

Improving Rectified Flow with Boundary Conditions

Xixi Hu^{1,2*}, Runlong Liao^{1*}, Keyang Xu²,
Bo Liu¹, Yeqing Li², Eugene Ie², Hongliang Fei² and Qiang Liu¹

¹University of Texas at Austin, ²Google

{hxixi, liaorl, bliu, lqiang}@cs.utexas.edu

{keyangxu, yeqing, eugeneie, hongliangfei}@google.com

Abstract

Rectified Flow offers a simple and effective approach to high-quality generative modeling by learning a velocity field. However, we identify a limitation in directly modeling the velocity with an unconstrained neural network: the learned velocity often fails to satisfy certain boundary conditions, leading to inaccurate velocity field estimations that deviate from the desired ODE. This issue is particularly critical during stochastic sampling at inference, as the score function’s errors are amplified near the boundary. To mitigate this, we propose a Boundary-enforced Rectified Flow Model (Boundary RF Model), in which we enforce boundary conditions with a minimal code modification. Boundary RF Model improves performance over vanilla RF model, demonstrating 8.01% improvement in FID score on ImageNet using ODE sampling and 8.98% improvement using SDE sampling.

1. Introduction

Diffusion models have achieved significant success as a powerful class of generative models, demonstrating impressive results across various domains [12, 37, 38]. More recently, flow-based generative models have emerged as a compelling alternative for generative modeling [1, 10, 21, 25]. Within this family of flow-based models, Rectified Flow (RF) [23, 25], also known as Flow Matching [1, 21] stands out by formulating generative modeling as solving an Ordinary Differential Equation (ODE) that transforms a noise distribution into the target data distribution.

The core of RF lies in learning a velocity field that governs this continuous transformation. This is achieved by modeling a linear interpolation between the noise and data distributions along the defined trajectory. During inference, novel data samples are synthesized through numerical integration of this learned ODE, taking advantage of the powerful generalization capabilities of deep neural networks

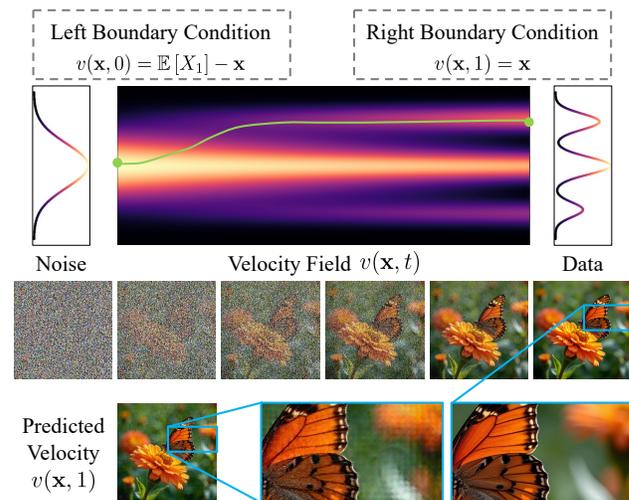


Figure 1. Boundary condition violation in Rectified Flow: predicted vs. expected velocity. Vanilla Rectified Flow model learns a velocity field $v(\mathbf{x}, t)$ to transform noise (left, $t = 0$) into data (right, $t = 1$). Ideally, this learned velocity field should satisfy defined boundary condition (top). However, as visualized, the predicted velocity at $t = 1$ (bottom) deviates from the expected data distribution and violate the right boundary condition $v(\mathbf{x}, 1) = \mathbf{x}$. This highlights a critical practical limitation of vanilla RF model. Please refer to Appendix A.2 for more examples.

in high-dimensional spaces. Recent advances have showcased the remarkable scalability of RF in complex generation tasks, including text-to-image and video generation, yielding compelling results [7, 18, 31]. Furthermore, the inherent property of RF to encourage straighter trajectories enables efficient numerical discretization, such as with Euler samplers, leading to accelerated sampling speeds – a crucial advantage for practical deployment.

Despite these advances, we observe a critical limitation when training RF models in practice: the learned velocity field often exhibits suboptimal behavior, particularly in violating a set of theoretically derived boundaries as shown in Figure 1 (upper panel). For example, it is theoretically expected that the velocity field reduces to the identity map

*Equal contribution

when applied to the clean images ($t = 1$), as shown Figure 1 (lower panel). Notably, this challenge persists even in large-scale text-to-image applications, where vanilla RF model often struggles to accurately satisfy the theoretical boundary conditions as we demonstrated in Appendix A.2.

This inaccurate velocity field estimation is particularly problematic in stochastic RF sampling, where the added Langevin dynamics term, containing a score estimation component, can diverge near the terminal time ($t = 1$), resulting in over-smoothed or cartoonish samples [14].

To address these limitations, we propose *Boundary-enforced Rectified Flow Model* (Boundary RF Model), a novel approach designed to explicitly enforce the boundary velocities by incorporating constraints and informed parameterizations. Our method effectively rectifies the terminal behavior of the velocity field, ensuring adherence to boundary conditions, while preserving the representational power of state-of-the-art architectures like U-Net and Diffusion Transformers [29]. Moreover, Boundary RF Model provides a simple and readily implementable parameterization that enforces boundary conditions, easily integrable with existing Rectified Flow models.

To validate the efficacy of Boundary RF Model, we conduct a comprehensive ablation study on image generation tasks. Our empirical evaluations demonstrate that Boundary RF Model and its variants, Subtraction-based Boundary RF Model, consistently enhance both deterministic and stochastic sampling performance. For instance, in ImageNet experiment, Boundary RF Model achieves a Fréchet Inception Distance (FID) of 6.32, outperforming the Rectified Flow baseline which yields an FID of 6.87. Moreover, by explicitly imposing the boundary conditions, Boundary RF Model also contribute to a more stable score function near $t = 1$, further contributing to performance gains of stochastic samplers. In summary, Boundary RF Model offers three primary advantages:

- **Easy to Implement and Adapt:** Boundary RF Model is straightforward to implement and integrates easily into existing Rectified Flow models, offering both mask-based and subtraction-based model variants for flexibility.
- **Improved Generation Quality:** Boundary RF Model consistently enhances image generation quality compared to vanilla RF model.
- **Stable Stochastic Sampling:** By enforcing boundary conditions, Boundary RF Model stabilizes the score function, leading to improved and more robust stochastic sampling results.

