

Summary

Leading Spatial Reasoning in NVIDIA Cosmos 3. Tech Lead of the Embodied Reasoning Project (Ember), advancing Robot System 2 models including Vesta, a state-of-the-art generalist embodied reasoning VLM spanning localization, navigation, memory, reasoning, tool use, and long-horizon planning. My research develops grounded multimodal systems and efficient vision architectures that connect language, perception, geometry, and action.

Professional Experience

NVIDIA Research Principal Research Scientist · Learning and Perception Research	Nov 2017 – Present Santa Clara, CA
- Leading Spatial Reasoning in NVIDIA Cosmos 3. - Tech Lead of the Embodied Reasoning Project (Ember), advancing Robot System 2 models including Vesta, a state-of-the-art generalist embodied reasoning VLM spanning localization, navigation, memory, reasoning, tool use, and long-horizon planning. - Develop efficient multimodal architectures, including GSPN-family spatial propagation and token-efficient high-resolution vision-language models.	
NVIDIA Research Research Intern · Learning and Perception Research	March 2017 - Aug 2017 Santa Clara, CA
The Chinese University of Hong Kong Visiting Scholar · Multimedia Laboratory (MMLab)	Jul 2016 - Dec 2016 Hong Kong
Baidu Applied Scientist Intern · Institute of Deep Learning; face parsing and computational photography.	May 2013 - Jan 2016 Beijing, China

Education

University of California, Merced Ph.D. in Electrical Engineering and Computer Science Advised by Ming-Hsuan Yang.	2017
University of Science and Technology of China M.S. in Electronic Engineering and Information Sciences Advised by Stan Z. Li and Bin Li.	2011
North China Electric Power University Bachelor's in Control Science	

Honors & Awards

Rising Stars in EECS	2019
NVIDIA Pioneering Research Award	2017
Baidu Graduate Fellowship	2013

Publications (2024–2026)

Complete 2024–2026 publication record. Sifei Liu is shown in bold.

2026

- 1. Vesta: A Generalist Embodied Reasoning Model**
Johan Bjorck, Zhiqi Li, Yunze Man, Jing Wang, An-Chieh Cheng, **Sifei Liu**, Shihao Wang, Zhiding Yu, Abhishek Badki, Stan Birchfield, Valts Blukis, Yevgen Chebotar, Siyi Chen, Sicong Leng, Yu-Cheng Chou, Tianli Ding, Boyi Li, Zhengyi Luo, Hang Su, Jonathan Tremblay, Tingwu Wang, Bowen Wen, Jimmy Wu, Xianghui Xie, Hanrong Ye, Hongxu Yin, K. R. Zentner, Liangyan Gui, Yu-Xiong Wang, Yuke Zhu, Linxi Fan, Jan Kautz
NVIDIA Technical Report, 2026 · arXiv
- 2. Reinforcing Dual-Path Reasoning in Spatial Vision Language Models**
Yatai Ji, An-Chieh Cheng, Yang Fu, Yukang Chen, Han Zhang, Zhaojing Yang, Wei Huang, Ka Chun Cheung, Song Han, Vidya Nariyambut Murali, Pavlo Molchanov, Simon See, Jan Kautz, Hongxu Yin, Ping Luo, **Sifei Liu**
arXiv, 2026 · arXiv · code
- 3. SpatialClaw: Rethinking Action Interface for Agentic Spatial Reasoning**
Seokju Cho, Ryo Hachiuma, Abhishek Badki, Hang Su, Byung-Kwan Lee, Chan Hee Song, **Sifei Liu**, Subhashree Radhakrishnan, Seungryong Kim, Yu-Chiang Frank Wang, Min-Hung Chen
arXiv, 2026 · arXiv · code

