Can AI-Generated Images Help Image Recognition?

GISLab Short-Term Course 2025 Summer

Zhenyuan Chen
School of Earth Science, Zhejiang University

2025

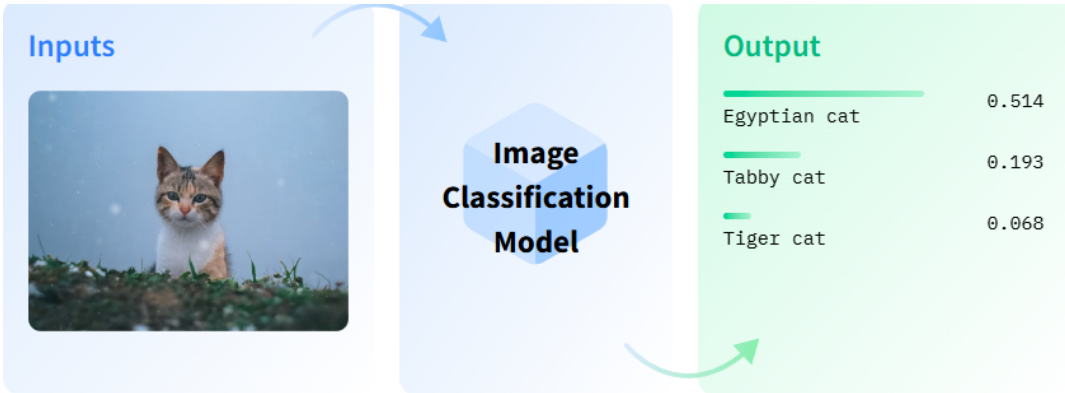
bili_sakura@zju.edu.cn

Outline

- What is Image Classification?
- How Can We Improve It? (Data Augmentation)
- Real-World Example: Disaster Images
- Project: Can AI-Generated Images Help?

What is Image Classification?

- Computers learn to recognize what's in a picture (e.g., cat, dog, airplane).
- We use lots of labeled images to teach the computer.
- Goal: Predict the correct label for new images.



End-to-End Neural Network

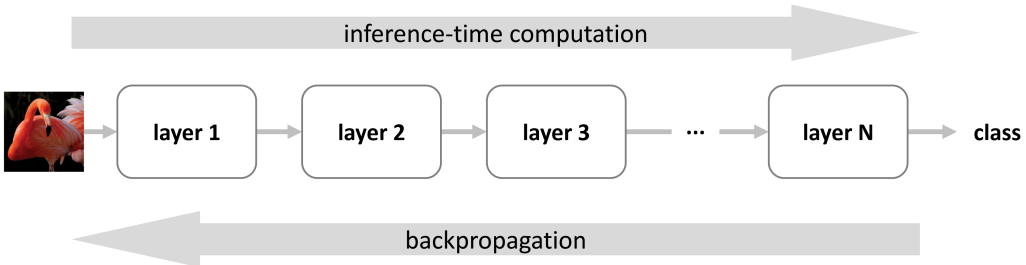


Figure: An end-to-end neural network for image classification. The image passes through multiple layers, and the network learns by backpropagation. Image Source: He, 2025.

He, Towards End-to-End Generative Modeling, Proceedings of the CVPR 2025 Workshop on Visual Generative Modeling: What's After Diffusion?, 2025.

How Does It Work?

Neural networks are special computer programs that learn patterns from many example images. Over time, they get better at telling images apart.

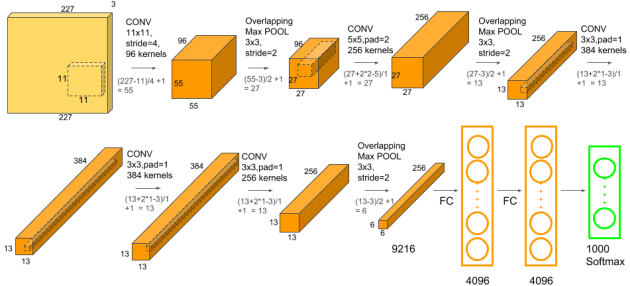


Figure: Example of a neural network for image classification: AlexNet (Krizhevsky, Sutskever, and Hinton, 2012).

Krizhevsky, et al. ImageNet Classification with Deep Convolutional Neural Networks, NeurIPS, 2012.

Modern Models: CLIP

- CLIP learns to match images and text.
- It can recognize new things by understanding descriptions.
- Useful for many tasks, not just classification.

