

# Hongxu (Danny) Yin

✉ danny@nvidia.com • 🌐 <https://hongxu-yin.github.io/>  
Google Scholar

## Education

---

### Princeton University

Ph.D. in Electrical Computer Engineering  
Research focus: Efficient and Secure Deep Learning  
Advisor: Prof. Niraj K. Jha

New Jersey, USA  
2015 - 2020

### Nanyang Technological University

B.Eng in Electronic & Electronics Engineering (GPA 3.9/4.0, dean's lister all four years)  
Minor in Business (GPA 4.0/4.0)  
Advisor: Prof. Bah Hwee Gwee and Prof. Zhiping Lin

Singapore, SG  
2011 - 2015

### University of California, Berkeley

Undergraduate summer exchange

California, USA  
2012

### University of Cambridge

High school elite exchange program

Cambridge, UK  
2007

## Academic Experience

---

### NVIDIA, Learning and Perception Research (LPR)

Senior Research Scientist (Team lead: Dr. Jan Kautz)

May 2022 - Now

### NVIDIA, Learning and Perception Research (LPR)

Research Scientist (Team lead: Dr. Jan Kautz)

May 2020 - Apr 2022

### NVIDIA, Learning and Perception Research (LPR)

Research Intern (Mentor: Dr. Pavlo Molchanov and Dr. Jan Kautz)

May 2019 - Nov 2019

### Alibaba U.S., Machine Learning Team

Research Intern (Mentor: Dr. Weifeng Zhang)

May 2018 - Nov 2018

## Selected Awards

---

- 36 Kr Top 100 Global Outstanding Chinese Awards 2022
- Forbes Top 60 Elite Chinese North America 2021
- Princeton ECE Best Dissertation Award Finalist (Top-3 in department) 2020
- Princeton Yan Huo \*94 Fellowship (Top-3 in department) 2019
- Princeton Natural Science and Foundation Fellowship 2015-2017
- Gold Medal - Defense Science and Technology 2015
- Gold Medal - Thomas Asia Pacific Holdings 2015
- Department Dean's Lister Award 2011-2015
- Nanyang Best Industrial Orientation Award 2014
- Nanyang Presidential Scholar with Highest Distinction 2012-2015

## Conference Publications

---

(\*: equal contribution; †: advised intern)

34. Divyam Madaan<sup>†</sup>, **Hongxu Yin**, Wonmin Byeon, Jan Kautz, Pavlo Molchanov  
*Heterogeneous continual learning*

Conference on Computer Vision and Pattern Recognition (CVPR), 2023  
(**Highlight - top 2.5% paper**)

