# Exploring SpaceNet Data with GeoPySpark

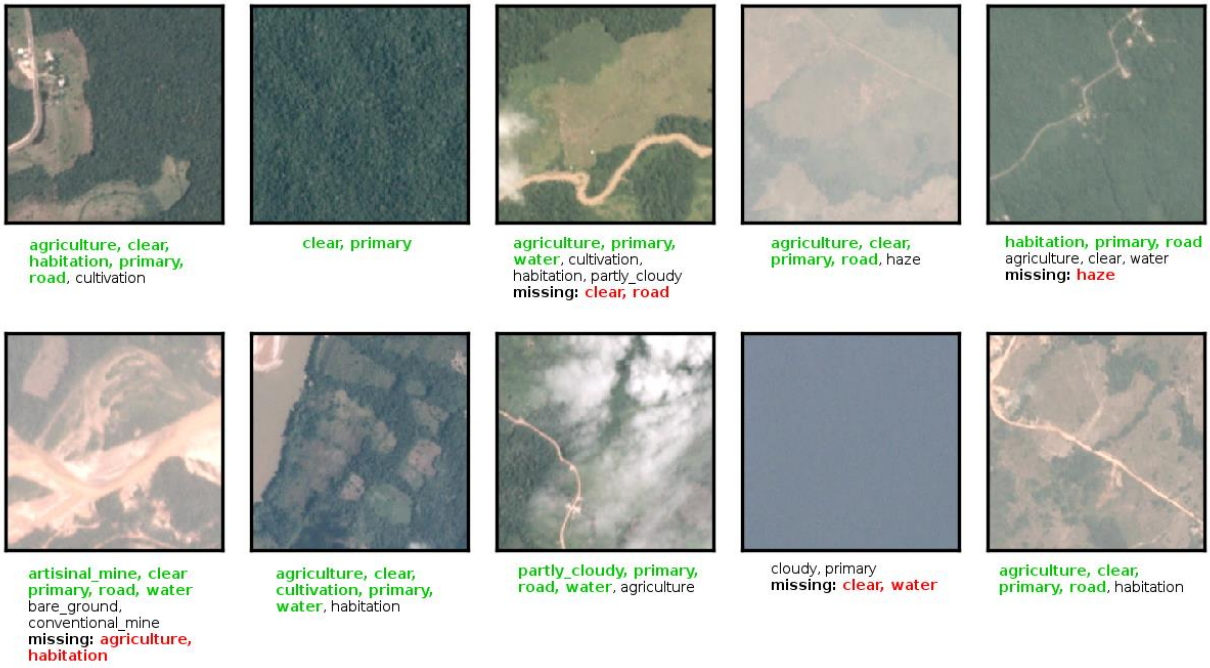
By Yoni Nachmany on August 11, 2017

---

*This post is part of Azavea's inaugural Open Source Software Fellowship Program, which matches software engineering fellows to open source projects at Azavea. To learn more about the Open Source Software Fellowship and subscribe to the mailing list, click here.*

---

On July 20th, Raster Vision, Azavea's system for analyzing aerial and satellite imagery using deep learning that works on a variety of computer vision tasks and datasets, placed 23rd out of 938 teams in a Kaggle competition called "Planet: Understanding the Amazon from Space".




In the above figure, the ground truth tags (ie. tagged by hand) for the Planet Kaggle dataset are bolded. Green bolded tags are correct. Unbolded and uncolored tags mean that they have been incorrectly predicted for the chip. Red bolded tags are ones missed by the network prediction. Photo: Raster Vision.

