SQUEEZE: Preserving Homeomorphism and Smooth in Higher-Dimensional Flows

Wangzi Yao[1,4,†] **Yue Sun**[2,†] **Rongmin Chen**[2]
Honglie Wang[4] **Jian Zhang**[2,3] **Tielin Zhang**[2,4,5,6,*]
[1]Shanghai Maiqi Xinzhi Technology Co., Ltd., Shanghai 200090, China
[2]Center for Excellence in Brain Science and Intelligence Technology,
State Key Laboratory of Brain Cognition and Brain-inspired Intelligence Technology,
Institute of Neuroscience, Chinese Academy of Sciences, Shanghai 200031, China
[3]University of Chinese Academy of Sciences, Beijing 100049, China
[4]Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[5]School of Artificial Intelligence, University of Chinese Academy of Sciences,
Beijing 100049, China
[6]The Key Laboratory of Cognition and Decision Intelligence for Complex Systems,
Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[†]Equal contribution. [*]Corresponding author: zhangtielin@ion.ac.cn

Abstract

Rectified flow (RF) is motivated as optimizing the velocity field through trajectory crossings. However, this explanation contradicts prior work showing that exact crossings cannot occur, when the dimension exceeds 2 and the source and target distributions are continuous. We revisit this question and make three contributions. 1) By relaxing the geometric definition of a crossing, we prove that **crossings can then occur, though their probability still drops rapidly with dimension**. 2) As a consequence, we observe **the homeomorphism of RF breaking down in high dimensions**. This effect is hard to notice, and coloring serves as a useful indicator. 3) To address this breakdown, we propose **Squeeze**, which corrects trajectories within the principal subspace of transport before lifting them back to the original space, thereby restoring smoothness and homeomorphism. Experiments confirm our analysis and show that **Squeeze alleviates the high-dimensional breakdown**.

1 Introduction

Continuous-time generative models seek to transport a simple source distribution to a complex data distribution by integrating a learned vector field. This perspective unifies a number of recent generative modeling paradigms, including normalizing flows [1], continuous normalizing flows built on neural ODEs [2, 3], and score-based diffusion models [4, 5]. Within this view, the geometric properties of the underlying transport, such as smoothness, invertibility, and the preservation of distributional and topological structure [6, 7], play a central role. Among continuous-time models, flow matching has emerged as a particularly tractable formulation, replacing simulation-based likelihood training with a regression objective on the velocity field [8–10]. Several lines of work have analyzed its statistical and approximation properties [11–13], including its adaptivity to low-dimensional manifold structure [14].

Rectified Flow (RF) is a particularly influential instance of this framework [15]. It learns an ODE whose trajectories follow straight paths between paired source and target samples, and recursively updates the coupling through reflow to obtain progressively straighter transport. A common low-dimensional intuition is that RF improves transport by resolving crossing or locally chaotic trajectories,

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

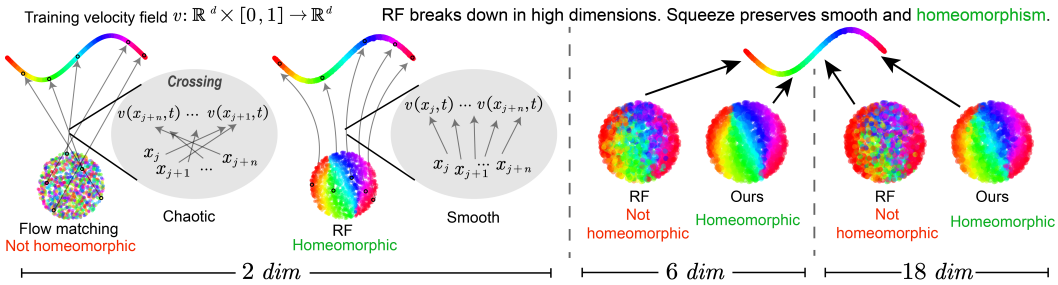


Figure 1: Generation process from a uniform spherical distribution to a sine-curve distribution, visualized through coloring. Trajectory endpoints on the sine curve are smoothly colored, and the starting points inside the sphere are colored the same as their corresponding endpoints. ***Smooth coloring is a diagnostic for flow homeomorphism***. The left panel shows the 2D case, where a chaotic velocity field (e.g., the training target of flow matching) induces a non-homeomorphic flow, while RF resolves the crossings to obtain a smooth velocity field, inducing a homeomorphic flow. ***From left to right, the dimension increases (2, 6, 18), and the homeomorphism of RF gradually breaks down. In contrast, our Squeeze method preserves both the homeomorphism of the flow and the smoothness of the velocity field.***

as shown in Figue 1 left. In two-dimensional examples, this picture is visually clear and often matches the behavior of reflow. However, recent work has shown that this intuition becomes fragile in higher dimensions. In particular, Reu et al. [16] prove that, when the source and target distributions admit continuous densities, exact interpolating trajectories almost surely do not cross for ambient dimension greater than two. This result appears at odds with the empirical effectiveness of RF in high-dimensional generation [15, 17], where reflow is still observed to straighten trajectories. It suggests that the strict mathematical notion of crossing may be too narrow to capture the effective local crossing that matter in practice.

Contributions. Motivated by this gap between exact crossing theory and the observed behavior of RF, we revisit crossing at finite resolution and study its effect on the geometry of the learned flow. In Section 2, we distinguish exact trajectory crossing from local crossing. We show that local crossing has nonzero probability under a local density condition, while its probability still decays rapidly with ambient dimension. In Section 3, we identify a high-dimensional degradation phenomenon in RF, where the learned velocity field becomes less smooth and the induced flow map becomes less homeomorphic, as shown in Figure 1. We further provide a coloring-based diagnostic, based on the composition of smooth invertible transformations, to visualize this otherwise non-intuitive geometric behavior. In Section 4, we propose Squeeze, which performs ODE simulation in a low-dimensional subspace spanned by the dominant directions of the velocity field and lifts the corrected trajectories back to the ambient space. Experiments verify our theoretical analysis and show that Squeeze alleviates the degradation of velocity-field smoothness and flow-map homeomorphism.

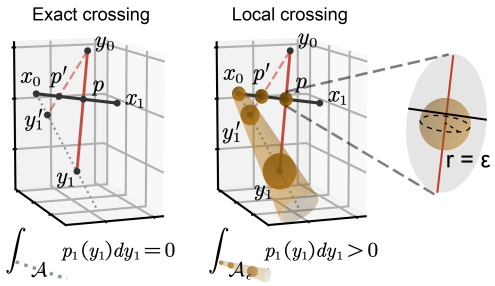


Figure 2: Illustration of crossing probabilities. The exact crossing probability is an integral over a ray (dashed line) and equals 0. For local crossing, it suffices that the distance d between x_0x_1 and y_0y_1 is less than ϵ (a small neighborhood), so the probability is an integral over a cone-like region (orange) and is nonzero under a local density condition (Proposition 2).

2 Local crossing probability is nonzero but decays with dimension

In this section, we first define exact crossings and local crossings. We then analyze their occurrence probabilities. Unlike exact crossings, local crossings have a nonzero probability of occurring under a

local density condition. Finally, we prove that the probability of local crossings still decreases rapidly as the dimension increases.

