

The Landscape of Deep Generative Learning

Diffusion Models for Image Generation in Remote Sensing

GISLab Short-Term Course 2025 Summer

Zhenyuan Chen

School of Earth Science, Zhejiang University

June 25, 2025

bili_sakura@zju.edu.cn

Bayesian Networks

Restricted Boltzmann Machines

Variational Autoencoders

Normalizing Flows

Energy-based Models

Generative Adversarial Networks

Autoregressive Models

Denoising Diffusion Models

Image Source: Vahdat Arash, Song, and Meng, 2023.

Outline

- ▶ Background: Generative Models for Image Synthesis
- ▶ Diffusion Models: Theory
- ▶ Applications in Remote Sensing
- ▶ Summary & Q/A

Background: Generative Models for Image Synthesis

Background: Generative Models for Image Synthesis

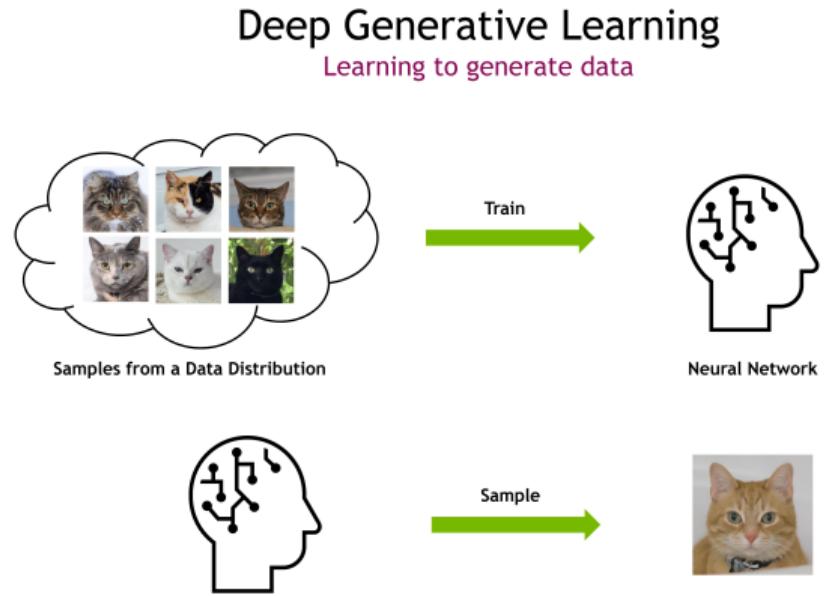


Figure: Illustration of generative modeling (Vahdat Arash, Song, and Meng, 2023).

2

Timeline of Generative Models

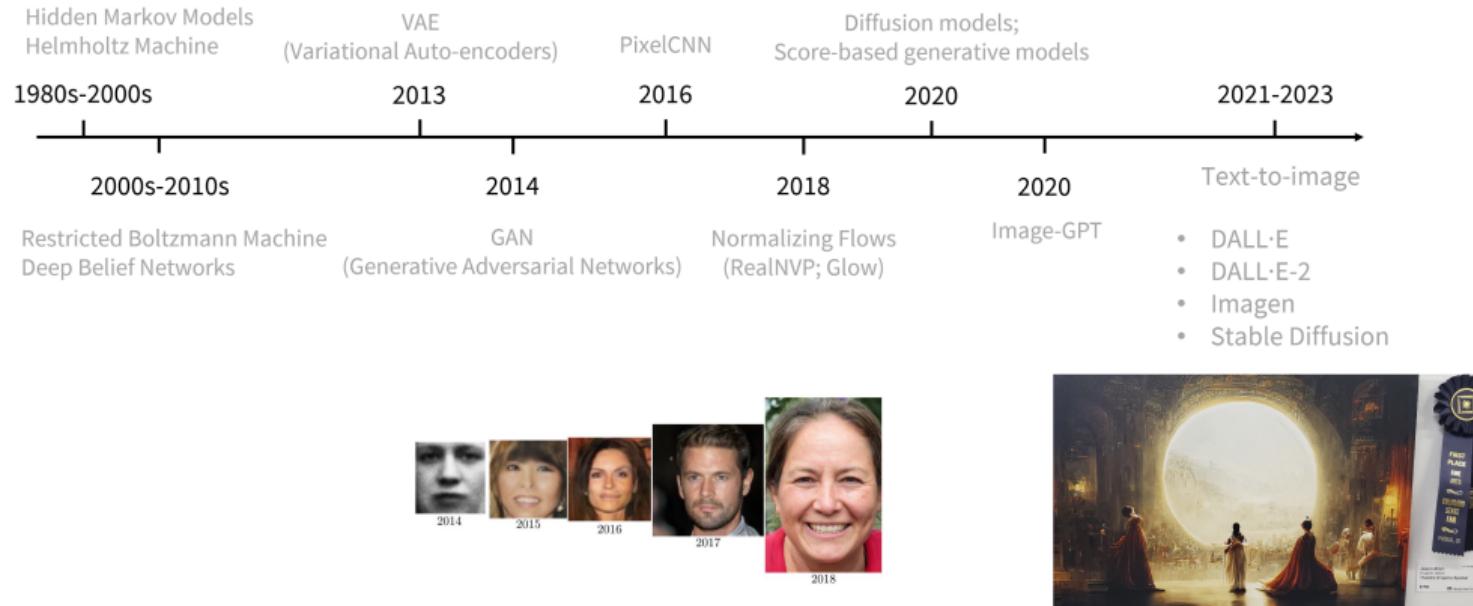


Figure: Timeline of key developments in generative models (Deng, 2024).

Diffusion Models: Theory

Denoising Diffusion Models

Denoising diffusion models consist of two processes:

- ▶ A forward diffusion process that gradually adds noise to the input.
- ▶ A reverse denoising process that learns to generate data by denoising.

