Balanced Conic Rectified Flow

Shin seong Kim Mingi Kwon Jaeseok Jeong Youngjung Uh

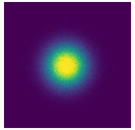
Yonsei University

VILAB
VISUAL INTELLIGENCE LAB

NEURAL INFORMATION PROCESSING SYSTEMS

Preliminaries: Flow matching for generative modeling

Source



Target

For time-dependent vector field,

$$v : [0,1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$$

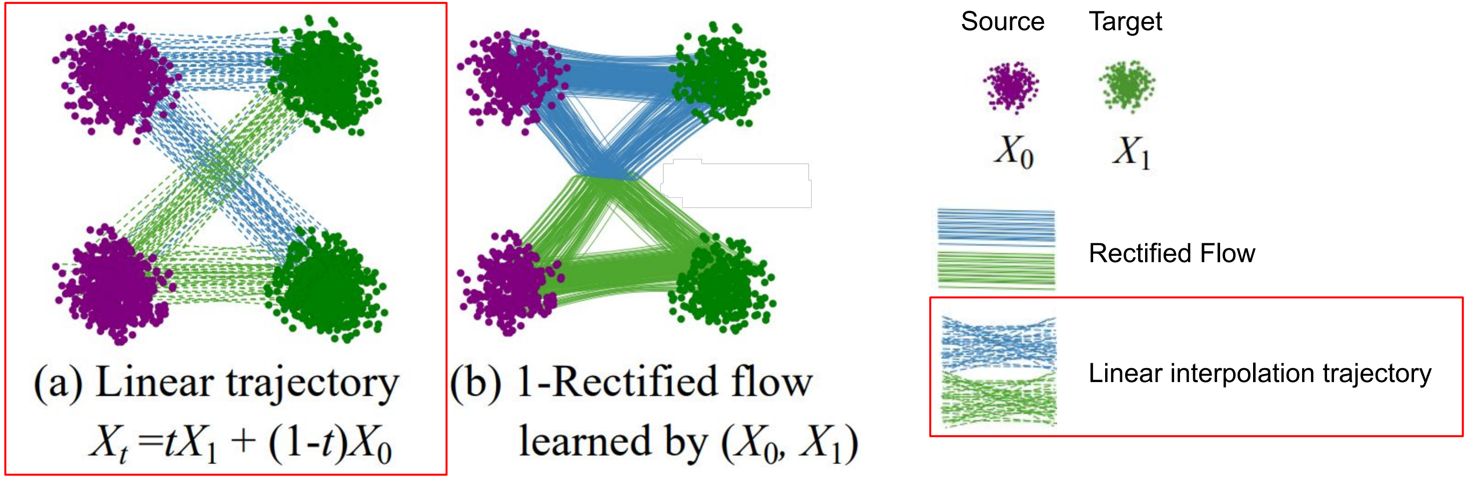
Vector field can construct time-dependent diffeomorphic map (flow) $\phi : [0,1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ defined via the ordinary differential equations (ODE) such that,

$$\frac{d}{dt}\phi_t(x) = v_t(\phi_t(x))$$

$$\phi_0(x) = x$$

Figure reference : https://mlg.eng.cam.ac.uk/blog/2024/01/20/flow-matching.html

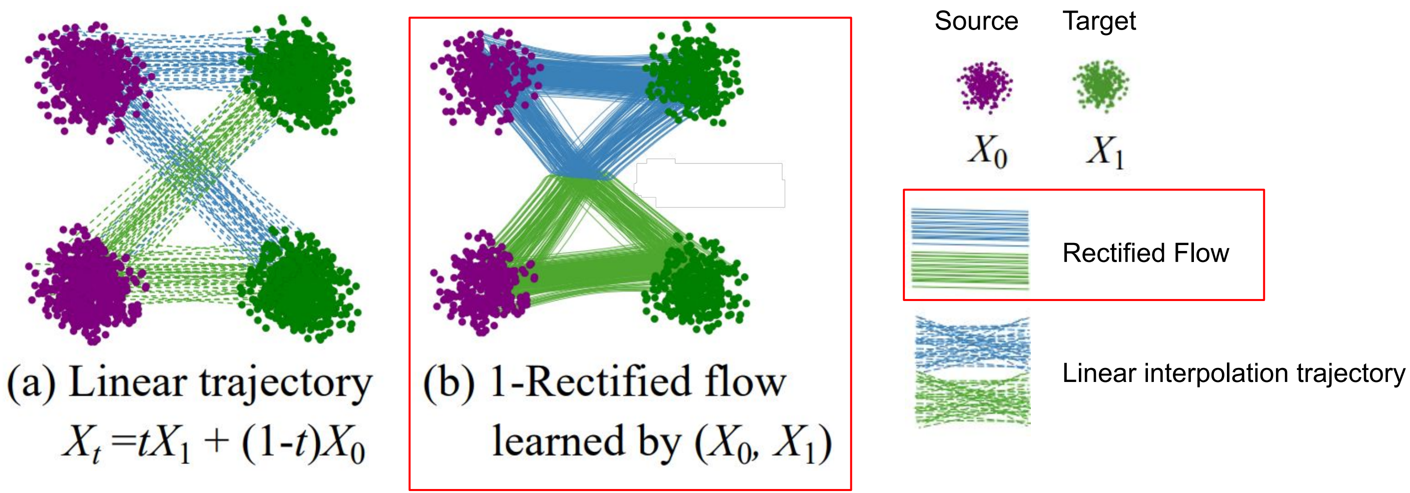
Preliminaries : Rectified flow



Consider the linear interpolation path between source and target

$$X_t = (1-t)X_0 + tX_1 \text{ for } t \in [0,1]$$

Preliminaries : Rectified flow



(a) Linear trajectory $X_t = tX_1 + (1\text{-}t)X_0$

(b) 1-Rectified flow learned by (X_0, X_1)

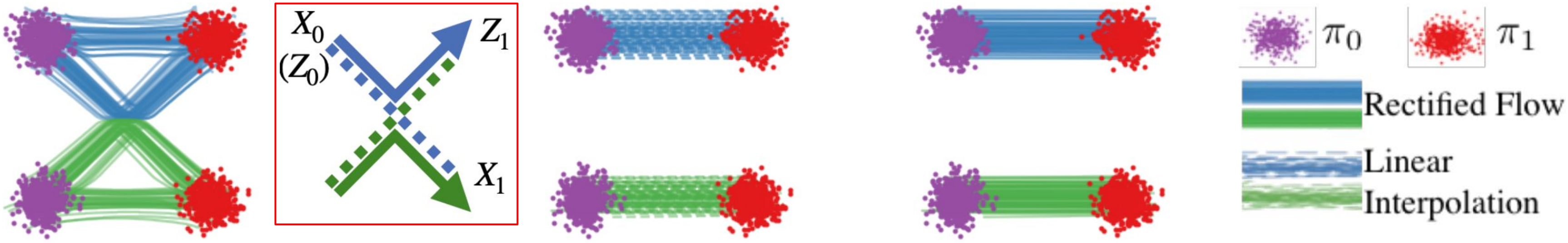
Use the time derivative of linear interpolation as supervision

$$\frac{dZ_t}{dt} = v_\theta(Z_t, t) := \frac{1}{t}(Z_t - \mathbb{E}[(X_1 - X_0)|X_t = Z_t])$$

