

Flow straight and fast: Learning to generate and transfer data with rectified flow

ICLR 2023 oral

Xingchao Liu*
University of Texas at Austin
xcliu@utexas.edu

Chengyue Gong*
University of Texas at Austin
cygong@cs.utexas.edu

Qiang Liu
University of Texas at Austin
lqiang@cs.utexas.edu

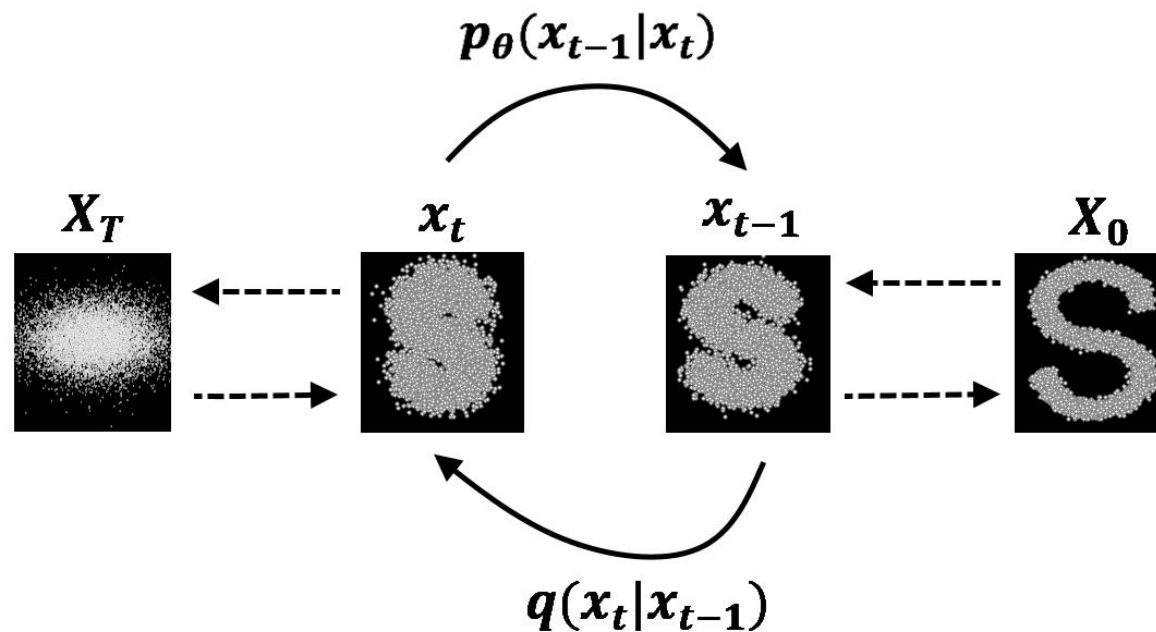
Presented by Zejia Fan
2024.6.2

Background

- Diffusion model迅速发展:
- 把一个简单的概率分布, 经过一系列"理论上无限步数、实践中有限步数"的变换, 转换为目标概率分布的过程

Background

- 阶段1: 传统意义上的diffusion model:
- 以最大似然作为优化目标, 用神经网络建模了加高斯噪声(即扩散)的逆过程的所需的未知参数



←----- Forward Diffusion
-----→ Reverse Diffusion

Background

- DDIM简化了DDPM中的约束，获得了更大的解空间，从而能够实现"多步并一步"的重建过程

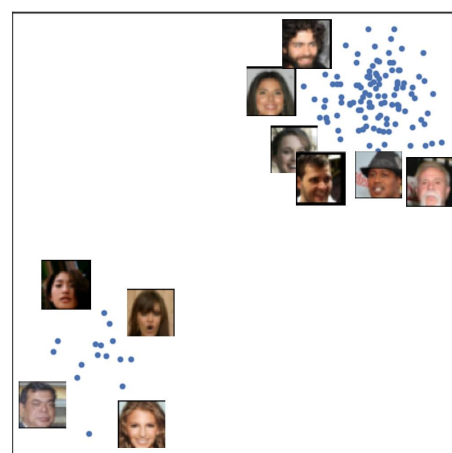
$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) \xrightarrow{\text{推导}} p(\mathbf{x}_t|\mathbf{x}_0) \xrightarrow{\text{推导}} p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \xrightarrow{\text{近似}} p(\mathbf{x}_{t-1}|\mathbf{x}_t)$$

- 损失函数只依赖于 $p(\mathbf{x}_t|\mathbf{x}_0)$
- 采样过程只依赖于 $p(\mathbf{x}_{t-1}|\mathbf{x}_t)$
- 更大的解空间：满足边际分布条件

$$\int p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)p(\mathbf{x}_t|\mathbf{x}_0)d\mathbf{x}_t = p(\mathbf{x}_{t-1}|\mathbf{x}_0)$$

Background

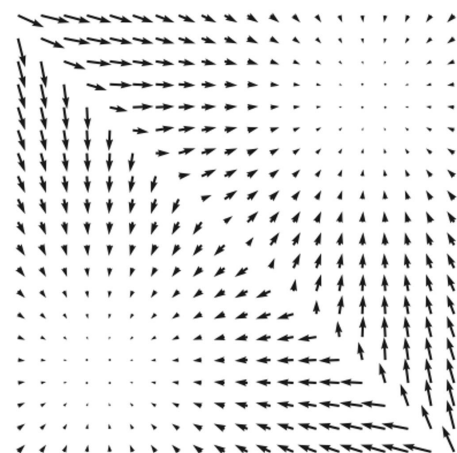
- 阶段2: 另一个分支路线是score based model
- 对概率分布的梯度——即分数(score), 作为建模对象。生成阶段: 以朗之万动力学为基础, 可以进行生成采样