Exact crossings, local crossings. We first give their formal definitions of exact crossings and local crossings. Let p_0 and p_1 be probability densities on $\mathbb{R}^d$ with respect to the d-dimensional Lebesgue measure λ_d. Let $x_0, y_0 \sim p_0, \qquad x_1, y_1 \sim p_1$ be sampled independently. For $a, b \in \mathbb{R}^d$, define the linear interpolant $\ell_{a,b}(t) = (1-t)a + tb, \qquad t \in (0,1)$. For fixed (x_0, x_1, y_0), let

$$\gamma(t) := \frac{(1-t)x_0 + tx_1 - (1-t)y_0}{t}, \quad t \in (0,1), \tag{1}$$

define the **exact crossing** endpoint set

$$A(x_0, x_1, y_0) = \{\gamma(t) : t \in (0,1)\}. \tag{2}$$

For practical **local crossings**, we define the state-space local crossing region in endpoint space by

$$A^{\mathrm{loc}}_{I,\varepsilon}(x_0, x_1, y_0) := \left\{ y \in \mathbb{R}^d : \exists t \in I \text{ such that } \|\ell_{y_0,y}(t) - \ell_{x_0,x_1}(t)\| \le \varepsilon \right\}, \tag{3}$$

where $I = [\alpha, \beta] \subset (0,1)$ is a compact time interval. Since $\ell_{y_0,y}(t) - \ell_{x_0,x_1}(t) = t\big(y - \gamma(t)\big)$, this is equivalent

$$A^{\mathrm{loc}}_{I,\varepsilon}(x_0, x_1, y_0) = \left\{ y \in \mathbb{R}^d : \exists t \in I \text{ such that } \|y - \gamma(t)\| \le \frac{\varepsilon}{t} \right\}. \tag{4}$$

Therefore

$$A^{\mathrm{loc}}_{I,\varepsilon}(x_0, x_1, y_0) \subseteq \left\{ y \in \mathbb{R}^d : \mathrm{dist}\left(y, A_I(x_0, x_1, y_0)\right) \le \frac{\varepsilon}{\alpha} \right\}, \tag{5}$$

where $A_I(x_0, x_1, y_0) := \{\gamma(t) : t \in I\}$. With the definitions of exact crossings and local crossings in place, we now analyze the probability of each.

Proposition 1 (The probability of exact crossing.)**.** *Assume that y_1 has a probability density p_1 with respect to the d-dimensional Lebesgue measure λ_d. Let $d \ge 2$. Then, for every fixed $(x_0, x_1, y_0) \in (\mathbb{R}^d)^3$,*

$$\mathbb{P}\big(y_1 \in A(x_0, x_1, y_0)\big) = \int_{A(x_0,x_1,y_0)} p_1(y)\, dy = 0. \tag{6}$$

The proof idea is as follows, with details provided in the Appendix C.1. For fixed (x_0, x_1, y_0). An exact same-time crossing means that there exists $t \in (0,1)$ such that $\ell_{x_0,x_1}(t) = \ell_{y_0,y_1}(t)$. It implies that there exists $t \in (0,1)$ satisfies

$$y_1 = \frac{(1-t)x_0 + tx_1 - (1-t)y_0}{t} = x_1 + \left(\frac{1}{t} - 1\right)(x_0 - y_0). \tag{7}$$

Thus the set of all endpoints y_1 that produce exact same-time crossings, which is defined as $A(x_0, x_1, y_0)$, contained in an affine line in $\mathbb{R}^d$. If $x_0 \neq y_0$, it is contained in a half-line; if $x_0 = y_0$, it reduces to the single point x_1. Therefore, for $d \ge 2$, $\lambda_d\big(A(x_0, x_1, y_0)\big) = 0$, it implies that Equation 6.

Local crossing probability under a local density condition. As shown above, the probability of an exact crossing is an integral over a lower-dimensional region and thus equals zero. This, however, changes in the calculation of the local crossing probability below. Figure 2 illustrates the difference.

Proposition 2 (Bounds on local crossing probability.)**.** *Assume $p_1 \in L^\infty(\mathbb{R}^d)$. For every fixed $I = [\alpha, \beta] \subset (0,1)$ and $(x_0, x_1, y_0) \in (\mathbb{R}^d)^3$, there exists a constant L_I such that*

$$\mathbb{P}\big(y_1 \in A^{\mathrm{loc}}_{I,\varepsilon}(x_0, x_1, y_0)\big) \le \|p_1\|_\infty \left[\omega_{d-1} L_I(x_0, y_0) \left(\frac{\varepsilon}{\alpha}\right)^{d-1} + \omega_d \left(\frac{\varepsilon}{\alpha}\right)^d\right], \tag{8}$$

where ω_m denotes the volume of the unit ball in $\mathbb{R}^m$. If, additionally, there exist $t_0 \in I$, $0 < r_0 \le \varepsilon/t_0$ and $m > 0$ such that $p_1(y) \ge m$ almost everywhere on $B(\gamma(t_0), r_0)$, then

$$\mathbb{P}\big(y_1 \in A^{\mathrm{loc}}_{I,\varepsilon}(x_0, x_1, y_0)\big) \ge m\omega_d r_0^d > 0.$$

Proof. We have already defined the local crossing region in Equation 5. Since $I = [\alpha, \beta]$ is a compact set, for convenience, we define the above right-hand side, which is the ε/α-tube, around a line segment of length at most $L_I(x_0, y_0)$. Hence, its d-dimensional Lebesgue measure is bounded by

$$\omega_{d-1} L_I(x_0, y_0) \left(\frac{\varepsilon}{\alpha}\right)^{d-1} + \omega_d \left(\frac{\varepsilon}{\alpha}\right)^d. \tag{9}$$

With the help of the assumption $p_1 \in L^\infty(\mathbb{R}^d)$, we obtain that

$$\begin{aligned} &\mathbb{P}\big(Y_1 \in A^{\mathrm{loc}}_{I,\varepsilon}(x_0, x_1, y_0)\big) = \int_{A^{\mathrm{loc}}_{I,\varepsilon}(x_0,x_1,y_0)} p_1(y)\, dy \\ \leq &\|p_1\|_\infty \lambda_d\big(A^{\mathrm{loc}}_{I,\varepsilon}(x_0, x_1, y_0)\big) \leq \|p_1\|_\infty \left[\omega_{d-1} L_I(x_0, y_0) \left(\frac{\varepsilon}{\alpha}\right)^{d-1} + \omega_d \left(\frac{\varepsilon}{\alpha}\right)^d\right]. \end{aligned} \tag{10}$$

This proves the upper bound. For the lower bound, taking $t = t_0$ gives $B(\gamma(t_0), r_0) \subseteq A^{\mathrm{loc}}_{I,\varepsilon}$, and integrating $p_1 \geq m$ over this ball gives $m\omega_d r_0^d$. The additional condition holds whenever p_1 is continuous and positive at some $\gamma(t_0)$. □

Remark 1 (Why the time truncation is natural)**.** *The restriction to a compact interval $I = [\alpha, \beta] \subset (0, 1)$ excludes degenerate endpoint regimes. This is a natural assumption for analyzing meaningful local trajectory crossing. Indeed, when the endpoint pairs are distinct, for example $x_0 \neq y_0$ and $x_1 \neq y_1$, the two trajectories are separated at both endpoints. By continuity of the trajectory, there exist small neighborhoods near $t = 0$ and $t = 1$ in which the two trajectories remain separated. Thus, possible local crossing occur in the interior part of the transport path. Therefore, it is reasonable to analyze local crossings on an interior interval $I = [\alpha, \beta] \subset (0, 1)$.*

Remark 2 (Why local crossings fade in high dimension)**.** *Proposition 2 shows that the local crossing region is controlled by an ε/α-tube around a one-dimensional set. Hence its leading volume term scales as*

$$O\left(\left(\frac{\varepsilon}{\alpha}\right)^{d-1}\right). \tag{11}$$

Therefore, for a small fixed neighborhood size $\varepsilon \ll 1$, the probability of local crossings decreases rapidly as d increases. This formalizes the intuition that local crossings become increasingly rare in high-dimensional spaces.

3 Breakdown of smooth and homeomorphic flow in high dimensions

The previous chapter showed that local crossings become increasingly unlikely in high-dimensional spaces. As a consequence, the rectifying effect of RF, which depends on trajectory crossings, is weakened. In this chapter we further observe that the smooth velocity fields and homeomorphic flow maps that RF achieves in low-dimensional settings break down in high dimensions. We first introduce the relationship between smooth velocity fields and homeomorphic flow maps. Then we use coloring as an intuitive visualization tool to examine whether the learned velocity fields is smooth and homeomorphic. Finally, we show that the smooth velocity fields and homeomorphic behavior observed in low-dimensional RF can be disrupted in high dimensions, and that our Squeeze method helps mitigate this breakdown.

