

Baoxiong Jia

CONTACT INFORMATION	Beijing Institute for General Artificial Intelligence Phone: (+86) 13910779217 Yiheyuan Road No.2, Beijing, China, 100080 (+1) 240-550-4292 Email: baoxiongjia@cs.ucla.edu Homepage: buzz-beater.github.io/	
EDUCATION	University of California, Los Angeles , Los Angeles, U.S. <i>Doctor of Philosophy (Ph.D.)</i> , Computer Science Sept. 2019 - Dec. 2022 Advisor: Prof. Song-Chun Zhu University of California, Los Angeles , Los Angeles, U.S. <i>Master of Science (M.S.)</i> , Computer Science Sept. 2017 - June 2019 Advisor: Prof. Song-Chun Zhu Peking University , Beijing, China <i>Bachelor of Science (B.S.)</i> with honor , Computer Science Sept. 2014 - July 2018 Advisor: Prof. Yao Guo	
RESEARCH INTEREST	Computer Vision Artificial Intelligence Machine Learning	Activity Recognition/Prediction, 4D Scene Understanding Planning and Inverse Planning, Intent Recognition Representation Learning, Neural-symbolic Methods
PUBLICATION	* denotes equal contribution, † denotes corresponding author.	
PREPRINTS	<ul style="list-style-type: none">[1] 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. <i>arXiv preprint arXiv:2401.09340</i> (arXiv) 2024.[2] 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. <i>arXiv preprint arXiv:2404.10220</i> (arXiv) 2024.	
JOURNAL	<ul style="list-style-type: none">[1] Chi Zhang, Baoxiong Jia, Song-Chun Zhu, Yixin Zhu. Human-level Few-shot Concept Induction through Minimax Entropy Learning. <i>Science Advances</i> 2024.[2] Siyuan Qi, Baoxiong Jia, Siyuan Huang, Ping Wei, Song-Chun Zhu. A Generalized Earley Parser for Human Activity Parsing and Prediction. <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> (TPAMI) 2020.[3] Yuanchun Li, Baoxiong Jia, Yao Guo, Xiangqun Chen. Mining User Reviews for Mobile App Comparisons. <i>Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies</i> (IMWUT) 2017. (presented at UbiComp17)	
CONFERENCE	<ul style="list-style-type: none">[1] 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. <i>International Conference on Machine Learning</i> (ICML) 2024.[2] Yandan Yang*, Baoxiong Jia*, Peiyuan Zhi, Siyuan Huang. PhyScene: Physically Interactable 3D Scene Synthesis for Embodied AI. <i>IEEE Conference on Computer Vision and Pattern Recognition</i> (CVPR) 2024. (Highlight)[3] 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. <i>IEEE Conference on Computer Vision and Pattern Recognition</i> (CVPR) 2024. (Highlight)	

- [4] 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.
- [5] 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.
- [6] 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)
- [7] 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.
- [8] 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.
- [9] **Baoxiong Jia***, Yu Liu*, Siyuan Huang. Unsupervised Object-Centric Learning with Bi-Level Optimized Query Slot Attention. *International Conference on Learning Representations* (ICLR) 2023.
- [10] **Baoxiong Jia**, Ting Lei, Song-Chun Zhu, Siyuan Huang. EgoTaskQA: Understanding Human Tasks in Egocentric Videos. *Advances in Neural Information Processing Systems* (NeurIPS) 2022.
- [11] 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.
- [12] 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. *International Conference on Machine Learning* (ICML) 2022.
- [13] 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.
- [14] Chi Zhang, **Baoxiong Jia**, Mark Edmonds, Song-Chun Zhu, Yixin Zhu. ACRE: Abstract Causal REasoning Beyond Covariation. *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR) 2021.
- [15] **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.
- [16] 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)
- [17] 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.
- [18] 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.
- [19] 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	Center for Vision, Cognition, Learning and Autonomy		UCLA, U.S.A.
	<i>Research Assistant</i> , advised by: Prof. Song-Chun Zhu		Sept. 2017 - Dec. 2022
	<ul style="list-style-type: none"> • 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. 		
	Beijing Institute for General Artificial Intelligence		BIGAI, P.R.C.
	<i>Research Intern</i> , advised by: Dr. Siyuan Huang		Oct. 2021 - Dec. 2022
	<ul style="list-style-type: none"> • 4D human activity understanding and prediction with common sense knowledge base. • Interactive learning of world dynamics and human intent. 		
	Alexa Research, Teachable AI Team		Amazon Inc., U.S.A.
	<i>Applied Scientist Intern</i> , advised by: Dr. Qing Ping		June 2021 - Sept. 2021
	<ul style="list-style-type: none"> • 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.
	<i>Software Engineering Intern</i> , mentored by: Tao Yuan		Apr. 2019 - Mar. 2020
	<ul style="list-style-type: none"> • Development of cognitive platform: 3D pose estimation, head pose and pointing gesture, modeling human beliefs. 		
	Operating System Lab		Peking University, P.R.C.
	<i>Research Intern</i> , advised by: Prof. Yao Guo		Feb. 2016 - May. 2018
	<ul style="list-style-type: none"> • 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, <i>Teaching Assistant</i>		Spring 2020
	COM SCI 131 Programming Languages, <i>Teaching Assistant</i>		Fall 2020
	COM SCI 31 Introduction to Computer Science I, <i>Teaching Assistant</i>		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 1st Workshop on New Trends in Multimodal Human Action Perception, Understanding, and Generation (MANGO-CVPR)	2024
	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