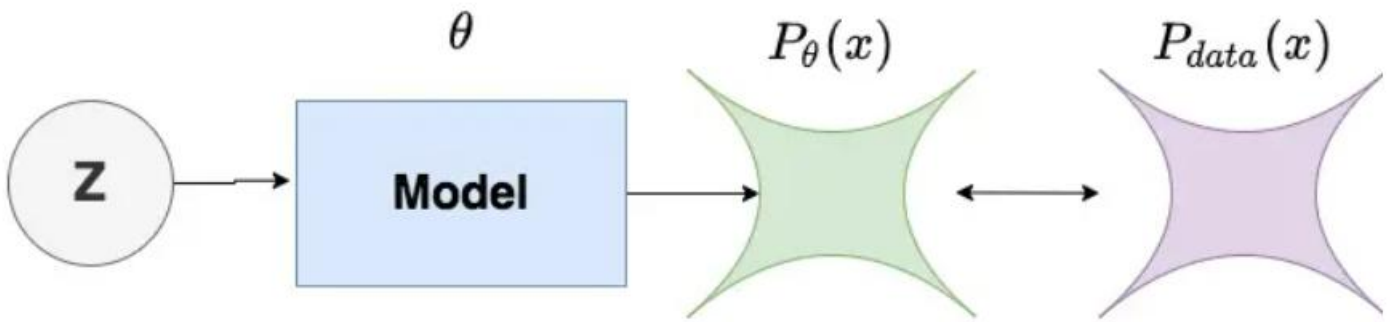
论文分享：时空扩散点过程

传统的图像扩散模型

扩散模型解决什么问题?

学习训练数据的**分布**，生成尽可能**符合训练数据分布**的真实图片



$P_\theta(x)$：模型所产生的图片的分布

$P_{data}(x)$：训练数据图片的分布

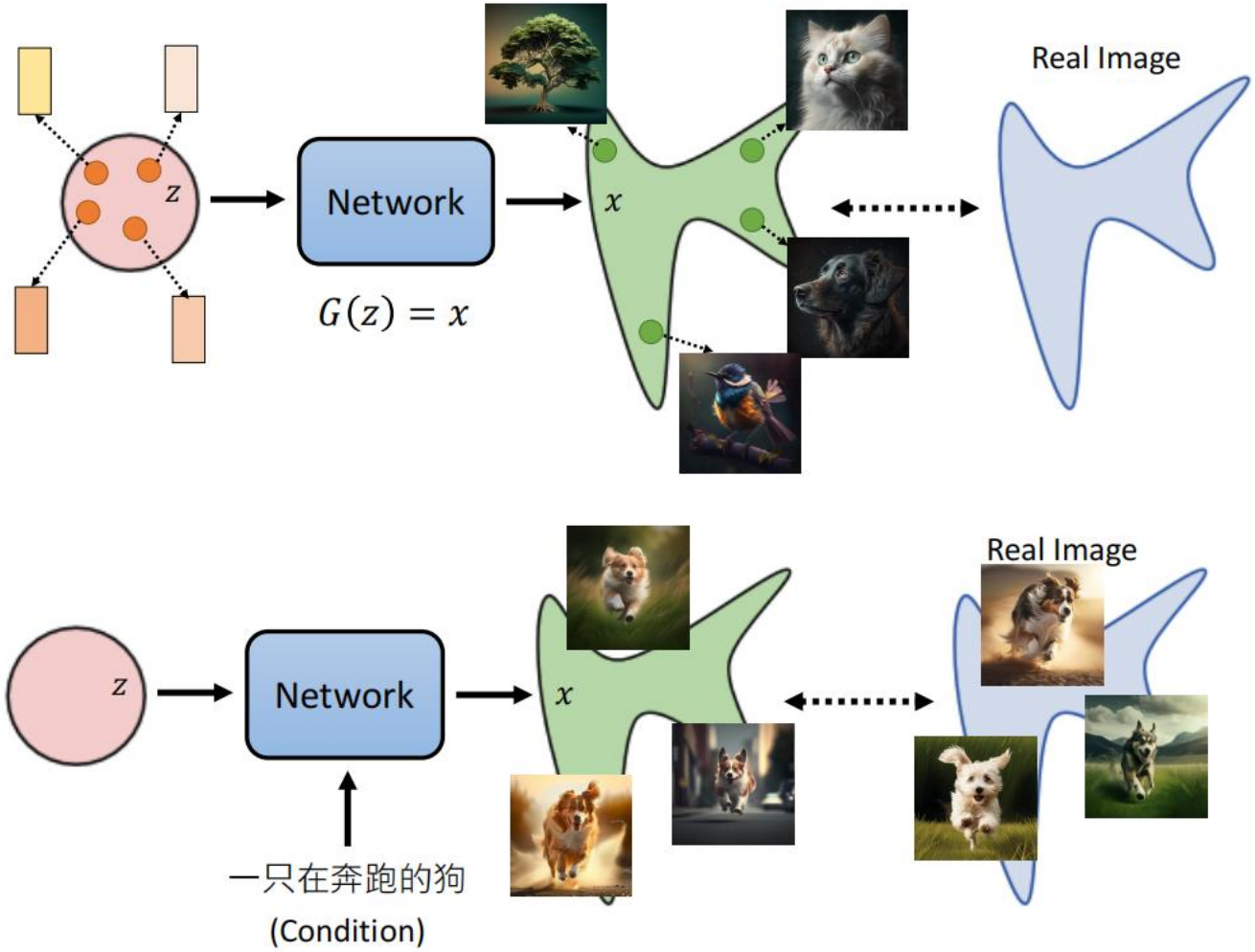
优化目标：$argmin_\theta KL(P_{data}||P_\theta) = argmax_\theta \prod P_\theta(x_i)$