2. Background

This section provides a concise introduction to Rectified Flow (RF) [23]. RF aims to construct an efficient transport mapping from a simple noise distribution π_0 to a complex data distribution π_1 . This framework is related to, but con-

ceptually simpler than, diffusion-based or score-based approaches [12, 37, 38], as it learns an ODE with a velocity field that matches a prescribed interpolation slope.

Formally, let $\mathbf{X}_0 \sim \pi_0$ denote noise samples (e.g., from a standard Gaussian) and $\mathbf{X}_1 \sim \pi_1$ denote data samples. Suppose we are given a coupling $(\mathbf{X}_0, \mathbf{X}_1)$ drawn independently from their respective distributions. Rectified Flow first specifies an interpolation process $\mathbf{X}_t = \mathbb{I}_t(\mathbf{X}_0, \mathbf{X}_1)$, $t \in [0, 1]$, that smoothly connects \mathbf{X}_0 and \mathbf{X}_1 . A common choice is the straight-line interpolation $\mathbf{X}_t = t\mathbf{X}_1 + (1-t)\mathbf{X}_0$ [1, 10, 21, 25], although different affine interpolations introduce essentially equivalent RF [9, 35?]. Note that we cannot directly generate data samples from the interpolation process itself, because sampling \mathbf{X}_t depends on knowing \mathbf{X}_1 . To make the procedure causal, RF then Rectify this interpolation process into an ODE process of the form

$$d\mathbf{Z}_t = \mathbf{v}_t(\mathbf{Z}_t) dt, \quad \text{starting from } \mathbf{Z}_0 \sim \pi_0, \quad (1)$$

by learning a velocity field $\mathbf{v}(\mathbf{x}, t): \mathbb{R}^d \times [0, 1] \mapsto \mathbb{R}^d$. The ideal velocity \mathbf{v}_t^* at each point \mathbf{x} and time t is the conditional expectation of the interpolation slope:

$$\mathbf{v}_t^*(\mathbf{x}) = \mathbb{E}[\dot{\mathbf{X}}_t \mid \mathbf{X}_t = \mathbf{x}], \quad (2)$$

where $\dot{\mathbf{X}}_t$ denotes the derivative $d\mathbf{X}_t/dt$. In practice, \mathbf{v} is parameterized by a neural network and is learned by minimizing the mean-squared error between $\mathbf{v}(\mathbf{X}_t, t)$ and the slope of \mathbf{X}_t :

$$\min_{\mathbf{v}} \int_0^1 \mathbb{E}_{(\mathbf{X}_0, \mathbf{X}_1) \sim \pi_0 \times \pi_1} [\eta_t \|\mathbf{v}(\mathbf{X}_t, t) - \dot{\mathbf{X}}_t\|^2] dt \quad (3)$$

where η_t is a time-dependent weighting factor. Upon convergence, the learned ODE trajectories $\{\mathbf{Z}_t\}$ share the same marginal distribution as $\{\mathbf{X}_t\}$ at each time t [25]. Consequently, \mathbf{Z}_1 should match the target distribution π_1 . In practice, starting from $\mathbf{Z}_0 = \mathbf{X}_0$, we can numerically integrate this ODE forward to obtain \mathbf{Z}_1 , which serves as generated samples from π_1 .

In practice, the objective function in Equation (3) is typically optimized using stochastic gradient descent. A common strategy is to sample time t from a uniform distribution over $[0, 1]$ or employ a re-weighted sampling scheme such as the logit-normal distribution [7]. With these approaches, time points are sampled independently for each training batch, and the expectation in Equation (3) is approximated by averaging over these sampled time points and data pairs $(\mathbf{X}_0, \mathbf{X}_1)$. However, such uniform or re-weighted sampling strategies may not explicitly emphasize the boundaries of the flow, i.e., $t \approx 0$ and $t \approx 1$. Consequently, the learned velocity field might be less accurate at these crucial boundary regions, potentially contributing to

the observed challenges in satisfying the theoretical boundary conditions in practical RF models.

Euler Sampler A standard way to solve RF ODE (1) is via the explicit Euler method. Let $0 = t_0 < t_1 < \dots < t_N = 1$ be N timesteps in the interval $[0, 1]$. The discretized trajectory $\{\tilde{\mathbf{Z}}_{t_k}\}$ evolves according to

$$\tilde{\mathbf{Z}}_{t_{k+1}} = \tilde{\mathbf{Z}}_{t_k} + (t_{k+1} - t_k) \mathbf{v}(\tilde{\mathbf{Z}}_{t_k}, t_k), \quad \tilde{\mathbf{Z}}_0 \sim \pi_0,$$

Stochastic Samplers When an ODE is solved numerically (e.g., via Euler), both model approximation and discrete-time updates can introduce errors. These errors accumulate over time, causing the estimated distribution to drift from its true counterpart. To mitigate this drift, one can add a feedback mechanism in the form of Langevin dynamics to the pretrained ODE model $\mathbf{v}(\mathbf{x}, t)$ [14, 24, 38]:

$$d\mathbf{Z}_t = \underbrace{\mathbf{v}(\mathbf{Z}_t, t) dt}_{\text{Flow}} + \underbrace{\sigma_t \nabla \log \rho_t(\mathbf{Z}_t) dt + \sqrt{2\sigma_t} d\mathbf{W}_t}_{\text{Langevin}}, \quad (4)$$

where ρ_t is the density function of \mathbf{Z}_t following the ODE model (1), σ_t is a non-negative sequence, and \mathbf{W}_t is a standard Brownian motion. This Langevin force nudges \mathbf{Z}_t toward its own distribution ρ_t but does not alter the marginal. As a result, \mathbf{Z}_t under this SDE retains the same marginal distribution as the original ODE (1).

