

The Landscape of Deep Generative Learning

扩散模型在遥感影像生成中的应用

GISLab 2025 夏季短期课程

陈振源

浙江大学地球科学学院

2025 年 6 月 25 日

bili_sakura@zju.edu.cn

Restricted
Boltzmann Machines

Bayesian Networks

Variational
Autoencoders

Normalizing
Flows

Energy-based
Models

Generative
Adversarial Networks

Autoregressive
Models

Denoising
Diffusion Models

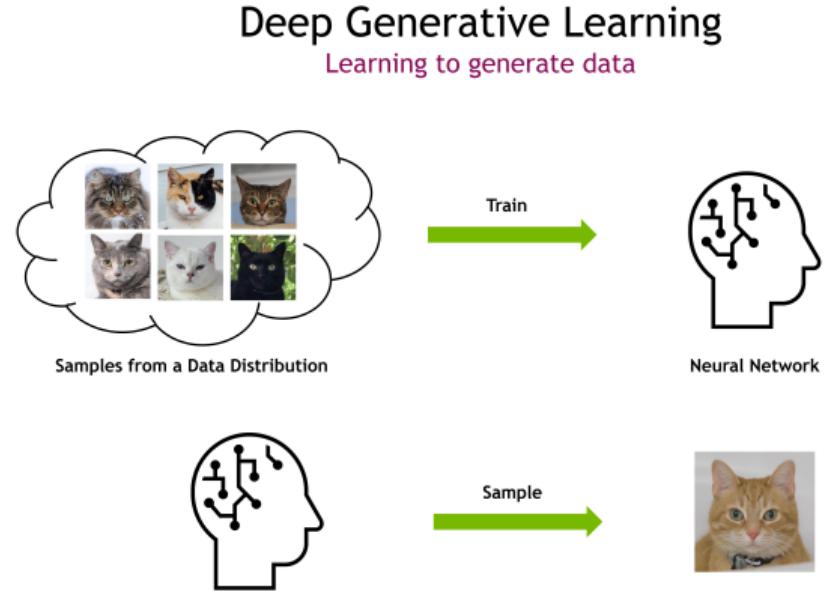
图片来源: Vahdat, Arash, Song, and Meng, 2023.

