Baoxiong Jia

Contact Information

Beijing Institute for General Artificial Intelligence
Yiheyuan Road No.2, Beijing, China, 100080
Email: baoxiongjia@gmail.com

Phone: (+86) 13910779217
(+1) 240-550-4292
Homepage: buzz-beater.github.io

Education

University of California, Los Angeles, Los Angeles, U.S.
Doctor of Philosophy (Ph.D.), Computer Science — Sept. 2019 - Dec. 2022
Advisor: Prof. Song-Chun Zhu

University of California, Los Angeles, Los Angeles, U.S.
Master of Science (M.S.), Computer Science — Sept. 2017 - June 2019
Advisor: Prof. Song-Chun Zhu

Peking University, Beijing, China
Bachelor of Science (B.S.) with **honor**, Computer Science — Sept. 2014 - July 2018
Advisor: Prof. Yao Guo

Research Interest

Computer Vision	Activity Recognition/Prediction, 4D Scene Understanding
Artificial Intelligence	Planning and Inverse Planning, Intent Recognition
Machine Learning	Representation Learning, Neural-symbolic Methods

Publication

*Equal contribution. †Corresponding author. ‡Project lead.

Preprints

[1] Haoran Geng*, Yuyang Li*, Jie Yang, Feishi Wang, Ran Gong, Peiyuan Zhi, Puhao Li, Ruimao Zhang, Yixin Zhu, **Baoxiong Jia**, Siyuan Huang. RoboVerse: A Unified Simulation Framework for Scaling Vision-Language Manipulation. *Embodied AI Workshop @ CVPR* (EAI-CVPR) 2024.

[2] Zhuofan Zhang*, Ziyu Zhu*, Pengxiang Li*, Tengyu Liu, Xiaojian Ma, Yixin Chen, **Baoxiong Jia**, Siyuan Huang, Qing Li. Task-oriented Sequential Grounding in 3D Scenes. *arXiv preprint arXiv:2408.04034* (arXiv) 2024.

Journal

[1] Chi Zhang, **Baoxiong Jia**, Song-Chun Zhu, Yixin Zhu. Human-level Few-shot Concept Induction through Minimax Entropy Learning. *Science Advances* 2024.

[2] Siyuan Qi, **Baoxiong Jia**, Siyuan Huang, Ping Wei, Song-Chun Zhu. A Generalized Earley Parser for Human Activity Parsing and Prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (TPAMI) 2020.

[3] Yuanchun Li, **Baoxiong Jia**, Yao Guo, Xiangqun Chen. Mining User Reviews for Mobile App Comparisons. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (IMWUT) 2017. (presented at UbiComp17)

Conference

[1] Peiyuan Zhi*, Zhiyuan Zhang*, Muzhi Han, Zeyu Zhang, Zhitian Li, Ziyuan Jiao, **Baoxiong Jia**†,‡, Siyuan Huang†. Closed-loop Open-vocabulary Mobile Manipulation with GPT-4V. *International Conference on Robotics and Automation* (ICRA) 2025.

[2] Rundong Luo, Haoran Geng, Congyue Deng, Puhao Li, Zan Wang, **Baoxiong Jia**, Leonidas Guibas, Siyuang Huang. PhysPart: Physically Plausible Part Completion for Interactable Objects. *International Conference on Robotics and Automation* (ICRA) 2025.

[3] Yu Liu*, Baoxiong Jia*,‡, Ruijie Lu, Junfeng Ni, Song-Chun Zhu, Siyuan Huang. ArtGS: Building Interactable Replicas of Complex Articulated Objects via Gaussian Splatting. *International Conference on Learning Representations* (ICLR) 2025.

[4] Xiongkun Linghu*, Jiangyong Huang*, Xuesong Niu*, Xiaojian Ma, **Baoxiong Jia**†,‡, Siyuan Huang†. Multi-modal Situated Reasoning in 3D Scenes. *Advances in Neural Information Processing Systems* (NeurIPS) 2024.

[5] **Baoxiong Jia***, Yixin Chen*, Huangyue Yu, Yan Wang, Xuesong Niu, Tengyu Liu, Qing Li, Siyuan Huang. SceneVerse: Scaling 3D Vision-Language Learning for Grounded Scene Understanding. *European Conference on Computer Vision* (ECCV) 2024.

[6] Yu Liu*, **Baoxiong Jia***,‡, Yixin Chen, Siyuan Huang. SlotLifter: Slot-guided Feature Lifiting for Learning Object-centric Radiance Fields. *European Conference on Computer Vision* (ECCV) 2024.

