Hao Tan

Long-Context Models, Generative 3D/4D and Multimodal AI

San Jose, CA | +1 919 519 6899 | haotan1994@gmail.com | Google Scholar | GitHub | Homepage

SUMMARY

AI researcher and technical leader currently focused on long-context modeling: using test-time training and fast-weight memory to scale language, video, and 3D systems beyond the limits of quadratic attention. Led Adobe's 3D generation and world-model research, co-led its video-generation effort, and helped establish distributed training infrastructure for large-scale multimodal models. Research spans long-context learning, the LRM family, video world models, and foundational vision-language systems including LXMERT.

EXPERIENCE

Senior Research Scientist **Aug 2021 - Present**
Adobe Research *San Jose, CA*

Long-Context and Test-Time Training Research Oct 2024 - Present

- **LastKV** (*last author; unpublished, 2026*) shares only the final-layer KV cache across completed chunks, reducing retained long-context memory by up to the model depth while preserving standard within-chunk training across language, video, and 3D.
- **Chimera** (*last author; arXiv preprint, 2026*) combines Kimi Delta Attention, compressed global attention, sparse MoE layers, and HeteroP scaling laws for efficient long-context image and video generation.
- **Going Sparser with Test-Time Training** (*last author; unpublished, 2026*) uses dynamically routed sparse fast-weight experts to scale TTT memory capacity without proportional compute, improving language modeling and novel-view synthesis.
- Published **Test-Time Training Done Right** (ICLR 2026), introducing efficient large-chunk test-time training for million-token context modeling across language, video, and 3D.

3D Generation and World Models Feb 2023 - Present

- Led Adobe's 3D-generation research from foundational development through product transfer; technology was integrated into 5+ Adobe products, including Photoshop, Firefly, and Substance 3D.
- Led the LRM research family across 3D/4D reconstruction, Gaussian splatting, relighting, mesh generation, and long-context scene modeling; built a cross-disciplinary team and delivered six ICLR papers with two Orals and three Spotlights.

Adobe Video Generation Mar 2024 - Sep 2024

- Co-led a video-generation effort that reached competitive state-of-the-art quality within six months and contributed to Adobe's first generation of video models released in late 2024.
- Authored the initial training framework, set training-system direction, and owned core data serving and curation while helping grow the team from 4 to 10+ contributors.

Distributed Training Infrastructure Nov 2021 - Oct 2022; Mar 2024 - Sep 2024

- Wrote and launched Adobe's first distributed training job, then optimized large-scale training for video and multimodal pretraining as well as long-context video training and inference.

EDUCATION

University of North Carolina at Chapel Hill 2016 - 2021
Ph.D. in Computer Science; advisor: Mohit Bansal. Bloomberg Data Science Ph.D. Fellow.

Shanghai Jiao Tong University 2012 - 2016
B.S. in Computer Science; ACM Honors Class.

TECHNICAL FOCUS

Research: Long-context modeling, test-time training, generative 3D/4D, world models, video generation, VLMs.
Systems: Distributed training, parallelism, data serving and curation, PyTorch, Python, C/C++.

SELECTED PUBLICATIONS

Full publication list available on Google Scholar.

2026 Last Layer Key-Value Cache is All You Need
Last author | Unpublished manuscript. Retains only final-layer KV for older chunks, cutting long-context memory across language, video, and 3D.

2026 Chimera: Designing and Chinchilla-Scaling Hybrid Visual Diffusion Transformers
Last author | arXiv preprint. Hybrid KDA/MLA attention, sparse MoE layers, and HeteroP scaling laws for efficient long-context visual generation.

2026 Going Sparser with Test-Time Training
Last author | Unpublished manuscript. Dynamically routed sparse fast-weight experts scale TTT memory capacity at comparable compute.

2026 Test-Time Training Done Right
Last author | ICLR 2026. Large-chunk test-time training across language, video, and 3D.

2025 Relic: Interactive Video World Model with Long-Horizon Memory
Last author | Preprint. Interactive generation with persistent long-horizon scene memory.

2025 RayZer: A Self-Supervised Large View Synthesis Model
ICCV 2025 Oral; Best Student Paper Honorable Mention. