The First Workshop on Efficient Spatial Reasoning (ICLR 2026 workshop proposal)

Tagline: Making spatial reasoning systems more adaptive, robust, and efficient for real-world deployment.
Organizers: Haozheng Luo, Yijiang Li, Zhenyu Pan, Weiyang Liu, Manling Li, Zhijian Liu, Ruiyang Qin, Nuno Vasconcelos
Website: https://sites.google.com/ucsd.edu/efficient-spatial-reasoning
Format: Hybrid with in-person and online.
Anticipated Attendance: We expect about 100-200 attendees in total from different fields such as vision, NLP and robotics.

1 Workshop Summary

Spatial reasoning refers to the capacity to comprehend, represent, and manipulate spatial relationships among objects, agents, and environments. It plays a fundamental role in human cognition and is essential for artificial intelligence (AI) systems operating in domains such as robotics, autonomous navigation, 3D vision, medical imaging, and scientific modeling. Traditional methods typically employ geometric algorithms or symbolic logic to address spatial reasoning tasks. In contrast, contemporary foundation models integrate spatial reasoning within large-scale learning frameworks, offering greater flexibility. The emergence of large foundation models has significantly advanced the field of spatial representation and reasoning, achieving impressive performance in tasks such as 3D reconstruction [3, 11], 3D understanding [12], and spatial reasoning in vision-language models [14, 9].

Despite these developments, current models frequently depend on large parameter counts or test-time scaling strategies, such as extended generation length or deeper architectures. These approaches introduce substantial inefficiencies during either training or inference. Furthermore, existing models often underperform on tasks that require multi-step reasoning and a more nuanced understanding of complex spatial relationships. A central challenge lies in the prevalence of unreliable reasoning paths, which compromise both efficiency and accuracy.

To solve these challenges, researchers widely explore efficient reasoning to improve accuracy and interpretability under limited computation, time, and token budgets by streamlining the reasoning process and reducing redundancy. However, prevailing approaches—such as prompt engineering [13, 2], supervised fine-tuning [5, 8], reinforcement learning [1, 7], and representation learning [10] —while markedly improving general reasoning efficiency, have not yet addressed the specific demands of spatial reasoning. In applications like robotics and embodied AI, there is a heightened requirement for spatial reasoning that is not only efficient but also robust [6] and safe [4]. Yet the corresponding evaluation protocols and impact analyses for safety and robustness in spatial settings are largely underdeveloped in the general efficient-reasoning literature, leaving a clear measurement gap.

To further accelerate progress, we plan to host a workshop that brings together researchers and practitioners from academia and industry to discuss the latest advancements, challenges, and future directions in efficient spatial reasoning for foundation models. The main goal of our workshop bring together researchers and practitioners to explore advances in efficient spatial reasoning—methods that achieve strong generalization and robustness while remaining computationally practical. Topics of interest include but are not limited to symbolic–neural integration, geometric deep learning, scalable reasoning architectures, and evaluation frameworks. Through invited talks, paper presentations, and interactive sessions, participants will discuss open challenges such as efficiency–accuracy trade-offs, cross-modal reasoning, and robustness in real-world environments. The workshop aims to catalyze collaboration across AI, cognitive science, and applied domains, advancing the design of efficient and generalizable spatial reasoning systems.

Driving Questions. This workshop will bridge spatial reasoning and efficiency, fostering progress in the emerging field of efficient spatial reasoning in the domains of NLP, Vision, Multi-modality, and Robotics, with the following driving questions:

- Are there specialized neural architectures uniquely suited for efficient spatial reasoning, beyond general-purpose foundation models?
- What deployment strategies (on-device inference, low-resource training, adaptive reasoning budgets) are most promising for practical adoption?
- How should we measure "reasoning efficiency" beyond speed and FLOPs—can we capture interpretability, robustness, or adaptability?
- How can we ensure fairness and safety when compressing or pruning reasoning steps in spatial decision-making systems?

- What real-world applications—robotics, AR/VR, scientific discovery—stand to benefit most from efficient spatial reasoning models?
- What benchmarks should be established to fairly evaluate efficiency–accuracy trade-offs in spatial reasoning tasks?