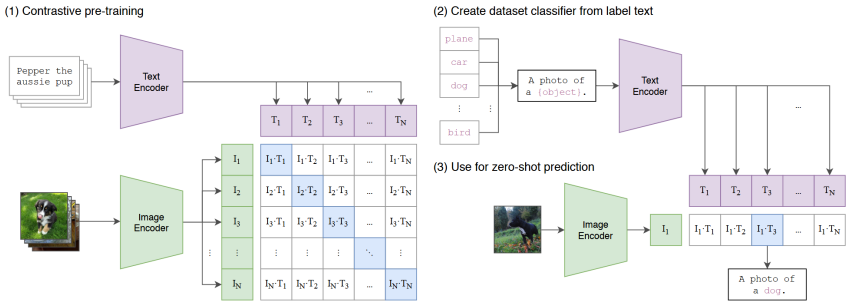


Figure: Summary of OpenAI CLIP ViT (Radford et al., 2021).

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

How Can We Improve Image Classification? (Data Augmentation)

Why Use Data Augmentation?

- Sometimes we do not have enough images to train a good model.
- Data augmentation means making new images from existing ones.
- It helps the model learn better and avoid mistakes.

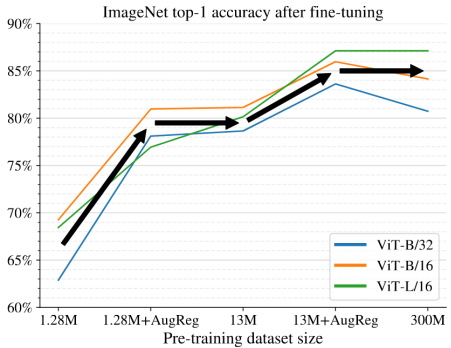


Figure. Adding the right amount of regularization and image augmentation can lead to similar gains as increasing the dataset size by an order of magnitude. (Steiner et al., 2022)

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

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).

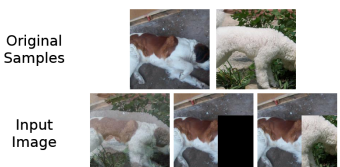


Figure: Illustration of modern augmentation methods. From Left to Right: Mixup (Zhang et al., 2018), Cutout (DeVries and Taylor, 2017), and CutMix (Yun et al., 2019).

Soft Label Example (CutMix):

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

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

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

Zhang, et al. Mixup: Beyond Empirical risk minimazation, 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.

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.

X_{pre}

+ suffer from volcano eruption

C_T

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

Generative Models

X_{post}

Real-World Example: Disaster Images (xBD)

xBD: A Large-Scale Disaster Damage Dataset

xBD (Gupta et al., 2019) is a bi-temporal remote sensing dataset covering 19 distinct disaster events. Pre-disaster imagery (top) and post-disaster imagery (bottom). From left to right: Hurricane Harvey; Joplin tornado; Lower Puna volcanic eruption; Sunda Strait tsunami.

Table. 19 Disaster events in xBD.

Disaster Type	Disaster Event	Event Date
Earthquake	Mexico City earthquake	Sep 19, 2017
Wildfire	Portugal wildfires	Jun 17–24, 2017
Wildfire	Santa Rosa wildfires	Oct 8–31, 2017
Wildfire	Carr wildfire	Jul 23–Aug 30, 2018
Wildfire	Woolsey fire	Nov 9–28, 2018
Wildfire	Piner fire	Nov 25–Dec 2, 2018
Volcano	Lower Puna volcanic eruption	May 23–Aug 14, 2018
Volcano	Guatemala Fuego volcanic eruption	Jun 3, 2018
Storm	Tuscaloosa, AL tornado	Apr 27, 2011
Storm	Joplin, MO tornado	May 22, 2011
Storm	Moore, OK tornado	May 20, 2013
Storm	Hurricane Matthew	Sep 28–Oct 10, 2016
Storm	Hurricane Florence	Sep 10–19, 2018
Flooding	Monsoon in Nepal, India, and Bangladesh	Aug 2017
Flooding	Hurricane Harvey	Aug 17–Sep 7, 2017
Flooding	Hurricane Michael	Oct 7–16, 2018
Flooding	Midwest US floods	Jan 3–May 31, 2019
Tsunami	Indonesia tsunami	Sep 18, 2018
Tsunami	Sunda Strait tsunami	Dec 22, 2018

Gupta, et al. xBD: A Dataset for Assessing Building Damage from Satellite Imagery, arXiv, CVPRW, 2019.

