

Xin Yu

CONTACT INFORMATION

Kahlert School of Computing
The University of Utah
Salt Lake City, UT, 84112

Phone: +1 8018158442
E-mail: yuxwind@gmail.com
Homepage: <https://yuxwind.github.io/>

RESEARCH INTERESTS

My research domain centers around developing **efficient vision representation** for large-scale scenes and **inference optimization** for large models (like LLMs and generative image modeling), thus to achieve **less computation/carbon, less memory, less data, less bias** especially for edge devices, such as robots, embodied AI. Some topics of interest are:

- **Topics:** 3D vision, including 3D reconstruction and SLAM for self-driving; Image generation, including mLLM, auto-regressive generative image models; inference optimization, including neural network pruning and quantization, sparse neural network expressivity; Efficient data, active learning, time series learning, and streaming tensor decomposing, etc.
- **Tools:** PyTorch, Scikit-learn, OpenCV, HuggingFace Transformers, Tensorflow, Matlab, c/c++.

EDUCATION

The University of Utah, Salt Lake City, Utah

Aug 2018 - Aug 2024

Ph.D. Student, Computer Science

- Dissertation Topic: "Efficient Pruning Algorithms for Deep Neural Networks"
- Advisor: Shandian Zhe, Srikumar Ramalingam

The University of Utah, Salt Lake City, Utah,

Aug 2016 - Aug 2018

M.S., Image Analysis, and transited to Ph.D program

EXPERIENCES

Chinese University of Hong Kong, Shenzhen(CUHKSZ), China

Mar 2025 - now

Visiting Researcher @ GAP (Generation and Analysis of Pixels, Points and Polygons) Lab

Develop continuous 3D perception including 3D reconstruction, and semantic segmentation simultaneously from images/videos for large-scale scenes.

National University of Singapore (NUS), Singapore

Jul 2024 - Feb 2025

Visiting Researcher @ Computer Vision and Robotic Perception (CVRP) Laboratory

Develop functional-space 3D scene representation with 3DGS for surface reconstruction and total scene understanding.

University of Utah, SLC, UT

Sep 2016 - Jun 2024

Research Assistant @ Vision group and Data group, School of Computing

Design, implement, and conduct experiments in vision, network pruning, and Bayesian learning.

Mitsubishi Electric Research Laboratories, Boston, MA

May 2020 - Aug 2020

Research Intern @ Vision group

Design and develop 3D face and body reconstruction.

Mitsubishi Electric Research Laboratories, Boston, MA

May 2019 - Nov 2019

Research Intern @ Vision group

Design and develop joint 3D human shape recovery and pose estimation from a single image with bilayer graph.

Highlight projects:

On 3D Vision Perception, Reconstruction and SLAM

- **Single view reconstruction: Joint 3D body reconstruction and Pose Estimation**
 - To improve both training efficiency and model performance, I proposed and built a dual-scale graph approach: a coarse graph to estimate the human's 3D pose and the dense graph to estimate

the 3D shape, where information about pose can guide to recover local shape detail and vice versa, which can converge at least four times faster than transformers. [3DV'21]

- **VLASE: Vehicle localization by aggregating semantic edges**
- To achieve a fast and light human-like location recognition, I proposed and built sparse semantic edge features instead of full-resolution images for image-based localization [10], leading to a state-of-the-art on-road localization framework [IROS'18].
- **Mapping of Sparse 3D Data using Alternating Projection**
- To achieve extreme sparse point cloud registration without correspondence and texture information, I proposed and built highly sparse point cloud registration by redefining it through line segments, where as few as seven pairs of intersections suffice. In Kinect, this methodology with data down-sampled by 100x more yet still outperforms methods that rely on full-resolution data [ACCV'20].

On Inference Optimization

- **Optimization for pruning: combinatorial importance of individual weights**
- The dependency between weights are important but no pruning research explored it before due to huge search space.
- I proposed and built effective post-training network pruning by removing multiple weights simultaneously with combinatorial optimization methods, outperforming the mainstream pruning methods without considering a combined effect, especially under high sparsity [ICML'22].
- **Light architecture search by pruning**
- To search optimal depths for ResNet, I designed ϵ -ResNet that allows us to achieve optimal depths by starting training super deep ResNet and automatically discarding redundant layers during training, which produces responses below a threshold ϵ , without any loss in performance [CVPR'18].
- **Level up interpretability: expressivity-based pruning algorithms**
- To verify redundancy exists in NN theoretically, I applied Mixed-Integer Linear Programming techniques to identify stable neurons for a rectifier network and get a smaller yet exact equivalent network from the large model, capable of producing identical outputs for any input. This lossless pruning ratio can be as high as 40% on a simple rectifier network trained on Cifar10. [NeurIPS'21].

