Jinkun Lin

Github: https://github.com/lazycal Email : jinkun.lin@nyu.edu

Education

- **New York University** — New York, USA
  Ph.D. in Computer Science — *09/2019 – Present*
- **Tsinghua University** — Beijing, China
  B.Eng. in Computer Science and Technology — *08/2015 – 07/2019*

Research interests

I am broadly interested in systems research, particularly in systems abstraction, reliability, debugging, and profiling aspects. My recent research focus is on Machine Learning Systems and Machine Learning Compilers.

Research & Internship Experience

- **NYU Systems Group, *New York University*** — New York, USA
  Research Assistant — *09/2019 – Present*
  Advisor: Prof. Aurojit Panda & Prof. Jinyang Li
  - *Automatic Operator Fusion*: Ongoing project aiming to automatically search for operator fusion strategies that better utilize the multi-layer memory hierarchy on GPUs.
  - *Testing for DNN compilers (NNSmith)*:
    - Encoded operator parameter specifications using Z3 to generate valid DNN models.
    - Encoded operator input ranges with loss functions to search numerically valid input with gradient-descend.
  - *SmartNIC Offloading (QingNiao)*: Helping with an ongoing project aiming to offload L7 dispatch.
  - *ML Explainability (AME)*: A metric to measure the contribution of each training data on ML predictions.
    - Designed an efficient AME estimator by exploiting the sparsity of data contributions using LASSO.
    - Applied Knockoffs to select the data with high contribution with a controlled false selection rate.
- **AML Group, *ByteDance*** — USA
  Research Intern — *05/2023 – 11/2023, 03/2024 – 08/2024*
  Advisor: Haibin Lin & Ziheng Jiang
  - *Straggler analysis in LLM training*:
    - Built and deployed a what-if based straggler analysis and monitoring tool for LLM training systems.
  - *Fault-tolerance of LLM training*:
    - Implemented asynchronous and remote memory checkpoint mechanisms.
    - Relaxed the rigid requirement of 3D-parallel training on the number of training nodes by allowing different pipelines to use different numbers of nodes, so that more remaining nodes can be used after node failures; Used dynamic programming to search for the optimal configuration.
- **AML Group, *ByteDance*** — USA
  Research Intern — *06/2021 – 09/2021*
  Advisor: Prof. Cheng Tan
  - *DNN Compilers Reliability*:
    - Formulated DNN operator computation in SMT and verified TVM's graph passes on ResNet.
    - Explored fuzz testing on deep learning compilers.
- **PACMAN Lab, *Tsinghua University*** — Beijing, China
  Research Assistant — *05/2018 – 07/2019*
  Advisor: Prof. Wenguang Chen
  - *Sparse Computation on GPUs*: Implemented efficient GPU kernels to compute sparse tensor addition stored in a compound format proposed in the paper "TACO: The Tensor Algebra Compiler".
- **ALCHEM Lab, *University of Southern California*** — Los Angeles, USA
  Research Intern — *07/2018 – 09/2018*
  Advisor: Prof. Xuehai Qian

- *Distributed training (HOP)*: Implemented a heterogeneity-aware decentralized training algorithm for machine learning using TensorFlow; Designed and conducted corresponding experiments.

Publications & Preprints

- **Stateful Large Language Model Serving with Pensieve.**
  Lingfan Yu, Jinkun Lin, Jinyang Li.
  Eurosys 2025.
- **Application-Defined Receive Side Dispatching on the NIC.**
  Tao Wang, Jinkun Lin, Gianni Antichi, Aurojit Panda, Anirudh Sivaraman.
  Preprint.
- **NNSmith: Generating Diverse and Valid Test Cases for Deep Learning Compilers.**
  Jiawei Liu, Jinkun Lin*(Equal Contribution), Fabian Ruffy, Cheng Tan, Jinyang Li, Aurojit Panda, Lingming Zhang.*
  ASPLOS 2023.
- **Measuring the Effect of Training Data on Deep Learning Predictions via Randomized Experiments.**
  Jinkun Lin, Anqi Zhang*(Equal Contribution), Mathias Lécuyer, Jinyang Li, Aurojit Panda, Siddhartha Sen.*
  ICML 2022.
- **HOP: Heterogeneity-aware Decentralized Training.**
  Qinyi Luo, Jinkun Lin, Youwei Zhuo, Xuehai Qian.
  ASPLOS 2019.
- **Deep Online Video Stabilization With Multi-Grid Warping Transformation Learning.**
  Miao Wang, Guoye Yang, Jinkun Lin, Shaoping Lu, Ariel Shamir, Shimin Hu.
  IEEE Transactions on Image Processing 2019.
- **BiggerSelfie: Selfie Video Expansion with Hand-held Camera.**
  Miao Wang, Ariel Shamir, Guoye Yang, Jinkun Lin, Guowei Yang, Shaoping Lu, Shimin Hu.
  IEEE Transactions on Image Processing 2018.

Services

- Member of the Artifact Evaluation Committee (AEC) of EuroSys 2023

Awards & Honors

- Distinguished Artifact Award, ASPLOS 2023 *2023*
- KDD Cup 2018 Honorable Prize (8th place) *2018*
- Yixin Scholarship *2018*
- Zheng Geru Scholarship *2016, 2017*
- Freshmen Second Prize Scholarship *2015*
- The 31st China's National Olympiad in Informatics Gold Medal *2014*

Teaching Experience

- Undergraduate TA for Fundamentals of Programming (Fall 2015), Tsinghua University
- Undergraduate TA for Computer Systems Organization (Fall 2021, CSCI-UA.201.007), New York University

Other Research Experience

- **SenseTime Group Limited** Beijing, China
  Research Intern, *Computer Vision* *07/2017 – 09/2017*
  Advisor: Dr. Ding Liang
  - Augmented training images with ray-tracing-based algorithms for optical character recognition tasks.

- **Graphics and Geometric Computing Lab, *Tsinghua University*** Beijing, China
  Research Assistant, *Computer Vision, Computer Graphics* *09/2016 – 01/2018*
  Advisor: Prof. Shi-Min Hu
  - *BiggerSelfie*: Expanding the background in a selfie video with another background video.
  - *StabNet*: First deep learning approach for video stabilization.
    * Participated in the design and implementation of the model architecture.
    * Generated the dataset, implemented the code, and trained the network for the feature loss alignment loss function.