Guodong Zhang

University of Toronto & Vector Institute
Department of Computer Science
Machine Learning Group

(+1) 647-675-8256
gdzhang.cs@gmail.com
www.cs.toronto.edu/∼gdzhang/

RESEARCH INTERESTS

Areas: deep learning dynamics, Bayesian deep learning, multi-agent optimization and efficient machine learning.

I study the theoretical foundations and practical algorithms for machine learning. As an academic, I aim to understand existing learning algorithms, build up theoretical guarantees for those algorithms. As an engineering academic, I believe better understanding and theory should improve practice. Based on deep understanding, I design algorithms and models that train faster, generalize better, and give well-calibrated uncertainty.

EDUCATION

Ph.D. Student, **University of Toronto, Canada** Jan. 2019 - present

- Department of Computer Science
- Advisor: Roger Grosse

M.Sc. Student, **University of Toronto, Canada** Sep. 2017 - Jan. 2019

- Department of Computer Science
- Advisor: Roger Grosse and David Duvenaud

B.Eng., Information Engineering, **Zhejiang University, China** Aug. 2013 - Jun. 2017

- Minor: Advanced Class of Engineering Education, Chu-Kochen Honors College.
- GPA: 92.71/100 (3.96/4.0), Rank: **1 / 182** (Three consecutive years)

RESEARCH VISITS

Visiting Graduate Student, **School of Mathematics, Institute for Advanced Study** Spring 2020

- Special Year on Optimization, Statistics, and Theoretical Machine Learning

Visiting Undergraduate Student, **University of California, Los Angeles** Summer 2016

- Cross-disciplinary Scholars in Science and Technology (**CSST**)
- Advisor: Song-Chun Zhu and Ying Nian Wu

PROFESSIONAL EXPERIENCE

Research Scientist Intern, **DeepMind**, London May. 2021 - Oct. 2021

- Host: James Martens

Research Intern, **Google Brain**, Toronto Jun. 2019 - Nov. 2019

- Host: Geoffrey Hinton

Student Researcher, **Google Brain**, Toronto Feb. 2019 - May. 2019

- Host: Geoffrey Hinton and Lala Li

Research Intern, **Microsoft Research Asia**, Beijing Nov. 2016 - Jun. 2017

- Host: Jifeng Dai

TEACHING

University of Toronto

Course Instructor

- CSC311: Introduction to Machine Learning (2021 Fall)

Guest Lecturer

- CSC2541: Neural Network Training Dynamics (2021 Winter)

Teaching Assistant

- CSC2541: Neural Network Training Dynamics (2021 Winter)
- CSC311: Introduction to Machine Learning (2020 Fall)
- CSC2515: Machine Learning (2019 Fall)
- CSC411: Machine Learning and Data Mining (2018 Fall)
- CSC321: Introduction to Neural Networks and Machine Learning (2018 Winter)
- CSC384: Introduction to Artificial Intelligence (2017 Fall, 2018 Summer, 2019 Winter)

PUBLICATIONS

Note: * below denotes equal contribution (co-first author).

Google Scholar and Semantic Scholar

Preprint or Workshop Publications

P1 Cem Anil, **Guodong Zhang**, Yuhuai Wu, Roger Grosse. Learning to Give Checkable Answers with Prover-Verifier Games. *arXiv, 2021*.

P2 Tingwu Wang, Xuchan Bao, Ignasi Clavera, Jerrick Hoang, Yeming Wen, Eric Langlois, Shunshi Zhang, **Guodong Zhang**, Pieter Abbeel, Jimmy Ba. Benchmarking Model-Based Reinforcement Learning. *arXiv, 2019*.

P3 Juhan Bae, **Guodong Zhang**, Roger Grosse. Eigenvalue Corrected Noisy Natural Gradient. *Bayesian Deep Learning Workshop, **NeurIPS** 2018*.

Journal Publications

J1 **Guodong Zhang**, Xuchan Bao, Laurent Lessard, Roger Grosse. A Unified Analysis of First-Order Methods for Smooth Games via Integral Quadratic Constraints. *Journal of Machine Learning Research, (**JMLR**), 2021*.

