Zhuoran Yu

San Francisco, CA, USA | zhuoran.yu336@gmail.com | www.zhuoranyu.com | Google Scholar

Education

Ph.D. in Computer Science Sep 2021 – May 2026
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
Graduated with Distinction & Dean's Honor List
Arthur Beaumont Memorial Scholarship

Research and Technical Expertise

Multimodal and Generative Foundation Models: Universal foundation models for 3D generation from heterogeneous visual and geometric inputs [E1]; and vision-language models for document understanding and reasoning [E2, P7].

Data-Centric Model Development: Multimodal data mixture design [E2, E3]; synthetic data generation [E4, P6, P10]; and semi-supervised learning [E5, P11, P12].

Evaluation and Model Analysis: Document and audiovisual benchmark development [P1, P3, P4]; robustness analysis for active speaker detection [P3, P5]; and multimodal LLM interpretability [P8].

Experience

[E1] Senior AI Research Scientist Jun 2026 – Present
Autodesk Research, San Francisco, CA

- Conduct research on a universal foundation model for B-rep generation conditioned on images, depth maps, sketches, point clouds, voxels, and combinations of these modalities.
- Develop multimodal learning and generation methods that unify heterogeneous visual and geometric inputs for controllable B-rep generation.

[E2] Research Scientist Intern May 2025 – Aug 2025
IBM Watson AI Lab, Cambridge, MA

- Contributed to the development of Granite vision-language models, focusing on multimodal data mixture design, document understanding, and reasoning over information-dense documents.

[E3] Research Scientist Intern May 2024 – Jan 2025
Meta FAIR, Language and Communication Team, Seattle, WA / Remote

- Investigated data mixture strategies for early-fusion multimodal large language models through controlled model training and analysis across visual and language tasks.

[E4] Research Scientist Intern May 2023 – Nov 2023
Meta, CORE AI Team, Menlo Park, CA / Remote

- Developed synthetic-data-based approaches for improving visual recognition models using pre-trained diffusion models, with a focus on scaling synthetic data without task-specific generative-model fine-tuning.

[E5] Applied Scientist Intern May 2022 – Aug 2022
AWS AI Rekognition, Seattle, WA

- Developed semi-supervised learning methods for 2D human pose estimation using denoised pseudo-heatmaps and uncertainty-guided pseudo-label selection under limited-label settings.

Publications

* denotes equal contribution

[P1] DocHop: Benchmarking Out-of-domain Multi-hop Reasoning in Information-Dense Documents
Zhuoran Yu, Le Thien Phuc Nguyen, Jaden Park, Xinyi Gu, Zexue He, Soochahn Lee, Rogerio Feris, Yong Jae Lee
*International Conference on Machine Learning (***ICML***)*, Seoul, South Korea, 2026.

[P2] Large Language Model Teaches Visual Students: Cross-Modality Transfer of Fine-Grained Conceptual Knowledge
Thomas Shih-Chao Liang*, **Zhuoran Yu***, Yong Jae Lee
*International Conference on Machine Learning (***ICML***)*, Seoul, South Korea, 2026.

[P3] 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, Tuan Khai Nguyen, Soochahn Lee, Yong Jae Lee
INTERSPEECH, Sydney, Australia, 2026. **Oral Presentation**.

[P4] 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
*Findings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (***CVPR Findings***)*, Denver, Colorado, 2026.

[P5] 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. **Oral Presentation**.

[P6] Composition-Grounded Instruction Synthesis for Visual Reasoning
Xinyi Gu, Jiayuan Mao, Zhang-Wei Hong, **Zhuoran Yu**, Pengyuan Li, Dhiraj Joshi, Rogerio Feris, Zexue He
*International Conference on Learning Representations (***ICLR***)*, Rio de Janeiro, Brazil, 2026.

[P7] DAVE: A VLM Vision Encoder for Document Understanding and Web Agents
Brandon Huang, Hang Hua, **Zhuoran Yu**, Trevor Darrell, Rogerio Feris, Roei Herzig
*International Conference on Learning Representations (***ICLR***)*, Rio de Janeiro, Brazil, 2026.

[P8] 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.

[P9] 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.

[P10] Diversify, Don't 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**.

[P11] 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.

[P12] 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.

[P13] 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.

[P14] 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.