Project: Can AI-Generated Images Help?

Project: What Will You Do?

- Test if adding AI-generated images helps the model learn.
- Try three ways:
  - Only real images
  - Only generated images
  - Both real and generated images
- See which way gives the best results!

Project: Dataset Details & How to Generate Images

- Use 100 real images per disaster type (6 types, 600 images total).
- For each type, generate 100–400 new images using AI.
- Try different mixes: 1:1, 1:2, 1:3, 1:4 (real:generated).
- Use commercial AI models (e.g., GPT-Image-1 (OpenAI, 2025), Gemini 2.5 Pro (Gemini Team, Google, 2025), SeedEdit 3.0 (Wang et al., 2025)) to generate images.
- Input: a "before" image and a short description (e.g., "make it suffer from flooding").
- Output: a new, realistic "after" image.

OpenAI. GPT-Image-1, 2025.
Gemini Team, Google (2025). *Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities.*
Wang, et al. SeedEdit 3.0: Fast and High-Quality Generative Image Editing. arXiv. 2025.

Which Models to Use & How to Measure Success

- **Try these models:**
  - OpenAI CLIP (Radford et al., 2021)
  - Google SigLip/SigLip2 (Tschannen et al., 2025; Zhai et al., 2023)
  - Both of the above use ViT configurations (see CLIP training example and ViT tutorials).
- **How to measure success:**
  - **Accuracy**: % of correct answers
  - **F1 Score**: balances precision and recall
  - **Confusion Matrix**: shows what gets confused
- Always test on real images the model hasn't seen before.
- Make simple bar or line charts to compare results.

Radford, et al. Learning Transferable Visual Models From Natural Language Supervision, ICML, 2021.
Tschannen, et al. (2025). *SigLIP 2: Multilingual Vision-Language Encoders with Improved Semantic Understanding, Localization, and Dense Features.*
Zhai, et al. Sigmoid Loss for Language Image Pre-Training, 2023.

Appendix

Background: Image Classification with Deep Learning

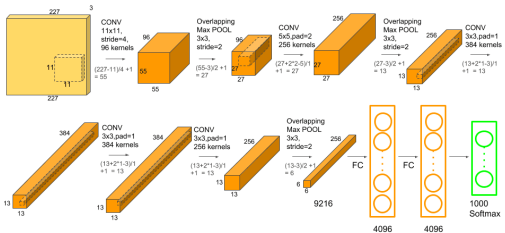
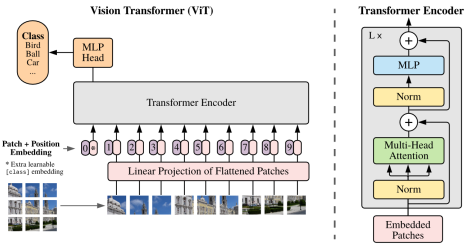


Figure: Left: AlexNet on ILSVRC-2010 (Berg, Deng, and Fei-Fei, 2010) Right: Architecture of AlexNet (Krizhevsky, Sutskever, and Hinton, 2012).

Berg, et al. Large Scale Visual Recognition Challenge 2010, 2010.
Krizhevsky, et al. ImageNet Classification with Deep Convolutional Neural Networks, NeurIPS, 2012.

Architecture Evolution of Image Classification

- **NeurIPS 2012: AlexNet (CNN)**
  (Krizhevsky, Sutskever, and Hinton, 2012)
- **CVPR 2016: ResNet (CNN)**
  (He, Zhang, et al., 2016)
- **ICLR 2021: Vision Transformers (ViT)**
  (Dosovitskiy et al., 2021)
- **ICCV 2021: Swin Transformer (ViT)**
  (Liu et al., 2021)
- **ICML 2021: CLIP (ViT)**
  (Radford et al., 2021)
- **CVPR 2022: MAE (ViT)**
  (He, Chen, et al., 2022)
- **TMLR 2022: CoCa (ViT)**
  (Yu et al., 2022)



Overview of Vision Transformer (Dosovitskiy et al., 2021).

Krizhevsky, et al. ImageNet Classification with Deep Convolutional Neural Networks, NeurIPS, 2012.
He, Zhang, et al. Deep Residual Learning for Image Recognition, CVPR, 2016.
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, Chen, et al. Masked Autoencoders Are Scalable Vision Learners, CVPR, 2022.
Yu, et al. CoCa: Contrastive Captioners Are Image-Text Foundation Models. TMLR. 2022.

