01 背景介绍

- 论文标题截图：
- 论文链接：https://arxiv.org/abs/1907.05600
- 录用信息：NeurIPS 2019

Generative Modeling by Estimating Gradients of the Data Distribution

score

核心就是 score
怎么预测 score
怎么用 score 来做生成

Yang Song
Stanford University
yangsong@cs.stanford.edu

Stefano Ermon
Stanford University
ermon@cs.stanford.edu

02 论文摘要

- **提出问题：**
- 提出一个全新的生成模型

- **解决方案：**
- 估计分布的梯度（score）
- 加入不同强度的高斯噪声，解决了ill-defined gradients的问题

- **优势&实验结果：**
- 支持灵活的的模型架构
- 训练中不需要任何的采样过程
- 不需要对抗学习策略
- 提供了和GAN可比的生成效果

- 论文摘要截图：

采样过程 sampling process
数据分布的梯度 score
全新的生成模型
解决 低密度区的梯度难以预测

Abstract

We introduce a new generative model where samples are produced via Langevin dynamics using gradients of the data distribution estimated with score matching. Because gradients can be ill-defined and hard to estimate when the data resides on low-dimensional manifolds, we perturb the data with different levels of Gaussian noise, and jointly estimate the corresponding scores, i.e., the vector fields of gradients of the perturbed data distribution for all noise levels. For sampling, we propose an annealed Langevin dynamics where we use gradients corresponding to gradually decreasing noise levels as the sampling process gets closer to the data manifold. Our framework allows flexible model architectures, requires no sampling during training or the use of adversarial methods, and provides a learning objective that can be used for principled model comparisons. Our models produce samples comparable to GANs on MNIST, CelebA and CIFAR-10 datasets, achieving a new state-of-the-art inception score of 8.87 on CIFAR-10. Additionally, we demonstrate that our models learn effective representations via image inpainting experiments.

退火的朗之万动力学 噪声强度不断递减 做数据扰动 灵活的架构 不需要任何对抗的方法 训练过程不需要采样

摘要逻辑：①提出了什么
②遇到什么问题
③怎么解决的
④整体方法有什么优势
⑤实验结果如何

03 相关工作

- 生成模型的分类：
- （1）likelihood-based models：
  - 直接对数据分布进行拟合（给一张图片，要求输出图和这张图片完全一样）
  - variational auto-encoders (VAEs)
  - normalizing flow models
  - 缺点：对于网络结构的设计有很大的限制

对模型性能是有损害的

UNet

- （2）implicit generative models：
  - 间接对数据分布进行拟合（输出的图片，经过判别，应该落在目标分布内）
  - generative adversarial networks (GANs)
  - 缺点：往往需要对抗学习，不好训练，容易崩

