合成数据是否有助于遥感下游任务的能力提升？

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

内容提要

- 生成模型与扩散方法
- 遥感图像生成应用
- 研究课题: 使用合成数据进行数据增广并应用于遥感图像分类和超分
  1. 背景
  2. 遥感图像分类
  3. 遥感图像超分
  4. 数据集和基线模型
  5. 预期结果

生成式建模

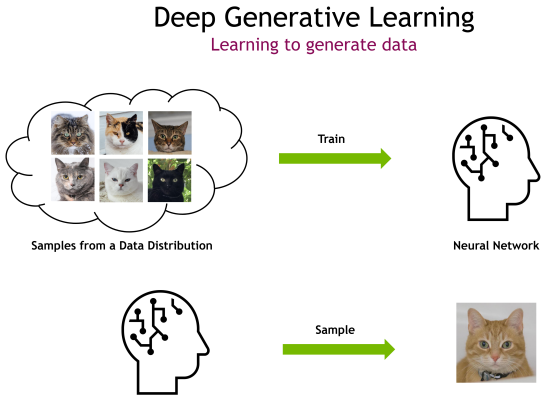

图: 生成式建模示意图 (Vahdat，Arash, Song, and Meng, 2023).

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

生成模型发展时间线

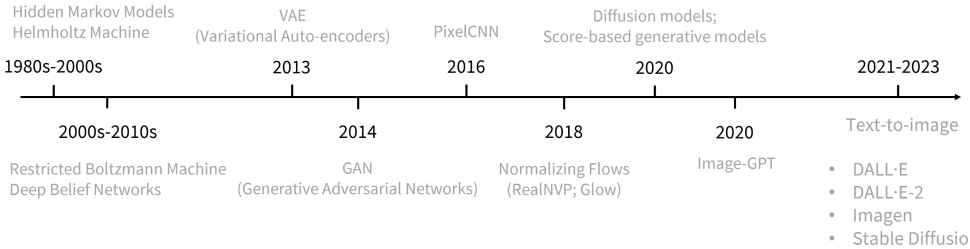

图: 生成模型关键进展时间线 (Deng, 2024).

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。

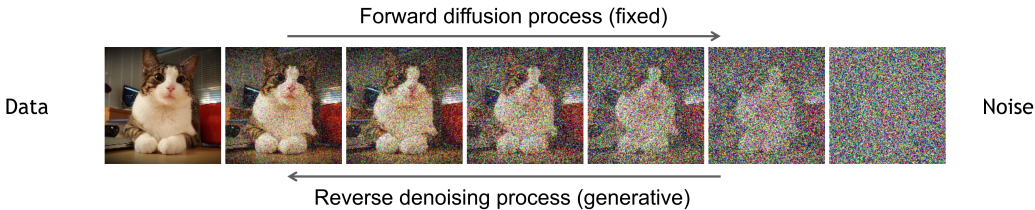

图: 扩散模型通过迭代去噪生成数据 (Sohl-Dickstein et al., 2015; Ho, Jain, and Abbeel, 2020).

Sohl-Dickstein, et al. Deep Unsupervised Learning using Nonequilibrium Thermodynamics, ICML, 2015.
Ho, et al. Denoising Diffusion Probabilistic Models, NeurIPS, 2020.

扩散模型：正向与反向过程

正向（扩散）过程：

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

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

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

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)

Sohl-Dickstein, et al. Deep Unsupervised Learning using Nonequilibrium Thermodynamics, ICML, 2015.
Ho, et al. Denoising Diffusion Probabilistic Models, NeurIPS, 2020.

扩散模型：训练与推理

训练目标：

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

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

推理（采样）：

- 从纯噪声开始：$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
- 对 $t = T, \ldots, 1$：
  - 预测噪声：$\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$
  - 计算均值：$\boldsymbol{\mu}_\theta(\mathbf{x}_t, t)$
  - 采样：$\mathbf{x}_{t-1} \sim \mathcal{N}(\boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t))$
- 重复直到得到 $\mathbf{x}_0$（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。

