

Minh Hoang | Curriculum Vitae

Lewis-Sigler Institute for Integrative Genomics, Princeton University

Princeton, NJ 08540, USA

✉ minhhoang@princeton.edu

Research Interests

Machine Learning and Computational Biology

Automated Algorithm Design, Sequence Sketching, Sequence Compression, and Protein Language Models

Education

Carnegie Mellon University

School of Computer Science

Ph.D. in Computer Science, Advisor: Carl Kingsford

2018–2023

Thesis: Practical Methods for Automated Algorithm Design in Machine Learning and Computational Biology

National University of Singapore

School of Computer Science

B.Comp. (First Class Honor)

2012–2017

Thesis: A Probabilistic Approach for Protein Function Prediction with Hierarchical Structured Outputs

Honor and Awards

Ian Lawson Van Toch Memorial Award for Outstanding Student Paper at ISMB, 2025

Lee Kwan Yew Gold Medal for Outstanding Graduates, 2017

Professional Experience

Princeton University

Postdoctoral Research Associate

2023–Present

Developing efficient computational frameworks for proteomics applications.

Co-Lecturer of [IW-04] Machine Learning and Algorithms for Medicine.

Carnegie Mellon University

Teaching Assistant

2021–2022

Courses: [07–300] Research and Innovation (Wehbe, Vasilescu), [10–716] Advanced Theory in Machine Learning (Ravikumar)

Amazon Web Services (AWS)

Applied Scientist Intern

2020

Developed deep ensemble methods for supervised learning with tabular data, contributing to AWS SageMaker ML services.

Conference Program Committee

ICML, NeurIPS, AAAI, ICLR, ICCV, IJCAI (Senior PC)

2018–Present

Singapore-MIT Alliance for Research and Technology

Research Assistant

2017–2018

Developed large-scale probabilistic optimization frameworks for urban traffic monitoring (Singapore Smart City Initiative).

A*STAR Singapore, Bioinformatics Institute

Research Intern

2016

Curated hand gesture dataset and developed classification algorithm for structured data with biomedical applications.

Selected Publications

[1]: Minh Hoang and Mona Singh (2025). *Locality-aware pooling enhances protein language model performance across varied applications*. 33rd International Conference on Intelligent Systems for Molecular Biology (ISMB). Ian Lawson Van Toch Memorial Award for Outstanding Student Paper.

- [2]: Pei-Yau Weng, **Minh Hoang**, and Trong Nghia Hoang (2024). *Probabilistic Federated Prompt-Tuning in Data Imbalance Settings*. 41st Conference on Neural Information Processing Systems (NeurIPS).
- [3]: **Minh Hoang**, Azza Fadhel, Aryan Deshwal, Janardhan Rao Doppa, and Trong Nghia Hoang (2024). *Learning Surrogates for Offline Black-Box Optimization via Gradient Matching*. 41st International Conference on Machine Learning (ICML).
- [4]: **Minh Hoang** and Carl Kingsford (2024). *Efficient Heterogeneous Meta-Learning via Channel Shuffling Modulation*. 12th International Conference on Learning Representations (ICLR).
- [5]: C. S. Elder, **Minh Hoang**, Mohsen Ferdosi, and Carl Kingsford (2024). *A Scalable Optimization Algorithm for Solving the Beltway and Turnpike Problems with Uncertain Measurements*. 28th International Conference on Research in Computational Molecular Biology (RECOMB).
- [6]: **Minh Hoang** and Nghia Hoang (2024). *Few-Shot Learning via Repurposing Ensemble of Black-Box Models*. 38th AAAI Conference on Artificial Intelligence (AAAI).
- [7]: **Minh Hoang**, Guillaume Marçais, and Carl Kingsford (2023). *Density and Conservation Optimization of the Generalized Masked-Minimizer Sketching*. Journal of Computational Biology (JCB).
- [8]: **Minh Hoang**, Hongyu Zheng, and Carl Kingsford (2022). *Differentiable Learning of Sequence-Specific Minimizer Schemes with DeepMinimizer*. Journal of Computational Biology (JCB).
- [9]: **Minh Hoang**, Hongyu Zheng, and Carl Kingsford (2022). *DeepMinimizer: A Differentiable Framework for Optimizing Sequence-Specific Minimizer Schemes*. 26th International Conference on Research in Computational Molecular Biology (RECOMB).
- [10]: **Minh Hoang** and Carl Kingsford (2020). *Optimizing Dynamic Structures with Bayesian Generative Search*. 37th International Conference on Machine Learning (ICML).
- [11]: **Minh Hoang**, Nghia Hoang, Hai Pham, and David P. Woodruff (2020). *Revisiting the Sample Complexity of Sparse Spectrum Approximation of Gaussian Processes*. 34th Conference on Neural Information Processing Systems (NeurIPS).
- [12]: **Minh Hoang**, Nghia Hoang, Brian Kian Hsiang Low, and Carl Kingsford (2019). *Collective Model Fusion for Multiple Black-box Experts*. 36th International Conference on Machine Learning (ICML).
- [13]: Nghia Hoang, **Minh Hoang**, and Brian Kian Hsiang Low (2019). *Collective Online Learning via Decentralized Gaussian Processes in Massive Multi-Agent Systems*. 33rd AAAI Conference on Artificial Intelligence (AAAI).
- [14]: Nghia Hoang, **Minh Hoang**, Ruofei Ouyang, and Brian Kian Hsiang Low (2018). *Decentralized High-Dimensional Bayesian Optimization with Factor Graphs*. 32nd AAAI Conference on Artificial Intelligence (AAAI).
- [15]: **Minh Hoang**, Trong Nghia Hoang, and Kian Hsiang Low (2017). *A Generalized Stochastic Variational Bayesian Hyperparameter Learning Framework for Sparse Spectrum Gaussian Process Regression*. 31st AAAI Conference on Artificial Intelligence (AAAI).
- [16]: Nghia Hoang, **Minh Hoang**, and Brian Kian Hsiang Low (2016). *A Distributed Variational Inference Framework for Unifying Parallel Sparse Gaussian Process Regression Models*. 33rd International Conference on Machine Learning (ICML).
- [17]: Nghia Hoang, **Minh Hoang**, and Brian Kian Hsiang Low (2015). *A Unifying Framework of Anytime Sparse Gaussian Process Regression Models with Stochastic Variational Inference for Big Data*. 32nd International Conference on Machine Learning (ICML).

References

Prof. Mona Singh: Princeton University, mona@cs.princeton.edu

Prof. Carl Kingsford: Carnegie Mellon University, carlk@cs.cmu.edu

Prof. David P. Woodruff: Carnegie Mellon University, dwoodruf@cs.cmu.edu

Prof. Bryan K. H. Low: National University of Singapore, lowkh@comp.nus.edu.sg