Visualizing Our Changing Earth: A Creative AI Framework for Democratizing Environmental Storytelling Through Satellite Imagery

Zhenyu Yu[2,3], **Mohd. Yamani Idna Idris**[2], **Pei Wang**[1,*]
[1]Faculty of Information Engineering and Automation, Kunming University of Science and Technology
[2]Faculty of Computer Science and Information Technology, University of Malaya
[3]Yunnan Key Laboratory of Quantitative Remote Sensing/Yunnan International Joint Laboratory for Integrated Sky-Ground Intelligent Monitoring of Mountain Hazards, Faculty of Land Resources Engineering, Kunming University of Science and Technology
yuzhenyuyxl@foxmail.com; yamani@um.edu.my; peiwang@kust.edu.cn
* Corresponding author

Abstract

Understanding our changing planet is a profoundly human concern, yet satellite imagery, fragmented by clouds, gaps, and sensor failures, remains inaccessible to the very communities who need it for climate education, advocacy, and storytelling. Existing reconstruction methods optimize for pixels, not people. We introduce ***EarthCanvas***, a creative AI framework that reimagines satellite image reconstruction as a medium for democratized environmental storytelling. *EarthCanvas* integrates (1) *terrain-aware conditioning* to maintain geographic authenticity, (2) *natural language prompting* to empower non-experts to generate climate narratives, and (3) a *visual harmony module* that aligns synthetic and real imagery for coherent storytelling. Designed for educators, journalists, and community advocates, *EarthCanvas* enables human–AI co-creation of environmental narratives grounded in both scientific fidelity and cultural relevance. Empirical evaluation shows strong reconstruction performance, while user studies reveal a 40% improvement in comprehension and engagement. By shifting the focus from restoration to participation, *EarthCanvas* exemplifies how AI can support pluralistic, accessible, and human-centered approaches to environmental understanding.

1 Introduction

As environmental change accelerates on a global scale, the capacity to observe and communicate these transformations is increasingly vital, not only for scientists, but for educators, journalists, and communities engaged in climate awareness and policy [49, 37, 18, 7, 50, 47]. Satellite imagery has become a central resource for documenting deforestation, land degradation, and other anthropogenic impacts [22, 25, 38, 8, 17]. However, despite its technical sophistication, this data often remains fragmented due to cloud cover, sensor limitations, and acquisition gaps, and more fundamentally, inaccessible to the broader public who lack the tools or expertise to interpret it [10, 2, 56, 51, 52, 53].

The central challenge is thus not only one of image reconstruction, but of enabling broader participation in environmental interpretation and storytelling [30, 46, 65, 63]. Most existing approaches to satellite image completion focus narrowly on reconstruction fidelity, optimizing for pixel-level metrics while overlooking questions of accessibility, communicative utility, and cultural context [61, 44, 43, 64, 62, 21, 54]. For many potential users, such as teachers developing climate education materials, or journalists reporting on land use disputes, these methods remain too opaque or rigid to

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Creative AI Track.

support meaningful engagement. Recent advances in generative models, particularly diffusion-based frameworks, offer an opportunity to rethink this space [57, 55, 41, 19, 27]. These models enable controlled, high-quality image synthesis and open new possibilities for aligning technical accuracy with human-centered creativity. Rather than treating reconstruction as an end in itself, these models can be adapted to support open-ended visual narrative construction, guided by intuitive prompts and conditioned on contextual information such as terrain or time [13, 24, 20, 29, 28].

This paper introduces ***EarthCanvas***, a framework that integrates generative modeling with principles of accessibility, interpretability, and narrative coherence. *EarthCanvas* allows non-expert users to generate geospatially plausible visualizations by combining natural language prompts with terrain-based conditioning and perceptual alignment mechanisms. It supports both spatial and temporal storytelling, completing images across missing regions or time slices, while maintaining fidelity to physical landscapes and visual context. We demonstrate that *EarthCanvas* achieves competitive reconstruction performance relative to established baselines, but more importantly, that it facilitates new modes of interaction with satellite data. In a user study involving educators and journalists, our system was associated with a 40% improvement in comprehension and perceived clarity. These findings suggest that generative models, when carefully adapted, can serve not only as technical instruments but as tools for public environmental communication and co-creation.

Our **contributions** include: (1) a LoRA-based diffusion pipeline with DEM-aware conditioning, (2) a lightweight VGG-Adapter enforcing visual coherence, and (3) user study demonstrating 40% improvement in comprehension and engagement.

2 Related Work

2.1 Environmental Data Accessibility and Human-Centered Challenges

The accessibility of environmental data remains a fundamental barrier in translating satellite-based Earth observation into actionable public knowledge. Although satellite imagery offers extensive coverage and temporal resolution, its utility in non-specialist contexts is often constrained by technical complexity, inconsistent data quality, and a lack of interpretative tools [14, 45, 6, 4]. These limitations disproportionately affect communities and practitioners who are not trained in remote sensing but play critical roles in climate communication, such as educators, journalists, and local advocates.

