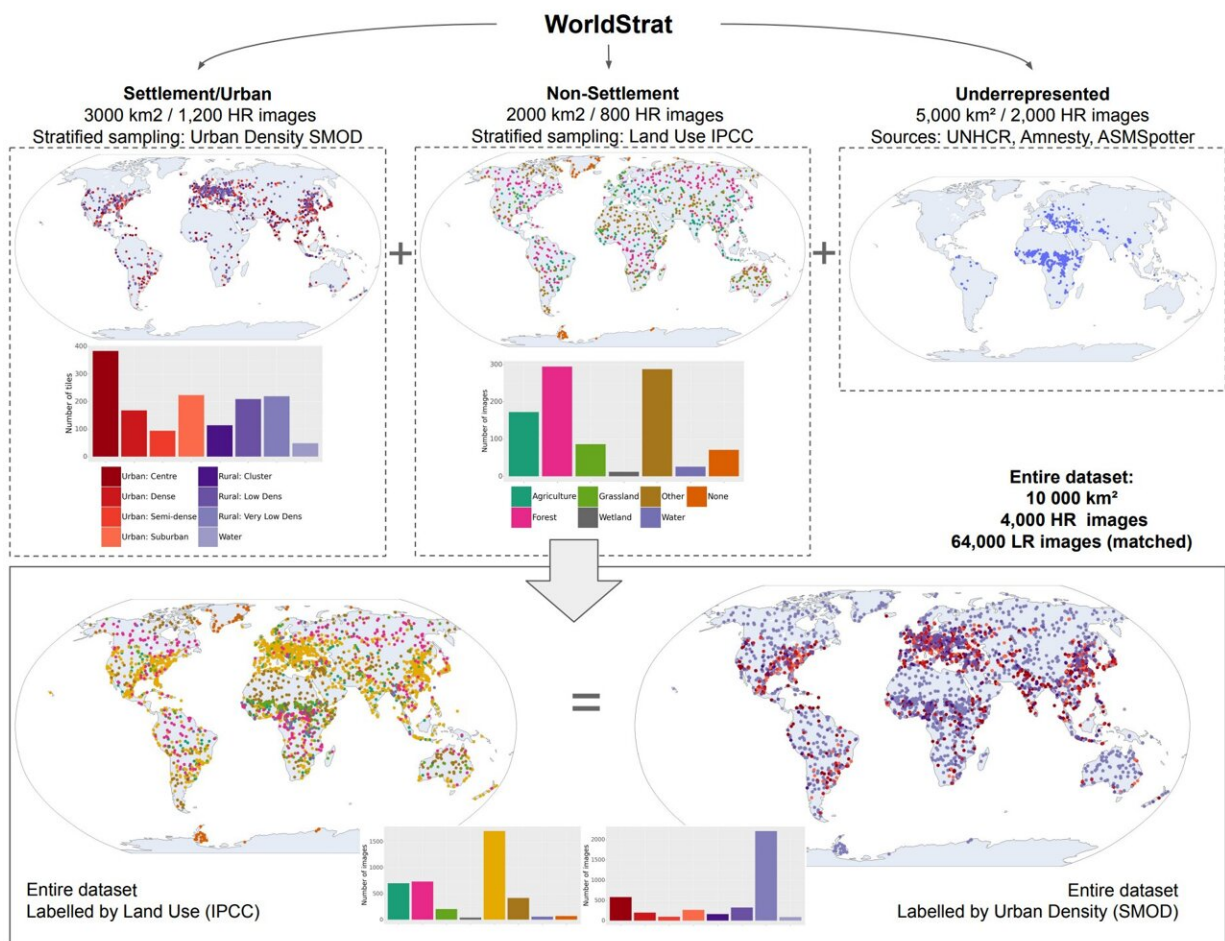


Worldwide dataset captures Earth in finest ever detail

November 18 2022



Summarizing the construction and classes of the WorldStrat dataset. Credit: *arXiv* (2022). DOI: 10.48550/arxiv.2207.06418

A global open-source dataset of high-resolution images of Earth—the most extensive and detailed of its kind—has been developed by experts led by UCL with data from the European Space Agency (ESA).

The free dataset, WorldStrat, will be presented at the NeurIPS 2022 conference in New Orleans. It includes nearly 10,000km² of free [satellite](#) images, showing every type of location, [urban area](#) and [land use](#) from agriculture, grasslands and forests to cities of every size and polar ice caps.

The dataset includes locations in the Global South and those needing [humanitarian aid](#), which are often underrepresented in [satellite imagery](#) because this is usually collected for commercial gain, therefore disproportionately featuring wealthier regions.

The scientists say the collection enables worldwide analysis of terrain to tackle global challenges such as responding to natural and man-made disasters, managing natural resources and urban planning.

Work on WorldStrat began in 2021, and since it launched in June 2022 it has been downloaded over 3,000 times.

Project lead, Dr. Julien Cornebise (UCL Computer Science) said, "The combination of high-resolution commercial imagery and [machine learning](#) has huge potential to enable planetwide analyses, which could help to tackle all kinds of global challenges—the problem is that commercial data are often locked behind a paywall."

"ESA's TPM program made our project possible by providing [free access](#) to data that would normally be very expensive."

The team used data from the Airbus SPOT 6 and SPOT 7 satellites, commissioned by the ESA and launched in 2012 and 2014 respectively.

The satellites can provide imagery at resolutions as high as 1.5m per pixel, meaning that each pixel represents a 1.5m by 1.5m area on the ground.

The scientists used around 4,000 highly detailed images from the SPOT satellites. Even though these images are high (spatial) resolution, they are low in temporal resolution, meaning in this context that each satellite doesn't revisit and recapture each site regularly. This is because images taken by the satellites were originally intended to be used for specific commercial applications rather than longer-term analyses.

To combat this, the team also used freely available, lower resolution images from the Copernicus Sentinel-2 satellite. These are at higher [temporal resolution](#), meaning they were captured at more regular time points every five days. They matched each SPOT image with 16 images from Copernicus Sentinel-2, using around 64,000 in total.

The researchers developed the dataset to also support the development of machine learning applications to extend and enhance it, for example to further improve the image resolution. To allow the development of further applications, the scientists have developed an artificial intelligence toolbox as well as the full source code, enabling developers to reproduce, extend and transform the work.

Dr. Cornebise continued, "Thousands of data users from around the world have already downloaded WorldStrat—and we look forward to seeing the ways in which they extend and improve it, using machine learning techniques."

A pre-print version of the research is available on *arXiv*.

More information: Julien Cornebise et al, Open High-Resolution Satellite Imagery: The WorldStrat Dataset—With Application to Super-

Resolution, *arXiv* (2022). [DOI: 10.48550/arxiv.2207.06418](https://doi.org/10.48550/arxiv.2207.06418)

GitHub dataset: worldstrat.github.io/

Provided by University College London

Citation: Worldwide dataset captures Earth in finest ever detail (2022, November 18) retrieved 29 January 2023 from <https://phys.org/news/2022-11-worldwide-dataset-captures-earth-finest.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.