International Cartographic Association
Association Cartographique Internationale

Land-cover Segmentation from Low-resolution Historical Labels via Diffusion Models

Yue Wu [a,b], Ling Zhang [a,b,*]

[a] *Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing 210023, China, Yue Wu - 221301030@njnu.edu.cn, Ling Zhang - lingzhang@njnu.edu.cn*
[b] *State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing, 210023, China*

* Corresponding author

Keywords: Land-cover; Semantic segmentation; Diffusion model; Remote sensing

Abstract:

Building on generic segmentation tasks, land-cover segmentation aims to generate accurate and reproducible results based on updated RS images as sensors and satellites developed. High-resolution (HR) images provide rich details of the land surface, allowing us to observe the change of global environment more clearly (Xinhao et al., 2024). Nevertheless, the unique up-to-date characteristics of RS images remain relatively under-explored when it comes to generating high-fidelity and diverse land-cover segmentation results (Zhiwei et al., 2024).

With the advancement of deep learning techniques, several HR land-cover mapping studies have successfully applied classical convolutional neural networks (CNNs) (Enzo et al., 2024) and popular vision transformers (ViTs) (Wen et al., 2023) for segmentation tasks. CNNs capture local information by large or dilated kernels, often leading to a loss of image information, while ViTs struggle with computational efficiency, local feature capture, and noise sensitivity (Krishna et al., 2021). Both rely on HR labels that need extensive manual annotation. Hybrid models like Paraformer and L2HNet combine CNNs and ViTs to capture both global and local information using LR labels for large-scale HR mapping (Zhuo et al., 2024), but their performance is often suboptimal due to incompatibility between the global features from ViTs and local features from CNNs. Recently, Denoising Diffusion Probabilistic Models (DDPMs) have shown promise in generating high-quality images, with many studies reporting state-of-the-art results in remote sensing. However, these methods often use classical UNet backbones, which struggle to capture fine details like small objects and complex edges, leading to reduced precision compared to CNN-based models (Aysim et al., 2024).

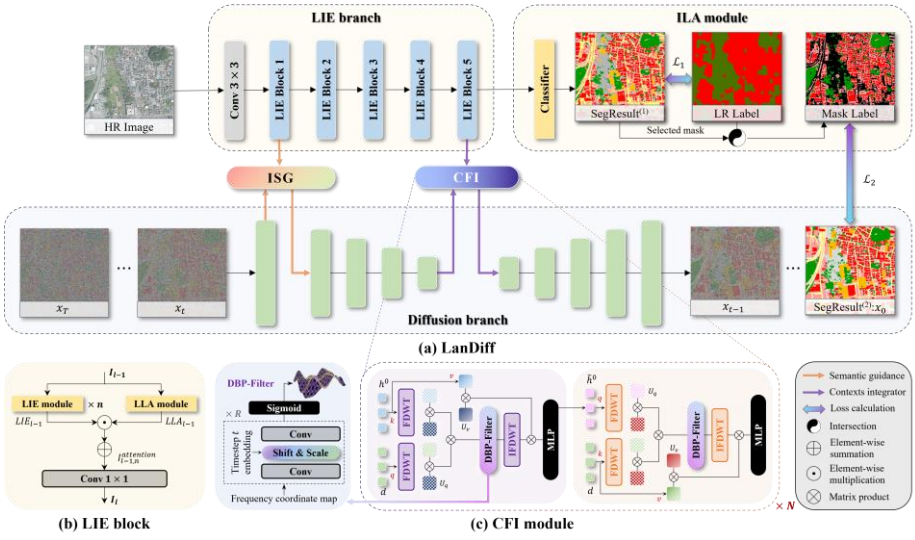

Figure 1. An illustration of LanDiff, starts from (a) an overview of the pipeline, and continues with zoomed-in diagrams of individual parts, including (b) LIE block, and (c) CFI module

To address these challenges, we propose a novel diffusion framework for land-cover segmentation, named LanDiff (Figure 1). The main idea is to parallelly hybridize the information extract branch and diffusion branch through multiple conditional methods, generating high-quality, diverse and accurate HR labels solely from HR images and LR labels.