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

遥感图像生成应用：Text2Earth

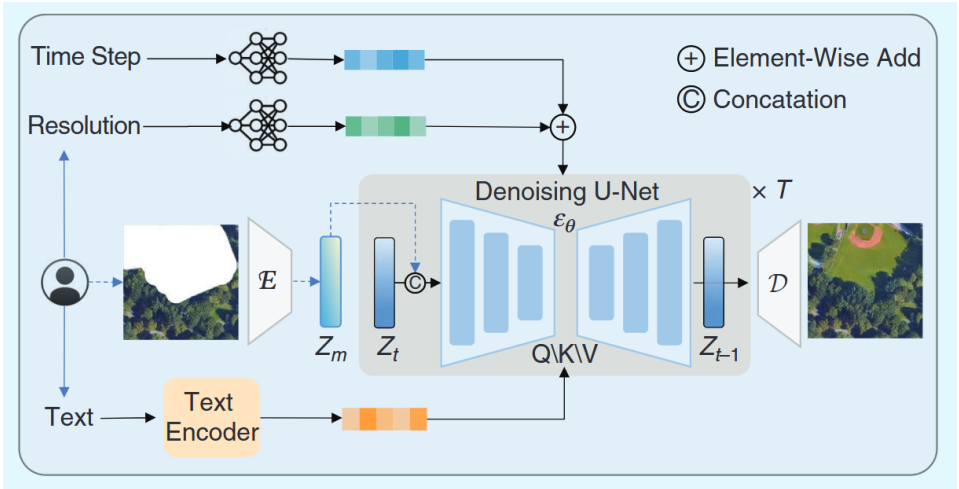

图: Text2Earth：面向文本驱动地球观测的基础模型 (Liu et al., 2025).

Liu, et al. Text2Earth: Unlocking text-driven remote sensing image generation with a global-scale dataset and a foundation model. GRSM. 2025.

Text2Earth：示例结果

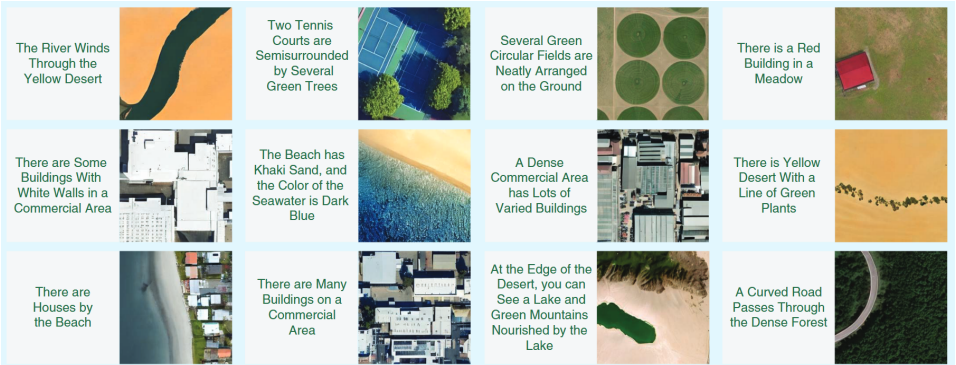

图: Text2Earth 生成的示例结果 (Liu et al., 2025).

Liu, et al. Text2Earth: Unlocking text-driven remote sensing image generation with a global-scale dataset and a foundation model. GRSM. 2025.

遥感图像生成应用：CRS-Diff

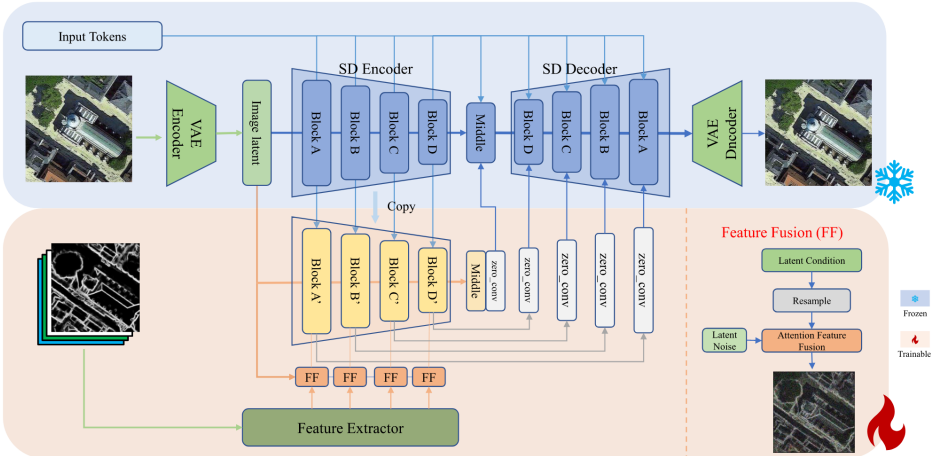

图: CRS-Diff：可控遥感图像生成框架 (Tang, Li, et al., 2024).