l_1/l_2

扩散模型：①网络结构灵活，输入输出尺寸相同即可

②不需要对抗，只需要学习如何去噪

04 提出方法

- score-based model不是直接学习概率分布，而是学习score

$p_\theta(\boldsymbol{x})$

Probability

概率

$\log p_\theta(\boldsymbol{x})$

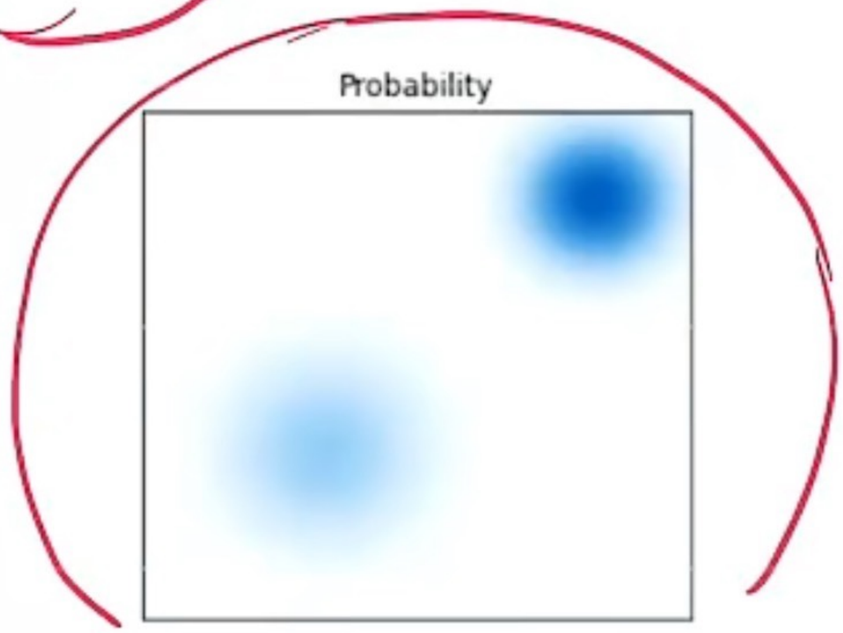
Log-probability

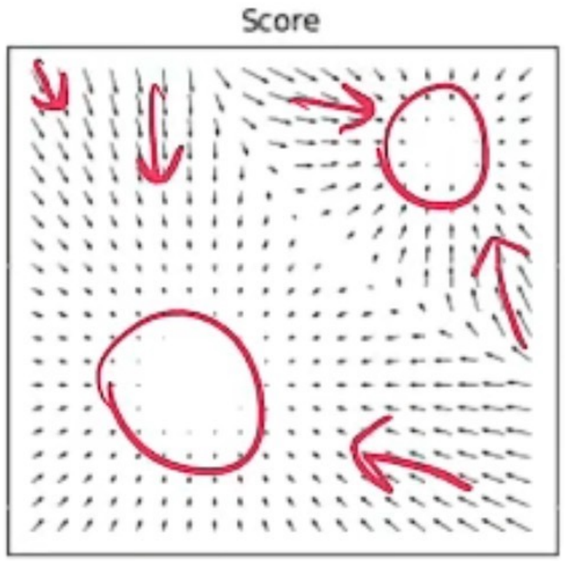
对数概率

$\nabla_{\boldsymbol{x}} \log p_\theta(\boldsymbol{x})$

"Score"

对数概率的梯度





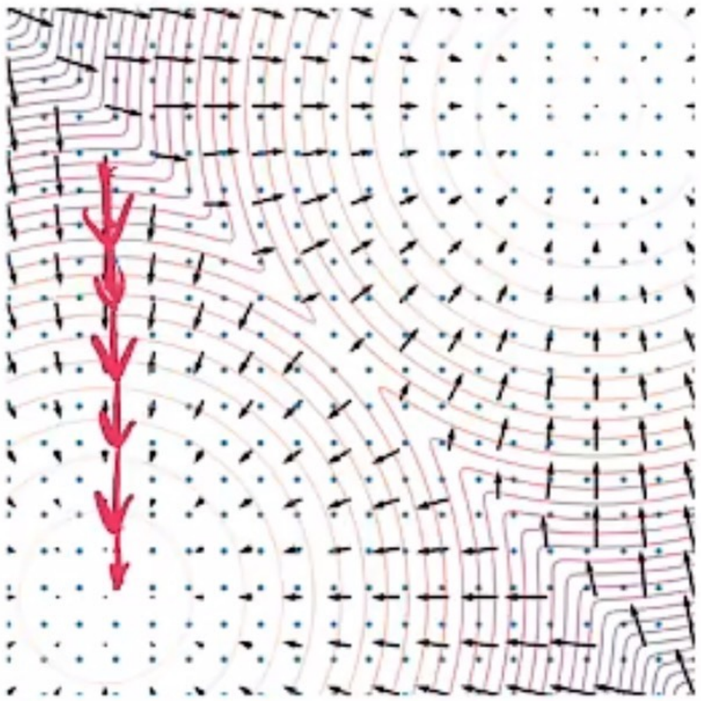
04 提出方法

怎么估计的，后面再说

- 假设我们通过某种方法（score matching）得到了估计score的模型 $\mathbf{s}_\theta(\mathbf{x}) \approx \nabla_{\mathbf{x}} \log p(\mathbf{x})$.
- 那就可以用朗之万动力学的迭代过程从一个任意分布走到目标分布，如下公式：