输入：隐向量 z

输出：对应的一个样本

传统的图像扩散模型

无条件生成vs有条件生成



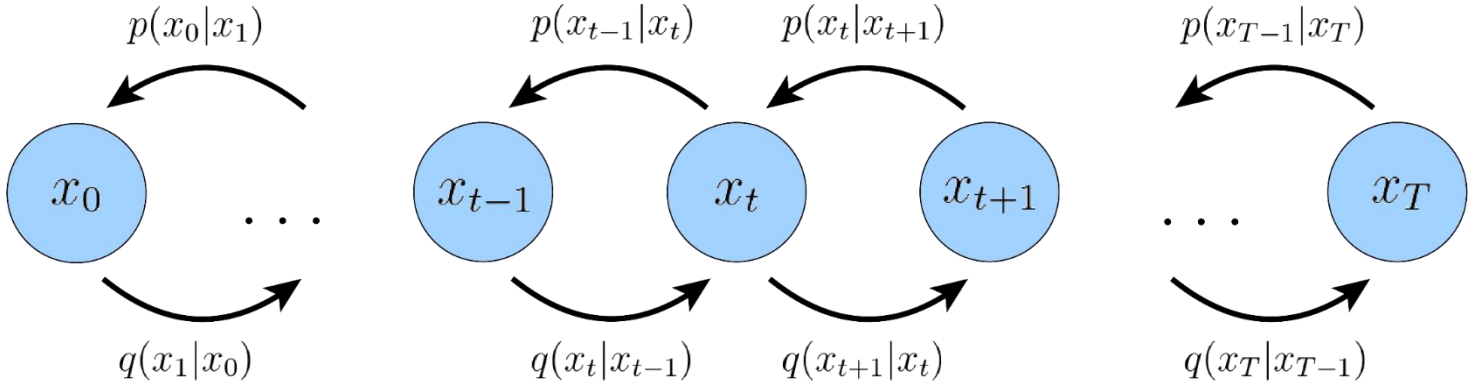
传统的图像扩散模型

前向过程，扩散过程，q过程

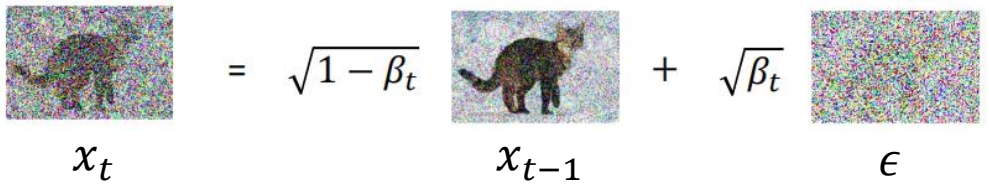
x_0：原图像。

不断向图像加入 $\sqrt{\beta_t}$ 的高斯噪声

x_T：标准正态分布



$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$



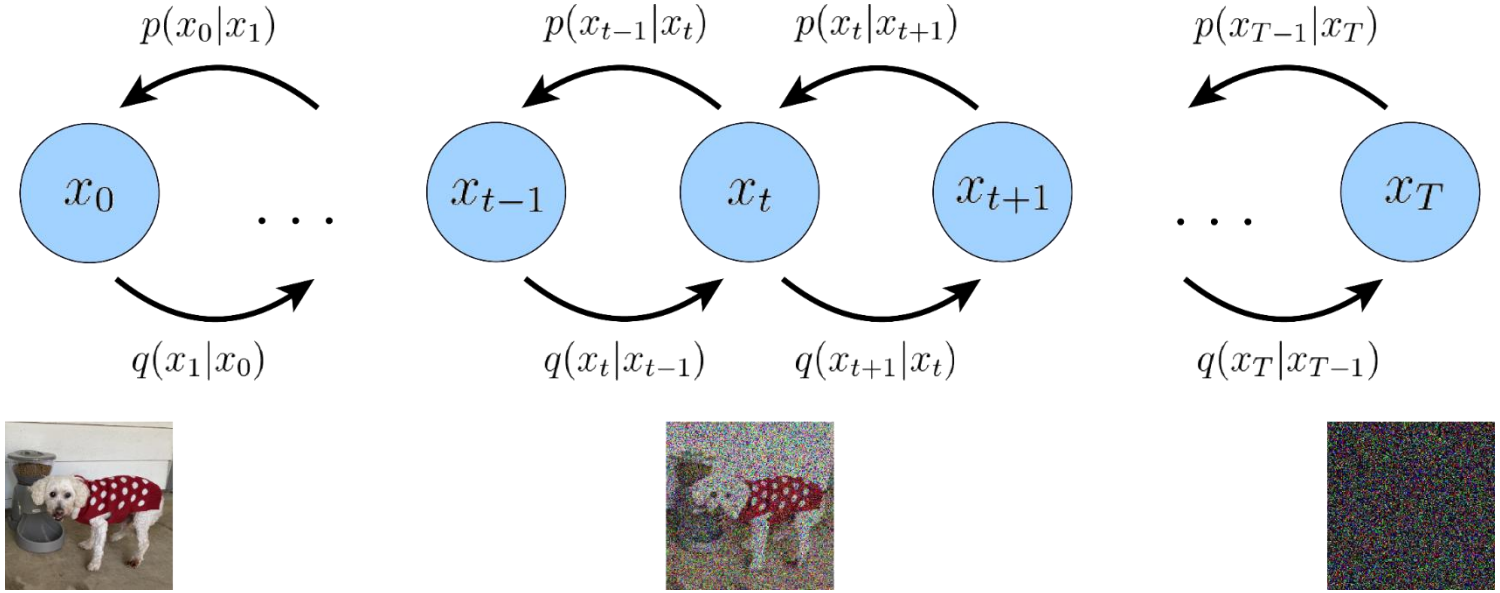
传统的图像扩散模型

反向过程，去噪过程，p过程

x_T：从高斯分布采样

不断根据 x_t 预测（采样） x_{t-1}

$p_\theta(x_{t-1}|x_t)$：使用神经网络预测均值