4. **4DP-QA: Scalable QA for 4D Perception in Vision Language Models**
Seokju Cho, Abhishek Badki, Hang Su, Jindong Jiang, Ziyao Zeng, Seungryong Kim, **Sifei Liu**, Orazio Gallo
CVPR, 2026 · [arXiv](#)
5. **Scaling Parallel Sequence Models to Vision Foundation Models**
Yitong Jiang, Collin McCarthy, Hongjun Wang, Hanrong Ye, Qi Dou, Tianfan Xue, Jinwei Gu, Jan Kautz, Hongxu Yin, Pavlo Molchanov, **Sifei Liu**
CVPR, 2026 · [arXiv](#)
6. **Grounded 3D-Aware Spatial Vision-Language Modeling**
An-Chieh Cheng, Yang Fu, Yatai Ji, Ligeng Zhu, Guanqi Zhan, Zhuoyang Zhang, Zhaojing Yang, Song Han, Yao Lu, Pavlo Molchanov, Vidya Nariyambut Murali, Jan Kautz, Xiaolong Wang, Hongxu Yin, **Sifei Liu**
CVPR, 2026 · [arXiv](#)
7. **Long-Horizon Manipulation via Trace-Conditioned VLA Planning**
Isabella Liu, An-Chieh Cheng, Rui Yan, Geng Chen, Ri-Zhao Qiu, Xueyan Zou, Sha Yi, Hongxu Yin, Xiaolong Wang, **Sifei Liu**
arXiv, 2026 · [arXiv](#)
8. **3D-Layout-R1: Structured Reasoning for Language-Instructed Spatial Editing**
Haoyu Zhen, Xiaolong Li, Yilin Zhao, Han Zhang, **Sifei Liu**, Kaichun Mo, Chuang Gan, Subhashree Radhakrishnan
arXiv, 2026 · [arXiv](#)
9. **4D-RGPT: Toward Region-level 4D Understanding via Perceptual Distillation**
Chiao-An Yang, Ryo Hachiuma, **Sifei Liu**, Subhashree Radhakrishnan, Raymond A. Yeh, Yu-Chiang Frank Wang, Min-Hung Chen
CVPR, 2026 · [arXiv](#) · [code](#)
10. **OmniVinci: Enhancing Architecture and Data for Omni-Modal Understanding LLM**
Hanrong Ye, Chao-Han Huck Yang, Arushi Goel, Wei Huang, Ligeng Zhu, Yuanhang Su, Sean Lin, An-Chieh Cheng, Zhen Wan, Jinchuan Tian, Yuming Lou, Dong Yang, Zhijian Liu, Yukang Chen, Ambrish Dantrey, Ehsan Jahangiri, Sreyan Ghosh, Daguang Xu, Ehsan Hosseini-Asl, Danial Mohseni Taheri, Vidya Nariyambut Murali, **Sifei Liu**, Yao Lu, Oluwatobi Olabiye, Yu-Chiang Frank Wang, Rafael Valle, Bryan Catanzaro, Andrew Tao, Song Han, Jan Kautz, Hongxu Yin, Pavlo Molchanov
ICLR, 2026 · [arXiv](#) · [code](#)
11. **QeRL: Beyond Efficiency—Quantization-Enhanced Reinforcement Learning for LLMs**
Wei Huang, Yi Ge, Shuai Yang, Yicheng Xiao, Huizi Mao, Yujun Lin, Hanrong Ye, **Sifei Liu**, Ka Chun Cheung, Hongxu Yin, Yao Lu, Xiaojuan Qi, Song Han, Yukang Chen
ICLR, 2026 · [arXiv](#)
12. **3D Aware Region Prompted Vision Language Model**
An-Chieh Cheng, Yang Fu, Yukang Chen, Zhijian Liu, Xiaolong Li, Subhashree Radhakrishnan, Song Han, Yao Lu, Jan Kautz, Pavlo Molchanov, Hongxu (Danny) Yin, Xiaolong Wang, **Sifei Liu**
ICLR, 2026 · [arXiv](#)