目录

- ▶ 背景：图像生成的生成式模型
- ▶ 扩散模型：原理
- ▶ 遥感中的应用
- ▶ 总结和思考

背景：图像生成的生成式模型

背景：图像生成的生成式模型



图：生成式建模示意图 (Vahdat, Arash, Song, and Meng, 2023)。

2

生成式模型发展时间线

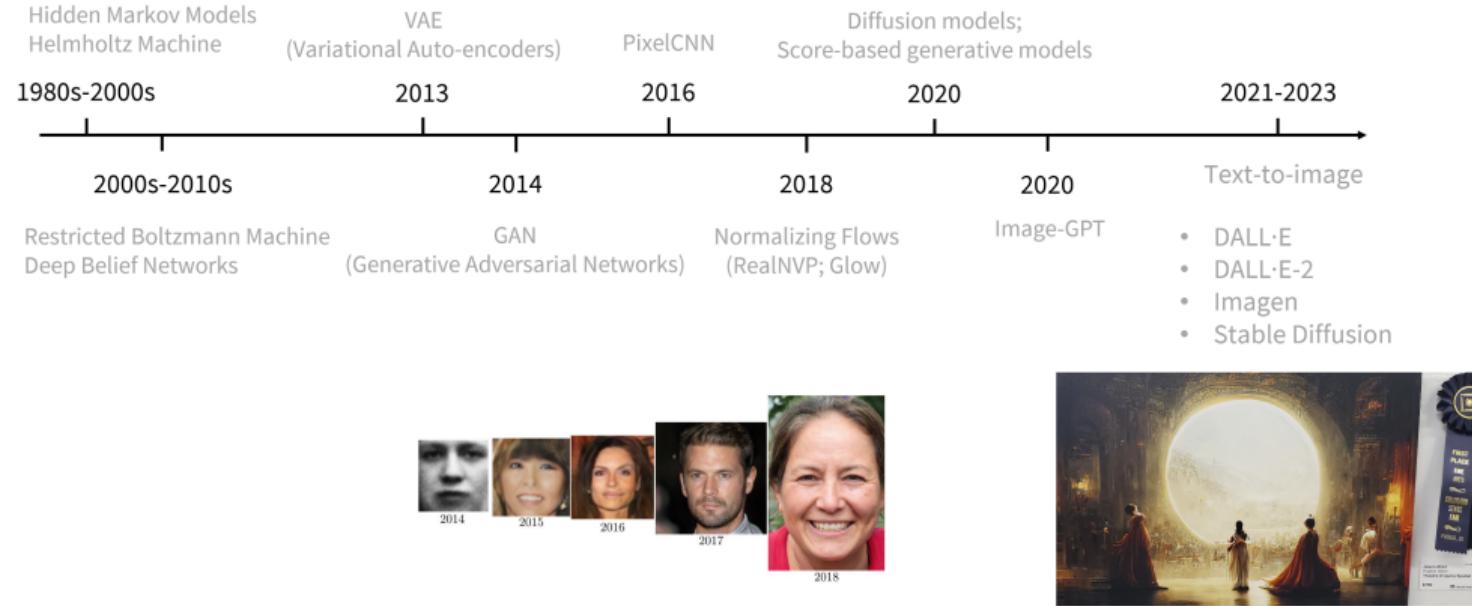


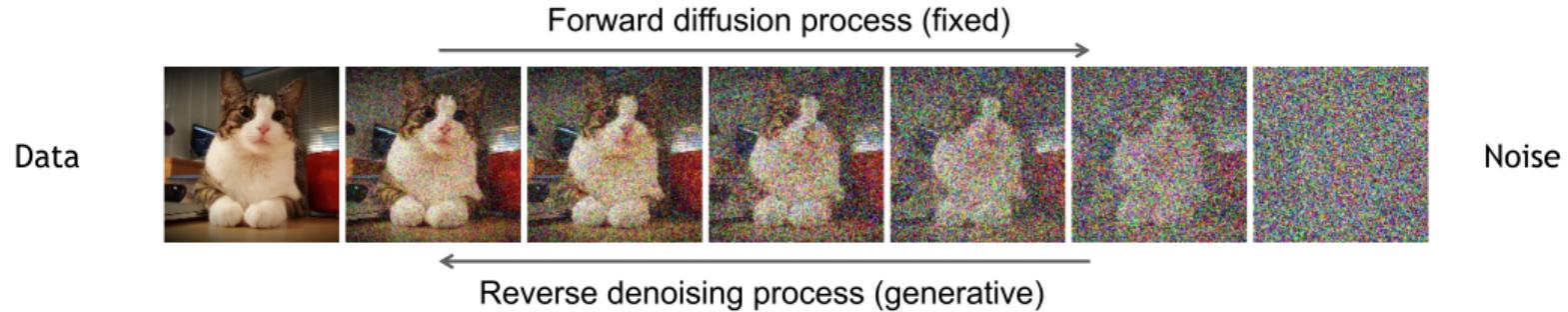
图: 生成式模型关键进展时间线 (Deng, 2024)。

扩散模型：原理

去噪扩散模型

去噪扩散模型包含两个过程：

- ▶ 正向扩散过程：逐步向输入中添加噪声。
- ▶ 反向去噪过程：通过去噪学习生成数据。



图：扩散模型通过迭代去噪生成数据 (Sohl-Dickstein et al., 2015; Ho, Jain, and Abbeel, 2020)。图片来源：Vahdat, Arash, Song, and Meng, 2023。

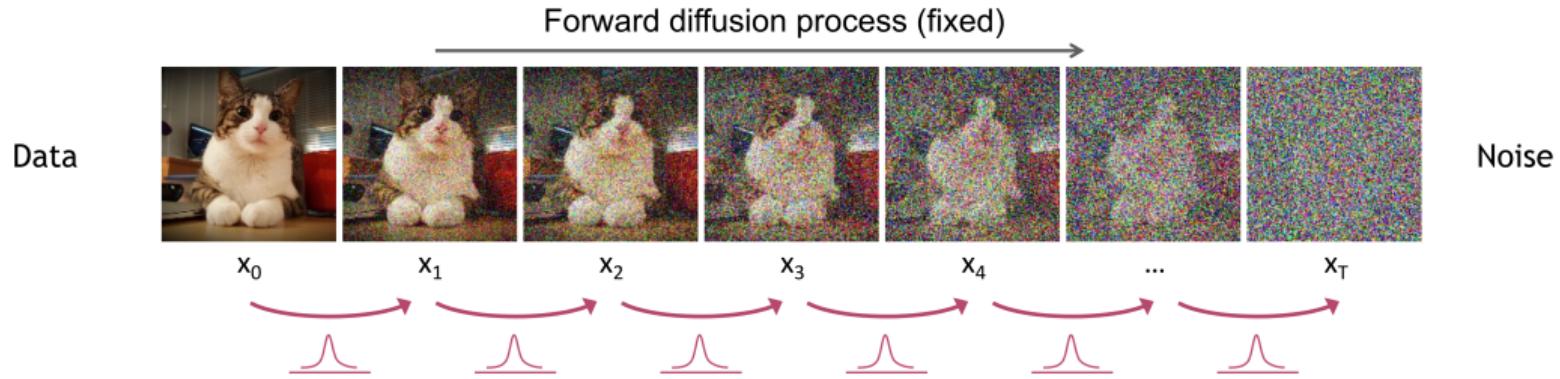
Sohl-Dickstein, et al. Deep Unsupervised Learning using Nonequilibrium Thermodynamics, ICML, 2015.

Ho, et al. Denoising Diffusion Probabilistic Models, NeurIPS, 2020.

Vahdat, Arash et al. CVPR 2023 Tutorial Denoising Diffusion-based Generative Modeling: Foundations and Applications, 2023.

正向扩散过程

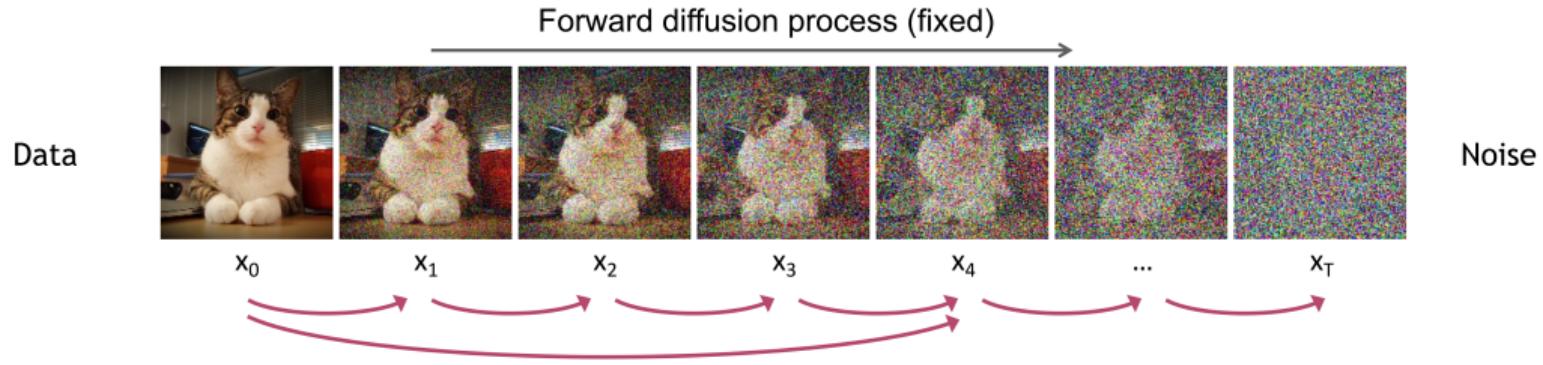
正向过程在 T 步内的形式化定义：



$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N} \left(\mathbf{x}_t; \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I} \right) \quad \Rightarrow \quad q(\mathbf{x}_{1:T} | \mathbf{x}_0) = \prod_{t=1}^T q(\mathbf{x}_t | \mathbf{x}_{t-1}) \quad (\text{联合分布})$$

图片来源：Vahdat, Arash, Song, and Meng, 2023。

扩散核



定义: $\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s) \quad \Rightarrow \quad q(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t} \mathbf{x}_0, (1 - \bar{\alpha}_t) \mathbf{I}) \quad (\text{扩散核})$

采样公式: $\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}$ 其中 $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

噪声调度 $\{\beta_t\}$ 的选择使得 $\bar{\alpha}_T \rightarrow 0$, 且 $q(\mathbf{x}_T \mid \mathbf{x}_0) \approx \mathcal{N}(\mathbf{x}_T; \mathbf{0}, \mathbf{I})$ 。

图片来源: Vahdat, Arash, Song, and Meng, 2023.