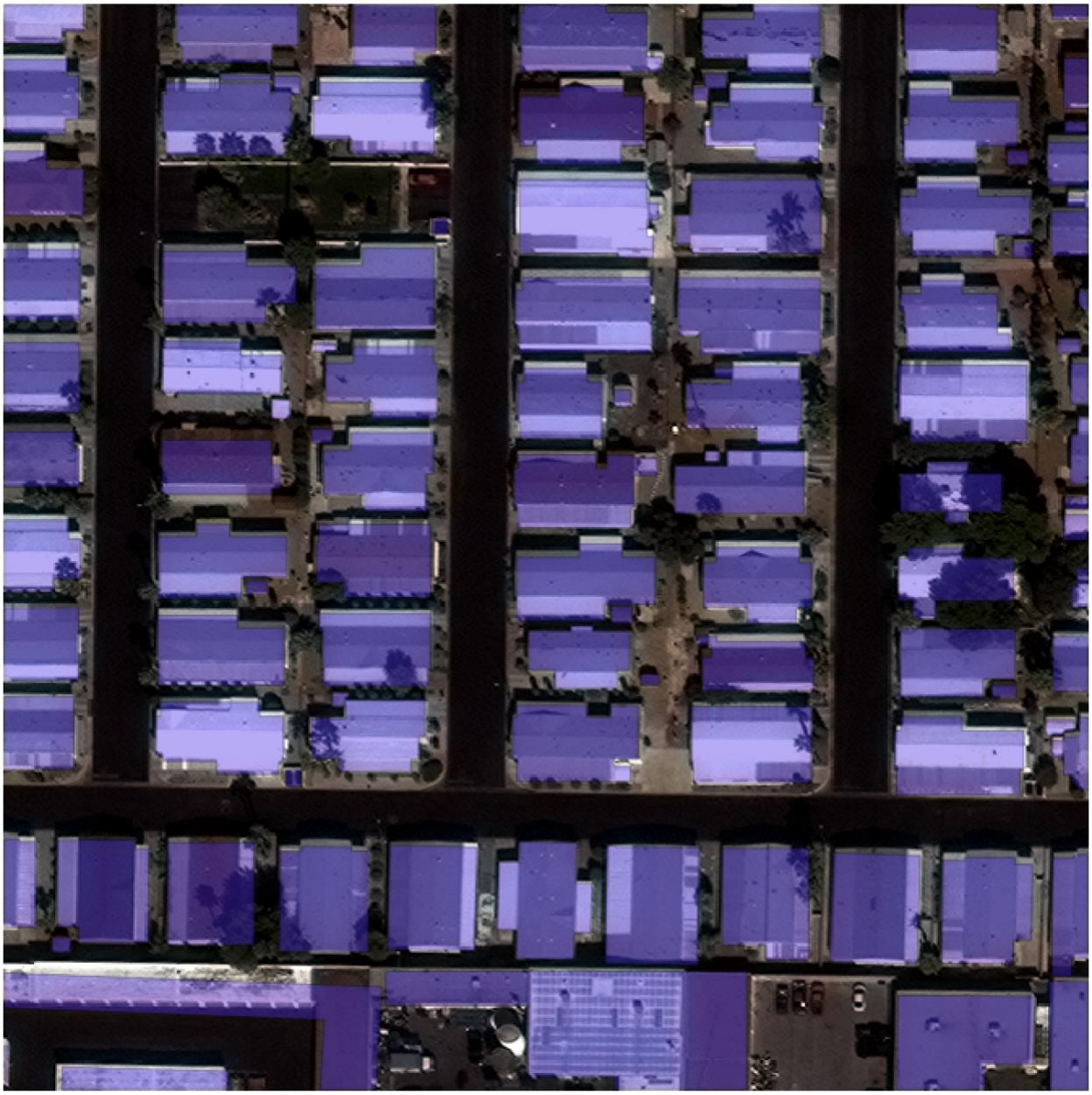
Raster Vision achieved a 93.154% Mean F-Score Beta on the data, which was provided by Planet. Over the last seven years, the steep accuracy increases of GPU-accelerated deep learning techniques for computer vision tasks have been driven largely by the ImageNet dataset, which contains 14 million photos labeled in over 20,000 categories.