$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

传统的图像扩散模型

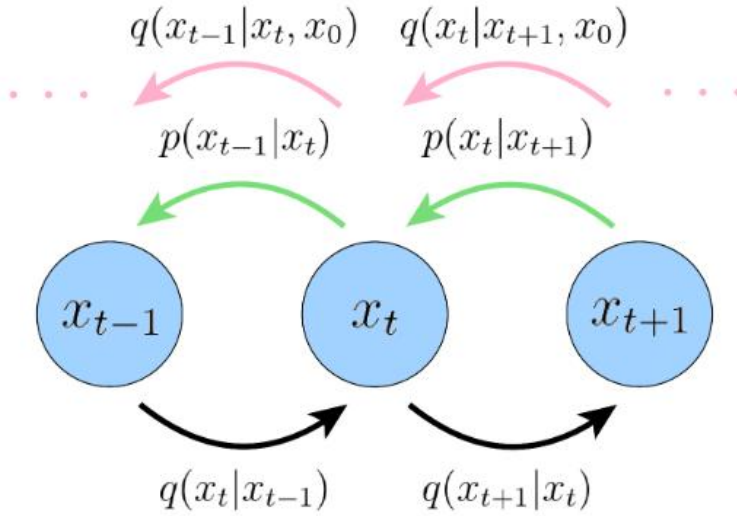
继续推导

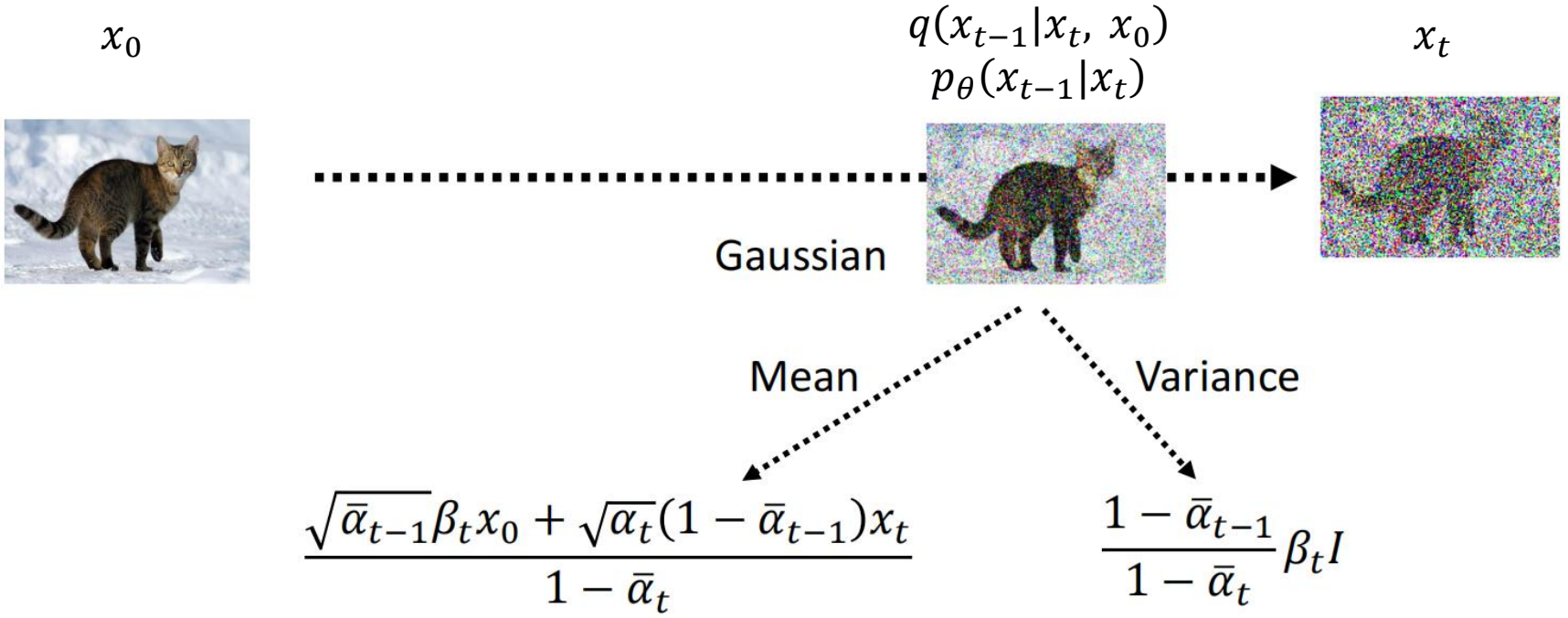
经过一些数学推导，优化的目标由 $argmin_\theta KL(P_{data}||P_\theta)$ 转化为

$$E_{q(x_t|x_0)}\, KL\big(q(x_{t-1}|x_t,\, x_0)||p_\theta(x_{t-1}|x_t)\big)$$

$q(x_{t-1}|x_t,\, x_0)$：高斯分布，均值和方差已知

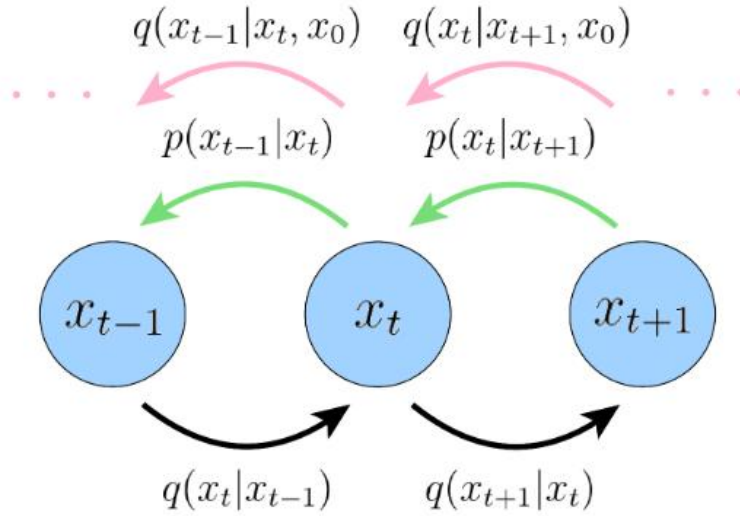
$p_\theta(x_{t-1}|x_t)$：高斯分布，$p_\theta(x_{t-1}|x_t)\sim N(\mu_\theta(x_t,\, t),\, \sigma_t I)$





传统的图像扩散模型

预测均值 → 预测噪声



接上页

$$q(x_{t-1}|x_t,\ x_0) = N\left(\frac{\sqrt{\overline{\alpha}_{t-1}}\beta_t x_0 + \sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})x_t}{1-\overline{\alpha}_t}, \sigma_t I\right)$$

由正态分布可加性，可以直接从 x_0 得出，所以不再一步步加噪。

将 x_t 表示为 x_0 加噪声 ϵ 的形式

$$x_t = \sqrt{\overline{\alpha}_t}\, x_0 + \sqrt{1-\overline{\alpha}_t}\, \epsilon,\ \ \epsilon \sim N(0, I)$$