Technical Scope. This workshop aims to serve as a platform for researchers and practitioners to discuss the latest advances, share insights across disciplines, and chart the path forward in the future for efficient spatial reasoning. We welcome submissions on a broad range of topics including, but not limited to the following research topics of interest:

- Novel architectures for efficient spatial reasoning, including sparse, modular, and geometry-aware designs.
- Training and optimization techniques such as distillation, token-budgeting, curriculum learning, and reinforcement learning for concise reasoning.
- Benchmarking and evaluation methods that jointly assess accuracy, efficiency, interpretability, and robustness in spatial reasoning tasks.
- Applications of efficient spatial reasoning in domains such as robotics, AR/VR, scientific discovery, urban planning, and simulation.
- Safety, fairness, and ethical considerations when compressing or pruning spatial reasoning processes.

2 Tentative Schedule

The workshop will include invited talks, contributed talks, and two poster sessions. See table 1 for a preliminary schedule. Each invited talk will take 45 minutes including QAs; each contributed talk will take 15 minutes. Although many of the invited speakers have extensively published on spatial reasoning work in major AI conferences and journals, We have requested that their presentations focus on novel ideas and unifying perspectives, rather than reiteration of already published work.

Table 1: Preliminary schedule.

Time	Event	Participants
09:00-09:10	Opening Remarks	
09:10-10:05	Invited Speaker: **Sewon Min**	
10:05-10:50	Invited Speaker: **Jiajun Wu**	
10:50-11:00	Coffee Break	☕🍦🍰
11:00-11:45	3 Contributed Talks	
11:45-13:00	Lunch	🍔🍕🍱
13:00-14:00	Poster Section I	
14:00-14:45	Invited Speaker: **Mengdi Wang**	
14:45-15:30	Invited Speaker: **Beidi Chen**	
15:30-15:40	Coffee Break	☕🍦🍰
15:40-16:25	Invited Speaker: **Soumalya Sarkar**	
16:25-17:25	4 Contributed Talks	
17:25-18:25	Poster Section II	
18:25-18:45	Conclusions & Outlook	

Attendance. Given the recent developments, exciting results, and renewed interest in efficient reasoning, spatial reasoning, vision–language models, and embodied AI, we anticipate around 150 participants. If **accepted**, we will promote the workshop through organizer and speaker professional networks, social media platforms (e.g., X/Twitter, LinkedIn, RedNote, Facebook), and affinity group mailing lists to ensure broad visibility and engagement.

Contributed Papers, Talks and Posters. We plan to allocate **seven** slots for oral contributed talks (**15 minutes each**), ensuring that contributed work receives nearly as much visibility as invited talks. The remaining accepted contributions will be presented during a **60-minute** poster session.
Submissions will be managed through **OpenReview**, with each paper receiving two reviews. Reviewers will be limited to a maximum of three submissions, and conflict-of-interest policies will follow ICLR guidelines (e.g., no reviews by recent coauthors or colleagues from the same institution). We anticipate approximately 80 contributed submissions.

We have already secured a sufficient pool of qualified reviewers from the program committee, organizing team. If necessary, we will invite additional expert reviewers to ensure sufficient attention on each submission and coverage of expertise. The tentative submission deadline is **February 7, 2026**, with final decisions announced by **March 1, 2026**. The workshop will be **non-archival**. All papers are required to use ICLR 2026 template and we will support two types of submissions: **long papers** and **short papers**.

Long papers. At the time of submission, the main text should be 9 pages or fewer. During the camera ready, the page limit will be increased to 10 pages to allow for new results/discussions. This limit will be strictly enforced. Papers with main body beyond the page limit will be desk-rejected. The list of references does not count towards the page limit, and unlimited additional pages are allowed for the bibliography/references. Authors may use as many pages of appendices (after the bibliography) as they wish, but reviewers are not required to read the appendix.

Short papers. Short submissions (4 pages, excluding references) highlighting preliminary ideas, negative results, position pieces, or ongoing work. These papers aim to lower the barrier to participation and encourage early-stage contributions from a diverse set of researchers. Tiny papers will primarily be presented as posters.

3 Invited Speakers

We are pleased that a group of researchers with diverse backgrounds, affiliations, and areas of expertise has agreed to give invited talks at our proposed workshop. Each speaker will bring a unique perspective to current developments of efficient spatial reasoning and its applications in the various domains. **Confirmed speakers include:**

1. **Sewon Min** (Assistant Professor at University of California Berkeley, female), Efficient text-based reasoning and retrieval
2. **Jiajun Wu** (Assistant Professor at Stanford, male), From 3D Perception to Spatial Reasoning: Building World Models Efficiently
3. **Soumalya Sarkar** (Senior Principal Scientist at RTRC, male), Efficient spatial reasoning in industrial production
4. **Beidi Chen** (Assistant Professor at CMU, female), Scaling Spatial Reasoning: Efficient Learning from Geometry to Generative Models
5. **Mengdi Wang** (Professor at Princeton University, female), Optimization and Reinforcement Learning Foundations for Efficient Spatial Reasoning

The final allocation of individual speakers to specific time slots will be **posted on the workshop website half a month before the event**, after confirming their specific time slot during the event day.