Vahdat, Arash et al. CVPR 2023 Tutorial Denoising Diffusion-based Generative Modeling: Foundations and Applications, 2023.

正向扩散过程与展开推导

正向扩散过程摘要：

$$x_t = \sqrt{\alpha_t}x_{t-1} + \sqrt{1 - \alpha_t}\epsilon_t, \quad q(x_t | x_{t-1}) \sim \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t \mathbf{I})$$

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, \quad \bar{\alpha}_t = \prod_{i=1}^t \alpha_i$$

详细展开推导：

$$x_t = \sqrt{\alpha_t}x_{t-1} + \sqrt{1 - \alpha_t}\epsilon_t$$

$$x_t = \sqrt{\alpha_t} \left(\sqrt{\alpha_{t-1}}x_{t-2} + \sqrt{1 - \alpha_{t-1}}\epsilon_{t-1} \right) + \sqrt{1 - \alpha_t}\epsilon_t$$

$$x_t = \sqrt{\alpha_t \alpha_{t-1}}x_{t-2} + \sqrt{\alpha_t} \sqrt{1 - \alpha_{t-1}}\epsilon_{t-1} + \sqrt{1 - \alpha_t}\epsilon_t$$

$$x_t = \sqrt{\alpha_t \alpha_{t-1} \alpha_{t-2}}x_{t-3} + \cdots + \sqrt{1 - \alpha_t \alpha_{t-1}}\epsilon$$

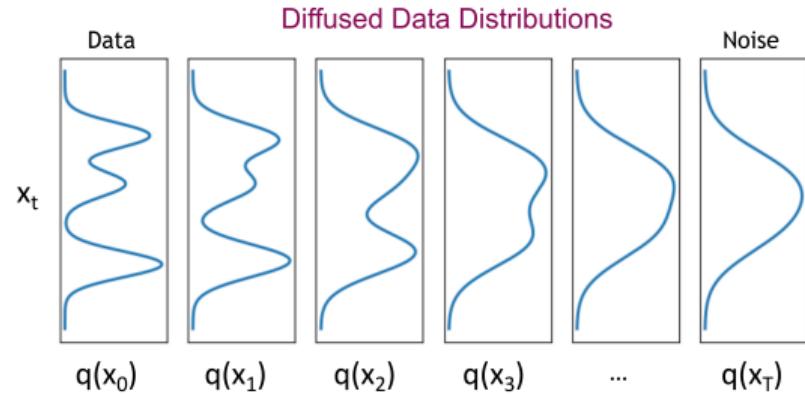
$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon$$

正向扩散中分布会发生什么？

到目前为止，我们讨论了扩散核 $q(\mathbf{x}_t|\mathbf{x}_0)$ ，但 $q(\mathbf{x}_t)$ 呢？

$$q(\mathbf{x}_t) = \underbrace{\int q(\mathbf{x}_0, \mathbf{x}_t) d\mathbf{x}_0}_{\text{Diffused data dist.}} = \underbrace{\int q(\mathbf{x}_0) \underbrace{q(\mathbf{x}_t|\mathbf{x}_0)}_{\text{Input data dist.}} d\mathbf{x}_0}_{\text{Joint dist.}} = \int q(\mathbf{x}_0) q(\mathbf{x}_t|\mathbf{x}_0) d\mathbf{x}_0$$

The diffusion kernel is Gaussian convolution.



我们可以先采样 $\mathbf{x}_0 \sim q(\mathbf{x}_0)$ ，再采样 $\mathbf{x}_t \sim q(\mathbf{x}_t|\mathbf{x}_0)$ ，从而得到 $\mathbf{x}_t \sim q(\mathbf{x}_t)$ （即祖先采样）。

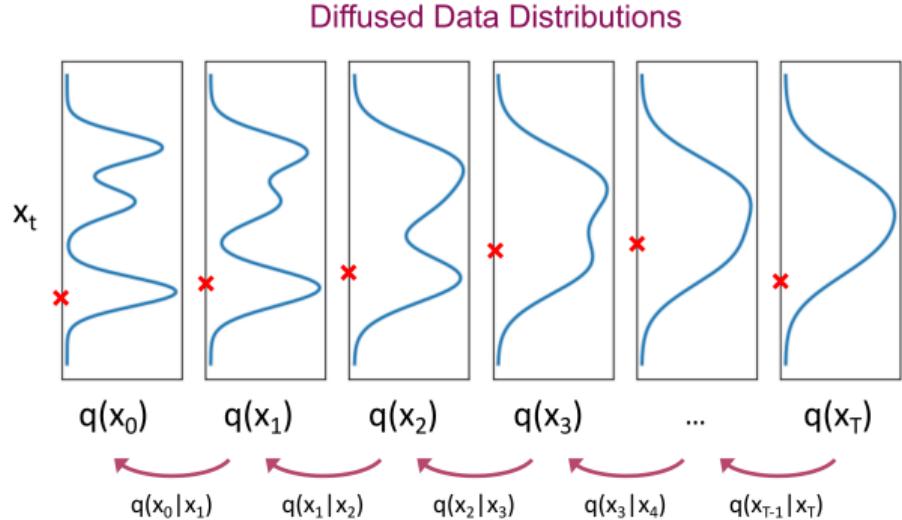
图片来源：Vahdat, Arash, Song, and Meng, 2023。

通过去噪进行生成式学习

回顾：扩散参数被设计为 $q(\mathbf{x}_T) \approx \mathcal{N}(\mathbf{x}_T; \mathbf{0}, \mathbf{I})$ 。

生成过程：

- ▶ 采样 $\mathbf{x}_T \sim \mathcal{N}(\mathbf{x}_T; \mathbf{0}, \mathbf{I})$
- ▶ 迭代采样 $\mathbf{x}_{t-1} \sim \underbrace{q(\mathbf{x}_{t-1} | \mathbf{x}_t)}_{\text{真实去噪分布}}$

