf{x}'$  and  $\mathbf{x}''$ , and it instead averages their directions. This averaging effect bends the local velocity field, producing curvature and consequently degrading the accuracy of one-step Euler sampling, which assumes the path to be straight.

To understand how often this phenomenon can happen, Lee et al. [24] show that an intersection implies

$$\mathbf{z}'' = \mathbf{z}' + \frac{1-t}{t}(\mathbf{x}' - \mathbf{x}'').$$

Under typical training, nearly all noise samples used to form 1-rectified couplings lie in the high-density region of the Gaussian prior. The  $\mathbf{z}''$  required above usually lies far outside that region unless  $\|\mathbf{x}' - \mathbf{x}''\|_2$  is extremely small or  $t$  is very close to 1. Therefore, intersections are statistically rare.

Further assuming that the 1-rectified flow is approximately  $L$ -Lipschitz,

$$\|\mathbf{x}' - \mathbf{x}''\|_2 \leq L \|\mathbf{z}' - \mathbf{z}''\|_2,$$

nearby noise samples cannot map to widely separated data points. Combining the rarity of intersections with this Lipschitz condition, the authors conclude that the optimal 2-rectified flow is nearly straight. Hence, in their view, one additional Reflow step is sufficient, and any remaining performance gap should be attributed primarily to training inefficiency rather than insufficient straightening.

**Relation to Our Work.** Our focus is complementary to the above perspective. While their analysis shows that intersections are rare for most couplings, our experiments reveal that in realistic settings, a small but influential subset of couplings can still exhibit noticeable curvature, particularly when the effective Lipschitz constant  $L$  becomes large due to geometric imbalance in the data distribution. Our method is designed to improve robustness in these challenging cases, providing stable one-step generation even when residual curvature persists after a single reflow step.

## A.3 Comparison with CMT [19]

CMT [19] is an important concurrent effort that also leverages synthetic trajectories generated by a pretrained sampler to stabilize few-step flow-map training. Conceptually, CMT introduces a dedicated mid-training stage that learns a full trajectory-to-endpoint mapping from solver-generated paths, which then serves as a trajectory-aligned initialization for a subsequent post-training flow-map stage. In contrast, our method adopts a fundamentally different design: rather than supervising on entire solver trajectories, we distill only the end-point couplings of rectified flows and learn the corresponding mean velocity in a single training stage. This distinction yields a practical advantage—our pipeline avoids the compounded complexity of CMT’s two-stage optimization, which is sensitive to hyperparameters at both stages and substantially more expensive to tune.

## B Implementation Details

In this section, we first describe the conditioning strategy of how Re-Meanflow utilizes previous pretrained models. We then outline the key design choices during training Re-Meanflow. We provide the training hyperparameters across all ImageNet resolutions in Tab. 4.

Table 4: Training settings of Re-Meanflow on ImageNet.

	Resolution		
	$64^2$	$256^2$	$512^2$
<b>Training details</b>			
Model backbone	EDM2-S[23]	SiT-XL[31]	EDM2-S[23]
Global batch size	128	128	128
Learning rate	1e-4	1e-4	1e-4
Adam $\beta_1$	0.9	0.9	0.9
Adam $\beta_2$	0.95	0.95	0.95
Model capacity (M params)	280.2	676.7	280.5
EMA rate	0.9999	0.9999	0.9999
<b>MeanFlow settings</b>			
Ratio of $r \neq t$	0.25	0.25	0.25
$p$ for adaptive weight	0.5	0.5	0.5
CFG effective scale $w'$	Uniform(1.0, 3.0)	Uniform(1.0, 3.0)	Uniform(1.0, 2.5)
Avoiding high-variance ( $t, r$ )	$t > 0.95 \wedge r < 0.4$	$t > 0.95 \wedge r < 0.4$	$t > 0.95 \wedge r < 0.4$
<b>Sampling details (rectified couplings from 1-rectified flow)</b>			
Pretrained model	EDM2-S[23]	SiT-XL[31]	EDM2-S[23]
Sampling number	5M	5M	5M
Guidance method	Autoguidance[22]	CFG[16]	Autoguidance[22]
Distance truncate strength	Top 10%	Top 10%	Top 10%