Train the velocity field using an ODE neural network

$$\arg\min_\theta \mathbb{E}\left[\left\|X_1 - X_0 - v(tX_1 + (1-t)X_0, t)\right\|^2\right]$$

Preliminaries : Reflow process



(b) Rectified flow Z_t induced by (X_0, X_1)

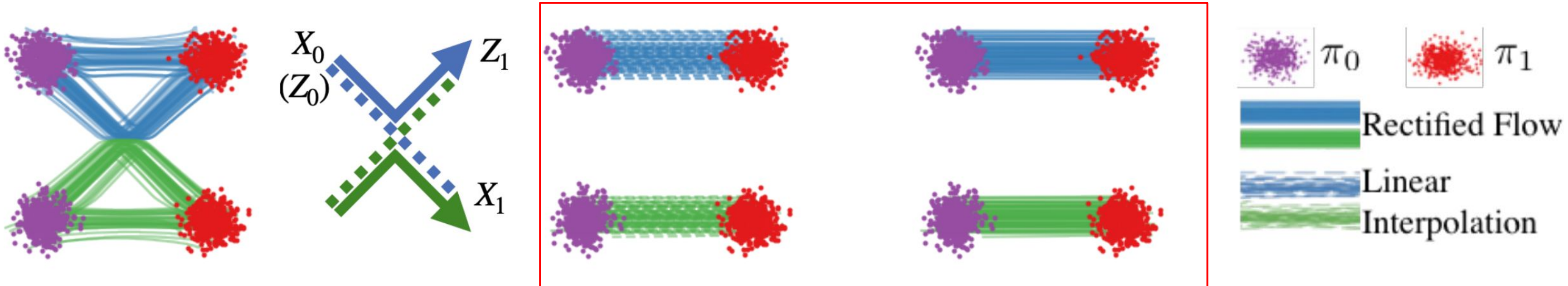
(c) Linear interpolation $Z_t = tZ_1 + (1-t)Z_0$

(d) Rectified flow Z'_t induced by (Z_0, Z_1)

Generate Fake pairs $(Z_{0,F}, Z_{1,F})$ using 1-Rectified models v_θ

Figure reference : Liu et al., "Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow," https://arxiv.org/abs/2209.03003 (2023).

Preliminaries : Reflow process



(b) Rectified flow Z_t induced by (X_0, X_1)

(c) Linear interpolation $Z_t = tZ_1 + (1-t)Z_0$

(d) Rectified flow Z'_t induced by (Z_0, Z_1)

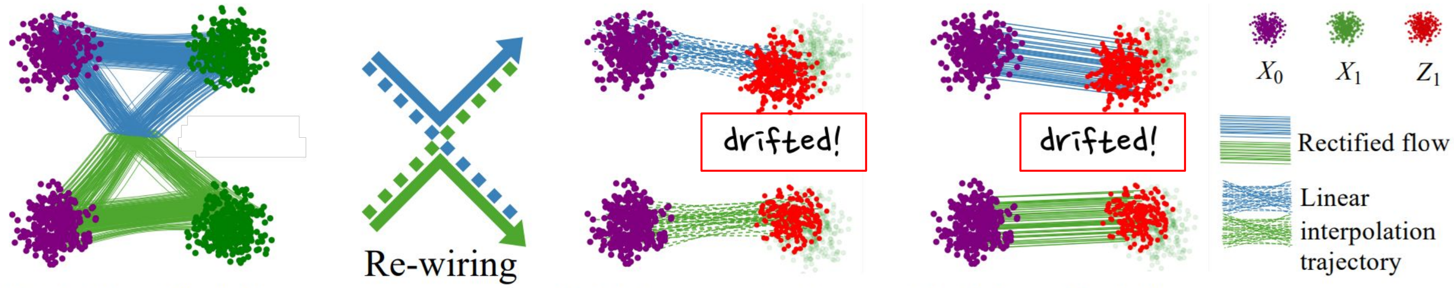
Generate Fake pairs $(Z_{0,F}, Z_{1,F})$ using 1-Rectified models v_θ

Fine tuning with 1-Rectified models using fake pairs

$$\arg\min_\theta \mathbb{E}\left[\|Z_{1,F} - Z_{0,F} - v_\theta(tZ_{1,F} + (1-t)Z_{0,F})\|^2\right],$$

Figure reference : Liu et al., "Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow," https://arxiv.org/abs/2209.03003 (2023).