Image Classification Dataset: RESISC45

We display 2 example images for each class from the NWPU-RESISC45 (Cheng, Han, and Lu, 2017), a remote sensing scene classification dataset which consists of 31500 remote sensing images divided into 45 scene classes. Each class includes 700 images with a size of 256 $\times$ 256 pixels in RGB.

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

CLIP for Image Classification

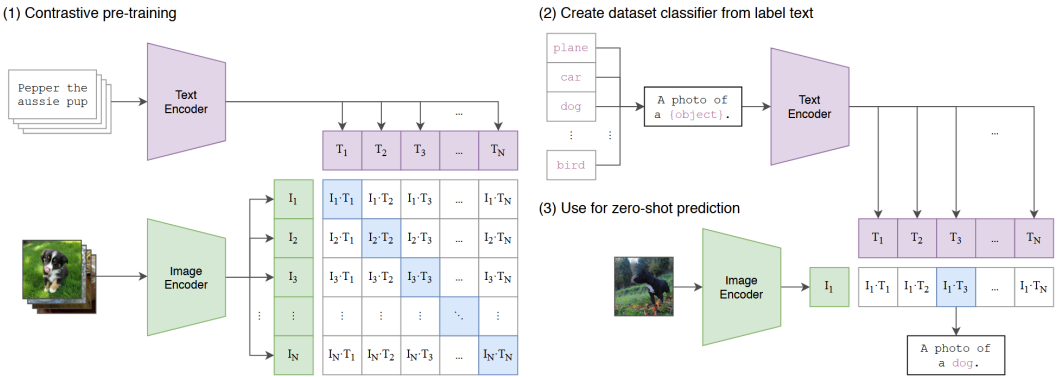


Figure: Summary of OpenAI CLIP ViT (Radford et al., 2021). Left: Training. Right: Inference.

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

Data Augmentation Methods

Traditional vs. Modern Data Augmentation Methods

Classic Methods:

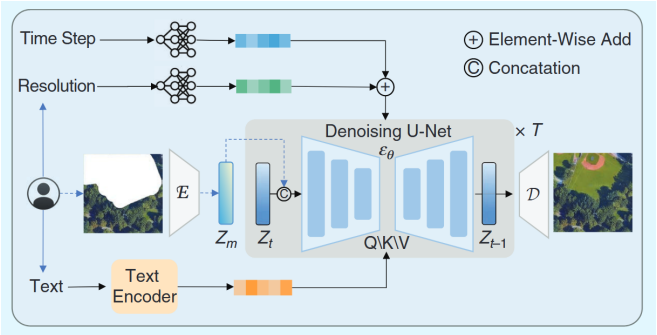
- **Geometric Transformations:** Rotation, Flipping, Scaling, Translation, Cropping

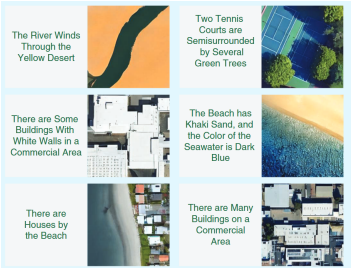
Modern Methods:

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).

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

Using Generative Model for Synthetic Dataset for Data Augmentation





Left: Text2Earth (Liu et al., 2025), a foundation model for text-driven remote sensing image generation observation. **Right:** Example results generated by Text2Earth. Generative models such as Text2Earth can be used to create synthetic remote sensing images, which can augment real datasets for tasks like image scene classification.

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

Remote Sensing Temporal Dataset for Disaster Events: xBD

Project: Explore whether generated images can benefit image classification

Project Assignment: Overview

Project: Can Generated Images Improve Remote Sensing Image Classification?

Objective:
Investigate whether combining real images and generated images can improve remote sensing image classification.

Pipeline:
In this project, you will experiment with three different training settings to evaluate the impact of generated data:

1. **Real Dataset Only:** Train the classification model using only real images from the xBD dataset.
2. **Generated Dataset Only:** Train the model using only synthetic images generated by commercial generative models.
3. **Combined Dataset:** Train the model using both real and generated images together.

Compare the classification performance across these three settings to analyze the effect of synthetic data.

Project Assignment: Dataset

Dataset: xBD Disaster Damage Dataset

- Use the **xBD** remote sensing disaster dataset.
- The dataset includes **6 disaster classes**.
- For each class, use **100 real images** (total: 600 real images).