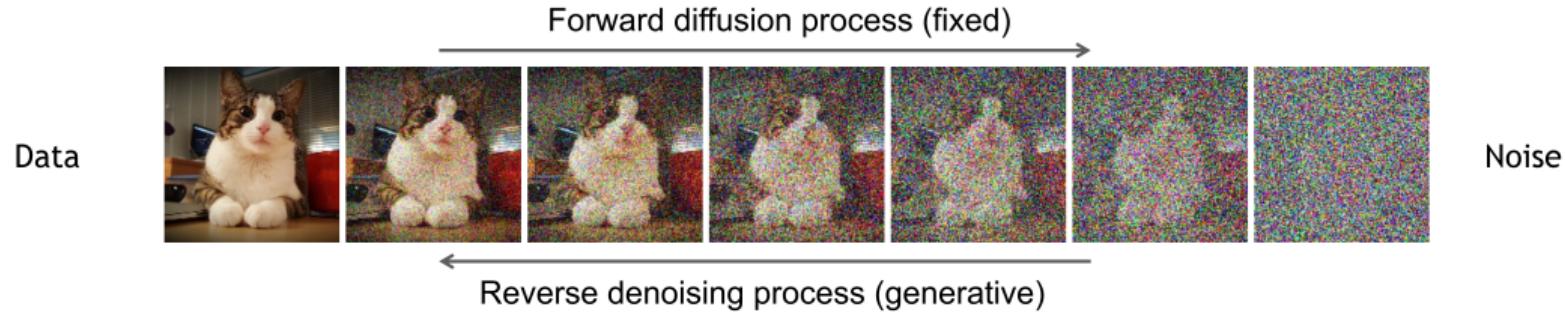


Figure: Diffusion models generate data through iterative denoising (Sohl-Dickstein et al., 2015; Ho, Jain, and Abbeel, 2020). Image credit: Vahdat Arash, Song, and Meng, 2023.

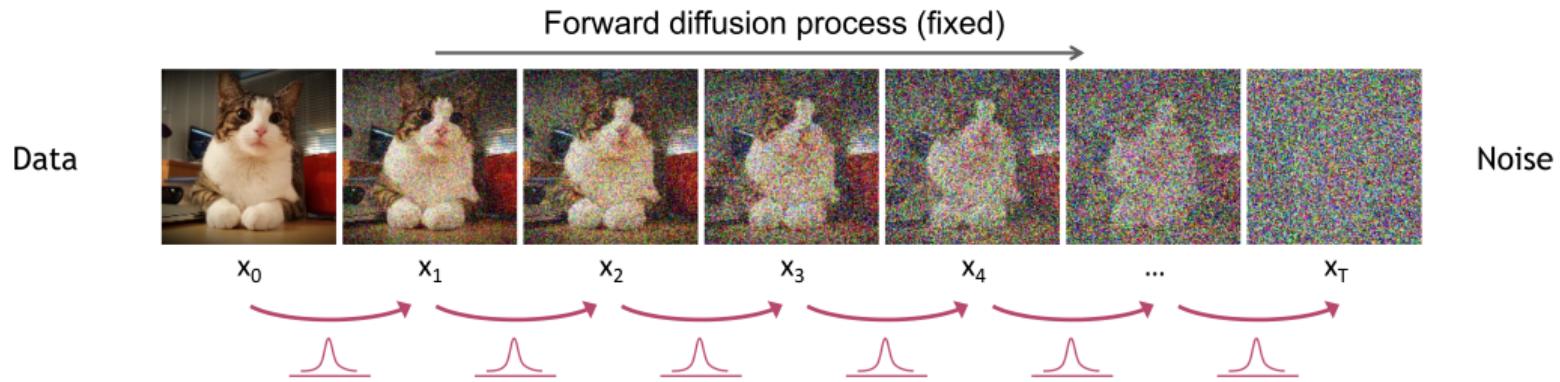
Sohl-Dickstein, et al. Deep Unsupervised Learning using Nonequilibrium Thermodynamics, ICML, 2015.

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.

Forward Diffusion Process

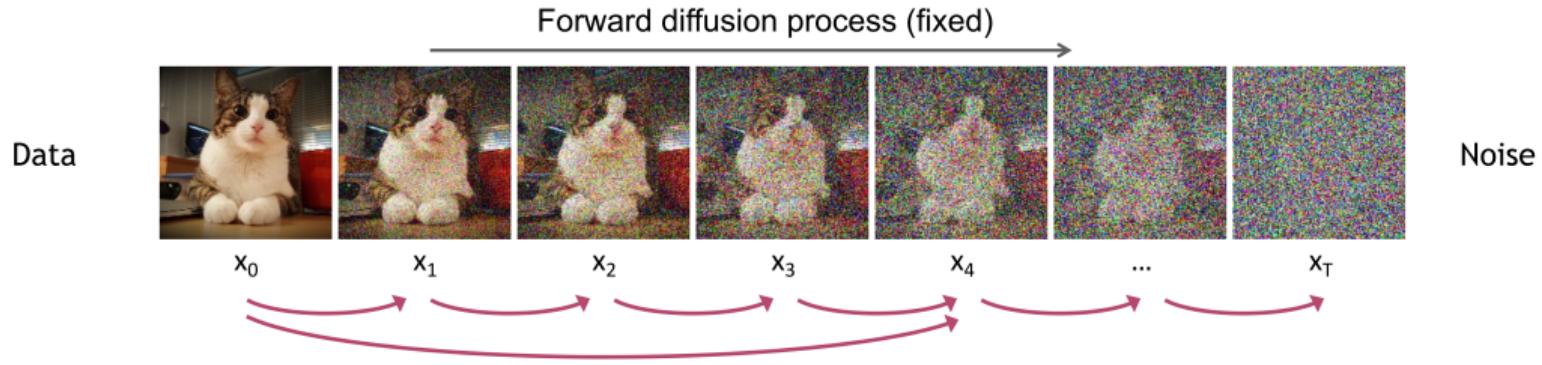
The formal definition of the forward process in T steps:



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

Image credit: Vahdat Arash, Song, and Meng, 2023.

Diffusion Kernel



Define: $\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s) \implies q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t} \mathbf{x}_0, (1 - \bar{\alpha}_t) \mathbf{I})$ (Diffusion Kernel)

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

The noise schedule $\{\beta_t\}$ is chosen so that $\bar{\alpha}_T \rightarrow 0$ and $q(\mathbf{x}_T | \mathbf{x}_0) \approx \mathcal{N}(\mathbf{x}_T; \mathbf{0}, \mathbf{I})$.

Image credit: Vahdat Arash, Song, and Meng, 2023.