The diffusion branch is conditioned by segmentation features extracted from HR images through the Land-cover Information Exact (LIE) branch. It utilizes two distinct conditioning methods: Image Semantic Guidance (ISG) and Contextual Features Integrator (CFI). Specifically, ISG imposes HR image guidance features onto the encoded features of the diffusion branch. This allows the diffusion process to be initialized by a rough guidance, which helps to reduce diffusion variances. Meanwhile, CFI integrates the HR image semantic feature embeddings into the embeddings of the diffusion branch. This integration bridges the gap between the noise and semantic embeddings, resulting in a more robust contextual representation. The deepest HR image semantic feature is also input to the ILA module, aiming to screen out uncertain samples and dig up reliable information from the LR labels. We experimentally show that LanDiff achieves state-of-the-art (SOTA) performance in both land-cover realism and accuracy on across two large-scale datasets (Figure 2). Furthermore, ablation study across the other datasets demonstrates the effectiveness of the proposed branches and modules in LanDiff.

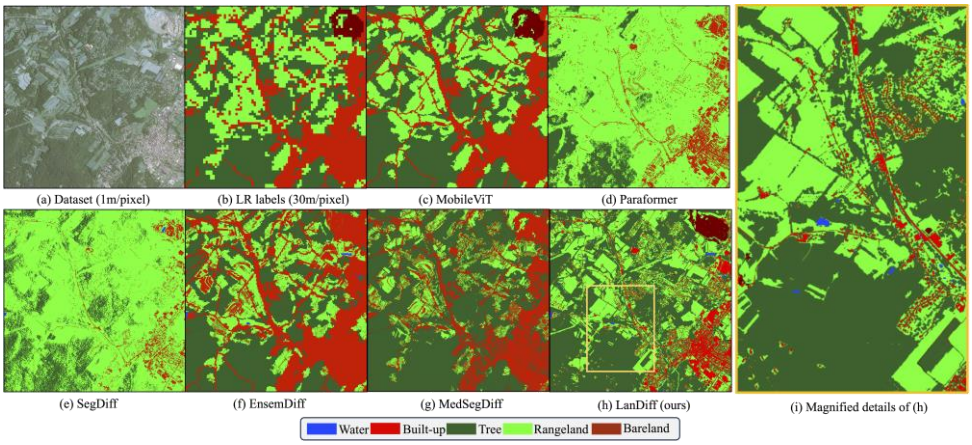

Figure 2. The visual comparison of segmentation models methods on HR image from the Chesapeake bay dataset.

In brief, the contributions of this paper are: (1) We are the first to integrate multi-scale convolution kernels with local attention into a diffusion-based model for generating HR Land-cover map only using HR images and LR labels. (2) We propose ISG and CFI modules to effectively bridge the gap between diffusion and conditional embeddings. (3) The ILA module iteratively intersects primal segmentation and LR labels to refine for guiding part of the framework training. (4) Our generated HR land-cover labels achieve SOTA performance in both image realism and accuracy.

References

Xinhao Cai, Qiuxia Lai, Yuwei Wang, Wenguan Wang, Zeren Sun, and Yazhou Yao. Poly kernel inception network for remote sensing detection. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 27706 – 27716, 2024.

Zhiwei Li, Qihao Weng, Yuhan Zhou, Peng Dou, and Xiaoli Ding. Learning spectral-indices-fused deep models for time series land use and land cover mapping in cloud-prone areas: The case of pearl river delta. Remote Sensing of Environment, 2024.

Enzo Pelous, Nicolas Me´ger, Alexandre Benoit, Abdourrahmane Mahamane Atto, Dino Ienco, Hermann Courteille, and Christophe Lin-Kwong-Chon. Explaining the decisions and the functioning of a convolutional spatiotemporal land-cover classifier with channel attention and redescription mining. ISPRS Journal of Photogrammetry and Remote Sensing, 2024.

Wen Zhou, Claudio Persello, Mengmeng Li, and Alfred Stein. Building use and mixed-use classification with a transformer-based network fusing satellite images and geospatial textual information. Remote Sensing of Environment, 2023.

Krishna Karra, Caitlin Kontgis, Zoe Statman-Weil, Joseph C. Mazzariello, Mark M. Mathis, and Steven P. Brumby. Global land use / land cover with sentinel 2 and deep learning. 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pages 4704 – 4707, 2021.

Zhuo Li, Wei He, Jiepan Li, Fangxiao Lu, and Hongyan Zhang. Learning without exact guidance: Updating large-scale high-resolution land cover maps from low-resolution historical labels. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 27717 – 27727, 2024.

Aysim Toker, Marvin Eisenberger, Daniel Cremers, and Laura Leal-Taix'e. Satsynth: Augmenting image-mask pairs through diffusion models for aerial semantic segmentation. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 27685 – 27695, 2024.

https://doi.org/10.5194/ica-abs-10-304-2025