Tang, et al. CRS-Diff: Controllable Remote Sensing Image Generation with Diffusion Model. TGRS. 2024.

CRS-Diff：示例结果

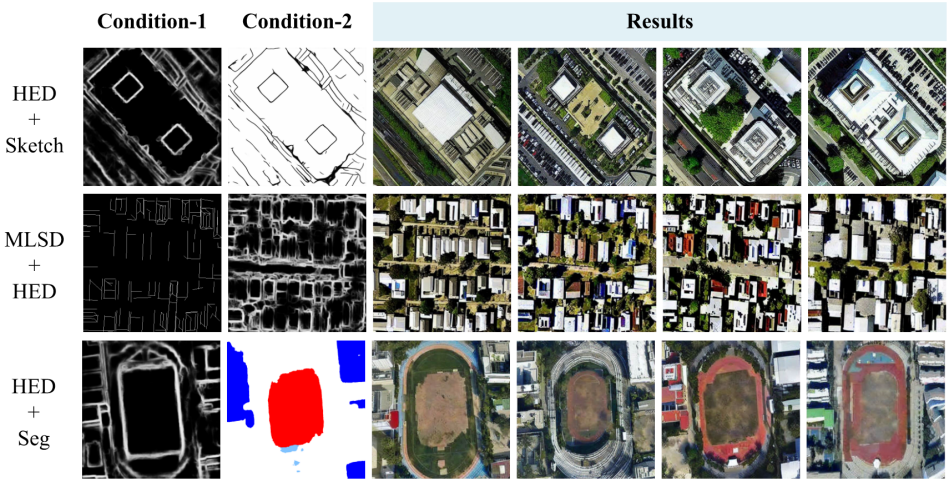

图: CRS-Diff 生成的示例结果 (Tang, Li, et al., 2024).

Tang, et al. CRS-Diff: Controllable Remote Sensing Image Generation with Diffusion Model. TGRS. 2024.

DiffusionSat：框架概览

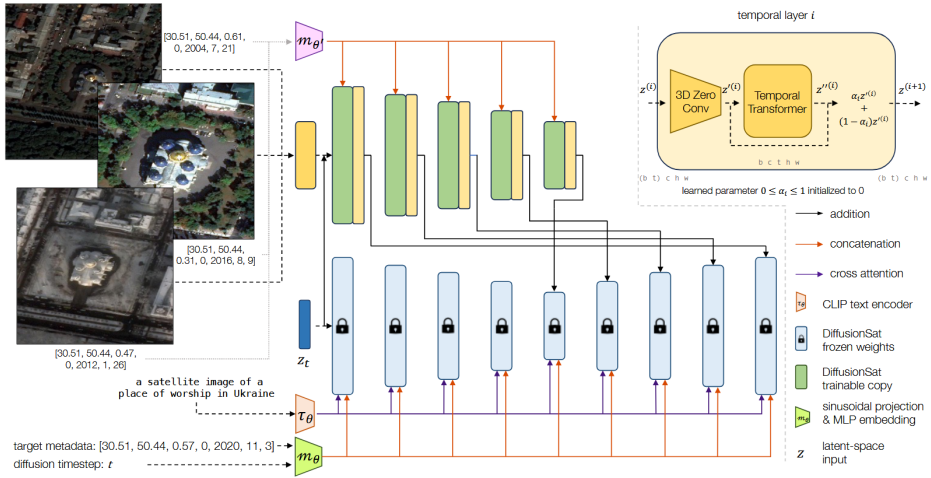

图: DiffusionSat：面向卫星影像的生成式基础模型 (Khanna et al., 2024).

Khanna, et al. DiffusionSat: A Generative Foundation Model for Satellite Imagery, ICLR, 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).

Khanna, et al. DiffusionSat: A Generative Foundation Model for Satellite Imagery, ICLR, 2024.