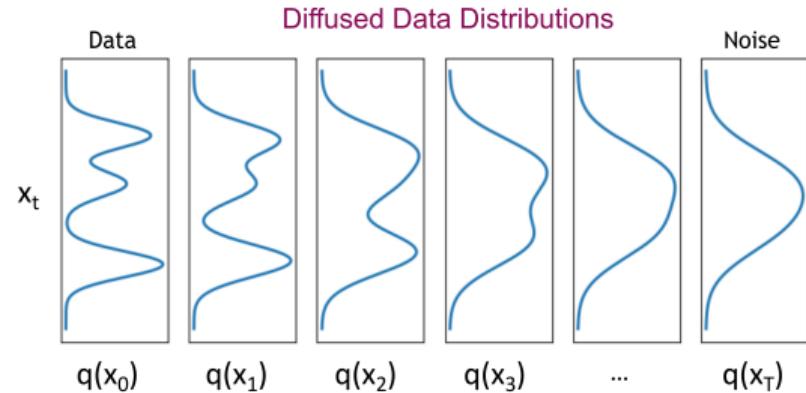
Vahdat Arash et al. CVPR 2023 Tutorial Denoising Diffusion-based Generative Modeling: Foundations and Applications, 2023.

What happens to a distribution in the forward diffusion?

So far, we discussed the diffusion kernel $q(\mathbf{x}_t|\mathbf{x}_0)$ but what about $q(\mathbf{x}_t)$?

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

The diffusion kernel is Gaussian convolution.



We can sample $\mathbf{x}_t \sim q(\mathbf{x}_t)$ by first sampling $\mathbf{x}_0 \sim q(\mathbf{x}_0)$ and then sampling $\mathbf{x}_t \sim q(\mathbf{x}_t|\mathbf{x}_0)$ (i.e., ancestral sampling).

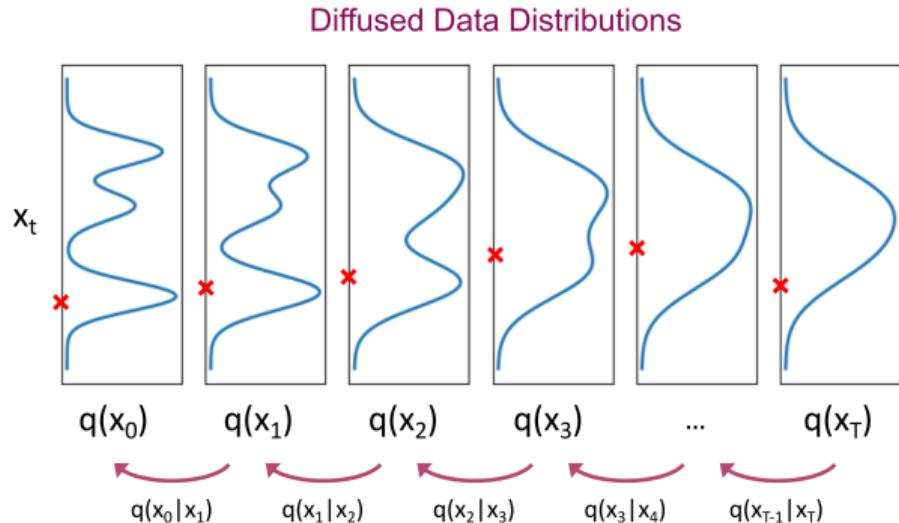
Image credit: Vahdat Arash, Song, and Meng, 2023.

Generative Learning by Denoising

Recall that the diffusion parameters are designed such that $q(\mathbf{x}_T) \approx \mathcal{N}(\mathbf{x}_T; 0, \mathbf{I})$.

Generation:

- ▶ Sample $\mathbf{x}_T \sim \mathcal{N}(\mathbf{x}_T; 0, \mathbf{I})$
- ▶ Iteratively sample
 $\mathbf{x}_{t-1} \sim \underbrace{q(\mathbf{x}_{t-1} | \mathbf{x}_t)}_{\text{True Denoising Dist.}}$

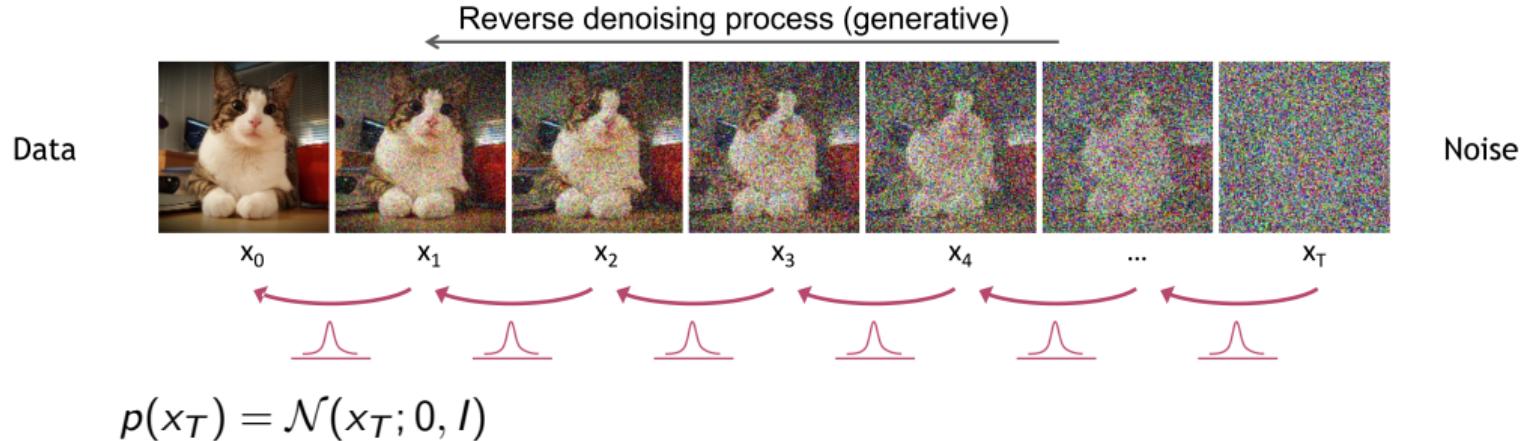


Can we approximate $q(\mathbf{x}_{t-1} | \mathbf{x}_t)$? Yes, we can use a **normal distribution** if β_t is small in each forward diffusion step.