## SpaceNet, the ImageNet of Satellite Imagery

Last August, DigitalGlobe, CosmiQ Works, NVIDIA, and Amazon Web Services (AWS) teamed up to launch SpaceNet, the ImageNet of high-resolution multi-spectral satellite imagery. In the announcement, Tony Frazier, the Senior Vice President at DigitalGlobe, said "SpaceNet is key to unlocking a huge explosion of new AI-driven applications that ultimately will help us better respond to natural disasters, counter global security threats, improve population health outcomes, and much more."

SpaceNet Rio Area of Interest (AOI). Photo: DigitalGlobe.

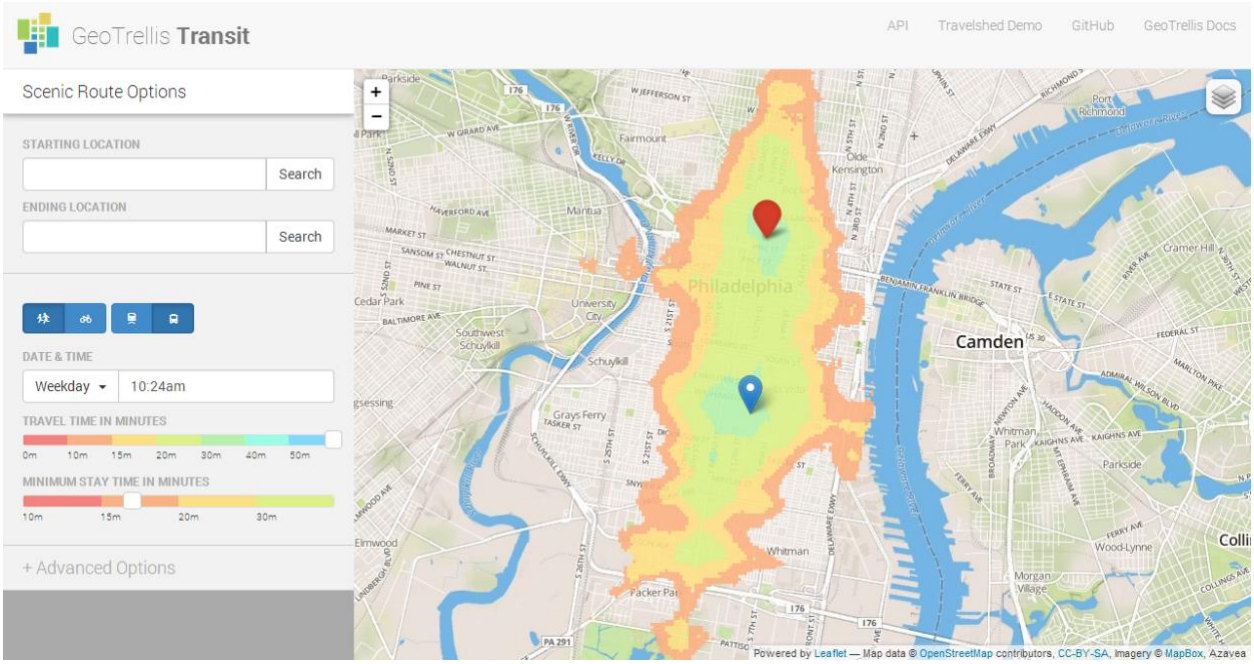
Initially, SpaceNet included 50cm 8-band imagery of Rio de Janeiro, collected from DigitalGlobe's WorldView-2 satellite, along with 200,000 associated building footprints. In the last year, SpaceNet has released a Points of Interest (POI) dataset for Rio, the winning implementations of Rio building footprint extraction from The SpaceNet Challenge, and a second round of The SpaceNet Challenge with improved training data, including the addition of four geographically diverse cities: Vegas, Paris, Shanghai, and Khartoum. Eventually, SpaceNet will have over 60 million labeled images, four times the current number of images in ImageNet.