## B.1 Pretrained Model Conditioning

In Re-Meanflow, the velocity network is conditioned on two time variables,  $t$  and  $r$ . We implement this conditioning by learning two separate embeddings,  $\text{emb}_t(t)$  and  $\text{emb}_r(r)$ , and summing them before passing the result to the rest of the network. When initializing from pretrained models, the original networks only contain a single time embedding for  $t$ . Replacing this embedding with our two-embedding design requires careful initialization to ensure the model initially behaves like the pretrained flow mode. Specifically, before MeanFlow fine-tuning, we want:

$$u(\mathbf{x}_t, t, r) \approx v(\mathbf{x}_t, t) \quad (7)$$

**On ImageNet  $64^2$  and  $512^2$ ,** which Re-Meanflow is initialized Re-Meanflow from the pretrained EDM2-S model [23], following AYS [37], we perform a short alignment stage in which we train the new embeddings to reproduce the original time-embedding output for the corresponding noise level. Specifically, for EDM2-S the original embedding depends on  $\log \sigma_t$ , where  $\sigma_t = \frac{t}{1-t}$ . We train the new embeddings via:

$$\mathbb{E}_{t,r}[\|\text{emb}_t(t) + \text{emb}_r(r) - \text{emb}_{\text{ori}}(\log \sigma_t)\|_2^2], \quad (8)$$

for 10k iterations with learning rate 1e-3. This process takes only a few minutes. We also convert the original VE diffusion parameterization into the flow-matching setting following Lee et al. [24].

**On ImageNet  $256^2$ ,** we initialize Re-Meanflow from SiT-XL [31], which is already a flow-based model. In this case, only the additional  $r$ -embedding needs to be introduced. We simply zero-initialize  $\text{emb}_r(r)$  and keep the original SiT time embedding for  $\text{emb}_t(t)$ .

## B.2 Time Distribution

We observe a similar loss-time profile to that reported in Lee et al. [24]: the training loss as a function of  $t$  closely matches that of 2-rectified flow++. Accordingly, for sampling  $t$  we adopt the same U-shaped distribution:

$$p_t(u) \propto \exp(au) + \exp(-au), \quad u \in [0, 1], a = 4.$$

Following AYF [37], after sampling  $t$  we draw the interval length  $|t - r|$  from a normal distribution  $\mathcal{N}(P_{\text{mean}}, P_{\text{std}})$  and apply a sigmoid transformation. We use the same parameters as AYF,

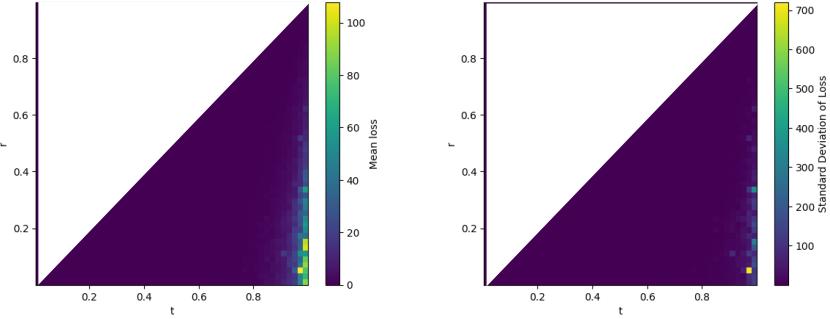


Figure 8: **Heatmaps of the mean loss (left) and the standard deviation of the loss (right)** for a trained Re-Meanflow model on ImageNet-512<sup>2</sup> under configuration (c).

$(P_{\text{mean}}, P_{\text{std}}) = (-0.8, 1.0)$ , which emphasize medium-length intervals and substantially improve stability. As shown in Table 3, this setting yields a noticeable improvement in FID and accelerates convergence.

