

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. <i>Doctor of Philosophy (Ph.D.)</i> , Computer Science Advisor: Prof. Song-Chun Zhu Sept. 2019 - Dec. 2022	
	University of California, Los Angeles , Los Angeles, U.S. <i>Master of Science (M.S.)</i> , Computer Science Advisor: Prof. Song-Chun Zhu Sept. 2017 - June 2019	
	Peking University , Beijing, China <i>Bachelor of Science (B.S.)</i> with honor , Computer Science Advisor: Prof. Yao Guo Sept. 2014 - July 2018	
SELECTED HONORS AND AWARDS	Best Paper Award (SceneWeaver), RoboGen@IROS Best Open-source Award (RoboVerse), RoboGen@IROS Best Paper Award (UniFP), CoRL Outstanding Reviewer Award , ICLR Graduate Division Award , UCLA Outstanding Reviewer Award , CVPR Excellent College Graduate Award , Peking University Kwang-Hua Scholarship , Peking University Award for Academic Excellence , Peking University	2025 2025 2025 2021 2020 2020 2018 2014-2015 2015-2016
RESEARCH INTEREST	Computer Vision Robotics Artificial Intelligence Machine Learning	4D Scene Reconstruction, Embodied Vision Humanoid Loco-Manipulation, Compliant Behaviors Planning and Inverse Planning, Intent Recognition Representation Learning, Neural-symbolic Methods
PUBLICATION	*Equal contribution. †Corresponding author. ‡Project lead.	
CONFERENCE	<ul style="list-style-type: none">[1] Yandan Yang*, Baoxiong Jia^{*,†,‡}, Shujie Zhang, Siyuan Huang[†]. SceneWeaver: All-in-One 3D Scene Synthesis with an Extensible and Self-Reflective Agent. <i>Advances in Neural Information Processing Systems</i> (NeurIPS) 2025. (Best Paper Award, RoboGen Workshop @ IROS 2025)[2] Peiyuan Zhi*, Peiyang Li*, Jianqin Yin, Baoxiong Jia^{†,‡}, Siyuan Huang[†]. Learning Unified Force and Position Control for Legged Loco-Manipulation. <i>Conference on Robot Learning</i> (CoRL) 2025. (Best Paper Award)[3] Guanxing Lu*, Baoxiong Jia^{*,†,‡}, Puhao Li*, Yixin Chen, Ziwei Wang, Yansong Tang[†], Siyuan Huang[†]. GWM: Toward Scalable Gaussian World Models for Robotic Manipulation. <i>International Conference on Computer Vision</i> (ICCV) 2025.[4] Ziyu Zhu, Xilin Wang, Yixuan Li, Zhuofan Zhang, Xiaojian Ma, Yixin Chen, Baoxiong Jia, Wei Liang, Qian Yu, Zhidong Deng, Siyuan Huang, Qing Li. Move to Understand a 3D Scene: Bridging Visual Grounding and Exploration for Efficient and Versatile Embodied Navigation. <i>International Conference on Computer Vision</i> (ICCV) 2025.	