4 Diversity Commitment

In the **selection of organizers and speakers**, we actively aimed for diversity in various forms. The final roster of organizers and speakers comprises individuals from varied genders (males and females), races, regions, interests and affiliations. To enable interdisciplinary discussions, we bring together speakers from six different fields including text, audio, biology, robotics and graphs. We also aimed to maintain a balanced representation of speakers across academia and industry, encompassing assistant, associate, and full professors, as well as researchers from industry. Our organizing team includes members at various levels, from junior and senior Ph.D. students to assistant and full professors.

To broaden the accessibility of our workshop and encourage participation from underrepresented communities and individuals with limited resources, we intend to provide registration fee waivers and travel grants for accepted papers. To support these initiatives, we will actively pursue sponsorship from leading organizations such as Google DeepMind, NVIDIA, Meta AI, Qualcomm, and Amazon.

To encourage a diverse group of reviewers, we plan to host a **reviewing mentorship program** aimed at supporting the growth of junior reviewers. Through this initiative, junior reviewers will be paired with senior mentors who will provide real-time feedback and guidance throughout the review process. This collaborative effort is designed to improve the quality of reviews while fostering the development of the next generation of expert reviewers in the field.

5 Organizational Experience

The organizing team represents a diverse group with varying levels of experience in event organization and strong expertise in machine learning and related fields.

Manling Li has served as organizing committee member for ACL 2025 (Virtual Infrastructure Chairs), NAACL 2025 (Publication Chairs), and EMNLP 2024 (Demo Track Chairs). She also led the organization of the KnowledgeFM Workshop at ACL 2024, AAAI 2025, and ACL 2025; Workshop on Foundation Models meet Embodied Agents at CVPR 2025; Challenge on LLM for Embodied Decision Making at NeurIPS 2025.

Weiyang Liu has served as areas chair for multiple conferences, such as ICLR and NeurIPS. He has co-organized the ECCV 2024 Workshop on Foundation Models for 3D Humans.

Ruiyang Qin organized three tutorials regarding accelerating personalized AI via emerging technologies on DAC'25, ASP-DAC'25, and MWSCAS'25. He also served as TPC at AAAI'24, ICCAD'25, and ICCD'25.

Zhijian Liu co-organized the ECCV 2020 Tutorial on "From HPO to NAS: Automated Deep Learning". He also served as the Travel Grant Chair for the Conference on Machine Learning and Systems (MLSys) 2025.

Yijiang Li co-founded GrowAI, a research organization and community, where he regularly hosted seminars, organizing panels and events to foster discussion and sharing of knowledge from multiple disciplinaries.

Haozheng Luo previously served as President of the Chinese Christian Fellowship, where he regularly organized seminars and coordinated special events such as hiking trips and welcome parties for new Chinese students. He also organized a discussion group focused on exploring foundation model applications in AI for Science (AI4Science) and AI for City (AI4City).

Zhenyu Pan contributed to program operations for the ICCV 2025 workshop *"Trust Before Use: Building Foundation Models that You Can Trust"* (Honolulu, Oct. 19–23, 2025), helping coordinate the peer-review workflow—drafting evaluation rubrics, recruiting and matching reviewers, tracking deadlines, synthesizing meta-reviews with area chairs, and communicating decisions and rebuttal guidance to authors.

6 Conflicts of Interest

We will strictly adhere to the ICLR Conflicts of Interest policy and Code of Conduct. Workshop organizers will not give talks and will only moderate the session. All reviewers and Program Committee members will be required to declare conflicts of interest in OpenReview, which will be used to ensure fair and unbiased reviewer assignments. The review process will be double-blind and conducted via OpenReview to reduce institutional and author bias. All acceptance decisions will strictly follow ICLR conflict-of-interest guidelines, with conflicted PC members recused from handling relevant submissions.