Data samples

$$\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \stackrel{\text{i.i.d.}}{\sim} p(\mathbf{x})$$

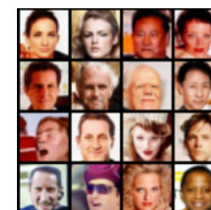
score
matching



Scores

$$\mathbf{s}_\theta(\mathbf{x}) \approx \nabla_{\mathbf{x}} \log p(\mathbf{x})$$

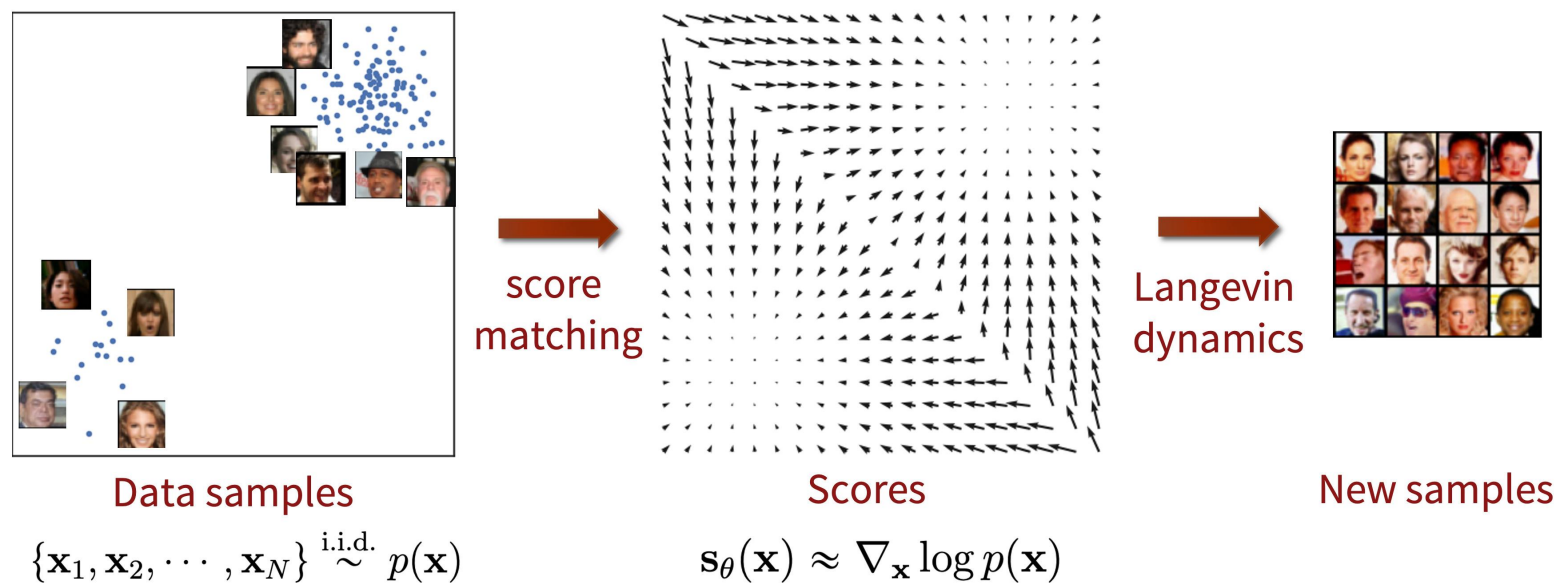
Langevin
dynamics



New samples

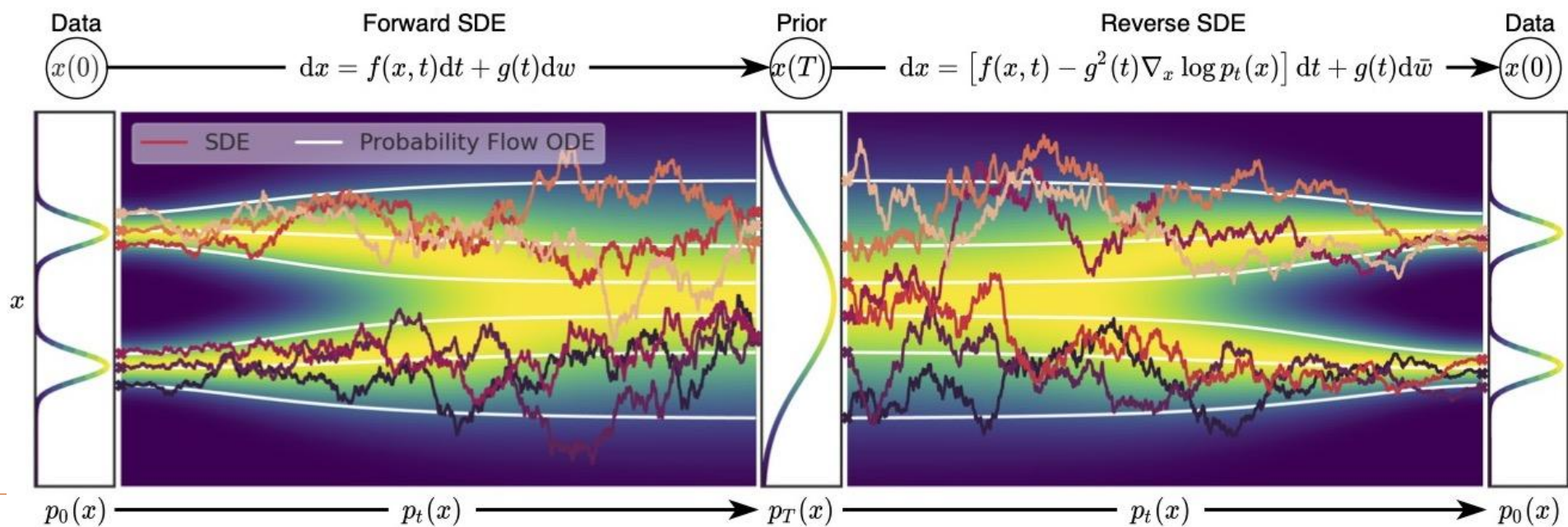
Background

- 动机不同，但从结果上看，其加噪去噪过程，和diffusion model很有相似之处



Background

- SDE统一了score based model和diffusion model
- Yang Song:随机微分方程(SDE)的视角, 统一了diffusion model和score based model, 他们都可以看做是更广义的随机微分方程的特例



Background

- 对于所有的扩散过程，都存在一个确定性过程，并且这个确定性过程和SDE有着相同的边缘概率密度。这个确定性过程符合一个ODE的表达形式
- SDE对应的ODE，被称为probability flow ODE

Background

- DDPM和SDE、 DDIM和ODE
- DDIM是一个probability flow ODE的一阶离散解
- Progressive Distillation for Fast Sampling of Diffusion Models
- 微分方程成为了更泛化的diffusion model, 通过不同的设计, 可以得到各种各样的diffusion model

Flow

- 更简单、更直接: flow
- flow matching使用了flow形式的ODE来描述简单分布到目标分布的变化过程
- 通过直接挑选生成过程中想要的概率流路线(越简单直接越好)作为人为约束条件, 将剩余部分交给神经网络学习, 达到简化建模、快速生成的目的
- 让训练和生成过程中的概率变化, 走直线

Flow

v_t : 向量场(*vector field*)

ϕ_t : 连续标准化流(*Continuous Normalizing Flow*)

- 任意变量对时间的变化常微分方程
- 表示不同时刻的 x

$$\frac{dx}{dt} = v(x)$$

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