In the standard RF setting, when $\pi_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\mathbf{X}_0 \perp \mathbf{X}_1$, the score function $\nabla \log \rho_t$ can be estimated from \mathbf{v}_t using Tweedie’s formula:

$$\nabla \log \rho_t(\mathbf{x}) = \frac{t\mathbf{v}(\mathbf{x}, t) - \mathbf{x}}{1 - t}. \quad (5)$$

Note, however, that if $\mathbf{v}(\mathbf{x}, t)$ is not accurately learned, relying on a relatively large σ_t can amplify the estimation errors for $\nabla \log \rho_t(\mathbf{x})$. In text-to-image generation tasks, choosing a relatively large σ_t often manifests as oversmoothing and loss of fine details in synthesized images [14].

3. Boundary-enforced Rectified Flow Models

In this section, we first analyze the boundary conditions under linear interpolation, revealing the inherent problem of unconstrained velocity fields. We then introduce two variants of Boundary RF Model: 1) Mask-based Boundary RF Model that explicitly enforces boundary constraints into the velocity parameterization, and 2) Subtraction-based Boundary RF Model, which offers design simplicity and strong empirical performance. We further discuss the advantage of Boundary RF Model in estimating the score function and stochastic sampling in generative models.

3.1. Boundary Conditions of Rectified Flow

As detailed in Section 2, the optimal velocity field for Rectified Flow, $\mathbf{X}_t = (1 - t)\mathbf{X}_0 + t\mathbf{X}_1$, is given by:

$$\mathbf{v}^*(\mathbf{x}, t) = \mathbb{E}_{\mathbf{X}_0, \mathbf{X}_1} [\mathbf{X}_1 - \mathbf{X}_0 \mid \mathbf{X}_t = \mathbf{x}]. \quad (6)$$

Given the assumption that the noise \mathbf{X}_0 and data \mathbf{X}_1 are independent variables, the boundary velocities at $t = 0$ and $t = 1$ are:

$$\begin{aligned} \mathbf{v}^*(\mathbf{x}, 0) &= \mathbb{E}[\mathbf{X}_1 - \mathbf{X}_0 \mid \mathbf{X}_0 = \mathbf{x}] = \mathbb{E}[\mathbf{X}_1] - \mathbf{x}, \\ \mathbf{v}^*(\mathbf{x}, 1) &= \mathbb{E}[\mathbf{X}_1 - \mathbf{X}_0 \mid \mathbf{X}_1 = \mathbf{x}] = \mathbf{x} - \mathbb{E}[\mathbf{X}_0] = \mathbf{x}, \end{aligned}$$

where the simplification at $t = 1$ leverages the standard RF setting $\mathbf{X}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, resulting in $\mathbb{E}[\mathbf{X}_0] = \mathbf{0}$.

However, in practice, a general neural network $\mathbf{v}_\theta(\mathbf{x}, t)$ is used to approximate $\mathbf{v}^*(\mathbf{x}, t)$, and this approximation often fails to adhere to the derived boundary conditions, as illustrated in Figure 1.

3.2. Mask-based Boundary RF Model

To ensure adherence to the boundary conditions of Rectified Flow, we propose a boundary-aware parameterization of the velocity field, $\mathbf{v}(\mathbf{x}, t)$. This parameterization enforces both boundary conditions: $\mathbf{v}(\mathbf{x}, 1) = \mathbf{x}$ and $\mathbf{v}(\mathbf{x}, 0) = C - \mathbf{x}$, where $C = \mathbb{E}[\mathbf{X}_1]$. We achieve this by structuring the velocity field as follows:

$$\mathbf{v}(\mathbf{x}, t) = g(t) \cdot (C - \mathbf{x}) + f(t) \cdot \mathbf{x} + h(t) \cdot m_\theta(\mathbf{x}, t), \quad (7)$$

where $m_\theta(\mathbf{x}, t) : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^d$ is a neural network (e.g., U-Net or DiT), and f, g , and $h : \mathbb{R} \rightarrow \mathbb{R}$ are scalar functions with the following properties:

$$\begin{aligned} g(1) &= f(0) = h(0) = h(1) = 0, \\ g(0) &= f(1) = 1. \end{aligned}$$

This construction guarantees that both boundary conditions are satisfied by design. While various functions can fulfill these conditions (see Section 5.3 for examples), a natural and effective choice is:

$$g(t) = \cos\left(\frac{\pi}{2}t\right), \quad f(t) = \sin\left(\frac{\pi}{2}t\right), \quad h(t) = \sin(\pi t). \quad (8)$$

This results in the loss function of Eq. (3) in vanilla RF model turning into

$$\min_{\theta} \int_0^1 \mathbb{E}_{(\mathbf{X}_0, \mathbf{X}_1) \sim \pi_0 \times \pi_1} \left[\eta_t \|\mathbf{v}(\mathbf{X}_t, t) - \dot{\mathbf{X}}_t\|^2 \right] dt, \quad (9)$$

where $\mathbf{v}(\mathbf{X}_t, t)$ is now parameterized by our boundary-aware velocity field from Eq. (7).

We further explore alternative function choices and their empirical performance in Section 5.3.

3.3. Subtraction-based Boundary RF Model

While our Mask-based Boundary RF Model effectively enforces both $t = 1$ and $t = 0$ boundary conditions, in practice, we found that only ensuring the $t = 1$ condition ($\mathbf{v}(\mathbf{x}, 1) = \mathbf{x}$) also empirically benefits high-quality generation. Motivated by this observation and the desire for a less

constrained parameterization, we introduce a subtraction-based variant:

$$v(\mathbf{x}, t) = x + m_\theta(\mathbf{x}, t) - m_\theta(\mathbf{x}, 1). \quad (10)$$

This construction inherently satisfies $v(\mathbf{x}, 1) = \mathbf{x}$ by design: $v(\mathbf{x}, 1) = \mathbf{x} + m_\theta(\mathbf{x}, 1) - m_\theta(\mathbf{x}, 1) = \mathbf{x}$, regardless of the neural network m_θ . This guarantee is achieved without carefully tuning scalar functions f, g, h , offering a more direct and flexible approach.