7 Difference from Previous Related Workshops

Several notable workshops have focused on diffusion and generative models. For example, "Workshop on Efficient Reasoning" (NeurIPS 2025) emphasizes efficient reasoning with LLMs in natural and social sciences. "Workshop on SPACE in Vision, Language, and Embodied AI" (NeurIPS 2025) primarily explores spatial representation. Our workshop uniquely distinguishes itself by focusing on the following topics:

- **Efficiency-Centric Spatial Reasoning:** Unlike previous workshops that separately tackle efficiency and spatial understanding, we emphasize their intersection—how reasoning models can achieve spatial awareness with minimal computational and token budgets. This includes designing architectures and algorithms that balance accuracy, latency, and energy efficiency.
- **Safety and Robustness:** As spatial reasoning models increasingly operate in embodied or real-world environments, efficiency must not compromise safety. We invite discussions on robust reasoning under uncertainty, alignment for embodied decision-making, and mitigation of failure modes in safety-critical applications such as autonomous navigation and human–robot interaction.
- **Cross-Disciplinary Integration:** By bridging traditionally separate communities in vision, language, robotics, and spatial simulation, our workshop aims to unify perspectives on spatial reasoning efficiency. This interdisciplinary forum encourages the exchange of ideas, benchmarks, and tools to advance scalable, generalizable, and efficient spatial reasoning across modalities and domains.
- **Cross-Domain Applications:** By connecting vision, robotics, simulation, and language reasoning communities, our workshop aims to catalyze new applications in embodied AI, scientific discovery, and geospatial modeling—where efficient spatial reasoning is both a computational and conceptual challenge.
- **Benchmarking Beyond Accuracy:** Unlike previous workshops that focus mainly on accuracy-driven benchmarks, we aim to design new evaluation protocols that jointly assess efficiency–accuracy trade-offs, interpretability, and robustness in spatial reasoning tasks.

8 Organizers

Haozheng Luo (Northwestern University)

- Email: hluo@u.northwestern.edu
- Webpage: https://www.luor.org/
- Google Scholar: https://scholar.google.com/citations?user=MzH7kAcAAAAJ&hl=en
- Bio: Haozheng Luo is a Ph.D. candidate in Computer Science at Northwestern University, advised by Professors Han Liu and Yan Chen. His research focuses on AI safety, efficient reasoning, and foundation models across text, vision, speech, and genomic domains. He has developed methods for reasoning efficiency, such as outlier-removal strategies and token-level attention pruning, and has proposed frameworks that enhance quantization robustness and low-rank adaptation in large models. His work has been published at premier venues including ICML, NeurIPS, ICLR, and USENIX Security. This is the only workshop at ICLR he is involved with organizing this year.

Yijiang Li (UC San Diego)

- Email: yijiangli@ucsd.edu
- Webpage: https://williamium3000.github.io/
- Google Scholar: https://scholar.google.com/citations?user=Dx3z0m8AAAAJ
- Bio: Yijiang Li is a PhD student at the University of California, San Diego. Prior to that, he earned his M.S.E. in Computer Science at Johns Hopkins University (JHU). His areas of research focuses on the learning aspects of AI - to enable efficient (e.g. label efficiency, sample efficiency) and robust learning in multi-modal, interactive and 3D embodied environments. He has published papers on the above topics in top machine learning conferences such as ICML, ICCV, ICLR, CVPR, and EMNLP.

Zhenyu Pan (Northwestern University)

- Email: zhenyupan@u.northwestern.edu
- Webpage: https://pzyseere.github.io/
- Google Scholar: https://scholar.google.com/citations?hl=en&user=a_K3Nk0AAAAJ
- Bio: Zhenyu Pan is a second-year Ph.D. student in Computer Science at Northwestern University, advised by Han Liu. His research spans LLMs/VLMs, diffusion models, and graph neural networks, with a current focus on code generation, information retrieval, and 3D world generation via spatially aware reasoning and multi-agent RL. His work has appeared at venues such as NeurIPS and ICLR, as well as in Nature Computational Science. Zhenyu received the NVIDIA Academic Grant and has held research internships at Amazon's AWS AI and Alibaba's Tongyi Lab.

Manling Li (Northwestern University)

- Email: manling.li@northwestern.edu
- Webpage: https://limanling.github.io/
- Google Scholar: https://scholar.google.com/citations?user=6U4SXnUAAAAJ&hl=en
- Bio: Manling Li is an Assistant Professor of Computer Science at Northwestern University and an Amazon Scholar. She was a postdoc at Stanford University, and obtained the PhD degree in Computer Science at University of Illinois Urbana-Champaign. She works on Reasoning and Planning, in the intersection of Language, Vision, and Robotics. Her work has been recognized as ACL 2025 Inaugural Dissertation Award Honorable Mention, ACL 2024 Outstanding Paper Award, NAACL 2021 Best Demo Paper Award, ACL 2020 Best Demo Paper Award, Microsoft Research PhD Fellowship, EECS Rising Star, MIT Tech Review Innovators Under 35, and oral/spotlight papers in multiple venues.