Smooth generative velocity fields induce homeomorphic flow maps. Flow matching [8] trains a time-dependent velocity field $v_\theta : \mathbb{R}^d \times [0, 1] \to \mathbb{R}^d$, parameterized by θ. This velocity field induces a flow map $\Phi_{s,\tau} : \mathbb{R}^d \to \mathbb{R}^d$ that transports a state at time s to its position at time τ via

$$\frac{dx_t}{dt} = v_\theta(x_t, t), \quad \Phi_{s,\tau}(x_s) = x_\tau. \tag{12}$$

By classical ODE theory, when v_θ is sufficiently regular, $\Phi_{s,\tau}$ is a diffeomorphism, and in particular a homeomorphism, of $\mathbb{R}^d$ [18]. In this paper we focus on the latter, weaker property.

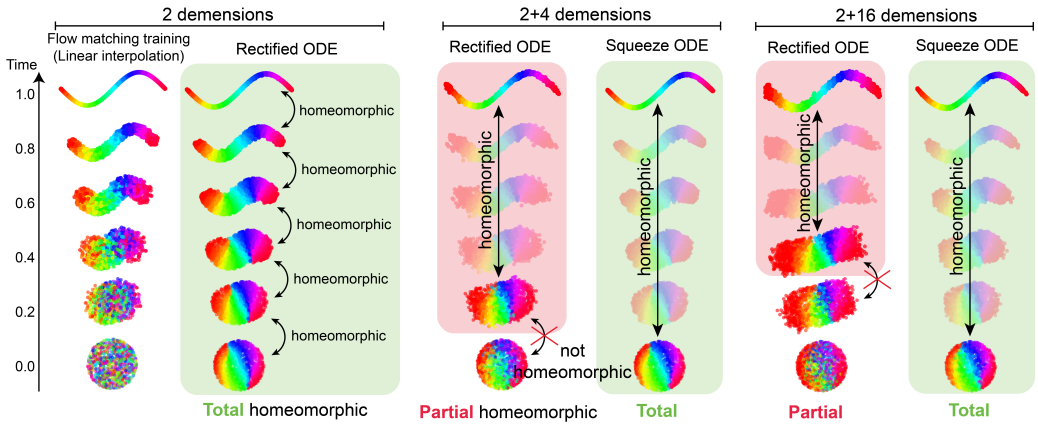


Figure 3: As the dimension increases, the homeomorphism breakdown appears earlier. Time goes from 0 to 1 from bottom to top, and the dimension increases from left to right. The coloring rule is the same as in Figure 1. We can see that RF preserves homeomorphism throughout the entire generation process in 2D, but as the dimension increases, the homeomorphism breakdown appears earlier and earlier. In contrast, our Squeeze method preserves homeomorphism throughout.

Coloring as a diagnostic of flow regularity. Directly checking whether a flow map is a homeomorphism, or whether its velocity field is smooth, is difficult. We instead adopt a coloring approach: assign a smooth color map $c_t : \Omega_t \to \mathbb{R}^k$ to a region Ω_t at time t, and pull it back along the flow to obtain $c_s = c_t \circ \Phi_{s,t}$ on Ω_s. The following composition property motivates this diagnostic.

Proposition 3 (Continuity and smoothness of pulled-back coloring)**.** *Let Ω_s, Ω_t be open subsets of $\mathbb{R}^d$. If $c_t : \Omega_t \to \mathbb{R}^k$ and $\Phi_{s,t} : \Omega_s \to \Omega_t$ are continuous, then $c_s = c_t \circ \Phi_{s,t}$ is continuous. If both maps are C^1, their composition is C^1.*

The proof follows from closure under composition. This motivates the coloring-coherence diagnostic; it does not test injectivity, and a finite sampled color plot cannot establish failure of continuity or C^1 regularity.

Returning to Figure 1, we smoothly color the sine curve at $t = 1$. For any $s < 1$, an irregularly colored sphere indicates reduced coloring coherence in the numerically realized flow.

Breakdown of smooth and homeomorphic behavior in high dimensions. With this indicator at hand, we now examine how RF behaves across different dimensions. In low-dimensional settings, RF can learn a smooth velocity field, and a smooth velocity field induces a homeomorphic flow map. As shown in the left part of Figure 1, the velocity field defined by the original flow matching objective is highly irregular around crossing regions. After coloring, the source distribution exhibits mixed and discontinuous color patterns. This illustrates reduced coloring coherence, as motivated by Proposition 3. In contrast, RF rectifies chaotic crossing regions into a smoother velocity field. As a result, the colored source distribution remains smooth. Although the coloring criterion is formally a necessary diagnostic rather than a sufficient proof, it suggests that RF learns a smooth velocity field in low dimensions, and that the induced flow map behaves homeomorphically.

However, as the dimension increases, the probability of trajectory crossing decreases, as discussed in Section 2. In this case, RF may leave some locally irregular regions in the velocity field, because these regions are not treated as crossings. These unresolved regions can make the learned velocity field non-smooth. As shown in the right part of Figure 1 and in Figure 3, when the dimension increases to 6 and 18, the colors on the source distribution become mixed and torn. This indicates that the velocity field does not induce a smooth identity-preserving homeomorphic flow map. Moreover, Figure 3 shows that the non-homeomorphic behavior appears earlier in 18 dimensions than in 6 dimensions, suggesting that this breakdown becomes more severe as the dimension increases. In contrast, our Squeeze method consistently preserves smooth color variation, indicating that it better maintains the smoothness of the velocity field and the homeomorphic behavior of the induced flow map in higher dimensions.

4 Our method

Motivation. The preceding result in Proposition 2 provides the key geometric intuition behind our method. The decay depends directly on the ambient dimension through the tube-volume factor ε^{d-1}. Thus, in the original high-dimensional space, the local crossing that induce velocity averaging in RF become much less frequent. A natural way to counteract this effect is to perform the correction in a lower-dimensional space. If the trajectories are projected to a k-dimensional subspace with $k \ll d$, the corresponding tube-volume factor becomes ε^{k-1} rather than ε^{d-1}. Hence, local crossing-like interactions are substantially more likely after projection.

This observation motivates Squeeze. Instead of correcting the coupling directly in the high-dimensional space, Squeeze first projects the transport into a low-dimensional subspace spanned by the principal transport directions. In this space, trajectories that are separated by weak or redundant high-dimensional coordinates can become locally close again, making the RF-style local averaging mechanism more effective. The corrected low-dimensional coupling is then lifted back to the original space for high-dimensional training. In this way, Squeeze increases the probability of useful local crossings while still producing a model in the original data space.

Squeeze. We give a brief description here, with the detailed pseudocode provided in Appendix B. Given a trained flow matching model v_θ^{fm} and a real dataset $\mathcal{D} = \{Y^{(j)}\}_{j=1}^N$, Squeeze proceeds in two stages. **Squeeze stage.** We first sample $X_0^{(j)} \sim \mathcal{N}(0, I_d)$, integrate the FM ODE to obtain endpoints, and pair them with real samples through the optimal permutation $\sigma_0 \in S_N$, yielding the coupling $\mathcal{C}_{\text{fm}} = \{(X_0^{(j)}, X_1^{(j)})\}_{j=1}^N$. We then take the SVD of the centered transport vectors $X_1^{(j)} - X_0^{(j)}$ and $X_0^{(j)} - X_1^{(j)}$ and retain the top-k directions $P_k \in \mathbb{R}^{d\times k}$. A flow matching model u_ϕ is trained on the projected targets $\{P_k^\top X_1^{(j)}\}$ in $\mathbb{R}^k$. **Lift stage.** We integrate u_ϕ from $P_k^\top X_0^{(j)}$ to obtain $z_1^{(j)}$, find the permutation σ_1 that best matches $z_1^{(j)}$ to $P_k^\top X_1^{(\sigma(j))}$, and form the lifted coupling $\mathcal{C}_{\text{sq}} = \{(X_0^{(j)}, X_1^{(\sigma_1(j))})\}$. Fine-tuning v_θ^{fm} on $\mathcal{C}_{\text{sq}}$ produces the final squeeze model v_θ^{sq}.

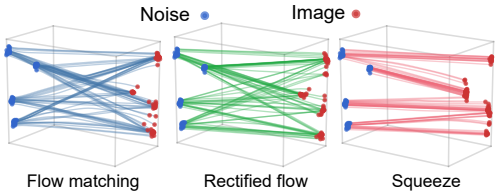


Figure 4: Generation trajectories on a 3D toy model. The trajectories of our Squeeze method are the most ordered, and the velocity field is the smoothest.