Original Reflow process drift target distribution



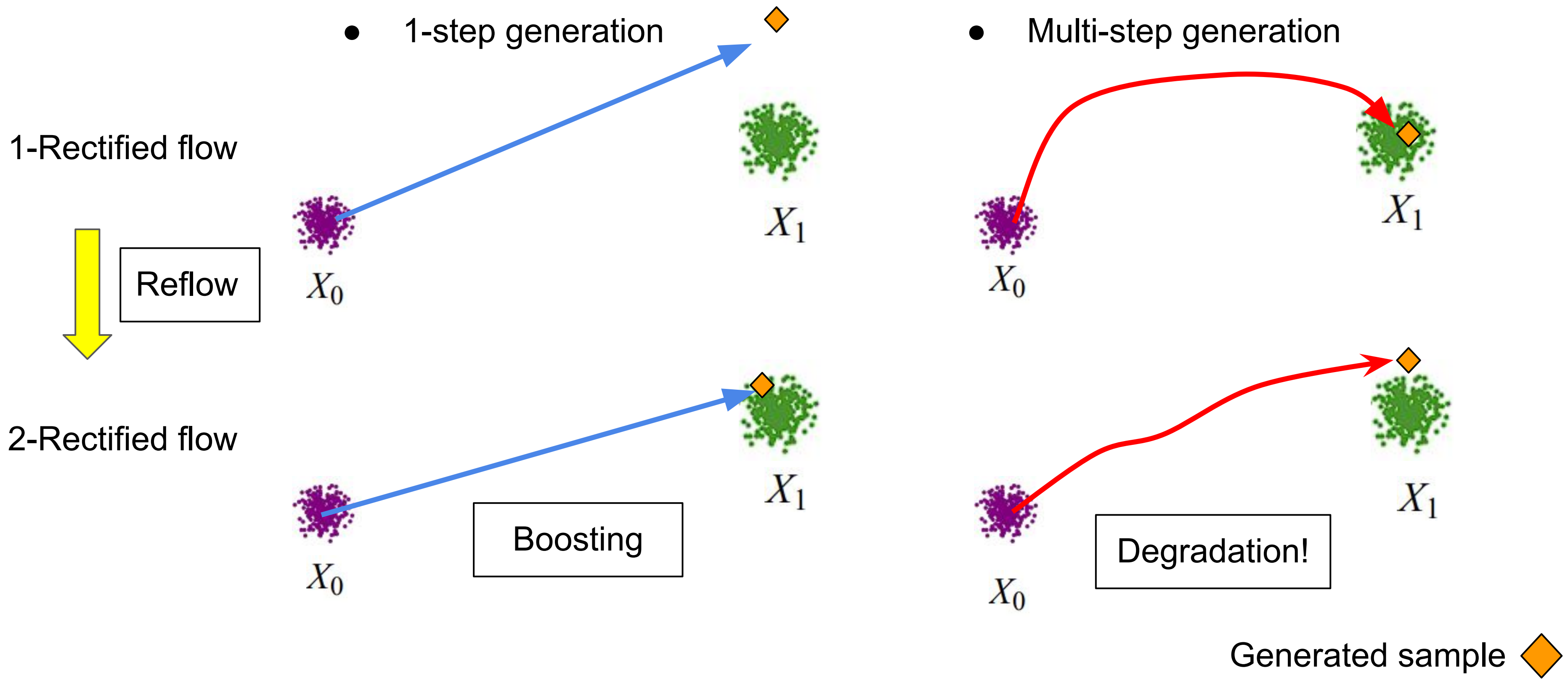
(b) 1-Rectified flow learned by (X_0, X_1)

(c) Linear trajectory $Z_t = tZ_1 + (1-t)Z_0$

(d) 2-Rectified flow learned by (Z_0, Z_1)

- Since 1-Rectified Flow cannot perfectly match the target distribution, Reflow using only fake pairs becomes biased toward generated samples.

Distribution drift degrades the quality of full-step generation

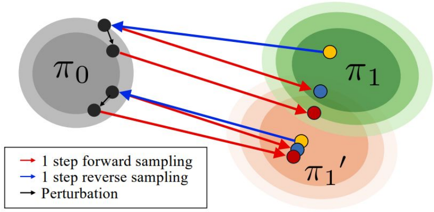


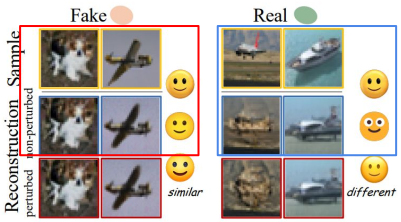
- Boosting 1-step generation quality (Blue)
- Degradation Multi-step generation quality (Red)

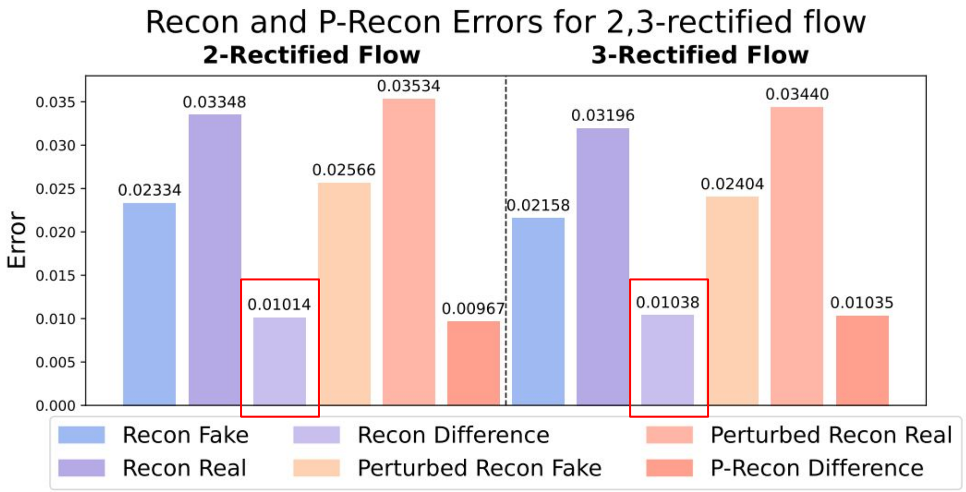
Drift of target distribution in the original Reflow process

Reconstruction Error

$$L_2^{\text{recon}}(X) = \mathbb{E}_{x \sim X}\left[\|x - v(v^{-1}(x))\|_2\right]$$





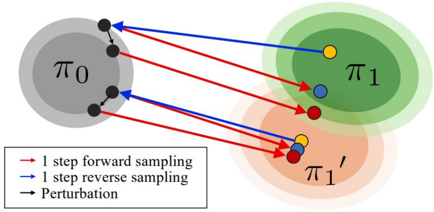


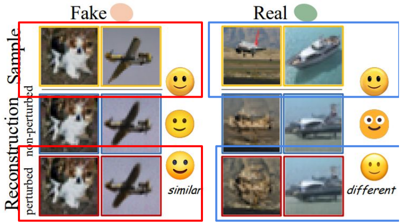
- Experimental results show a clear discrepancy in reconstruction and perturbed reconstruction errors between real and generated images

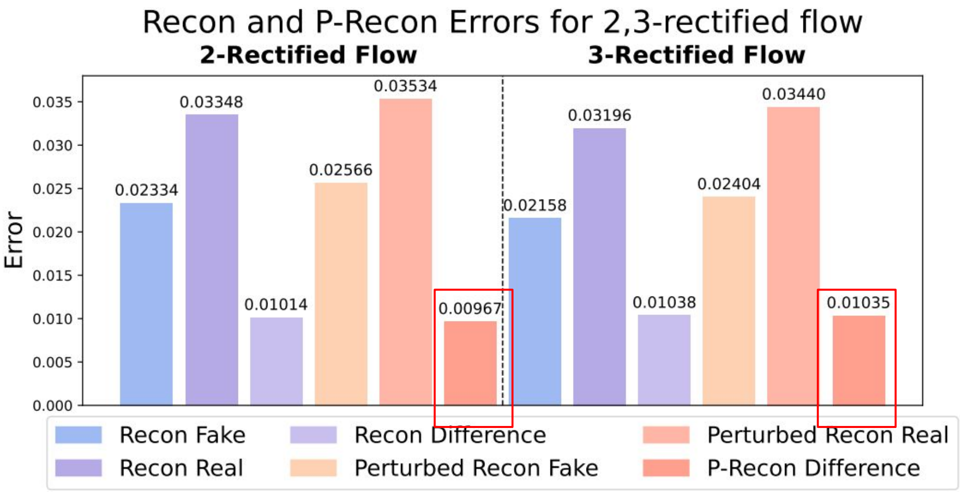
Drift of target distribution in the original Reflow process