- 经过 $\phi_t(x)$ 变换, 对任意连续时刻 t , 变量 x 的概率分布 p_t 与初始时刻 x 的概率分布 p_0 的关系是

$$p_t = p_0(\phi_t^{-1}(x)) \det \left[\frac{\partial \phi_t^{-1}}{\partial x}(x) \right]$$

Flow

- p_t 是概率分布, $\varphi_t(x)$ 是 p_t 变化的路径, $v_t(x)$ 是推动 p_t 沿着 $\varphi_t(x)$ 这条路径往前走的“动力”
- 假设 x 在0时刻服从高斯分布(纯噪声), 根据公式3, 并离散化之后, 并用 x_t 代替 $\varphi_t(x)$, (他们的含义是相同的), 应该有

$$x_{t+\Delta t} = x_t + \Delta t * v_t(x_t)$$

- 小步迭代, 逐步沿着概率路径, 生成符合目标分布的变量

Flow

- 通过网络学习目标向量场 $u_t(x)$ ，即通过 MSE loss 让下式作为优化目标

$$\mathcal{L}_{FM}(\theta) = \mathbb{E}_{t, p_t(x)} \| v_t(x; \theta) - u_t(x) \|^2$$

- p_t 是未知的， $v_t(x)$ 也是未知的

Flow

- Denoising score matching 的方法

$$q_{\sigma}(\tilde{x}) = \int q_{\sigma}(\tilde{x}|x)p_{data}(x)dx$$

- mixture of gaussian probability拟合任何一个复杂分布
- 将分布表示为边缘分布的积分，设定边缘分布为高斯分布

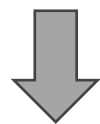
Flow

- 类似的，推导 *flow matching* $p_t(x) = \int p_t(x|x_1)q(x_1)dx_1$
- 定义一个“条件 *vector field*”： $u_t(x|x_1)$
- 从目标分布 $q_1(x)$ 中采样出来的一个样本 x_1 作为 *condition*，分布路径 $p_t(x|x_1)$ 的“推动力”

Flow

$$\mathcal{L}_{FM}(\theta) = \mathbb{E}_{t,p_t(x)} \| v_t(x; \theta) - u_t(x) \|^2$$

- 使用边缘分布:



$$\mathcal{L}_{CFM}(\theta) = \mathbb{E}_{t,q_1(x),p_t(x|x_1)} \| v_t(x; \theta) - u_t(x|x_1) \|^2$$