$$\mathbf{x}_{i+1} \leftarrow \mathbf{x}_i + \epsilon \nabla_{\mathbf{x}} \log p(\mathbf{x}) + \sqrt{2\epsilon}\mathbf{z}_i, \quad i = 0, 1, \cdots, K,$$

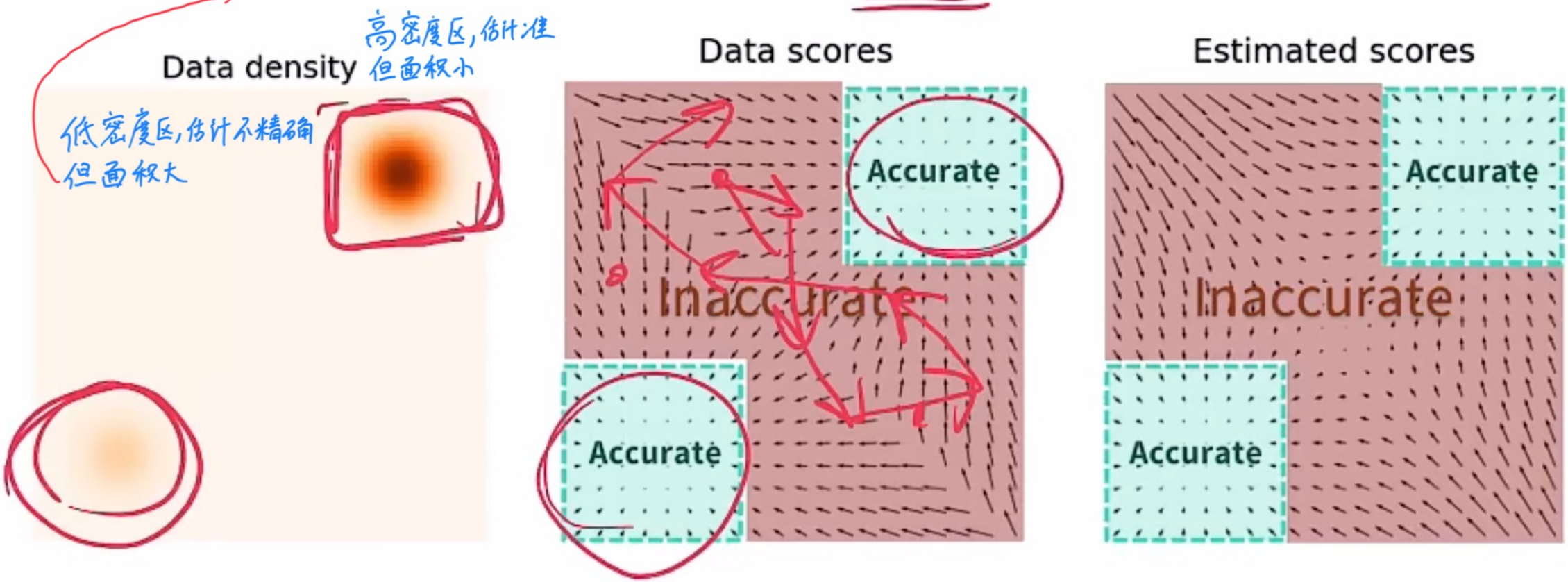
- 可以说上式给出了一种从随机采样噪声出发一步步逼近目标数据的方法，如果我们可以准确求score的话
- 示意图（gif格式）：



最重要的内容

04 提出方法

- 但是根据以上方法实现的生成过程，依然是存在问题的
- 问题是：**在数据密度较低的位置，score的估计往往是不准确的**
- 这一问题导致**推理的早期，模型很容易根据错误的梯度而“脱轨”**，不容易获得较好的结果