Perturbed Reconstruction Error

$$L_2^{\text{p-recon}}(X, \varepsilon) = \mathbb{E}_{x \sim X, z \sim \pi_0} \|x - v(v^{-1}(x) + \varepsilon z)\|_2,$$







- Experimental results show a clear discrepancy in reconstruction and perturbed reconstruction errors between real and generated images

Drift of target distribution in the original Reflow process

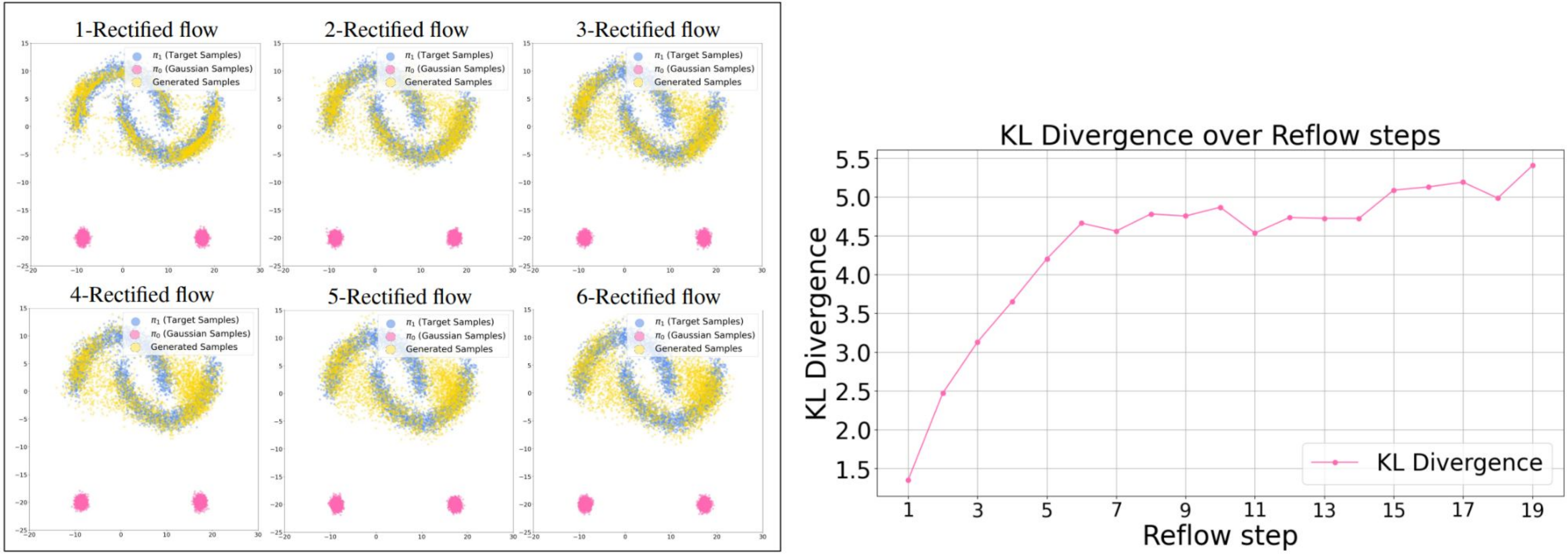
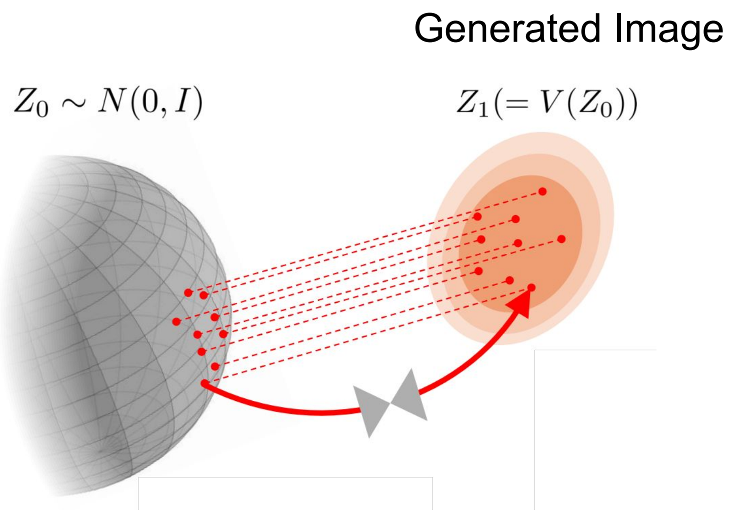
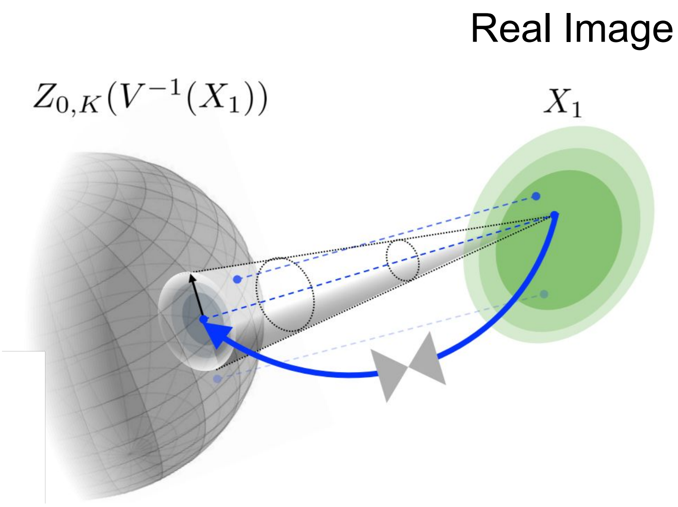


Figure 3: **(a)** As reflow steps increase, generated samples diverge from the target distribution. **(b)** This drift is further evidenced by the rising KL divergence from the real data distribution.