Compared to the double-boundary parameterization, the Subtraction-based Boundary RF Model provides design simplicity by directly leveraging a standard neural network output. However, this simplicity introduces a computational trade-off: Each evaluation of $v(\mathbf{x}, t)$ requires computing $m_\theta(\mathbf{x}, 1)$ in an additional forward pass. We will empirically compare the performance and characteristics of both the Mask-based Boundary RF Model and Subtraction-based Boundary RF Model in Section 5.3.

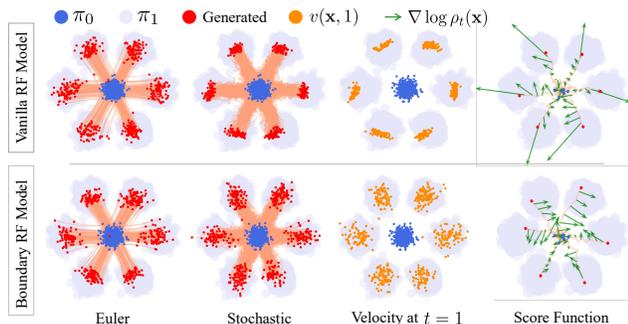


Figure 2. **Toy example: Boundary RF Model stabilizes stochastic sampling and score function.** We visualize the behavior of vanilla RF model and Boundary RF Model when learning to map from noise π_0 to data π_1 . From left to right, we visualize the followings: 1) Euler Sampling (Deterministic): Both models learn effective ODE trajectories, generating similar samples via deterministic Euler sampler. 2) Stochastic Sampling: vanilla RF model produces more concentrated samples due to score function instability. In contrast, Boundary RF Model, by enforcing boundary conditions, generates samples that retain the shape of the target distribution. 3) Velocity at $t = 1$ Boundary: vanilla RF model velocities deviate from the data, violating $v(\mathbf{x}, 1) = \mathbf{x}$, while Boundary RF Model adheres to this boundary condition. 4) Score Function: We visualize the score function computed using Eq. (5). Visualization of the score function near $t = 1$ shows that vanilla RF model exhibits an unstable and divergent score field. In contrast, Boundary RF Model demonstrates a stable and well-behaved score function, preventing the unboundedness that leads to concentration effect in stochastic sampling.

3.4. Impact on Stochastic Sampling

As discussed, stochastic samplers for Rectified Flow inherently rely on the score function, which, via Tweedie’s formula (Eq. (5)), is inversely proportional to $(1 - t)$. Conse-

quently, if the boundary condition $v(\mathbf{x}, 1) = \mathbf{x}$ is not enforced, the score function estimate will become unbounded as $t \rightarrow 1$. This unboundedness destabilizes stochastic sampling, particularly when using larger step sizes (or σ_t) for faster generation. Such instability can manifest as over-smoothing and loss of detail in generated samples [14]. In contrast, Boundary RF Model, by enforcing $v(\mathbf{x}, 1) = \mathbf{x}$, prevents the score function from diverging. This leads to a more stable score estimate, enabling robust stochastic sampling with larger σ_t .

To visually illustrate this stabilization, we present a toy example in Figure 2, directly comparing constrained and unconstrained models under stochastic sampling and demonstrating the improved score function behavior. We further showcase the benefits of Boundary RF Model in the more complex domain of image generation. As shown in Section 5.4, Boundary RF Model consistently produces sharper and more detailed image samples compared to vanilla RF model when using stochastic samplers, effectively alleviating over-smoothing artifacts.

4. Related Work

Generative modeling with diffusion and flow models.

Diffusion models [12, 37, 38] have revolutionized the field of generative modeling by introducing iterative refinement for content generation. These models have demonstrated the capacity to generate samples from complex distributions across various data modalities, including images, videos, audio, and point clouds [2, 3, 6, 13, 30, 32]. A key characteristic of diffusion models is their reliance on iterative sampling over numerous steps, with the continuous limit being described by a Stochastic Differential Equation (SDE) process. In contrast, Rectified Flow [1, 10, 21, 25], also known as Flow Matching, InterFlow, or IADB, presents an alternative approach to generative modeling based on Ordinary Differential Equations (ODEs). This ODE-based formulation is conceptually simpler and has proven to be successful in diverse practical applications, achieving performance comparable to or exceeding that of diffusion models in many cases [7, 8, 19, 31]. Recent research has increasingly recognized the conceptual connections and transferability between diffusion models and flow-based models [9, 16, 22, 24]. Consequently, sampling approaches, training techniques, and other methodological details are being explored for interchangeable use between these frameworks [14, 24]. In this work, we primarily focus on Rectified Flow due to its current efficacy and practical advantages in various applications. However, we note that the core concepts and contributions of our approach could also be potentially extended to diffusion models with minor adaptations.

Training enhancements for diffusion and flow models.

Recent advances have improved flow and diffusion mod-

els’ training by adjusting the interpolation path or the noise schedule [16, 20, 28], or by explicitly re-weighting the loss across noise levels [7, 15, 16]. A series of analyses [9, 16, 22, 24] reveals that the noise schedule itself adds no expressiveness: different choices merely impose alternative weightings over noise levels. For instance, EDM [15] decouples the noise schedule from the auxiliary time variable, enabling sampling at arbitrary noise scales during inference. While other works [9, 16, 24] prove that any monotonic noise schedule reduces to a log-SNR-weighted loss. These methods, however, do not focus on the study of boundary conditions on RF. By contrast, our approach explicitly enforces the theoretical boundary conditions of the velocity network, remedying a shortcoming of standard RF parameterizations, offering a distinct perspective from existing training refinements.