- 分布设计:

$$p_t(x|x_1) = \mathcal{N}(x|\mu_t(x_1), \sigma_t(x_1)^2 I),$$

$$\mu_1(x_1) = x_1$$

$$\sigma_1(x_1) = \sigma_{min}$$

$$\mu_0(x_1) = 0$$

$$\sigma_0(x_1) = 1$$

Flow

- 考虑概率流形式

$$\psi_t(x) = \sigma_t(x_1)x + \mu_t(x_1)$$

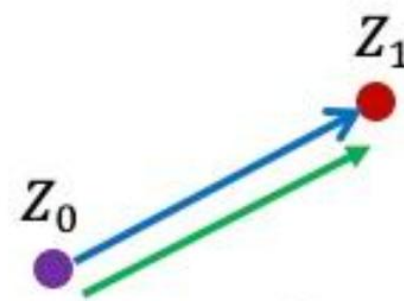
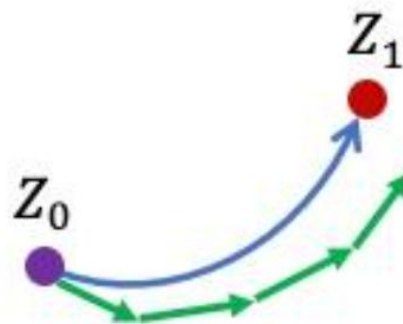
- 代入损失

$$\mathcal{L}_{CFM}(\theta) = \mathbb{E}_{t, q_1(x), p(x_0)} \| v_t(\psi_t(x_0); \theta) - u_t(x_t|x_1) \|^2$$

- 变换得到向量场

$$u_t(x_t|x_1) = \frac{\sigma'_t(x_1)}{\sigma_t(x_1)} (x_t - \mu_t(x_1)) + \mu'_t(x_1)$$

Flow



- “最优传输” 路径
- 通过设计概率流为“直线”，条件概率 $p_t(x|x_1)$ 的中心沿着一条直线路径收敛到目标分布 $q_1(x)$

$$\mu_t(x) = tx_1, \text{ and } \sigma_t(x) = 1 - (1 - \sigma_{min})t$$

$$u_t(x_t|x_1) = \frac{x_1 - (1 - \sigma_{min}x_t)}{1 - (1 - \sigma_{min})t}$$

$$\psi_t(x) = (1 - (1 - \sigma_{min})t)x + tx_1$$

$$\mathcal{L}_{CFM}(\theta) = \mathbb{E}_{t, q_1(x), p(x_0)} \| v_t(\psi_t(x_0); \theta) - (x_1 - (1 - \sigma_{min})x_0) \|^2$$

对比

DDPM

$$\begin{aligned} L_{\text{VLB}} &= \mathbb{E}_{q(\mathbf{x}_{0:T})} \left[\log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T})} \right] \\ &= \mathbb{E}_q \left[\log \frac{\prod_{t=1}^T q(\mathbf{x}_t|\mathbf{x}_{t-1})}{p_\theta(\mathbf{x}_T) \prod_{t=1}^T p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} \right] \\ &= \mathbb{E}_q \left[-\log p_\theta(\mathbf{x}_T) + \sum_{t=1}^T \log \frac{q(\mathbf{x}_t|\mathbf{x}_{t-1})}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} \right] \\ &= \mathbb{E}_q \left[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \frac{q(\mathbf{x}_t|\mathbf{x}_{t-1})}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)} \right] \\ &= \mathbb{E}_q \left[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \left(\frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} \cdot \frac{q(\mathbf{x}_t|\mathbf{x}_0)}{q(\mathbf{x}_{t-1}|\mathbf{x}_0)} \right) + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)} \right] \\ &= \mathbb{E}_q \left[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} + \sum_{t=2}^T \log \frac{q(\mathbf{x}_t|\mathbf{x}_0)}{q(\mathbf{x}_{t-1}|\mathbf{x}_0)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)} \right] \\ &= \mathbb{E}_q \left[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} + \log \frac{q(\mathbf{x}_T|\mathbf{x}_0)}{q(\mathbf{x}_1|\mathbf{x}_0)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)} \right] \\ &= \mathbb{E}_q \left[\log \frac{q(\mathbf{x}_T|\mathbf{x}_0)}{p_\theta(\mathbf{x}_T)} + \sum_{t=2}^T \log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} - \log p_\theta(\mathbf{x}_0|\mathbf{x}_1) \right] \\ &= \mathbb{E}_q \left[\underbrace{D_{\text{KL}}(q(\mathbf{x}_T|\mathbf{x}_0) \parallel p_\theta(\mathbf{x}_T))}_{L_T} + \sum_{t=2}^T \underbrace{D_{\text{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \parallel p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}} - \log p_\theta(\mathbf{x}_0|\mathbf{x}_1) \right] \end{aligned}$$