Fake pair and Real Pair





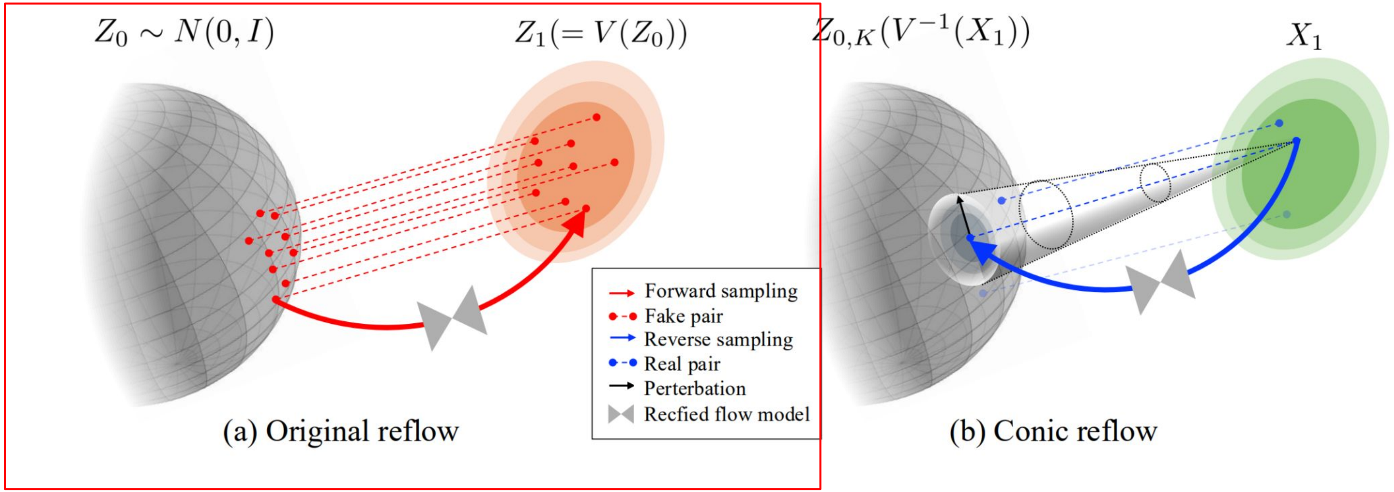
- Fake Pair

$$(Z_0, v_\theta(Z_0)) := (Z_{0,F}, Z_{1,F})$$

- Real Pair

$$(Z_{0,R}, X_1) := (v^{-1}(X_1), X_1)$$

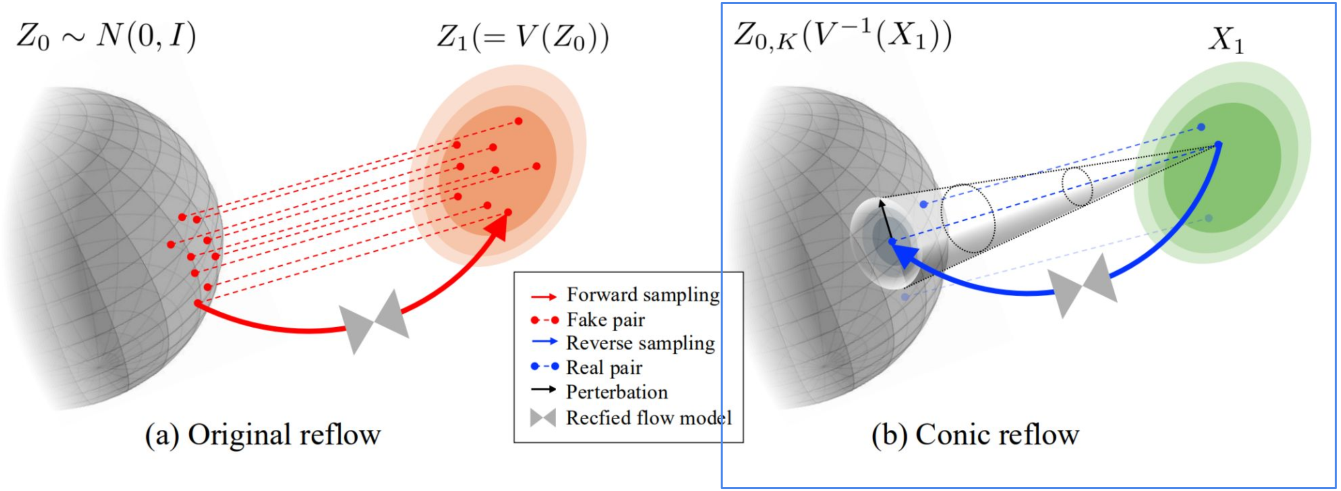
Methods - Conic Reflow



Original reflow for Fake pairs

$$\arg\min_{\theta} \mathbb{E}\left[\|Z_{1,F} - Z_{0,F} - v_\theta(tZ_{1,F} + (1-t)Z_{0,F})\|^2\right]$$

Methods - Conic Reflow



Perturbed based supervision for Real pairs

$$\hat{\theta} = \arg\min_{\theta} \int_0^1 \mathbb{E}\left[w_t \left\| X_1 - \mathrm{slerp}(Z_{0,R}, \epsilon, \zeta) - v_\theta\big(\mathrm{Conic}(X_1, \epsilon, \zeta, t)\big) \right\|^2 \right] dt$$

where Conic operator defined as,

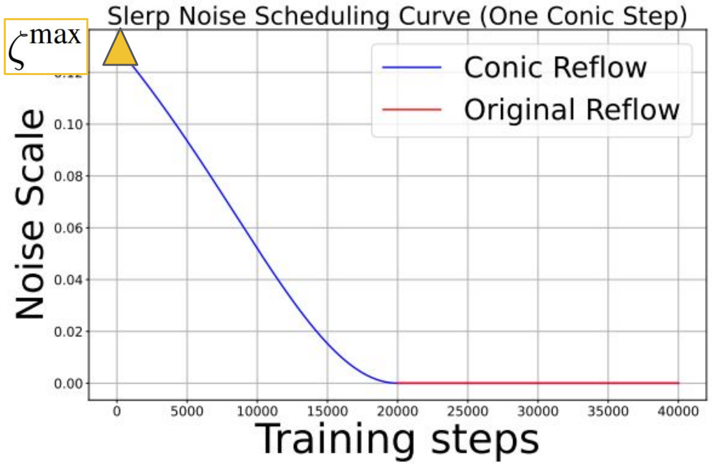
$\mathrm{Conic}(X_1, \epsilon, \zeta, t) = tX_1 + (1-t)\mathrm{slerp}(Z_{0,R}, \epsilon, \zeta)$, $\epsilon \sim \mathcal{N}(0, I)$, ζ: Perturbation intensity via Slerp

