

生成模型合成数据已准备好用于图像识别了吗？

GISLab 2025 年暑期短课程

陈振源

浙江大学地球科学学院

2025

bili_sakura@zju.edu.cn

提纲

- ▶ 1. 基于深度学习的图像分类简介
- ▶ 2. 传统数据增强方法
- ▶ 3. 用于数据增强的生成模型
- ▶ 4. 灾害事件遥感数据集：xBD
- ▶ 项目 - **探索生成图像是否能提升图像分类效果**

图像分类：概述



图：图像分类概述。

背景：深度学习下的图像分类

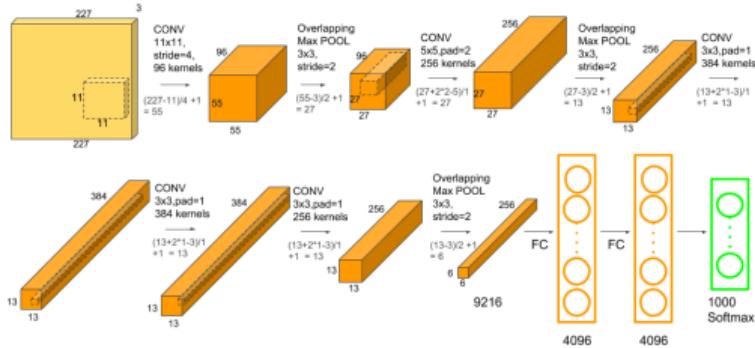
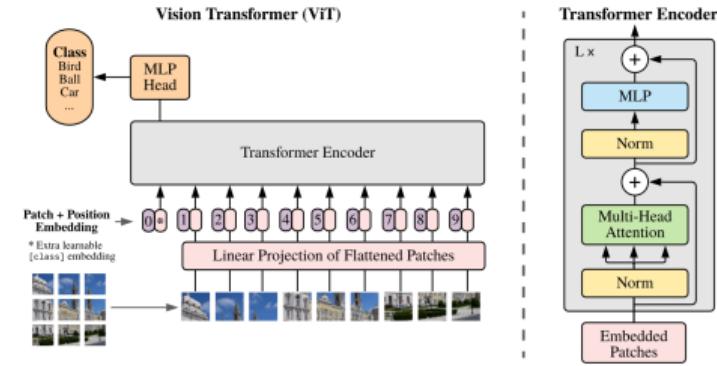


图: 左: ILSVRC-2010 上的 AlexNet (Berg, Deng, and Fei-Fei, 2010) 右: AlexNet 结构 (Krizhevsky, Sutskever, and Hinton, 2012)。

图像分类架构演进

- ▶ 2012: AlexNet, 2016: ResNet
- ▶ 2021: ViT
(Dosovitskiy et al., 2021)
- ▶ 2021: Swin Transformer
(Liu et al., 2021)
- ▶ 2021: CLIP-ViT
(Radford et al., 2021)
- ▶ 2022: MAE-ViT
(He et al., 2022)
- ▶ 2022: CoCa-ViT
(Yu et al., 2022)



Vision Transformer 结构概览
(Dosovitskiy et al., 2021).

Dosovitskiy, et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale, ICLR, 2021.

Liu, et al. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows, ICCV, 2021.

Radford, et al. Learning Transferable Visual Models From Natural Language Supervision, ICML, 2021.

He, et al. Masked Autoencoders Are Scalable Vision Learners, CVPR, 2022.

Yu, et al. CoCa: Contrastive Captioners Are Image-Text Foundation Models. TMLR, 2022.

图像分类数据集：RESISC45



airplane



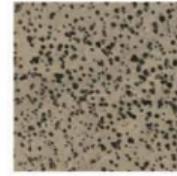
airport



baseball diamond



bridge



chaparral



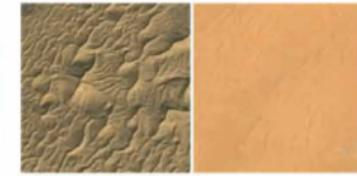
church



commercial area



dense residential



desert

RESISC45 遥感场景分类数据集示例图像 (Cheng, Han, and Lu, 2017)。

Cheng, et al. Remote Sensing Image Scene Classification: Benchmark and State of the Art. Proceedings of the IEEE. 2017.