Stochastic sampling methods for diffusion and flow models. Deterministic sampling methods, often based on ODE solvers like the Euler method, are commonly employed in both diffusion models [15, 26, 36, 38] and Rectified Flow models [14]. Stochastic sampling methods offer an alternative by introducing randomness into the sampling process, potentially generating higher quality samples [14, 15, 27]. These techniques, often leveraging Langevin dynamics or related stochastic differential equation (SDE) solvers, implicitly benefit from accurate score function estimation. Our Boundary-enforced Rectified Flow Model directly enhances the stability of the score function by explicitly enforcing boundary conditions. This improved score function behavior, as we demonstrate, is particularly advantageous for stochastic samplers, enabling them to operate more robustly, especially with larger step sizes for faster sampling.

5. Experiments

In this section, we present comprehensive experiments to empirically validate the effectiveness of Boundary RF Model and address the following key research questions.

- **Efficacy of Boundary Model with Rectified Flow:** Does the integration of the boundary condition improve the performance of Rectified Flow (Section 5.2)?
- **Ablation Study of Boundary Model Variants:** How do different variations of the proposed Boundary RF Model impact the generative performance (Section 5.3)?
- **Robustness to Stochastic Sampling:** Does Boundary RF Model enhance the stability and quality of stochastic sampling in Rectified Flow (Section 5.4)?
- **Scalability and High-Resolution Performance:** How does Boundary RF Model scale to larger models and higher-resolution image generation tasks (Section 5.5)?

5.1. Experimental Setup

Datasets. We evaluated our model on CIFAR-10 [17] (32×32 resolution) and ImageNet [5] (256×256 and 512×512 resolutions), two widely used image datasets.

Evaluation Metrics. To quantitatively assess the generative performance, we employ standard metrics commonly used in image generation literature. These include Fréchet Inception Distance (FID) [11], Inception Score (IS) [34], and Precision and Recall (P/R) [33].

Model Architecture and Training Details. We build upon the Diffusion Transformer (DiT) architecture [29] for our generative models. Unless otherwise specified, the main results are reported using the DiT-B/2 configuration, which comprises 12 layers, a hidden dimension of 768, and 12 attention heads. All models were implemented using JAX [4] and trained on TPU-v5 pods. To ensure a fair comparison across all experiments, we maintain consistent random seeds and data preprocessing procedures. For stochastic sampling, we primarily employ the Stochastic Curved Euler Sampler [24] throughout our main experiments, which subsequently referred to as the “SDE sampler”. Meanwhile, we also explore Overshooting Sampler [14] which offers controllable stochasticity strength. Further details regarding training hyperparameters and implementation specifics are provided in Appendix A.1.

5.2. Comparison with Vanilla RF Model

In this section, we present a comparative analysis of our proposed method, Boundary RF Model, against the vanilla RF model. We evaluate both quantitatively and qualitatively, demonstrating the better generative performance of Boundary RF Model.

Quantitative Evaluation. Table 1 summarizes the quantitative performance of Boundary RF Model and vanilla RF model on CIFAR-10 and ImageNet 256×256 . As shown, Boundary RF Model exhibit strong quantitative performance, often outperforming vanilla RF model, especially in FID across both datasets and sampler types. across both datasets and evaluation metrics, including FID, IS, Precision, and Recall. On ImageNet 256×256 , Mask-based Boundary RF Model achieves a notable FID of 6.32, surpassing the 6.87 FID obtained by the vanilla RF model. In addition, we also compare against Rectified Flow employing mode sampling [7]. Mode sampling is a timestep sampling strategy designed to concentrate sampling probability towards the boundaries of the timestep range (i.e., near 0 and 1). However, even when compared to Rectified Flow with this enhanced mode sampling strategy, Boundary RF Model maintains its performance advantage, highlighting the effectiveness of our boundary-aware approach in addressing the inherent limitations of vanilla RF model.

	Euler Sampler				SDE Sampler			
	FID (\downarrow)	IS (\uparrow)	Precision (\uparrow)	Recall (\uparrow)	FID (\downarrow)	IS (\uparrow)	Precision (\uparrow)	Recall (\uparrow)
CIFAR-10								
Vanilla RF Model	3.94	9.23	0.71	0.57	3.63	9.61	0.73	0.56
Mask-based Boundary RF Model	3.75	9.36	0.71	0.57	3.53	9.67	0.73	0.56
Subtraction-based Boundary RF Model	3.47	9.37	0.70	0.58	3.04	9.56	0.72	0.57
ImageNet 256×256								
Vanilla RF Model	6.87	130.08	0.67	0.58	6.79	141.86	0.67	0.57
RF with Mode Sampling	7.48	114.84	0.67	0.57	7.56	120.97	0.67	0.56
Mask-based Boundary RF Model	6.63	122.35	0.67	0.60	6.18	141.43	0.69	0.56
Subtraction-based Boundary RF Model	6.32	137.28	0.66	0.59	6.38	149.00	0.66	0.59
ImageNet 256×256 with CFG								
Vanilla RF Model	6.11	201.64	0.74	0.52	7.20	224.36	0.75	0.50
RF with Mode Sampling	5.60	185.37	0.75	0.51	6.60	197.11	0.75	0.49
Mask-based Boundary RF Model	4.94	192.14	0.75	0.53	6.52	222.46	0.77	0.49
Subtraction-based Boundary RF Model	5.68	216.45	0.75	0.53	6.94	238.38	0.75	0.51

Table 1. Quantitative evaluation of Boundary RF Model against the vanilla RF model across different datasets and metrics. Lower FID and higher IS, Precision, and Recall indicate better generative performance.

Moreover, Subtraction-based Boundary RF Model performs the best among all methods in terms of IS score.

Figure 3 examines the impact of sampling step count on FID and IS scores for Boundary RF Model and vanilla RF model. Evaluations are conducted across 50, 100, and 200 steps using both Euler and SDE samplers on ImageNet dataset. The results demonstrate that Boundary RF Model exhibits better performance at all step counts. While quantitative metrics may indicate a slight performance degradation with increasing sampling steps, this trend may not fully reflect human perceptual quality. In practice, a higher number of steps often yields images with richer content and finer details, potentially surpassing what standard metrics capture [7, 14]. Also note that while the Subtraction-based Boundary RF Model achieves the best performance overall, it requires approximately twice the computation compared to the single-pass Mask-based Boundary RF Model and vanilla RF model.