- [5] Haoran Geng*, Feishi Wang*, Songlin Wei*, Yuyang Li*, Bangjun Wang*, Boshi An*, Charlie Tianyue Cheng*, Haozhe Lou, Peihao Li, Yen-Jen Wang, Yutong Liang, Dylan Goetting, Chaoyi Xu, Haozhe Chen, Yuxin Qian, Yiran Geng, Jiageng Mao, Weikang Wan, Mingtong Zhang, Jiangran Lyu, Siheng Zhao, Jiazhaoh Zhang, Jialiang Zhang, Chengyang Zhao, Haoran Lu, Yufei Ding, Ran Gong, Yuran Wang, Yuxuan Kuang, Ruihai Wu, **Baoxiong Jia**, Carlo Sferrazza, Hao Dong, Siyuan Huang[†], Yue Wang[†], Jitendra Malik[†], Pieter Abbeel[†]. RoboVerse: Towards a Unified Platform, Dataset and Benchmark for Scalable and Generalizable Robot Learning. *Robotics: Science and Systems (RSS)* 2025. (**Best Open-source Award, RoboGen Workshop @ IROS 2025**)
- [6] Huangyue Yu*, **Baoxiong Jia**^{*,†}, Yixin Chen*, Yandan Yang, Puhao Li, Rongpeng Su, Jiabin Li, Qing Li, Wei Liang, Song-Chun Zhu, Tengyu Liu, Siyuan Huang. MetaScenes: Towards Automated Replica Creation for Real-world 3D Scans. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2025.
- [7] Jiangyong Huang*, **Baoxiong Jia**^{*,†}, Yan Wang*, Ziyu Zhu*, Xiongkun Linghu, Qing Li, Song-Chun Zhu, Siyuan Huang. Unveiling the Mist over 3D Vision-Language Understanding: Object-Centric Evaluation with Chain-of-Analysis. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2025.
- [8] Yan Wang*, **Baoxiong Jia**^{*,†}, Ziyu Zhu, Siyuan Huang. MPEC: Masked Point-Entity Contrast for Open-Vocabulary 3D Scene Understanding. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2025.
- [9] Ruijie Lu*, Yixin Chen*, Junfeng Ni, **Baoxiong Jia**, Yu Liu, Diwen Wan, Gang Zeng, Siyuan Huang. MOVIS: Enhancing Multi-Object Novel View Synthesis for Indoor Scenes. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2025.
- [10] Peiyuan Zhi*, Zhiyuan Zhang*, Yu Zhao, 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.
- [11] Rundong Luo, Haoran Geng, Congyue Deng, Puhao Li, Zan Wang, **Baoxiong Jia**, Leonidas Guibas, Siyuan Huang. PhysPart: Physically Plausible Part Completion for Interactable Objects. *International Conference on Robotics and Automation (ICRA)* 2025.
- [12] 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.
- [13] 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.
- [14] **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.
- [15] Yu Liu*, **Baoxiong Jia**^{*,†}, Yixin Chen, Siyuan Huang. SlotLifter: Slot-guided Feature Lifting for Learning Object-centric Radiance Fields. *European Conference on Computer Vision (ECCV)* 2024.
- [16] 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.
- [17] 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.

- [18] 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**)
- [19] 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**)
- [20] 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.
- [21] 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.
- [22] 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**)
- [23] 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.
- [24] 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.
- [25] **Baoxiong Jia***, Yu Liu*, Siyuan Huang. Unsupervised Object-Centric Learning with Bi-Level Optimized Query Slot Attention. *International Conference on Learning Representations (ICLR)* 2023.
- [26] **Baoxiong Jia**, Ting Lei, Song-Chun Zhu, Siyuan Huang. EgoTaskQA: Understanding Human Tasks in Egocentric Videos. *Advances in Neural Information Processing Systems (NeurIPS)* 2022.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] **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.
- [32] 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**)

- [33] Chi Zhang*, Feng Gao*, **Baoxiong Jia**, Yixin Zhu, Song-Chun Zhu. RAVEN: A Dataset for Relational and Analogical Visual rEcognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2019.
- [34] 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.
- [35] 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.

JOURNAL

- [1] Fan Wu, Ziyuan Jiao, Wanlin Li, Zeyu Zhang, Hang Li, Jiahao Wu, **Baoxiong Jia**, Shaopeng Dong. A VR-Based Robotic Teleoperation System With Haptic Feedback and Adaptive Collision Avoidance. *IEEE Transactions on Consumer Electronics* 2024.
- [2] Chi Zhang, **Baoxiong Jia**, Song-Chun Zhu, Yixin Zhu. Human-level Few-shot Concept Induction through Minimax Entropy Learning. *Science Advances* 2024.
- [3] 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.
- [4] 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)