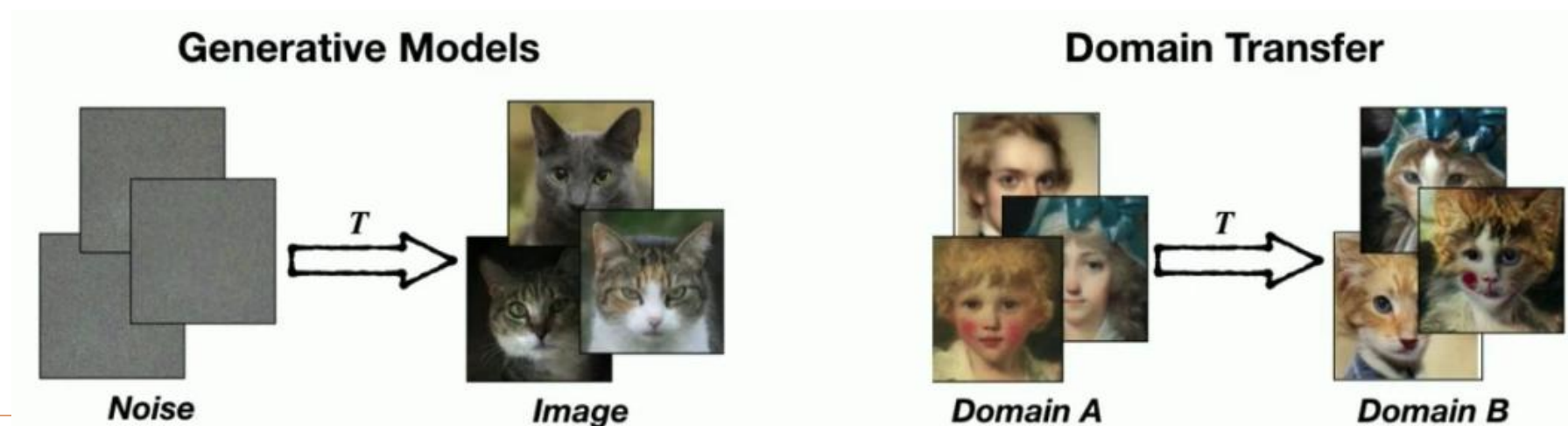
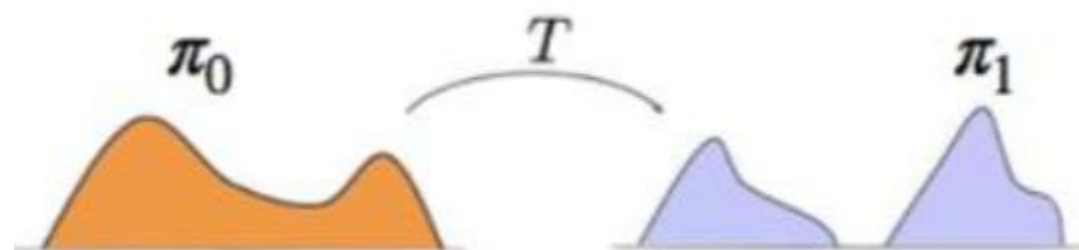
Rectified Flow

$$L = \int_0^1 \mathbb{E}_{\substack{X_0 \sim \text{noise} \\ X_1 \sim \text{data}}} [\| (X_1 - X_0) - v(X_t, t) \|^2] dt,$$

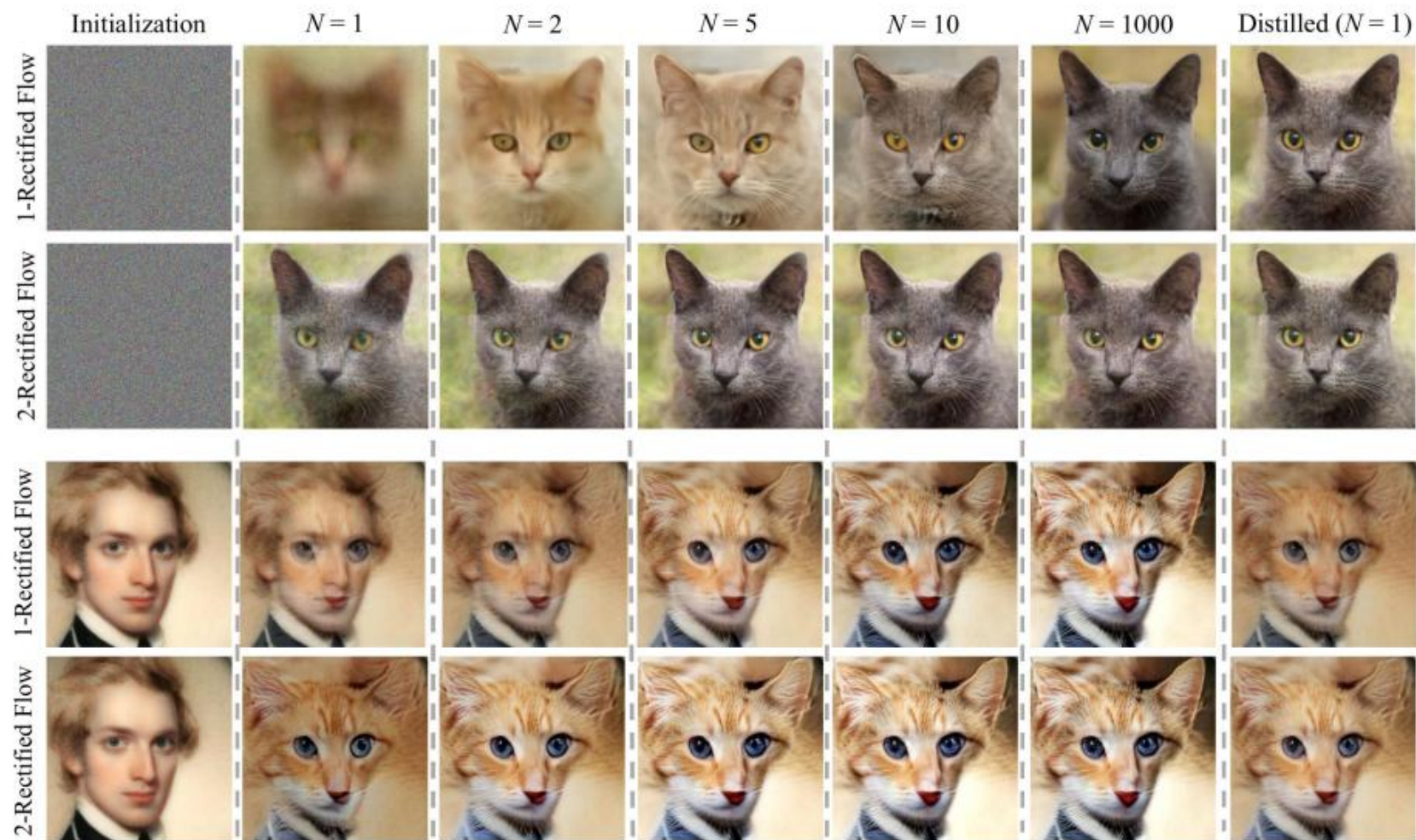
$$\text{with } X_t = t X_1 + (1 - t) X_0$$