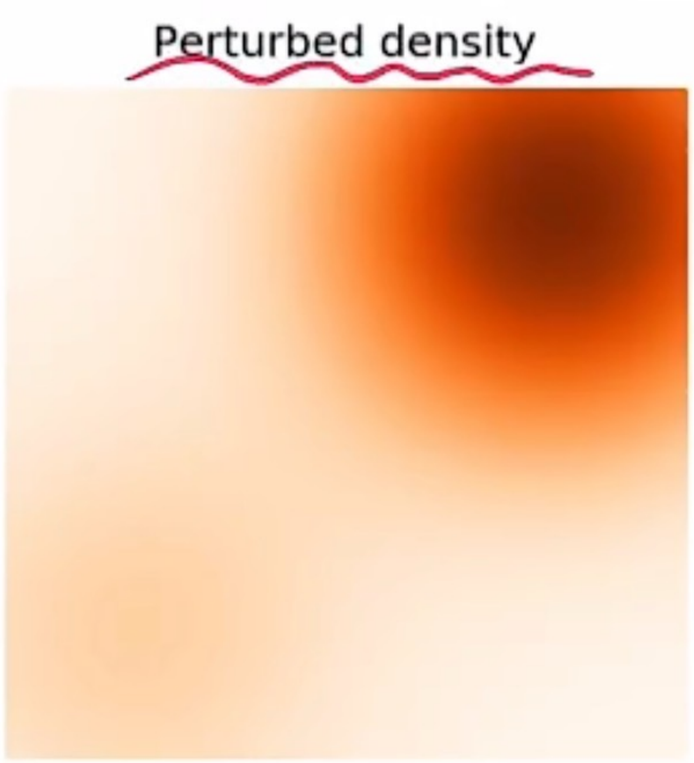


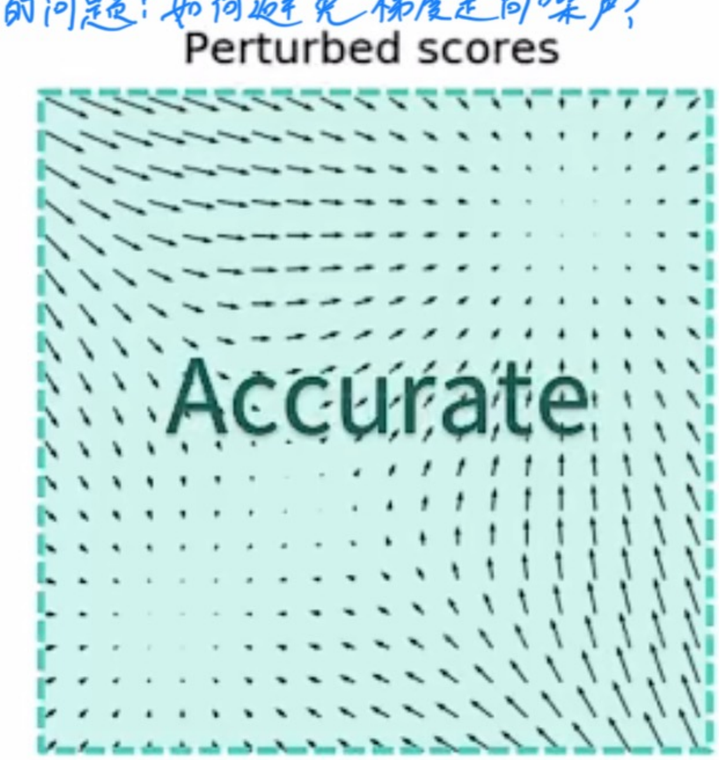
04 提出方法

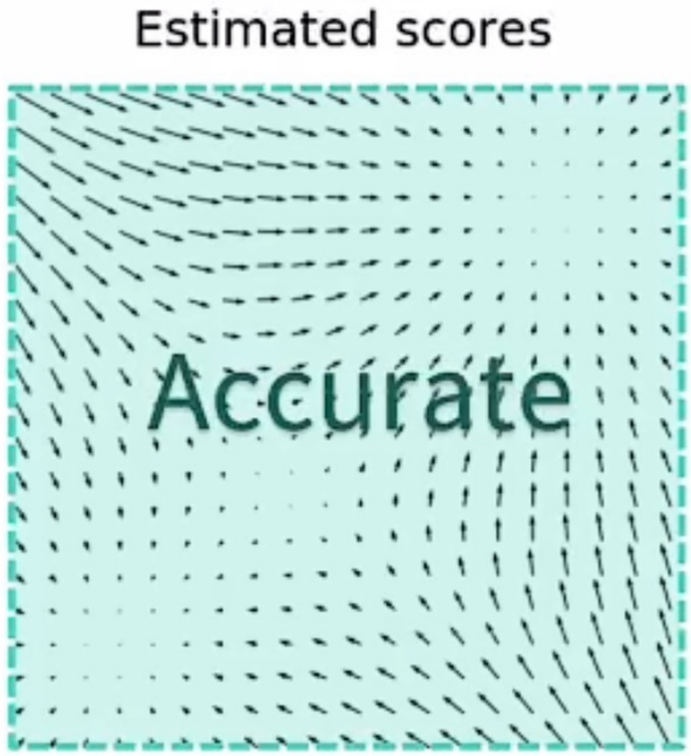
- 解决方法：对数据加噪声，以此来扩大数据范围，从而让原本低密度的数据区域“膨胀”，这样能够比较准确估计score的区域就能扩大很多了
- 所以score-based models回答了一个问题：**我们为什么要给原始数据加噪声？**
- 答案是：**为了更准确的估计score**

为了解决低密度区面积大，score难以准确估计，影响早期估计准确性的问题

新的问题：如何避免梯度走向噪声？







04 提出方法

- 新问题：**但是如何选择所加的噪声强度？**
- （1）较强的噪声——更多的区域可以准确估计score——更严重损害原本的数据分布
- （2）较小的噪声——避免损害太多原本的数据分布——无法在大多区域估计出准确的score

- 所以解决的思路就是：**在推理的不同阶段加不同强度的噪声，噪声从大到小**

减少噪声损害

让早期估计更准

Algorithm 1 Annealed Langevin dynamics.

Require: $\{\sigma_i\}_{i=1}^{L}, \epsilon, T$.

1: Initialize $\tilde{\mathbf{x}}_0$
2: **for** $i \leftarrow 1$ to L **do**
3: $\quad \alpha_i \leftarrow \epsilon \cdot \sigma_i^2 / \sigma_L^2$ ▷ α_i is the step size.
4: $\quad$ **for** $t \leftarrow 1$ to T **do**
5: $\quad\quad$ Draw $\mathbf{z}_t \sim \mathcal{N}(0, I)$
6: $\quad\quad \tilde{\mathbf{x}}_t \leftarrow \tilde{\mathbf{x}}_{t-1} + \frac{\alpha_i}{2} \mathbf{s}_{\boldsymbol{\theta}}(\tilde{\mathbf{x}}_{t-1}, \sigma_i) + \sqrt{\alpha_i}\, \mathbf{z}_t$
7: $\quad$ **end for**
8: $\quad \tilde{\mathbf{x}}_0 \leftarrow \tilde{\mathbf{x}}_T$
9: **end for**
return $\tilde{\mathbf{x}}_T$