2025

13. **Real Deep Research for AI, Robotics and Beyond**
Xueyan Zou, Jianglong Ye, Hao Zhang, Xiaoyu Xiang, Mingyu Ding, Zhaojing Yang, Yong Jae Lee, Zhuowen Tu, **Sifei Liu**, Xiaolong Wang
arXiv, 2025 · [arXiv](#)
14. **EgoVLA: Learning Vision-Language-Action Models from Egocentric Human Videos**
Ruihan Yang, Qinxu Yu, Yecheng Wu, Rui Yan, Borui Li, An-Chieh Cheng, Xueyan Zou, Yunhao Fang, Xuxin Cheng, Ri-Zhao Qiu, Hongxu Yin, **Sifei Liu**, Song Han, Yao Lu, Xiaolong Wang
arXiv, 2025 · [arXiv](#) · [code](#)
15. **Scaling RL to Long Videos**
Yukang Chen, Wei Huang, Baifeng Shi, Qinghao Hu, Hanrong Ye, Ligeng Zhu, Zhijian Liu, Pavlo Molchanov, Jan Kautz, Xiaojuan Qi, **Sifei Liu**, Hongxu Yin, Yao Lu, Song Han
NeurIPS, 2025 · [arXiv](#) · [code](#)
16. **SSE: Multimodal Semantic Data Selection and Enrichment for Industrial-scale Data Assimilation**
Maying Shen, Nadine Chang, **Sifei Liu**, Jose M. Alvarez
KDD, 2025 · [arXiv](#)
17. **M3: 3D-Spatial MultiModal Memory**
Xueyan Zou, Yuchen Song, Ri-Zhao Qiu, Xuanbin Peng, Jianglong Ye, **Sifei Liu**, Xiaolong Wang
ICLR, 2025 · [arXiv](#)
18. **Describe Anything: Detailed Localized Image and Video Captioning**
Long Lian, Yifan Ding, Yunhao Ge, **Sifei Liu**, Hanzi Mao, Boyi Li, Marco Pavone, Ming-Yu Liu, Trevor Darrell, Adam Yala, Yin Cui
ICCV, 2025 · [arXiv](#) · [code](#)
19. **Token-Efficient VLM: High-Resolution Image Understanding via Dynamic Region Proposal**
Yitong Jiang, Jingwei Gu, Tianfan Xue, Ka Chun Cheung, Pavlo Molchanov, Hongxu (Danny) Yin, **Sifei Liu**
ICCV, 2025
20. **Parallel Sequence Modeling via Generalized Spatial Propagation Network**
Hongjun Wang, Wonmin Byeon, Jiarui Xu, Jingwei Gu, Ka Chun Cheung, Xiaolong Wang, Kai Han, Jan Kautz, **Sifei Liu**
CVPR, 2025 · [arXiv](#) · [code](#)
21. **Omni-RGPT: Unifying Image and Video Region-level Understanding via Token Marks**
Miran Heo, Min-Hung Chen, De-An Huang, **Sifei Liu**, Subhashree Radhakrishnan, Seon Joo Kim, Yu-Chiang Frank Wang, Ryo Hachiuma
CVPR, 2025 · [arXiv](#)
22. **BlobGEN-Vid: Compositional Text-to-Video Generation with Blob Video Representations**
Weixi Feng, Chao Liu, **Sifei Liu**, William Yang Wang, Arash Vahdat, Weili Nie
CVPR, 2025 · [arXiv](#)
23. **Scaling Vision Pre-Training to 4K Resolution**
Baifeng Shi, Boyi Li, Han Cai, Yao Lu, **Sifei Liu**, Marco Pavone, Jan Kautz, Song Han, Trevor Darrell, Pavlo Molchanov, Hongxu Yin
CVPR, 2025 · [arXiv](#)
24. **NaVILA: Legged Robot Vision-Language-Action Model for Navigation**
An-Chieh Cheng, Yandong Ji, Zhaojing Yang, Zaitian Gongye, Xueyan Zou, Jan Kautz, Erdem Biyik, Hongxu (Danny) Yin, **Sifei Liu**, Xiaolong Wang
RSS, 2025 · [arXiv](#) · [code](#)