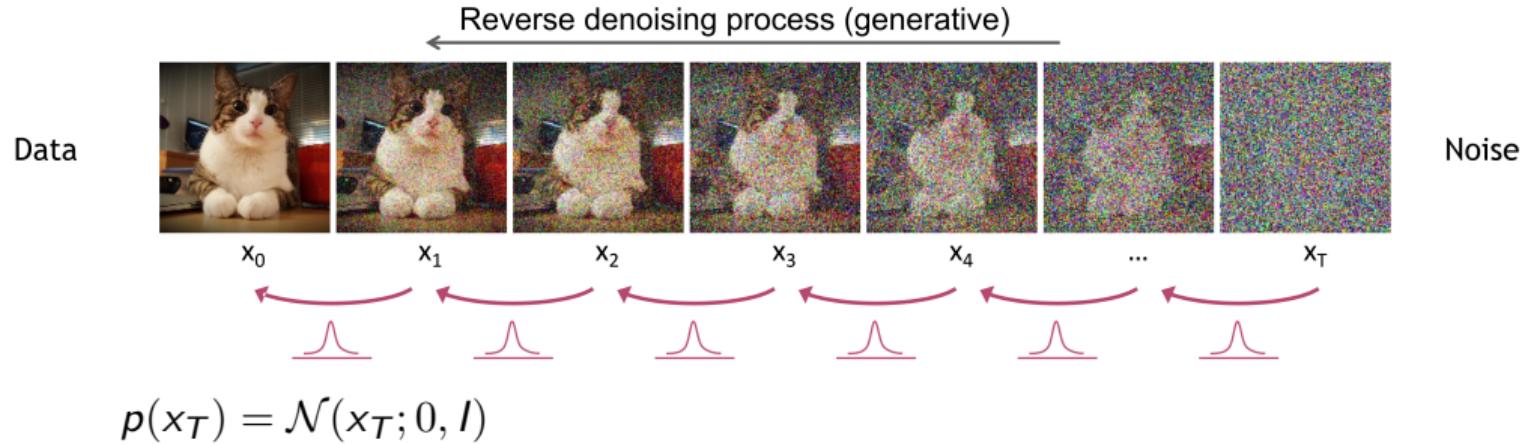


我们能否近似 $q(\mathbf{x}_{t-1} | \mathbf{x}_t)$? 答案是可以的，如果每一步正向扩散中的 β_t 很小，则可以用高斯分布近似。

图片来源：Vahdat, Arash, Song, and Meng, 2023。

反向去噪过程

正向与反向过程在 T 步内的形式化定义：



$$p_\theta(x_{t-1} | x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2 I) \quad \Rightarrow \quad p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^T p_\theta(x_{t-1} | x_t).$$

其中 $\mu_\theta(x_t, t)$ 是可训练的神经网络（如 U-Net、去噪自编码器）
图片来源：Vahdat, Arash, Song, and Meng, 2023。

去噪模型的学习

变分上界

- ▶ 训练时，采用变分上界（类似 VAE）：

$$\mathbb{E}_{q_\lambda} [\log p_\theta(\mathbf{x})] \leq \mathbb{E}_{q_\lambda} \left[\log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q_\lambda(\mathbf{z}|\mathbf{x})} \right] = L$$

- ▶ $\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}$, 均值参数化方式见 (Ho, Jain, and Abbeel, 2020)：

$$\mu_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\bar{\alpha}_t}} \left(\mathbf{x}_t - \frac{1 - \bar{\alpha}_t}{\sqrt{1 - \bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right)$$

- ▶ 变分目标函数：

$$L = \mathbb{E}_{q(\mathbf{x}_0, \boldsymbol{\epsilon})} \left[\sum_{t=1}^T \lambda_t \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)} \left[\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}, t)\|^2 \right] \right]$$

- ▶ 设置 $\lambda_t = 1$ 对所有 t 效果最佳 (Ho, Jain, and Abbeel, 2020)。

小结

训练与采样过程

Algorithm1-Training

```
1: 重复
2:  x0 ~ q(x0)
3:  t ~ Uniform({1, ..., T})
4:  ε ~ N(0, I)
5:  对以下目标做梯度下降
    ∇θ \|ε - εθ (sqrt(αt)x0 + sqrt(1 - αt)ε, t)\|^2
6:  直到收敛
```

Algorithm2-Sampling

```
1:  xT ~ N(0, I)
2:  for t = T, ..., 1 do
3:      z ~ N(0, I)
4:      xt-1 = 1/sqrt(αt) (xt - (1 - αt)/sqrt(1 - αt) εθ(xt, t)) + σtz
5:  end for
6:  return x0
```

算法来源：去噪概率扩散模型 DDPM (Ho, Jain, and Abbeel, 2020)

遥感中的应用

DiffusionSat 简介

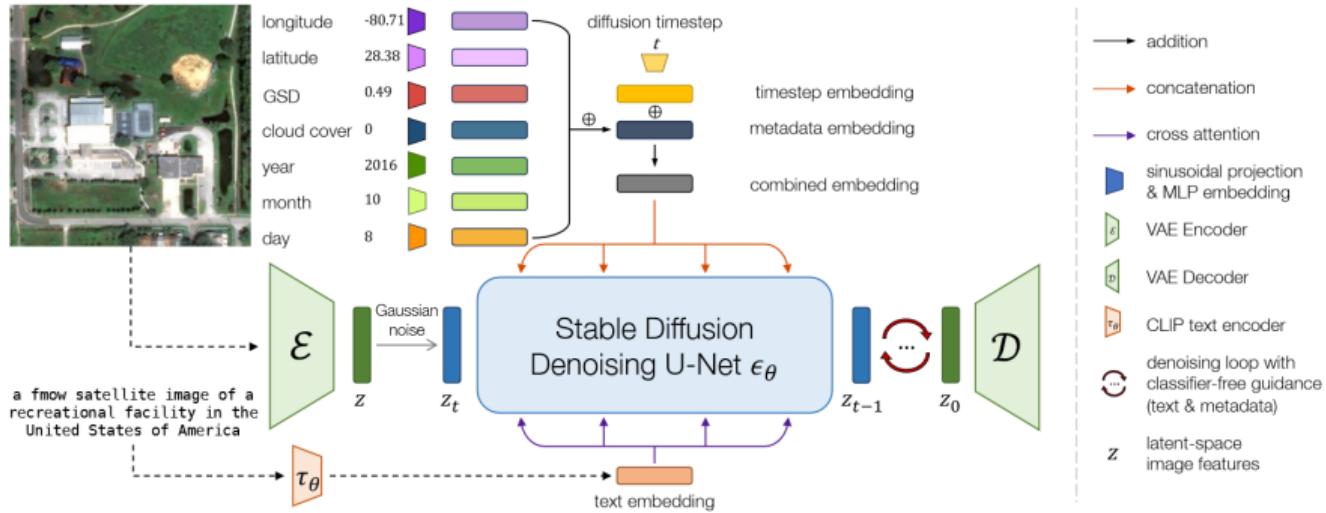


图: DiffusionSat base 基础模型的整体架构, 展示了如何将自由获取的元数据 (传感器类型、日期、位置) 与扩散主干网络融合, 以生成高保真卫星影像 (Khanna et al., 2024)。