33. Huanrui Yang<sup>†</sup>, **Hongxu Yin**, Pavlo Molchanov, Hai Li, Jan Kautz  
*NViT: Vision transformer compression and parameter redistribution*  
Conference on Computer Vision and Pattern Recognition (CVPR), 2023
32. Paul Micaelli<sup>†</sup>, Pavlo Molchanov, Arash Vahdat, **Hongxu Yin**, Jan Kautz  
*Recurrence without recurrence: stable video landmark detection with deep equilibrium models*  
Conference on Computer Vision and Pattern Recognition (CVPR), 2023
31. Xin Dong<sup>†</sup>, **Hongxu Yin**, Jose Alvarez, Jan Kautz, Pavlo Molchanov  
*Privacy vulnerability of split computing to data-free model inversion attacks*  
British Machine Vision Conference (BMVC), 2022
30. Maying Shen\*, **Hongxu Yin\***, Pavlo Molchanov, Lei Mao, Jianna Liu, Jose Alvarez  
*Structural pruning via latency-saliency Knapsack*  
Advances in Neural Information Processing Systems (NeurIPS), 2022
29. **Hongxu Yin**, Arash Vahdat, Jose Alvarez, Arun Mallya, Jan Kautz, Pavlo Molchanov  
*A-ViT: Adaptive tokens for efficient vision transformer*  
Conference on Computer Vision and Pattern Recognition (CVPR), 2022  
(**Oral Presentation**)
28. Ali Hatamizadeh\*, **Hongxu Yin\***, Holger Roth, Wenqi Li, Jan Kautz, Daguang Xu, Pavlo Molchanov  
*GradViT: Gradient inversion of vision transformers*  
Conference on Computer Vision and Pattern Recognition (CVPR), 2022
27. Maying Shen, Pavlo Molchanov, **Hongxu Yin**, Jose Alvarez  
*When to prune? A policy towards early structural pruning*  
Conference on Computer Vision and Pattern Recognition (CVPR), 2022
26. Xin Dong<sup>†</sup>, **Hongxu Yin**, Jose Alvarez, Jan Kautz, Pavlo Molchanov  
*Deep neural networks are surprisingly reversible: A baseline for zero-shot inversion*  
Conference on Computer Vision and Pattern Recognition Workshop (CVPR Workshop), 2022
25. Ali Hatamizadeh\*, **Hongxu Yin\***, Holger Roth, Wenqi Li, Jan Kautz, Daguang Xu, Pavlo Molchanov  
*Gradient inversion of vision transformers*  
Conference on Computer Vision and Pattern Recognition Workshop (CVPR Workshop), 2022
24. Pavlo Molchanov\*, Jimmy Hall\*, **Hongxu Yin\***, Jan Kautz, Nicolo Fusi, Arash Vahdat  
*HANT: Hardware-aware network transformation*  
European Conference on Computer Vision (ECCV), 2022
23. **Hongxu Yin**, Arun Mallya, Arash Vahdat, Jose Alvarez, Jan Kautz, Pavlo Molchanov  
*See through gradients: Image batch recovery via GradInversion*  
Conference on Computer Vision and Pattern Recognition (CVPR), 2021
22. Yerlan Idelbayev<sup>†</sup>, Pavlo Molchanov, Maying Shen, **Hongxu Yin**, M. C. Perpinan, Jose Alvarez  
*Optimal quantization using scaled codebook*  
Conference on Computer Vision and Pattern Recognition (CVPR), 2021
21. Akshay Chawla<sup>†</sup>, **Hongxu Yin**, Pavlo Molchanov, Jose Alvarez  
*Data-free knowledge distillation for object detection*  
Winter Conference on Applications of Computer Vision (WACV), 2021
20. **Hongxu Yin**, Arun Mallya, Arash Vahdat, Jose Alvarez, Jan Kautz, Pavlo Molchanov  
*Dreaming to distill: Data-free knowledge transfer via DeepInversion*  
Conference on Computer Vision and Pattern Recognition (CVPR), 2020  
(**Oral Presentation**)

19. Wenhan Xia, **Hongxu Yin**, Niraj K. Jha  
*Efficient synthesis of compact deep neural networks*  
IEEE Design Automation Conference (DAC), 2020
18. Xiaoliang Dai, Peizhao Zhang, Bichen Wu, **Hongxu Yin**, Fei Sun, Yanghan Wang, Marat Dukhan, Yunqing Hu, Yiming Wu, Yangqing Jia, Peter Vajda, Matt Uyttendaele, Niraj K. Jha  
*ChamNet: Towards efficient network design through platform-aware model adaptation*  
Conference on Computer Vision and Pattern Recognition (CVPR), 2019
17. Ozge Akmandor, **Hongxu Yin**, and Niraj K. Jha  
*Simultaneously ensuring smartness, security, and energy efficiency in Internet-of-Things sensors*  
IEEE Custom Integrated Circuits Conference (CICC), 2017
16. **Hongxu Yin**, Bah Hwee Gwee, Zhiping Lin, Kumar Anil, Galul R. Sirajudeen, and Choo M. S. See  
*Novel real-time system design for floating-point sub-Nyquist multi-coset signal blind reconstruction*  
IEEE Int. Symp. on Circuits and Systems (ISCAS), 2015  
(**Oral Presentation**)

