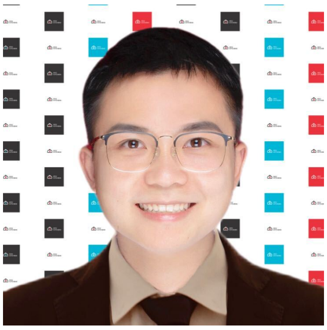
Bin ZHU

School of Computing and Information Systems
Singapore Management University (SMU)
80 Stamford Road
Singapore 178902

Email: binzhu@smu.edu.sg



Education

PhD, City University of Hong Kong, China, 2021

Master of Engineering, Zhejiang University, China, 2018

Bachelor of Engineering, Southeast University, China, 2015

Academic Appointments

Assistant Professor of Computer Science, School of Computing and Information Systems, SMU, Jan 2024 - Present

RESEARCH

Research Interests

My research interest lies in Human Centered Multimedia Analysis, including Cross-modal Search and Creation, Egocentric Video Understanding, Multi-modal Large Language Model and AI for Healthcare. Specifically, the objective is to conduct frontier research and develop cutting-edge technologies for processing, modeling, analyzing, and understanding multimedia content, that facilitate natural and immersive human experience and exert positive impact for our society.

Research and Project Areas

Multimedia Analysis
Generative AI
Egocentric Video Understanding
Multi-modal Large Language Model
AI for Healthcare

Publications

Journal Articles [Refereed]

CookingDiffusion: Cooking procedural image generation with Stable Diffusion, by WANG, Yuan; ZHU, Bin;

HAO, Yanbin; NGO, Chong-Wah; TAN, Yi; WANG, Xiang. (2025). *ACM Transactions on Multimedia Computing, Communications and Applications,* 1-23. https://doi.org/10.1145/3771995 (Advance Online)

Large lithium-ion battery model for secure shared e-bike battery in smart cities, by DING, Donghui; LI, Zhao; LUO, Linhao; JIN, Ming; ZHU, Bin; ZHONG, Yichen; HU, Junhao; CAI, Peng; HU, Huiqi. (2025). *Nature Communications, 16* (8415), 1-12. https://www.nature.com/articles/s41467-025-63678-7 (Published)

FoodLMM: A versatile food assistant using large multi-modal model, by YIN, Yuehao; QI, Huiyan; ZHU, Bin; CHEN, Jingjing; JIANG, Yu-Gang; NGO, Chong-Wah. (2025). *IEEE Transactions on Multimedia,* 1-38. https://doi.org/10.48550/arXiv.2312.14991 (Advance Online)

Text-driven video prediction, by SONG, Xue; CHEN, Jingjing; ZHU, Bin; JIANG, Yu-gang. (2024). *ACM Transactions on Multimedia Computing, Communications and Applications, 20* (9), 1-15. https://doi.org/10.1145/3675171 (Published)

From canteen food to daily meals: generalizing food recognition to more practical scenarios, by LIU, Guoshan; JIAO, Yang; CHEN, Jingjing; ZHU, Bin; JIANG, Yu-Gang. (2024). *IEEE Transactions on Multimedia, 27* 2724-2733. https://doi.org/10.1109/TMM.2024.3371212 (Published)

Efficient unsupervised video hashing with contextual modeling and structural controlling, by DUAN, Jingru; HAO, Yanbin; ZHU, Bin; CHENG, Lechao; ZHOU, Pengyuan; WANG, Xiang. (2024). *IEEE Transactions on Multimedia, 26* 1-13. https://doi.org/10.1109/TMM.2024.3368924 (Advance Online)

Learning from web recipe-image pairs for food recognition: Problem, baselines and performance, by ZHU, Bin; NGO, Chong-Wah; CHAN, Wing-Kwong. (2022). *IEEE Transactions on Multimedia, 24* 1175-1185. https://doi.org/10.1109/TMM.2021.3123474 (Published)

Learning to match anchor-target video pairs with dual attentional holographic networks, by HAO, Yan Bin; NGO, Chong-Wah; ZHU, Bin. (2021). *IEEE Transactions on Image Processing, 30* 8130-8143. (Published)