On Efficient data: Abstract Representation for time-series data and Active Learning

- **Temporal decomposition algorithms to distill key features from time-series data**
- Expanded the Tucker decomposition for continuous-indexed data with the Functional Bayesian Tucker Decomposition (FunBaT) which treats continuous-indexed data as the interaction between the Tucker core and a set of latent functions, offering a more generalized solution [ICLR'24].
- **Active learning for multi-fidelity data with high acquisition costs**
- Built a probabilistic multi-resolution Fourier Neural Operator (FNO) and implemented an effective posterior inference algorithm using ensemble Monte Carlo methods [AISTATS'24].

PUBLICATIONS
(*: EQUAL
CONTRIBUTION)

Li Z., Liu S., **Yu X.**, Bhavya K., Cao J., Daniel D., Bremer P. and ascucci V. **Understanding Robustness Lottery: A Comparative Visual Analysis of Neural Network Pruning Approaches.** ((*TVCG 2024*)).

Cai J., Nguyen K., Shrestha N., Good A., Tu R., **Yu X.**, & Serra T. **Getting away with more network pruning: From sparsity to geometry and linear regions.** *International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research (CPAIOR 2023)*.

Yu X.*, Serra T.*, Ramalingam S., Zhe S. **The Combinatorial Brain Surgeon: Pruning Weights That Cancel One Another in Neural Networks.** In *International Conference on Machine Learning (ICML 2022)*.

Good A.*, Lin J.*, **Yu X.***, Sieg H., Ferguson M., Zhe S., Wiecezore J., & Serra T. **Recall Distortion in Neural Network Pruning and the Undecayed Pruning Algorithm**. *Advances in Neural Information Processing Systems (NeurIPS 2022)*

Serra T., **Yu X.**, Kumar A., Ramalingam S. **Scaling Up Exact Neural Network Compression by ReLU Stability**. *Advances in Neural Information Processing Systems (NeurIPS 2021)*

Yu X., Yu Z., & Ramalingam S.(2018 June). **Learning strict identity mappings in deep residual networks**. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018)*.

Li S., **Yu X.**, Xing W., Kirby M., Narayan A., & Zhe S. **Multi-Resolution Active Learning of Fourier Neural Operators** *27th International Conference on Artificial Intelligence and Statistics(AISTATS 2024)*.

Fang S., **Yu X.**, Wang Z., Li S., Kirby R., Zhe S. **Functional Bayesian Tucker Decomposition for Continuous-indexed Tensor Data**. *Twelfth International Conference on Learning Representation(ICLR 2024)*.

Fang, S., **Yu, X.**, Li, S., Wang, Z., Kirby R., & Zhe, S. **Streaming Factor Trajectory Learning for Temporal Tensor Decomposition**. *Advances in Neural Information Processing Systems (NeurIPS 2023)* .

Li, S., Phillips, J. M., **Yu, X.**, Kirby, R., & Zhe, S. **Batch Multi-Fidelity Active Learning with Budget Constraints**. *Advances in Neural Information Processing Systems (NeurIPS 2022)*.

Yu X., & Baar J, Chen S. **Joint 3D Human Shape Recovery and Pose Estimation from A Single Image with Bilayer-Graph**. *in International Conference on 3D Vision (3DV 2021)*.

Ranade S.*, **Yu X.***, Kakkar K., Miraldo P., & Ramalingam S. **Mapping of Sparse 3D Data using Alternating Projection**. *in Proceedings of the Asian Conference on Computer Vision (ACCV 2020)*.

Yu X.*, Sagar C.*, Feng C., Taguchi Y., Lee T., Fernandes C., & Ramalingam S. **Vlase: Vehicle localization by aggregating semantic edges**. *International Conference on Intelligent Robots and Systems (IROS 2018)*.

Paul D., Li F., Teja M., **Yu X.**, Frost R. **Compass: Spatio temporal sentiment analysis of US election what twitter says!** *in Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining (KDD 2017)*.

POSTERS AND PRESENTATIONS

The Combinatorial Brain Surgeon: Pruning Weights That Cancel One Another in Neural Networks
The Mixed Integer Programming Workshop (MIPs), Rutgers University, Newark, NJ, 2022

Recall Distortion in Neural Network Pruning and the Undecayed Pruning Algorithm
Sparsity in Neural Networks(SNN) workshop, virtual, 2022

Joint 3D Human Shape Recovery and Pose Estimation from A Single Image with Bilayer-Graph
Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, 2019

Learning strict identity mappings in deep residual networks
College of Engineering, University of Idaho, Idaho Falls, ID, 2018

ACADEMIC SERVICES

Conference Reviewer

NeurIPS 2025, ICML 2025, ICML 2024, ICLR 2024, ICML 2023, NeurIPS 2023, NeurIPS 2022, IROS 2021, ICVGIP 2021

TEACHING

The University of Utah
Teaching Mentorships

- **CS6190 Probability Modeling** (Shandian Zhe) Spring 2023
- **CS6320 Computer Vision** (Tucker Hermans) Spring 2020
- **CS6320 Computer Vision** (Srikumar Ramalingam) Spring 2019
- **CS6320 Computer Vision** (Srikumar Ramalingam) Spring 2018