传统数据增强方法

- ▶ **几何变换：旋转、翻转（水平/垂直）、缩放、平移、裁剪**
- ▶ **颜色扰动：**调整亮度、对比度、饱和度和色调
- ▶ **噪声注入：**向图像中添加随机噪声
- ▶ **Cutout** (DeVries and Taylor, 2017)
- ▶ **CutMix** (Yun et al., 2019)
- ▶ **Copy-Paste** (Ghiasi et al., 2021)

还有一项综合研究《How to train your ViT? Data, Augmentation, and Regularization in Vision Transformers》(Steiner et al., 2022)。

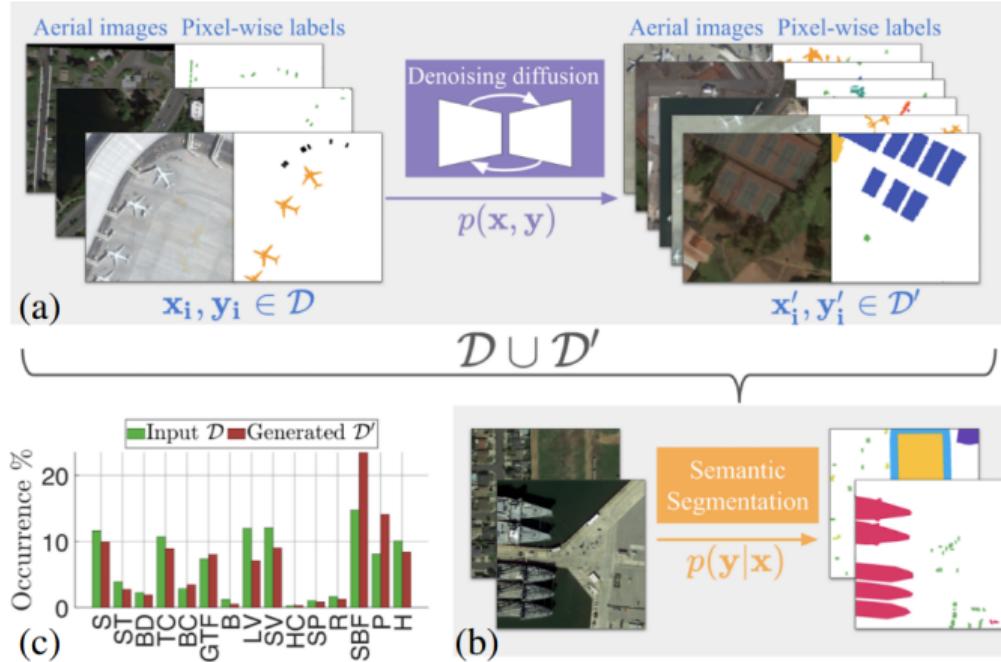
DeVries, et al. Improved Regularization of Convolutional Neural Networks with Cutout, arXiv, 2017.

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

Ghiasi, et al. Simple copy-paste is a strong data augmentation method for instance segmentation, CVPR, 2021.

Steiner, et al. How to Train Your ViT? Data, Augmentation, and Regularization in Vision Transformers. TMLR. 2022.

用于数据增强的生成模型



SatSyn (Toker et al., 2024) 提出了一种生成模型(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增强方法带来了显著的定量提升。

生成的文本-图像数据集提升图像分类

Dataset	Task	CLIP-RN50	CLIP-RN50+SYN	CLIP-ViT-B/16	CLIP-ViT-B/16+SYN
CIFAR-10	o	70.31	80.06 (+9.75)	90.80	92.37 (+1.57)
CIFAR-100	o	35.35	45.69 (+10.34)	68.22	70.71 (+2.49)
Caltech101	o	86.09	87.74 (+1.65)	92.98	94.16 (+1.18)
Caltech256	o	73.36	75.74 (+2.38)	80.14	81.43 (+1.29)
ImageNet	o	60.33	60.78 (+0.45)	68.75	69.16 (+0.41)
SUN397	s	58.51	60.07 (+1.56)	62.51	63.79 (+1.28)
Aircraft	f	17.34	21.94 (+4.60)	24.81	30.78 (+5.97)
Birdsnap	f	34.33	38.05 (+3.72)	41.90	46.84 (+4.94)
Cars	f	55.63	56.93 (+1.30)	65.23	66.86 (+1.63)
CUB	f	46.69	56.94 (+10.25)	55.23	63.79 (+8.56)
Flower	f	66.08	67.05 (+0.97)	71.30	72.60 (+1.30)
Food	f	80.34	80.35 (+0.01)	88.75	88.83 (+0.08)
Pets	f	85.80	86.81 (+1.01)	89.10	90.41 (+1.31)
DTD	t	42.23	43.19 (+0.96)	44.39	44.92 (+0.53)
EuroSAT	si	37.51	55.37 (+17.86)	47.77	59.86 (+12.09)
ImageNet-Sketch	r	33.29	36.55 (+3.26)	46.20	48.47 (+2.27)
ImageNet-R	r	56.16	59.37 (+3.21)	74.01	76.41 (+2.40)
Average	/	55.13	59.47 (+4.31)	65.42	68.32 (+2.90)

Table 1: **Main Results on Zero-shot Image Recognition.** All results are top-1 accuracy on test set.
o: object-level. s: scene-level. f: fine-grained. t: textures. si: satellite images. r: robustness.