Zhijian Liu (NVIDIA)

- Email: zhijian@ucsd.edu
- Webpage: https://zhijianliu.com/
- Google Scholar: https://scholar.google.com/citations?user=mwzYYPgAAAAJ

- Bio: Zhijian Liu is a research scientist at NVIDIA and an incoming assistant professor at UCSD. Previously, he received his Ph.D. and S.M. from MIT and his B.Eng. from Shanghai Jiao Tong University. His research focuses on efficient machine learning and systems. He has developed efficient ML algorithms and provided them with effective system support. He has also contributed to accelerating computation-intensive AI applications in computer vision, natural language processing, and scientific discovery. His work has been featured as oral and spotlight presentations at conferences such as NeurIPS, ICLR, and CVPR. He was selected as the recipient of the Qualcomm Innovation Fellowship. He was also recognized as a Rising Star in ML and Systems by MLCommons and a Rising Star in Data Science by UChicago and UCSD.

Ruiyang Qin (Villanova University)

- Email: rqin@villanova.edu
- Webpage: https://computingx-lab.github.io/
- Google Scholar: https://scholar.google.com/citations?user=c26FtWIAAAAJ&hl=en
- Bio: Ruiyang Qin is currently an assistant professor in the Department of Electrical and Computer Engineering at Villanova University, where he direct the ComputingX Lab. He received his B.S./M.S. in CS from Georgia Tech and Ph.D. in CSE from the University of Notre Dame. In the very early stages of his career, his work has been nominated William J. McCalla Best Paper in ICCAD 2024 and Spotlight Paper in ICLR 2025. He also received Edison Innovation Fellowship from IDEA center Notre Dame, IAD Academic Fellowship from University at Buffalo, and DAC Young Fellow.

Weiyang Liu (CUHK)

- Email: wyliu@cse.cuhk.edu.hk
- Webpage: https://wyliu.com/
- Google Scholar: https://scholar.google.com/citations?user=DMjROf0AAAAJ&hl=en
- Bio: Weiyang Liu is an Assistant Professor in the Department of Computer Science and Engineering at The Chinese University of Hong Kong, where he leads the Scalable Principles for Learning and Reasoning Lab (SphereLab). He finished his postdoc at the Max Planck Institute for Intelligent Systems, advised by Bernhard Schölkopf. He received a Ph.D. in Machine Learning from the University of Cambridge and a Ph.D. in Computer Science from the Georgia Institute of Technology. He has also held research positions at Google, Nvidia, and MERL. He has received the Baidu Fellowship, Hitachi Fellowship, and was a Qualcomm Innovation Fellowship Finalist. His work has been recognized by the 2023 IEEE Signal Processing Society Best Paper Award.

Nuno Vasconcelos (UC San Diego)

- Email: nuno@ucsd.edu
- Webpage: http://www.svcl.ucsd.edu/people/nuno/
- Google Scholar: https://scholar.google.com/citations?user=Fykyo9gAAAAJ
- Bio: Nuno is a Professor at the Electrical and Computer Engineering Department of the University of California, San Diego, where he heads the Statistical Visual Computing Laboratory. Before joining UCSD, he was a member of the research staff at the Compaq Cambridge Research Laboratory, which later became the HP Cambridge Research Laboratory. He received a PhD from MIT in 2000 and his areas of research interest are computer vision, statistical signal processing, machine learning, and multimedia. He is a Fellow of IEEE and the recipient of NSF CAREER award and Hellman Fellowship.

9 Program Committee

All organizers listed above will also serve as members of the Program Committee. In addition, we plan to recruit more qualified reviewers to handle an anticipated 100 submissions, leveraging outreach through social media and professional networks. We are actively expanding the committee to ensure balanced representation across academia and industry. Members marked with (✓) have already confirmed their participation: Chengwei Xu(✓), Haoyu He(✓), Zhuolin Jiang(✓), Guo Ye(✓), Chenghao Qiu(✓), Chingyuen Huang(✓), Hao Xu(✓), Eric Jiang(✓), Jiansu Zhang(✓),

10 Plan to Get an Audience

To reach a broad and diverse group of potential participants and audience members, we plan to take the following steps:

- Develop a dedicated workshop website containing all relevant information;
- Share the workshop through the professional networks of the organizers and invited speakers;
- Distribute the call for papers via mailing lists relevant to both academic and industry communities, such as universities.;
- Promote the event through social media platforms (e.g., X/Twitter, Bluesky, LinkedIn, RedNote, Facebook);
- Contact institutions, organizations and communities that are traditionally underrepresented in the field.

