

# YIFAN (EVELYN) GONG

(she / her / hers) ◊ 140 The Fenway, Boston, MA 02115

☎ (857)891-0509   ✉ gong.yifa@northeastern.edu   🌐 Personal Website   **in** Yifan Gong

## EDUCATION

---

### Northeastern University

Boston, MA

Ph.D. Candidate in Computer Engineering, advised by Prof. Yanzhi Wang

Sep 2019 – current

- With a focus on **energy-efficient deep learning and artificial intelligence systems, accelerations of deep neural networks including large-scale models for AIGC, trustworthy machine learning**

### University of Toronto

Toronto, ON, Canada

Master of Applied Science (Thesis-based with Fellowship)

Sep 2017 – Sep 2019

- With a focus on deep reinforcement learning and its applications

### Xidian University

Xi'an, Shaanxi, China

Bachelor of Engineering (**Valedictorian, with highest honor**), GPA: 3.83/4.0 (rank **1<sup>st</sup>**)

Sep 2013 – Jun 2017

Education Experimental Class (**Undergraduate honor program**)

## RESEARCH INTERESTS

---

- Hardware and Software Co-Design for Artificial Intelligence Acceleration
- Energy-Efficient Deep Learning and Artificial Intelligence Systems
- Accelerations of Emerging Large-Scale Models for AIGC such as Diffusion Models and LLMs
- Model Compression
- Efficient and Robust Deep Learning

## PUBLICATIONS

---

**Submitted**, † means equal contribution.

[I4] **Yifan Gong**, Zheng Zhan, Yanyu Li, Yerlan Idelbayev, Andrey Zharkov, Kfir Aberman, Sergey Tulyakov, Yanzhi Wang, Jian Ren, "Efficient Training with Denoised Model Weights", *under review*.

[I3] **Yifan Gong**, Zheng Zhan, Yushu Wu, Pu Zhao, Chao Wu, Minghai Qin, Caiwen Ding, Yanzhi Wang, "An Automatic Framework for Adaptive Deep Neural Network Deployment with DVFS on the Edge", *under review* for TCAD.

[I2] **Yifan Gong**<sup>†</sup>, Chao Wu<sup>†</sup>, Liangkai Liu<sup>†</sup>, et al, "AyE-Edge: Automated Deployment Space Search Empowering Accuracy yet Efficient Real-Time Object Detection on the Edge", *under review*.

[I1] Yuguang Yao<sup>†</sup>, Jiancheng Liu<sup>†</sup>, **Yifan Gong**<sup>†</sup>, Xiaoming Liu, Yanzhi Wang, Xue Lin, Sijia Liu, "Can Adversarial Examples Be Parsed to Reveal Victim Model Information?", *under review*.

**Conference Proceedings**, † means equal contribution.

[C17] **Yifan Gong**, Zheng Zhan, Qing Jin, Yanyu Li, Yerlan Idelbayev, Xian Liu, Andrey Zharkov, Kfir Aberman, Sergey Tulyakov, Yanzhi Wang, Jian Ren, "E<sup>2</sup>GAN: Efficient Training of Efficient GANs for Image-to-Image Translation", in **ICML** 2024. (**Acceptance rate: 27.5%**)

[C16] **Yifan Gong**, Yushu Wu, Zheng Zhan, Pu Zhao, Liangkai Liu, Chao Wu, Xulong Tang, Yanzhi Wang, "Lotus: learning-based online thermal and latency variation management for two-stage detectors on edge devices", in **DAC** 2024. (**Acceptance rate: 23%**)

