

EDUCATION

Stanford University <i>MS in Statistics</i>	Stanford, CA <i>June 2022</i>
Stanford University <i>BS in Mathematical and Computational Science</i>	Stanford, CA <i>June 2021</i>

EXPERIENCE

Technical Program Manager, Data Engineering <i>Stanford University School of Medicine</i>	July 2022 – Present <i>Stanford, CA</i>
<ul style="list-style-type: none">Developed a containerized, distributed, cloud-native data pipeline to process hundreds of millions of hectares of high-resolution satellite imagery daily (Docker, Kubernetes, Redis, RabbitMQ, Google Cloud Platform, AWS).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 <i>Stanford Human Trafficking Data Lab</i>	June 2021 – June 2022 <i>Stanford, CA</i>
<ul style="list-style-type: none">Designed, trained, and deployed computer vision algorithms to identify remote commodity production sites using satellite imagery (PyTorch, GDAL).Papers forthcoming.	
Research Assistant <i>Stanford Center for Ocean Solutions</i>	June 2020 – June 2021 <i>Stanford, CA</i>
<ul style="list-style-type: none">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 <i>Python, C++</i>
<ul style="list-style-type: none">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"
<ul style="list-style-type: none">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
-------------------------------	------------