Image credit: Vahdat Arash, Song, and Meng, 2023.

Reverse Denoising Process

Formal definition of forward and reverse processes in T steps:



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

where $\mu_\theta(x_t, t)$ is a trainable network (e.g., U-Net, Denoising Autoencoder)
Image credit: Vahdat Arash, Song, and Meng, 2023.

Learning Denoising Model

Variational upper bound

- ▶ For training, use a variational upper bound (as in VAEs):

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

- ▶ $\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}$, mean parameterized as (Ho, Jain, and Abbeel, 2020):

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

- ▶ Variational objective:

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

- ▶ Set $\lambda_t = 1$ for all t works best (Ho, Jain, and Abbeel, 2020).

Summary

Training and Sample Generation

Algorithm1-Training

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

Algorithm2-Sampling

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

Algorithms are derived from (Ho, Jain, and Abbeel, 2020)

Applications in Remote Sensing

DiffusionSat: Framework Overview

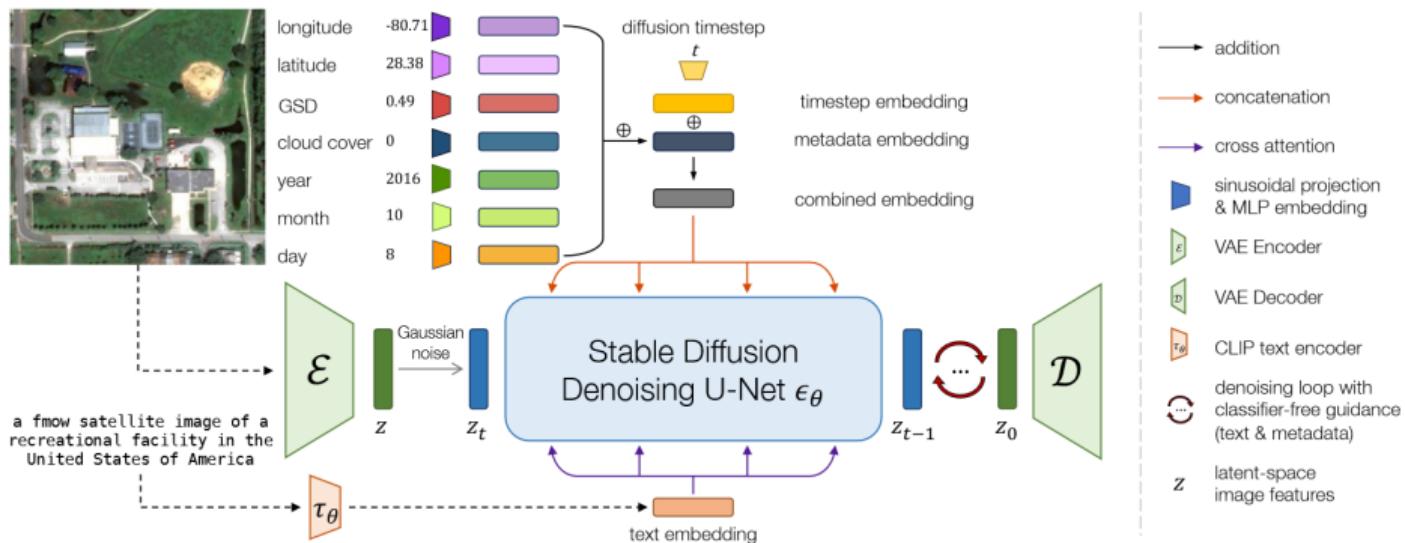


Figure: Conditioning on freely available metadata and using large, publicly available satellite imagery datasets shows DiffusionSat is a powerful generative foundation model for remote sensing data. (Khanna et al., 2024).