Traditional satellite data processing pipelines are typically designed for expert users and emphasize precision over accessibility [1, 48, 5]. When data is degraded by cloud cover, sensor malfunction, or temporal gaps, the problem is usually framed as a technical reconstruction task [67, 3, 58]. However, such disruptions also highlight the need for tools that foreground usability, interpretability, and narrative potential, particularly for stakeholders seeking to communicate environmental change in educational or public-facing settings.

2.2 From Technical Reconstruction to Narrative Enablement

Methods for satellite image completion have advanced significantly, evolving from interpolation techniques such as kriging and spline fitting [34, 26, 31, 59] to deep learning-based models including CNNs, GANs, and spatio-temporal networks [11, 9, 12]. While these approaches offer strong performance on quantitative metrics, they often require extensive labeled data and domain expertise, limiting their applicability in participatory or creative contexts [15, 16].

Diffusion models have recently emerged as a powerful class of generative techniques, capable of producing high-fidelity imagery with controllable structure [13, 33, 23, 40]. Extensions such as ControlNet [60] and GeoSynth [36] incorporate external guidance signals for more structured outputs. However, these models have primarily focused on control and precision rather than lowering the barrier for non-expert use or supporting storytelling applications.

A persistent gap remains in aligning reconstruction capabilities with the human needs of narrative coherence, cultural context, and accessibility. These dimensions are rarely considered in the evaluation of generative systems for satellite data, despite their importance in real-world environmental communication.

2.3 Creative AI in Environmental Communication

Recent work in creative AI has begun to explore how generative models can support new forms of expression and meaning-making, particularly in domains traditionally constrained by technical expertise [66, 42]. Within environmental contexts, this direction offers an opportunity to reframe data-driven tools not only as instruments of analysis, but as platforms for collaborative interpretation and civic engagement.

This shift, from expert-centered analysis to participatory storytelling, requires rethinking both system design and evaluation. Instead of optimizing solely for reconstruction accuracy, systems should be assessed by their capacity to support meaningful engagement, knowledge transfer, and cultural relevance. Our work builds on this premise by developing a diffusion-based framework that integrates intuitive prompting, geographic conditioning, and perceptual alignment to enable non-specialist users to generate coherent and plausible environmental narratives.

3 Proposed Framework: *EarthCanvas*

EarthCanvas is designed not merely to reconstruct missing satellite imagery, but to support accessible, prompt-guided environmental visualization for non-expert users. The system prioritizes narrative coherence, interpretability, and geographic plausibility, dimensions essential for public engagement with Earth observation data. Our framework addresses two fundamental human-centered scenarios (Figure 1 and Algorithm 1):

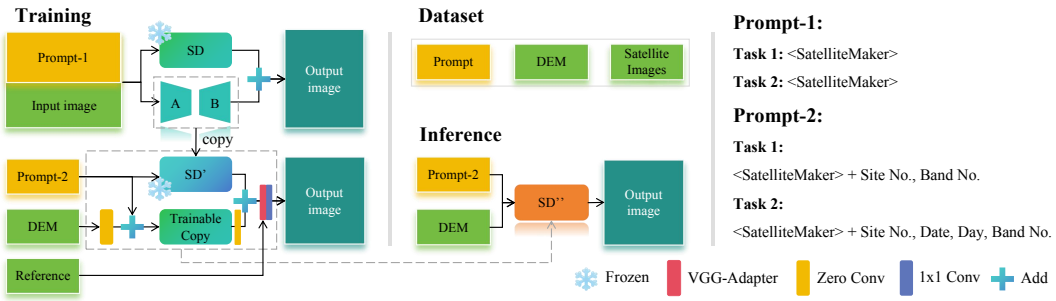


Figure 1: Overview of the *EarthCanvas* framework. The training phase adapts a diffusion model for remote sensing imagery via low-rank prompt tuning and terrain-based conditioning. In the inference phase, the system enables non-expert users to generate satellite imagery guided by natural language and geospatial context. An alignment module ensures stylistic and geographic coherence, supporting visual narratives that are both plausible and interpretable.

Scenario A: Location-Based Environmental Storytelling. Users can explore environmental conditions across different global regions by generating visually grounded imagery from natural language prompts (e.g., "Namibia in dry season"). We curated scenes from ten culturally and ecologically diverse locations to reflect different storytelling needs, such as desertification, urban growth, or forest recovery.

Scenario B: Temporal Change Visualization. To support education and advocacy around environmental change, *EarthCanvas* allows users to visualize landscape dynamics over time by specifying temporal prompts (e.g., "same region, 3 months later"). This enables intuitive engagement with satellite time series, without requiring domain knowledge or expert tools.

