Richard Correro
Email | Website | GitHub | LinkedIn

Education

Stanford University — Stanford, CA
MS in Statistics — *June 2022*
Stanford University — Stanford, CA
BS in Mathematical and Computational Science — *June 2021*

Experience

Technical Program Manager, Data Engineering — July 2022 – Present
Stanford University School of Medicine — *Stanford, CA*

- Developed a containerized, distributed, cloud-native data pipeline to process **hundreds of millions of hectares of high-resolution satellite imagery daily using deep-learning-based computer vision algorithms** (Docker, Kubernetes, Redis, RabbitMQ, Google Cloud Platform, AWS).
- Developed novel computer vision algorithms for image classification, image segmentation, and object detection.
- **Applied mathematical and statistical theories, techniques, and methods**, to perform statistical inference on massive data sets produced using the pipeline.
- System is being used by the Brazilian Federal Labor Prosecution Office to target inspections to identify and prevent **modern slavery** and **illegal deforestation in the Amazon rainforest**.
- Article about my work here.

Graduate Research Assistant — June 2021 – June 2022
Stanford Human Trafficking Data Lab — *Stanford, CA*

- **Designed, trained, and deployed** computer vision algorithms to identify remote commodity production sites using satellite imagery (PyTorch, GDAL).
- Papers forthcoming.

Research Assistant — June 2020 – June 2021
Stanford Center for Ocean Solutions — *Stanford, CA*

- Created a deep-learning-based computer vision algorithm to identify small fishing vessels in satellite imagery (PyTorch, GDAL, OpenCV).
- Analyzed entire near-shore region of the Peruvian EEZ and identified previously unknown locations where **illegal, unreported, or unregulated fishing** was occurring (Google Cloud Platform, Statsmodels, R).
- Code available here.
- Article about my work here.
- Paper accepted for publication (forthcoming).

Projects

Light-Pipe | *Python, C++*

- Extensible, light-weight, open-source Python framework for **data pipelines that scale.**
- Provides a set of intuitive abstractions designed to decouple pipeline implementation from the operations they perform.
- Scales effortlessly, being built from the ground-up to support concurrency in all its forms.
- Super fast and efficient, **used to perform critical geospatial data processing tasks at least an order of magnitude faster than existing systems.**
- Talk I presented about Light-Pipe at **Google's Geo for Good Summit in Mountain View available here.**
- Code available here.

"Weak Supervision with Incremental Source Accuracy Estimation"

- **Developed an algorithm** to estimate the dependency structure and accuracy of weak supervision sources incrementally using precision matrices and robust principal components analysis.
- Allows for model training with weakly-supervised training data in on-line settings.
- Preprint available here.
- Code available here.

Skills

Languages: Python, C++, SQL, R, BASH

Tools: Git, Docker, Kubernetes, Apache Beam, Apache Spark, PostgreSQL, PostGIS, Rabbit MQ, Redis, Google Cloud Platform, Amazon Web Services (AWS), RESTful APIs, QGIS

Libraries: GDAL, Rasterio, PyTorch, Tensorflow, Scikit-Learn, OpenCV, Statsmodels, NumPy, Pandas, Flask, Celery, PyTest

Awards

National Merit Scholar — April 2017