# MINGYU YANG

Homepage: mingyuyng.github.io
(+1)7342728559 ◇ mingyuy@umich.edu

## EDUCATION

**University of Michigan, Ann Arbor, MI, USA** *Seq 2019 - Dec 2024*
Doctor of Philosophy in Electrical and Computer Engineering GPA: 4.0/4.0
Advisor: *Prof. Hun-Seok Kim*

**University of Michigan, Ann Arbor, MI, USA** *Sep 2017 - Apr 2019*
Master of Science in Electrical and Computer Engineering GPA: 4.0/4.0
Major: Signal & Image Processing and Machine Learning

**Beijing University of Technology, Beijing, China**
**University College Dublin, Dublin, Ireland** *Sep 2013 - June 2017*
Bachelor of Engineering in Internet of Things. GPA: 4.19/4.2

## WORK EXPERIENCE

**AMD, San Jose, CA** *Jan 2025 - present*
*MTS Software Application Eng.*

- Acceleration and optimization on large-scale LLM training and inference.
- Research on efficient LLM architecture and KV cache compression.

**Samsung Research America, Mountain View, CA** *May 2024 - August 2024*
*Research Intern - AI Center*

- Explored time series foundation models (TSFM) and their usage in time series classification.
- Performed multiple fine-tuning techniques (e.g., Linear Probing, Full Fine-tuning, LoRA, etc) on multiple cutting-edge transformer-based TSFMs such as Moment, UniTS, and Chronos.

**Meta, Seattle, WA** *May 2022 - August 2022*
*PhD Software Engineer Intern*

- Worked on ML solutions for BM Abuse and Compromise Detection using user activity sequences.
- Developed the first end-to-end sequential model for BM compromise detection using CNN-based TIES model and outperformed the baseline (frequency of grams) by 57% and 187% in AUROC and AUPRC.
- Proposed the first learning-based method to interpret the importance of different business activities using a two-layer Transformer.

## PUBLICATIONS

19. Diffusion-Aided Joint Source Channel Coding For High Realism Wireless Image Transmission
*IEEE Transactions on Machine Learning in Communications and Networking (TMLCN) 2025*
Mingyu Yang, Bowen Liu, Boyang Wang, Hun-Seok Kim

18. Zebra-Llama: Towards Extremely Efficient Hybrid Models
*Neural Information Processing Systems (NeurIPS) 2025*
Mingyu Yang*, Mehdi Rezagholizadeh*, Guihong Li*, Vikram Appia, Emad Barsoum

17. X-EcoMLA: Upcycling Pre-Trained Attention into MLA for Efficient and Extreme KV Compression
*Conference on Language Modeling (COLM) 2025*
Guihong Li*, Mehdi Rezagholizadeh*, Mingyu Yang*, Vikram Appia, Emad Barsoum

16. NBLoc: A Narrowband RF Localization System for Wide-area Indoor Applications
*IEEE Transactions on Mobile Computing (TMC) 2025*
Demba Komma, Chien-Wei Tseng, Andrea Bejarano-Carbo, Mingyu Yang, et al.

15. SAM-guided Pseudo Label Enhancement for Multi-modal 3D Semantic Segmentation
*International Conference on Robotics and Automation (ICRA) 2025*
Mingyu Yang, Jitong Lu, Hun-Seok Kim

14. H-PCC: Point Cloud Compression with Hybrid Mode Selection and Content Adaptive Downsampling
*IEEE Robotics and Automation Letters (RAL) 2025*
Bowen Liu, Yu Chen, Boyang Wang, Mingyu Yang, Hun-Seok Kim

13. Efficient Computation Sharing for Multi-Task Visual Scene Understanding
*International Conference on Computer Vision (ICCV) 2023*
Sara Shoouri, Mingyu Yang, Zichen Fan, Hun-Seok Kim

12. Search for Efficient Deep Visual-Inertial Odometry Through Neural Architecture Search
*IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2023*
Yu Chen, Mingyu Yang, Hun-Seok Kim

11. Efficient Deep Visual and Inertial Odometry with Adaptive Visual Modality Selection
*European Conference on Computer Vision (ECCV) 2022*
Mingyu Yang, Yu Chen, Hun-Seok Kim

10. Siamese Learning-based Monarch Butterfly Localization
*IEEE Data Science and Learning Workshop (DSLW) 2022*
Sara Shoouri, Mingyu Yang, Gordy Carichner, et al.