We also experimented with sampling  $t$  uniformly. Despite its simplicity, uniform sampling performs competitively, and in many cases better than commonly used log-normal time distributions in diffusion and flow-matching models [21, 23, 11], which prioritize mid-range timesteps to avoid high variance near  $t \approx 0$  and  $t \approx 1$ . We attribute the strong performance of uniform sampling to the significantly lower variance of rectified trajectories: after one reflow step, early-time and high-noise regions become much more stable, allowing us to allocate more samples to these challenging regimes without the usual degradation observed when training on independent couplings.

### B.3 Avoiding High-Variance Time Regions

As discussed in Sec. 5.2, we observe that Re-Meanflow exhibits unusually high variance when the noise level  $t$  is large while the reference time  $r$  is close to zero. Intuitively, this corresponds to asking the model to predict a *slightly denoised* sample ( $r \in [0, 0.4]$ ) from an input that remains heavily corrupted ( $t \approx 1$ ). To quantify this effect, we sample 100k pairs of  $(t, r)$  uniformly and evaluate a trained Re-Meanflow model under configuration (c) in Table 3. As shown in Fig. 8, the resulting loss landscape displays a clear spike in both error and variance within this region, often even higher than the loss incurred when predicting directly from noise to a clean target.

Inspired by the truncation strategy in TCM [25], we adopt a simple yet effective rule to avoid this problematic regime: whenever a sampled pair satisfies  $t > 0.95$  and  $r < 0.4$ , we set  $r = 0$ . Empirically, this improves both training stability and FID. Our hypothesis is that predicting a *clean* image from pure noise ( $r = 0, t \approx 1$ ) is substantially easier than predicting a lightly corrupted target: the latter requires the model to determine which noise components should be preserved, introducing ambiguity and variance at high  $t$ . By redirecting training toward these easier high- $t$  targets, the model can allocate more capacity to learning accurate one-step predictions. This modification not only improves FID relative to configuration (c), but also accelerates convergence: in the high- $t$  regime, more updates involve  $r = 0$ , allowing the model to refine its one-step outputs more quickly and reliably.

### B.4 Training with Guidance

Classifier-free guidance (CFG) [16] is widely used to boost the performance of diffusion and flow-based generative models. To incorporate CFG into the MeanFlow stage, we train Re-Meanflow on the CFG-enhanced velocity field:

$$v^{\text{cfg}}(\mathbf{z}_t, t | c) \triangleq \omega v(\mathbf{z}_t, t | c) + (1 - \omega) v(\mathbf{z}_t, t). \quad (9)$$

MeanFlow [11] further introduced an improved CFG method that mixes conditional and unconditional mean-velocity predictions:

$$v^{\text{cfg}}(\mathbf{z}_t, t | c) = \omega v(\mathbf{z}_t, t | c) + \kappa u^{\text{cfg}}(\mathbf{z}_t, t, t | c) + (1 - \omega + \kappa) u^{\text{cfg}}(\mathbf{z}_t, t, t). \quad (10)$$

which is equivalent to using an effective guidance of  $\omega' = \frac{\omega}{1-\kappa}$ . We adopt this improved CFG formulation for all experiments.

Empirically, we found that directly training MeanFlow on the CFG field results in instability, consistent with observations in Hu et al. [19]. To mitigate this, we use a simple two-stage strategy: first train  $u_\theta$  on the unconditional flow, then on the CFG-modified flow. Usually, allocating half of the total training budget to each stage provides a good balance between stability and final quality.

We also found that sampling a random CFG scale  $\omega'$  from a uniform distribution (rather than fixing it) gives better results. Large values of  $\kappa$  is also important for stable training. In practice, we sample  $\omega'$  from a uniform distribution, then set  $\kappa = \max(1.0, \omega' - 1)$ , and finally compute the corresponding value for  $\omega = \frac{\omega'}{1-\kappa}$ .