## Journal Publications

---

15. Ali Hatamizadeh, **Hongxu Yin**, Pavlo Molchanov, Andriy Myronenko, Wenqi Li, Prerna Dogra, Andrew Feng, Mona G Flores, Jan Kautz, Daguang Xu, Holger R. Roth  
*Do gradient inversion attacks make federated learning unsafe?*  
IEEE Transactions on Medical Imaging, 2023
14. Shayan Hassantabar, Joe Zhang, **Hongxu Yin**, Niraj K. Jha  
*MHDeep: Mental health disorder detection system based on body-area and deep neural networks*  
ACM Transactions on Embedded Computing Systems, 2022
13. **Hongxu Yin**, Guoyang Chen, Yingmin Li, Shuai Che, Weifeng Zhang, and Niraj K. Jha  
*Hardware-guided symbiotic training for compact, accurate, yet execution-efficient LSTMs*  
IEEE Trans. Emerging Topics in Computing, 2021
12. Wenhan Xia, **Hongxu Yin**, Xiaoliang Dai, Niraj K. Jha  
*Fully dynamic inference with deep neural networks*  
IEEE Trans. Emerging Topics in Computing, 2021
11. Xiaoliang Dai\*, **Hongxu Yin\***, and Niraj K. Jha  
*Grow and prune compact, fast, and accurate LSTMs*  
IEEE Trans. Computers, 2020
10. **Hongxu Yin**, Bilal Mukadam, Xiaoliang Dai, and Niraj K. Jha  
*DiabDeep: Pervasive diabetes diagnosis based on wearable medical sensors and efficient neural networks*  
IEEE Trans. Emerging Topics in Computing, 2020
9. Xiaoliang Dai, **Hongxu Yin**, and Niraj K. Jha  
*Incremental learning using a grow-and-prune paradigm with efficient neural networks*  
IEEE Trans. Computers, 2020
8. Xiaoliang Dai, **Hongxu Yin**, and Niraj K. Jha  
*NeST: A neural network synthesis tool based on a grow-and-prune paradigm*  
IEEE Trans. Computers, 2019
7. **Hongxu Yin**, Zeyu Wang, and Niraj K. Jha  
*A hierarchical inference model for Internet-of-Things*  
IEEE Trans. Multi-scale Computing Systems, 2018
6. **Hongxu Yin** and Niraj K. Jha  
*A health decision support system for disease diagnosis based on wearable medical sensors and machine learning*

*ensembles*

IEEE Trans. Multi-scale Computing Systems, 2017

5. Ozge Akmandor, **Hongxu Yin** and Niraj K. Jha  
*Smart, secure, yet energy-efficient, Internet-of-Things sensors*  
IEEE Trans. Multi-scale Computing Systems, 2017

## Book Chapter

---

4. **Hongxu Yin**, Ozge Akmandor, Arsalan Mosenia, and Niraj K. Jha  
*Smart healthcare*  
Foundations and Trends, 2017

## Preprint (publicly available & under review)

---

3. Xinglong Sun<sup>†</sup>, Maying Shen, **Hongxu Yin**, Lei Mao, Pavlo Molchanov, Jose M Alvarez  
*Towards dynamic sparsification by iterative prune-grow lookAheads*  
preprint, 2023
2. Ali Hatamizadeh, **Hongxu Yin**, Jan Kautz, Pavlo Molchanov  
*Global context vision transformer*  
preprint, 2022
1. Zhen Dong<sup>†</sup>, **Hongxu Yin**, Arash Vahdat, Jan Kautz, Pavlo Molchanov  
*Efficient transformation of architectures through hardware-aware nonlinear optimization*  
preprint, 2022

## Workshop & Tutorial Organizer

---

- *Full-Stack, GPU-based Acceleration of Deep Learning*  
CVPR 2023 2023
- *Transformers for Vision*  
CVPR 2022 2022

## Invited Keynote & Talk

---

- *Efficient Deep Learning*  
Invited Panelist, Open Compute Project (OCP) Global Summit Oct. 2022
- *Towards Efficient and Secure Deep Learning*  
Invited Keynote, Design Automation Conference (DAC'60) Jul. 2022
- *Towards Efficient and Secure Deep Nets*  
University of British Columbia ECE Department May 2022
- *Inverting Deep Nets*  
Princeton University, Department of Computer Science research groups Aug. 2021
- *See through Gradients*  
Europe ML meeting Apr. 2021
- *Dreaming to Distill*  
Synced AI (largest AI media in Asia) Jul. 2020
- *Dreaming to Distill*  
Facebook AR/VR Jun. 2020
- *Making Neural Networks Efficient*  
Alibaba Cloud / Platform AI group Feb. 2020

- *Efficient Neural Networks*  
NVIDIA Research, Facebook Research Dec. 2019
- *Efficient Neural Networks*  
Baidu Research, ByteDance A.I. Lab US Dec. 2019
- *Efficient Neural Networks*  
Alibaba A.I. Research, Kwai Lab Nov. 2019
- *Applied Machine Learning: From Theory to Practice*  
Invited Keynote, IEEE Circuits and Systems Society (Singapore Chapter) Feb. 2018
- *A Health Decision Support System for Disease Diagnosis*  
New Jersey Tech Council Jun. 2016