DiffusionSat 的文本编码器

- ▶ **输入提示:** "A satellite image of a farmland"
- ▶ **分词:**
 - ▶ 拆分为子词 → token
 - ▶ 示例映射:
{"A": 101, "satellite": 564, ..., <EOS>: 102}
- ▶ **CLIP 文本编码器:**
 - ▶ Token ID → 512 维嵌入
 - ▶ 捕捉语义特征

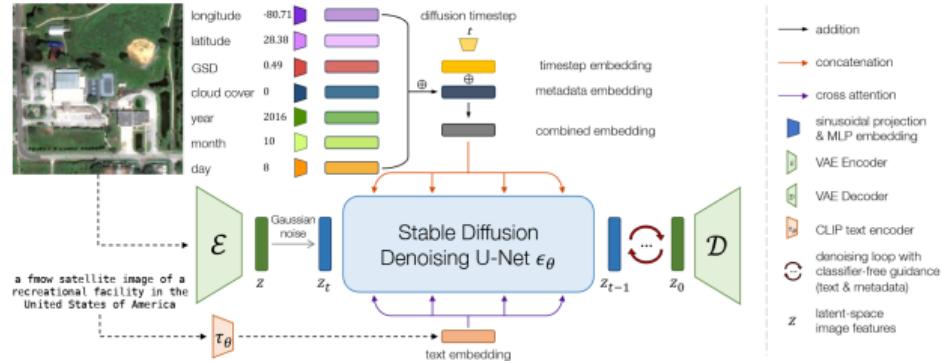


图: 文本编码器模块 (OpenAI CLIP) 对输入提示进行分词，并生成 512 维嵌入用于条件扩散主干网络 (Khanna et al., 2024)。

DiffusionSat 的元数据编码器

▶ 输入元数据示例:

- ▶ 传感器: Sentinel-2
- ▶ 位置: (纬度: 37.7749, 经度: -122.4194)
- ▶ 日期: 2022-06-01
- ▶ GSD: 10 m
- ▶ 云量: 5%

▶ 处理模块: 元数据编码器

▶ 输出: 512 维条件嵌入

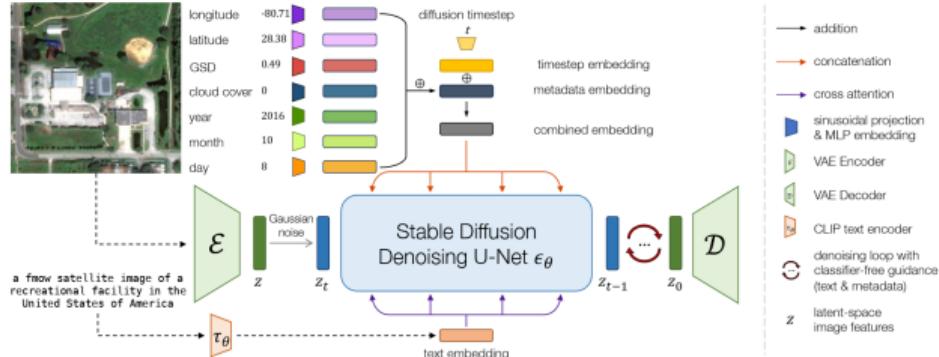


图: 元数据编码器将原始卫星元数据转换为定长嵌入，用于条件扩散主干网络 (Khanna et al., 2024)。

图像处理与扩散步骤

▶ 训练流程:

- ▶ 输入: 干净的卫星影像
- ▶ 编码: 图像编码器 → 隐变量
- ▶ 正向扩散: 在 T 步中加入高斯噪声
- ▶ 条件: 注入文本和元数据嵌入
- ▶ 学习: UNet 预测并去除噪声

▶ 推理流程:

- ▶ 输入: 随机噪声隐变量 $x_T \sim \mathcal{N}(0, I)$
- ▶ 反向扩散: 通过条件 UNet 迭代去噪
- ▶ 解码: 隐变量解码器 → 最终高保真影像

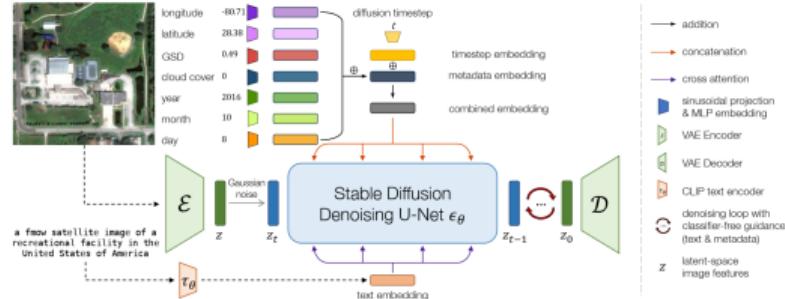


图: DiffusionSat 基础模型中训练 (正向扩散) 与推理 (反向去噪) 流程的对比, 展示了影像如何经过编码器、扩散主干和解码器 (Khanna et al., 2024)。

反向扩散：采样

▶ **生成:** $x \sim p(\text{noise}, c_{\text{text}}, c_{\text{meta}})$

▶ **其中:**

- ▶ noise: $x_T \sim \mathcal{N}(0, I)$
- ▶ c_{text} : 文本 (嵌入)
- ▶ c_{meta} : 元数据 (嵌入)