- [C15] **Yifan Gong**<sup>†</sup>, Yushu Wu<sup>†</sup>, Zheng Zhan, Geng Yuan, Yanyu Li, Qi Wang, Chao Wu, Yanzhi Wang, "MOC: Multi-Objective Mobile CPU-GPU Co-optimization for Power-efficient DNN Inference", in **ICCAD** 2023. (Acceptance rate: 22.9%)
- [C14] **Yifan Gong**, Pu Zhao, Zheng Zhan, Yushu Wu, Chao Wu, Zhenglun Kong, Minghai Qin, Caiwen Ding, Yanzhi Wang, "Condense: A Framework for Device and Frequency Adaptive Neural Network Models on the Edge", in **DAC** 2023. (Acceptance rate: 23%)
- [C13] **Yifan Gong**, Zheng Zhan, Pu Zhao, Yushu Wu, Chao Wu, Caiwen Ding, Weiwen Jiang, Minghai Qin, Yanzhi Wang, "All-in-One: A Highly Representative DNN Pruning Framework for Edge Devices with Dynamic Power Management", in **ICCAD** 2022. (Acceptance rate: 22.5%)
- [C12] **Yifan Gong**, Yuguang Yao, Yize Li, Yimeng Zhang, Xiaoming Liu, Xue Lin, Sijia Liu, "Reverse Engineering of Imperceptible Adversarial Image Perturbations", in **ICLR** 2022. (Acceptance rate: 32.2%)
- [C11] **Yifan Gong**<sup>†</sup>, Yushu Wu<sup>†</sup>, Pu Zhao, Yanyu Li, Zheng Zhan, Wei Niu, Hao Tang, Minghai Qin, Bin Ren, Yanzhi Wang, "Compiler-Aware Neural Architecture Search for On-Mobile Real-time Super-Resolution", in **ECCV** 2022. (Acceptance rate: 28%)
- [C10] **Yifan Gong**<sup>†</sup>, Zheng Zhan<sup>†</sup>, Pu Zhao, Geng Yuan, Wei Niu, Yushu Wu, Tianyun Zhang, Malith Jayaweera, David Kaeli, Bin Ren, Xue Lin, Yanzhi Wang, "Achieving on-Mobile Real-Time Super-Resolution with Neural Architecture and Pruning Search", in **ICCV** 2021. (Acceptance rate: 25.9%)
- [C9] **Yifan Gong**, Zheng Zhan, Zhengang Li, Wei Niu, Xiaolong Ma, Wenhao Wang, Bin Ren, Caiwen Ding, Xue Lin, Xiaolin Xu, Yanzhi Wang, "A Privacy-Preserving-Oriented DNN Pruning and Mobile Acceleration Framework", in **GLSVLSI** (invited) 2020.
- [C8] **Yifan Gong**, Baochun Li, Ben Liang, Zheng Zhan, "Chic: Experience-driven Scheduling in Machine Learning Clusters", in **IWQoS** 2019.
- [C7] Peiyan Dong, Zhenglun Kong, Xin Meng, Pinrui Yu, **Yifan Gong**, Geng Yuan, Hao Tang, Yanzhi Wang, "HotBEV: Hardware-oriented Transformer-based Multi-View 3D Detector for BEV Perception", in **NeurIPS** 2023.
- [C6] Zifeng Wang, Zheng Zhan, **Yifan Gong**, Yucai Shao, Stratis Ioannidis, Yanzhi Wang, Jennifer Dy, "DualHSIC: HSIC-Bottleneck and Alignment for Continual Learning", in **ICML** 2023.
- [C5] Sizhe Chen, Geng Yuan, Xinwen Cheng, **Yifan Gong**, Minghai Qin, Yanzhi Wang, Xiaolin Huang, "Self-Ensemble Protection: Training Checkpoints Are Good Data Protectors", in **ICLR** 2023.
- [C4] Xiaolong Ma, Geng Yuan, Zhengang Li, **Yifan Gong**, Tianyun Zhang, Wei Niu, Zheng Zhan, Pu Zhao, Ning Liu, Jian Tang, Xue Lin, Bin Ren, Yanzhi Wang, "Blcr: Towards Real-Time DNN Execution with Block-based Reweighted Pruning", in **ISQED** 2022.
- [C3] Zifeng Wang, Zheng Zhan, **Yifan Gong**, Geng Yuan, Wei Niu, Tong Jian, Bin Ren, Stratis Ioannidis, Yanzhi Wang, Jennifer Dy, "Sparcl: Sparse continual learning on the edge", in **NeurIPS** 2022.
- [C2] Geng Yuan, Xiaolong Ma, Wei Niu, Zhengang Li, Zhenglun Kong, Ning Liu, **Yifan Gong**, Zheng Zhan, Chaoyang He, Qing Jin, Siyue Wang, Minghai Qin, Bin Ren, Yanzhi Wang, Sijia Liu, Xue Lin, "Mest: Accurate and fast memory-economic sparse training framework on the edge", in **NeurIPS** 2021.
- [C1] Peiyan Dong, Siyue Wang, Wei Niu, Chengming Zhang, Sheng Lin, Zhengang Li, **Yifan Gong**, Bin Ren, Xue Lin, Dingwen Tao, "RTMobile: Beyond Real-Time Mobile Acceleration of RNNs for Speech Recognition", in **DAC** 2020.