由生成模型合成的文本-图像数据集可以显著提升图像分类性能，如 (He et al., 2023) 所示。

xBD: 大规模灾害损失数据集

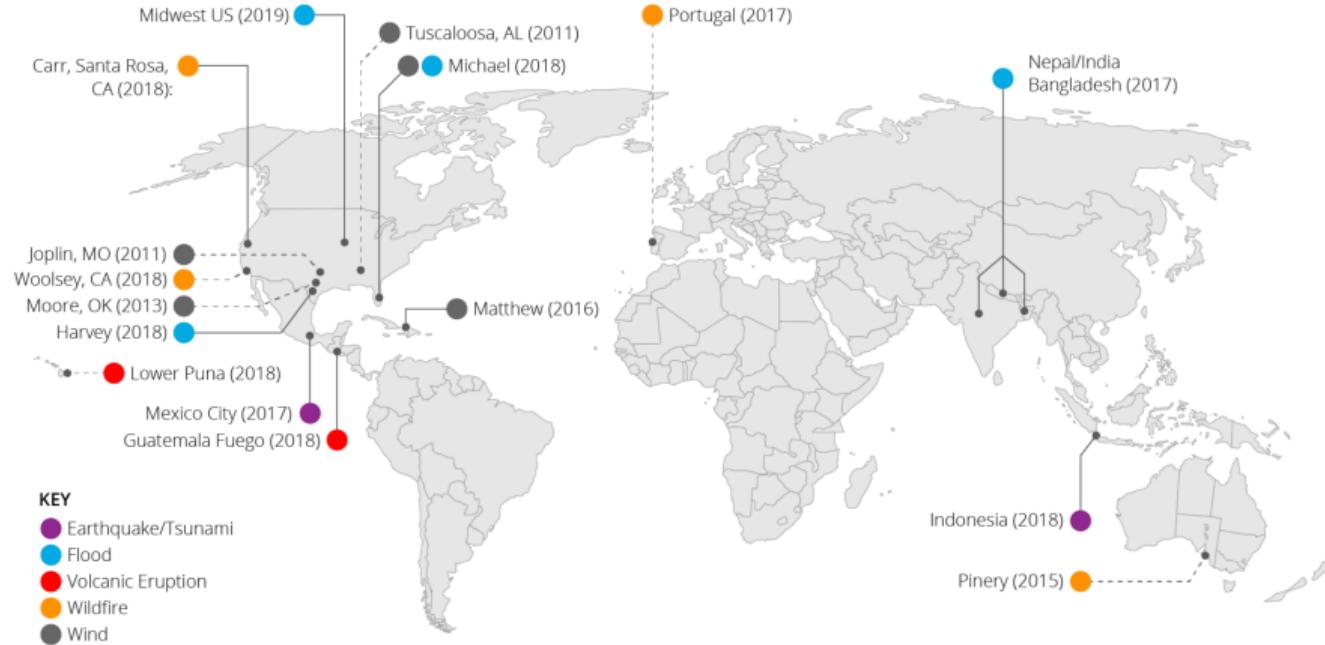


灾前影像 (上) 与灾后影像 (下)。从左到右依次为：哈维飓风、乔普林龙卷风、下普纳火山喷发、巽他海峡海啸。

影像来源：DigitalGlobe。

xBD (Gupta et al., 2019)

xBD: 全球灾害类型覆盖



xBD 数据集在全球范围内涵盖的灾害类型及事件。
xBD (Gupta et al., 2019)

项目作业：总体介绍

项目：生成图像能否提升遥感图像分类？

目标：

探索将真实图像与生成图像结合，是否能提升遥感图像分类效果。

实验流程：

本项目将通过三种不同训练设置，评估生成数据的影响：

1. **仅用真实数据集**: 仅用xBD 数据集中的真实图像训练分类模型。
2. **仅用生成数据集**: 仅用商业生成模型合成的图像训练模型。
3. **真实+生成数据集**: 同时用真实图像和生成图像训练模型。

比较三种设置下的分类性能，分析合成数据的作用。

项目作业：数据集

数据集：xBD 灾害损失数据集

- ▶ 使用 **xBD** 遥感灾害数据集。
- ▶ 数据集包含 **6 类灾害**。
- ▶ 每类灾害选取 **100 张真实图像** (共 600 张真实图像)。

项目作业：生成模型

图像生成：

- ▶ 生成图像视为**文本引导的图像编辑**结果：每个案例输入**灾前图像和文本描述**（如“洪水”、“建筑倒塌”），模型生成灾后图像。
- ▶ 可使用商业生成模型，如**GPT-4o 图像生成 (GPT-Image-1)** (OpenAI, 2025)、**Gemini-2** (Google, 2024) 或 **SeedEdit 3.0** (Wang et al., 2025)，为每类灾害生成合成图像。

项目作业：分类模型

推荐基线模型：

- ▶ **OpenAI CLIP** (Radford et al., 2021) - 模型
- ▶ **RemoteCLIP** (Liu, Chen, Guan, et al., 2024) - 模型
- ▶ **Git-RSCLIP** (Liu, Chen, Zhao, et al., 2025) - 模型

以上模型均为 ViT 结构。代码与教程参考：

- ▶ [CLIP 训练示例](#)
- ▶ [ViT 教程](#)
- ▶ [更多遥感基础模型见 **huggingface 合集**](#)