代表级别

步长 噪声强度

$i: 1 \sim L \quad \alpha_i \quad \sigma_i$

在每个级别里，采样T步

$t: 1 \sim T$

$L \cdot T = 1000$

$L=1000 \quad T=1 \Rightarrow DDPM$

04 提出方法

$s_\theta(x_t) \approx \nabla_{x_t} \log p(x_t)$

- 最后一个问题：**怎么求score?**
- 这里讲一种最简单的思路：**denoising score matching**

- 和原文略有出入，这里我们套用DDPM的噪声假设 $\mathbf{x}_t \sim \mathcal{N}(\sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$
- 写出其分布函数为： $p(\mathbf{x}_t) \propto \exp\{-\frac{(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0)^\top(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0)}{2(1-\bar{\alpha}_t)}\}$
- 我们需要求的score为 $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$
- 推导可得其表达式： $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) = -\frac{\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0}{1-\bar{\alpha}_t}$ → $\sqrt{1-\bar{\alpha}_t}\,\varepsilon$

和噪声就差了一个系数

- 观察可知score和加在原图上的噪声只是相差一个系数的关系，可以用一个噪声估计网络来估计

- 自此，这篇工作回答了第二个问题：**为什么要估计噪声?**
- 答案是：**估计噪声就是估计score，也就是估计数据分布的对数梯度**

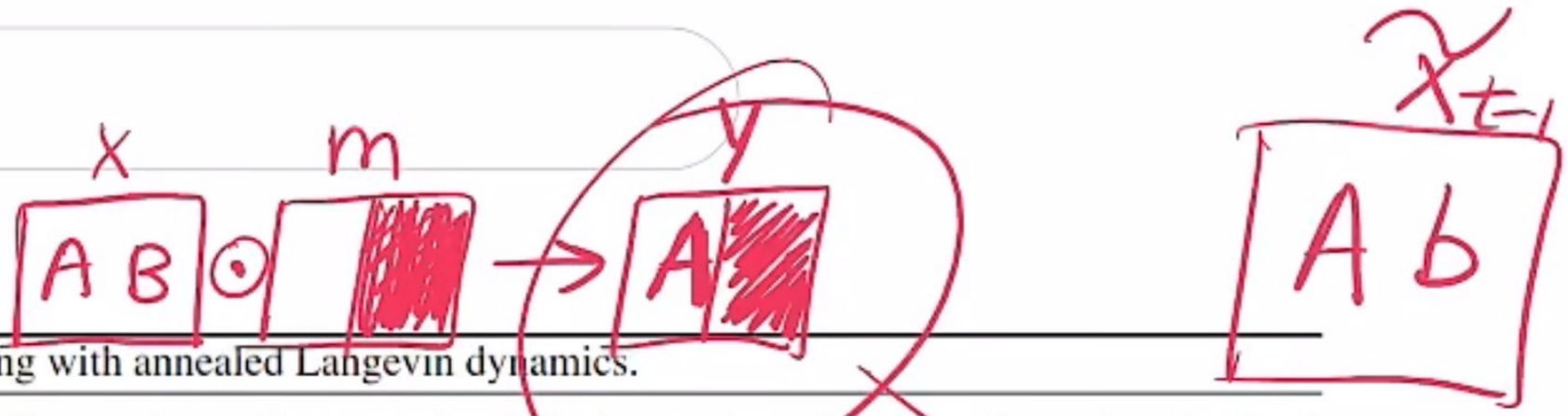
核心：① 为什么加噪声?
② 为什么要估计噪声?