DiffusionSat：图像修复结果

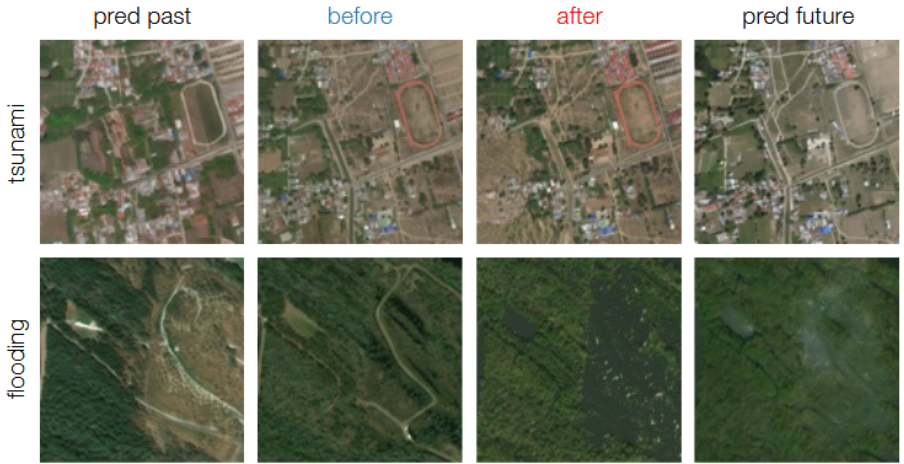

图: DiffusionSat 用于遥感图像修复的示例结果 (Khanna et al., 2024).

Khanna, et al. DiffusionSat: A Generative Foundation Model for Satellite Imagery, ICLR, 2024.

问答环节

项目作业：研究主题

核心问题：

生成模型合成数据是否已准备好用于图像识别？

- 生成模型合成数据是当前流行的数据增强新方式 (He et al., 2023; Toker et al., 2024).
- 本项目将探索此类增强是否有助于遥感领域的下游任务，如文本-图像检索、图像场景分类和超分辨率等。

He, et al. Is synthetic data from generative models ready for image recognition?, ICLR, 2023.
Toker, et al. SatSynth: Augmenting Image-Mask Pairs through Diffusion Models for Aerial Semantic Segmentation, CVPR, 2024.

背景：为什么要用合成数据?

- 手工数据采集与标注成本高、耗时长。
- 生成模型合成数据可实现大规模数据增强。
- 合成数据对遥感下游任务的有效性仍在积极研究中。

零样本设置下不同合成数据策略的可视化 (He et al., 2023). 展示了真实数据与不同策略 (基础 B、语言增强 LE、语言增强+CLIP 过滤 LE+CF) 合成图像的可视化对比。

He, et al. Is synthetic data from generative models ready for image recognition?, ICLR, 2023.

实验设置：下游任务的生成式数据

- 我们利用最先进的生成模型（如 **DiffusionSat** (Khanna et al., 2024) 和 **Text2Earth** (Liu et al., 2025)）进行合成数据增强。
- **两种主要策略：**
  1. **文本到图像生成：**
     - 生成文本-图像对。
     - 支持图像场景分类、图文检索等任务。
  2. **超分辨率生成：**
     - 生成低分辨率（LR）与高分辨率（HR）配对图像。
     - 支持遥感超分辨率任务。

Khanna, et al. DiffusionSat: A Generative Foundation Model for Satellite Imagery, ICLR, 2024.
Liu, et al. Text2Earth: Unlocking text-driven remote sensing image generation with a global-scale dataset and a foundation model. GRSM. 2025.