11 Workshop Materials and Outcomes

All relevant materials and outcomes, including recordings of invited and contributed talks, will be made available on the workshop website for participants who cannot attend in person or follow the live stream. All accepted papers will also be posted online. Submissions of work already published at major machine learning venues (ICLR, ICML, NeurIPS), will be explicitly discouraged in the call for papers. Additionally, a Zoom link will be provided to ensure full participation for remote attendees.

12 LLM Usage Policy

LLMs may be used as general-purpose assistance tools. However, if they contributed substantially to research ideation or writing, authors must include a separate "LLM Usage" section—typically in the appendix and excluded from the page limit—detailing their role. Failure to disclose significant use may result in desk rejection. Regardless of usage, authors bear full responsibility for all content, including any text generated by LLMs, which cannot be credited as authors.

References

[1] Gongfan Fang, Xinyin Ma, and Xinchao Wang. Thinkless: Llm learns when to think. *arXiv preprint arXiv:2505.13379*, 2025.

[2] Tingxu Han, Zhenting Wang, Chunrong Fang, Shiyu Zhao, Shiqing Ma, and Zhenyu Chen. Token-budget-aware llm reasoning. *arXiv preprint arXiv:2412.18547*, 2024.

[3] Yicong Hong, Kai Zhang, Jiuxiang Gu, Sai Bi, Yang Zhou, Difan Liu, Feng Liu, Kalyan Sunkavalli, Trung Bui, and Hao Tan. Lrm: Large reconstruction model for single image to 3d. *arXiv preprint arXiv:2311.04400*, 2023.

[4] Yeonjun In, Wonjoong Kim, Sangwu Park, and Chanyoung Park. R1-act: Efficient reasoning model safety alignment by activating safety knowledge. *arXiv preprint arXiv:2508.00324*, 2025.

[5] Yuxuan Jiang, Dawei Li, and Frank Ferraro. Drp: Distilled reasoning pruning with skill-aware step decomposition for efficient large reasoning models. *arXiv preprint arXiv:2505.13975*, 2025.

[6] Gang Li, Yan Chen, Ming Lin, and Tianbao Yang. Drpo: Efficient reasoning via decoupled reward policy optimization, 2025.

[7] Zheng Li, Qingxiu Dong, Jingyuan Ma, Di Zhang, Kai Jia, and Zhifang Sui. Selfbudgeter: Adaptive token allocation for efficient llm reasoning. *arXiv preprint arXiv:2505.11274*, 2025.

[8] Xinyin Ma, Guangnian Wan, Runpeng Yu, Gongfan Fang, and Xinchao Wang. Cot-valve: Length-compressible chain-of-thought tuning. *arXiv preprint arXiv:2502.09601*, 2025.

[9] Kun Ouyang, Yuanxin Liu, Haoning Wu, Yi Liu, Hao Zhou, Jie Zhou, Fandong Meng, and Xu Sun. Spacer: Reinforcing mllms in video spatial reasoning. *arXiv preprint arXiv:2504.01805*, 2025.

[10] Xuan Shen, Yizhou Wang, Xiangxi Shi, Yanzhi Wang, Pu Zhao, and Jiuxiang Gu. Efficient reasoning with hidden thinking. *arXiv preprint arXiv:2501.19201*, 2025.

[11] Jiaxiang Tang, Zhaoxi Chen, Xiaokang Chen, Tengfei Wang, Gang Zeng, and Ziwei Liu. Lgm: Large multi-view gaussian model for high-resolution 3d content creation. In *European Conference on Computer Vision*, pages 1–18. Springer, 2024.

[12] Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. Vggt: Visual geometry grounded transformer. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 5294–5306, 2025.

[13] Silei Xu, Wenhao Xie, Lingxiao Zhao, and Pengcheng He. Chain of draft: Thinking faster by writing less. *arXiv preprint arXiv:2502.18600*, 2025.

[14] Jihan Yang, Shusheng Yang, Anjali W Gupta, Rilyn Han, Li Fei-Fei, and Saining Xie. Thinking in space: How multimodal large language models see, remember, and recall spaces. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 10632–10643, 2025.