A study of multi-task and region-wise deep learning for food ingredient recognition, by CHEN, Jingjing; ZHU, Bin; NGO, Chong-Wah; CHUA, Tat-Seng; JIANG, Yu-Gang. (2021). *IEEE Transactions on Image Processing, 30* 1514-1526. https://doi.org/10.1109/TIP.2020.3045639 (Published)

Conference Proceedings

Look before you decide: Prompting active deduction of MLLMs for assumptive reasoning, by LI, Yian; TIAN, Wentao; JIAO, Yang; CHEN, Jingjing; QIAN, Tianwen; ZHU, Bin; ZHAO, Na; JIANG, YuGang. (2025.0). *MM '25: The 33rd ACM International Conference on Multimedia, Dublin Ireland, October 27-31,* (pp. 2713-2722) New York: ACM. https://doi.org/10.1145/3746027.3754720 (Published)

Exploring object status recognition for recipe progress tracking in non-visual cooking, by LI, Franklin Mingzhe; NG, Kaitlyn; ZHU, Bin; CARRINGTON, Patrick. (2025.0). *ASSETS '25: Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility, Denver, Colorado USA, October 26-29,* (pp. 1-15) New York : ACM. https://doi.org/10.1145/3663547.3746318 (Published)

From holistic to localized: Local enhanced adapters for efficient visual instruction fine-tuning, by JIAO, Pengkun; ZHU, Bin; CHEN, Jingjing; NGO, Chong-Wah; JIANG, Yugang. (2025.0). *Proceedings of the 2025 International Conference on Computer Vision, Honolulu, Hawaii, October 19-23,* (pp. 1-10) Honolulu, Hawai'i, USA: (Published)

Efficient prompt tuning for hierarchical ingredient recognition, by GUI, Yinxuan; ZHU, Bin; CHEN, Jingjing; NGO, Chong-Wah. (2025.0). *Proceedings of the 2025 IEEE International Conference on Multimedia and Expo (ICME 2025), Nantes, France, June 30 - July 4,* (pp. 1-6) Piscataway, NJ: IEEE. https://doi.org/10.48550/arXiv.2504.10322 (Presented)

Advancing food nutrition estimation via visual-ingredient feature fusion, by QI, Huiyan; ZHU, Bin; NGO, Chong-Wah; CHEN, Jingjing; LIM, Ee-Peng. (2025.0). *Proceedings of the 2025 International Conference on Multimedia Retrieval, Chicago, IL, USA, June 30 - July 3 ,* (pp. 1091-1099) New York: ACM. https://doi.org/10.1145/3731715.3733269 (Published)

HD-EPIC: A highly-detailed egocentric video dataset, by PERRETT, Toby; DARKHALIL, Ahmad; SINHA, Saptarshi; EMARA, Omar; POLLARD, Sam; PARIDA, Kranti Kumar; LIU, Kaiting; GATTI, Prajwal; BANSAL, Siddhant; FLANAGAN, Kevin; CHALK, Jacob; ZHU, Zhifan; GUERRIER, Rhodri; ABDELAZIM, Fahd; ZHU, Bin; MOLTISANTI, Davide; WRAY, Michael; DOUGHTY, Hazel; DAMEN, Dima. (2025.0). *Proceedings of the 2025*

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, June 10-17 , (pp. 23901-23913) Piscataway, NJ: IEEE. https://doi.org/10.1109/CVPR52734.2025.02226 (Published)

OSCAR: Object status and contextual awareness for recipes to support non-visual cooking, by LI, Franklin Mingzhe; NG, Kaitlyn; ZHU, Bin; CARRINGTON, Patrick. (2025.0). *CHI EA '25: Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, April 26 - May 1,* (pp. 1-6) New York: ACM. https://doi.org/10.1145/3706599.372017 (Published)