Project Assignment: Generative Models

Image Generation:

- We consider generated images as the result of **text-guided image editing**: for each case, you input a **pre-event image** and a **text description** (e.g., "flooded", "collapsed building"), and the model yields a generated (post-event) image.
- Use commercial generative models such as **GPT-4o Image Generation(GPT-Image-1)** (OpenAI, 2025), **Gemini-2** (Gemini Team, Google, 2025), or **SeedEdit 3.0** (Wang et al., 2025) to create synthetic images for each disaster class.

OpenAI. GPT-Image-1, 2025.
Gemini Team, Google (2025). *Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities.*
Wang, et al. SeedEdit 3.0: Fast and High-Quality Generative Image Editing. arXiv. 2025.

Project Assignment: Classification Models

Recommended Baseline Models:

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

All of the above follow ViT configurations. For code and tutorials:

- CLIP training example
- ViT tutorials
- For more Remote Sensing Foundation Models, ref to huggingface collection.
- The latest strong baseline RSFM 'SkySense-O' (Zhu et al., 2025). GitHub

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.
Zhu, et al. SkySense-O: Towards Open-World Remote Sensing Interpretation with Vision-Centric Visual-Language Modeling, CVPR, 2025.

Project Assignment: Data Augmentation Protocol

- For each disaster class, generate **1**×**–4**× synthetic images (i.e., 100, 200, 300, or 400 synthetic images per class).
- Explore and compare different ratios of real to synthetic images (e.g., 1:1, 1:2, 1:3, 1:4).
- The augmented dataset for each class will range from **200 to 500 images**.

Project Assignment: Evaluation

Evaluation:

- Use **standard accuracy**, **F1 score**, and **confusion matrix** to measure performance.
- Always evaluate on a **held-out real (unseen) test set**.
- Include **curve or bar plots** comparing classification performance across different real:synthetic ratios.

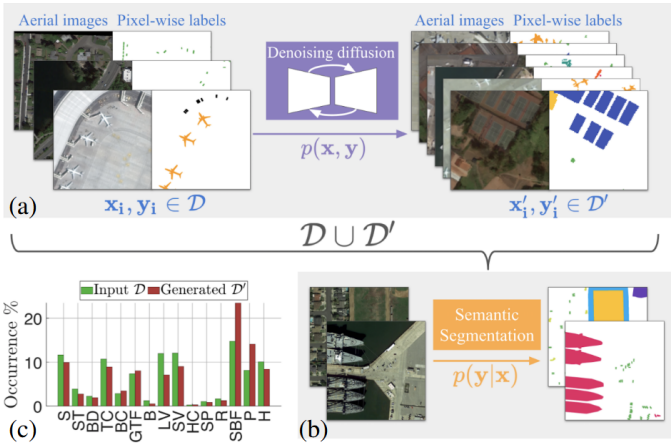
Traditional Data Augmentation Methods

- **Geometric Transformations: Rotation**, **Flipping** (horizontal/vertical), **Scaling**, **Translation**, **Cropping**
- **Color Jittering:** Adjusting brightness, contrast, saturation, and hue
- **Noise Injection:** Adding random noise to images
- **Cutout** (DeVries and Taylor, 2017)
- **CutMix** (Yun et al., 2019)
- **Copy-Paste** (Ghiasi et al., 2021)

There is also a comprehensive study entitled '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.

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.

Toker, et al. SatSynth: Augmenting Image-Mask Pairs through Diffusion Models for Aerial Semantic Segmentation, CVPR, 2024.

Generated Text-Image Dataset Improving Image Classification

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.

Synthetic text-image datasets generated by generative models can significantly improve image classification performance, as demonstrated in (He et al., 2023).

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

Additonal Text-Image Remote Sensing Datasets

Text-to-Image Generation:

- **RSICD** (Lu et al., 2018): Remote Sensing Image Captioning Dataset with 10,921 images and five captions per image.
- **RSICap** (Hu et al., 2025): High-quality dataset with 2,585 human-annotated image-caption pairs.
- **UCM-Captions** (Qu et al., 2016): Derived from the UC Merced Land Use Dataset, containing 2,100 images with five captions each.
- **RESISC45** (Cheng, Han, and Lu, 2017): It is a publicly available benchmark for REmote Sensing Image Scene Classification (RESISC), created by Northwestern Polytechnical University (NWPU). This data set contains 31 500 images, covering 45 scene classes with 700 images in each class.

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.