DiffusionSat: Single-Image Generation

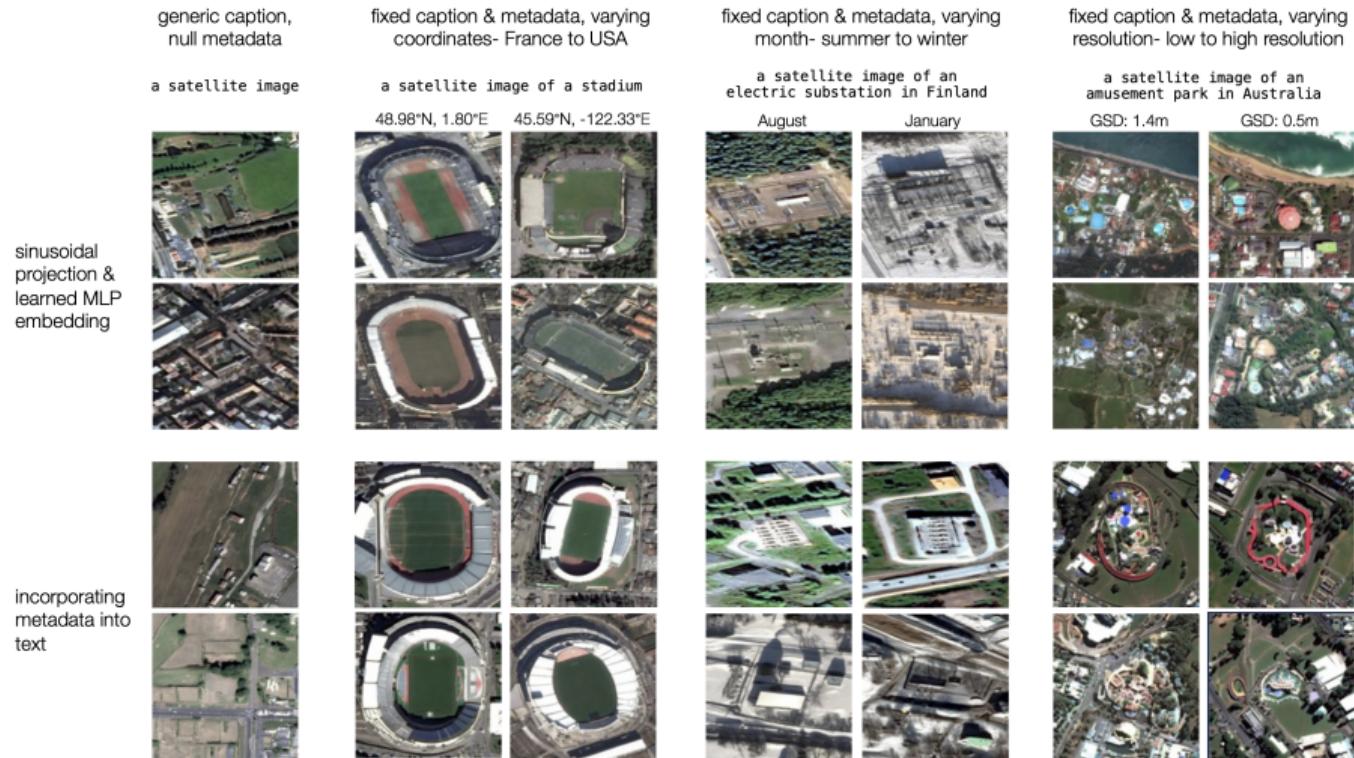


Figure: Here we generate samples from single-image DiffusionSat (Khanna et al., 2024).

DiffusionSat: Framework Overview

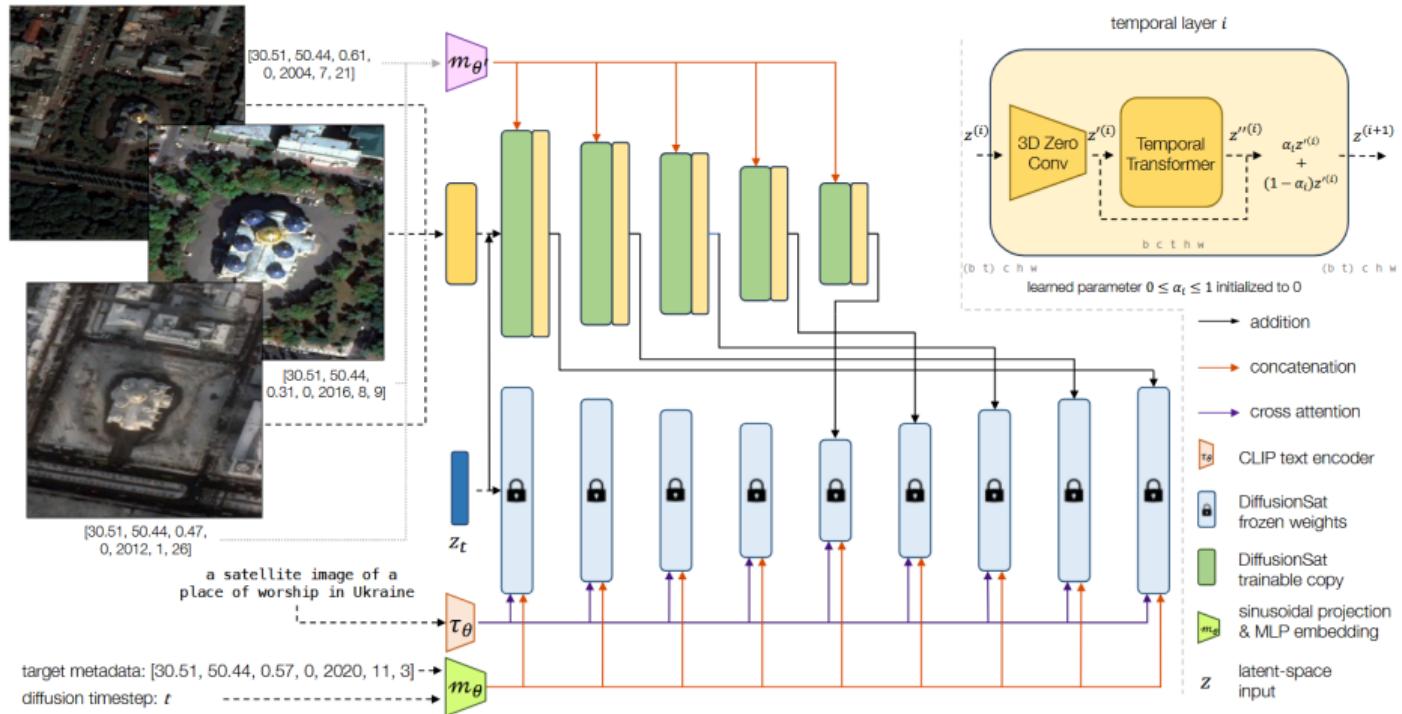


Figure: 3DControlNet in DiffusionSat (Khanna et al., 2024).

DiffusionSat: Temporal Prediction Results



Figure: Generated samples from the fMoW-temporal, for temporal prediction (Khanna et al., 2024).

| Model | $t' > t$ | | | $t' < t$ | | |
|-----------------------------|---------------|----------------|---------------|---------------|----------------|---------------|
| | 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 |

Table: Table 4: Sample quality quantitative results on fMoW-temporal validation data. $t' > t$ represents generating an image in the past given a future image, and $t' < t$ is the task for generating a future image given a past image.