## Journal Papers

- [J3] Yuguang Yao, Xiao Guo, Vishal Asnani, **Yifan Gong**, Jiancheng Liu, Xue Lin, Xiaoming Liu, Sijia Liu, "Reverse Engineering of Deceptions on Machine- and Human-Centric Attacks", Foundations and Trends® in Privacy and Security, 2024.

[J2] **Yifan Gong**, Geng Yuan, Zheng Zhan, Wei Niu, Zhengang Li, Pu Zhao, Yuxuan Cai, Sijia Liu, Bin Ren, Xue Lin, Xulong Tang, Yanzhi Wang, "Automatic Mapping of the Best-Suited DNN Pruning Schemes for Real-Time Mobile Acceleration", ACM Transactions on Design Automation of Electronic Systems (**TODAES**), 2021.

[J1] Tong Jian, **Yifan Gong**, Zheng Zhan, Runbin Shi, Nasim Soltani, Zifeng Wang, Jennifer Dy, Kaushik Chowdhury, Yanzhi Wang, Stratis Ioannidis, "Radio Frequency Fingerprinting on the Edge", IEEE Transactions on Mobile Computing (**TMC**), 2021.

## Patents

[P1] Yanzhi Wang, **Yifan Gong**, Zheng Zhan, Computer-implemented methods and systems for privacy-preserving deep neural network model compression, US Patent App. 17/176,340, 2021.

## INTERNSHIP

---

### Snap Inc.

Santa Monica, CA

#### *Ph.D. Research Intern @ Creative Vision Group*

May 2023 – Aug 2023

- *Project: Model Generation with Knowledge from Diffusion Models*

Mentor: Jian Ren, Sergey Tulyakov

Content: Worked on efficient distillation of GANs from diffusion models (**E<sup>2</sup>GAN**).

- Proposed a novel knowledge transfer framework to train efficient GANs with knowledge from diffusion models
- Created a large-scale weight vector dataset for weight generator training in the scale of millions
- Built model weight generation pipeline

### IBM Research

Cambridge, MA

#### *Ph.D. Research Intern @ MIT-IBM Watson AI Lab*

May 2021 – Aug 2021

- *Project: Improving Vision Transformers by Attention Graph*

Mentor: Quanfu Fan

Content: Worked on improving the performance of vision transformers by incorporating the interpretability of an image with structural information.

## RESEARCH EXPERIENCE

---

### Northeastern University

Boston, MA

#### *Research Assistant advised by Prof. Yanzhi Wang @ College of Engineering*

Sep 2019 – present

- *Project: Mobile CPU-GPU Co-Optimization for Power-efficient DNN Inference*

Feb 2023 – present

Content: Worked on power-efficient DNN inference by optimizing CPU-GPU configurations including active CPU core, CPU frequency, and GPU frequency (**ICCAD-23**)

- Developed demand-resource matching-based models (feed-forward ceiling model and CPU-memory ceiling model) to classify the DNNs into five categories
- Proposed a DRL approach based on the model category information to derive CPU-GPU configurations for different stages of diverse DNN inference

- *Project: Effective Compression-DVFS Co-design*

Feb 2022 – present

Content: Worked on reducing the runtime variation of DNNs on edge devices under dynamic power management with DVFS (**DAC-23**, **ICCAD-22**)

- Developed a framework to get multiple subnets in one DNN to reduce latency variation for different hardware frequency levels with DVFS (**ICCAD-22**)
- Proposed a two-level algorithm for obtaining subnets with arbitrary ratios in a single model with theoretical proof for a more automatic framework that works for arbitrary devices (**DAC-23**)

- *Project: Intelligent Diagnosis for Machine and Human-Centric Adversaries*

Jan 2021 – Mar 2023

Content: Explored a new adversarial learning paradigm-Reverse Engineering of Deceptions (**ICLR-22**).

