

# Kaizhao Liang

(650)7720082 | [kyleliang919@gmail.com](mailto:kyleliang919@gmail.com) | <https://github.com/kyleliang919>

## EDUCATION

---

### University of Illinois, Urbana Champaign

*Bachelor of Science in Computer Science (Highest Honors)*

Urbana, IL

Aug. 2016 – May 2020

## EXPERIENCE

---

### Senior Principal Software Engineer (ML)

*SambaNova Systems*

May 2024 – Now

Austin, TX

- Composition of Experts (Samba-1)
- Speculative Decoding for super fast inference
- Memory efficient Optimizers
- Long context reasoning

### Principal Software Engineer (ML)

*SambaNova Systems*

March 2022 – May 2024

Palo Alto, CA

- Developing and testing software frameworks for large scale neural network training, such as large language models (GPT3, GPT-Neox, Bloom and Llama), text-image pretraining (CLIP, Stable diffusion) as well as other vision models (ViT and ConvNext).
- Designing novel and efficient algorithms on specialized hardwares (Dataflow). (Structured sparsity for fast inference).
- Leading a cross-functional team ranging from compiler to data collection for research (publications) as well as product development

### Senior Software Engineer (ML)

*SambaNova Systems*

March 2021 – March 2022

Palo Alto, CA

- Mega High Resolution Convolutions Algorithms on Dataflow (100K x 100K)
- Sparsity is all you need (Pixelated Butterfly)

### Software Engineer (ML)

*SambaNova Systems*

June 2020 – March 2021

Palo Alto, CA

- Numerical testing and verification of basic modules for deep learning
- Model building and beta testing compiler stack

### Undergraduate Research Assistant

*Secure Learning Lab advised by Prof. Bo Li*

June 2018 – May 2020

Urbana Champaign, IL

- Investigating various aspects of deep neural networks, including adversarial training, adversarial robustness and attack, transfer learning
- Exploring improving intrinsic bias in convolution neural networks.

## PUBLICATIONS

---

### Composition of Experts on the SN40L Reconfigurable Dataflow Unit [IEEE Micro]

Raghu Prabhakar, Ram Sivaramakrishnan, Darshan Gandhi, Yun Du, Mingran Wang, Xiangyu Song, Kejie Zhang, Tianren Gao, Angela Wang, Karen Li, Yongning Sheng, Joshua Brot, Denis Sokolov, Apurv Vivek, Calvin Leung, Arjun Sabnis, Jiayu Bai, Tuowen Zhao, Mark Gottscho, David Jackson, Mark Luttrell, Manish K. Shah, Edison Chen, Kaizhao Liang\*, Swayambhoo Jain, Urmish Thakker, Dawei Huang, Sumti Jairath, Kevin J. Brown, Kunle Olukotun

### Memory-Efficient LLM Training with Online Subspace Descent [NeurIPS 2024]

Kaizhao Liang\*, Bo Liu, Lizhang Chen, Qiang Liu

### Communication Efficient Distributed Training with Distributed Lion [NeurIPS 2024]

Bo Liu, Lemeng Wu, Lizhang Chen, Kaizhao Liang\*, Jiaxu Zhu, Chen Liang, Raghuraman Krishnamoorthi, Qiang Liu

### Lion Secretly Solves Constrained Optimization: As Lyapunov Predicts [ICLR 2024]

Lizhang Chen, Bo Liu, Kaizhao Liang\*, Qiang Liu

## Simulating Disease Progression via Progressive Image Editing

Kaizhao Liang\*, Xu Cao, Kuei-Da Liao, Tianren Gao, Zhengyu Chen, Tejas Nama

## Uncovering the Connections Between Adversarial Transferability and Knowledge Transferability [ICML 2021]

Kaizhao Liang\*, Jacky. Y. Zhang\*, Boxin Wang, Zhuolin Yang, Sanmi Koyejo, Bo Li

## Unrestricted Adversarial Examples via Semantic Manipulation [ICLR 2020]

Anand Bhattad\*, Min Jin Chong\*, Kaizhao Liang, Bo Li, David A. Forsyth

## Adversarial Mutual Information for Text Generation [ICML 2020]

Boyuan Pan, Yazheng Yang, Kaizhao Liang, Bhavya Kailkhura, Zhongming Jin, Xian-Sheng Hua, Deng Cai, Bo Li

## Pixelated Butterfly: Simple and Efficient Sparse training for Neural Network Models [ICLR 2022]

Beidi Chen, Tri Dao, Kaizhao Liang, Jiaming Yang, Zhao Song, Atri Rudra, Christopher Re

## A Unified Knowledge Distillation Framework for Deep Directed Graphical Models [CVPR 2023]

Yizhuo Chen, Kaizhao Liang, Zhe Zeng, Shuochao Yao, Huajie Shao

## PATENTS (US)

---

### Lossless Tiling In Convolution Networks

Published Feb 2022

- TILING CONFIGURATION (US 11195080 B1)
- Section Boundary (US 11227207 B1)
- Read-Modify-Write in Backward Pass(US 11250061 B1)
- Weight Gradient Calculation(US 11232360 B1)

## WORK AUTHORIZATIONS (US)

---

I am currently on O1-A visa (Individuals with Extraordinary Ability or Achievement). To switch job, a transfer request I-129 needs to be filed by any new potential employers.