Maximum perturbation intensity and noise schedule

Adaptively compute the maximum perturbation intensity

$$\zeta^{\max} := \max_{\zeta \in (0,0.5]} \mathbb{E}_{\substack{x \sim X_1 \\ z \sim Z_{1,F}}} \left[\left\| v_\theta \left(\mathrm{Slerp}(z_{0,R}, \epsilon, \zeta) \right) - x \right\|_2 - \left\| v_\theta \left(\mathrm{Slerp}(z_{0,F}, \epsilon, \zeta) \right) - z \right\|_2 \right]$$

where $\epsilon \sim \mathcal{N}(0, I)$, with $z_{0,R} = v_\theta^{-1}(x)$ and $z_{0,F} = v_\theta^{-1}(z_{1,F})$



The perturbation intensity progressively decrease during training

$$\zeta(t') := \zeta^{\max} \cdot \frac{2t'^2}{1+t'^2}, \quad t' \in [0, 1]$$

Comparison with baselines

Method	NFE (↓)	IS (↑)	FID (↓)
One-Step Generation (Euler solver, N=1)			
1-Rectified Flow	1	1.13 (9.08)	378 (6.18)
2-Rectified Flow			
Original (+*Distill*)	1	8.08 (9.01)	12.21 (4.85)
Ours (+*Distill*)	1	**8.79 (9.11)**	**5.98 (4.16)**
Rf++† [24]	1	8.87	4.43
Rf++† (**+ours**)	1	8.87	**4.22**
3-Rectified Flow			
Original (+*Distill*)	1	8.47 (8.79)	8.15 (5.21)
Ours (+*Distill*)	1	**8.84 (8.96)**	**5.48 (4.68)**
Full Simulation (Runge–Kutta (RK45), Adaptive N)			
1-Rectified Flow	127	**9.60**	**2.58**
2-Rectified Flow			
Original	110	9.24	3.36
Ours	104	**9.30**	**3.24**
3-Rectified Flow			
Original	104	9.01	3.96
Ours	98	**9.14**	**3.70**

Table 1: One-step and full-simulation comparison of 2,3 Rectified Flows on CIFAR-10.

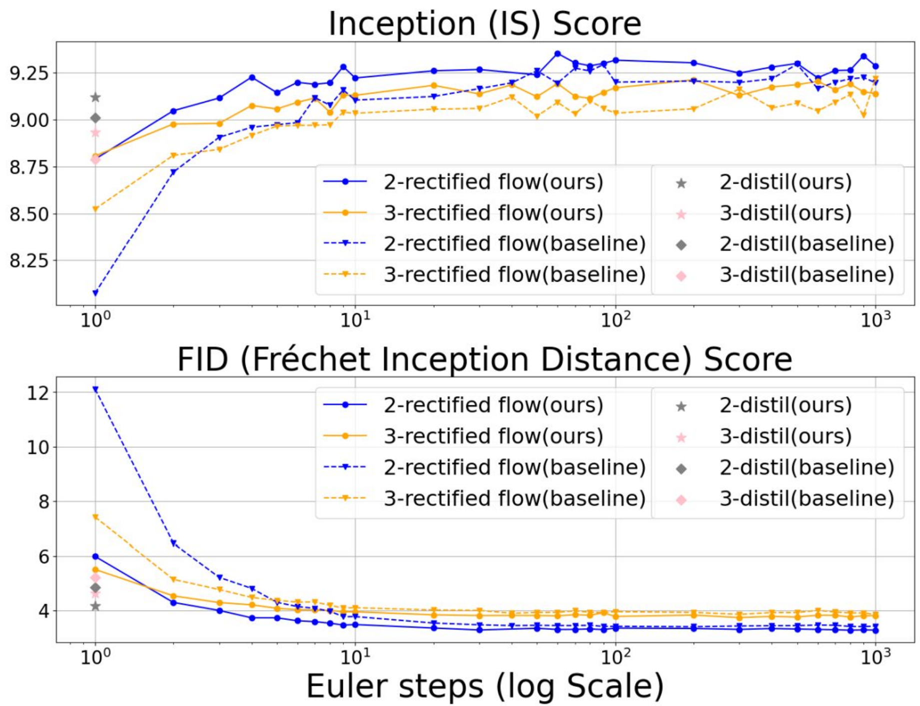
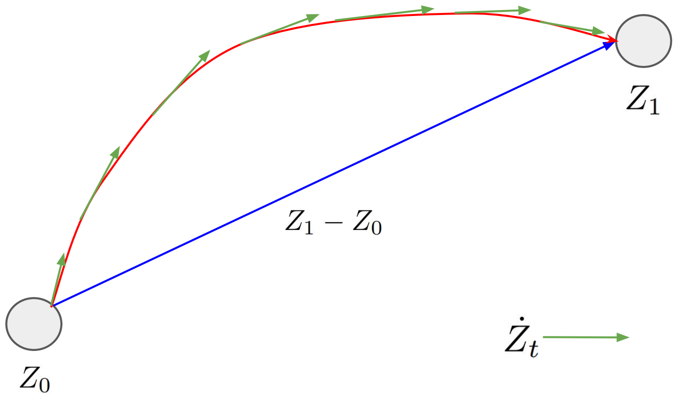


Figure 6: CIFAR-10 generation quality across Euler steps.

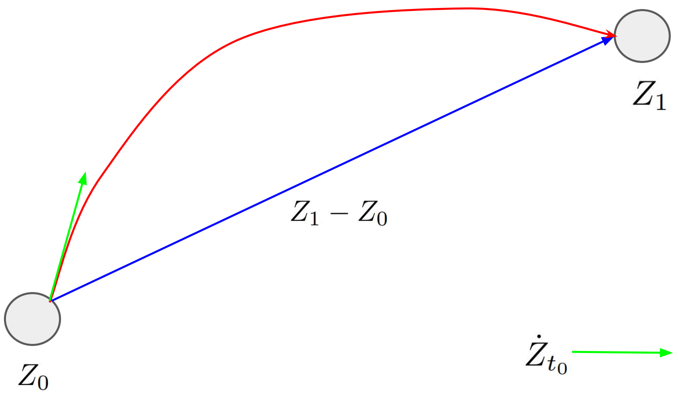
Curvature and Initial Velocity Delta (IVD)