项目作业：数据增强方案

- ▶ 每类灾害生成 $1\times\sim 4\times$ 数量的合成图像（即每类 100、200、300 或 400 张合成图像）。
- ▶ 探索并比较不同真实与合成图像比例（如 1:1、1:2、1:3、1:4）。
- ▶ 每类灾害增强后数据集规模为 **200 到 500 张**。

项目作业：评估

评估方式：

- ▶ 使用**标准准确率、F1 分数和混淆矩阵**衡量性能。
- ▶ 始终在**保留的真实（未见过的）测试集**上评估。
- ▶ 绘制**曲线或柱状图**，比较不同真实: 合成比例下的分类性能。

附录

更多文本-图像遥感数据集

文本到图像生成：

- ▶ **RSICD** (Lu et al., 2018): 遥感图像描述数据集，含 10,921 张图像，每张配有 5 条描述。
- ▶ **RSICap** (Hu et al., 2025): 高质量数据集，包含 2,585 个人工标注的图像-描述对。
- ▶ **UCM-Captions** (Qu et al., 2016): 基于 UC Merced 土地利用数据集，含 2,100 张图像，每张配 5 条描述。
- ▶ **RESISC45** (Cheng, Han, and Lu, 2017): 公开的遥感场景分类基准数据集，由西北工业大学创建，共 31,500 张图像，覆盖 45 类场景，每类 700 张。

Lu, et al. Exploring Models and Data for Remote Sensing Image Caption Generation. TGRS. 2018.

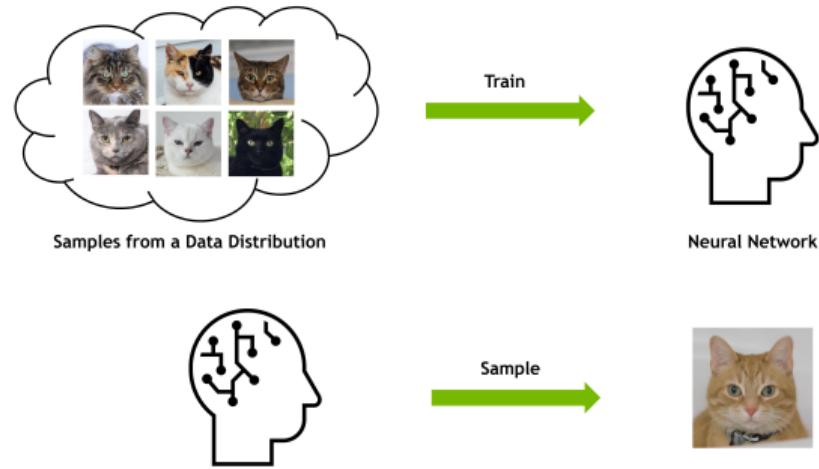
Hu, et al. RSGPT: A remote sensing vision language model and benchmark. ISPRS. 2025.

Qu, et al. Deep semantic understanding of high resolution remote sensing image, CITS, 2016.

Cheng, et al. Remote Sensing Image Scene Classification: Benchmark and State of the Art. Proceedings of the IEEE. 2017.

生成建模

Deep Generative Learning Learning to generate data



2

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

生成模型发展时间线

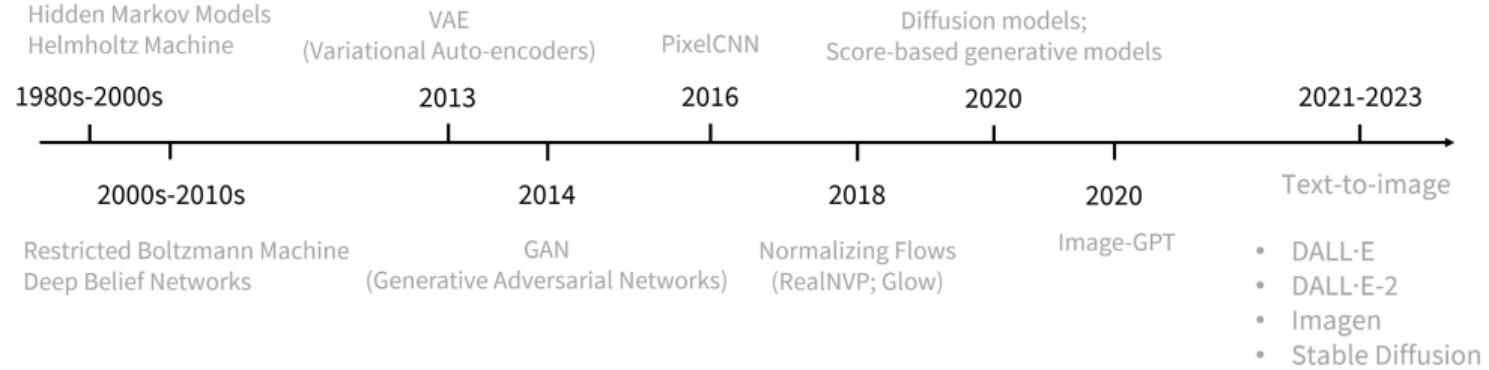
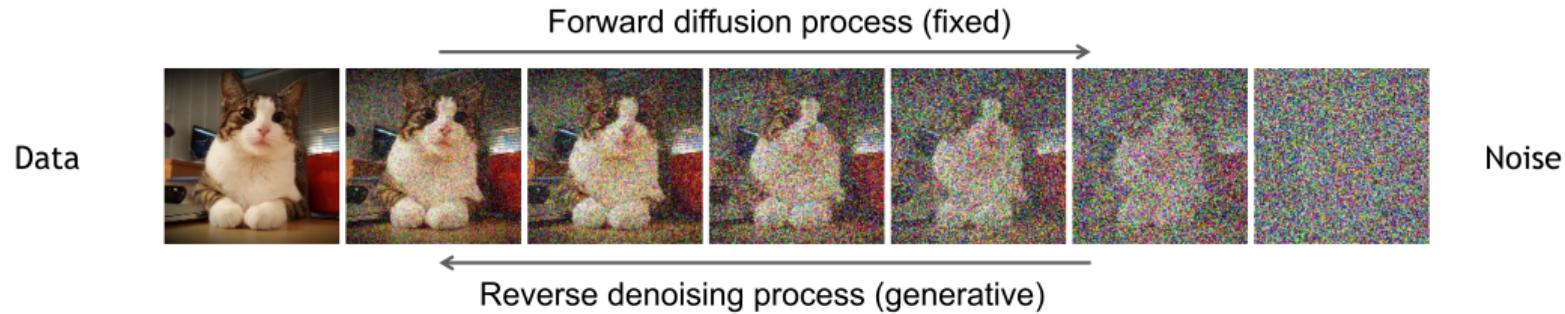