Flow

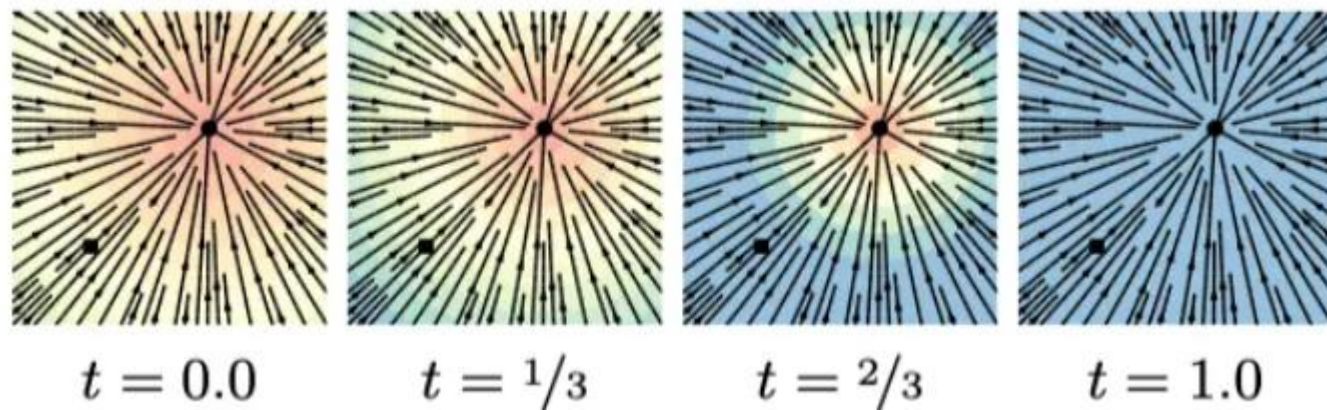
- Flow matching
- 通过flow建模概率的“搬运”



示例



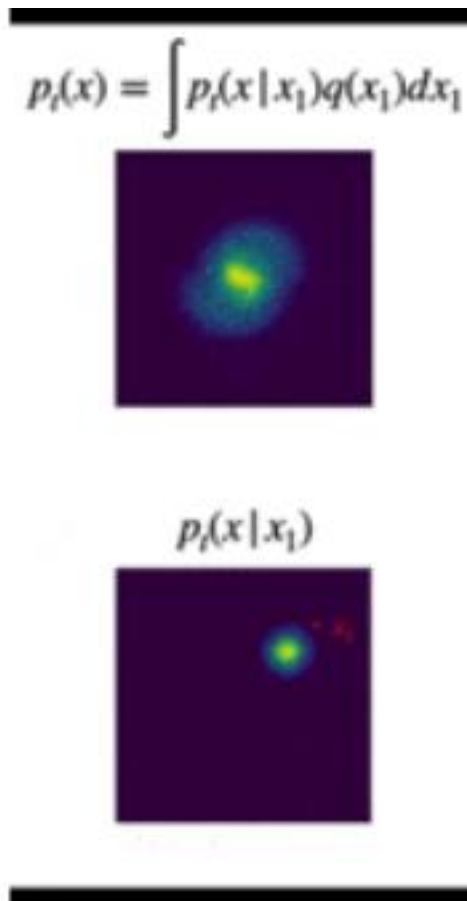
示例



Diffusion



OT



Idea相似

- Flow Matching for Generative Modeling
- Building Normalizing Flows with Stochastic Interpolants
- Iterative α (de)Blending: Learning a Deterministic Mapping Between Arbitrary Densities
- Action Matching: A Variational Method for Learning Stochastic Dynamics from Samples