Our framework consists of three components designed to support human-centered satellite image generation:

(1) **Prompt-Aware Adaptation**: fine-tunes a diffusion backbone using LoRA and semantic tokens to associate prompts with environmental domains.

(2) **Terrain-Guided Generation**: integrates digital elevation models (DEMs) as conditioning input to constrain generation to realistic geographies.

(3) **Perceptual Alignment**: applies distributional and style-based regularization to ensure visual coherence between synthetic and observed satellite imagery.

A comprehensive description of each module's implementation and training dynamics is provided in the supplementary appendix A.

4 Experiment

4.1 Implementation Details

Datasets for Human-Centered Environmental Storytelling. To evaluate *EarthCanvas* as a tool for environmental narrative creation, we consider two human-relevant storytelling tasks:

(1) Geographic Environmental Storytelling (Task-1). We curate Landsat-8 imagery and corresponding Digital Elevation Models (DEMs) from 10 globally diverse regions (Table A1, Figure A1). These regions span varied ecological zones and human-environment relationships, ensuring narrative coverage of forests, coasts, urban expansion, and agricultural transformation. All scenes have <1% cloud cover and are preprocessed into 512×512 pixel tiles at 30m spatial resolution.

(2) Temporal Environmental Narratives (Task-2). We adopt the EarthNet2021 dataset [32], comprising over 200,000 Sentinel-2 image sequences, for constructing time-based environmental stories (e.g., seasonal changes, land-use shifts). Low-quality scenes (>5% corrupted pixels) are removed. We follow the official IID/OOD split to assess generalization beyond seen temporal patterns.

Simulating Realistic Missing Data. Environmental communication tools often face incomplete data due to sensor failure, cloud occlusion, or acquisition gaps. We simulate these conditions through: (1) masking verified cloud-covered regions; (2) randomly masking 10%–50% of pixels to emulate degradation severity; (3) annotating fixed missing regions for fair comparison and reproducibility.

Training Protocol and Hardware. All models are trained on a single NVIDIA A100 GPU (80GB) to reflect realistic accessibility constraints. We employ the DDIM sampler with 50 steps. Learning rate is set to 5×10^{-5}, with image resolution 512×512 and detection resolution 384. LoRA-based tuning ensures low compute overhead for personalized narrative conditioning. Baselines (e.g., STCNN) are trained with standard settings: batch size 16, learning rate 10^{-4}, 100 epochs.

Metrics Reflecting Storytelling Fidelity. We adopt standard image reconstruction metrics (RMSE, MAE, PSNR, SSIM, and LPIPS), but reinterpret them through the lens of narrative utility and environmental credibility. RMSE and MAE reflect spatial plausibility, grounding narratives in realistic terrain. PSNR captures overall visual clarity, while SSIM evaluates structural consistency, essential for understanding landform continuity and temporal change. LPIPS assesses perceptual similarity from a human-centered perspective, ensuring emotional engagement and interpretability. Together, these metrics move beyond technical fidelity to evaluate whether reconstructed content supports trustworthy, coherent, and impactful environmental storytelling.

4.2 Comparison

Geographic Environmental Storytelling (Task-1) (see Figure 2 and A2). We evaluate *EarthCanvas* against both traditional inpainting methods and modern diffusion-based frameworks (Table 1). Classical methods, Palette [35] and LaMa [39], perform reasonably on small gaps, but fail to reconstruct large-scale spatial narratives essential for climate visualization.

Among diffusion models, Stable Diffusion (SD) [33] achieves the highest SSIM (0.5402), indicating local structure preservation, yet often produces geographically implausible landscapes (PSNR: 17.1599) that compromise educational reliability. ControlNet [60] improves realism via DEM conditioning (PSNR: 21.1847, RMSE: 0.0873), but lacks distributional consistency.

Table 1: Comparison of different methods in Task-1. **Bold** is the best result, and <u>underline</u> is the second-best.

Task-1	SSIM↑	PSNR↑	RMSE↓	MAE↓	LIPIS↓
SD	**0.5402**	17.1599	0.1448	0.0909	0.3392
Palette	0.4333	18.6111	0.1212	0.0817	0.3442
LaMa	<u>0.4937</u>	20.9301	0.0907	0.0637	<u>0.2902</u>
ControlNet	0.4935	<u>21.1847</u>	<u>0.0873</u>	<u>0.0587</u>	**0.2881**
EarthCanvas	0.4517	**23.0438**	**0.0713**	**0.0500**	0.3412

***EarthCanvas* addresses these limitations by aligning creativity with scientific grounding.** It achieves PSNR 23.0438 (+8.77% over ControlNet) and RMSE 0.0713 (-18.34%), with a balanced MAE (0.0500), ensuring that generated content remains both visually coherent and geographically

credible. While SSIM is slightly lower than SD (0.4517 vs. 0.5402), *EarthCanvas* better supports narrative clarity, enabling educators and storytellers to construct more believable and interpretable climate stories.