- Curvature



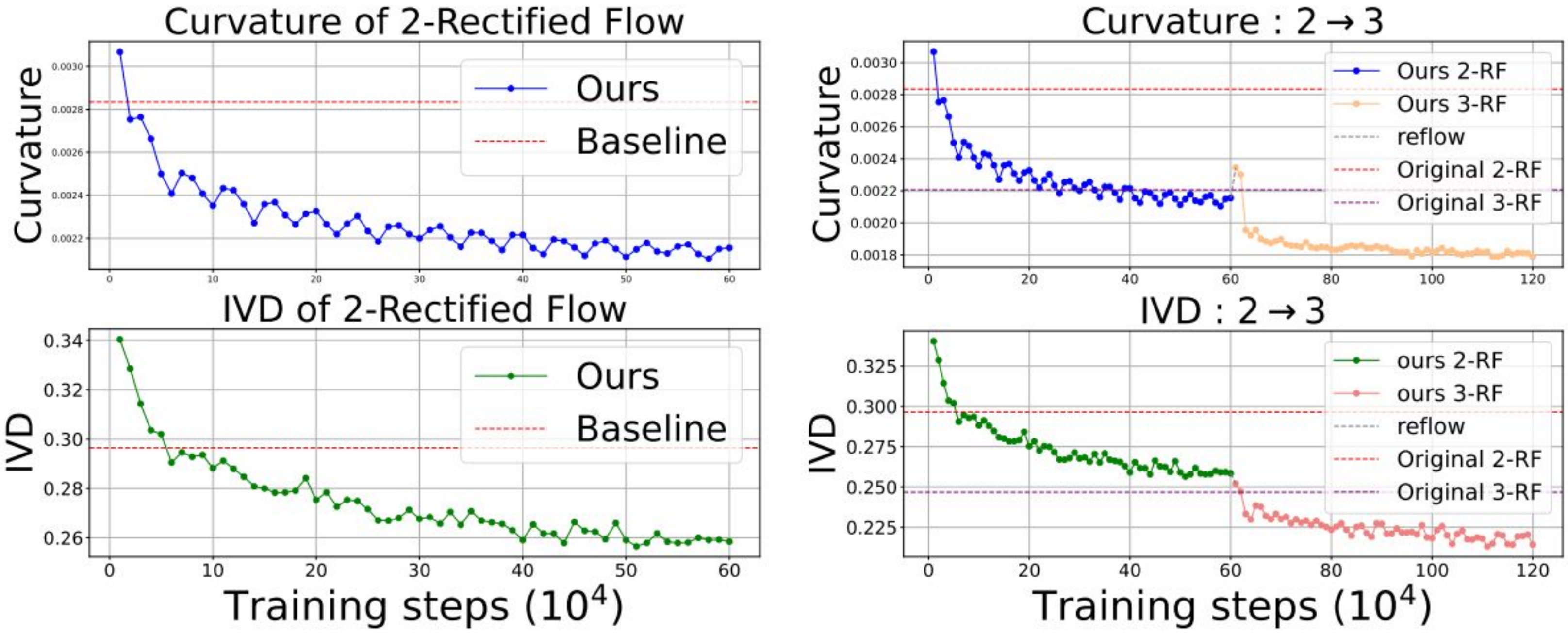
$$S(\mathbf{Z}) = \int_0^1 \mathbb{E}\left[\left\|(Z_1 - Z_0) - \dot{Z}_t\right\|^2\right] dt$$

- Initial Velocity Delta (IVD)



$$IVD(\mathbf{Z}, t_0) = \mathbb{E}\left[\left\|(Z_1 - Z_0) - \dot{Z}_{t_0}\right\|^2\right]$$

Straightness and Initial Velocity Delta (IVD)



Our methods reduces rap betweed real and fake image during reflow process

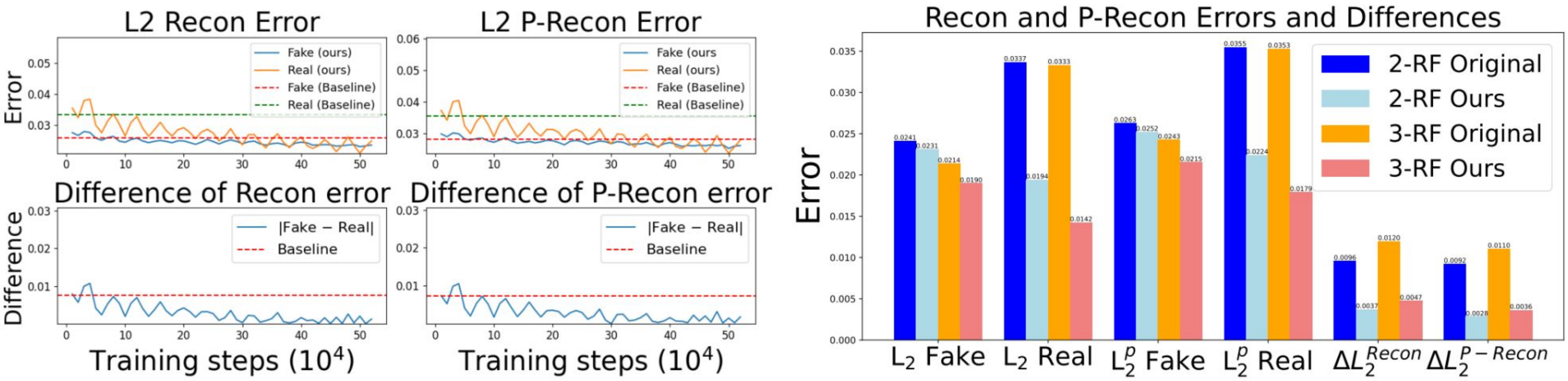


Figure 7: Reconstruction and perturbed reconstruction error across training iterations.

