David Chemaly

✉ dc824@cam.ac.uk   📞 +44 7535666432   🔗 david-chemaly.github.io/

Profile

Experienced Quantitative Researcher with a strong foundation in astrophysics, statistics, and machine learning. Skilled in advanced data analysis, mathematical modeling, and handling large datasets. Demonstrated ability to develop and implement algorithms for data-driven discoveries. Over 500 hours of teaching experience, highlighting strong communication and leadership skills. Eager to leverage interdisciplinary expertise to drive innovative, data-driven solutions.

Education

PhD   **University of Cambridge**, Astronomy   2023 – 2026
- Supervisor: Prof. Vasily Belokurov
- Fully funded by the Harding Distinguished Postgraduate Scholarship ↗

M.Sc.   **University of Montreal**, Astrophysics (GPA: 4.3/4.3)   2021 – 2023
- Supervisor: Prof. Yashar Hezaveh
- Fellowship at the Institute for Data Valorization ↗
- Funded by the Natural Sciences Research Council of Canada ↗

B.Sc.   **University of Montreal**, Physics (GPA: 4.2/4.3)   2018 – 2021
- *Summa Cum Laude*
- Funded by the Quebec Research Fund in Nature and Technologies ↗

Research Experience

Machine Learning-Based Automatic Discovery and Classification of Low Surface Brightness Features
- Efficiently managed and preprocessed over 1 PB of photometric data to fine-tune a foundational model
- Developed a novel combined loss function for simultaneous discovery and classification of dwarf galaxies, achieving 97% accuracy with a 2% false positive rate
- Accelerated state-of-the-art automated processes by over 1,000 times without compromising performance

Hierarchical Bayesian Inference: Constraining the Population Distribution of Dark Matter Halos
- Optimized data analysis pipelines for high-frequency photometric data, enhancing processing efficiency
- Reduced computational time complexity from $\mathcal{O}(n^2)$ to $\mathcal{O}(n)$ by developing an improved population likelihood
- Utilized both Nested Sampling and MCMC to improve robustness and accuracy of probabilistic models

The Multimodal Universe: Enabling Large-Scale Machine Learning with 70 TB of Astronomical Data
- Led the processing, preprocessing, and benchmarking of over 5 TB of time series data
- Trained advanced machine learning and deep learning models, setting new performance benchmarks in classification and regression tasks
- Co-authored a paper accepted at NeurIPS 2024 Datasets and Benchmarks track

Supervisor & Teaching Assistant
- Led teaching assistant roles for 10 courses, demonstrating strong leadership and communication skills
- Mentored and coordinated over 250 students, enhancing comprehension of complex subjects

Skills and Interests

Technical: Python, MATLAB, PyTorch, Lightning, HuggingFace, TensorFlow, Keras
Languages: English, French, Arabic
Interests: Chess, Competitive Rowing