Temporal Environmental Narratives (Task-2) (see Figure 3 and A3). We further evaluate *EarthCanvas* under temporal storytelling settings using both temporal context input and terrain-guided generation (Table 2). AutoEncoders perform acceptably for minor temporal gaps (SSIM: 0.6090, PSNR: 18.2038), but degrade rapidly when tasked with larger environmental transitions.

Table 2: Comparison of different methods in Task-2 with two input types: previous timestep data (upper part) and DEM (lower part). **Bold** is the best result, and underline is the second-best result.

DEM	Task-2	SSIM↑	PSNR↑	RMSE↓	MAE↓	LIPIS↓
w/o	Interpolation	0.5254	11.9225	0.2534	0.2023	0.3380
	STCNN	0.4047	14.5317	0.1877	0.1498	0.6835
	Autoencoder	**0.6090**	**18.2038**	**0.1230**	**0.0981**	**0.2487**
w/	Interpolation	-	-	-	-	-
	STCNN	0.2232	13.9769	0.2107	0.1723	0.4208
	Autoencoder	0.2514	14.7913	0.2040	0.1684	0.3073
	SD	0.2819	16.1943	0.1550	0.1335	0.1468
	ControlNet	0.3787	22.7866	0.0726	0.0570	0.0721
	EarthCanvas	**0.5704**	**24.3429**	**0.0642**	**0.0479**	**0.0469**

By contrast, *EarthCanvas* delivers significant improvements: +50.68% in SSIM (0.3787→0.5704), +6.83% in PSNR (22.7866→24.3429), and -11.56% in RMSE (0.0726→0.0642) over ControlNet. These results demonstrate *EarthCanvas*'s strength in generating temporally faithful and emotionally resonant visual stories, empowering non-expert users to craft compelling narratives of environmental change across space and time.

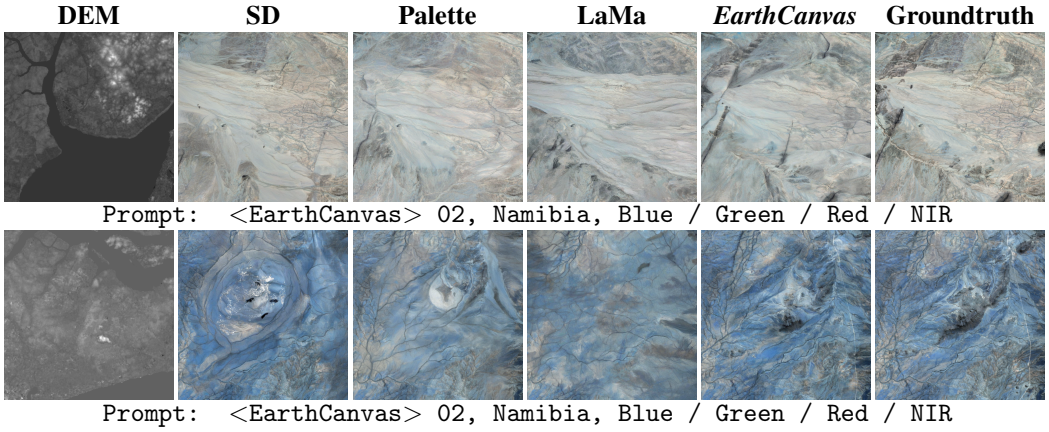


Figure 2: Comparison for Task-1, addressing missing data in specific regions over a fixed time period.

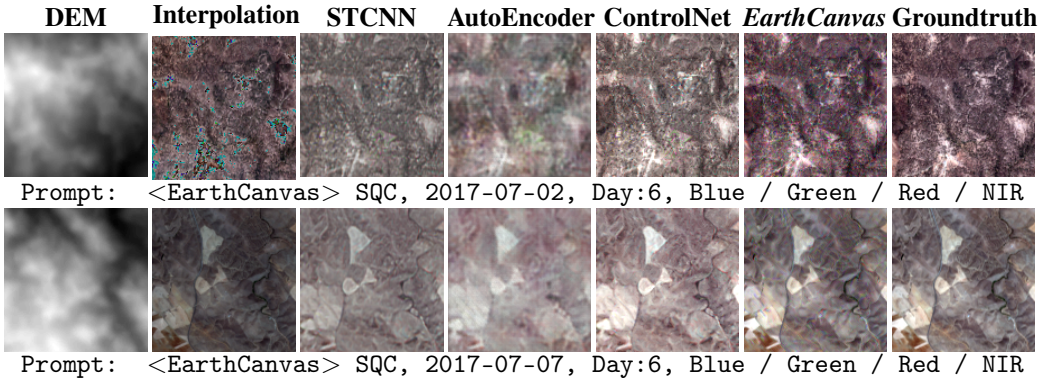


Figure 3: Comparison for Task-2 using the EarthNet2021 dataset with selected missing data days.