Marginal preservation under squeeze. Squeeze corrects the transport in a low-dimensional projected space. We first show that the projected Squeeze flow satisfies the same marginal-preserving principle as rectified flow. Assume that X is a path-wise continuously differentiable random process, the projected path has velocity $\dot{x}_t = P^\top \dot{X}_t$. We define the projected conditional velocity field by

$$u^P(\xi, t) = \mathbb{E}[\dot{x}_t \mid x_t = \xi] = \mathbb{E}[P^\top \dot{X}_t \mid P^\top X_t = \xi], \qquad \xi \in \mathbb{R}^k. \tag{13}$$

The ideal projected Squeeze flow is the ODE

$$\frac{d}{dt} z_t = u^P(z_t, t), \qquad z_0 = P^\top X_0. \tag{14}$$

Theorem 1. *[Projected marginal preservation] Assume X is path-wise continuously differentiable and z is the corresponding rectified flow in a low-dimensional projected space. Then* $\text{Law}(z_t) = \text{Law}(P^\top X_t)$ *for any* $t \in [0, 1]$.

The proof of this theorem is the projected analogue of the marginal-preserving property of rectified flow. More details can be found in the Appendix C.2. Moreover, the lift step restores the connection to the original data space, and the lift step preserves the marginals by construction. Since σ_1 is a permutation, the lifted targets $\{X_1^{(\sigma_1(j))}\}_{j=1}^N$ are exactly the original target samples $\{X_1^{(j)}\}_{j=1}^N$, only reordered. Therefore, Squeeze changes the pairing between source and target samples, but not the source or target marginals.

5 Experiments

We first empirically verify the theoretical prediction that the probability of local crossings decreases as the ambient dimension grows. We then show that, despite the challenges introduced by higher dimensions, Squeeze better preserves both the smoothness of the velocity field and the homeomorphism of the induced flow map. We further use quantitative metrics to confirm that the velocity field learned by Squeeze is less chaotic. Next, we evaluate few-step generation, comparing Squeeze with RF and several classical flow-based methods. Finally, we discuss the choice of the small auxiliary model and its additional computational overhead. Implementation details are provided in Appendix E. For dimensionality reduction we use a linear method, since nonlinear alternatives (e.g., kernel PCA, SNE) gave similar results at higher cost. We omit ablations because Squeeze is the only module added on top of RF.

5.1 Crossing probability decreases with dimension

Proportion 2 shows that the probability of local crossings decreases as the ambient dimension grows. We verify this prediction on a full-dimensional Gaussian toy dataset and CIFAR-10. For each setting, we draw $10{,}000$ samples from the initial noise distribution and $10{,}000$ samples from the target distribution, then form $10{,}000$ randomly paired linear trajectories. For a neighborhood radius ε, we count a local crossing when two trajectories pass within distance ε at the same interpolation time.

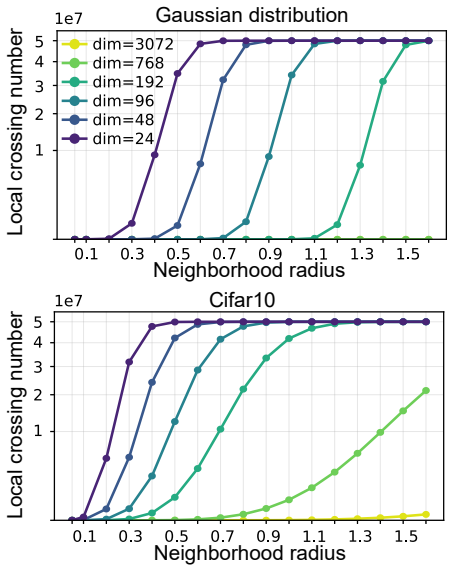


Figure 5: Number of crossings. The horizontal axis is the neighborhood radius used to determine crossings. At the same radius, the number of intersections decreases as the dimension increases, which is consistent with our proof.

Figure 5 reports the number of local crossings as a function of ε. The horizontal axis is the neighborhood radius, and the color changes from purple to yellow as the dimension increases. For CIFAR-10, dimension k means that k coordinates are randomly selected from the flattened $32 \times 32 \times 3$ image space before computing distances and crossings.

The results match the analysis. When the neighborhood radius is fixed, the number of crossings decreases as the dimension increases. When the neighborhood radius becomes smaller, the crossing condition becomes stricter, so the crossing count also decreases. We also observe fewer crossings on the Gaussian toy dataset than on CIFAR-10 at comparable dimensions [19]. Since the Gaussian fills the ambient space while CIFAR-10 concentrates near a lower-dimensional manifold, its effective dimension is higher, again consistent with our prediction.

5.2 Squeeze preserves homeomorphism in high dimensions

We compare the degree of homeomorphism of RF and Squeeze on a toy distribution, CIFAR-10, CIFAR-100, and ImageNet, using the coloring-coherence diagnostic in Proposition 3 for visualization.

Figure 3 shows the toy experiment. The initial distribution is uniform on a ball. The target distribution has its first two coordinates uniformly supported on a sine-curve segment, while the remaining padding dimensions are independent Gaussian coordinates. For visualization, all high-dimensional samples are projected to two dimensions. In 2 dimensions, RF preserves a fully smooth coloring across time. The distributions at different times can be transformed into one another. This indicates that the learned flow is close to homeomorphic.

This behavior degrades as the dimension increases. At dimension 6, RF only preserves homeomorphic behavior from $t = 0.2$ to $t = 1$, while a topological break appears near $t = 0.2$. At dimension 18,

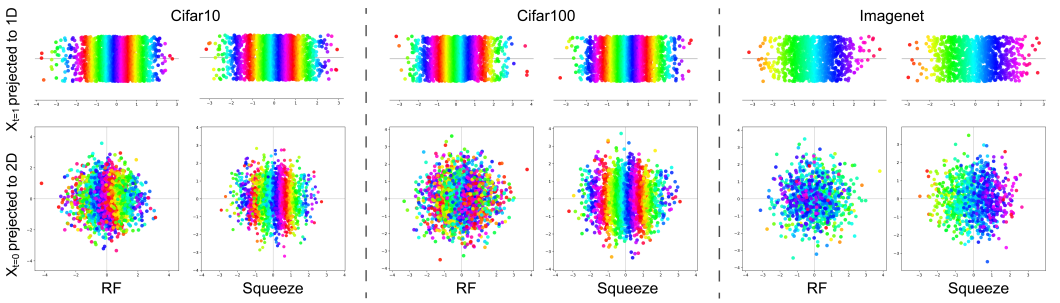


Figure 6: Degree of homeomorphism of the flows trained by Squeeze and RF on real datasets. We can see that the coloring of starting points is always smoother for Squeeze, indicating a stronger homeomorphism.

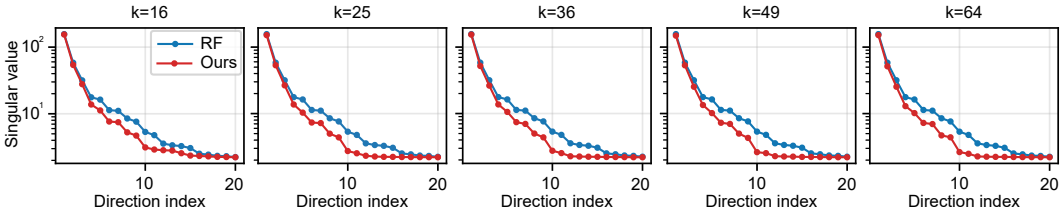


Figure 7: Chaotic analysis of generation trajectories along different directions. The vertical axis is the singular value, where larger values indicate more chaos. The horizontal axis lists different directions, sorted by singular value in descending order.

the non-homeomorphic behavior appears earlier, around $t = 0.4$. These results show that increasing dimension gradually damages the homeomorphic structure of RF. In contrast, Squeeze keeps the coloring smooth at these dimensions, indicating that it preserves a smoother and more homeomorphic transport.