## B.5 2D Toy Example Setting

We consider a controlled two-dimensional transport task designed to highlight curvature effects and the behavior of different flow-learning methods. Both the source and target distributions are mixtures of two Gaussians. The *source* distribution is a balanced mixture with equal component weights, while the *target* distribution is an imbalanced mixture with weights (0.6, 0.4). Component means are placed symmetrically on the left (source) and right (target) sides of the plane, and all components use identity covariance. This setup ensures that samples associated with the lighter target mode must traverse longer paths, creating the geometric imbalance that induces curved trajectories.

All velocity and MeanFlow models are implemented as small multilayer perceptrons (MLPs) with identical architectures across methods. Training is performed for a fixed budget of 20k steps using Adam with learning rate  $10^{-3}$  and batch size 1024.

We compare three approaches under this shared budget:

1. **2-rectified Flow**: two successive reflow steps applied to independently sampled couplings, each trained for 10k iterations.
2. **MeanFlow**: trained for the full 20k iterations directly on independent couplings.
3. **Re-Meanflow (ours)**: a velocity model is first trained for 10k iterations to obtain a 1-rectified flow; a MeanFlow model is then trained for another 10k iterations on the resulting rectified couplings.

Figure 3 visualizes the learned trajectories and one-step Euler sampling results for all three settings. As discussed in the main text, 2-rectified flow retains noticeable outliers due to residual curvature, and MeanFlow trained on independent couplings fails to converge within the budget. In contrast, Re-Meanflow combines moderate rectification with stable MeanFlow training to achieve accurate one-step generation.

## C More Experiment Details and Results

In this section, we provide additional experimental details, computational analyses, and extended results. We describe how we estimate FLOPs and GPU-hours for all compared methods, and we provide additional distance-error correlation for ImageNet 64<sup>2</sup> and 256<sup>2</sup>.

### C.1 Computation Estimation of Each Method

**FLOPS Estimation.** In Fig 7, we report the efficiency in terms of estimated exaFLOPs (Eflops). To ensure comparability, we estimate total training and sampling compute for each method based on their reported FLOPs per forward pass. Specifically, we use the following assumptions:

- The FLOPs of a forward pass are reported by prior works (e.g., EDM[21]: 100 GFLOPs, EDM2-S[23]: 102 GFLOPs, SiT-XL[31]: 118.64 GFLOPs).
- The FLOPs of a backward pass are measured empirically and are approximately 2× the cost of a forward pass. (One JVP operation is also counted as a backward pass), say for our example on the first stage where we will perform one forward of the model and one JVP

operation and one back propagation with JVP counted as one back propagation we have total flop amount of  $1 + (2 \times 2) = 1 + 4$  forward flops.

- For training, the total compute is computed as:

$$\text{Total Train FLOPs} = (\#\text{iters}) \times (\text{batch size}) \times (\text{forward+backward}) \times (\text{GFLOPs per fwd}).$$

- For sampling in the reflow process, the total compute is computed as:

$$\text{Total Sample FLOPs} = (\#\text{samples}) \times (\#\text{steps}) \times (\text{forward passes per step}) \times (\text{GFLOPs per fwd}).$$

- *Example: Re-Meanflow (Ours) on ImageNet-64<sup>2</sup>.* For sampling, we require:

$$\underbrace{5 \times 10^6}_{\#\text{samples}} \times \underbrace{63}_{\text{steps}} \times \underbrace{2}_{\text{fwd/step (auto-guidance)}} \times \underbrace{102}_{\text{GFLOPs/fwd}} \approx 6.43 \times 10^{10} \text{ GFLOPs} \approx 64 \text{ Eflops.}$$

For training, we have two stages, with the first stage trained on the original flow and the second stage trained on the CFG velocity field:

$$\underbrace{50,000}_{\text{iters}} \times \underbrace{128}_{\text{batch}} \times \underbrace{(1+4)}_{\text{fwd+back}} \times \underbrace{102}_{\text{GFLOPs/fwd}} + \underbrace{50,000}_{\text{iters}} \times \underbrace{128}_{\text{batch}} \times \underbrace{(3+4)}_{\text{fwd+back}} \times \underbrace{102}_{\text{GFLOPs/fwd}} \approx 8 \text{ Eflops.}$$