Qualitative Evaluation Figure 4 provides a qualitative side-by-side comparison of image generation results from Mask-based Boundary RF Model, Subtraction-based Boundary RF Model and vanilla RF model on the ImageNet 256×256 dataset. For controlled experimentation, sample generation for all models was conducted using a consistent random seed. Qualitative inspection reveals that images generated by Mask-based Boundary RF Model and Subtraction-based Boundary RF Model exhibit better object fidelity and fewer artifacts.

5.3. Ablation Studies

To investigate the contribution of different components within our Boundary RF Model framework, we conduct a detailed ablation study. We primarily focus on evaluating the impact of different design choices in Boundary RF Model velocity parameterization, including the choice of

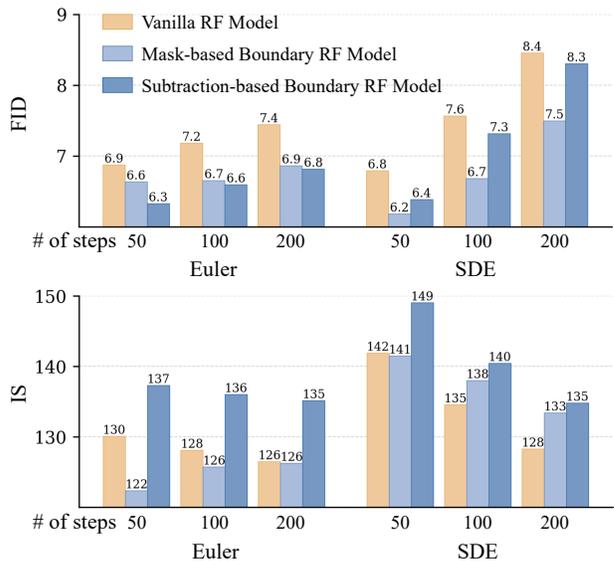


Figure 3. Performance comparison of Boundary RF Model and vanilla RF model v.s. sampling steps (50, 100 and 200 steps) on ImageNet dataset. Boundary RF Model consistently outperforms vanilla RF model across varying numbers of sampling steps, exhibiting a more substantial performance gain at higher step counts.

boundary function and the single- versus double-boundary enforcement strategy. all ablation experiments were conducted on ImageNet 256×256 .

Impact of Boundary Functions (f , g , h). In Mask-based Boundary RF Model (Eq. (7)), the functions $f(t)$, $g(t)$, and $h(t)$ play a crucial role in enforcing the boundary conditions. While we proposed $g(t) = \cos(\frac{\pi}{2}t)$, $f(t) = \sin(\frac{\pi}{2}t)$, and $h(t) = \sin(\pi t)$ as a natural choice, other functions satisfying the boundary conditions outlined in Section 3 are possible. To assess the sensitivity of our model to these function choices, we experiment with several alternatives



Figure 4. **Qualitative comparison of image generation results on ImageNet 256×256 dataset.** We present paired examples generated by vanilla RF model, Mask-based Boundary RF Model and Subtraction-based Boundary RF Model. We use the same random seed during training and evaluation for all models to ensure a fair visual comparison. Our approaches consistently generate images with better structures and improved visual fidelity compared to vanilla RF model. Additional visual examples are provided in Appendix A.3.

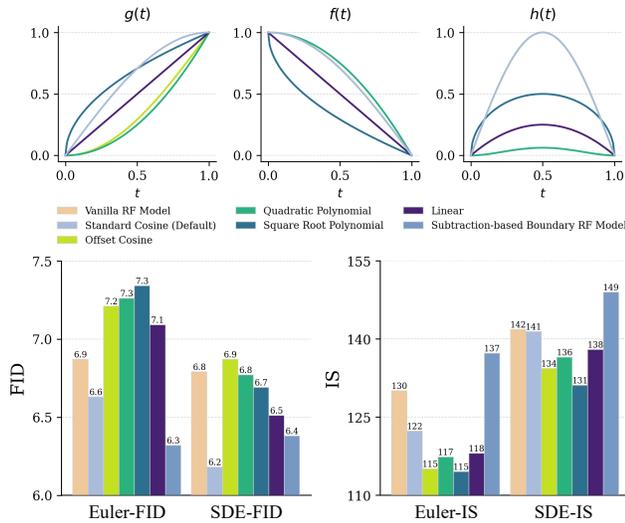


Figure 5. Ablation study on boundary functions. Quantitative comparison of different choices for functions ($f(t)$, $g(t)$, $h(t)$) in the double-boundary model. Metrics are evaluated on ImageNet 256×256 dataset.

for $f(t), g(t), h(t)$. Specifically, we compare the performance using:

- **Standard Cosine (Default):** $g(t) = \cos(\frac{\pi}{2}t)$, $f(t) = \sin(\frac{\pi}{2}t)$, $h(t) = \sin(\pi t)$.
- **Offset Cosine:** $g(t) = \cos(\frac{\pi}{2}t)$, $f(t) = 1 - \cos(\frac{\pi}{2}t)$, $h(t) = \sin(\pi t)$.
- **Quadratic Polynomial:** $g(t) = 1 - t^2$, $f(t) = t^2$, $h(t) = t^2(1 - t^2)$.
- **Square Root Polynomial:** $g(t) = (1 - \sqrt{t})$, $f(t) = \sqrt{t}$, $h(t) = \sqrt{t}(1 - \sqrt{t})$.

- **Linear:** $g(t) = (1 - t)$, $f(t) = t$, $h(t) = t(1 - t)$.

We train models using each of these function sets on the ImageNet dataset and evaluate their performance using FID and IS score. The quantitative results, summarized in Figure 5, reveal that the Standard Cosine (Default) function choice consistently achieves the best performance. We provide qualitative samples generated by these models, visualized in Appendix A.4. The default Standard Cosine model exhibits better visual fidelity compared to those from other function choices.