## Patents (till Jun. 2022)

---

25. *Pruning Neural Networks*  
NVIDIA 2022
24. *Neural Network Training Technique*  
NVIDIA 2022
23. *Techniques to Identify Data used to Train One or More Neural Networks*  
NVIDIA 2022
22. *Pruning Vision Transformers under Latency Budget and a Method to Distribute Parameters across Layers*  
NVIDIA 2022
21. *GradViT: Gradient Inversion of Vision Transformers*  
NVIDIA 2022
20. *Adaptive Token Depth Adjustment Algorithm for Networks with Transformer Blocks*  
NVIDIA 2022
19. *Global Context Model for Transformer Neural Networks*  
NVIDIA 2022
18. *Towards Understanding the Risks of Gradient Inversion in Federated Learning*  
NVIDIA 2022
17. *When to Prune? A Policy for Early Structural Pruning*  
NVIDIA 2021
16. *See Through Gradients: Image Batch Recovery via GradInversion*  
NVIDIA 2021
15. *Network similarity metric as a Pruning Indicator*  
NVIDIA 2021
14. *Zero-shot Model Inversion for Data-free Distillation*  
NVIDIA 2021
13. *MHDeep: Mental Health Disorder Detection System based on Body-Area and Deep Neural Networks*  
Princeton University 2019
12. *Optimal MSE Quantization with Fixed Codebook and Rescaling*  
NVIDIA 2020
11. *Dreaming Data for Continual Learning*  
NVIDIA 2020
10. *Data-Free Knowledge Distillation for Object Detection*  
NVIDIA 2020

9. <i>Hardware-aware Latency Neural Network Pruning</i> NVIDIA	2020
8. <i>Image Generation for Data Free Pruning</i> NVIDIA	2019
7. <i>Hardware-guided Symbiotic Training for Compact, Accurate, yet Execution-efficient LSTMs</i> Alibaba	2019
6. <i>Incremental Learning using a Grow-and-prune Paradigm with Efficient Neural Networks</i> Princeton University	2019
5. <i>DiabDeep: Pervasive Diabetes Diagnosis based on Wearable Medical Sensors and Efficient Neural Networks</i> Princeton University	2019
4. <i>Smart, Secure, yet Energy-efficient Internet-of-Things Sensors</i> Princeton University	2019
3. <i>NeST: A Neural Network Synthesis Tool based on a Grow-and-prune Paradigm</i> Princeton University	2018
2. <i>Grow and Prune Compact, Fast, yet Accurate LSTMs</i> Princeton University	2018
1. <i>A Hierarchical Health Decision support System based on Wearable Medical Sensors and Machine Learning Ensembles</i> Princeton University	2017

## Academic Services

---

### Teaching Assistant - Princeton University

ELE 364, Machine Learning for Predictive Data Analytics

Fall, 17-18

ELE464, Embedded Computing

Spring, 16-17

### Conference Reviewer & Committee

Computer Vision and Pattern Recognition (CVPR)

International Conference on Computer Vision (ICCV)

Conference on Neural Information Processing Systems (NeurIPS)

International Conference on Machine Learning (ICML)

European Conference on Computer Vision (ECCV)

British Machine Vision Conference (BMVC)

Winter Conference on Applications of Computer Vision (WACV)

AAAI Conference on Artificial Intelligence (AAAI)

Design Automation Conference (DAC)

High-Performance Computer Architecture (HPCA)

### Journal Reviewer & Committee

IEEE Transactions on Pattern Analysis and Machine Intelligence

IEEE Transactions on Neural Networks and Learning Systems

International Journal of Computer Vision

IEEE Journal of Biomedical and Health Informatics

IEEE Journal of Selected Topics in Signal Processing

IEEE Sensors Journal

IEEE Consumer Electronics Magazine

International Journal on Artificial Intelligence Tools

International Journal of Systems Architecture

International Journal of Healthcare Technology and Management

International Journal of Electronic Imaging

# Mentorship

---

## Princeton Senior Thesis Mentees

Joe Zhang, now Ph.D. at Stanford	2019-2020
Hari Santhanam, now Ph.D. at University of Pennsylvania	2019-2020
Frederick Hertan, now at SIG Trading	2018-2019
Kyle Johnson, now at Princeton University	2018-2019
Bilal Mukadam, now at Microsoft	2018-2019
Chloe Song, now at Astra Inc.	2017-2018

## NVIDIA Research Mentees

Huanrui Yang, Duke University	2021-2022
Zhen Dong, University of California, Berkeley	2021-2022
Xin Dong, Harvard University	2021-2022
Paul Micaelli, University of Edingburgh	2021-2022
Yerlan Idelbayev, University of California, Merced	2020-2021
Vu Nguyen, Stony Brooks University	2020-2021
Akshay Chawla, Carnegie Mellon University	2020-2021
Divyam Madaan, New York University	2022-now
Shixing Yu, University of Texas, Austin	2022-now
Annamarie Bair, Carnegie Mellon University	2022-now
Alex Sun, University of Illinois Urbana-Champaign	2022-now