DiffusionSat: Super-Resolution Results

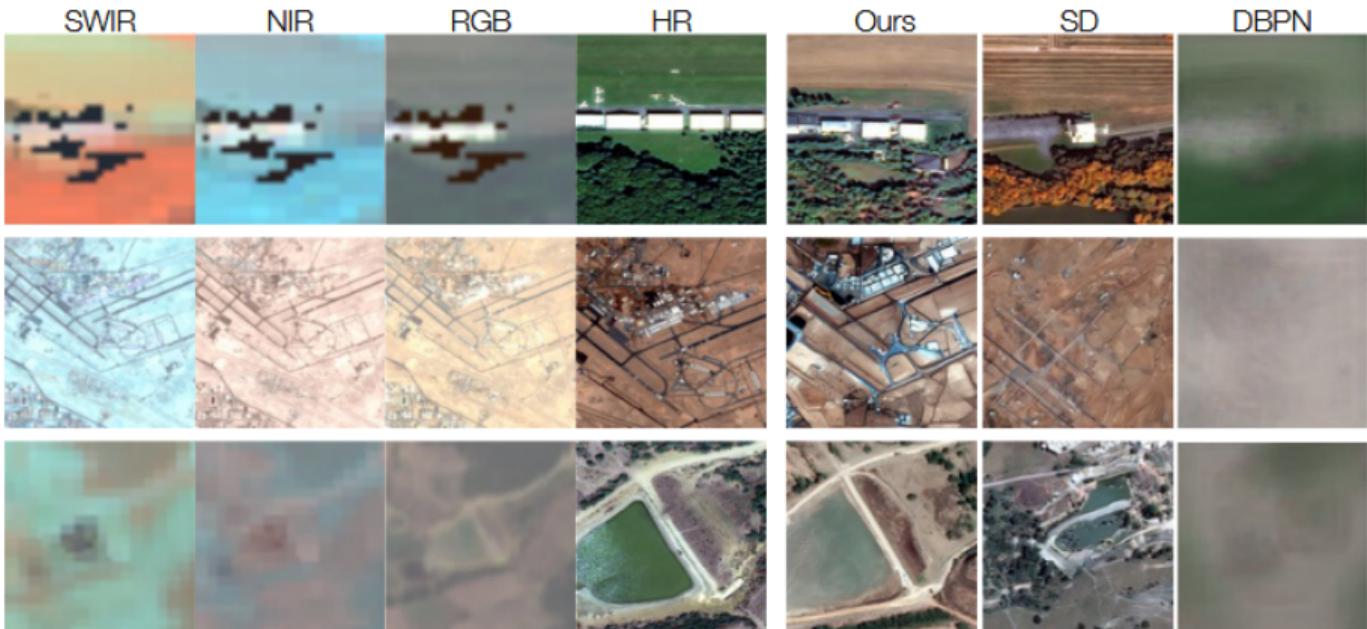
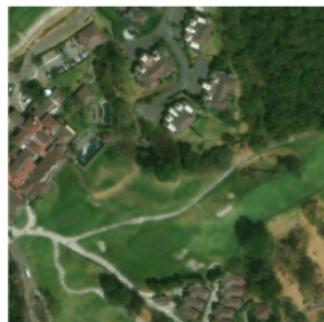


Figure: Example results: DiffusionSat for multi-spectral super-resolution (Khanna et al., 2024).

Further Discussion

Using Gen AI to Make New Images

- ▶ Generative models can create new, realistic images.
- ▶ We can use them to make more training data.
- ▶ Example: Give a “before” image and a description, get a new “after” image.



+ suffer from
volcano eruption

C_T

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

Generative Models



Supplementary

How Do We Augment Data?

Classic Methods:

- ▶ Flip, rotate, crop, change colors, etc.

Modern Methods:

- ▶ Mix two images together (Mixup) (Zhang et al., 2018).
- ▶ Cut and paste parts of images (CutMix) (Yun et al., 2019).



Figure: Illustration of modern augmentation methods. From Left to Right: Mixup (Zhang et al., 2018), Cutout (DeVries and Taylor, 2017), and CutMix (Yun et al., 2019).

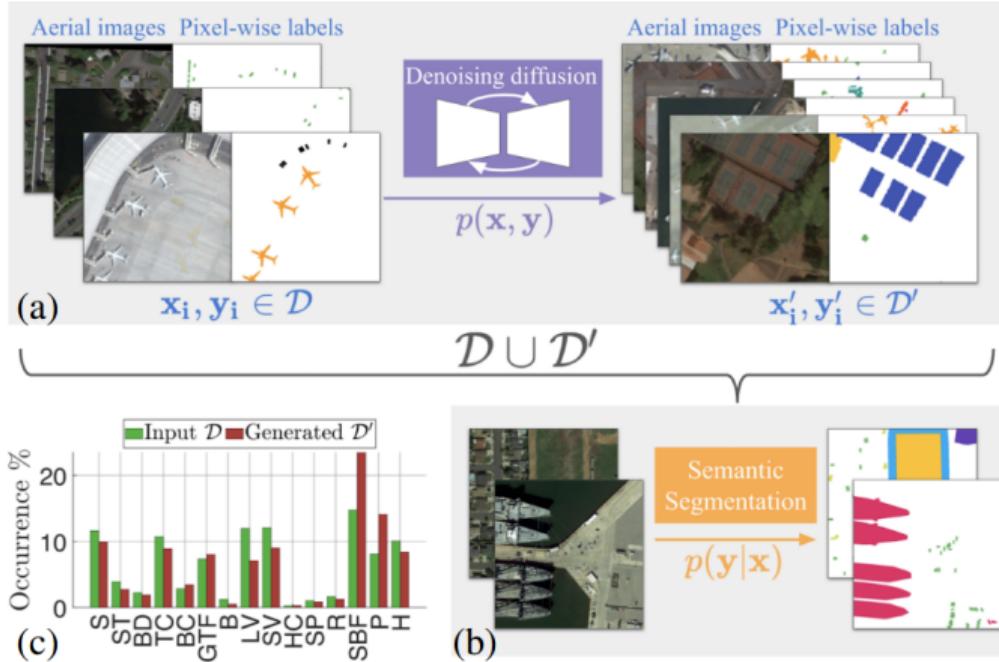
Soft Label Example (CutMix):

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

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

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

Generative Models for Data Augmentation



SatSyn (Toker et al., 2024) proposes a generative model (diffusion model) to generate both images and corresponding masks for satellite segmentation. The synthetic dataset is used for data augmentation, yielding significant quantitative improvements in satellite semantic segmentation compared to other data augmentation methods.

