Joshua Fan

Email: jyf6@cornell.edu – **Website**: http://joshuafan.github.io

Summary: I develop novel AI techniques to tackle real-world challenges in sustainability and agriculture. I am interested in representation learning for spatio-temporal data (including satellite imagery), contrastive learning, weakly-supervised learning with few labels, and integrating scientific knowledge with data-driven learning.

EDUCATION

Ph.D. Computer Science, Cornell University, *Ithaca, NY, USA (in progress)* 2019 – present
- Advisor: Prof. Carla Gomes
- GPA: 4.11/4.3

M.S. Computer Science, University of Washington, *Seattle, WA, USA* 2017 – 2019
- Advisor: Prof. Sreeram Kannan
- GPA: 3.84/4.0

B.S. Computer Science, University of Washington, *Seattle, WA, USA* 2013 – 2017
- GPA: 3.97/4.0 *(summa cum laude)*

RESEARCH EXPERIENCE

Research Assistant, Cornell University

Institute for Computational Sustainability *(advised by Prof. Carla Gomes)* **Aug. 2019 – present**
- Developing techniques to integrate machine learning with scientific (process-based) models, and impose smoothness constraints to improve generalization (applications in carbon cycle modeling)
- Designed novel contrastive learning methods for remote sensing imagery (including augmentations to emphasize shape and remove background information), with applications in fish pond detection
- Created an interpretable spatiotemporal deep learning framework for crop yield forecasting from weather/soil data, achieving accuracy comparable to USDA forecasts. Working on hypernetworks to handle spatial heterogeneity. Designing an improved Transformer architecture for continuous time series.
- Developed Coarsely Supervised Smooth U-Net, a weakly-supervised deep learning method. Predicts SIF (a proxy for vegetation productivity) at a fine spatial resolution (30m) from remote sensing images, even though training labels are only available at a very coarse resolution (3km).

Research Assistant, University of Washington

Information Theory Lab *(advised by Prof. Sreeram Kannan)* **Mar. 2017 – Jun. 2018**
- Developed scalable algorithms inspired by Latent Dirichlet Allocation and matrix factorization to discover cell types and find structure in large single-cell RNA-seq datasets (over 1 million cells) (Poster, Paper, Code)

Computing for Development Lab *(advised by Prof. Richard Anderson)* **Mar. 2015 – Jun. 2016**
- Redesigned a survey app which helps health workers collect patient data and suggest follow-up actions

PEER-REVIEWED PUBLICATIONS *(* denotes equal contribution)*

Monitoring Vegetation from Space at Extremely Fine Resolutions via Coarsely-Supervised Smooth U-Net.
Joshua Fan, Di Chen, Jiaming Wen, Ying Sun, Carla Gomes.
IJCAI-22: *International Joint Conference on Artificial Intelligence* (AI for Good track), 2022.
- *Also presented at the "Tackling Climate Change with Machine Learning" workshop at NeurIPS 2021*

A GNN-RNN Approach for Harnessing Geospatial and Temporal Information: Application to Crop Yield Prediction.
Joshua Fan*, Junwen Bai*, Zhiyun Li*, Ariel Ortiz-Bobea, Carla Gomes.
AAAI-22: *AAAI Conference on Artificial Intelligence (AI for Social Impact track),* 2022.
- *A workshop version received **best paper award (ML Innovation)** at "Tackling Climate Change with Machine Learning" workshop at NeurIPS 2021*

Scalable preprocessing for sparse scRNA-seq data exploiting prior knowledge.
Sumit Mukherjee, Yue Zhang, **Joshua Fan**, Georg Seelig, Sreeram Kannan.
Bioinformatics, 2018

WORKSHOP PAPERS

Detecting Aquaculture with Deep Learning in a Low-Data Setting. *Fragile Earth workshop at KDD 2023.*
L. Greenstreet, **J. Fan,** F. Pacheco, Y. Bai, M. Ummus, C. Doria, N. Barros, B. Forsberg, X. Xu, A. Flecker, C. Gomes.

INVITED TALKS

Near Real-Time Crop Yield Forecasting with Interpretable Time-Series Deep Learning.
The Workshop on Environmental Economics and Data Science (TWEEDS), Eugene, OR, USA. Oct 2022.

Using Deep Learning to Monitor and Forecast Vegetation Growth.
Soil and Crop Sciences Seminar, Cornell University, Ithaca, NY, USA. Feb 2022.

Predicting Iron Bioavailability in Yellow Beans with Hyperspectral Imaging and Machine Learning.
Yellow Bean Conference, Online. Dec 2020.

INDUSTRY EXPERIENCE

NLP Research Intern, Docugami *(AI Document Engineering startup)* **Summer 2018, Summer 2019**
- Researched and implemented state-of-the-art NLP algorithms (including topic modeling, clustering, and question-answering techniques), and adapted them in novel ways for enterprise document analysis

Software Engineer Intern, Facebook *(Integrity Computer Vision Team)* **Fall 2018**
- Trained a clip-based convolutional neural network to detect graphic and violent content in videos
- Created new datasets and achieved higher accuracy for violence detection than previous approaches

Software Engineer Intern, Facebook *(Search, Whole Page Ranking Team)* **Fall 2017**
- Trained a sequence classification neural network to predict which search result module the user will click on, based on recent query history; improved quality of search ranking and click rate

Software Engineer Intern, Facebook *(Search Indexing Team)* **Summer 2016**
- Built a web tool to help engineers debug and simulate changes to the search indexing pipeline

Software Design Engineer Intern, BitTitan **Summer 2015**
- Built infrastructure to allow mailbox migrations to be tested in memory; optimized SQL queries

TEACHING EXPERIENCE

- **TA, Application of Machine Learning to Plant Science** (Cornell University, PLSCI 7202): Fall 2022
- **TA, Intro to Artificial Intelligence** (Cornell University, CS 4700): Fall 2019
- **TA, Probability & Statistics** (UW, CSE 312): Fall 2015, Winter 2016, Spring 2017, Winter 2018, Winter 2019
- **TA, Foundations of Computing I/Discrete Math** (Univ of Washington, CSE 311): Fall 2016, Spring 2018
- **TA, Intro to Machine Learning for Non-Majors** (Univ of Washington, CSE 416): Spring 2019

ADVISING

Kaitlyn Chen (undergraduate, BS Cornell 2023)
- Topic: Improved Transformer architecture with inductive biases for continuous time-series regression.

Saraswathy Amjith (high school, project won 1st place in Earth & Envir. Sciences at WA State Science & Engineering Fair)
- Topic: Detecting deforestation using deep learning with radar and optical satellite imagery

SERVICE

Reviewer for:
- *AAAI* (AI for Social Impact track), 2024
- *NeurIPS Computational Sustainability workshop*, 2023
- *Environmental Research Letters* (2 journal papers), 2022-2023

Organizer for Cornell Computer Science Visit Day, 2020

TECHNICAL SKILLS

- **Significant experience:** Python, Java, C#, SQL, C++, PHP/Hack
- **Working knowledge:** R, Matlab, Julia, HTML/CSS, JavaScript, JQuery
- **Libraries/tools:** PyTorch, Pandas, Matplotlib, Eclipse, Git, Visual Studio, Linux, Nuclide

HONORS & AWARDS

- **National Science Foundation Research Training (NRT) Fellowship** in Digital Plant Science, 2022-2025
- **Bob Bandes Memorial Excellence in Teaching Award**, University of Washington, 2019 (one of 3 winners)