25. NVILA: Efficient Frontier Visual Language Models

Zhijian Liu, Ligeng Zhu, Baifeng Shi, Zhuoyang Zhang, Yuming Lou, Shang Yang, Haocheng Xi, Shiyi Cao, Yuxian Gu, Dacheng Li, Xiuyu Li, Haotian Tang, Yunhao Fang, Yukang Chen, Cheng-Yu Hsieh, De-An Huang, An-Chieh Cheng, Jinyi Hu, **Sifei Liu**, Ranjay Krishna, Pavlo Molchanov, Jan Kautz, Hongxu Yin, Song Han, Yao Lu
CVPR, 2025 · [arXiv](#) · [code](#)

26. No Pose, No Problem: Surprisingly Simple 3D Gaussian Splats from Sparse Unposed Images

Botao Ye, **Sifei Liu**, Haofei Xu, Xueting Li, Marc Pollefeys, Ming-Hsuan Yang, Songyou Peng
ICLR, 2025 · [arXiv](#) · [code](#)

2024

27. Compositional Text-to-Image Generation with Feedforward Layout Generation

Sifei Liu, Weili Nie, An-Chieh Cheng, Morteza Mardani, Chao Liu, Benjamin Eckart, Arash Vahdat
ECCV Workshops, 2024

28. GroPrompt: Efficient Grounded Prompting and Adaptation for Referring Video Object Segmentation

Ci-Siang Lin, I-Jieh Liu, Min-Hung Chen, Chien-Yi Wang, **Sifei Liu**, Yu-Chiang Frank Wang
CVPR Workshops, 2024 · [arXiv](#)

29. RGBD Objects in the Wild: Scaling Real-World 3D Object Learning from RGB-D Videos

Hongchi Xia, Yang Fu, **Sifei Liu**, Xiaolong Wang
CVPR, 2024 · [arXiv](#)

30. AGG: Amortized Generative 3D Gaussians for Single Image to 3D

Dejia Xu, Ye Yuan, Morteza Mardani, **Sifei Liu**, Jiaming Song, Zhangyang Wang, Arash Vahdat
TMLR, 2024 · [arXiv](#)

31. CamCo: Camera-Controllable 3D-Consistent Image-to-Video Generation

Dejia Xu, Weili Nie, Chao Liu, **Sifei Liu**, Jan Kautz, Zhangyang Wang, Arash Vahdat
arXiv, 2024 · [arXiv](#)

32. BlobGEN-3D: Compositional 3D-Consistent Freeview Image Generation with 3D Blobs

Chao Liu, Weili Nie, **Sifei Liu**, Abhishek Badki, Hang Su, Morteza Mardani, Benjamin Eckart, Arash Vahdat
ACM Transactions on Graphics (SIGGRAPH Asia), 2024

33. Compositional Text-to-Image Generation with Dense Blob Representations

Weili Nie, **Sifei Liu**, Morteza Mardani, Chao Liu, Benjamin Eckart, Arash Vahdat
ICML, 2024 · [arXiv](#)

34. COLMAP-Free 3D Gaussian Splatting

Yang Fu, **Sifei Liu**, Amey Kulkarni, Jan Kautz, Alexei A. Efros, Xiaolong Wang
CVPR, 2024 · [arXiv](#) · [code](#)

35. CosAE: Learnable Fourier Series for Image Restoration

Sifei Liu, Shalini De Mello, Jan Kautz
NeurIPS, 2024

36. SpatialRGPT: Grounded Spatial Reasoning in Vision-Language Models

An-Chieh Cheng, Hongxu (Danny) Yin, Yang Fu, Qiushan Guo, Ruihan Yang, Jan Kautz, Xiaolong Wang, **Sifei Liu**
NeurIPS, 2024 · [arXiv](#) · [code](#)

37. RegionGPT: Towards Region Understanding Vision Language Model

Qiushan Guo, Shalini De Mello, Hongxu Yin, Wonmin Byeon, Ka Chun Cheung, Yizhou Yu, Ping Luo, **Sifei Liu**
CVPR, 2024 · [arXiv](#)

38. Dream-in-4D: A Unified Approach for Text- and Image-guided 4D Scene Generation

Yufeng Zheng, Xueting Li, Koki Nagano, **Sifei Liu**, Otmar Hilliges, Shalini De Mello
CVPR, 2024 · [arXiv](#)