- Formulated the Reverse Engineering of Deceptions (RED) problem to estimate adversarial perturbations and provided the feasibility of inferring the adversary intention
- Identified a series of RED principles and built a comprehensive evaluation pipeline
- **Recognized and valued by the community, we had the privilege of hosting the CVPR'23 tutorial on Reverse Engineering of Deceptions (RED) based on my two works on RED against machine-centric attacks**
- *Project: Compression-Compilation Co-design (CoCoPIE)* Feb 2020 – present  
Content: Optimizing AI models for the implementation on edge devices ([ICCV-21](#), [ECCV-22](#)).
  - Worked on achieving real-time super-resolution on mobile platform, we are **the first** to achieve real-time super-resolution inference for implementing 720p resolution with competitive image quality on mobile platforms

**University of Toronto**

Toronto, ON, Canada

**Research Assistant advised by Prof. Baochun Li @ Department of ECE**

Sep 2017 – Sep 2019

- *Project: Scheduling Machine Learning Jobs with Reinforcement Learning*  
Content: Proposed a scheduler to find the scheduling decision for distributed machine learning workloads to minimize the average completion time based on reinforcement learning ([IWQoS-19](#)).
  - Modeled the scheduling problem for reinforcement learning agent and simulated the results to compare with SOTA methods

TEACHING EXPERIENCES

---

**Teaching Assistant of Advances in Deep Learning @ Northeastern University**

- Taught a 2-hour course about how to use SOTA deep learning frameworks such as Pytorch and Tensorflow
- Prepared course materials and final project topics
- Held office hours to address the questions of students in the class
- Assigned and graded homework, quiz, and final projects

**Teaching Assistant of Operating System and Computer Fundamentals @ University of Toronto**

- Gave lab demonstrations to students
- Marked and graded tests and exams

INVITED TALKS

---

**Tutorials**

[T4] "Reverse Engineering of Deceptions: Foundations and Applications", @ [CVPR'23](#).

**Invited Seminars**

[T3] "Automatic Mapping of the Best-Suited DNN Pruning Schemes for Real-Time Mobile Acceleration", in ROAD4NN @ [DAC'21](#).

[T2] "A Privacy-Preserving-Oriented DNN Pruning and Mobile Acceleration Framework", @ [GLSVLSI'20](#).

[T1] "Towards Best Possible Deep Learning Acceleration on the Edge - A Compression-Compilation Co-Design Framework", in MGHPCC @ [SC'20](#).

PROFESSIONAL SERVICES

---

**Conference Reviewer:** ICLR'23, NeurIPS'23, ICCV'23, CVPR'23, ISCAS'23, AICAS'23, AdvML'22

**Journal Reviewer:** TCAD

SELECTED SCHOLARSHIP, HONORS AND AWARDS

---

CVPR Travel Grant	04/2024
DAC Young Fellow	04/2024
ICCAD Student Scholar Award	09/2023
College of Engineering <b>Outstanding TA Awards</b> of Northeastern University	04/2023
College of Engineering <b>Dean's Fellowship</b> of Northeastern University	2019-2020
ECE Student Fellowship of University of Toronto	2017-2019
<b>Valedictorian</b> of Xidian University	06/2017
<b>Excellent Graduate</b> of Xidian University ( <b>10 of 5180</b> )	06/2017
Goodix Scholarship for Science and Technology	12/2016
<b>National Scholarship (1%)</b>	10/2015, 10/2016
<b>Special Prize</b> in National English Competition for College Students ( <b>0.1%</b> )	05/2016
<b>Role Model Outstanding Student</b>	11/2014, 11/2015
<b>Provincial 1<sup>st</sup> Prize</b> in CUMCM	11/2015

## REFERENCES

---

### **Dr. Yanzhi Wang**

Associate Professor, Faculty Fellow  
Northeastern University  
yanz.wang@northeastern.edu

### **Dr. David R. Kaeli**

COE Distinguished Professor  
Northeastern University  
kaeli@ece.neu.edu

### **Dr. Stratis Ioannidis**

Professor  
Northeastern University  
ioannidis@ece.neu.edu

### **Dr. Xiaoming Liu**

MSU Foundation Professor  
Michigan State University  
liuxm@cse.msu.edu

### **Dr. Jennifer Dy**

Professor, EAI director  
Northeastern University  
jdy@ece.northeastern.edu

### **Dr. Baochun Li**

Professor, Associate Chair  
University of Toronto  
bli@ece.toronto.edu