PREPRINTS

- [1] Yushi Du*, Yixuan Li*, **Baoxiong Jia**^{*,†,‡}, Yutang Lin, Pei Zhou, Wei Liang[†], Yanchao Yang[†], Siyuan Huang[†]. COLA: Learning Human-Humanoid Coordination for Collaborative Object Carrying. *arXiv preprint arXiv:2510.14293v1* (arXiv) 2025
- [2] Xiongkun Linghu, Jiangyong Huang, Ziyu Zhu, Baoxiong Jia[†], Siyuan Huang[†]. SceneCOT: Eliciting Grounded Chain-of-Thought Reasoning in 3D Scenes. *arXiv preprint arXiv:2510.16714v1* (arXiv) 2025.
- [3] Zan Wang*, Jingze Zhang*, Yixin Chen, **Baoxiong Jia**, Wei Liang, Siyuan Huang. Spatial-Temporal Multi-Scale Quantization for Flexible Motion Generation. *arXiv preprint arXiv:2508.08991* (arXiv) 2025.
- [4] Wentao Jiang, Jingya Wang, Haotao Lu, Kaiyang Ji, Baoxiong Jia, Siyuan Huang, Ye Shi. ARFlow: Human Action-Reaction Flow Matching with Physical Guidance. *arXiv preprint arXiv:2503.16973* (arXiv) 2025.
- [5] Jiangyong Huang, Xiaojian Ma, Xiongkun Linghu, Yue Fan, Junchao He, Wenxin Tan, Qing Li, Song-Chun Zhu, Yixin Chen, **Baoxiong Jia**[†], Siyuan Huang[†]. LEO-VL: Towards 3D Vision-Language Generalists via Data Scaling with Efficient Representation. *arXiv preprint arXiv:2506.09935* (arXiv) 2025.
- [6] 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.

RESEARCH EXPERIENCE

- | | |
|--|---|
| <p>Beijing Institute for General Artificial Intelligence
 Senior Research Scientist</p> <ul style="list-style-type: none"> • 4D dynamic scene reconstruction, semantic understanding, and multi-modal learning. • Embodied generalist agents and mobile manipulation. | <p>BIGAI, P.R.C.
 Feb. 2023 - Now</p> |
| <p>Beijing Institute for General Artificial Intelligence
 Research Intern, advised by: Prof. Song-Chun Zhu</p> <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. | <p>BIGAI, P.R.C.
 Oct. 2021 - Feb. 2023</p> |

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. 	
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. 	

INVITED TALKS	Agentic Frameworks for 3D Scene Generation	
	<i>at RoboGen Workshop, IROS 2025</i>	Oct. 2025
	Force-inspired Learning for Legged Robots	
	<i>at 3D Vision Factory</i>	Oct. 2025
	Learning Unified Force and Position Control in Legged Loco-Manipulation	
	<i>at HKU Data 8017</i>	Oct. 2025
	Building Interactable 3D Scenes for Embodied AI	
	<i>at Robotics AI Institute, Boston Dynamics</i>	Mar. 2025
	3D Scene Understanding, Generation, and Interaction for Embodied AI	
	<i>at ChinaGraph 2024 & ZhiDX</i>	Oct. 2024
	Perceive, Ground, Interact, and Act	
	<i>at BIGAI Summer Class</i>	July 2024

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

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 Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)	2025
	Reviewer	IEEE Robotics and Automation Letters (RA-L)	2025
	Reviewer	IEEE Transactions on Image Processing (TIP)	2021
	Reviewer	International Conference on Intelligent Robots and Systems	Since 2025
	Reviewer	International Conference on Machine Learning (ICML)	Since 2021
	Reviewer	International Conference on Learning Representation (ICLR)	Since 2021
	Reviewer	International Conference on Computer Vision (ICCV)	Since 2021
	Reviewer	AAAI Conference on Artificial Intelligence (AAAI)	2020-2021
	Reviewer	Neural Information Processing Systems (NeurIPS)	Since 2020

Reviewer	European Conference on Computer Vision (ECCV)	Since 2020
Reviewer	Computer Vision and Pattern Recognition (CVPR)	Since 2019