Application in Remote Sensing Image Generation: Text2Earth

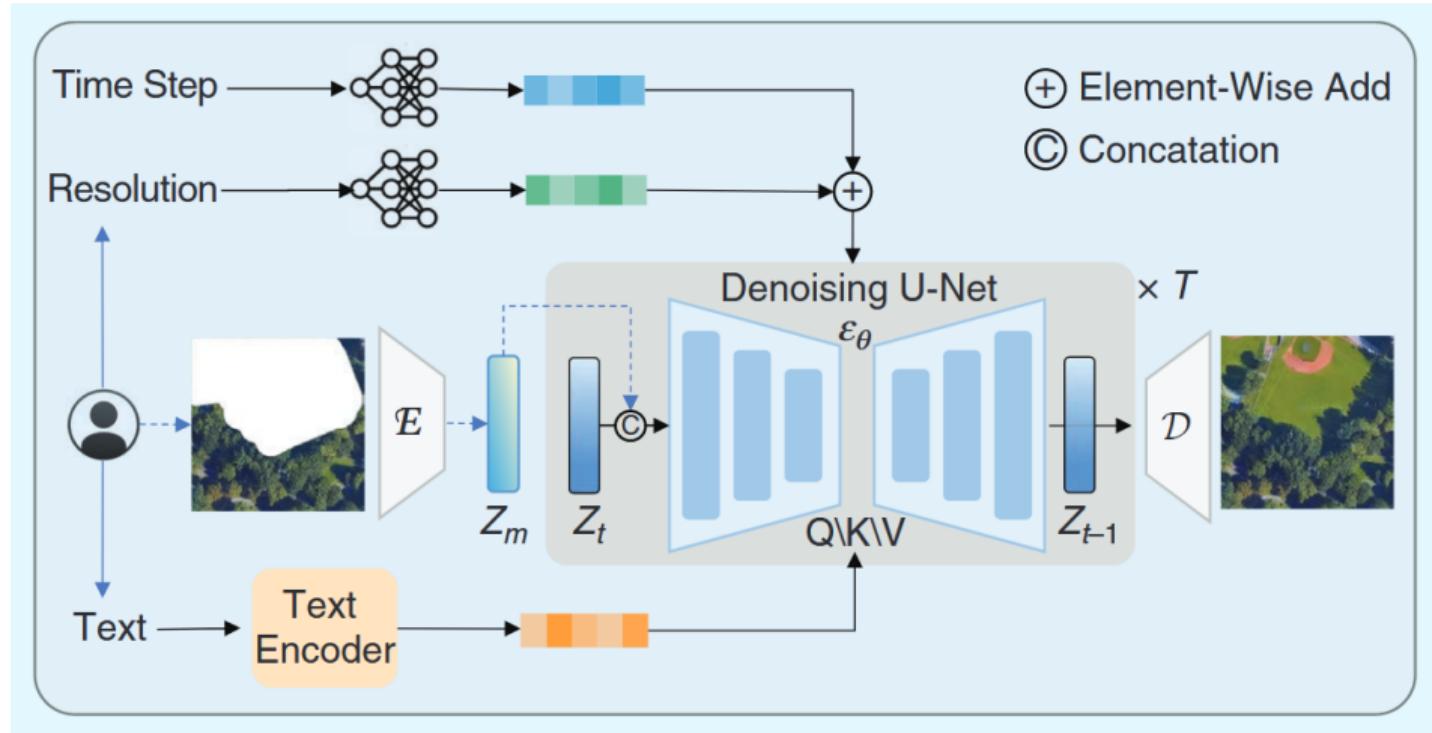


Figure: Text2Earth: Foundation model for text-driven Earth observation (Liu et al., 2025).

Text2Earth: Example Results

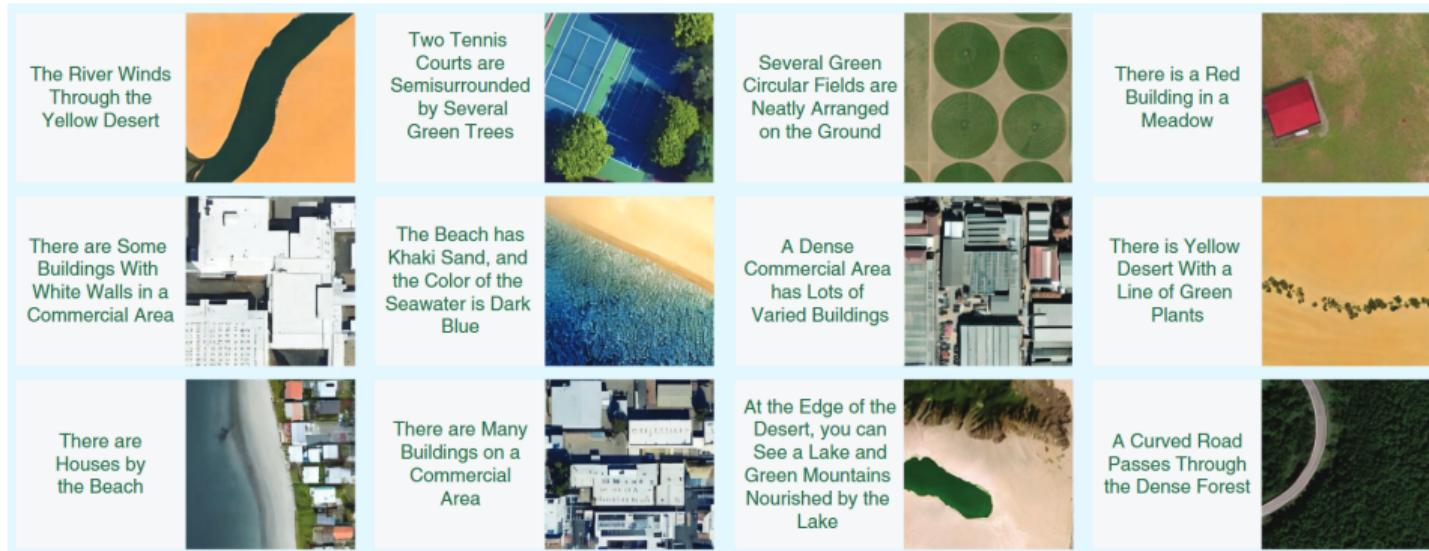


Figure: Example results generated by Text2Earth (Liu et al., 2025).