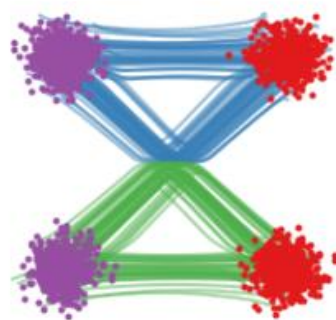
Rectified flow

- 使用采样的配对数据(噪声与生成结果)继续finetune



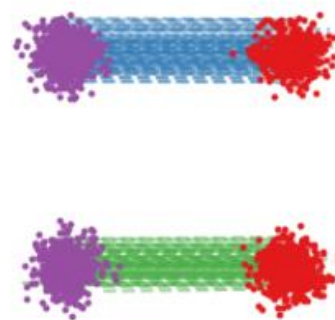
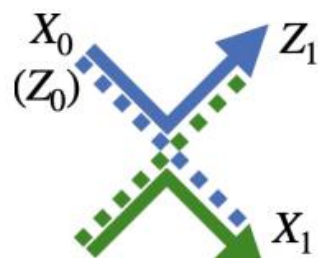
(a) Linear interpolation

$$X_t = tX_1 + (1 - t)X_0$$



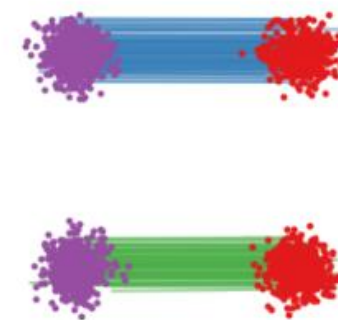
(b) Rectified flow Z_t

induced by (X_0, X_1)



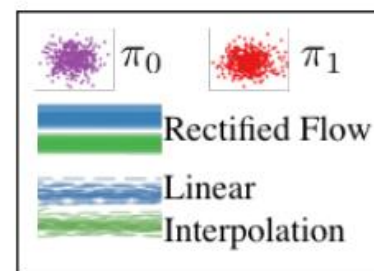
(c) Linear interpolation

$$Z_t = tZ_1 + (1 - t)Z_0$$



(d) Rectified flow Z'_t

induced by (Z_0, Z_1)



Rectified flow

- Distillation和Reflow有本质的区别
- 如果 (X_0, X_1) 的配对是随机的，Distillation最多只能得到 X_1 在给定 X_0 时的条件平均，并不能成功地完全匹配

$$\mathbb{E}[X_1|X_0] \approx X_0 + v(X_0, 0)$$

- 而Rectified flow试图找到最优配对
- 每次Reflow都可以降低两个分布之间的传输代价

Rectified flow

- 使用采样的配对数据(噪声与生成结果)继续finetune

随机配对：

$$X_0 \sim \pi_0, X_1 \sim \pi_1$$



交叉概率：46.39%
传输代价(L2)：21.20

1-Rectified Flow配对：

$$X_0 \sim \pi_0, X_1 = \text{Flow}_1(X_0) \sim \pi_1$$



交叉概率：0.76%
传输代价(L2)：20.16

2-Rectified Flow配对：

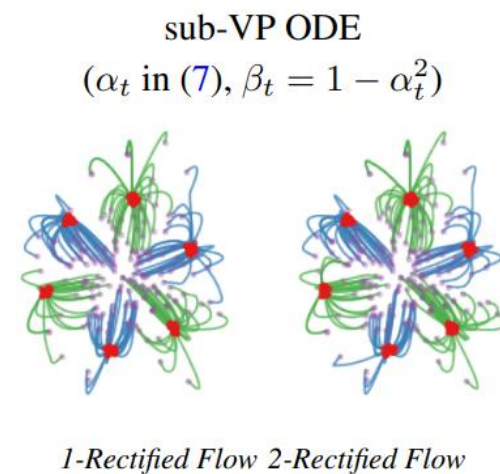
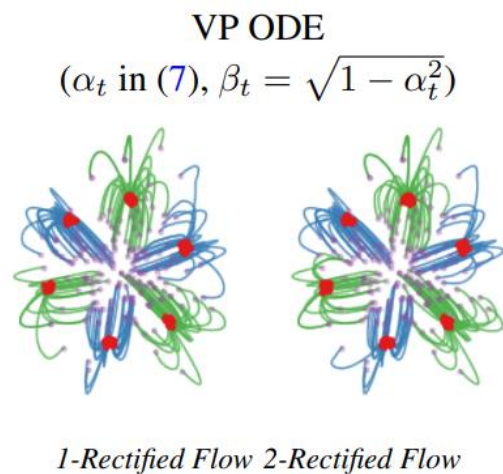
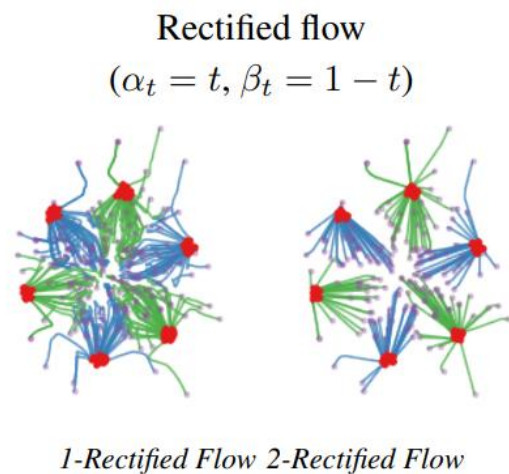
$$X_0 \sim \pi_0, X_1 = \text{Flow}_2(X_0) \sim \pi_1$$