We further evaluate RF and Squeeze on CIFAR-10, CIFAR-100, and ImageNet [20], as shown in Figure 6. Across these datasets, Squeeze produces smoother color transitions than RF. This suggests that the flow induced by Squeeze is more homeomorphic and has a smoother velocity field. ImageNet is harder than CIFAR-10 and CIFAR-100 because its representation has higher ambient dimension. After VAE encoding, ImageNet uses a $32 \times 32 \times 4$ representation, while CIFAR-10 and CIFAR-100 use $32 \times 32 \times 3$. This higher dimension further reduces effective crossings, making homeomorphic transport harder to preserve.

Table 1: Generation results on CIFAR-10 and CIFAR-100 under different NFEs. Model is Unet. Entries are five-seed means; standard deviations are in Table 10.

NFE	Methods	CIFAR-10			CIFAR-100		
		FID ↓	IS ↑	Recall ↑	FID ↓	IS ↑	Recall ↑
1	FM	355.92	1.03	0.02	349.02	1.02	0.02
	RF	135.51	2.29	17.64	125.57	2.62	23.05
	OT-FM	224.14	2.22	0.15	240.69	2.28	0.32
	Squeeze (Ours)	**79.27**	**3.51**	**38.77**	**76.14**	**4.04**	**41.41**
2	FM	191.97	1.48	6.01	200.13	1.54	8.03
	RF	38.62	4.57	58.22	46.11	4.78	55.76
	OT-FM	95.66	5.09	6.23	92.76	4.85	18.33
	Squeeze (Ours)	**23.59**	**5.90**	**77.05**	**30.87**	**6.07**	**69.25**
4	FM	57.36	3.50	38.72	60.36	3.48	40.73
	RF	13.27	6.64	85.01	20.83	6.64	76.92
	OT-FM	32.23	**7.38**	58.12	37.97	6.83	40.14
	Squeeze (Ours)	**11.35**	7.23	**89.16**	**17.31**	**7.43**	**81.57**
200	FM	**3.11**	8.53	**96.02**	**7.19**	8.68	**90.00**
	RF	7.76	8.40	93.08	12.27	8.82	85.79
	OT-FM	6.06	**8.89**	66.69	12.16	8.14	53.88
	Squeeze (Ours)	6.54	8.36	93.78	11.34	**8.90**	87.21

5.3 Less chaotic velocity fields

We compare the chaoticity of the velocity fields learned by RF and Squeeze. Since their trajectories are approximately linear, each trajectory is summarized by its displacement vector $\Delta x^{(i)} = x_1^{(i)} - x_0^{(i)}$, where $x_0^{(i)}$ and $x_1^{(i)}$ are its endpoints. We stack all displacement vectors into a matrix D, center it, and apply singular value decomposition. The leading singular vectors identify the directions of largest trajectory variation, and the corresponding singular values quantify the disorder along these directions—larger values indicating a more disordered field. As shown in Figure 7, the singular values of Squeeze lie consistently below those of RF, indicating a

less chaotic velocity field. Different values of k correspond to different hyperparameter settings of Squeeze, all of which achieve favorable results.

5.4 Few-step generation

The analysis above suggests that the smoother velocity field and more homeomorphic flow map produced by Squeeze should benefit few-step generation. We evaluate this on CIFAR-10, CIFAR-100, and ImageNet, with U-Net and DiT-B/4 backbones, varying the number of function evaluations (NFEs). Baselines include Flow Matching [8], RF [15], and OT-FM [21]. Our goal is not to match specialized NFE methods but to isolate the effect of smoothness and homeomorphism. Since Squeeze is derived from RF, RF serves as the most direct comparison.

Table 2: FID of generation results on ImageNet256 under different NFEs. Model is DiT-B/4.

Methods	NFE				
	1	2	4	8	50
FM	375.24	306.26	219.53	87.53	**64.13**
RF	305.77	211.44	128.00	85.98	67.69
OT-FM	342.18	190.36	151.47	103.84	83.72
Squeeze (Ours)	**295.14**	**168.91**	**119.87**	**84.22**	66.55

On CIFAR-10/100, we train the FM U-Net for 500 epochs (CIFAR-10 FID 3.05, matching [21]), then fine-tune RF and Squeeze from this checkpoint for 10 epochs; OT-FM follows the FM protocol for 500 epochs. On ImageNet, we train DiT-B/4 with FM for 400K iterations (FID 64.13, close to [22]) and fine-tune RF and Squeeze for 10K iterations. Hyperparameters are detailed in Appendix E.

Tables 1 and 2 show that Squeeze consistently outperforms RF in the low-NFE regime (1, 2, 4 steps) across FID, IS, and Recall. At high NFE, both RF and Squeeze fall slightly behind FM, which is expected: with enough sampling steps, even a more complex velocity field is well approximated by the solver. These results show measurable gains for few-step generation alongside the smoother velocity field and more homeomorphic flow map of Squeeze.

5.5 The additional time and space overhead of Squeeze is negligible

Both Squeeze and RF are fine-tuning methods built on FM. The only addition in Squeeze is the squeeze stage of Section 1, which trains an extra small model. We now show that the compute and time cost of this auxiliary model is negligible.

Table 3: Computational overhead.

Resource	U-Net FM	Small flow	Ratio
Parameters	55.68M	0.092M	$605\times$
Training GPU-hours	$\sim$ 110–130 A100·h	0.35 A100·h	$\sim 350\times$
Training wall-clock	16h $\times$ 8 GPUs	21min $\times$ 1 GPU	$\sim 46\times$

Squeeze uses a small MLP for the auxiliary low-dimensional flow, since this stage only learns the velocity field after projection rather than generating images in the original data space. To verify that this capacity suffices, we compare two distributions on CIFAR-10 PCA-64 under the sliced Wasserstein distance [23]: samples from a strong U-Net FM model (FID 3.05) projected to the low-dimensional space yield $\mathrm{SW}_2 = 0.1128$, while our small MLP trained directly in that space for 300K iterations yields $\mathrm{SW}_2 = 0.1140$. Since the former is known to be of high quality and the latter nearly matches it (both far from the noise baseline of about 0.2), the MLP is adequate for the low-dimensional task. The auxiliary stage is also far cheaper than the image-space model. As shown in Table 3, the small flow has only 0.092M parameters and requires 0.35 A100 GPU-hours, compared with 55.68M parameters and roughly 110–130 A100 GPU-hours for the U-Net FM model. Hence, the overhead of Squeeze is essentially negligible relative to the main model training.

6 Related work and limitation

On the term "homeomorphism". We use "homeomorphism" as an empirical descriptor. Classical flow theory [24] guarantees that a continuous-time flow induced by a sufficiently regular velocity field is itself a homeomorphism, so the idealized rectified-flow ODE does not violate this property. Our claims about RF being "non-homeomorphic" refer to the *learned and numerically integrated* flow map under finite capacity, finite training, and discrete integration, a regime in which the realized flow is known to depart from its continuous-time ideal [25, 26]. We empirically observe folding-like artifacts and near-non-invertible regions inconsistent with a smooth homeomorphic flow.

Baseline selection in few-step generation. We focus on whether Squeeze's homeomorphism brings gains in few-step generation. To isolate this effect, we strictly control for variables. Our comparisons are limited to methods that share our architecture and training pipeline, differing only in the Squeeze component. Therefore we do not compare with the current strong baselines: their training paradigms differ too substantially from ours, and the additional variables they introduce would make it impossible to attribute performance differences cleanly to the homeomorphism property. For instance, consistency models [27] rely on a specifically designed boundary-condition parameterization, an EMA target network, and a carefully tuned discretization schedule; distillation-based methods such as InstaFlow [17], ADD [28], and DMD [29] depend on a pretrained teacher, so their performance is tightly coupled with the teacher's capacity and the choice of distillation objective; more recent approaches such as Shortcut Models [30] and MeanFlow [31] further require additional time conditioning, JVP computations, and self-consistency constraints. Each of these design choices substantially alters the inductive bias and optimization landscape of the model, so a direct comparison would not allow us to isolate the effect of the homeomorphism property itself.