- For training, the total compute is computed as:

$$\text{Total Train FLOPs} = (\#\text{iters}) \times (\text{batch size}) \times (\text{forward+backward}) \times (\text{GFLOPs per fwd}).$$

- For sampling in the reflow process, the total compute is computed as:

$$\text{Total Sample FLOPs} = (\#\text{samples}) \times (\#\text{steps}) \times (\text{forward passes per step}) \times (\text{GFLOPs per fwd}).$$

- *Example: Re-Meanflow (Ours) on ImageNet-64<sup>2</sup>.* For sampling, we require:

$$\underbrace{5 \times 10^6}_{\#\text{samples}} \times \underbrace{63}_{\text{steps}} \times \underbrace{2}_{\text{fwd/step (auto-guidance)}} \times \underbrace{102}_{\text{GFLOPs/fwd}} \approx 6.43 \times 10^{10} \text{ GFLOPs} \approx 64 \text{ Eflops.}$$

For training, we have two stages, with the first stage trained on the original flow and the second stage trained on the CFG velocity field:

$$\underbrace{50,000}_{\text{iters}} \times \underbrace{128}_{\text{batch}} \times \underbrace{(1+4)}_{\text{fwd+back}} \times \underbrace{102}_{\text{GFLOPs/fwd}} + \underbrace{50,000}_{\text{iters}} \times \underbrace{128}_{\text{batch}} \times \underbrace{(3+4)}_{\text{fwd+back}} \times \underbrace{102}_{\text{GFLOPs/fwd}} \approx 8 \text{ Eflops.}$$

- *Example: AYF [37].* AYF does not require sampling, so only the training compute is considered:

$$\underbrace{50,000}_{\text{iters}} \times \underbrace{2048}_{\text{batch}} \times \underbrace{(4+4)}_{\text{fwd+back}} \times \underbrace{102}_{\text{GFLOPs/fwd}} \approx 8.36 \times 10^{10} \text{ GFLOPs} \approx 84 \text{ Eflops.}$$

**GPU Hours Estimation.** In Fig.7, we also estimate the total GPU hours for AYF[37] and 2-rectified flow++ [24]. For all methods, including ours, we follow the standard convention of computing GPU hours as

$$\text{GPU Hours} = (\#\text{ of GPUs}) \times (\text{wall-clock training time}).$$

For example, Re-Meanflow requires 66 hours of wall-clock time on 8 A100 GPUs, yielding  $8 \times 66 = 528$  GPU hours.

## C.2 Distance-Error Correlation at Other Resolutions

In Fig. 9, we replicate the distance-error correlation analysis of Fig. 4 on ImageNet 256<sup>2</sup> (Left) and 64<sup>2</sup> (Right), respectively. In both settings, we observe the same phenomenon: couplings with large data-noise  $\ell_2$  distances form a pronounced high-error tail of the flow regression objective. This mirrors our ImageNet 512<sup>2</sup> results.

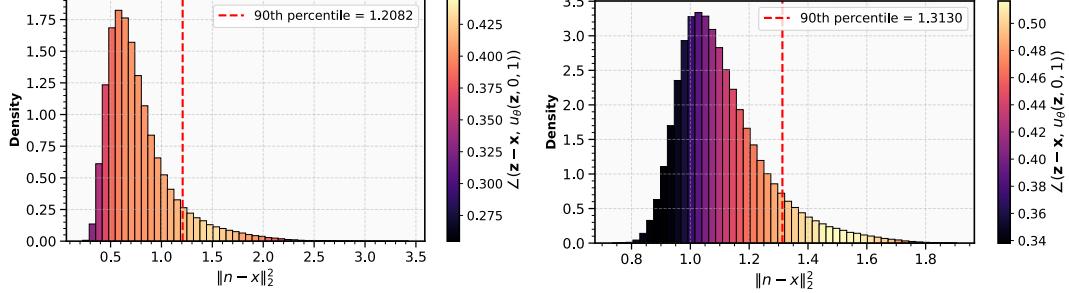


Figure 9: **Distance-Error Correlation.** Histogram of data-noise  $\ell_2$  distances on ImageNet 256<sup>2</sup> (Left) and ImageNet 64<sup>2</sup> (Right), analogous to Fig. 4. Similar high-distance, high-error tails (90th percentile marked) are observed.