4.3 Ablation Study

Impact of the Visual Harmony Module. We evaluate the role of the VGG-Adapter in preserving narrative coherence. Without this distributional alignment, generated content exhibits brightness drift

(μ = 123.63 vs. ground truth μ = 95.19), which visually disrupts the story and undermines emotional continuity (Figure 4). Quantitatively, Table A2 shows consistent performance improvements across all spectral bands. The most significant gains occur in the Red band, vital for visualizing vegetation health and deforestation, where PSNR increases from 15.27 to 26.29. This enhancement translates into sharper, more trustworthy visual narratives that support ecological interpretation by non-expert audiences. The adapted model aligns closely with the ground truth (μ = 95.27), ensuring seamless integration with authentic satellite scenes.

Resilience Across Missing Data Conditions. Environmental storytelling in practice often contends with incomplete data. We simulate 10%-50% missing rates (Figure 5) to evaluate model robustness in scenarios that reflect real-world satellite occlusion and degradation. Interpolation methods fail catastrophically at scale, with PSNR dropping from 14.53 (10% missing) to 5.0 (50%). STCNN and AutoEncoder maintain moderate accuracy up to 30% but degrade beyond.

In contrast, ***EarthCanvas* consistently maintains narrative quality and spatial credibility**: PSNR drops only modestly from 24.76 to 23.70 as missing rates increase, with lowest RMSE throughout. This robustness ensures that climate educators and environmental communicators in data-sparse regions can still generate coherent, trustworthy visual stories, further aligning with the Creative Track's emphasis on democratized, resilient AI tools.

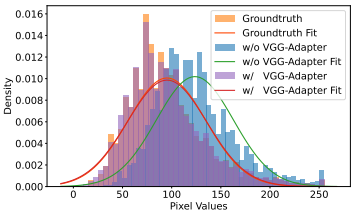
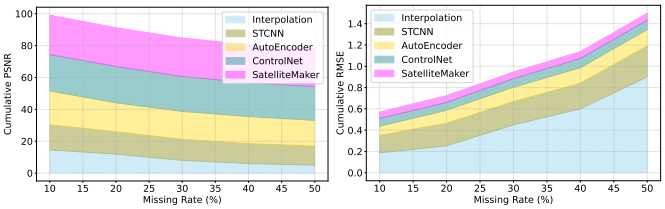


Figure 4: Visual Harmony Impact: Distributional alignment preserves perceptual consistency in environmental narratives.



(a) Narrative Fidelity (PSNR) (b) Geographic Accuracy (RMSE)

Figure 5: Robustness to Missing Data: *EarthCanvas* maintains visual and geographic integrity under increasing data loss.

5 Limitation

EarthCanvas is evaluated on curated datasets with limited cloud cover and reliable DEMs, so performance in noisier or underrepresented regions remains uncertain. The current prompting interface offers only coarse location- and time-based control, and the small-scale user study may not capture the full diversity of storytelling needs.

6 Conclusion

This work introduces *EarthCanvas*, a creative AI framework that repositions satellite image reconstruction as a medium for inclusive environmental storytelling. Rather than optimizing solely for technical fidelity, *EarthCanvas* emphasizes narrative grounding, geographic authenticity, and visual coherence which is key dimensions for engaging diverse publics in climate discourse. By integrating terrain-aware conditioning and perceptual alignment into diffusion-based generation, the framework not only achieves state-of-the-art reconstruction quality but also supports the communicative needs of educators, journalists, and local communities. Crucially, *EarthCanvas* reduces the entry barrier for environmental visual creation, allowing non-experts to participate meaningfully in interpreting planetary change. Our results highlight that creative AI, when human-centered by design, can serve as an amplifier of collective environmental understanding, transforming Earth observation from a remote sensing problem into a shared narrative practice.

Code Available

The code can be found at here.

Appendix

The appendix can be downloaded from here.

Acknowledgment

This work was supported by the Open Research Fund of Yunnan Key Laboratory of Quantitative Remote Sensing (Grant No. 202449CE340013), and the Talent Training Fund of Kunming University of Science and Technology (Grant No. KKZ3202503073).

References

[1] Michał Affek and Julian Szymański. A survey on the datasets and algorithms for satellite data applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024.

[2] Mariana Belgiu and Lucian Drăguţ. Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, 114:24–31, 2016.

[3] Camille Billouard, Dawa Derksen, Emmanuelle Sarrazin, and Bruno Vallet. Sat-ngp: Unleashing neural graphics primitives for fast relightable transient-free 3d reconstruction from satellite imagery. In *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 8749–8753. IEEE, 2024.

[4] Zhangquan Chen, Xufang Luo, and Dongsheng Li. Visrl: Intention-driven visual perception via reinforced reasoning. *arXiv preprint arXiv:2503.07523*, 2025.