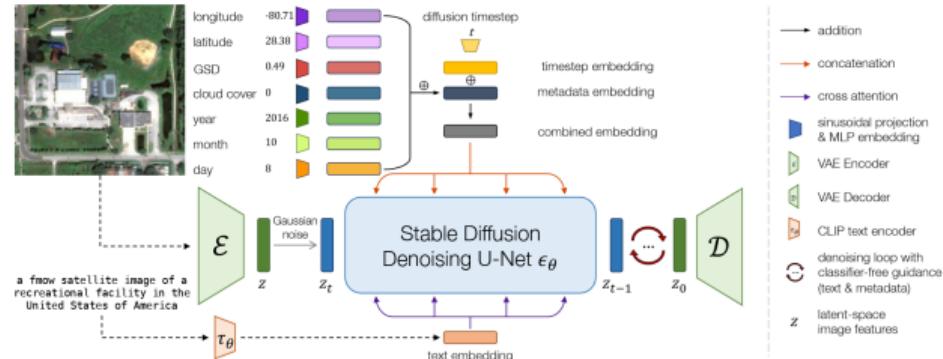


图: DiffusionSat 基础模型示意图 (Khanna et al., 2024)。

DiffusionSat+3DControlNet：框架概览

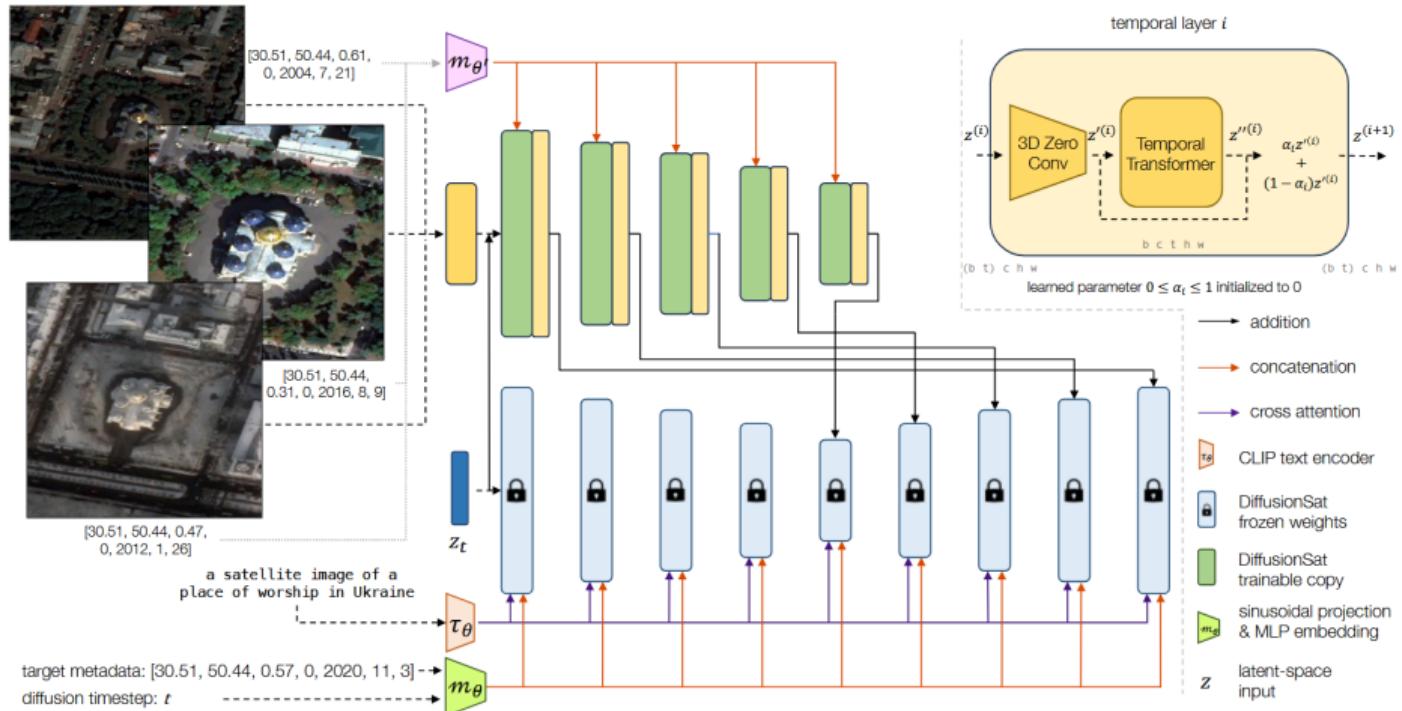


图: DiffusionSat 中的 3DControlNet。 (Khanna et al., 2024)

DiffusionSat+3DControlNet: 时序预测结果



图: fMoW-temporal 数据集上的时序预测生成样例。 (Khanna et al., 2024)

模型	$t' > t$			$t' < t$		
	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
SD + 3D CN	0.2027	11.0536	0.5523	0.2181	11.3004	0.5342
DiffusionSat + CN	0.3297	13.6938	0.5062	0.2862	12.4990	0.5307
DiffusionSat + 3D CN	0.3983	13.7886	0.4304	0.4293	14.8699	0.3937

表: 表 4: fMoW-temporal 验证集上的样本质量定量结果。 $t' > t$ 表示已知未来影像生成过去影像, $t' < t$ 表示已知过去影像生成未来影像。

DiffusionSat+3DControlNet：超分辨率结果

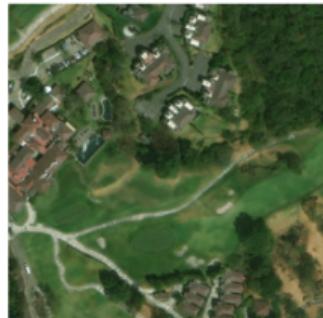


图: DiffusionSat 在多光谱超分辨率任务中的示例结果。([Khanna et al., 2024](#))

进一步讨论

利用生成式人工智能生成新图像

- ▶ 生成式模型可以生成新的、逼真的图像。
- ▶ 我们可以用它们来扩充训练数据。
- ▶ 示例：给定“前”图像和描述，生成新的“后”图像。



+ suffer from
volcano eruption

C_T

$$X_{post} \sim p(X|X_{pre}, C_T)$$



Generative Models



补充内容

我们如何进行数据增强？

经典方法：

- ▶ 翻转、旋转、裁剪、改变颜色等。

现代方法：

- ▶ 融合两张图片 (Mixup) (Zhang et al., 2018)。
- ▶ 剪切并粘贴图片部分 (CutMix) (Yun et al., 2019)。



图：现代数据增强方法示意图。从左到右依次为：Mixup (Zhang et al., 2018), Cutout (DeVries and Taylor, 2017), CutMix (Yun et al., 2019)。

软标签示例 (CutMix)：

$$\text{cutmix_label} = \lambda \cdot \text{label}_A + (1 - \lambda) \cdot \text{label}_B$$