图: 生成模型关键进展时间线 (Deng, 2024)。

背景：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)。

扩散模型：正向与反向过程

正向（扩散）过程：

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

$$q(\mathbf{x}_{1:T} \mid \mathbf{x}_0) = \prod_{t=1}^T q(\mathbf{x}_t \mid \mathbf{x}_{t-1})$$

等价于 $\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

反向（去噪）过程：

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

$$p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T) \prod_{t=1}^T p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$$

其中 \mathbf{x}_0 为数据, β_t 为噪声调度, $\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s)$, $p(\mathbf{x}_T) = \mathcal{N}(\mathbf{0}, \mathbf{I})$ 。

扩散模型通过学习逆转逐步加噪过程来生成数据。 (Sohl-Dickstein et al., 2015; Ho, Jain, and Abbeel, 2020)

扩散模型：训练与推理

训练目标：

$$\mathcal{L}_{\text{simple}} = \mathbb{E}_{x_0, \epsilon, t} \left[\|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, t)\|^2 \right]$$

其中 $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s)$ 。

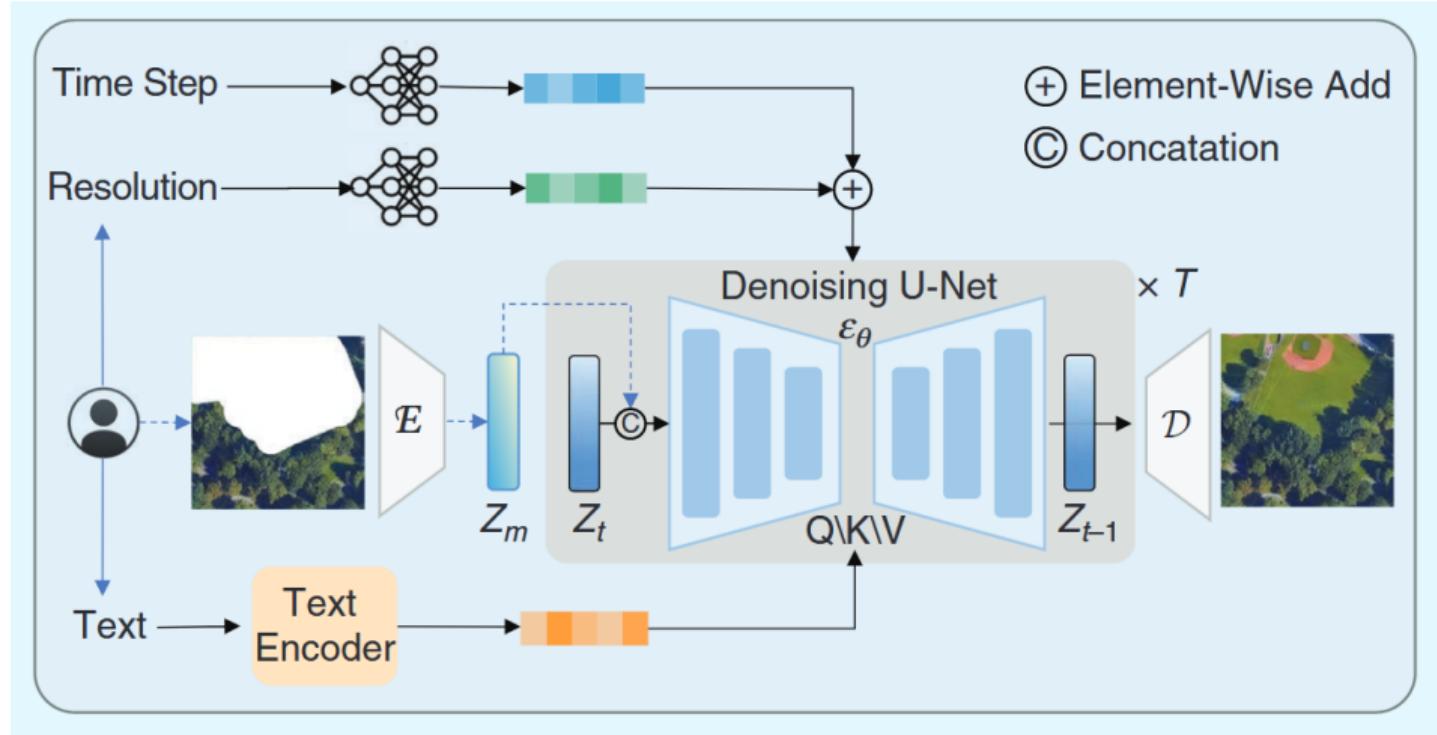
推理（采样）：

- ▶ 从纯噪声开始: $x_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
- ▶ 对于 $t = T, \dots, 1$:
 - ▶ 预测噪声: $\epsilon_\theta(x_t, t)$
 - ▶ 计算均值: $\mu_\theta(x_t, t)$
 - ▶ 采样: $x_{t-1} \sim \mathcal{N}(\mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$
- ▶ 重复直到 x_0 (生成样本)

