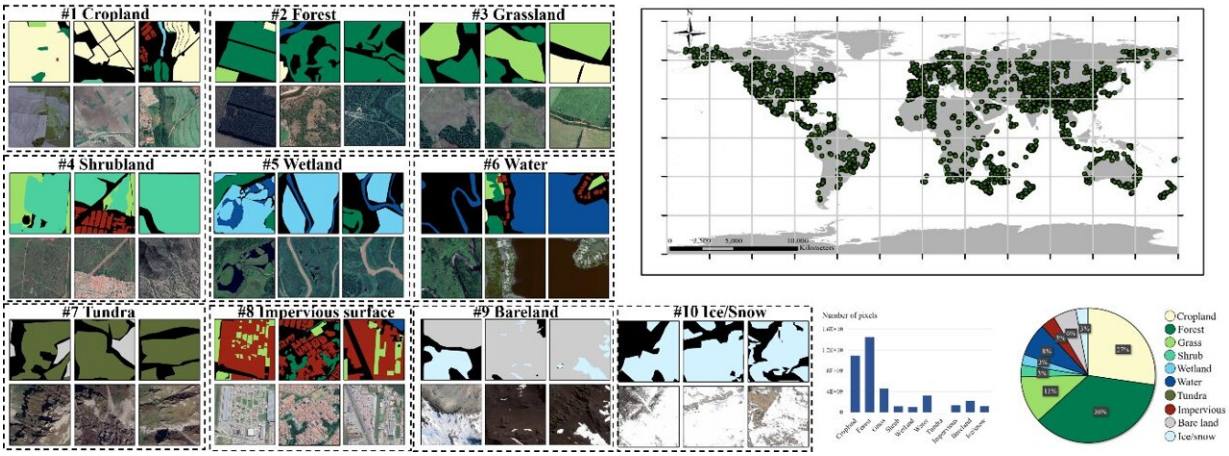
New remote sensing dataset improves global land change tracking

November 22 2023



Scientists from Sun Yat-Sen University developed a large-scale annotated dataset (Globe230k) for high generalized global land cover mapping. The annotated patches provide cues to help classification tools distinguish cropland, forest, wetland, grassland, and more. Credit: [Qian Shi, Sun Yat-Sen University]; [Da He, Sun Yat-Sen University]; [Zhengyu Liu, Sun Yat-Sen University]; [Xiaoping Liu, Sun Yat-Sen University]; [Jingqian Xue, Sun Yat-Sen University]

Tracking unprecedented changes in land use over the past century, global land cover maps provide key insights into the impact of human settlement on the environment. Researchers from Sun Yat-sen University created a large-scale remote sensing annotation dataset to support Earth observation research and provide new insight into the dynamic monitoring of global land cover.

In their study, published in the *Journal of Remote Sensing*, the team examined how global land use/landcover (LULC) has undergone dramatic changes with the advancement of industrialization and urbanization, including deforestation and flooding.

"We urgently need high-frequency, high-resolution monitoring of LULC to mitigate the impact of human activities on the climate and the environment," said Qian Shi, a professor from Sun Yat-sen University.

Global LULC monitoring relies on automatic classification algorithms that classify satellite remote sensing images pixel by pixel. Data-driven deep learning methods extract intrinsic features from the remote sensing images and estimate the LULC label of each pixel.

In recent years, researchers have increasingly employed a method called semantic segmentation for remote sensing image classification tasks in deep learning for global land cover mapping. Instead of classifying images as a whole, semantic segmentation classifies every pixel or element with certain labels.

"Different from recognizing the commercial scene or residential scene in an image, the semantic segmentation network can delineate the boundaries of each land object in the scene and help us understand how land is being used," Shi said.

This sort of high-level semantic understanding cannot be achieved without the context information of each pixel; geographical objects are closely connected to the surrounding scenes, which can provide cues for the prediction of each pixel. For example, airplanes berth in airports, ships dock in harbors, and mangroves generally grow shoreside.

However, the performance of semantic segmentation is limited by the number and quality of training data, and the existing annotation data are

usually insufficient in quantity, quality, and spatial resolution, according to Shi.

To top things off, the datasets are usually sampled regionally and lack diversity and variability, making data-driven models difficult to scale globally.

To address these drawbacks, the research team proposed a large-scale annotation dataset, Globe230k, for semantic segmentation of remote sensing images. The dataset has three advantages:

- Scale—the Globe230k dataset includes 232,819 annotated images with adequate size and a spatial resolution;
- Diversity—the annotated images are sampled from worldwide regions with coverage area of over 60,000 square kilometers, indicating a high variability and diversity;
- Multimodal features—the Globe230k dataset not only contains RGB bands but also other important features for Earth system research such as vegetation, elevation, and polarization indices.

The team tested the Globe230k dataset on several state-of-the-art semantic segmentation algorithms and found that it was able to evaluate algorithms crucial to characterizing land cover, including multiscale modeling, detail reconstruction, and generalization ability.

"We believe that the Globe230k dataset could support further Earth observation research and provide new insights into global land cover dynamic monitoring," Shi said.

The dataset has been made public and can be used as a benchmark to promote further development of global land cover mapping and semantic segmentation algorithm development.

Other contributors include Da He, Zhengyu, Liu, Xiaoping Liu and Jingqian Xue all from Sun Yat-sen University and the Guangdong Provincial Key Laboratory for Urbanization and Geo-simulation.

More information: Qian Shi et al, Globe230k: A Benchmark Dense-Pixel Annotation Dataset for Global Land Cover Mapping, *Journal of Remote Sensing* (2023). DOI: 10.34133/remotesensing.0078