Retrieval augmented recipe generation, by LIU, Guoshan; YIN, Hailong; ZHU, Bin; CHEN, Jingjing; NGO, Chong-Wah; JIANG, Yu-Gang. (2025.0). *IEEE/CVF Winter Conference on Applications of Computer Vision 2025, Tucson, Arizona, February 28 - March 4,* (pp. 1-11) Tucson, Arizona: (Accepted)

RAGG: Retrieval-augmented grasp generation model, by TANG, Zhenhua; ZHU, Bin; HAO, Yanbin; NGO, Chong-Wah; HONG, Richang. (2025.0). *AAAI'25/IAAI'25/EAAI'25: Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence, Philadelphia, USA, February 25 - March 4,* (pp. 7311-7319) New York: ACM. https://doi.org/10.1609/aaai.v39i7.32786 (Published)

Hand1000: Generating realistic hands from text with only 1,000 images, by ZHANG, Haozhuo; ZHU, Bin; CAO, Yu; HAO, Yanbin. (2025.0). *AAAI'25/IAAI'25/EAAI'25: Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence, Philadelphia, USA, February 25 - March 4,* (pp. 9905-9913) New York: ACM. https://doi.org/aaai.v39i9.33074 (Published)

Navigating weight prediction with diet diary, by GUI, Yinxuan; ZHU, Bin; CHEN, Jingjing; NGO, Chong-Wah; JIANG, Yu-Gang. (2024.0). *MM '24: Proceedings of the 32nd ACM International Conference on Multimedia, Melbourne, Australia, October 28 - November 1,* (pp. 127-136) New York: ACM. https://doi.org/10.1145/3664647.3680977 (Published)

Video editing for video retrieval, by ZHU, Bin; FLANAGAN, Kevin; FRAGOMENI, Adriano; WRAY, Michael; DAMEN, Dima. (2024.0). *Computer vision: ECCV 2024 Workshops: Milan, September 29-October 4: Proceedings,* (pp. 236-252) Cham: Springer. https://doi.org/10.1007/978-3-031-92591-7_15 (Published)

Enhancing recipe retrieval with foundation models: A data augmentation perspective, by SONG, Fangzhou; ZHU, Bin; HAO, Yanbin; WANG, Shuo. (2024.0). *Proceedings of the18th European Conference, Milan, Italy, 2024 September 29-October 4,* (pp. 111-127) Cham: Springer. https://doi.org/10.1007/978-3-031-72983-6_7 (Published)

CgT-GAN: CLIP-guided text GAN for image captioning, by YU, Jiarui; LI, Haoran; HAO, Yanbin; ZHU, Bin; XU, Tong; HE, Xiangnan. (2023.0). *MM'23: Proceedings of the 31st ACM International Conference on Multimedia, Ottawa, Canada, October 29 - November 3,* (pp. 2252-2263) New York: ACM. https://doi.org/10.1145/3581783.3611891 (Published)

EPIC-KITCHENS VISOR benchmark: Video segmentations and object relations, by DAR KHALIL, Ahmad AK; SHAN, Dandan; ZHU, Bin; MA, Jian; KAR, Amlan; HIGGINS, Richard; FOUHEY, David; FIDLER, Sanja; DAMEN, Dima. (2022.0). *Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS 2022) Track on Datasets and Benchmarks, Virtual Conference, 2022 November 28,* (pp. 1-14) New Orleans, USA: (Published)

Mix-DANN and dynamic-modal-distillation for video domain adaptation, by YIN, Yuehao; ZHU, Bin; CHEN, Jingjing; CHENG, Lechao; JIANG, Yu-Gang. (2022.0). *MM '22: Proceedings of the 30th ACM International Conference on Multimedia, Lisboa, Portugal, October 10-14,* (pp. 3224-3233) New York: ACM. https://doi.org/10.1145/3503161.3548313 (Published)

Unsupervised video hashing with multi-granularity contextualization and multi-structure preservation, by HAO, Yanbin; DUAN, Jingru; ZHANG, Hao; ZHU, Bin; ZHOU, Pengyuan; HE, Xiangnan. (2022.0). *Proceedings of the 30th ACM International Conference on Multimedia, Lisboa, Portugal, 2022 October 10 - 14,* (pp. 3754-3763) New York: ACM. https://doi.org/10.1145/3503161.3547836 (Published)

