Shaojie Bai

Machine Learning Department, 4902 Forbes Ave. – Pittsburgh, PA – USA
✉ shaojieb@cs.cmu.edu • jerrybai1995.github.io • Google Scholar

Education

Carnegie Mellon University **Pittsburgh, PA**
Doctor of Philosophy (Ph.D.) in Machine Learning (Advisor: J. Zico Kolter) *Aug. 2017–May 2022*

- Doctoral Thesis: Equilibrium Approaches to Modern Deep Learning

Carnegie Mellon University **Pittsburgh, PA**
Bachelor of Science (B.S.) in Computer Science, B.S. in Mathematics, University Honor *Aug. 2013–May 2017*
Cumulative GPA: 4.0/4.0

Select Publications

- **Shaojie Bai***, Zhengyang Geng*, Yash Savani and J. Zico Kolter. Deep Equilibrium Optical Flow Estimation. *Conference on Computer Vision and Pattern Recognition (CVPR) 2022*.
- **Shaojie Bai**, Vladlen Koltun and J. Zico Kolter. Neural Deep Equilibrium Solvers. *International Conference on Learning Representations (ICLR) 2022*.
- Zaccharie Ramzi, Florian Mannel, **Shaojie Bai**, Jean-Luc Starck, Philippe Ciuciu and Thomas Moreau. SHINE: SHaring the INverse Estimate from the forward pass for bi-level optimization and implicit models. *International Conference on Learning Representations (ICLR) 2022*.**(Spotlight Oral, 5% acceptance rate)**
- Zhengyang Geng*, Xin-Yu Zhang*, **Shaojie Bai**, Yisen Wang and Zhouchen Lin. On Training Implicit Models. *Neural Information Processing Systems (NeurIPS) 2021*.
- Zhichun Huang, **Shaojie Bai** and J. Zico Kolter. Implicit^2: Implicit Layers for Implicit Representations. *Neural Information Processing Systems (NeurIPS) 2021*.
- Swaminathan Gurumurthy, **Shaojie Bai**, Zachary Manchester and J. Zico Kolter. Joint Inference and Input Optimization in Equilibrium Networks. *Neural Information Processing Systems (NeurIPS) 2021*.
- **Shaojie Bai**, Vladlen Koltun and J. Zico Kolter. Stabilizing Equilibrium Models by Jacobian Regularization. *International Conference on Machine Learning (ICML) 2021*.
- Lars A Bratholm, Will Gerrard, Brandon Anderson, **Shaojie Bai** et al. A Community-powered Search of Machine Learning Strategy Space to Find NMR Property Prediction Models. *PLOS ONE 16(7), 2021*.
- **Shaojie Bai**, Vladlen Koltun and J. Zico Kolter. Multiscale Deep Equilibrium Models. *Neural Information Processing Systems (NeurIPS) 2020*. **(Oral, 1% acceptance rate)**
- **Shaojie Bai**, J. Zico Kolter and Vladlen Koltun. Deep Equilibrium Models. *Neural Information Processing Systems (NeurIPS) 2019*. **(Spotlight Oral, 3% acceptance rate; also led to a NeurIPS 2020 Tutorial)**
  Code: https://github.com/locuslab/deq
- Yao-Hung Hubert Tsai, **Shaojie Bai**, Makoto Yamada, Louis-Philippe Morency and Ruslan Salakhutdinov. Transformer Dissection: An Unified Understanding for Transformer's Attention via the Lens of Kernel. *Conference on Empirical Methods in Natural Language Processing (EMNLP) 2019*.
- Yao-Hung Hubert Tsai*, **Shaojie Bai***, Paul Pu Liang, J. Zico Kolter, Louis-Philippe Morency and Ruslan Salakhutdinov. Multimodal Transformer for Unaligned Multimodal Language Sequences. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL) 2019*. (*equal contribution)
  Code: https://github.com/yaohungt/Multimodal-Transformer
- **Shaojie Bai**, J. Zico Kolter and Vladlen Koltun. Trellis Networks for Sequence Modeling. *Interntional Conference on Learning Representations (ICLR) 2019*.
  Code: https://github.com/locuslab/trellisnet

- **Shaojie Bai**, J. Zico Kolter and Vladlen Koltun. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv:1803.01271*.
  Code: https://github.com/locuslab/TCN (★ 3,023 on GitHub by 11/11/2021)

- **Shaojie Bai**, J. Zico Kolter and Vladlen Koltun. Convolutional Sequence Modeling Revisited. *International Confernce on Learning Representations (ICLR) 2018 Workshop*.

Professional Services

Reviewer: IEEE-TPAMI (journal), IEEE-TIP (journal), ICML 2019/2020/2021/2022, ICLR 2020/2021/2022, NeurIPS 2020/2021/2022, ICRA 2021

Organizer/Coordinator: CMU Artificial Intelligence Seminar 2021-2022

Skills

Research Interest: Deep Learning, Sequence Modeling, Implicit Networks, Optimizations

Programming: Python, Java, C, C++, Javascript, Go, Standard ML, Common Lisp, LaTeX

Frameworks & Softwares: PyTorch, Tensorflow, Keras, CUDA, ASP.NET, Matlab

Relevant Courses

- Computer Vision
- Parallel Computer Architecture and Programming
- Distributed Systems
- Probabilistic Graphical Models
- Convex Optimization
- Statistical Machine Learning

Employment

Facebook AI Research (FAIR) **(remote) Pittsburgh, PA**
Research Intern (AI) *May 2020–Aug. 2020*
- I work with Michael Auli, Jiatao Gu, Zhouhan Lin and Alexei Baevski at FAIR on improving self-attention mechanism using cross-scale features of NLP data.

