

# YIFAN (EVELYN) GONG

140 The Fenway, Boston, MA 02115

☎ (857)891-0509    ✉ gong.yifa@northeastern.edu

## EDUCATION

---

### Northeastern University

Boston, MA

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

Sep 2019 – current

- With a focus on **Efficient and Trustworthy Machine Learning Systems**

### 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**)

## EXPERIENCE

---

### SnapInc.

Santa Monica, CA

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

May 2023 – Aug 2023

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

Content: Worked on building a model weight generation framework to transfer knowledge from diffusion models to lightweight GANs (**GAN-Adapter**).

- Proposed a novel knowledge transfer framework where the generated images from diffusion models are utilized to train a GAN that can run efficiently on devices
- Built model weight generation pipeline based on diffusion models

### IBM Research

Cambridge, MA

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

May 2021 – Aug 2021

- *Project: Improving Vision Transformers by Attention Graph*

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

### Northeastern University

Boston, MA

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

Sep 2019 – present

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

Feb 2022 – present

Content: Worked on reducing 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 neural network 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 SR 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

## SELECTED PUBLICATIONS

---

**In Submission**, <sup>†</sup> means equal contribution.

[3] Chao Wu<sup>†</sup>, **Yifan Gong**<sup>†</sup>, Liangkai Liu<sup>†</sup>, Yushu Wu, et al, "PEFO: a Power Efficient Framework for Real-Time Object Detection on the Edge", *under review*.

[2] **Yifan Gong**, Zheng Zhan, Qing Jin, et al, "GAN-Adapter: Efficient GAN Adapting for Image Translation with Knowledge from Diffusion Models", *under review*.

[1] Yuguang Yao<sup>†</sup>, Jiancheng Liu<sup>†</sup>, **Yifan Gong**<sup>†</sup>, et al, "Can Adversarial Examples Be Parsed to Reveal Victim Model Information?", *under review*.

**Conference Proceedings**, <sup>†</sup> means equal contribution.

[C7] Yushu Wu<sup>†</sup>, **Yifan Gong**<sup>†</sup>, Zheng Zhan et al, "MOC: Multi-Objective Mobile CPU-GPU Co-optimization for Power-efficient DNN Inference", **ICCAD** 2023.

[C6] **Yifan Gong**, Pu Zhao, Zheng Zhan, et al, "Condense: A Framework for Device and Frequency Adaptive Neural Network Models on the Edge", in **DAC** 2023. (**Acceptance rate: 23%**)

[C5] **Yifan Gong**, Zheng Zhan, Pu Zhao, et al, "All-in-One: A Highly Representative DNN Pruning Framework for Edge Devices with Dynamic Power Management", in **ICCAD** 2022. (**Acceptance rate: 22.5%**)

[C4] **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%**)

[C3] Yushu Wu<sup>†</sup>, **Yifan Gong**<sup>†</sup>, Pu Zhao, et al, "Compiler-Aware Neural Architecture Search for On-Mobile Real-time Super-Resolution", in **ECCV** 2022. (**Acceptance rate: 28%**)

[C2] Zheng Zhan<sup>†</sup>, **Yifan Gong**<sup>†</sup>, Pu Zhao, et al, "Achieving on-Mobile Real-Time Super-Resolution with Neural Architecture and Pruning Search", in **ICCV** 2021. (**Acceptance rate: 25.9%**)

[C1] **Yifan Gong**, Baochun Li, Ben Liang, Zheng Zhan, "Chic: Experience-driven Scheduling in Machine Learning Clusters", in **IWQoS** 2019.

## Journal Papers

[J2] **Yifan Gong**, Geng Yuan, et al, "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**, et al, "Radio Frequency Fingerprinting on the Edge", *IEEE Transactions on Mobile Computing*, 2021.

## 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](#).

## REVIEW SERVICES AND SKILLS

---

Review services: NeurIPS'23, ICCV'23, CVPR'23, ISCAS'23, AICAS'23, AdvML'22, TCAD'22

Research interests: Model Compression, Computer Vision, Robust and Efficient Deep Learning

## SELECTED SCHOLARSHIP, HONORS AND AWARDS

---

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
<b>National Scholarship (1 %)</b>	10/2015, 10/2016
<b>Role Model Outstanding Student</b>	11/2014, 11/2015