交叉概率：0.11%
传输代价(L2)：20.14

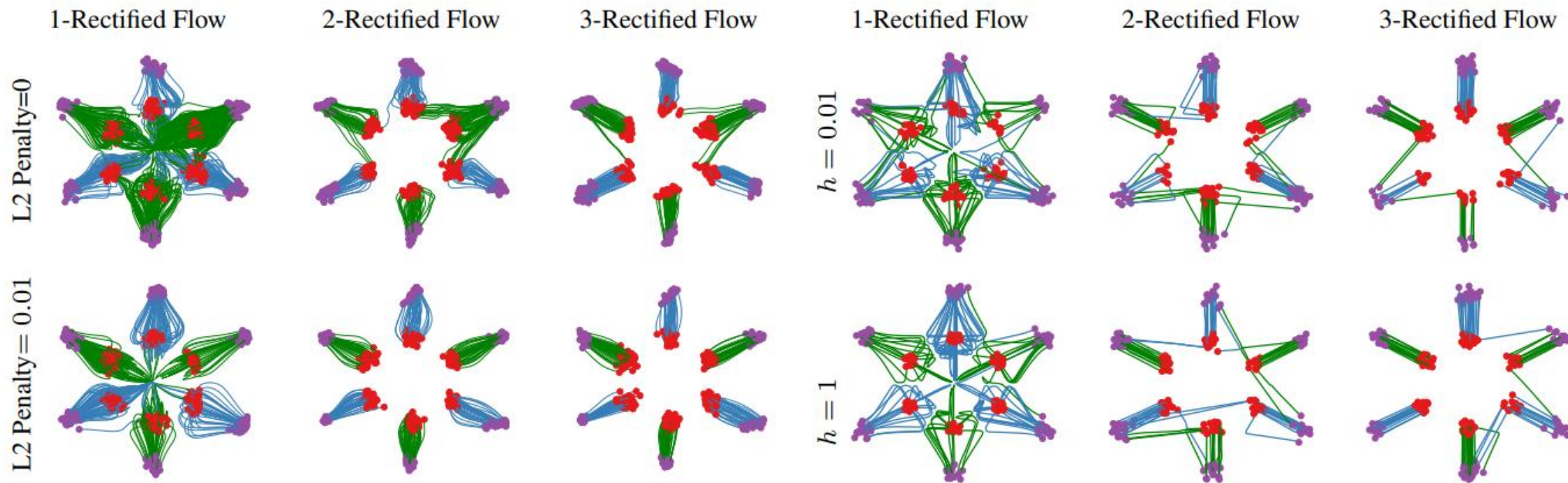
对比示例

- 走直线的重要性



Time-Discretization Steps N	Rectified Flow $\alpha_t = t, \beta_t = 1 - t$	VP ODE $\alpha_t \text{ in (7)}, \beta_t = \sqrt{1 - \alpha_t^2}$	sub-VP ODE $\alpha_t \text{ in (7)}, \beta_t = 1 - \alpha_t^2$	VP ODE (const speed) $\alpha_t = t, \beta_t = \sqrt{1 - \alpha_t^2}$
$N = 1$				
$N = 2$				
$N = 5$				
$N = 100$				

实验结果



实验结果

• Cifar-10结果

Method	NFE(↓)	IS (↑)	FID (↓)	Recall (↑)
<i>ODE</i>	<i>One-Step Generation (Euler solver, N=1)</i>			
1-Rectified Flow (+Distill)	1	1.13 (9.08)	378 (6.18)	0.0 (0.45)
2-Rectified Flow (+Distill)	1	8.08 (9.01)	12.21 (4.85)	0.34 (0.50)
3-Rectified Flow (+Distill)	1	8.47 (8.79)	8.15 (5.21)	0.41 (0.51)
VP ODE [73] (+Distill)	1	1.20 (8.73)	451 (16.23)	0.0 (0.29)
sub-VP ODE [73] (+Distill)	1	1.21 (8.80)	451 (14.32)	0.0 (0.35)
<i>ODE</i>	<i>Full Simulation (Runge–Kutta (RK45), Adaptive N)</i>			
1-Rectified Flow	127	9.60	2.58	0.57
2-Rectified Flow	110	9.24	3.36	0.54
3-Rectified Flow	104	9.01	3.96	0.53
VP ODE [73]	140	9.37	3.93	0.51
sub-VP ODE [73]	146	9.46	3.16	0.55
<i>SDE</i>	<i>Full Simulation (Euler solver, N=2000)</i>			
VP SDE [73]	2000	9.58	2.55	0.58
sub-VP SDE [73]	2000	9.56	2.61	0.58

实验结果

- 图像生成效果



(A) LSUN Church



(B) CelebA HQ



(C) LSUN Bedroom



(D) AFHQ Cat

实验结果

- 无监督图像转换



(A) Cat \rightarrow Wild Animals



(B) Wild Animals \rightarrow Cat



(C) MetFace \rightarrow CelebA Face



(D) CelebA Face \rightarrow MetFace

总结

- 回顾diffusion的发展与flow matching的提出
- 基于直线的最优传输, rectified flow
- 实用的生成加速手段

Thanks for your listening!