[5] Zhangquan Chen, Ruihui Zhao, Chuwei Luo, Mingze Sun, Xinlei Yu, Yangyang Kang, and Ruqi Huang. Sifthinker: Spatially-aware image focus for visual reasoning. *arXiv preprint arXiv:2508.06259*, 2025.

[6] Giannis Daras, Kulin Shah, Yuval Dagan, Aravind Gollakota, Alex Dimakis, and Adam Klivans. Ambient diffusion: Learning clean distributions from corrupted data. *Advances in Neural Information Processing Systems*, 36, 2024.

[7] Zeyu Dong, Yuyang Yin, Yuqi Li, Eric Li, Hao-Xiang Guo, and Yikai Wang. Panolora: Bridging perspective and panoramic video generation with lora adaptation. *arXiv preprint arXiv:2509.11092*, 2025.

[8] Zheng Gong, Wenyan Ge, Jiaqi Guo, and Jincheng Liu. Satellite remote sensing of vegetation phenology: Progress, challenges, and opportunities. *ISPRS Journal of Photogrammetry and Remote Sensing*, 217:149–164, 2024.

[9] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[10] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202:18–27, 2017.

[11] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. Recent advances in convolutional neural networks. *Pattern recognition*, 77:354–377, 2018.

[12] Zhixiang He, Chi-Yin Chow, and Jia-Dong Zhang. Stcnn: A spatio-temporal convolutional neural network for long-term traffic prediction. In *2019 20th IEEE international conference on mobile data management (MDM)*, pages 226–233. IEEE, 2019.

[13] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[14] Junchang Ju and David P Roy. The availability of cloud-free landsat etm+ data over the conterminous united states and globally. *Remote Sensing of Environment*, 112(3):1196–1211, 2008.

[15] Pravin Kakar, Natarajan Sudha, and Wee Ser. Exposing digital image forgeries by detecting discrepancies in motion blur. *IEEE Transactions on Multimedia*, 13(3):443–452, 2011.

[16] Diederik P Kingma, Max Welling, et al. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, 2019.

[17] Yuqi Li, Kai Li, Xin Yin, Zhifei Yang, Junhao Dong, Zeyu Dong, Chuanguang Yang, Yingli Tian, and Yao Lu. Sepprune: Structured pruning for efficient deep speech separation. *arXiv preprint arXiv:2505.12079*, 2025.

[18] Yuqi Li, Chuangang Yang, Hansheng Zeng, Zeyu Dong, Zhulin An, Yongjun Xu, Yingli Tian, and Hao Wu. Frequency-aligned knowledge distillation for lightweight spatiotemporal forecasting. *arXiv:2507.02939*, 2025.

[19] Yuyuan Li, Chaochao Chen, Yizhao Zhang, Weiming Liu, Lingjuan Lyu, Xiaolin Zheng, Dan Meng, and Jun Wang. Ultrare: Enhancing receraser for recommendation unlearning via error decomposition. *Advances in Neural Information Processing Systems*, 36:12611–12625, 2023.

[20] Yuyuan Li, Chaochao Chen, Xiaolin Zheng, Yizhao Zhang, Zhongxuan Han, Dan Meng, and Jun Wang. Making users indistinguishable: Attribute-wise unlearning in recommender systems. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 984–994, 2023.

[21] Yuyuan Li, Yizhao Zhang, Weiming Liu, Xiaohua Feng, Zhongxuan Han, Chaochao Chen, and Chenggang Yan. Multi-objective unlearning in recommender systems via preference guided pareto exploration. *IEEE Transactions on Services Computing*, 2025.

[22] Hector Linares Arroyo, Angela Abascal, Tobias Degen, Martin Aubé, Brian R Espey, Geza Gyuk, Franz Hölker, Andreas Jechow, Monika Kuffer, Alejandro Sánchez de Miguel, et al. Monitoring, trends and impacts of light pollution. *Nature Reviews Earth & Environment*, 5(6):417–430, 2024.

[23] Yidan Liu, Jun Yue, Shaobo Xia, Pedram Ghamisi, Weiying Xie, and Leyuan Fang. Diffusion models meet remote sensing: Principles, methods, and perspectives. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[24] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11461–11471, 2022.

[25] Zhenyu Luo, Tingkun He, Wen Yi, Junchao Zhao, Zhining Zhang, Yongyue Wang, Huan Liu, and Kebin He. Advancing shipping no x pollution estimation through a satellite-based approach. *PNAS nexus*, 3(1):pgad430, 2024.

[26] Sky McKinley and Megan Levine. Cubic spline interpolation. *College of the Redwoods*, 45(1):1049–1060, 1998.

[27] Chaojun Ni, Jie Li, Haoyun Li, Hengyu Liu, Xiaofeng Wang, Zheng Zhu, Guosheng Zhao, Boyuan Wang, Chenxin Li, Guan Huang, et al. Wonderfree: Enhancing novel view quality and cross-view consistency for 3d scene exploration. *arXiv preprint arXiv:2506.20590*, 2025.