Recent advances in RF. Beyond the original formulation, RF has been refined and extended along several axes. Stable Diffusion 3 [32] demonstrates its scalability to large-scale text-to-image synthesis. The framework has further been extended to Riemannian manifolds [33], discrete data [34], and equivariant settings for molecular generation [35]. On the coupling side, OT-CFM and multisample flow matching [21, 36] leverage optimal-transport couplings to straighten trajectories, and variational RF [37] models multi-modal velocity fields to handle ambiguity in coupled samples. Recent analyses further examine the geometry and training dynamics of RF trajectories [38]. Our analysis of homeomorphic behavior is largely orthogonal to these directions, and may suggest an interesting new perspective worth further investigation.

Acknowledgments and Disclosure of Funding

This work was supported by the Brain Science and Brain-like Intelligence Technology – National Science and Technology Major Project (2025ZD0217200), the Science and Technology Commission of Shanghai Municipality (25511102500), the Shanghai Municipal Science and Technology Project (Grant Nos. 25ZR1401370 and 25LN3200400), the CAS Project for Young Scientists in Basic Research (YSBR-116), the Youth Innovation Promotion Association CAS, the Shanghai Leading Talent Program of Eastern Talent Plan, and the Lingang Laboratory Fund (Grant Nos. LG-GG-202402-06-07 and LGL-1987-09). This work was also supported by Shanghai Maiqi Xinzhi Technology Co., Ltd. The numerical calculations in this study were carried out on the ORISE Supercomputer.

The authors declare that they have no competing interests.

References

[1] Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR, 2015.

[2] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

[3] Will Grathwohl, Ricky TQ Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. *arXiv preprint arXiv:1810.01367*, 2018.

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

[5] 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.

[6] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

[7] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64, 2021.

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

[9] Michael S Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. *arXiv preprint arXiv:2209.15571*, 2022.

[10] Michael Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209): 1–80, 2025.

[11] Yuan Gao, Jian Huang, and Yuling Jiao. Gaussian interpolation flows. *Journal of Machine Learning Research*, 25(253):1–52, 2024.

[12] Kenji Fukumizu, Taiji Suzuki, Noboru Isobe, Kazusato Oko, and Masanori Koyama. Flow matching achieves almost minimax optimal convergence. *arXiv preprint arXiv:2405.20879*, 2024.

[13] Lea Kunkel. Distribution estimation via flow matching with lipschitz guarantees. *arXiv preprint arXiv:2509.02337*, 2025.

[14] Shivam Kumar, Yixin Wang, and Lizhen Lin. Flow matching is adaptive to manifold structures. *arXiv preprint arXiv:2602.22486*, 2026.

[15] Xingchao Liu, Chengyue Gong, et al. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*.

[16] Teodora Reu, Sixtine Dromigny, Michael M Bronstein, and Francisco Vargas. Flows don't cross in high dimension. In *ICLR 2025 Workshop on Deep Generative Model in Machine Learning: Theory, Principle and Efficacy*, 2025.

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

[18] Philip Hartman. *Ordinary differential equations*. SIAM, 2002.

[19] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

[20] Jia Deng. A large-scale hierarchical image database. *Proc. of IEEE Computer Vision and Pattern Recognition, 2009*, 2009.

[21] 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.

[22] 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.

[23] Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. In *International conference on scale space and variational methods in computer vision*, pages 435–446. Springer, 2011.

[24] Laurent Younes. *Shapes and diffeomorphisms*, volume 171. Springer, 2010.

[25] Jens Behrmann, Paul Vicol, Kuan-Chieh Wang, Roger Grosse, and Jörn-Henrik Jacobsen. Understanding and mitigating exploding inverses in invertible neural networks. In *International Conference on Artificial Intelligence and Statistics*, pages 1792–1800. PMLR, 2021.

[26] Emilien Dupont, Arnaud Doucet, and Yee Whye Teh. Augmented neural odes. *Advances in neural information processing systems*, 32, 2019.

[27] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. 2023.

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

[29] Tianwei Yin, Michaël Gharbi, Richard Zhang, Eli Shechtman, Fredo 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*, pages 6613–6623, 2024.

[30] Kevin Frans, Danijar Hafner, Sergey Levine, and Pieter Abbeel. One step diffusion via shortcut models. *arXiv preprint arXiv:2410.12557*, 2024.

[31] 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.

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

[33] Ricky TQ Chen and Yaron Lipman. Flow matching on general geometries. *arXiv preprint arXiv:2302.03660*, 2023.

[34] Itai Gat, Tal Remez, Neta Shaul, Felix Kreuk, Ricky TQ Chen, Gabriel Synnaeve, Yossi Adi, and Yaron Lipman. Discrete flow matching. *Advances in Neural Information Processing Systems*, 37:133345–133385, 2024.

[35] Leon Klein, Andreas Krämer, and Frank Noé. Equivariant flow matching. *Advances in Neural Information Processing Systems*, 36:59886–59910, 2023.

[36] Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky TQ Chen. Multisample flow matching: Straightening flows with minibatch couplings. *arXiv preprint arXiv:2304.14772*, 2023.

[37] Pengsheng Guo and Alex Schwing. Variational rectified flow matching. In *International Conference on Machine Learning*, pages 20921–20940. PMLR, 2025.

[38] Sangyun Lee, Zinan Lin, and Giulia Fanti. Improving the training of rectified flows. *Advances in neural information processing systems*, 37:63082–63109, 2024.

[39] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pages 8162–8171. PMLR, 2021.

[40] 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.

A Notation

Table 4: Notation used throughout the paper.

Symbol	Meaning
Spaces and dimensions	
d	Ambient (data) dimension.
k	Subspace dimension used by Squeeze ($k \ll d$).
λ_d	d-dimensional Lebesgue measure on $\mathbb{R}^d$.
ω_m	Volume of the unit ball in $\mathbb{R}^m$.
Distributions and samples	
p_0, p_1	Source and target densities on $\mathbb{R}^d$.
$\mathcal{N}(0, I_d)$	Standard Gaussian source distribution.
$x_0, y_0 \sim p_0;\ x_1, y_1 \sim p_1$	Independent endpoint samples.
$\rho_t,\ \mu_t$	Densities of X_t on $\mathbb{R}^d$ and of PX_t on $\mathbb{R}^k$.
$\mathcal{D} = \{Y^{(j)}\}_{j=1}^N$	Real training dataset of size N.
Trajectories and crossings (Sec. 2)	
$\ell_{a,b}(t) = (1-t)a + tb$	Linear interpolant between $a, b \in \mathbb{R}^d$.
$\gamma(t)$	Candidate y_1 producing an exact crossing for fixed (x_0, x_1, y_0).
$A(x_0, x_1, y_0)$	Exact-crossing endpoint set $\{\gamma(t) : t \in (0,1)\}$.
$I = [\alpha, \beta] \subset (0,1)$	Compact interior time interval.
$\varepsilon > 0$	Local-crossing neighborhood radius.
$A^{\mathrm{loc}}_{I,\varepsilon}(x_0, x_1, y_0)$	Local-crossing endpoint set on I with tolerance ε.
Velocity field and flow (Sec. 3)	
$v_\theta(x, t)$	Parameterized velocity field, $v_\theta : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d$.
$\Phi_{s,t}$	Flow map sending $x_s \mapsto x_t$ along v_θ.
Ω_s, Ω_t	Regions with $\Omega_t = \Phi_{s,t}(\Omega_s)$.
$c_t : \Omega_t \to \mathbb{R}^m$	Smooth coloring at time t.
$c_s = c_t \circ \Phi_{s,t}$	Pulled-back coloring on Ω_s.
Squeeze method (Sec. 4)	
$v_\theta^{\mathrm{fm}},\ v_\theta^{\mathrm{sq}}$	Pretrained FM model and final Squeeze model.
$X_0^{(j)}, X_1^{(j)}$	Source and target endpoints of the j-th coupling pair.
$\sigma_0, \sigma_1 \in S_N$	Optimal permutations from the assignment problems.
$\mathcal{C}_{\mathrm{fm}}, \mathcal{C}_{\mathrm{sq}}$	FM-induced coupling and squeezed coupling.
$P_k \in \mathbb{R}^{d \times k}$	Top-k orthonormal basis of principal transport directions.
u_ϕ	Auxiliary low-dimensional flow on $\mathbb{R}^k$.
Marginal-preservation theorem (Theorem 1)	
$X_t,\ \dot{X}_t$	Random variable on $\mathbb{R}^d$ and its time derivative.
$P \in \mathbb{R}^{d \times k}$	Projection matrix; equivalent to P_k.
$v_t(\xi) = \mathbb{E}[\dot{x}_t \mid x_t = \xi]$	Conditional projected velocity field on $\mathbb{R}^k$.
z_t	Solution of $\dot{z}_t = v_t(z_t)$ with $z_0 = P^\top X_0$.
$C_c^1(\mathbb{R}^k)$	Compactly supported C^1 functions on $\mathbb{R}^k$.