示例: $\lambda = 0.5$, $\text{label}_A = [1, 0]$, $\text{label}_B = [0, 1]$

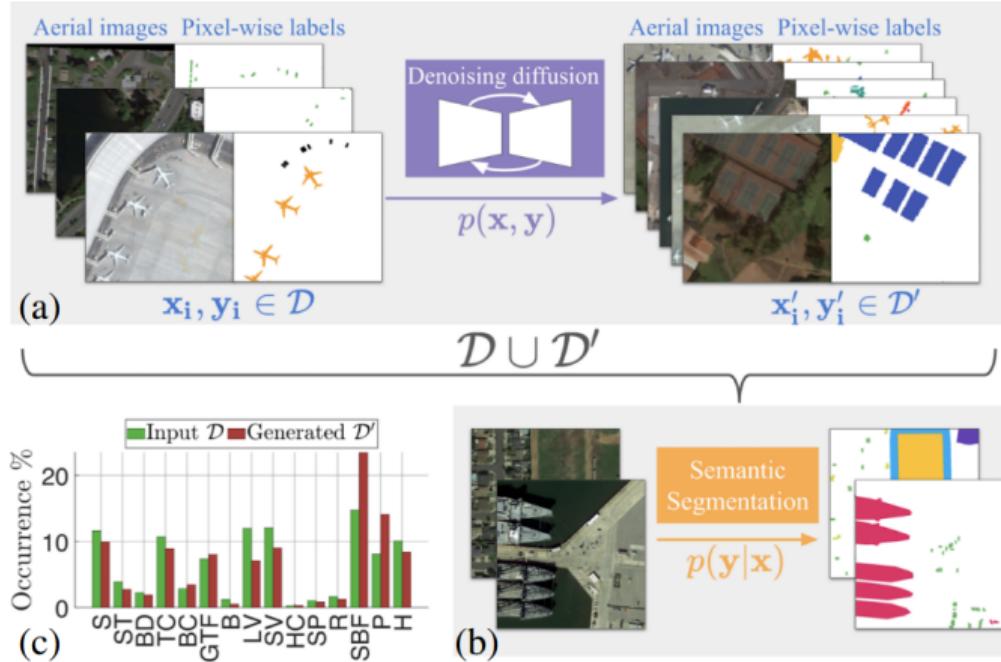
$$\text{cutmix_label} = 0.5 \times [1, 0] + 0.5 \times [0, 1] = [0.5, 0.5]$$

Zhang, et al. Mixup: Beyond Empirical risk minimization, ICLR, 2018.

Yun, et al. CutMix: Regularization Strategy to Train Strong Classifiers With Localizable Features, ICCV, 2019.

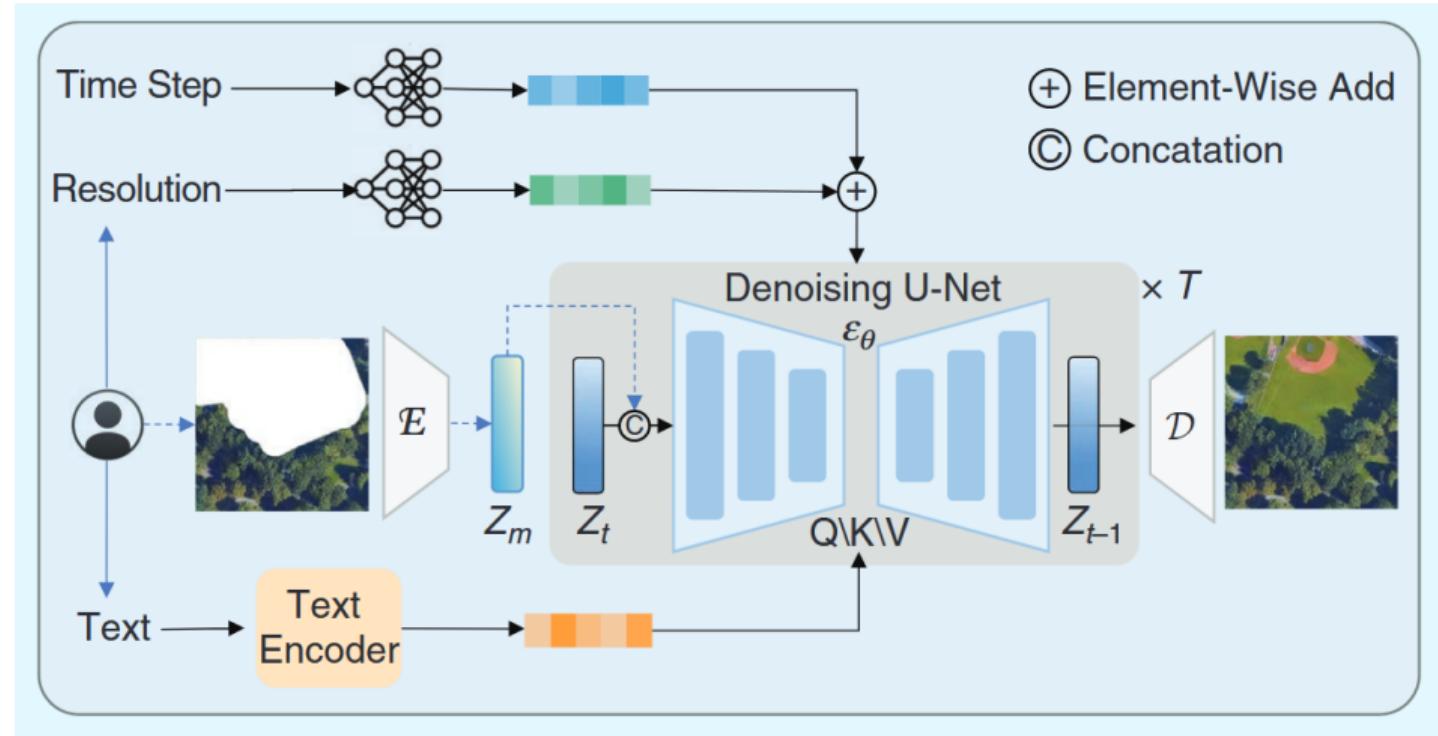
DeVries, et al. Improved Regularization of Convolutional Neural Networks with Cutout, arXiv, 2017.