[7] Ziyu Zhu*, Zhuofan Zhang*, Xiaojian Ma, Xuesong Niu, Yixin Chen, **Baoxiong Jia**, Zhidong Deng, Siyuan Huang, Qing Li. Unifying 3D Vision-Language Understanding via Promptable Queries. *European Conference on Computer Vision* (ECCV) 2024.

[8] Jiangyong Huang*, Silong Yong*, Xiaojian Ma*, Xiongkun Linghu*, Puhao Li, Yan Wang, Qing Li, Song-Chun Zhu, **Baoxiong Jia**, Siyuan Huang. An Embodied Generalist Agent in 3D World. *International Conference on Machine Learning* (ICML) 2024.

[9] Yandan Yang*, **Baoxiong Jia***,‡, Peiyuan Zhi, Siyuan Huang. PhyScene: Physically Interactable 3D Scene Synthesis for Embodied AI. *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR) 2024. (Highlight)

[10] Zan Wang, Yixin Chen, **Baoxiong Jia**, Puhao Li, Jinlu Zhang, Jingze Zhang, Tengyu Liu, Yixin Zhu†, Wei Liang†, Siyuan Huang†. Move as You Say, Interact as You Can: Language-guided Human Motion Generation with Scene Affordance. *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR) 2024. (Highlight)

[11] Jieming Cui*, Ziren Gong*, **Baoxiong Jia***,‡, Siyuan Huang, Zilong Zheng, Jianzhu Ma, Yixin Zhu. Probio: A Protocol-guided Multimodal Dataset for Molecular Biology Lab. *Advances in Neural Information Processing Systems* (NeurIPS) 2023.

[12] Ran Gong*, Jiangyong Huang*, Yizhou Zhao, Haoran Geng, Xiaofeng Gao, Qingyang Wu, Wensi Ai, Ziheng Zhou, Demetri Terzopoulos, Song-Chun Zhu, **Baoxiong Jia**†,‡, Siyuan Huang†. ARNOLD: A Benchmark for Language-Grounded Task Learning with Continuous States in Realistic Scenes. *International Conference on Computer Vision* (ICCV) 2023.

[13] Bo Dai, Linge Wang, **Baoxiong Jia**, Zeyu Zhang, Chi Zhang, Yixin Zhu, Song-Chun Zhu. X-VoE: Measuring eXplanatory Violation of Expectation in Physical events. *International Conference on Computer Vision* (ICCV) 2023. (Oral)

[14] Zeyu Zhang*, Muzhi Han*, **Baoxiong Jia**, Ziyuan Jiao, Yixin Zhu, Song-Chun Zhu, Hangxin Liu. Learning Spatial and Causal Transitions in Object Cutting. *International Conference on Intelligent Robots and Systems* (IROS) 2023.

[15] Siyuan Huang*, Zan Wang*, Puhao Li, **Baoxiong Jia**, Tengyu Liu, Yixin Zhu, Wei Liang, Song-Chun Zhu. Diffusion-based Generation, Optimization, and Planning in 3D Scenes. *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR) 2023.

[16] **Baoxiong Jia***, Yu Liu*, Siyuan Huang. Unsupervised Object-Centric Learning with Bi-Level Optimized Query Slot Attention. *International Conference on Learning Representations* (ICLR) 2023.

[17] **Baoxiong Jia**, Ting Lei, Song-Chun Zhu, Siyuan Huang. EgoTaskQA: Understanding Human Tasks in Egocentric Videos. *Advances in Neural Information Processing Systems* (NeurIPS) 2022.

[18] Chi Zhang*, Sirui Xie*, **Baoxiong Jia***, Yixin Zhu, Ying Nian Wu, Song-Chun Zhu. Learning Algebraic Representation for Systematic Generalization in Abstract Reasoning. *European Conference on Computer Vision* (ECCV) 2022.

[19] Peiyu Yu, Sirui Xie, Xiaojian Ma, **Baoxiong Jia**, Bo Pang, Ruiqi Gao, Yixin Zhu, Song-Chun Zhu, Ying Nian Wu. Latent Diffusion Energy-Based Model for Interpretable Text Modeling. *Interational Conference on Machine Learning* (ICML) 2022.

[20] Chi Zhang*, **Baoxiong Jia***, Song-Chun Zhu, Yixin Zhu. Abstract Spatial-Temporal Reasoning via Probabilistic Abduction and Execution. *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR) 2021.

[21] Chi Zhang, **Baoxiong Jia**, Mark Edmonds, Song-Chun Zhu, Yixin Zhua. ACRE: Abstract Causal REasoning Beyond Covariation. *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR) 2021.

[22] **Baoxiong Jia**, Yixin Chen, Siyuan Huang, Yixin Zhu, Song-Chun Zhu. LEMMA: A Multiview Dataset for LEarning Multi-agent Multi-task Activities. *European Conference on Computer Vision* (ECCV) 2020.