Generative Modeling

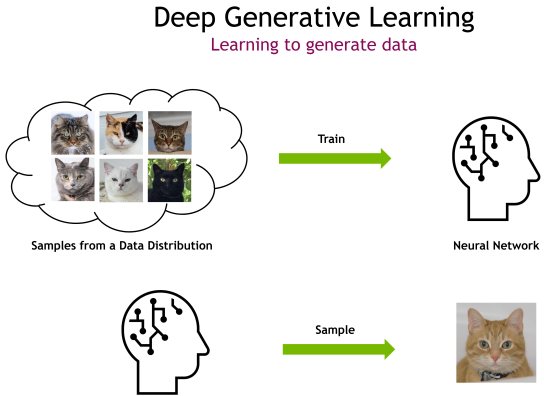


Figure: Illustration of generative modeling (Vahdat Arash, Song, and Meng, 2023).

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

Timeline of Generative Models

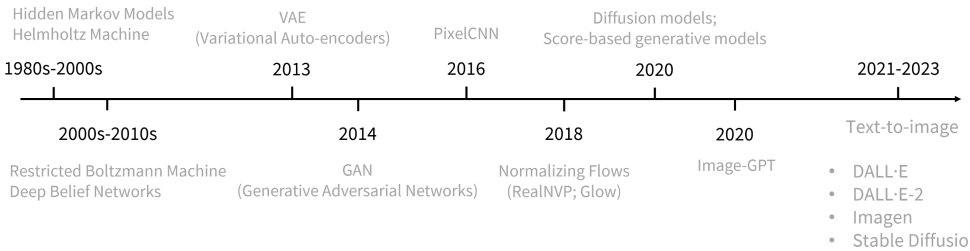


Figure: Timeline of key developments in generative models (Deng, 2024).

Background: 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.

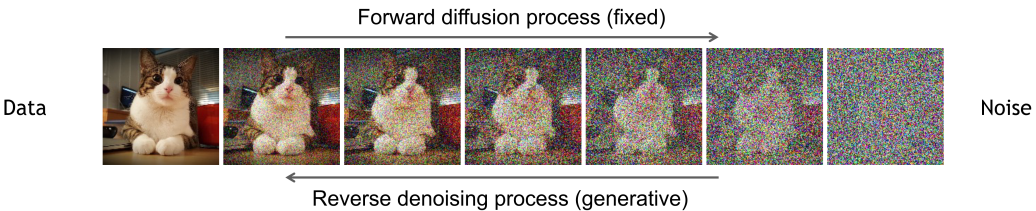


Figure: Diffusion models generate data through iterative denoising (Sohl-Dickstein et al., 2015; Ho, Jain, and Abbeel, 2020).

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

Diffusion Models: Forward and Reverse Processes

Forward (Diffusion) Process:

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

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

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

Reverse (Denoising) Process:

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

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

where $\mathbf{x}_0$ is the data, β_t is the noise schedule, and $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$. $p(\mathbf{x}_T) = \mathcal{N}(\mathbf{0}, \mathbf{I})$.
Diffusion models generate data by learning to reverse a gradual noising process. (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.

Diffusion Models: Training and Inference

Training Objective:

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

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

Inference (Sampling):

- Start from pure noise: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
- For $t = T, \ldots, 1$:
  - Predict noise: $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$
  - Compute mean: $\boldsymbol{\mu}_\theta(\mathbf{x}_t, t)$
  - Sample: $\mathbf{x}_{t-1} \sim \mathcal{N}(\boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t))$
- Repeat until $\mathbf{x}_0$ (generated sample)

Training: Minimize the simplified objective (Ho, Jain, and Abbeel, 2020).
Inference: Iteratively denoise from random noise to generate data.