Right Boundary	Left Boundary	2-Pass Model	Euler-FID (\downarrow)	Euler-IS (\uparrow)	SDE-FID (\downarrow)	SDE-IS (\uparrow)
✗	✗	✗	6.87	130.08	6.79	141.86
✓	✗	✗	6.80	125.44	6.75	134.37
✓	✓	✗	6.63	122.35	6.18	141.43
✓	✗	✓	6.32	137.28	6.38	149.00

Table 2. **Boundary enforcement ablation results on ImageNet.** We present FID, IS, Precision, and Recall scores for an ablation study comparing different boundary enforcement methods within Rectified Flow on ImageNet 256×256 . We evaluate: vanilla RF model (None), Right-Boundary Model ($t=1$ only), Double-Boundary Model ($t=0$ and $t=1$), and Subtraction-based Boundary RF Model ($t=1$, subtraction-based). The results demonstrate that boundary enforcement consistently improves performance over vanilla RF model. The Subtraction-based Boundary RF Model achieves the best FID and IS, indicating the effectiveness of enforcing at least the right ($t=1$) boundary, even with a simpler parameterization.

Single- vs. Double-Boundary Enforcement. Next, we compare the performance of our single and double-boundary model and the subtraction-based model. This comparison isolates the impact of enforcing both boundary conditions versus primarily focusing on the $t = 1$ con-

straint. We train all model variants using the default Standard Cosine functions under identical settings and evaluate them using FID and IS score.

The results, shown in Table 2, indicate that while both Mask-based Boundary RF Model outperform the vanilla RF model baseline in terms of FID, the one with double-boundary condition achieves slightly superior performance across most metrics than its counterpart with only right boundary, particularly in FID. The Subtraction-based Boundary RF Model provides a significant improvement over the baseline, demonstrating the effectiveness of enforcing at least the $t = 1$ boundary condition. We note that Subtraction-based Boundary RF Model introduces approximately $2\times$ increase in inference cost compared to the Mask-based Boundary RF Model, due to the extra forward pass for $t = 1$.

Noise Scale c	0	1	2	5
Vanilla RF Model	7.18 / 128.09	7.64 / 132.89	8.26 / 135.49	11.76 / 126.24
Ours (Mask-based)	6.65 / 125.71	6.68 / 137.94	7.27 / 141.34	10.68 / 129.91
Ours (Subtraction-based)	6.59 / 135.97	7.48 / 142.31	8.04 / 142.01	11.67 / 130.74

Table 3. **Impact of noise scale c of overshooting sampler.** FID and IS scores for vanilla RF model and Boundary RF Model with varying stochasticity strength of Overshooting sampler ($c = 1.0, 2.0, 5.0$) on ImageNet dataset. Boundary RF Model demonstrates more stable performance and less degradation in sample quality for varying c . Samples are generated using 100 steps. Numbers are shown in format FID/IS.

5.4. Impact on Stochastic Sampling

To empirically validate the impact of boundary conditions on stochastic sampling stability, we conduct experiments to compare the performance of vanilla RF model and Boundary RF Model under varying degrees of stochasticity in the sampling process. We focus on evaluating the sample quality generated by both models when using stochastic samplers with different noise scales.

We compared vanilla RF model and Boundary RF Model using the Overshooting sampler [14] with varying noise scale c , controlling the strength of stochasticity. We evaluated FID and IS scores across different c values, maintaining consistent settings (ImageNet 256x256, 100 steps).

Table 3 shows that Boundary RF Model generally outperforms vanilla RF model across different c values. Importantly, as stochasticity c increases, Mask-based Boundary RF Model and Subtraction-based Boundary RF Model demonstrate better performance than vanilla RF model. This highlights that boundary conditions mitigate the stochastic sampling instability by fixing the score function. We present qualitative differences in the appendix A.5.

5.5. Scalability to Large-Scale Models and High-Resolution Images

Having demonstrated the effectiveness of boundary-aware Rectified Flow and analyzed model variants, we now assess

the scalability of Boundary RF Model to larger models and high-resolution image generation.

Table 4 shows the performance gain achieved, confirming a consistent trend with our DiT-B 256×256 findings. Our subtraction-based boundary model remains universally superior to vanilla RF when scaled to DiT-L/XL and 512×512 . The mask-based boundary model also generally outperforms vanilla RF when scaling up, though IS slightly decreases for DiT-L with Euler sampler, similar to the observations on DiT-B. These results demonstrate the scalability and consistent advantages of our method.

We show qualitative samples in Appendix A.6, which further illustrates the visual superiority of Boundary RF Model at larger model and higher resolution.

Method	Euler-FID (\downarrow)	Euler-IS (\uparrow)	SDE-FID (\downarrow)	SDE-IS (\uparrow)
DiT-L, 256×256				
Vanilla RF Model	4.72	165.44	6.77	165.31
Ours (Mask-based)	4.23	162.34	4.86	174.34
Ours (Subtraction-based)	3.75	187.93	4.52	191.49
DiT-XL, 256×256				
Vanilla RF Model	4.04	165.86	4.36	171.30
Ours (Mask-based)	3.85	170.18	4.17	183.59
Ours (Subtraction-based)	3.38	187.88	4.12	190.34
DiT-B, 512×512				
Vanilla RF Model	16.35	69.92	16.70	72.93
Ours (Mask-based)	15.21	70.35	15.69	77.25
Ours (Subtraction-based)	15.08	75.96	14.58	80.92

Table 4. **Results on larger model size (DiT-L/2 and DiT-XL/2) and higher resolutions (512×512).** Results on the larger model and higher resolutions show consistent performance improvement than vanilla Rectified Flow models.

6. Conclusion

In this work, we identified and addressed a critical limitation in standard Rectified Flow: the violation of boundary conditions in learned velocity fields. To mitigate this, we introduced Boundary RF Model, a novel boundary-aware parameterization that explicitly enforces these conditions through mask-based and subtraction-based variants. Extensive experiments on CIFAR-10 and ImageNet datasets demonstrate the superior performance of Boundary RF Model over vanilla RF model across various quantitative metrics and qualitative assessments. Ablation studies further validated the effectiveness of our boundary-aware design. These results underscore the importance of proper boundary condition handling in Rectified Flow and establish Boundary RF Model as a robust approach for high-quality generative modeling. Future work could explore the application of boundary-aware parameterizations to other flow-based or diffusion generative models and investigate adaptive boundary enforcement strategies for further performance improvements and task-specific customization.

