Building a living map of the world: petabyte-scale real-time multi-sensor image analysis in commercial cloud.

Steven Brumby¹, Michael S Warren¹, Rick Chartrand¹, Eduardo Franco¹, Ryan Keisler¹, Tim Kelton¹, Mark Matthis¹, Daniela Moody¹, Sam Skillman¹

¹Descartes Labs

Multi-decadal Earth and planetary remote sensing global datasets at the petabyte (8×10¹⁵ bits) scale are now available in commercial clouds (e.g., Google Earth Engine and Amazon NASA NEX), and new commercial satellite constellations (e.g., Planet Labs Dove constellation) are planning to generate petabytes of images per year, providing daily global coverage at a few meters per pixel. Cloud storage with adjacent high-bandwidth compute, combined with recent advances in machine learning for computer vision, is enabling understanding of the world at a scale and at a level of space and time granularity never before feasible. This provides the opportunity to build a continuously updated map of the world to support decision making for commerce, industry, and policy makers.

We report here on a computation processing over a petabyte of compressed raw data from 2.8 quadrillion pixels (2.8 petapixels) acquired by the US Landsat over the past 40 years and MODIS programs over the last 15 years. This data and analysis system is sensor agnostic, enabling simultaneous analysis of image data from many sensors. Using commodity cloud computing resources, we convert the imagery to a calibrated, georeferenced, multi-resolution tiled format suited for deep temporal machine-learning analysis. We believe ours is the first application to process, in less than a day, on generally available resources, over a petabyte of scientific image data. We report on work using this reprocessed dataset for experiments demonstrating country-scale and global-scale agricultural forecasting, environmental monitoring and disaster detection. We apply remote sensing and statistical machine learning algorithms to detect and classify agricultural crops and then estimate crop yields (indicator for famine early warning) and detect other threats to food security (e.g., flooding and hail storms). The software platform and analysis methodology also support monitoring water resources and forestry resources that are general indicators of environmental health, and detecting growth and changes in cities that are displacing traditional agricultural zones. We present a live demonstration of the prototype platform running in the Google Cloud.