9. Deep Joint Source Channel Coding for Wireless Image Transmission with Adaptive Rate Control
*IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2022*
Mingyu Yang, Hun-Seok Kim

8. Tracking the Migration of the Monarch Butterflies with the World's Smallest Computer
*GetMobile: Mobile Computing and Communications, 2022*
Inhee Lee, Roger Hsiao, Gordy Carichner, Chin-Wei Hsu, Mingyu Yang, et al.

7. OFDM-guided Deep Joint Source Channel Coding for Wireless Multipath Fading Channels
*IEEE Transactions on Cognitive Communications and Networking (TCCN), 2022*
Mingyu Yang, Chenghong Bian, Hun-Seok Kim

6. Deep Learning Based Near-Orthogonal Superposition Code for Short Message Transmission
*IEEE International Conference on Communications (ICC) 2022*
Chenghong Bian, Mingyu Yang, Chin-Wei Hsu, Hun-Seok Kim

5. Deep Joint Source Channel Coding for Wireless Image Transmission with OFDM
*IEEE International Conference on Communications (ICC) 2021*
Mingyu Yang, Chenghong Bian, Hun-Seok Kim

4. mSAIL: Milligram-Scale Multi-Modal Sensor Platform for Monarch Butterfly Migration Tracking
*International Conference On Mobile Computing And Networking (Mobicom) 2021*
Inhee Lee, Roger Hsiao, Gordy Carichner, Chin-Wei Hsu, Mingyu Yang, et al.

3. Super-Resolution Time-of-Arrival Estimation using Neural Networks
*European Signal Processing Conference (EUSIPCO) 2020*
Mingyu Yang*, Yao-Shan Hsiao*, Hun-Seok Kim

2. Migrating Monarch Butterfly Localization Using Multi-Modal Sensor Fusion Neural Networks
*European Signal Processing Conference (EUSIPCO) 2020*

Mingyu Yang, Roger Hsiao, Gordy Carichner, Katherine Ernst, et al.

1. iLPS: Local Positioning System with Simultaneous Localization and Wireless Communication
*IEEE International Conference on Computer Communications (INFOCOM) 2019*
Mingyu Yang, Li-Xuan Chuo, Karan Suri, Lu Liu, Hao Zheng, Hun-Seok Kim

## PATENTS

”Low-Power, Long-Range RF Localization System And Method”, Application US16654547

## REVIEWER SERVICE

IEEE Journal on Selected Areas in Communications (JSAC)
IEEE Transactions on Wireless Communications (TWC)
IEEE Transactions on Communications (TCOM)
IEEE Transactions on Cognitive Communications and Networking (TCCN)
IEEE Transactions on Green Communications and Networking (TGCN)
IEEE Wireless Communications Letters (WCL)
IEEE Communications Letters (CL)
IEEE Transactions on Mobile Computing (TMC)
IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)
IEEE Transactions on Signal Processing (TSP)
IEEE Signal Processing Letters (SPL)
IEEE Transactions on Vehicular Technology (TVT)
IEEE Robotics and Automation Letters (RAL)
Neural Information Processing Systems (NeurIPS) ’23’24’25
International Conference on Learning Representations (ICLR) ’24’25
International Conference on Machine Learning (ICML) ’24’25
Association for the Advancement of Artificial Intelligence (AAAI) ’24’25’26
Computer Vision and Pattern Recognition Conference (CVPR) ’24’25’26
International Conference on Computer Vision (ICCV) ’25
IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) ’24’25
The International Conference on Robotics and Automation (ICRA) ’24’25
IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) ’25

## PROJECT EXPERIENCE

| | |
|---|---|
| Point Cloud Completion with Transformer | *2021* |
| Drift-Aware Predictive Coding for Adaptation in Changing Environments | *2020* |
| Classify MEG signals into Musicians and Non-Musicians using graph-based CNN | *2019* |
| Semantic Image Inpainting with Generative Models | *2018* |
| A Map Construction Robot Based on ORB-SLAM | *2018* |

## TEACHING ASSISTANT

| | |
|---|---|
| Beijing University of Technology, EEEN3003J, Signals and Systems | *2017* |
| Beijing University of Technology, EEEN3006J, Communication Theory | *2017* |
| Beijing University of Technology, COMP2003J, Data Structure and Algorithms | *2017* |

## ACHIECEMENTS

| | |
|---|---|
| Beijing University of Technology, Best 10 Graduates | *2017* |
| Beijing University of Technology, President Scholarship (10/27000) | *2016* |
| Beijing University of Technology, National Scholarship (Top 1%) | *2016* |

Beijing University of Technology, Kitagawa Scholarship (Top 5%) *2014 - 2016*

Beijing University of Technology, University-level Science and Technology Practice Award *2016*