J2 **Guodong Zhang**, Xiaojin Gong. Non-Negative Matrix Co-factorization for Weakly Supervised Image Parsing. *IEEE SIGNAL PROCESSING LETTERS, 2016*.

Conference Publications

C1 **Guodong Zhang**, Alex Botev, James Martens. Deep Learning without Shortcuts: Shaping the Kernel with Tailored Rectifiers. *International Conference on Learning Representations (**ICLR**), 2022*.

C2 **Guodong Zhang**, Yuanhao Wang, Laurent Lessard, Roger Grosse. Near-optimal Local Convergence of Alternating Gradient Descent-Ascent for Minimax Optimization. *International Conference on Artificial Intelligence and Statistics (**AISTATS**), 2022*.

C3 **Guodong Zhang**, Kyle Hsu, Jianing Li, Chelsea Finn, Roger Grosse. Differentiable Annealed Importance Sampling and the Perils of Gradient Noise. *Neural Information Processing Systems (**NeurIPS**), 2021*.

C4 **Guodong Zhang**, Yuanhao Wang. On the Suboptimality of Negative Momentum for Minimax Optimization. *International Conference on Artificial Intelligence and Statistics (**AISTATS**), 2021*.

C5 Yuanhao Wang*, **Guodong Zhang***, Jimmy Ba. On Solving Minimax Optimization Locally: A Follow-the-Ridge Approach. *International Conference on Learning Representations (**ICLR**), 2020*.

C6 Chaoqi Wang, **Guodong Zhang**, Roger Grosse. Picking Winning Tickets Before Training by Preserving Gradient Flow. *International Conference on Learning Representations (**ICLR**), 2020*.

C7 Yeming Wen*, Kevin Luk*, Maxime Gazeau*, **Guodong Zhang**, Harris Chan, Jimmy Ba. An Empirical Study of Large-Batch Stochastic Gradient Descent with Structured Covariance Noise. *International Conference on Artificial Intelligence and Statistics (**AISTATS**), 2020*.

C8 **Guodong Zhang**, Lala Li, Zachary Nado, James Martens, Sushant Sachdeva, George E. Dahl, Christopher J. Shallue, Roger Grosse. Which Algorithmic Choices Matter at Which Batch Sizes? Insights From a Noisy Quadratic Model. *Neural Information Processing Systems (**NeurIPS**), 2019*.

C9 **Guodong Zhang**, James Martens, Roger Grosse. Fast Convergence of Natural Gradient Descent for Overparameterized Networks. *Neural Information Processing Systems (**NeurIPS**), 2019*.

C10 Chaoqi Wang, Roger Grosse, Sanja Fidler, **Guodong Zhang**. EigenDamage: Structured Pruning in the Kronecker-Factored Eigenbasis. *International Conference on Machine Learning (**ICML**), 2019*.

C11 **Guodong Zhang**, Chaoqi Wang, Bowen Xu, Roger Grosse. Three Mechanisms for Weight Decay Regularization. *International Conference on Learning Representations (**ICLR**), 2019*.

C12 Shengyang Sun*, **Guodong Zhang***, Jiaxin Shi*, Roger Grosse. Functional Variational Bayesian Neural Networks. *International Conference on Learning Representations (**ICLR**), 2019*.

C13 **Guodong Zhang***, Shengyang Sun*, David Duvenaud, Roger Grosse. Noisy Natural Gradient as Variational Inference. *International Conference on Machine Learning (**ICML**), 2018*.

C14 Shengyang Sun, **Guodong Zhang**, Chaoqi Wang, Wenyuan Zeng, Jiaman Li, Roger Grosse. Differentiable Compositional Kernel Learning for Gaussian Processes. *International Conference on Machine Learning (**ICML**), 2018*.

C15 Jifeng Dai*, Haozhi Qi*, Yuwen Xiong*, Yi Li*, **Guodong Zhang***, Han Hu, Yichen Wei. Deformable Convolutional Network. *International Conference on Computer Vision (**ICCV**), 2017*.