## D Algorithm for Re-Meanflow

---

### Algorithm 1 Re-Meanflow Training

---

**Require:** dataset  $\mathcal{D}$ ; prior  $p_z$ ; time distributions  $p_t$  and  $p_{r,t}$ ; number of rectified-flow iterations  $T_{\text{flow}}$ ; number of meanflow iterations  $T_{\text{MF}}$ ; number of coupling samples  $N_{\text{pairs}}$ ; truncation ratio  $k\%$ .

```

1: Stage 1: Train 1-rectified Flow on Independent Couplings
2: for iter = 1 to  $T_{\text{flow}}$  do
3:   Sample  $\mathbf{x} \sim \mathcal{D}$ ,  $\mathbf{z} \sim p_z$ ,  $t \sim p_t$ 
4:    $\mathbf{z}_t \leftarrow (1-t)\mathbf{x} + t\mathbf{z}$ 
5:    $\dot{\mathbf{z}}_t \leftarrow \mathbf{z} - \mathbf{x}$ 
6:   Minimize  $\|\dot{\mathbf{z}}_t - v_\theta(\mathbf{z}_t, t)\|_2^2$  w.r.t.  $\theta$ 
7: end for
8: Freeze the resulting 1-rectified velocity field  $v_\theta^1$ 

9: Stage 2: Build Rectified Couplings + Distance Truncation
10: for  $n = 1$  to  $N_{\text{pairs}}$  do
11:   Sample  $\mathbf{x} \sim \mathcal{D}$ 
12:   Solve backward ODE with  $v_\theta^1$  from  $t = 1$  to  $t = 0$  to obtain  $\mathbf{z}$ 
13:    $d \leftarrow \|\mathbf{x} - \mathbf{z}\|_2$ 
14:   Store  $(\mathbf{x}, \mathbf{z}, d)$ 
15: end for
16: Let  $q$  be the  $(100 - k)$ th percentile of distances  $\{d_i\}$ 
17:  $\mathcal{D}_{\text{rect}} \leftarrow \{(\mathbf{x}_i, \mathbf{z}_i) : d_i \leq q\}$ 

18: Stage 3: Train MeanFlow on Rectified, Truncated Couplings
19: for iter = 1 to  $T_{\text{MF}}$  do
20:   Sample  $(\mathbf{x}, \mathbf{z}) \sim \mathcal{D}_{\text{rect}}$ ,  $(r, t) \sim p_{r,t}$ 
21:    $\mathbf{z}_t \leftarrow (1-t)\mathbf{x} + t\mathbf{z}$ 
22:   Compute  $g = (t-r) \frac{d}{dt} u_\phi(\mathbf{z}_t, r, t)$  ▷ JVP
23:    $u_{\text{tgt}} \leftarrow v_\theta^1(\mathbf{z}_t, t) - g$ 
24:   Minimize  $\|u_\phi(\mathbf{z}_t, r, t) - \text{sg}(u_{\text{tgt}})\|_2^2$  w.r.t.  $\phi$ 
25: end for
Ensure: trained meanflow model  $u_\phi$ 

```

---

## E More Qualitative Results

In this section, we present additional selected qualitative samples generated by Re-Meanflow across all ImageNet resolutions.