带入上式

$$q(x_{t-1}|\epsilon,\ x_t) = N\left(\frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon\right), \sigma_t I\right)$$

预测均值 $\mu_\theta(x_t,\ t)$ → 预测噪声 $\epsilon_\theta(x_t,\ t)$

传统的图像扩散模型

Training

- 从原始数据中采样 x_0
- 采样时间步 t，噪声 ϵ
- 优化 $||\epsilon - \epsilon_\theta(x_t, t)||^2$
- 其中 $x_t = \sqrt{\overline{\alpha_t}}\, x_0 + \sqrt{1 - \overline{\alpha_t}}\, \epsilon$

Algorithm 1 Training

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \text{Uniform}(\{1, \dots, T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: Take gradient descent step on
$$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$$
6: **until** converged

传统的图像扩散模型

Inference(Sampling)

- 从噪声中采样 x_T
- $for\ t = \mathrm{T},\ \ldots,\ 1\ do$
- 采样噪声 z
- $x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t,\ t)\right) + \sigma_t z$

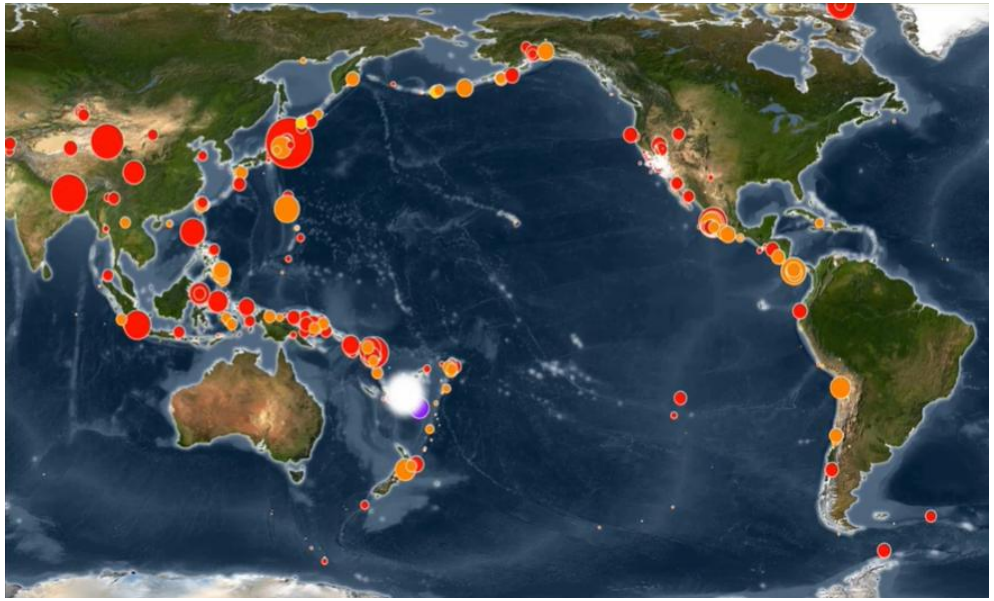
Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: $\quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\quad \mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