SELECTED HONORS & AWARDS

- Apple PhD Fellowship 2022
- Ontario Graduate Scholarship 2020
- Borealis AI Fellowships 2020
- Computer Science 50th Anniversary Graduate Scholarship, UofT 2019, 2020
- CHU Kochen Scholarship (Highest distinction in Zhejiang University) 2016
- Cross-disciplinary Scholars in Science and Technology, UCLA 2016
- National Scholarship in China (**1.5%**) 2014, 2015, 2016
- Meritorious Winner, Interdisciplinary Contest in Modeling (ICM) 2016
- 1^{st} Prize, China Undergraduate Mathematical Contest in Modeling (**1.5%**) (CUMCM) 2015

TALKS

T1 Training Deep Neural Networks without Shortcuts. DeepMind. Oct 2021.

T2 Differentiable Game Dynamics. Theory Meets Practice Seminar. DeepMind. Sep 2021.

T3 Differentiable Game Dynamics: Hardness and Complexity of Equilibrium Learning. Math ML Seminar. Max Planck Institute for Mathematics in the Sciences. Aug 2021.

T4 Differentiable Games. CSC2541: Neural Network Training Dynamics Guest Lecturer. University of Toronto. Apr 2021.

T5 Control Theory for Machine Learning Tutorial. University of Toronto. Nov 2020.

T6 Integral Quadratic Constraints for Smooth Games. University of Toronto. Sep 2020.

T7 Which Algorithmic Choices Matters at Which Batch Sizes. Vector Institute Workshop for Machine Learning Systems. Toronto. Oct 2019.

T8 Large-batch Neural Network Training. Google Brain Toronto Journal Club. Toronto. Sep 2019.

T9 Which Algorithmic Choices Matters at Which Batch Sizes. Google Brain, Online Weekly Seminar. Aug 2019.

T10 Three Mechanisms of Weight Decay. Vector Institute. Toronto. Jan 2019.

T11 Natural Gradient Methods. Optimization Seminar, University of Toronto. Nov 2018.

T12 Noisy Natural Gradient. Google Brain Toronto Journal Club. Toronto. Oct 2018.

T13 Noisy Natural Gradient as Variational Inference. ICML 2018. Stockholm. July 2018.

T14 Noisy Natural Gradient as Variational Inference. NeurIPS 2017 Bayesian Deep Learning Workshop. Long Beach. Dec 2017.

PROFESSIONAL SERVICE

I am/was a reviewer for

- Conference on Neural Information Processing Systems (NeurIPS) 2018-2021
- International Conference on Machine Learning (ICML) 2019-2021
- International Conference on Learning Representations (ICLR) 2018-2021
- International Conference on Artificial Intelligence and Statistics (AISTATS) 2020-2021
- Uncertainty in Artificial Intelligence (UAI) 2018
- Journal of Machine Learning Research (JMLR)

STUDENT SUPERVISION & MENTORSHIP

- Jakob Kelly (UofT undergraduate) 2021
- Muhammad Adil Asif (UofT undergraduate) 2021
- Yao Kuang (UofT undergraduate) 2021
- Ruth Crasto (UofT undergraduate $\rightarrow$ SDE at Microsoft) 2021
- Jianing (Robert) Li (UofT undergraduate $\rightarrow$ M.Sc student at UofT) 2020-2021
- Farzaneh Mahdisoltani (UofT PhD) 2020
- Yuanhao Wang (Tsinghua undergraduate $\rightarrow$ PhD student at Princeton University) 2019
- Kyle Hsu (UofT undergraduate $\rightarrow$ PhD student at Stanford) 2019
- Chaoqi Wang (UofT M.Sc $\rightarrow$ PhD student at University of Chicago) 2019-2020
- Xuchan (Jenny) Bao (UofT undergraduate $\rightarrow$ PhD student at UofT) 2018
- Juhan Bae (UofT undergraduate $\rightarrow$ PhD student at UofT) 2018