Future work could explore extending Boundary RF Model to larger models/higher resolutions on text-to-image/video generation task.

References

- [1] Michael S Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. *arXiv preprint arXiv:2209.15571*, 2022. 1, 2, 4
- [2] Jason Baldridge, Jakob Bauer, Mukul Bhutani, Nicole Brich-tova, Andrew Bunner, Kelvin Chan, Yichang Chen, Sander Dieleman, Yuqing Du, Zach Eaton-Rosen, et al. Imagen 3. *arXiv preprint arXiv:2408.07009*, 2024. 4
- [3] James Betker, Gabriel Goh, Li Jing, Tim Brooks, Jianfeng Wang, Linjie Li, Long Ouyang, Juntang Zhuang, Joyce Lee, Yufei Guo, et al. Improving image generation with better captions. *Computer Science*. [https://cdn. openai. com/papers/dall-e-3. pdf](https://cdn.openai.com/papers/dall-e-3.pdf), 2(3):8, 2023. 4
- [4] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, et al. Jax: composable transformations of python+ numpy programs. 2018. 5
- [5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 5
- [6] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021. 4
- [7] 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. 1, 2, 4, 5, 6, 11
- [8] Zhengcong Fei, Mingyuan Fan, Changqian Yu, and Junshi Huang. Flux that plays music. *arXiv preprint arXiv:2409.00587*, 2024. 4
- [9] Ruiqi Gao, Emiel Hoogeboom, Jonathan Heek, Valentin De Bortoli, Kevin P. Murphy, and Tim Salimans. Diffusion meets flow matching: Two sides of the same coin. 2024. 2, 4, 5
- [10] Eric Heitz, Laurent Belcour, and Thomas Chambon. Iterative α -(de) blending: A minimalist deterministic diffusion model. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–8, 2023. 1, 2, 4
- [11] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017. 5
- [12] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 1, 2, 4
- [13] Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J Fleet. Video diffusion models. *Advances in Neural Information Processing Systems*, 35:8633–8646, 2022. 4
- [14] Xixi Hu, Keyang Xu, Bo Liu, Qiang Liu, and Hongliang Fei. Amo sampler: Enhancing text rendering with overshooting. *arXiv preprint arXiv:2411.19415*, 2024. 2, 3, 4, 5, 6, 8, 11
- [15] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35:26565–26577, 2022. 5
- [16] Diederik Kingma and Ruiqi Gao. Understanding diffusion objectives as the elbo with simple data augmentation. *Advances in Neural Information Processing Systems*, 36: 65484–65516, 2023. 4, 5
- [17] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 5
- [18] Black Forest Labs. Flux. <https://github.com/black-forest-labs/flux>, 2023. GitHub repository. 1, 11
- [19] Matthew Le, Apoorv Vyas, Bowen Shi, Brian Karrer, Leda Sari, Rashel Moritz, Mary Williamson, Vimal Manohar, Yossi Adi, Jay Mahadeokar, et al. Voicebox: Text-guided multilingual universal speech generation at scale. *Advances in neural information processing systems*, 36, 2024. 4
- [20] Sangyun Lee, Zinan Lin, and Giulia Fanti. Improving the training of rectified flows. *Advances in neural information processing systems*, 37:63082–63109, 2024. 5
- [21] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022. 1, 2, 4
- [22] Yaron Lipman, Marton Havasi, Peter Holderrieth, Neta Shaul, Matt Le, Brian Karrer, Ricky TQ Chen, David Lopez-Paz, Heli Ben-Hamu, and Itai Gat. Flow matching guide and code. *arXiv preprint arXiv:2412.06264*, 2024. 4, 5
- [23] Qiang Liu. Rectified flow: A marginal preserving approach to optimal transport. *arXiv preprint arXiv:2209.14577*, 2022. 1, 2
- [24] Qiang Liu. *Let us Flow Together*. Dr. Qiang Liu’s Website, 2025. https://www.cs.utexas.edu/~lqiang/PDF/flow_book.pdf. 3, 4, 5, 11
- [25] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2022. 1, 2, 4
- [26] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *Advances in Neural Information Processing Systems*, 35:5775–5787, 2022. 5
- [27] Chenlin Meng, Robin Rombach, Ruiqi Gao, Diederik Kingma, Stefano Ermon, Jonathan Ho, and Tim Salimans. On distillation of guided diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14297–14306, 2023. 5
- [28] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pages 8162–8171. PMLR, 2021. 5
- [29] 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. 2, 5

- [30] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis. *arXiv preprint arXiv:2307.01952*, 2023. 4
- [31] Adam Polyak, Amit Zohar, Andrew Brown, Andros Tjandra, Animesh Sinha, Ann Lee, Apoorv Vyas, Bowen Shi, Chih-Yao Ma, Ching-Yao Chuang, et al. Movie gen: A cast of media foundation models. *arXiv preprint arXiv:2410.13720*, 2024. 1, 4
- [32] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in neural information processing systems*, 35:36479–36494, 2022. 4
- [33] Mehdi SM Sajjadi, Olivier Bachem, Mario Lucic, Olivier Bousquet, and Sylvain Gelly. Assessing generative models via precision and recall. *Advances in neural information processing systems*, 31, 2018. 5
- [34] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *Advances in neural information processing systems*, 29, 2016. 5
- [35] Neta Shaul, Juan Perez, Ricky TQ Chen, Ali Thabet, Albert Pumarola, and Yaron Lipman. Bespoke solvers for generative flow models. *arXiv preprint arXiv:2310.19075*, 2023. 2
- [36] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. 5
- [37] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019. 1, 2, 4
- [38] 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. 1, 2, 3, 4, 5