Hao Liu

KAUST, Thuwal, Saudi Arabia

hao.liu@kaust.edu.sa | liuhao-97.github.io | github.com/liuhao-97 | hao-liu-44686022b | Google Scholar | DBLP

Personal Profile

Hao Liu is a Ph.D candidate in the Accelerated Connected Computing Lab (ACCL, `https://cemse.kaust.edu.sa/accl`) at KAUST. He mainly works on optimizing DNN inference in the edge and near-edge scenarios. He also works on split computing, reconfigurable computing systems, and has explored machine learning model compression and physically unclonable functions.
He received his Bachelor's degree in Electronic and Information Engineering and his Master's degree in Cyber Science and Technology from Beihang University, Beijing, China in 2019 and 2021, respectively.

Education

King Abdullah University of Science and Technology (KAUST) *Thuwal, Saudi Arabia*
Ph.D student in Computer Science (Advisor: Prof. Suhaib A. Fahmy) *Jan 2022 - Present*

Beihang University (BUAA) *Beijing, China*
Master of Cyber Science and Technology (Advisor: Prof. Zhenyu Guan) *Sept 2019 - Jan 2022*

Beihang University (BUAA) *Beijing, China*
Bachelor of Electronic and Information Engineering *Sept 2015 - Jun 2019*

Publications

Ask the Expert: Collaborative Inference for Vision Transformers with Near-Edge Accelerators
Hao Liu, Suhaib A. Fahmy
under review, 2026

Practical Modeling for Split DNN Inference on Near-Edge Accelerators
Hao Liu, Mohammed E. Fouda, Ahmed M. Eltawil, Suhaib A. Fahmy
under review, 2026

Split DNN Inference for Exploiting Near-Edge Accelerators
Hao Liu, Mohammed E. Fouda, Ahmed M. Eltawil, Suhaib A. Fahmy
IEEE International Conference on Edge Computing and Communications (IEEE EDGE), 2024

A Filter Rank Based Pruning Method for Convolutional Neural Networks
Hao Liu, Zhenyu Guan, Peng Lei
IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), 2021

A Wearable Ad Hoc Device for Situational Awareness and Trusted Collaboration
Zhenyu Guan, Jiawei Li, **Hao Liu**, Dawei Li
Smart Blockchain: Second International Conference, 2019

Physical unclonable functions for IoT device authentication
Zhenyu Guan, **Hao Liu**, Yuyao Qin
Journal of Communications and Information Networks (2019). 2019

Research Experience

Collaborative Inference for Vision Transformers with Near-Edge Accelerators *Jeddah, Saudi Arabia*
KAUST (Advisor: Prof. Suhaib A. Fahmy) *Jun 2025 - Jan 2026*

- Propose a novel collaborative inference framework for ViTs with a general formulation, leveraging a lightweight edge model and multiple near-edge experts to enhance inference efficiency and accuracy.
- Design a robust routing mechanism that exploits the Top-k predictions of the edge model to determine expert activation without overhead.
- Propose a progressive specialist training strategy, which explicitly enforces expert specialization while preserving general accuracy.
- Experiments on CIFAR-100 dataset using a real-world edge and near-edge testbed.

Practical Modeling for Split DNN Inference on Near-Edge Accelerators *Jeddah, Saudi Arabia*
KAUST (Advisor: Prof. Suhaib A. Fahmy) *Jun 2024 - May 2025*

- Show experimentally that autoencoder-based communication compression offers higher accuracy and compression ratio than quantization and tensor low-rank decomposition.
- Propose a precise computation and communication formulation that matches measurements from practical deployments, considering system characteristics of GPU data ingestion and packet-based network data transfer.
- Build a real system comprising multiple accelerators and validate our proposed split DNN approach with the VGG16 and ResNet50 models on the CIFAR-100 dataset in a single-split setup with different batch sizes.
- Show use of the performance model to guide multi-split, dynamic bandwidth, and multi-tenant search scenarios.

Network Architecture Search for Split Computing *Beijing, China*
Tsinghua University Internship (Advisor: Prof. Yu Wang) *Jun 2023 - Sept 2023*

- Learn and extend an existing project of Network Architecture Search (NAS) aw_nas (`https://github.com/walkerning/aw_nas`)

Split Computing for Deep Neural Networks *Jeddah, Saudi Arabia*

KAUST (Advisor: Prof. Suhaib A. Fahmy) *Oct 2022 - May 2024*

- Formulate the problem of multi-split on DNNs into an optimization problem considering the accuracy, computing and transmission overhead.
- Propose to insert multiple autoencoders to split DNNs and execute different partitions of DNNs on different devices to reach the trade-off between latency and energy.
- Experiment on ResNet50 and VGG16 with CIFAR100 and ImageNet dataset.

A Filter Rank Based Pruning Method for Convolutional Neural Networks *Beijing, China*

Beihang University (Advisor: Prof. Zhenyu Guan) *Jun 2020 - Dec 2021*

Physical Unclonable Functions for IoT Device Authentication *Beijing, China*

Beihang University (Advisor: Prof. Zhenyu Guan) *Mar 2019 - Dec 2019*

A Wearable Ad Hoc Device for Situational Awareness and Trusted Collaboration *Beijing, China*

Beihang University (Advisor: Prof. Zhenyu Guan) *Sept 2018 - Feb 2019*

Teaching Experience

Saudi Ministry of Interior Program *Riyadh, Saudi Arabia*

Teaching Assistant for **ECE 265P: AI Training**. *April 2025*

KAUST *Jeddah, Saudi Arabia*

Teaching Assistant for **CS256: Digital Design and Computer Architecture** *Sept 2022*

Beihang University *Beijing, China*

Teaching Assistant for the **Digital Circuits and Systems** course. *Sept 2020*

Skills

Programming Verilog, Python (PyTorch), C Programming Language, MATLAB

English IELTS Overall Band 7 (Academic, L7.5 R8.5 W6.5 S6.0; taken on 7/31/2021)

Achievements

2019, 2020 **Second class,** The Graduate Academic Scholarships *Beijing, China*

2018 **Second Prize,** RMB 30,000 funding supported by Student Research Training Program *Beijing, China*

2018 **Second prize,** The 11th National College Student Information Security Contest (Captain) *Wuhan, China*

2018 **Second prize,** The 28th Feng Ru Cup Student Academic Competition of Beihang University *Beijing, China*

2017 **Third prize,** The 8th Lan Qiao Cup in MCU Design and Development Competition *Beijing, China*