### E.1 ImageNet-64<sup>2</sup>

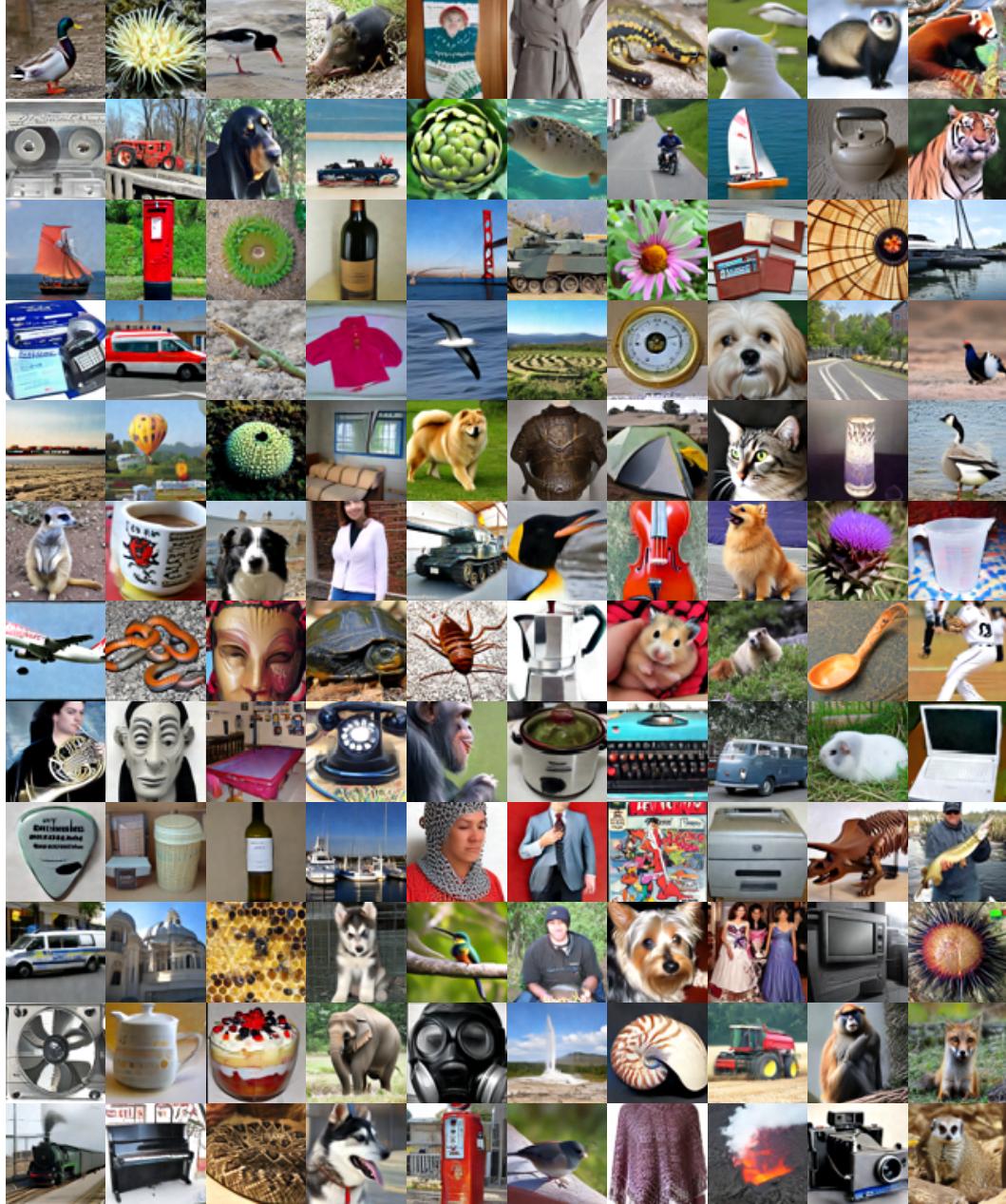


Figure 10: Selected qualitative results for Re-Meanflow (NFE=1) on ImageNet 64<sup>2</sup>.

## E.2 ImageNet-256<sup>2</sup>



Figure 11: Selected qualitative results for Re-Meanflow (NFE=1) on ImageNet 256<sup>2</sup>.

### E.3 ImageNet-512<sup>2</sup>



Figure 12: Selected qualitative results for Re-Meanflow (NFE=1) on ImageNet 512<sup>2</sup>.