CLIP 场景分类

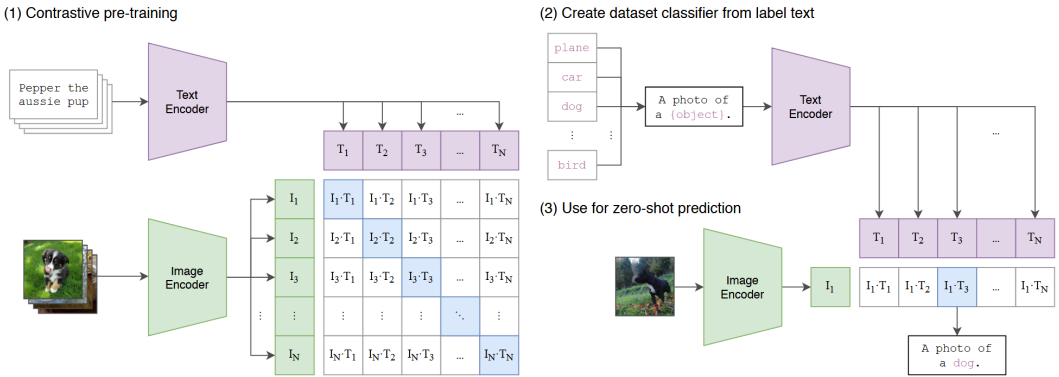

图: CLIP (Radford et al., 2021) 模型概览：图像与文本编码器联合训练，实现跨模态理解。

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

Real-ESRGAN 超分辨率

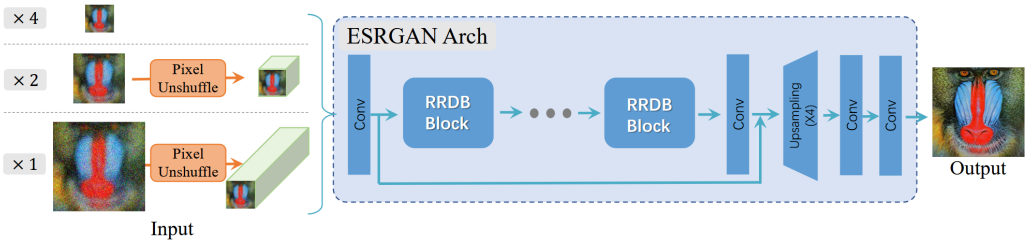

图: Real-ESRGAN (Wang et al., 2021) 框架：面向真实世界图像超分辨率的架构。

Wang, et al. Real-ESRGAN: Training Real-World Blind Super-Resolution With Pure Synthetic Data, ICCV, 2021.

场景分类与超分辨率基线模型

场景图像分类：

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

超分辨率：

- **Real-ESRGAN** (Wang, Xie, et al., 2021)
- **StableSR** (Wang, Yue, et al., 2024)
- **FaithDiff** (Chen, Pan, and Dong, 2025)
- **EResShift** (Yue, Wang, and Loy, 2025)

Radford, et al. Learning Transferable Visual Models From Natural Language Supervision, ICML, 2021.
Liu, Chen, Guan, et al. RemoteCLIP: A Vision Language Foundation Model for Remote Sensing. TGRS. 2024.
Liu, Chen, Zhao, et al. Text2Earth: Unlocking text-driven remote sensing image generation with a global-scale dataset and a foundation model. GRSM. 2025.
Wang, Xie, et al. Real-ESRGAN: Training Real-World Blind Super-Resolution With Pure Synthetic Data, ICCV, 2021.
Wang, Yue, et al. Exploiting Diffusion Prior for Real-World Image Super-Resolution. IJCV. 2024.
Chen, et al. FaithDiff: Unleashing Diffusion Priors for Faithful Image Super-resolution, CVPR, 2025.
Yue, et al. Efficient Diffusion Model for Image Restoration by Residual Shifting. TPAMI. 2025.

下游任务数据集

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 个人工标注的图文对。
- **UCM-Captions** (Qu et al., 2016)：源自 UC Merced 土地利用数据集，含 2,100 张图像，每张配有 5 条描述。

超分辨率：

- **fMoW**：Sentinel-2（10m GSD）(Cong et al., 2022) 与 fMoW-RGB (0.3m) (Christie et al., 2018) 配对数据集。

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.
Cong, et al. Functional Map of the World - Sentinel-2 corresponding images. 2022.
Christie, et al. Functional Map of the World, CVPR, 2018.

可能的结果

- (He et al., 2023) 的结果表明，合成数据增强在遥感基准上带来了显著的准确率提升。

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.

图: 使用合成数据增强后分类性能提升 (He et al., 2023).

He, et al. Is synthetic data from generative models ready for image recognition?, ICLR, 2023.

基于扩散的合成数据增强更优

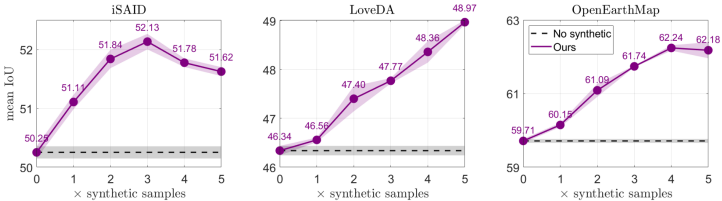

图: SatSynth 表格：基于扩散的合成数据增强在分割精度上优于传统增强方法 (Toker et al., 2024).

表: 定量对比：iSAID (Waqas Zamir et al., 2019) 数据集上的平均 IoU。

无增强	本方法	Cutout (DeVries and Taylor, 2017)	CutMix (Yun et al., 2019)	Copy-Paste (Ghiasi et al., 2021)
50.25	**51.11**	50.47	50.60	50.51

Toker, et al. SatSynth: Augmenting Image-Mask Pairs through Diffusion Models for Aerial Semantic Segmentation, CVPR, 2024.
Waqas Zamir, et al. isaid: A large-scale dataset for instance segmentation in aerial images, CVPRW, 2019.
DeVries, et al. Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552. 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.