时空扩散点过程

时空点过程

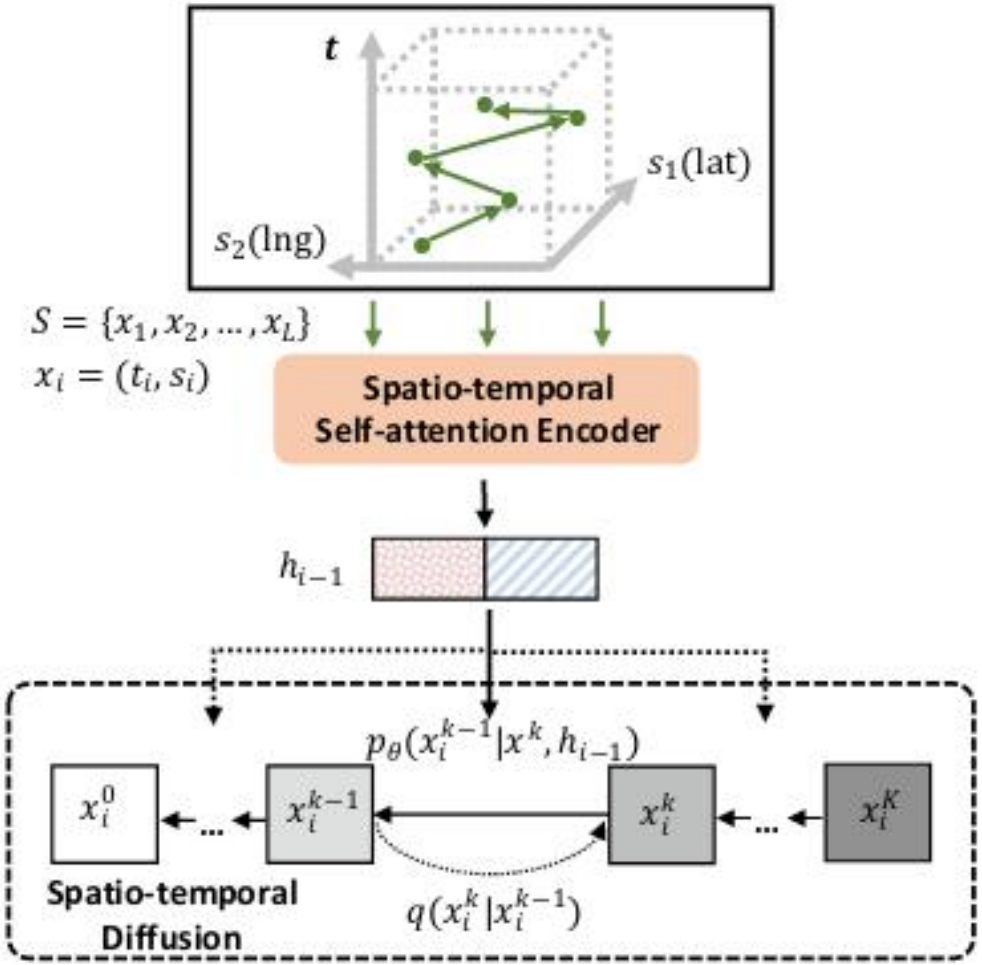
- 第 i 个时空事件：$x_i = (t_i, s_i)$
- 时空点过程：$S = \{x_1, x_2, \ldots, x_L\}$
- 过去已发生事件：$H_t = \{x_i | t_i < t, x_i \in S\}$
- 使用扩散模型建模时空点过程为 $p(t, s|H_t)$
- e.g. 地震、大脑神经元活动、人类活动



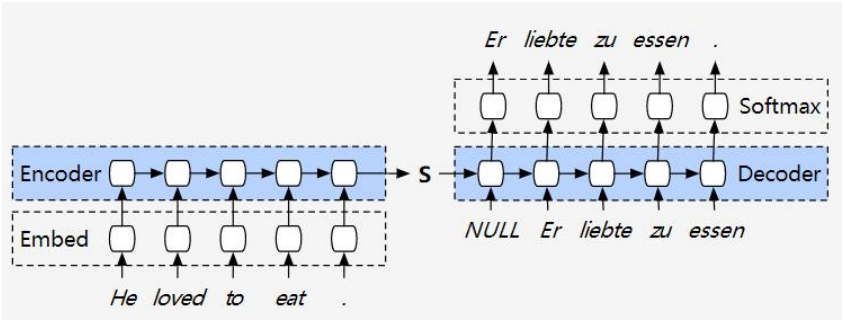
时空扩散点过程

Overview

- **两个模块**
- **时空Encoder（建模 H_t）**：学习事件历史的有效表示，充当条件以支持时空去噪扩散过程。
- **扩散模型与噪声预测网络（建模 $p(t, s|H_t)$）**：基于H_t预测下一事件。