30cm WorldView-3 imagery and building footprints in Las Vegas, Nevada. Photo: DigitalGlobe.

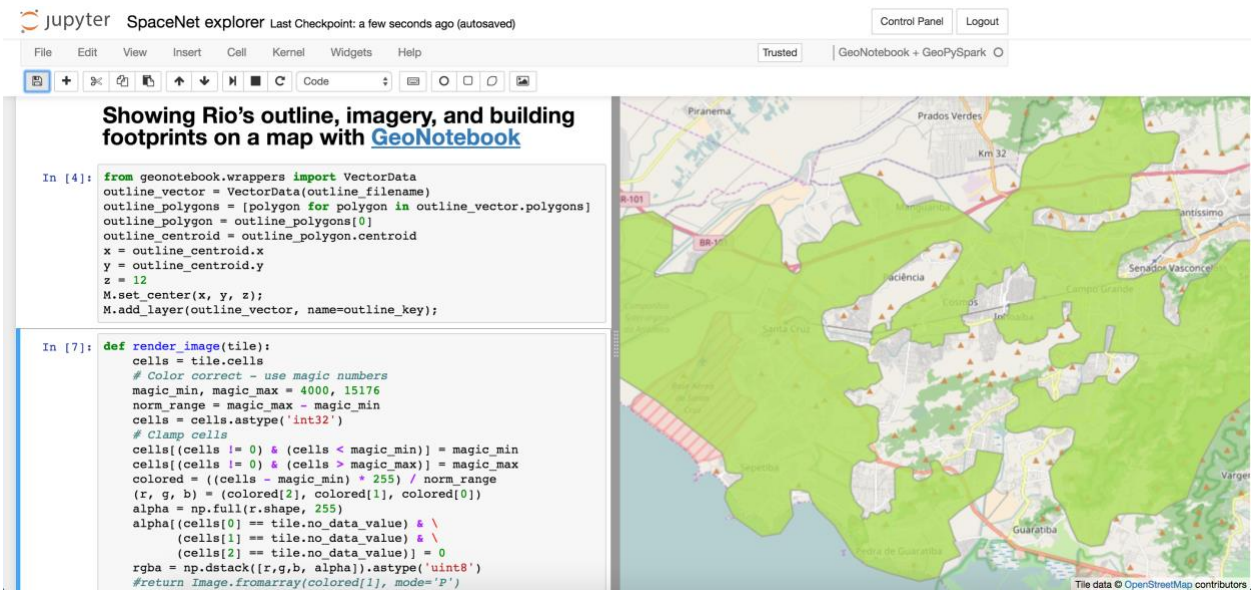
## Exploring SpaceNet data in a Jupyter notebook with GeoPySpark

To start exploring datasets, stacks of ready-to-run Jupyter applications in Docker provide isolated, functional, and easy-to-use environments. However, processing big raster data, such as SpaceNet satellite imagery, requires different tools, like GeoTrellis, a geographic data processing engine developed by Azavea. GeoPySpark, the Python binding of GeoTrellis, bridges the raster capabilities of GeoTrellis and the data science capabilities of Docker and Jupyter.




Scenic Route Demo on GeoTrellis Transit. Photo: Azavea.

Using a GeoDocker container with Jupyter and GeoPySpark, a single notebook can access the SpaceNet data on Amazon Web Services, download Rio raster and vector data with Boto, wrangle imagery with GDAL, ingest imagery for fast viewing with GeoPySpark, and show Rio's outline, imagery, and building footprints on a map with GeoNotebook.




## Next Steps

SpaceNet provides high-volume, high-veracity, and high-variety satellite data with high value potential. Raster Vision, the backbone of deep learning at Azavea, is continually adding new

datasets and computer vision capabilities over those datasets. Integrating SpaceNet into Raster Vision has the potential to benefit downstream Azavea applications like Raster Foundry and open up new use cases to help people and organizations make informed decisions.