Zhuoran Yu

Madison, WI, USA | zhuoran.yu@wisc.edu | www.zhuoranyu.com | Google Scholar

Education

Ph.D. in Computer Science Sep 2021 – May 2026 (Expected)
University of Wisconsin–Madison, Madison, WI
Advisor: Yong Jae Lee

M.S. in Computer Science Aug 2018 – Aug 2020
Georgia Institute of Technology, Atlanta, GA
GPA: 4.0/4.0

B.Math in Computer Science and Applied Mathematics Jan 2014 – Jun 2018
University of Waterloo, Waterloo, Canada
Graduate with Distinction & Dean's Honor List
Arthur Beaumont Memorial Scholarship

Technical Expertise

Multimodal LLM Training

- One of the initiators of IBM's next-generation Granite Vision model.
- Investigated multimodal data mixture strategies for early-fusion pretraining (Chameleon).

Evaluation & Benchmarking

- Currently leading the development of audiovisual and chart/document understanding benchmarks focusing on tasks that require strong multimodal reasoning and integration.

Publications & Preprints

* denotes equal contribution

Under Review

See, Hear, and Understand: Benchmarking Audiovisual Human Speech Understanding in Multimodal Large Language Models
Le Thien Phuc Nguyen*, **Zhuoran Yu***, Samuel Low Yu Hang, Subin An, Jeongik Lee, Yohan Ban, SeungEun Chung, Thanh-Huy Nguyen, JuWan Maeng, Soochahn Lee, Yong Jae Lee
Under review.

Revisiting Active Speaker Detection: An In-the-Wild Benchmark for Generalization and Robustness
Le Thien Phuc Nguyen*, **Zhuoran Yu***, Khoa Quang Nhat Cao, Yuwei Guo, Tu Ho Manh Pham, Tuan Tai Nguyen, Toan Ngo Duc Vo, Lucas Poon, Soochahn Lee, Yong Jae Lee
Under review.

Large Language Model Teaches Visual Students: Cross-Modality Transfer of Conceptual Knowledge
Thomas Liang*, **Zhuoran Yu***, Yong Jae Lee
Under review.

Tracing the Effects of Multimodal Data Mixture
Zhuoran Yu, Bhargavi Paranjape, Yong Jae Lee, Srini Iyer
Under review.

Published

How Multimodal LLMs Solve Image Tasks: A Lens on Visual Grounding, Task Reasoning, and Answer Decoding
Zhuoran Yu, Yong Jae Lee
*Conference on Language Modeling (***COLM***)*, Montreal, Canada, 2025.

LASER: Lip Landmark Assisted Speaker Detection for Robustness
Le Thien Phuc Nguyen*, **Zhuoran Yu***, Yong Jae Lee
*IEEE/CVF Winter Conference on Applications of Computer Vision (***WACV***)*, Tucson, Arizona, 2026.

CuRe: Cultural Gaps in the Long Tail of Text-to-Image Systems
Aniket Rege, Zinnia Nie, Mahesh Ramesh, Unmesh Raskar, **Zhuoran Yu**, Aditya Kusupati, Yong Jae Lee, Ramya Korlakai Vinayak
*IEEE/CVF International Conference on Computer Vision (***ICCV***)*, Honolulu, Hawaii, 2025.

Diversify, Dont Fine-Tune: Scaling Up Visual Recognition Training with Synthetic Images
Zhuoran Yu, Chenchen Zhu, Sean Culatana, Raghuraman Krishnamoorthi, Fanyi Xiao, Yong Jae Lee
*Transactions on Machine Learning Research (***TMLR***)*, 2025. Featured Certificate.

Denoising and Selecting Pseudo-Heatmaps for Semi-Supervised Human Pose Estimation
Zhuoran Yu*, Manchen Wang*, Yanbei Chen, Paolo Favaro, Davide Modolo
*IEEE/CVF Winter Conference on Applications of Computer Vision (***WACV***)*, Waikoloa, Hawaii, 2024.

InPL: Pseudo-labeling the Inliers First for Imbalanced Semi-Supervised Learning
Zhuoran Yu, Yin Li, Yong Jae Lee
*International Conference on Learning Representations (***ICLR***)*, Kigali, Rwanda, 2023.

Group R-CNN for Weakly Semi-supervised Object Detection with Points
Shilong Zhang*, **Zhuoran Yu***, Liyang Liu*, Xinjiang Wang, Aojun Zhou, Kai Chen
*IEEE/CVF Conference on Computer Vision and Pattern Recognition (***CVPR***)*, New Orleans, 2022.

Scale-Equalizing Pyramid Convolution for Object Detection
Xinjiang Wang*, Shilong Zhang*, **Zhuoran Yu**, Litong Feng, Wei Zhang
*IEEE/CVF Conference on Computer Vision and Pattern Recognition (***CVPR***)*, Virtual, 2020.

Experience

Research Scientist Intern May 2025 – Aug 2025
IBM Watson AI Lab, Cambridge, MA

- Led the initial effort on Granite-4 Vision development.

Research Scientist Intern May 2024 – Aug 2024
Meta (FAIR), Language and Communication Team, Seattle, WA

- Explored budget-constrained solutions for data mixture practices in early-fusion multimodal LLMs.

Research Scientist Intern May 2023 – Aug 2023
Meta, CORE AI Team, Menlo Park, CA

- Improved recognition models with synthetic data from pre-trained diffusion models.

Applied Scientist Intern May 2022 – Aug 2022
AWS AI Rekognition, Seattle, WA

- Developped semi-supervised learning approaches for 2D human pose estimation.

Research Intern May 2019 – Dec 2019
SenseTime Research, Shenzhen, China

- Conducted research on efficient deep learning to reduce latency of neural networks.
- Worked on object detection for both industrial and academic applications.