Application in Remote Sensing Image Generation: CRS-Diff

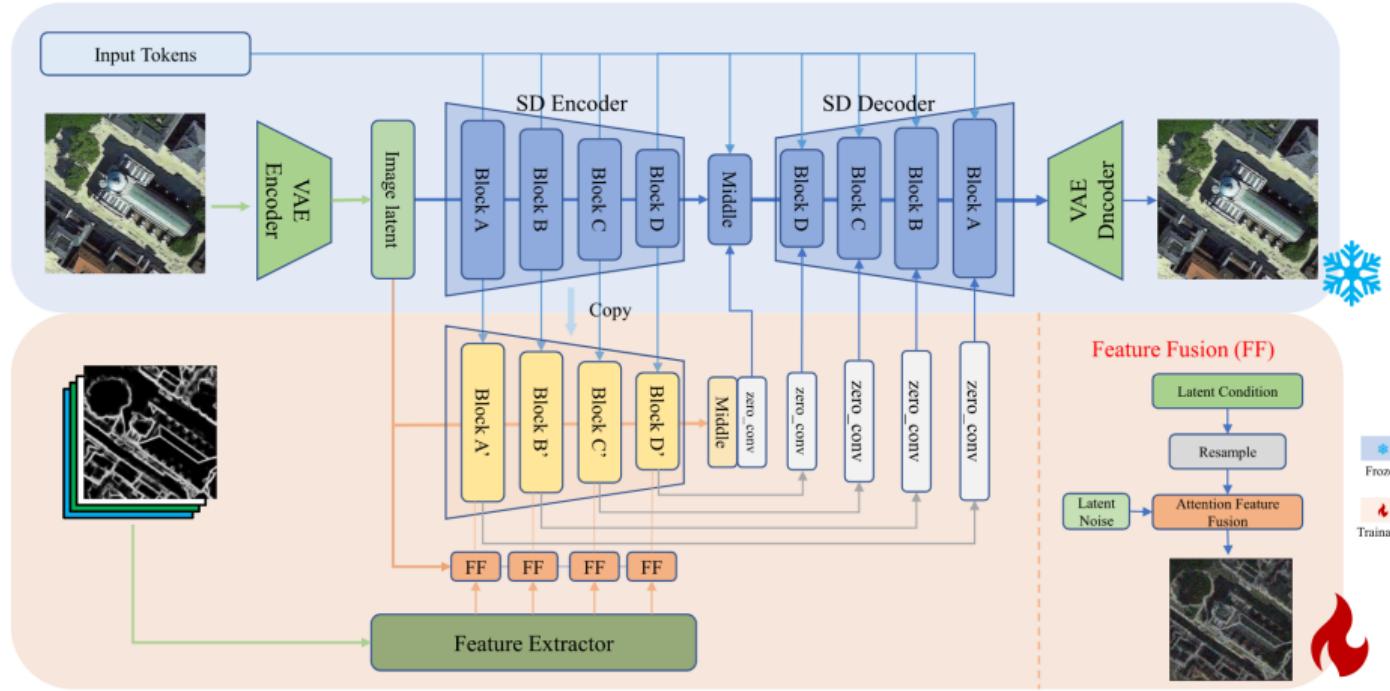


Figure: CRS-Diff: Controllable remote sensing image generation framework (Tang, Li, et al., 2024).

CRS-Diff: Example Results

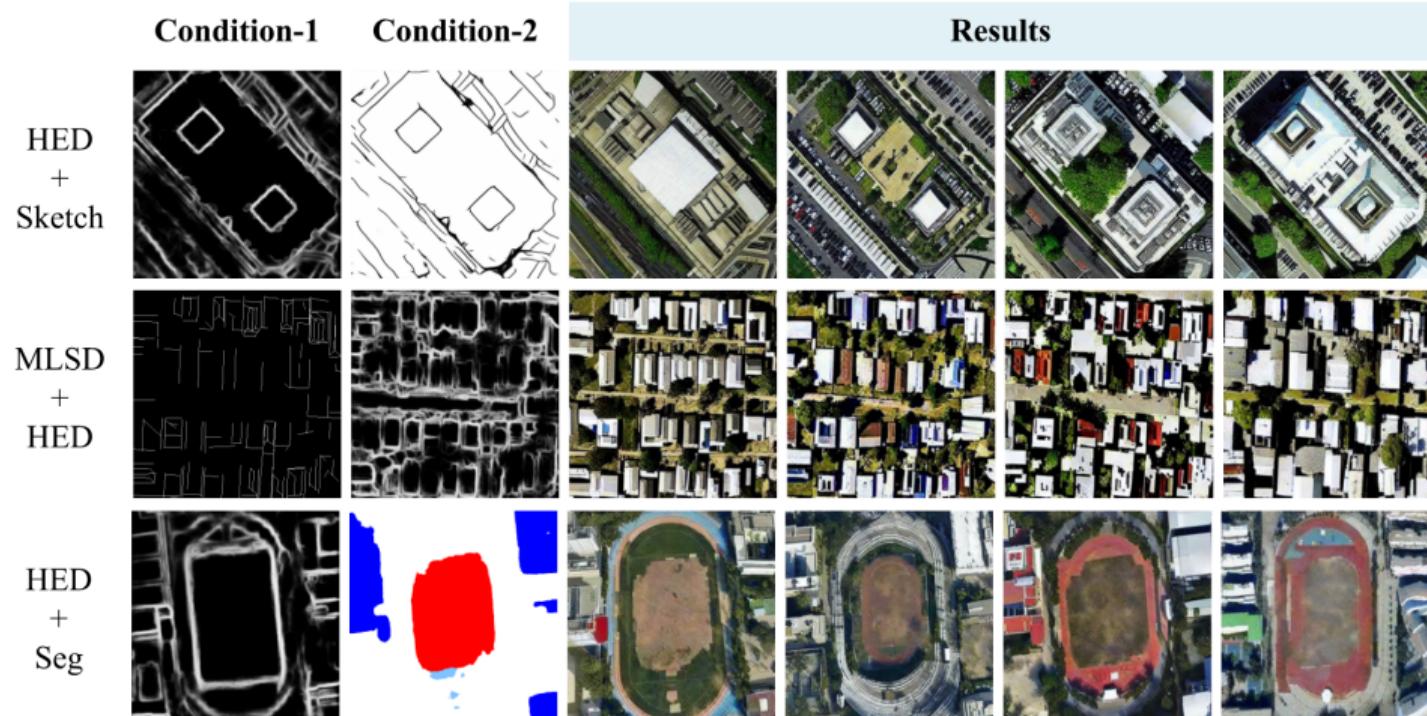


Figure: Example results generated by CRS-Diff (Tang, Li, et al., 2024).