[23] Chi Zhang*, **Baoxiong Jia***, Feng Gao, Yixin Zhu, Hongjing Lu, Song-Chun Zhu. Learning Perceptual Inference by Contrasting. *Advances in Neural Information Processing Systems* (NeurIPS) 2019. (Spotlight)

[24] Chi Zhang*, Feng Gao*, **Baoxiong Jia**, Yixin Zhu, Song-Chun Zhu. RAVEN: A Dataset for Relational and Analogical Visual rEasoNing. *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR) 2019.

[25] Siyuan Qi*, Wenguan Wang*, **Baoxiong Jia**, Jianbing Shen, Song-Chun Zhu. Learning Human-Object Interactions by Graph Parsing Neural Networks. *European Conference on Computer Vision* (ECCV) 2018.

[26] Siyuan Qi, **Baoxiong Jia**, Song-Chun Zhu. 2018. Generalized Earley Parser: Bridging Symbolic Grammars and Sequence Data for Future Prediction *International Conference on Machine Learning* (ICML) 2018.

Research Experience

Beijing Institute for General Artificial Intelligence BIGAI, P.R.C.
Senior Research Scientist Feb. 2023 - Now

- 4D scene dynamic reconstruction, semantic understanding, and multi-modal learning.
- Embodied generalist agents and mobile manipulation.

Beijing Institute for General Artificial Intelligence BIGAI, P.R.C.
Research Intern, advised by: Dr. Siyuan Huang Oct. 2021 - Feb. 2023

- 4D human activity understanding and prediction with common sense knowledge base.
- Interactive learning of world dynamics and human intent.

Center for Vision, Cognition, Learning and Autonomy UCLA, U.S.A.
Research Assistant, advised by: Prof. Song-Chun Zhu Sept. 2017 - Dec. 2022

- 4D understanding of human activities and forecasting of both actions and scenes.
- Intention prediction and inverse planning based on stochastic grammar parsing, inverse reinforcement learning and theory of mind theories.
- Visual reasoning and induction for analogy in Raven Progressive Matrices.

Alexa Research, Teachable AI Team Amazon Inc., U.S.A.
Applied Scientist Intern, advised by: Dr. Qing Ping June 2021 - Sept. 2021

- Conducted research on spatial-temporal reasoning for video question answering with a special focus on leveraging video-language models for generating spatial-temporal grounding and compositional methods for reasoning.

Research and Development Department DMAI Inc., U.S.A.
Software Engineering Intern, mentored by: Tao Yuan Apr. 2019 - Mar. 2020

- Development of cognitive platform: 3D pose estimation, head pose and pointing gesture, modeling human beliefs.

Operating System Lab Peking University, P.R.C.
Research Intern, advised by: Prof. Yao Guo Feb. 2016 - May. 2018

- Automatic app comparison generation by mining comparative user reviews from app markets and applying sentiment analysis methods.

Teaching Experience

University of California, Los Angeles, Department of Computer Science

COM SCI 32 Introduction to Computer Science II, *Teaching Assistant*	Spring 2020
COM SCI 131 Programming Languages, *Teaching Assistant*	Fall 2020
COM SCI 31 Introduction to Computer Science I, *Teaching Assistant*	Spring 2021

Selected Honors and Awards

Outstanding Reviewer Award, ICLR	2021
Graduate Division Award, UCLA	2020
Outstanding Reviewer Award, CVPR	2020
NeurIPS Travel Award, NeurIPS	2019
Excellent College Graduate Award, Peking University	2018
Kwang-Hua Scholarship, Peking University	2014-2015
Award for Academic Excellence, Peking University	2015-2016

Services

Organizer	The 5th Workshop on 3D Scene Understanding for Vision, Graphics, and Robotics (3DSUN-CVPR)	2025
Organizer	The 1st Workshop on New Trends in Multimodal Human Action Perception, Understanding, and Generation (MANGO-CVPR)	2024
Reviewer	IEEE Robotics and Automation Letters (RA-L)	2025
Reviewer	IEEE Transactions on Image Processing (TIP)	2021
Reviewer	International Conference on Machine Learning (ICML)	Since 2021
Reviewer	Computer Vision and Pattern Recognition (CVPR)	Since 2019
Reviewer	International Conference on Learning Representation (ICLR)	Since 2021
Reviewer	Neural Information Processing Systems (NeurIPS)	Since 2020
Reviewer	European Conference on Computer Vision (ECCV)	Since 2020
Reviewer	International Conference on Computer Vision (ICCV)	Since 2021
Reviewer	AAAI Conference on Artificial Intelligence (AAAI)	2020-2021