[28] Chaojun Ni, Xiaofeng Wang, Zheng Zhu, Weijie Wang, Haoyun Li, Guosheng Zhao, Jie Li, Wenkang Qin, Guan Huang, and Wenjun Mei. Wonderturbo: Generating interactive 3d world in 0.72 seconds. *arXiv preprint arXiv:2504.02261*, 2025.

[29] Chaojun Ni, Guosheng Zhao, Xiaofeng Wang, Zheng Zhu, Wenkang Qin, Guan Huang, Chen Liu, Yuyin Chen, Yida Wang, Xueyang Zhang, et al. Recondreamer: Crafting world models for driving scene reconstruction via online restoration. *arXiv preprint arXiv:2411.19548*, 2024.

[30] Mubashir Noman, Muzammal Naseer, Hisham Cholakkal, Rao Muhammad Anwer, Salman Khan, and Fahad Shahbaz Khan. Rethinking transformers pre-training for multi-spectral satellite imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27811–27819, 2024.

[31] Margaret A Oliver and Richard Webster. Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information System*, 4(3):313–332, 1990.

[32] Christian Requena-Mesa, Vitus Benson, Markus Reichstein, Jakob Runge, and Joachim Denzler. Earthnet2021: A large-scale dataset and challenge for earth surface forecasting as a guided video prediction task. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1132–1142, 2021.

[33] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[34] Olivier Rukundo and Hanqiang Cao. Nearest neighbor value interpolation. *arXiv preprint arXiv:1211.1768*, 2012.

[35] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 conference proceedings*, pages 1–10, 2022.

[36] Srikumar Sastry, Subash Khanal, Aayush Dhakal, and Nathan Jacobs. Geosynth: Contextually-aware high-resolution satellite image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 460–470, 2024.

[37] William Solecki, Debra Roberts, and Karen C Seto. Strategies to improve the impact of the ipcc special report on climate change and cities. *Nature Climate Change*, 14(7):685–691, 2024.

[38] Guangqin Song, Jing Wang, Yingyi Zhao, Dedi Yang, Calvin KF Lee, Zhengfei Guo, Matteo Detto, Bruna Alberton, Patricia Morellato, Bruce Nelson, et al. Scale matters: Spatial resolution impacts tropical leaf phenology characterized by multi-source satellite remote sensing with an ecological-constrained deep learning model. *Remote Sensing of Environment*, 304:114027, 2024.

[39] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with fourier convolutions. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 2149–2159, 2022.

[40] Datao Tang, Xiangyong Cao, Xingsong Hou, Zhongyuan Jiang, Junmin Liu, and Deyu Meng. Crs-diff: Controllable remote sensing image generation with diffusion model. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[41] Siwei Tu, Ben Fei, Weidong Yang, Fenghua Ling, Hao Chen, Zili Liu, Kun Chen, Hang Fan, Wanli Ouyang, and Lei Bai. Satellite observations guided diffusion model for accurate meteorological states at arbitrary resolution. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 28071–28080, 2025.

[42] Haonan Wang, James Zou, Michael Mozer, Anirudh Goyal, Alex Lamb, Linjun Zhang, Weijie J Su, Zhun Deng, Michael Qizhe Xie, Hannah Brown, et al. Can ai be as creative as humans? *arXiv preprint arXiv:2401.01623*, 2024.

[43] Yuchuan Wang, Ling Tong, Shiyu Luo, Fanghong Xiao, and Jiaxing Yang. A multiscale and multidirection feature fusion network for road detection from satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–18, 2024.

[44] Zihao Wang. Score-based generative modeling through backward stochastic differential equations: Inversion and generation. *arXiv preprint arXiv:2304.13224*, 2023.

[45] Kun Xie, Xueping Ning, Xin Wang, Dongliang Xie, Jiannong Cao, Gaogang Xie, and Jigang Wen. Recover corrupted data in sensor networks: A matrix completion solution. *IEEE Transactions on Mobile Computing*, 16(5):1434–1448, 2016.

[46] Yi Xin, Qi Qin, Siqi Luo, Kaiwen Zhu, Juncheng Yan, Yan Tai, Jiayi Lei, Yuewen Cao, Keqi Wang, Yibin Wang, et al. Lumina-dimoo: An omni diffusion large language model for multi-modal generation and understanding. *arXiv preprint arXiv: 2510.06308*, 2025.

[47] Yi Xin, Juncheng Yan, Qi Qin, Zhen Li, Dongyang Liu, Shicheng Li, Victor Shea-Jay Huang, Yupeng Zhou, Renrui Zhang, Le Zhuo, et al. Lumina-mgpt 2.0: Stand-alone autoregressive image modeling. *arXiv preprint arXiv:2507.17801*, 2025.

