**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

**Graduate Research Assistant** — June 2021 – Present
*Stanford Human Trafficking Data Lab* — *Stanford, CA*
- Developed a containerized, distributed, cloud-native data pipeline to process hundreds of thousands of square kilometers of satellite imagery daily (Docker, Kubernetes, Redis, RabbitMQ, Google Cloud Platform).
- Developed a computer vision algorithm to identify remote commodity production sites using satellite imagery (PyTorch, GDAL).
- System will be used by law enforcement to proactively identify locations where forced labor is being used.
- Article about our work here.
- 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 forthcoming.

**Mathematics Tutor** — September 2018 – June 2019
*Stanford University Mathematical Organization* — *Stanford, CA*
- Tutored Stanford students taking classes in linear algebra, vector calculus, and differential equations.

## Projects

**Light-Pipe** | *Python, Flask, GDAL, PyTorch, Tensorflow, Celery*
- A Python framework designed to facilitate the deployment of geospatial data pipelines at scale.
- Provides platform-agnostic API for efficient ETL with cloud-optimized geotiffs.
- Allows for model training with heterogeneous geospatial data sources and formats.
- Handles task scheduling and queuing to facilitate distributed processing of geospatial data.
- Facilitates data ingest from major data providers.

**"Weak Supervision with Incremental Source Accuracy Estimation"**
- Developed a method 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.

**"Effect of Behavioral Factors on Average Heart Rate During Cardiovascular Exercise"**
- Designed a resolution IV fractional factorial experiment to determine whether certain behavioral factors affect average heart rate during cardiovascular exercise.
- Found that two of the four factors tested were associated with a significant effect on average heart rate during exercise.
- Paper, data, and code available here.

## Skills

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

**Tools:** Git, Docker, Kubernetes, Apache Airflow, Apache Spark, Argo, PostgreSQL, PostGIS, Rabbit MQ, Redis, Google Cloud Platform, RESTful APIs

**Libraries:** Flask, Celery, GDAL, Rasterio, Pytorch, Tensorflow, Scikit-Learn, OpenCV, Statsmodels, Numpy, Pandas

## Awards

**National Merit Scholar** — April 2017