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

Application in Remote Sensing Image Generation: Text2Earth

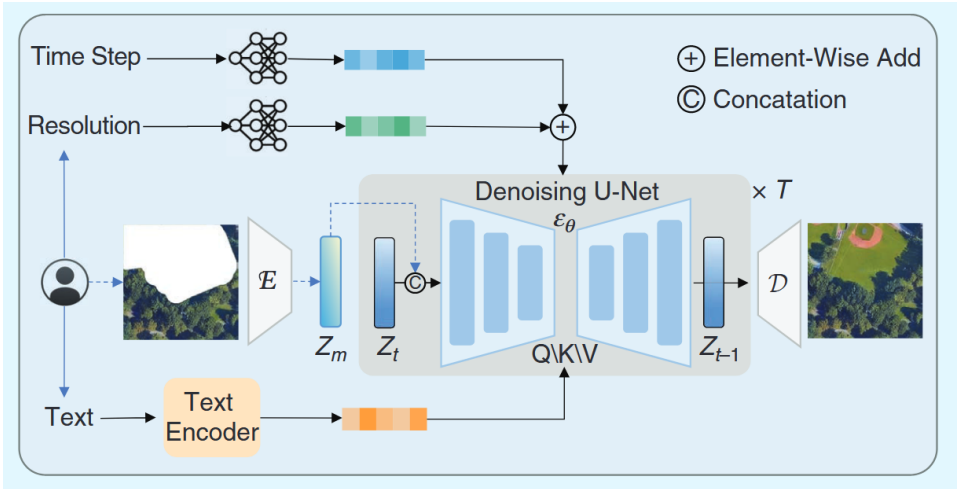


Figure: Text2Earth: Foundation model for text-driven Earth observation (Liu et al., 2025).

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

Text2Earth: Example Results

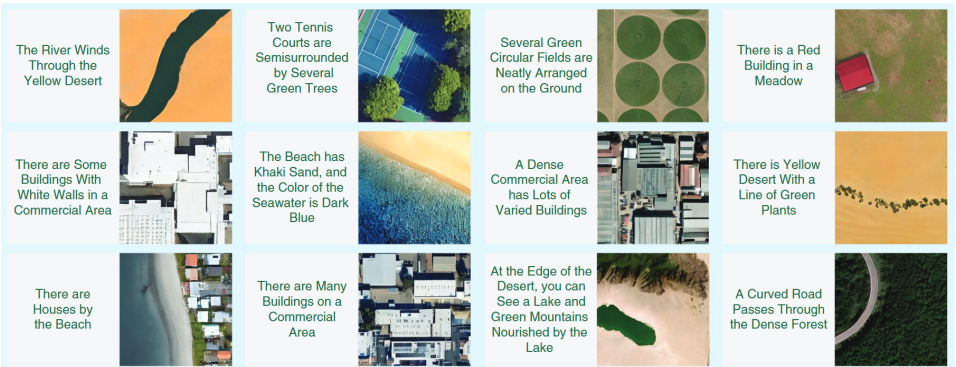


Figure: Example results generated by Text2Earth (Liu et al., 2025).

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

Application in Remote Sensing Image Generation: CRS-Diff

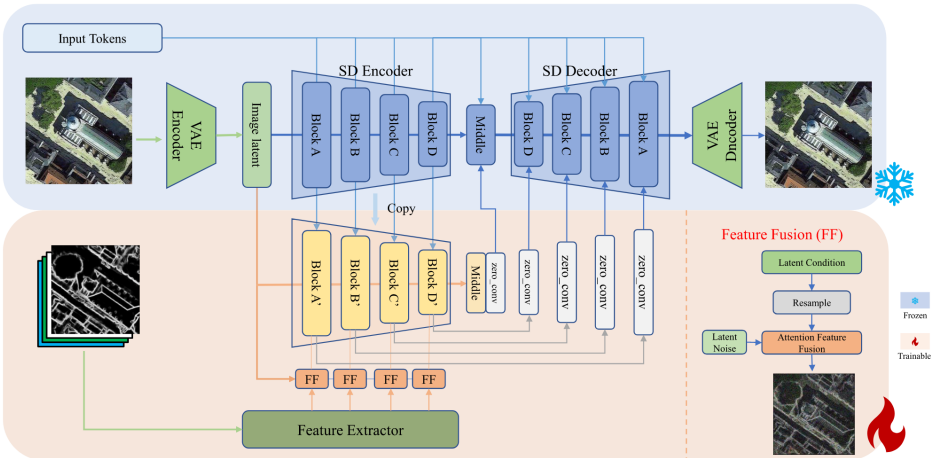


Figure: CRS-Diff: Controllable remote sensing image generation framework (Tang, Li, et al., 2024).

