Joshua Fan

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

EDUCATION

Ph.D. Computer Science, Cornell University, *Ithaca, NY, USA (in progress)* **Aug. 2019 – present**
- Advisor: Prof. Carla Gomes
- Topics: Machine learning with coarse-scale data, spatio-temporal data, applications to agriculture/sustainability

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

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

PEER-REVIEWED PUBLICATIONS

Joshua Fan, Di Chen, Jiaming Wen, Ying Sun, Carla Gomes. "Monitoring Vegetation from Space at Extremely Fine Resolutions via Coarsely-Supervised Smooth U-Net." *IJCAI* 2022, *AI for Good* track.

Joshua Fan*, Junwen Bai*, Zhiyun Li*, Ariel Ortiz-Bobea, Carla Gomes. "A GNN-RNN Approach for Harnessing Geospatial and Temporal Information: Application to Crop Yield Prediction." *AAAI* 2022, *AI for Social Impact* track.

Sumit Mukherjee, Yue Zhang, **Joshua Fan**, Georg Seelig, and Sreeram Kannan. "Scalable preprocessing for sparse scRNA-seq data exploiting prior knowledge." *Bioinformatics*, 34, 2018, i124–i132.

WORKSHOPS AND TALKS

Joshua Fan. "Using Deep Learning to Monitor and Forecast Vegetation Growth." Soil and Crop Sciences Seminar, Cornell University.

Joshua Fan, Di Chen, Jiaming Wen, Ying Sun, Carla Gomes. "Super Fine-Resolution SIF via Coarsely-Supervised U-Net Regression." Tackling Climate Change with Machine Learning workshop at NeurIPS 2021.

Joshua Fan*, Junwen Bai*, Zhiyun Li*, Ariel Ortiz-Bobea, Carla Gomes. "A GNN-RNN Approach for Harnessing Geospatial and Temporal Information: Application to Crop Yield Prediction." Tackling Climate Change with Machine Learning workshop at NeurIPS 2021. **Received best paper award (ML Innovation).**

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

RESEARCH EXPERIENCE

Research Assistant, Cornell Institute for Computational Sustainability **Aug. 2019 – present**
(supervised by Prof. Carla Gomes)
- Designed an innovative deep learning framework involving graph, convolutional, and recurrent neural networks to harness spatiotemporal structure, for crop yield forecasting from weather/soil data
- Designed coarsely-supervised deep learning techniques for predicting SIF (solar-induced chlorophyll fluorescence) at a fine spatial resolution from remote sensing images, given noisy and coarse-resolution labels
- Adapted self-supervised contrastive representation learning techniques to predict phenotypes (iron bioavailability, cooking time) from hyperspectral images of beans

Research Assistant, UW Information Theory Lab **Mar. 2017 – Jun. 2018**
(supervised by Prof. Sreeram Kannan)
- 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)

- Used Siamese Seq2Seq neural networks and deep reinforcement learning to learn an edit embedding (which approximates Levenshtein distance between strings) (Poster, Report)

Research Assistant, UW Computing for Development Lab **Mar. 2015 – Jun. 2016**
(supervised by Prof. Richard Anderson)
- Helped redesign a survey app which helps public health workers collect data according to medical protocol
- Collaborated with PATH (global health company) and field-testers to improve interactions and user experience

INDUSTRY EXPERIENCE

NLP Research Intern, Docugami **Jun. – Sep. 2018, Jun. – Aug. 2019 (full-time)**
(AI Document Engineering startup) ***Mar. – Jun. 2018, Jan. – Jun. 2019 (part-time)***
- Researched and implemented state-of-the-art NLP algorithms (including topic models, embeddings, clustering, and question-answering techniques), and applied them in novel ways for enterprise document analysis

Software Engineer Intern, Facebook *(Integrity Computer Vision Team)* **Sep. – Dec. 2018**
- Trained and evaluated 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)* **Sep. – Dec. 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)* **Jun. – Aug. 2016**
- Built a web tool to help engineers debug and test changes to the search indexing pipeline
- Created a back-end C++ Thrift service to query data stores and generate expected indexing output

Software Design Engineer Intern, BitTitan **Jun. – Sep. 2015**
- Implemented infrastructure in C# to allow mailbox migrations to be simulated and tested in memory
- Improved speed of a key method by around 60% by optimizing SQL queries and calculations

TEACHING EXPERIENCE

Teaching Assistant, Cornell University
- Course: **Introduction to Artificial Intelligence** (CS 4700): Fall 2019

Teaching Assistant, University of Washington
- Courses:
  - **Probability & Statistics** (CSE 312): Fall 2015, Winter 2016, Spring 2017, Winter 2018, Winter 2019
  - **Foundations of Computing I/Discrete Math** (CSE 311): Fall 2016, Spring 2018
  - **Introduction to Machine Learning for Non-Majors** (CSE 416): Spring 2019
- In addition to teaching sections, holding office hours, and grading, I took initiative to create additional resources and host extra review sessions to clarify concepts

ADDITIONAL PROJECTS *(more info at http://joshuafan.github.io/Projects.html)*

Storage and Retrieval of Robotic Laser Range Data (Poster, Report) ***(Course: Graduate Databases)***
- Implemented database algorithms for laser-range scans to allow for efficient content-based retrieval of images

Contextual Bandits Notes (Notes) ***(Course: Online and Adaptive Machine Learning)***
- Surveyed recent research on contextual bandits and created a report synthesizing important results/algorithms

LANGUAGES AND TECHNOLOGIES

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

HONORS & AWARDS

- National Science Foundation Research Training (NRT) Fellowship, area of Digital Plant Science, 2021
- **Bob Bandes Memorial Excellence in Teaching Award**, University of Washington, 2019 (one of 3 winners)