B Pseudocode of Squeeze

Algorithm 1 Squeeze

Require: Trained FM model v_θ^{fm}, real dataset $\mathcal{D} = \{Y^{(j)}\}_{j=1}^N$, subspace dimension k
Ensure: Squeezed model v_θ^{sq}

// Notation. S_N: symmetric group on $[N]$ (set of all permutations $\sigma : [N] \to [N]$)
// Squeeze
$X_0^{(j)} \sim \mathcal{N}(0, I_d)$; $\sigma_0 \leftarrow \arg\min_{\sigma \in S_N} \sum_j \big\|\mathrm{ODE}(v_\theta^{\mathrm{fm}},\, X_0^{(j)}) - Y^{(\sigma(j))}\big\|_2^2$; $X_1^{(j)} \leftarrow Y^{(\sigma_0(j))}$; form $\mathcal{C}_{\mathrm{fm}} = \big\{(X_0^{(j)},\, X_1^{(j)})\big\}_{j=1}^N$;
SVD of centered transport vectors $X_1^{(j)} - X_0^{(j)}$ and $X_0^{(j)} - X_1^{(j)}$; retain top-k directions $P_k \in \mathbb{R}^{d\times k}$
Train FM on projected targets $\big\{P_k^\top X_1^{(j)}\big\}_{j=1}^N$ in $\mathbb{R}^k \;\to$ obtain u_ϕ
// Lift
$z_1^{(j)} \leftarrow \mathrm{ODE}\big(u_\phi,\, P_k^\top X_0^{(j)}\big)$ for each j
$\sigma_1 \leftarrow \arg\min_{\sigma \in S_N} \sum_j \big\|z_1^{(j)} - P_k^\top X_1^{(\sigma(j))}\big\|_2^2$; form $\mathcal{C}_{\mathrm{sq}} = \big\{(X_0^{(j)},\, X_1^{(\sigma_1(j))})\big\}_{j=1}^N$
Fine-tune v_θ^{fm} on $\mathcal{C}_{\mathrm{sq}} \;\to\; v_\theta^{\mathrm{sq}}$
return v_θ^{sq}

C Proof

C.1 Proofs for the probability of exact crossing

Proof. For fixed (x_0, x_1, y_0). An exact same-time crossing means that there exists $t \in (0,1)$ such that $\ell_{x_0,x_1}(t) = \ell_{y_0,y_1}(t)$. It implies that there exists $t \in (0,1)$ satisfies

$$y_1 = \frac{(1-t)x_0 + tx_1 - (1-t)y_0}{t} = x_1 + \left(\frac{1}{t} - 1\right)(x_0 - y_0).$$

Thus the set of all endpoints y_1 that produce exact same-time crossings, which is defined as $A(x_0, x_1, y_0)$, contained in an affine line in $\mathbb{R}^d$. If $x_0 \neq y_0$, it is contained in a half-line; if $x_0 = y_0$, it reduces to the single point x_1. Therefore, for $d \geq 2$, $\lambda_d\big(A(x_0, x_1, y_0)\big) = 0$, it implies that

$$\mathbb{P}\big(y_1 \in A(x_0, x_1, y_0)\big) = \int_{A(x_0,x_1,y_0)} p_1(y)\, dy = 0.$$

□

C.2 Proof of Theorem 1

Proof. Let $h \in C_c^1(\mathbb{R}^k)$ be an arbitrary compactly supported continuously differentiable test function. Since x_t is path-wise continuously differentiable, the chain rule gives

$$\frac{d}{dt}\mathbb{E}[h(x_t)] = \mathbb{E}[\nabla h(x_t)^\top \dot{x}_t].$$

Using the defining property of conditional expectation and the fact that $\nabla h(x_t)$ is measurable with respect to x_t, we obtain

$$\mathbb{E}[\nabla h(x_t)^\top \dot{x}_t] = \mathbb{E}\big[\nabla h(x_t)^\top \mathbb{E}[\dot{x}_t \mid x_t]\big] = \mathbb{E}[\nabla h(x_t)^\top u^P(x_t, t)].$$

For convenience, we define $\mu_t := \mathrm{Law}(x_t) = \mathrm{Law}(P^\top X_t)$, then μ_t solves the continuity equation

$$\partial_t \mu_t + \nabla \cdot (u^P(\cdot, t)\mu_t) = 0.$$

Moreover, let $\nu_t := \mathrm{Law}(z_t)$, since z_t solves $dz_t/dt = u^P(z_t, t)$, the same calculation obtains that ν_t solves the same continuity equation

$$\partial_t \nu_t + \nabla \cdot (u^P(\cdot, t)\nu_t) = 0.$$

Moreover, the two laws have the same initial condition:

$$\nu_0 = \mathrm{Law}(z_0) = \mathrm{Law}(P^\top X_0) = \mu_0.$$

By the same method as illustrated in rectified flows, the uniqueness of the distributional solution to this continuity equation implies that

$$\nu_t = \mathrm{Law}(z_t) = \mathrm{Law}(P^\top X_t) = \mu_t, \quad \forall t \in [0,1].$$

□

D Local crossing number experimental data

Table 5: Local crossing number on CIFAR-10 target under different segment-segment minimum distance thresholds d.

Dim	Threshold d																
	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5	1.6
3072	0	0	0	0	0	0	0	1	3	8	52	276	1005	3177	8345	20188	44293
768	0	0	0	0	3	59	1067	8802	46348	176905	532393	1343994	2945945	5695532	9813887	15193137	21373809
192	0	0	9	2138	67736	672613	3400046	10542173	21837871	33407597	41931463	46790568	48998977	49771985	49962286	49992250	49994885
96	0	2	1678	171096	2492919	12369970	28645648	41596997	47711337	49592288	49954176	49992873	49994931	49995000	49995000	49995000	49995000
48	0	119	163311	5051068	24150309	42163548	48698664	49882975	49990491	49994931	49995000	49995000	49995000	49995000	49995000	49995000	49995000
24	32	13254	4877546	31790539	47684911	49886517	49993491	49994998	49995000	49995000	49995000	49995000	49995000	49995000	49995000	49995000	49995000

Table 6: Local crossing number on Gaussian target under different segment-segment minimum distance thresholds d.

Dim	Threshold d																
	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5	1.6
3072	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
768	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
192	0	0	0	0	0	0	0	0	0	0	0	1257	272760	6962151	31623920	47921680	49948726
96	0	0	0	0	0	0	0	1833	391868	8675905	34204435	48396291	49958840	49994793	49995000	49995000	49995000
48	0	0	0	0	519	241709	7214143	32268821	47965908	49938597	49994341	49994991	49995000	49995000	49995000	49995000	49995000
24	0	0	392	324100	9038191	34851130	48419628	49953224	49994614	49994998	49995000	49995000	49995000	49995000	49995000	49995000	49995000

E Implementation details

E.1 Training hyperparameters

We train three flow models per dataset: the standard online-OT Flow Matching (**FM**) baseline, its Rectified Flow (**RF**) finetune, and our **Squeeze** module — a low-capacity MLP that operates entirely inside a 64-d latent of the data distribution. On CIFAR-10/100 the FM/RF backbone is the U-Net described in E.2 (Table7); on ImageNet (256×256 SD-VAE latent space) it is the DiT-B/4 of E.3 (Table8).

Table 7: CIFAR-10 / CIFAR-100 training hyperparameters. "–" indicates not applicable. All optimizers are AdamW.