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

CRS-Diff: Example Results

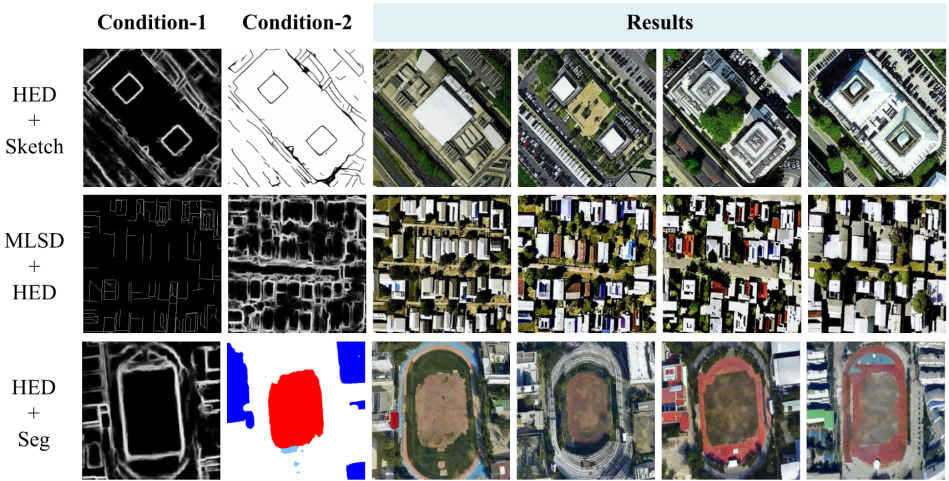


Figure: Example results generated by CRS-Diff (Tang, Li, et al., 2024).

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

DiffusionSat: Framework Overview

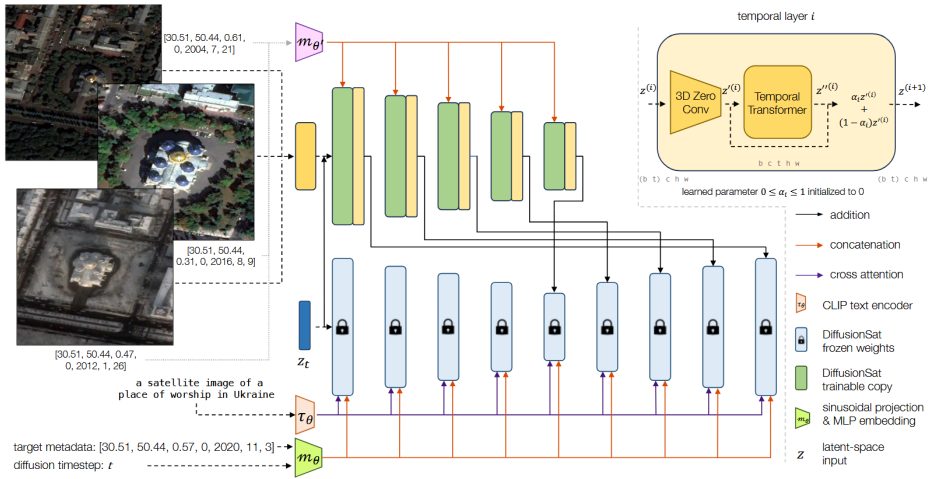


Figure: DiffusionSat: A generative foundation model for satellite imagery (Khanna et al., 2024).

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

DiffusionSat: Super-Resolution Results

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

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

DiffusionSat: Inpainting Results

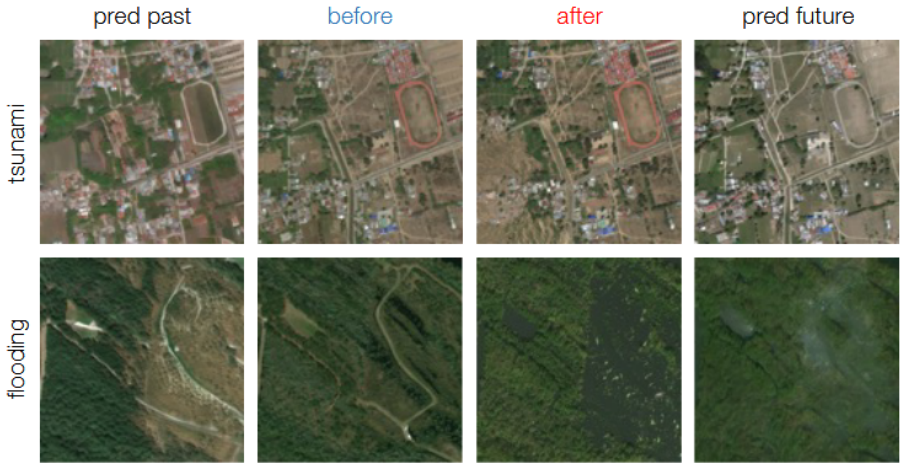


Figure: Example results: DiffusionSat for remote sensing image inpainting (Khanna et al., 2024).

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