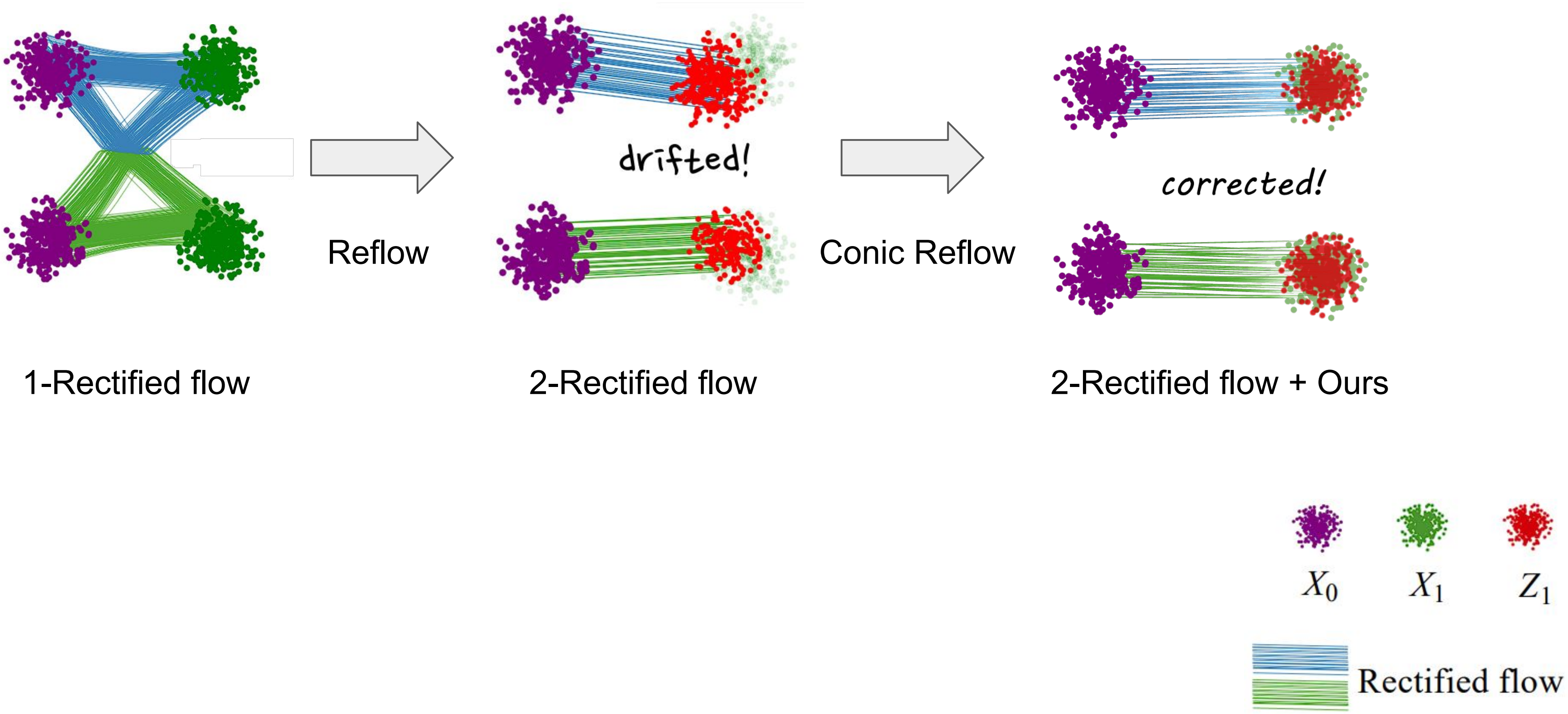
High resolution dataset (Lsun-Bedroom 256x256)

Euler 1-step Euler 2-step RK - 45

Solver	FID (Original/Ours)	Precision	Recall
1-step Euler	139.98 / **26.54**	0.0290 / **0.4822**	0.0220 / **0.2274**
RK	24.76 / **24.14**	0.4525 / **0.4703**	0.2386 / **0.2388**
	IVD, Recon, P-Recon error (Fake/Real)		
	IVD	L_2^{recon}	$L_2^{p-recon}$
Original	1.1790	0.0822 / 0.1147	0.0820 / 0.1146
Ours	**0.9103**	**0.0487 / 0.0405**	**0.0486 / 0.0407**
	ΔL_2^{recon}	$\Delta L_2^{p-recon}$	
Original	0.0325	0.0326	
Ours	**0.0083**	**0.0079**	

Figure 8: Visual and quantitative comparison on LSUN. **Left**: 2-row layout showing original (**top**) and ours (**bottom**) for each solver (1-step, 2-step, RK). **Right**: evaluation metrics.

Our method realigns the drift in 2-Rectified Flow



Our method realigns the drift in 2-Rectified Flow

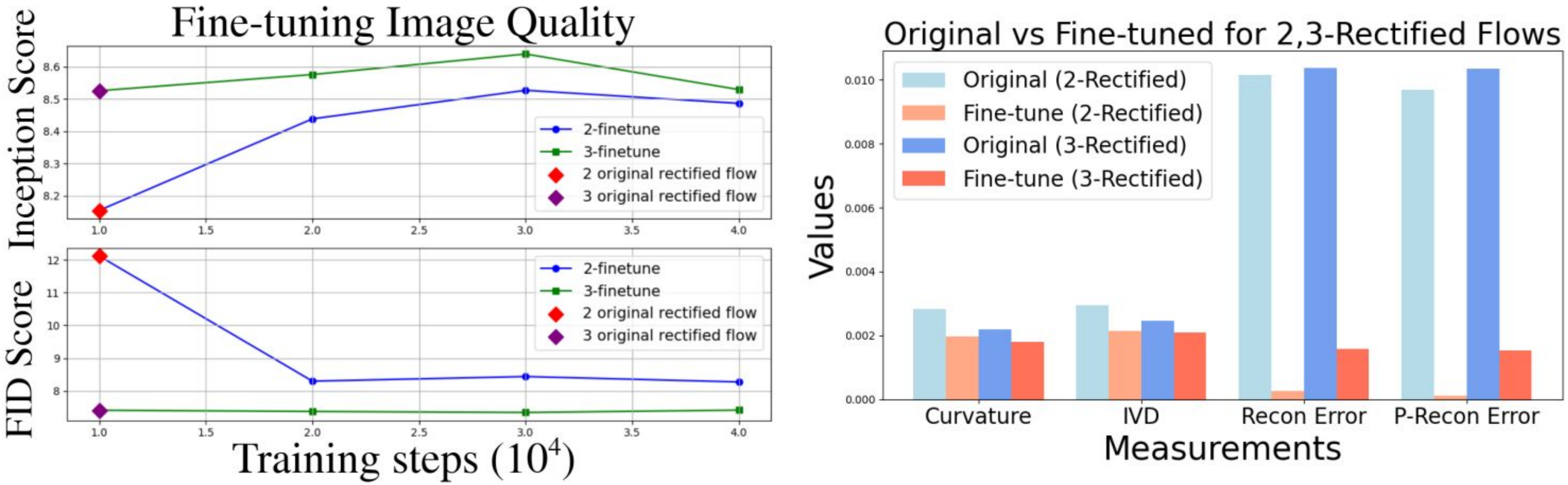


Figure A2: (a) Comparison of image quality between the original rectified model and our fine-tuned model across training steps **(left)**. (b) Comparison of measurements for original and fine-tuned models for 2- and 3-rectified flows **(right)**.

Balanced Conic Rectified flow

- Mitigated distribution drift of the target distribution during Reflow process

- Achieves better generation quality than the original Reflow process even with a small number of fake pairs.

- Produces a straighter solution trajectory than original 2-Rectified flow.