	CIFAR-10			**CIFAR-100**		
	FM	RF	Squeeze	FM	RF	Squeeze
Backbone	U-Net (55.7 M)	U-Net	ResMLP-h64 (93 504)	U-Net (55.7 M)	U-Net	ResMLP-h64 (93 504)
Input size	$32\times32\times3$	$32\times32\times3$	64	$32\times32\times3$	$32\times32\times3$	64
Epochs	500	10	10	500	10	10
Global batch size	256 (32×8)	256 (32×8)	256	256 (32×8)	256 (32×8)	256
Optimizer	AdamW	AdamW	AdamW	AdamW	AdamW	AdamW
Peak LR	5×10^{-4}	1×10^{-4}	1×10^{-4}	5×10^{-4}	1×10^{-4}	1×10^{-4}
LR warmup steps	45 000	—	—	45 000	—	—
LR schedule	cosine $\to 0$	cosine $\to 0$	cosine $\to 0$	cosine $\to 0$	cosine $\to 0$	cosine $\to 0$
Sampling steps	100 (Euler)	100 (Euler)	50 (Euler)	100 (Euler)	100 (Euler)	50 (Euler)
Seed	42	42	42	42	42	42

E.2 U-Net architecture (CIFAR)

For both CIFAR-10 and CIFAR-100 we use the standard ADM/improved-DDPM U-Net [39], configured for $32\times32\times3$ inputs: 3-input/output channels, base width 128, channel multipliers $[2, 2, 2]$

Table 8: ImageNet training hyperparameters. FM and RF operate in the SD-VAE latent space (`stabilityai/sd-vae-ft-ema`, $4\times32\times32$ latent); the Squeeze module operates in a 64 of those latents. "–" indicates not applicable. All optimizers are AdamW.

	FM	**RF**	**Squeeze**
Backbone	DiT-B/4 (130 M)	DiT-B/4	ResMLP-h64 (93 504)
Input size	$32\times32\times4$	$32\times32\times4$	64
Total optimizer steps	400 000	10 000	10 000
Global batch size	256	256	256
Optimizer	AdamW	AdamW	AdamW
Peak LR	1×10^{-4}	2×10^{-5}	1×10^{-3}
LR warmup steps	—	—	—
LR schedule	constant	constant	cosine $\to 0$
Sampling steps	250 (DDPM)	250 (DDPM)	50 (Euler)
Seed	42	42	0

across three resolution levels (so 256 channels at every level), and 4 residual blocks per level. Self-attention is inserted at the second resolution level (i.e. 16×16) with a single head and "new attention order" (channels-first QKV). The time embedding is a 128-dim sinusoidal vector projected to 512. We do not use label conditioning, Fourier features, or learned up/down convolutions. The total parameter count is **55.68 M**.

The same architecture is used for both FM and RF on CIFAR; only the training objective and coupling differ.

E.3 DiT architecture (ImageNet)

For ImageNet (256×256) we operate in the SD-VAE latent space [40]: each image is encoded by the frozen `stabilityai/sd-vae-ft-ema` encoder to a $4\times32\times32$ latent map. The flow backbone is a **DiT-B/4**: 12 transformer blocks, hidden size 768, 12 attention heads, MLP expansion ratio 4, patch size 4×4 (yielding $8\times8 = 64$ tokens per latent).

E.4 Hardware environment

All experiments are run on a server with $8\times$ NVIDIA A100 80 GB PCIe GPUs, dual Intel Xeon Gold 5318Y CPUs (24 cores each, 2.10 GHz base), and 512 GB DDR4 system memory. The seed settings above describe the original runs; the five-seed results in Table 1 use seeds 0–4.

F Additional experimental results

Neighbourhood statistics. Table 9 reports tearing (source neighbours becoming separated) and folding (target neighbours that were separated at the source) on the real-data settings. These finite-sample diagnostics complement the coloring plots. The geometric measurements and few-step generation gains are observed together; the experiments do not isolate a causal contribution of geometry to generation quality.

Table 9: Tear and fold rates. Lower is better.

Dataset	Tear (RF)	Tear (Squeeze)	Fold (RF)	Fold (Squeeze)
CIFAR-10	0.1715	0.0010	0.1587	0.0015
CIFAR-100	0.3040	0.0080	0.1935	0.0015
ImageNet	0.2742	0.0040	0.1320	0.0037

Variation across seeds. Table 10 gives the mean and standard deviation over seeds 0–4 for the CIFAR results in Table 1.

Table 10: CIFAR generation results: mean $\pm$ standard deviation over five seeds. Recall is in percent.

NFE	Method	CIFAR-10			CIFAR-100		
		FID $\downarrow$	IS $\uparrow$	Recall $\uparrow$	FID $\downarrow$	IS $\uparrow$	Recall $\uparrow$
1	FM	355.92 ± 3.90	1.03 ± 0.00	0.02 ± 0.00	349.02 ± 2.45	1.02 ± 0.01	0.02 ± 0.00
	RF	135.51 ± 3.47	2.29 ± 0.03	17.64 ± 0.58	125.57 ± 2.28	2.62 ± 0.04	23.05 ± 1.08
	OT-FM	224.14 ± 1.95	2.22 ± 0.09	0.15 ± 0.05	240.69 ± 2.00	2.28 ± 0.06	0.32 ± 0.04
	Squeeze	79.27 ± 1.71	3.51 ± 0.04	38.77 ± 0.25	76.14 ± 2.16	4.04 ± 0.10	41.41 ± 1.05
2	FM	191.97 ± 3.05	1.48 ± 0.02	6.01 ± 0.14	200.13 ± 3.08	1.54 ± 0.00	8.03 ± 0.61
	RF	38.62 ± 0.59	4.57 ± 0.04	58.22 ± 0.74	46.11 ± 0.71	4.78 ± 0.04	55.76 ± 0.25
	OT-FM	95.66 ± 1.40	5.09 ± 0.06	6.23 ± 0.21	92.76 ± 1.77	4.85 ± 0.04	18.33 ± 0.36
	Squeeze	23.59 ± 0.43	5.90 ± 0.02	77.05 ± 0.71	30.87 ± 0.77	6.07 ± 0.06	69.25 ± 1.07
4	FM	57.36 ± 0.07	3.50 ± 0.04	38.72 ± 0.56	60.36 ± 0.90	3.48 ± 0.03	40.73 ± 1.11
	RF	13.27 ± 0.11	6.64 ± 0.02	85.01 ± 0.20	20.83 ± 0.12	6.64 ± 0.02	76.92 ± 0.19
	OT-FM	32.23 ± 0.46	7.38 ± 0.05	58.12 ± 0.33	37.97 ± 0.59	6.83 ± 0.03	40.14 ± 0.38
	Squeeze	11.35 ± 0.23	7.23 ± 0.02	89.16 ± 0.41	17.31 ± 0.23	7.43 ± 0.11	81.57 ± 0.32
200	FM	3.11 ± 0.04	8.53 ± 0.01	96.02 ± 0.25	7.19 ± 0.42	8.68 ± 0.17	90.00 ± 0.06
	RF	7.76 ± 0.15	8.40 ± 0.05	93.08 ± 0.50	12.27 ± 0.06	8.82 ± 0.12	85.79 ± 0.20
	OT-FM	6.06 ± 0.13	8.89 ± 0.11	66.69 ± 1.11	12.16 ± 0.19	8.14 ± 0.07	53.88 ± 0.26
	Squeeze	6.54 ± 0.21	8.36 ± 0.06	93.78 ± 0.20	11.34 ± 0.33	8.90 ± 0.04	87.21 ± 0.39

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction accurately reflect the paper's contributions and scope.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The limitations are discussed in the paper.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: Full assumptions and proofs are provided.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: All information needed to reproduce the main results is provided.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: The code, dataset, and model weights are not publicly released at submission time, but we commit to releasing the full data, models, and code within two months.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: All training and test details are specified.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Table 1 reports five-seed means, and Table 10 reports the corresponding standard deviations over seeds 0–4. These error bars apply to CIFAR-10 and CIFAR-100; no formal significance test is claimed.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: All compute resources are reported.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conforms with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: There is no societal impact of the work performed.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: the paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Existing assets are cited with original papers.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [No]

    Justification: We plan to release the code with documentation upon acceptance.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.