39. 3D Reconstruction with Generalizable Neural Fields using Scene Priors

Yang Fu, Shalini De Mello, Xueting Li, Amey Kulkarni, Jan Kautz, Xiaolong Wang, **Sifei Liu**
ICLR, 2024 · [arXiv](#)

40. HOIDiffusion: Generating Realistic 3D Hand-Object Interaction Data

Mengqi Zhang, Yang Fu, Zheng Ding, **Sifei Liu**, Zhuowen Tu, Xiaolong Wang
CVPR, 2024 · [arXiv](#)

Selected Publications (2023 and Earlier)

1. TUVF: Learning Generalizable Texture UV Radiance Fields

An-Chieh Cheng, Xueting Li, **Sifei Liu**, Xiaolong Wang
arXiv, 2023

2. Affordance diffusion: Synthesizing hand-object interactions

Yufei (Judy) Ye, Xueting Li, Abhinav Gupta, Shalini De Mello, Stan Birchfield, Jiaming Song, Shubham Tulsiani, **Sifei Liu**
CVPR, 2023

3. Open-vocabulary panoptic segmentation with text-to-image diffusion models

Jiarui Xu, **Sifei Liu**, Arash Vahdat, Wonmin Byeon, Xiaolong Wang, Shalini De Mello
CVPR, 2023

4. CoordGAN: Self-Supervised Dense Correspondences Emerge from GANs

Jiteng Mu, Shalini De Mello, Zhiding Yu, Nuno Vasconcelos, Xiaolong Wang, Jan Kautz, **Sifei Liu**
CVPR, 2022

5. GroupViT: Semantic Segmentation Emerges from Text Supervision

Jiarui Xu, Shalini De Mello, **Sifei Liu**, Wonmin Byeon, Thomas B Breuel, Jan Kautz, Xiaolong Wang
CVPR, 2022

6. Learning Continuous Image Representation with Local Implicit Image Function

Yinbo Chen, **Sifei Liu**, Xiaolong Wang
CVPR, 2021

7. Video Autoencoder: self-supervised disentanglement of static 3D structure and motion

Zihang Lai, **Sifei Liu**, Alexei A. Efros, Xiaolong Wang
ICCV, 2021

8. Learning 3D Dense Correspondence via Canonical Point Autoencoder

An-Chieh Cheng, Xueting Li, Min Sun, Ming-Hsuan Yang, **Sifei Liu**
NeurIPS, 2021

9. Joint-task self-supervised learning for temporal correspondence

Xueting Li, **Sifei Liu**, Shalini De Mello, Xiaolong Wang, Jan Kautz, Ming-Hsuan Yang
NeurIPS, 2019

10. Learning Propagation for Arbitrarily-structured Data

Sifei Liu, Xueting Li, Varun Jampani, Shalini De Mello, Jan Kautz
ICCV, 2019

11. Switchable temporal propagation network

Sifei Liu, Guangyu Zhong, Shalini De Mello, Jingwei Gu, Varun Jampani, Ming-Hsuan Yang, Jan Kautz
ECCV, 2018

12. Learning Affinity via Spatial Propagation Networks

Sifei Liu, Shalini De Mello, Jingwei Gu, Guangyu Zhong, Ming-Hsuan Yang, Jan Kautz
NeurIPS, 2017

Selected Workshop & Tutorial Organization

Efficient Deep Learning for Foundation Models · ECCV 2024 (Co-organizer)

4D Hand Object Interaction: Geometric Understanding and Applications in Dexterous Manipulation · CVPR 2023 (Co-organizer)

Human-centric Trustworthy Computer Vision From Research to Applications · ICCV 2021 (Co-organizer)

Sensing, Understanding and Synthesizing Humans · ECCV 2020 (Co-organizer)

New Frontiers for Learning with Limited Labels or Data · ECCV 2020 (Co-organizer and Speaker)

Learning Representations via Graph-structured Networks Tutorial · CVPR 2019 and 2020 (Co-organizer and Speaker)