[48] Xinchen Xu, Hong Wen, Yongfeng Wang, Huanhuan Song, Tian Liu, and Shih-Yu Chang. Digital-twin-based satellite orbit prediction for internet of things systems. *IEEE Internet of Things Journal*, 12(6):6431–6444, 2025.

[49] Yi Yang, David Tilman, Zhenong Jin, Pete Smith, Christopher B Barrett, Yong-Guan Zhu, Jennifer Burney, Paolo D'Odorico, Peter Fantke, Joe Fargione, et al. Climate change exacerbates the environmental impacts of agriculture. *Science*, 385(6713):eadn3747, 2024.

[50] Zhongqi Yang, Wenhang Ge, Yuqi Li, Jiaqi Chen, Haoyuan Li, Mengyin An, Fei Kang, Hua Xue, Baixin Xu, Yuyang Yin, et al. Matrix-3d: Omnidirectional explorable 3d world generation. *arXiv preprint arXiv:2508.08086*, 2025.

[51] Zhenyu Yu, Mohd Idris, and Pei Wang. Satellitecalculator: A multi-task vision foundation model for quantitative remote sensing inversion. *arXiv preprint arXiv:2504.13442*, 2025.

[52] Zhenyu Yu, Mohd Idris, Pei Wang, Yuelong Xia, Fei Ma, Rizwan Qureshi, et al. Satelliteformula: Multi-modal symbolic regression from remote sensing imagery for physics discovery. *arXiv preprint arXiv:2506.06176*, 2025.

[53] Zhenyu Yu, Mohd Yamani Idna Idris, Hua Wang, Pei Wang, Junyi Chen, and Kun Wang. From physics to foundation models: A review of ai-driven quantitative remote sensing inversion. *arXiv preprint arXiv:2507.09081*, 2025.

[54] Zhenyu Yu, MOHD YAMANI IDNA IDRIS, and Pei Wang. Physics-constrained symbolic regression from imagery. In *2nd AI for Math Workshop@ ICML 2025*, 2025.

[55] Zhenyu Yu, Mohd Yamani Inda Idris, and Pei Wang. Satellitemaker: A diffusion-based framework for terrain-aware remote sensing image reconstruction. *arXiv preprint arXiv:2504.12112*, 2025.

[56] Zhenyu Yu, Jinnian Wang, Hanqing Chen, and Mohd Yamani Idna Idris. Qrs-trs: Style transfer-based image-to-image translation for carbon stock estimation in quantitative remote sensing. *IEEE Access*, 2025.

[57] Zhenyu Yu, Jinnian Wang, and Mohd Yamani Idna Idris. Iidm: Improved implicit diffusion model with knowledge distillation to estimate the spatial distribution density of carbon stock in remote sensing imagery. *arXiv preprint arXiv:2411.17973*, 2024.

[58] Shuang Zeng, Xinyuan Chang, Mengwei Xie, Xinran Liu, Yifan Bai, Zheng Pan, Mu Xu, and Xing Wei. Futuresightdrive: Thinking visually with spatio-temporal cot for autonomous driving. *arXiv preprint arXiv:2505.17685*, 2025.

[59] Shuang Zeng, Dekang Qi, Xinyuan Chang, Feng Xiong, Shichao Xie, Xiaolong Wu, Shiyi Liang, Mu Xu, and Xing Wei. Janusvln: Decoupling semantics and spatiality with dual implicit memory for vision-language navigation. *arXiv preprint arXiv:2509.22548*, 2025.

[60] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023.

[61] Wendong Zhang, Yunbo Wang, Bingbing Ni, and Xiaokang Yang. Fully context-aware image inpainting with a learned semantic pyramid. *Pattern Recognition*, 143:109741, 2023.

[62] Zhenjun Zhao. Balf: Simple and efficient blur aware local feature detector. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3362–3372, 2024.

[63] Zhenjun Zhao and Ben M Chen. Benchmark for evaluating initialization of visual-inertial odometry. In *2023 42nd Chinese Control Conference (CCC)*, pages 3935–3940. IEEE, 2023.

[64] Yaozong Zheng, Bineng Zhong, Qihua Liang, Zhiyi Mo, Shengping Zhang, and Xianxian Li. Odtrack: Online dense temporal token learning for visual tracking. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 7588–7596, 2024.

[65] Yaozong Zheng, Bineng Zhong, Qihua Liang, Shengping Zhang, Guorong Li, Xianxian Li, and Rongrong Ji. Towards universal modal tracking with online dense temporal token learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

[66] Eric Zhou and Dokyun Lee. Generative artificial intelligence, human creativity, and art. *PNAS nexus*, 3(3):pgae052, 2024.

[67] Xin Zhou, Yang Wang, Daoyu Lin, Zehao Cao, Biqing Li, and Junyi Liu. Satelliterf: Accelerating 3d reconstruction in multi-view satellite images with efficient neural radiance fields. *Applied Sciences*, 14(7):2729, 2024.