- **类比序列模型：**
  - **时空Encoder ↔ RNN**
  - **扩散模型 ↔ 分类头**



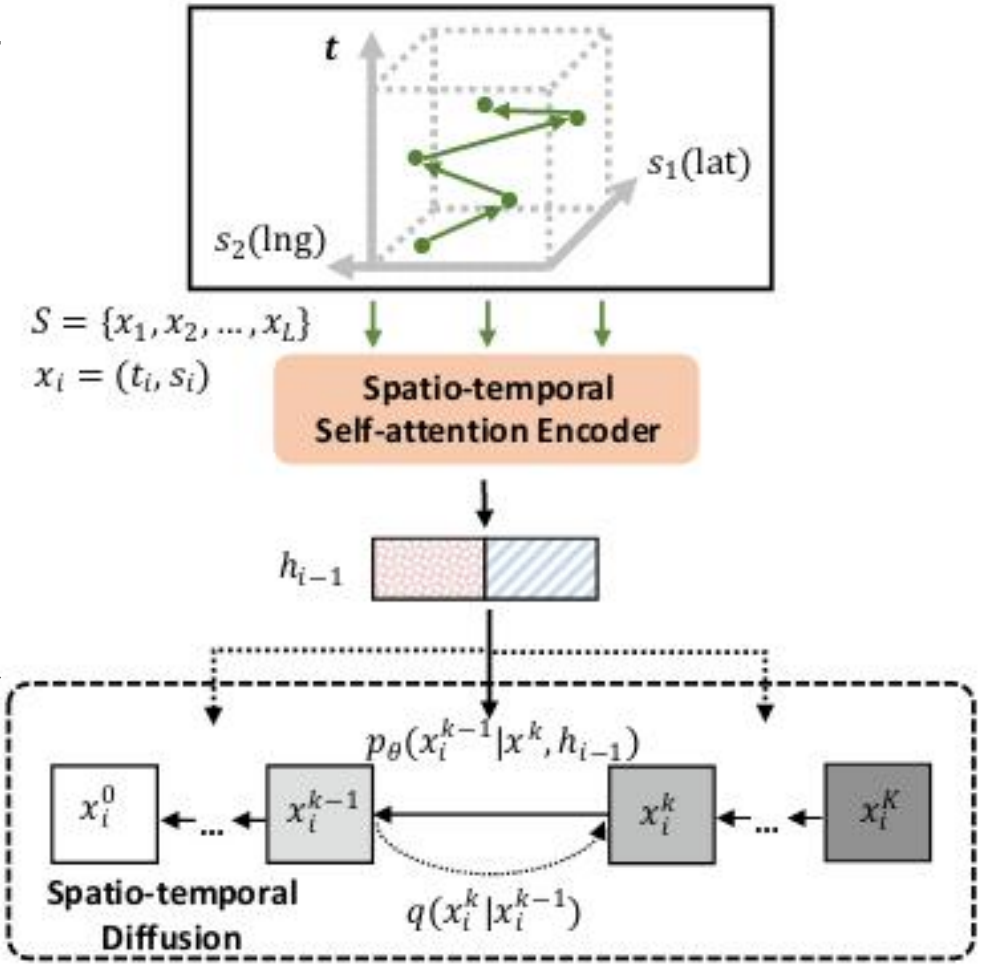
时空扩散点过程

时空Encoder

- 学习事件历史的有效表示，充当条件以支持时空去噪扩散过程。
- 时间 Embedding，E_t：

$$[e_t]_j = \begin{cases} cos(t/10000^{\frac{j-1}{M}}) & \text{if } j \text{ is odd} \\ sin(t/10000^{\frac{j-1}{M}}) & \text{if } j \text{ is even} \end{cases}$$

- 空间 Embedding，E_s：线性投影
- 联合 Embedding，$E_{st} = E_t + E_s$
- 将上述三个 Embedding 经过多头自注意力机制，得到隐状态（扩散模型的条件）： h_s、h_t、h_{st}



时空扩散点过程

扩散模型与噪声预测网络

- s_i^k, τ_i^k，待去噪的空间与时间差 ↔ 3 像素图像
- $\epsilon_s^k = \epsilon_\theta(x_i^k, h_{i-1}, k)$, $\epsilon_\tau^k = \epsilon_\theta(x_i^k, h_{i-1}, k)$，预测的噪声
- k：正弦位置编码
- 条件： $h_i = [h_s, h_t, h_{st}]$

$$x_i = [x_{s,i}, x_{t,i}] \ ,$$

$$\epsilon_{s,i}^k = \sum \alpha_s x_i \ , \epsilon_{t,i}^k = \sum \alpha_t x_i,$$

$$s_i^{k-1} = \frac{1}{\sqrt{\alpha_k}}(s_i^k - \frac{\beta_k}{\sqrt{1-\bar{\alpha}_k}}\epsilon_\theta(x_i^k, h_{i-1}, k)) + \sqrt{\beta_k} z_s$$

$$\tau_i^{k-1} = \frac{1}{\sqrt{\alpha_k}}(\tau_i^k - \frac{\beta_k}{\sqrt{1-\bar{\alpha}_k}}\epsilon_\theta(x_i^k, h_{i-1}, k)) + \sqrt{\beta_k} z_t$$