训练: 最小化简化目标 (Ho, Jain, and Abbeel, 2020)。

推理: 通过迭代去噪从随机噪声生成数据。

遥感图像生成应用：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

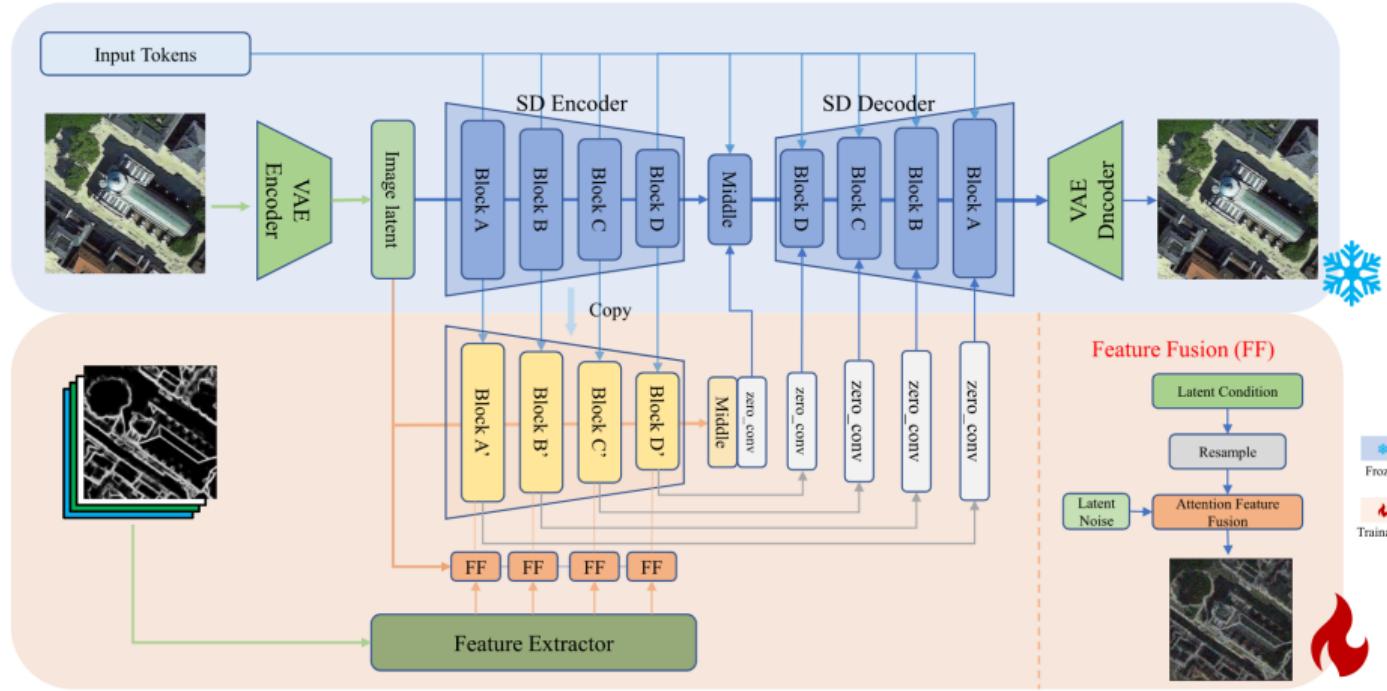


图: CRS-Diff: 可控遥感图像生成框架 (Tang, Li, et al., 2024)。

