

Brian Matejek

COMPUTER SCIENCE · ALGORITHMS · MACHINE LEARNING · COMPRESSION · DATA SCIENCE

☎ (+1) 609-751-4830 | ✉ bmatejek@gmail.com | 🏠 brianmatejek.com | 📷 bmatejek | 💻 brian-matejek | 🐦 @brianmatejek

Summary

I am a research scientist interested in algorithms, machine learning, compression, and data science. I received my Ph.D. from Harvard University, working under the guidance of Hanspeter Pfister in the Visual Computing Group. My doctoral research concerned biologically-aware algorithms for connectomics, the interdisciplinary field between computer science and neuroscience. Specifically, I focused on the compression, error correction, and volume processing of the terabyte label volumes and graph analysis on the neuronal wiring diagrams.

Education

Harvard University

Cambridge, MA 02138

PH.D. IN COMPUTER SCIENCE

Aug 2016 - May 2021

- Thesis: Biologically-Aware Algorithms for Connectomics
- Advisor: Hanspeter Pfister
- Committee Members: Michael Mitzenmacher, Todd Zickler
- GPA: 4.00 / 4.00

Princeton University

Princeton, NJ 08544

M.S.E. IN COMPUTER SCIENCE

Sep 2014 - Jun 2016

- Thesis: Learning Global Features for Neuron Reconstruction in EM Images
- Advisor: Thomas Funkhouser
- GPA: 3.85 / 4.00

B.S.E. IN COMPUTER SCIENCE

Sep 2010 - Jun 2014

- Independent Research: Detecting Objects Using Google Street View Data
- Independent Research: A Computational Analysis of Arbitrage Opportunities in Sports Gambling
- GPA: 3.79 / 4.00, High Honors

Research

Computer Vision

- Introduced a new step to the typical connectomics workflow to identify and correct errors produced by state-of-the-art automatic segmentation methods.
- Devised a novel method that uses biological shape priors to identify potential errors in an input segmentation. Significantly reduced the error search space to increase throughput by $9\text{-}10\times$ over other automated error correction methods.
- Trained a 3D convolutional neural network to learn neuronal shapes and predict if two segments belonged to the same neuron, achieving accuracies of 93-97%.
- Improved on state-of-the-art input segmentations, reducing the variation of information with expert labeled ground truth by 15-20%.

Graph Analysis

- Implemented a subgraph enumeration algorithm to identify frequently occurring motifs in the wiring diagram of the brain. Introduced methods to differentiate subgraphs based on the type of edges (e.g., excitatory or inhibitory synaptic connection).
- Parallelized subgraph enumeration with an effective load balancing heuristic to enable the processing of large (20,000 nodes) and extremely dense (avg. degree ~ 20) graphs. Distributed computation across a cluster to enumerate over 26 trillion subgraphs (10 years of computation time) in less than two weeks. Reduced wall time by $11\times$ by using the load balancing heuristic.
- Assembled and published the largest dataset of enumerated subgraphs in a connectome (26 trillion) to encourage future longitudinal studies between wiring diagrams of different animal species.

Compression

- Designed the *Compresso* algorithm that is specifically tailored to compress terabyte label volumes. Achieved compression ratios exceeding $700\times$, an $8\times$ improvement over the currently used general-purpose solutions (gzip).
- Implemented *Compresso* in C++, Javascript, and Python to encourage widespread adoption in the community, processing over 100-200 megabytes of data per second.

Volume Processing

- Created a novel block-based topological thinning algorithm to generate centerlines, or skeletons, for connectomic label volumes. Connectomic researchers frequently use skeletons for segmentation evaluation, analysis, visualization, and error correction.
- Divided the most computationally expensive steps in the pipeline into small parallelizable chunks enabling an overall throughput of one million voxels per second per CPU. Validated our design by processing a hundred gigavoxel connectomic volume in under two hours on a computer cluster.
- Improved over existing state-of-the-art skeleton generation techniques on three metrics, including a $5\times$ improvement on the fidelity of a skeleton to the underlying neuronal circuitry.

Freelance Consulting

Electrical Distribution Company

DATA ANALYST

December 2018 - Present

- Developed a “pricing matrix” that recommends to salespeople target margins for each transaction based on their region of the country, the product class, and the customer type.
- Created a web-based interface with a Django backend and PostgreSQL database that enables the company to analyze millions of transactions over the previous eight years and identify revenue and margin trends by product, customer, and salesperson.
- Automated the creation of several reports that summarize salesperson adherence to the matrix and the effects of incremental changes to individual matrix entries, among others.
- Participated in weekly meetings over several months with a company president and others to devise an incentive program that encourages salespeople to adhere to the new pricing matrix.

Military-Focused 3D Data Technology Startup

TECHNICAL & PRODUCT CONSULTANT

June 2021 - September 2021

- Connected the company’s existing desktop technology to a new web interface and added features enabling users to upload gigabyte LIDAR scans to an Amazon S3 bucket. On file upload, a lambda function converts the file into a format suitable for web-based visualization.
- Helped draft a 30+ page provisional patent for the company that included more than a dozen figures and schematics of the end-to-end pipeline. Summarized dozens of existing patents in the same domain to help differentiate the proposed technology from the prior art.
- Engaged in 3-4 hour long meetings per week with potential customers and collaborators in the Air Force, Navy, Coast Guard, and private sector. Led several meetings with research scientists in the defense sector to discuss the current state and future vision of the technology.
- Co-led a demonstration of the company’s visualization software for three potential clients on a WWII-era museum battleship.

Leadership and Service

Horizon Inspires

Remote

MENTOR

July 2021 - Present

- Mentored three high school students on research projects on topics in computer vision. Student projects included: diagnosing pneumonia and COVID-19 from chest X-Rays, detecting smoke from wildfires in satellite imagery, and determining human stress levels from video.
- Led 3 — 5 hours of meetings per week (20 hours over 12 — 16 weeks per student) to help direct their research. During sessions, helped the students debug their code, understand existing academic literature, and discover new research paths to explore.

Harvard Mind Brain Behavior

Cambridge MA, 02138

GRADUATE STUDENT COMMITTEE MEMBER

September 2019 - May 2020

- Planned with four other students a several-part lecture series headlining professors from around the country that investigate computer science, neuroscience, and biology.
- Coordinated with the undergraduate Mind Brain Behavior committee to develop a mentorship program that pairs undergraduates interested in academic research with Ph.D. candidates.

Harvard School of Engineering and Applied Sciences Student Council

Cambridge, MA 02138

SOCIAL COORDINATOR

April 2018 - March 2020

- Helped coordinate three parties per year for over 500 engineering students. Managed with two other volunteers the acquisition of all food and drinks (\$2,500 per party).
- Attended weekly meetings with the entire engineering student council to plan welcome weekends, guest lectures, and career events.

Skills

Programming Python, C++, MATLAB, Java, JavaScript, Julia, HTML5, C

Other Tools Cython, Keras, Numpy, Scipy, Pandas, Matplotlib, Django, CSS, Sass, JQuery, Linux, SLURM, LaTeX