生成式模型用于数据增强



SatSyn (Toker et al., 2024) 提出了一种生成式模型（扩散模型），可同时生成卫星分割的图像和对应掩码。该合成数据集用于数据增强，在卫星语义分割任务中，相比其他数据增强方法带来了显著的定量提升。

遥感图像生成中的应用：Text2Earth



图：Text2Earth：面向文本驱动地球观测的基础模型 (Liu et al., 2025)。

Text2Earth: 示例结果



图: Text2Earth 生成的示例结果 (Liu et al., 2025)。

遥感图像生成中的应用：CRS-Diff

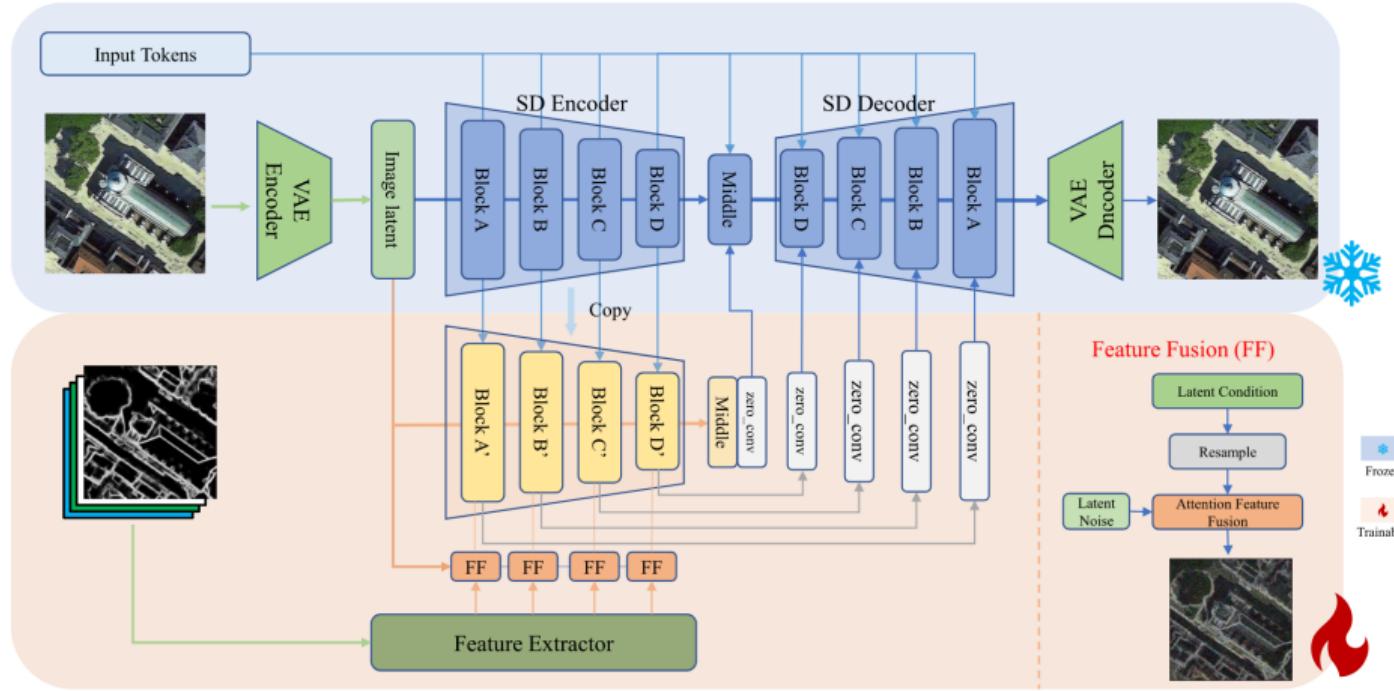


图: CRS-Diff: 可控遥感图像生成框架 (Tang, Li, et al., 2024)。

CRS-Diff: 示例结果

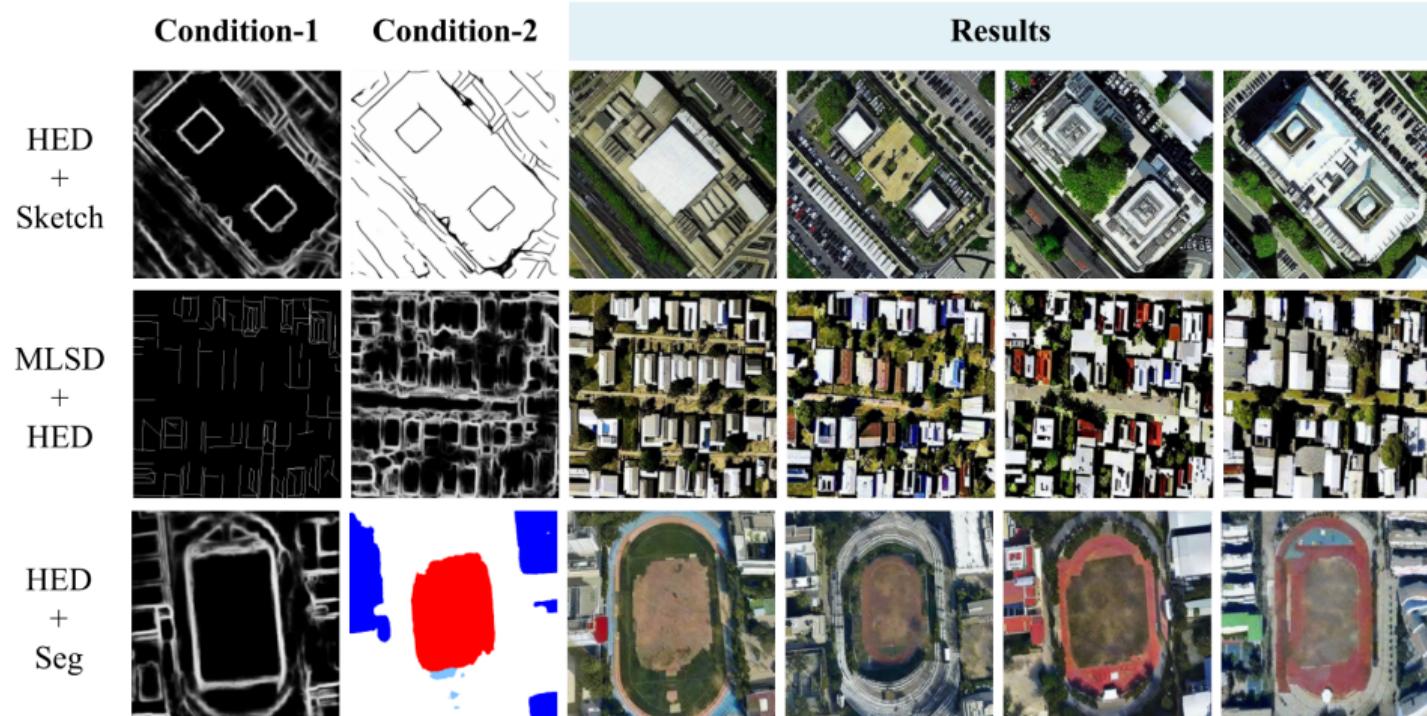


图: CRS-Diff 生成的示例结果 (Tang, Li, et al., 2024)。