Cross-lingual adaptation for recipe retrieval with mixup, by ZHU, Bin; NGO, Chong-Wah; CHEN, Jingjing; CHAN, Wing-Kwong. (2022.0). *ICMR '22: Proceedings of the 2022 International Conference on Multimedia Retrieval, Newark, NJ, June 27-30,* (pp. 258-267) New York: ACM. https://doi.org/10.1145/3512527.3531375 (Published)

Cross-domain cross-modal food transfer, by ZHU, Bin; NGO, Chong-Wah ; CHEN, Jingjing. (2020.0).

Proceedings of the 28th ACM International Conference on Multimedia, MM 2020, Seattle, October 12–16, (pp. 3762-3770) Virtual Conference: Association for Computing Machinery, Inc. (Published)

Person-level action recognition in complex events via TSD-TSM networks, by HAO, Yanbin; LIU, Zi-Niu; ZHANG, Hao; ZHU, Bin; CHEN, Jingjing; JIANG, Yu-Gang; NGO, Chong-Wah. (2020.0). *Proceedings of the 28th ACM International Conference on Multimedia, MM 2020, Seattle, October 12–16,* (pp. 4699-4702) Virtual Conference: Association for Computing Machinery, Inc. (Published)

CookGAN: Causality based Text-to-Image Synthesis, by ZHU, Bin; NGO, Chong-Wah. (2020.0). *Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, June 13-19,* (pp. 5518-5526) Virtual Conference: IEEE Computer Society. (Published)

R2GAN: Cross-modal recipe retrieval with generative adversarial network, by ZHU, Bin; NGO, Chong-Wah; CHEN, Jingjing; HAO, Yanbin. (2019.0). *Proceedings of the 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, California, June 16-21,* (pp. 11469-11478) Long Beach: IEEE Computer Society. (Published)

Research Grants

Singapore Management University

Quantity-aware Embedding for Dietary Suggestion Generation, SMU Internal Grant, Ministry of Education (MOE) Tier 1 , PI (Project Level): Bin ZHU, 2024, S$120,000

Other Institutions

Self-Adaptive Planning with Environmental Awareness for Embodied Agents, Ministry of Education (MOE) Tier 2, Ministry of Education (MOE) Tier 2 PI (Project Level): Bin ZHU, Co-PI (Project Level): HARA, Kotaro, SGD959,166

TEACHING

Courses Taught

Singapore Management University

Undergraduate Programmes :

Object Oriented Programming

THESES AND DISSERTATIONS

Theses and Dissertations Assessed

Other Institutions

External Examiner, "Towards Computationally-Efficient Solutions for Real-World Challenges in Egocentric Vision", Thesis by Gabriele Goletto, Polytechnic University of Turin, 2025

EXTERNAL SERVICE – PROFESSIONAL

Committee Chair, Program Co-chair, ACM International Conference on Multimedia Retrieval 2027, 2027 - Present

Other, Area Chair, IEEE International Conference on Multimedia & Expo (ICME) 2026, 2025 - Present

Other, Reviewer, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025 - Present

Special Session Organizer, IEEE International Conference on Multimedia & Expo 2025, 2024 - Present

Reviewer Conference Paper, European Conference on Computer Vision, 2024 - Present

Reviewer Conference Paper, IEEE/CVF Computer Vision and Pattern Recognition, 2023 - Present

Reviewer Conference Paper, Annual AAAI Conference on Artificial Intelligence, 2023 - 2024

Reviewer Conference Paper, IEEE/CVF Winter Conference on Applications of Computer Vision, 2023 - Present

Reviewer Conference Paper, ACM International Conference on Multimedia , 2023 - Present

Reviewer Conference Paper, International Conference on Computer Vision, 2023 - Present

Reviewer Journal Article, ACM Transactions on Multimedia Computing, Communications, and Applications, 2022 - Present

Reviewer Journal Article, IEEE Transactions on Multimedia, 2021 - Present