Bosch Center for AI (BCAI) **Pittsburgh (PA) USA and Renningen, Germany**
Machine Learning Research Intern *May 2019–Aug. 2019*
- I work with the chief scientist of AI research at BCAI (who is also my advisor), Zico Kolter, on applying autoregressive deep sequence models on practical control problems. I spent part of the internship at the headquarter in Germany.

Intel Labs **Santa Clara, CA**
Graduate Research Intern *May 2018–Aug. 2018*
- I work with Vladlen Koltun at Intel Labs (Intelligent Systems Lab) on sequence models and general machine learning (work accepted to ICLR 2019).

PayPal **San Jose, CA**
Software Engineer Intern (Machine Learning Research) *May 2016–Aug. 2016*
- As a part of PayPal merchant-side's machine learning (ML) group, I was involved in its development of new Partner Detection Service (PDS). PDS was a web-crawling based machine learning project that automated the e-commerce detection task during merchants' onboarding flow.

Epic Systems **Verona, WI**
Software Developer Intern *May 2015–Aug. 2015*
- During my internship at Epic, I mainly worked on developing their next-gen patient registration system, *Welcome*, in web-framework so as to provide more user interactions and support basic Q&As.

Teaching

Carnegie Mellon University, Machine Learning Department **Pittsburgh, PA**
TA for course 10-701: Machine Learning for PhD - Instructor: M. Veloso and P. Ravikumar *Jan. 2018 - May 2018*

Carnegie Mellon University, Computer Science Department **Pittsburgh, PA**
TA for course 15-440/640: Distributed Systems - Instructor: M. Satyanarayanan *Jan. 2017 - May 2017*

Carnegie Mellon University, Machine Learning Department **Pittsburgh, PA**
TA for course 10-601: Machine Learning for Graduates - Instructor: R. Rosenfeld *Aug. 2016 - Dec. 2016*

Carnegie Mellon University, Computer Science Department **Pittsburgh, PA**
TA for course 15-440/640: Distributed Systems - Instructor: M. Satyanarayanan *Jan. 2016 - May 2016*

Projects & Competitions

Kaggle Competition - Predicting Molecular Properties (Rank: 1/2749, top 0.05%) **Pittsburgh, PA**
Kaggle Team with BCAI *July 2019 - Aug. 2019*

- This challenge aims to predict the magnetic interactions between atoms, whose state-of-the-art computation method depends on days of quantum mechanics calculations. Using a graph transformer network, my team is ranked 1/2749 (top 0.05%) on both the public and private leaderboards.

Link: https://www.kaggle.com/c/champs-scalar-coupling (team name: *hybrid*)
Code: https://github.com/boschresearch/BCAI_kaggle_CHAMPS

The Effect of Pre-ReLU Input Distribution on DNN's Performance **Pittsburgh, PA**
Undergraduate Thesis at CMU - Advisor: J. Zico Kolter *Sep. 2016 - May 2017*

- This research delves into how the input distribution to ReLU activation could influence the performance of deep neural networks (DNN). An important focus of the research included exploring the effect of batch-normalization (BN). We attempted to look at different simulations of such behavior, including approximating and experimenting with methods such as optimizations in the Chebyshev space as well as Copula transformation.

Invited Talks & Presentations

- 2021/10: **Extrality.ai and INRIA**, *in Pittsburgh (PA), USA [remote]*
- 2021/8: **TechBeat**, *in Pittsburgh (PA), USA [remote]*
- 2021/6: **French Institute for Research in CS and Automation (aka. INRIA)**, *in Pittsburgh (PA), USA [remote]*
- 2021/1: **Scientific Machine Learning Webinar Series (MIT & CMU)**, *in Pittsburgh (PA), USA [remote]*
- 2020/12: **Tsinghua University (AITime)**, *in Pittsburgh (PA), USA [remote]*
- 2020/12: **Neural Information Processing Main Conference (NeurIPS oral)**, *in Pittsburgh (PA), USA [remote]*
- 2020/10: **Bosch Center for AI**, *in Pittsburgh (PA), USA [remote]*
- 2020/7: **University of California, Berkeley**, *in Pittsburgh (PA), USA [remote]*
- 2020/5: **Intel Intelligent Systems Lab**, *in Pittsburgh (PA), USA [remote]*
- 2019/12: **Neural Information Processing Main Conference (NeurIPS spotlight)**, *in Vancouver, Canada*
- 2019/11: **RIKEN Center for Advanced Intelligence Project (AIP) and Kyoto University**, *in Kyoto, Japan*
- 2018/8: **Intel Intelligent Systems Lab**, *in Santa Clara (CA), USA*

Honors & Awards

- **J.P. Morgan AI Research Fellowship 2020**
- **1st Place in the *CHAMPS - Predicting Molecular Property* Kaggle Competition**: Aug 2019
- **Carnegie Mellon University School of Computer Science College Honor**: May 2017
- **Carnegie Mellon University Mellon College of Science College Honor**: May 2017