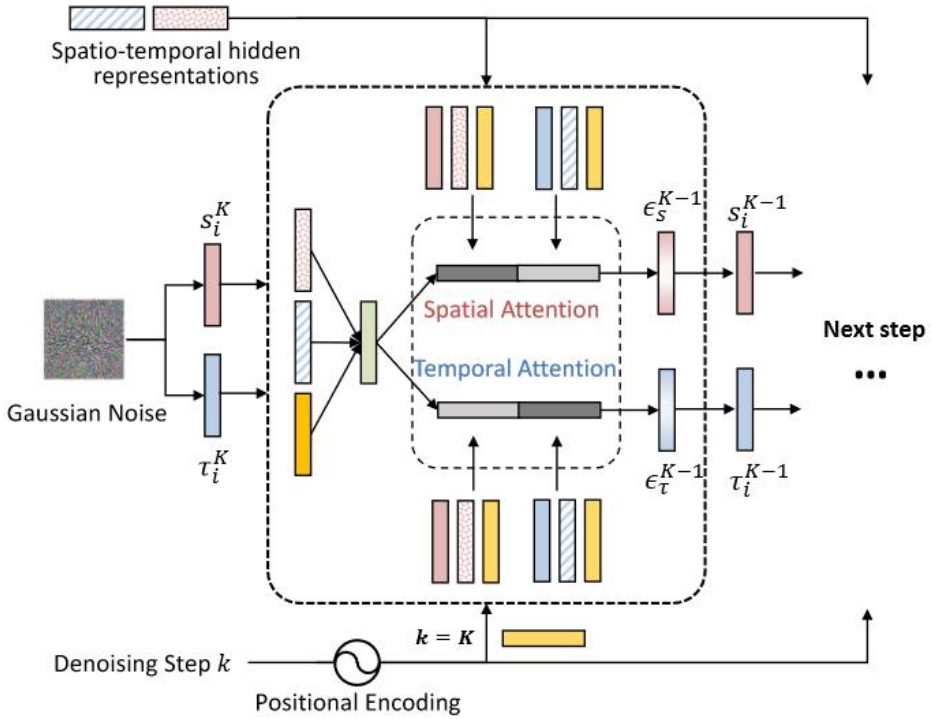
$$e_k = \text{SinusoidalPosEmb}(k) \ ,$$

$$\alpha_s = \text{Softmax}(W_{sa}\text{Concat}(h_{i-1}, e_k) + b_{sa})$$

$$\alpha_t = \text{Softmax}(W_{ta}\text{Concat}(h_{i-1}, e_k) + b_{ta})$$

$$x_{s,i} = \sigma(W_s s_i^{k+1} + b_s + W_{sh} h_{s,i-1} + b_{sh} + e_k)$$

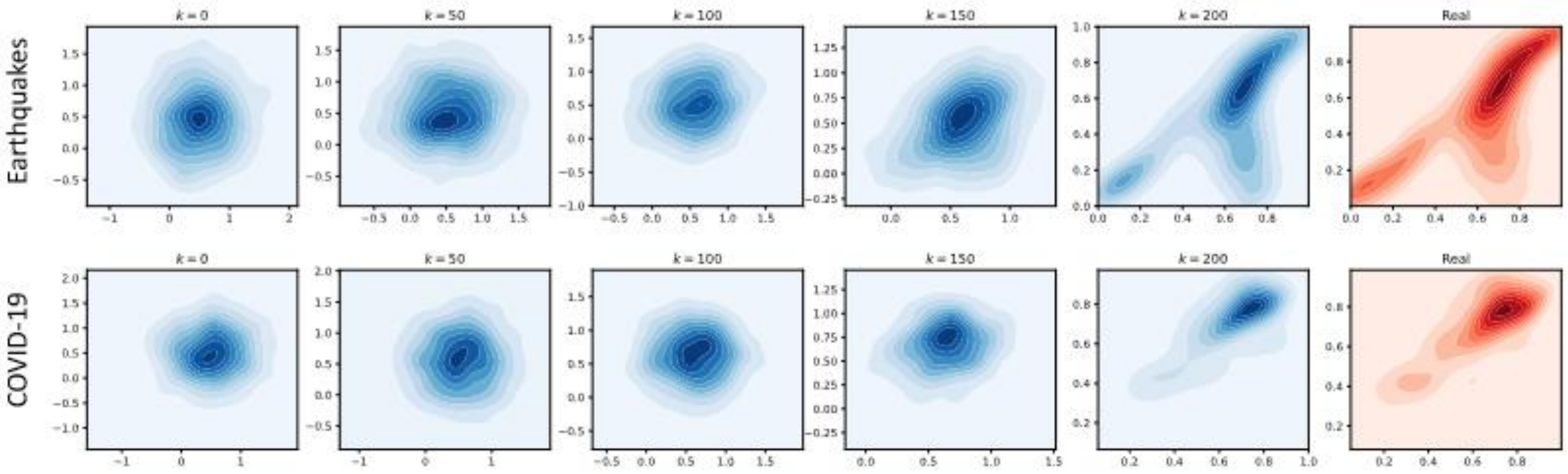
$$x_{t,i} = \sigma(W_t \tau_i^{k+1} + b_t + W_{th} h_{t,i-1} + b_{th} + e_k)$$



时空扩散点过程

可视化

- 与图像的扩散模型不同，这里的每张图片都是多次采样的结果综合出来的。
- 在预测下一事件时，采样300次，得到300个事件 （事件，位置）
- 然后用这300个位置绘制等密度曲线图。



感谢聆听