04 提出方法

- inpainting的思路：

随机训练的模型

不需要在inpainting方面做finetune



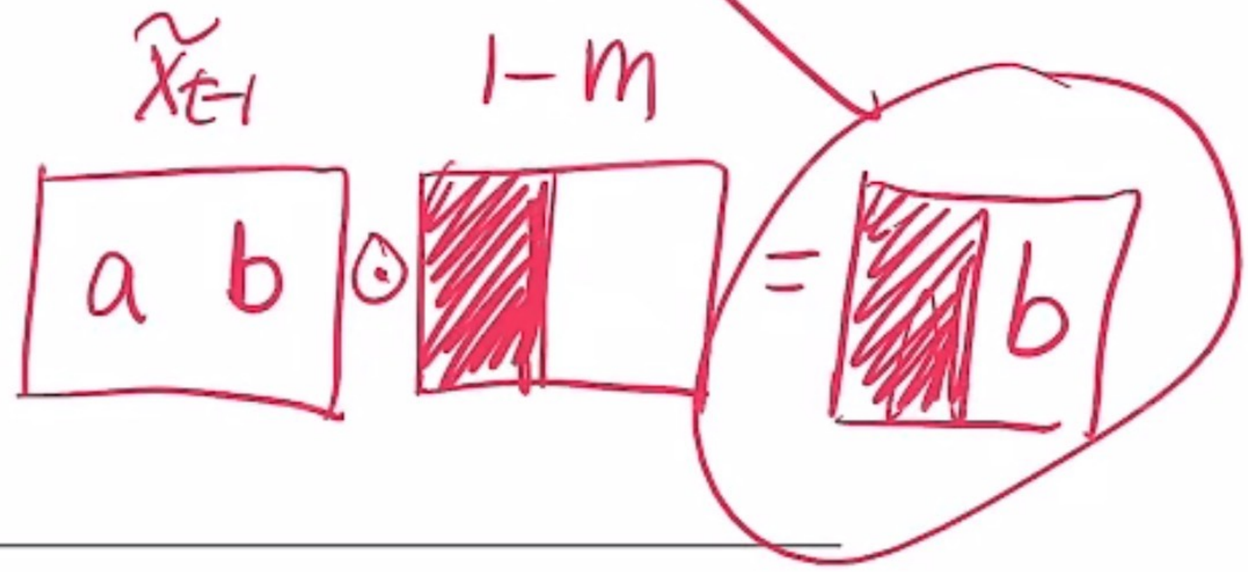
Algorithm 2 Inpainting with annealed Langevin dynamics.

Require: $\{\sigma_i\}_{i=1}^{L}, \epsilon, T$ ▷ ϵ is smallest step size; T is the number of iteration for each noise level.
Require: $\mathbf{m}, \mathbf{x}$ ▷ $\mathbf{m}$ is a mask to indicate regions not occluded; $\mathbf{x}$ is the given image.

1: Initialize $\tilde{\mathbf{x}}_0$
2: **for** $i \leftarrow 1$ to L **do**
3: $\quad \alpha_i \leftarrow \epsilon \cdot \sigma_i^2/\sigma_L^2$ ▷ α_i is the step size.
4: $\quad$ Draw $\tilde{\mathbf{z}} \sim \mathcal{N}(0, \sigma_i^2)$
5: $\quad \mathbf{y} \leftarrow \mathbf{x} + \tilde{\mathbf{z}}$
6: $\quad$ **for** $t \leftarrow 1$ to T **do**
7: $\quad\quad$ Draw $\mathbf{z}_t \sim \mathcal{N}(0, I)$
8: $\quad\quad \tilde{\mathbf{x}}_t \leftarrow \tilde{\mathbf{x}}_{t-1} + \frac{\alpha_i}{2}\mathbf{s}_{\boldsymbol{\theta}}(\tilde{\mathbf{x}}_{t-1}, \sigma_i) + \sqrt{\alpha_i}\,\mathbf{z}_t$
9: $\quad\quad \tilde{\mathbf{x}}_t \leftarrow \tilde{\mathbf{x}}_t \odot (1 - \mathbf{m}) + \mathbf{y} \odot \mathbf{m}$
10: $\quad$ **end for**
11: $\quad \tilde{\mathbf{x}}_0 \leftarrow \tilde{\mathbf{x}}_T$
12: **end for**
return $\tilde{\mathbf{x}}_T$

只加了这行代码

Mask



05 实验结果

(a) MNIST (b) CelebA (c) CIFAR-10

Figure 5: Uncurated samples on MNIST, CelebA, and CIFAR-10 datasets.

06 总结与收获

① 解答DDPM中 为什么加噪声?

② 解答DDPM中 为什么要学习噪声?

③ inpainting