CRS-Diff: 示例结果

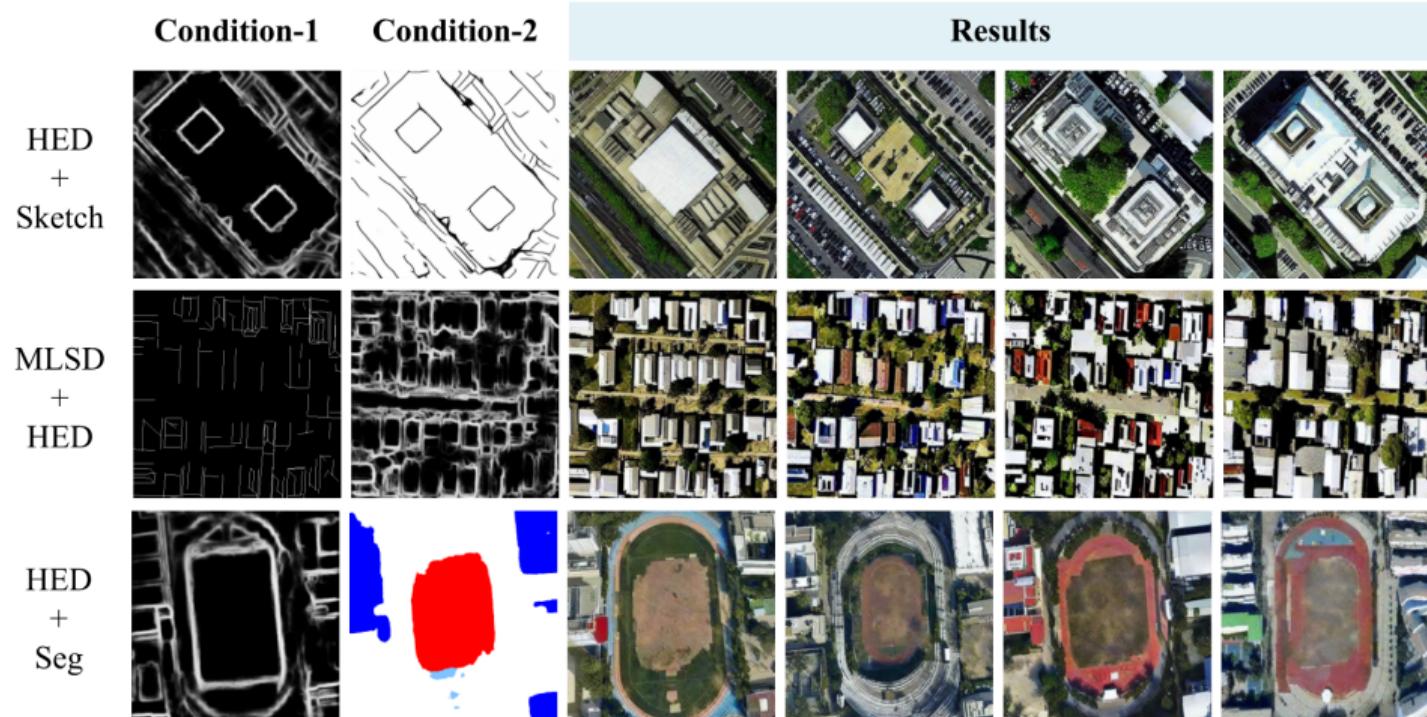


图: CRS-Diff 生成的示例结果 (Tang, Li, et al., 2024)。

DiffusionSat: 框架概览

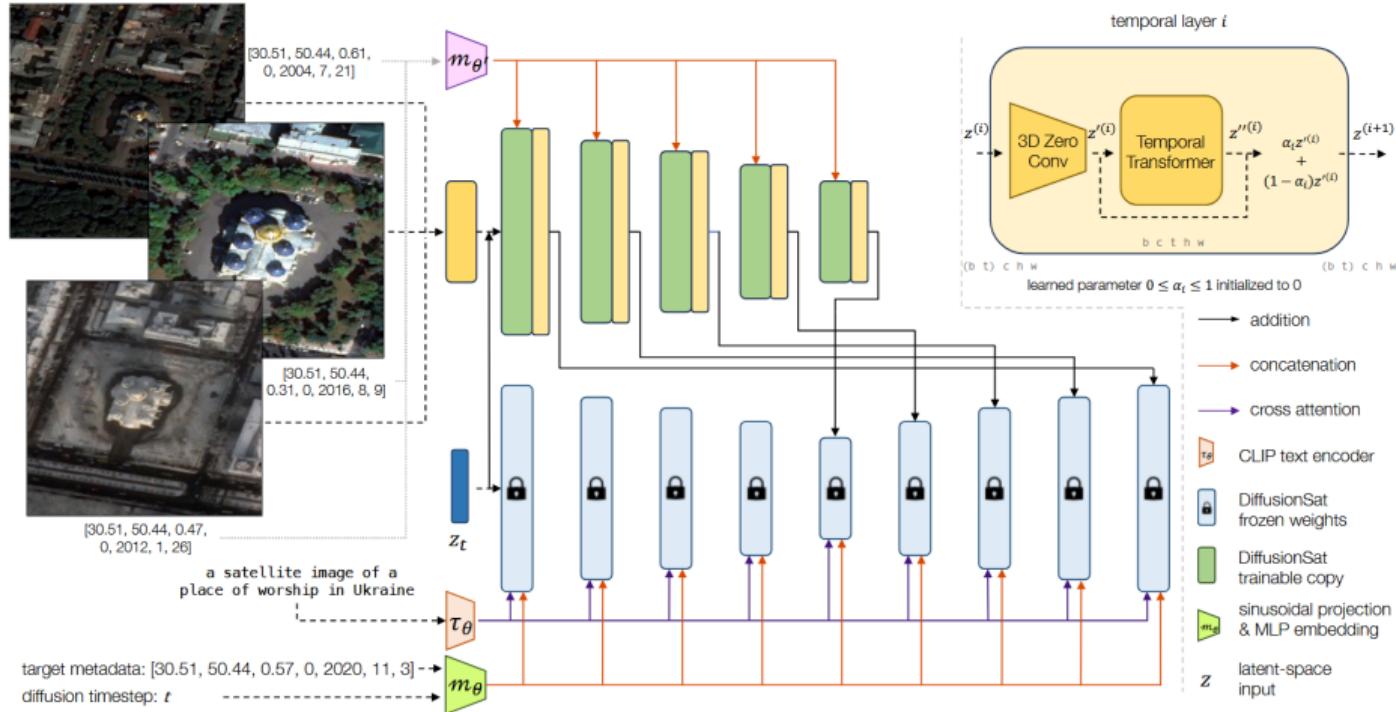
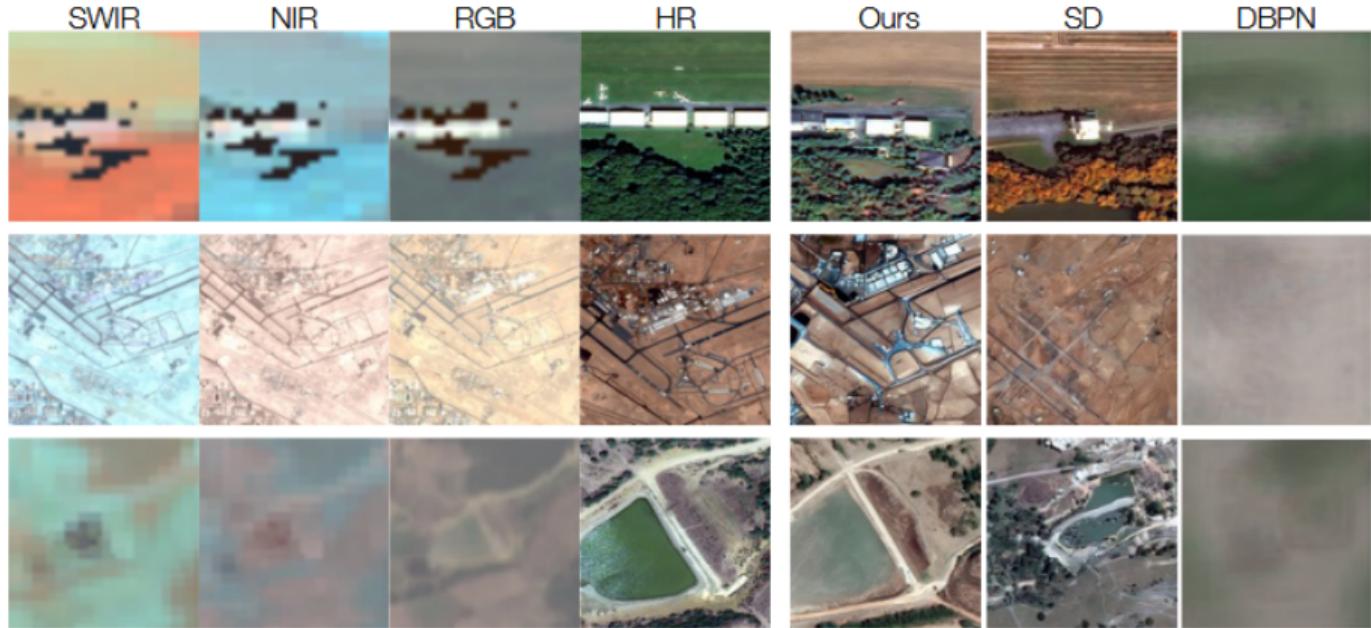


图: DiffusionSat: 卫星影像生成基础模型 (Khanna et al., 2024)。

DiffusionSat：超分辨率结果



图：DiffusionSat 多光谱超分辨率示例结果 (Khanna et al., 2024)。

DiffusionSat: 修复结果



图: DiffusionSat 遥感图像修复示例结果 (Khanna et al., 2024)。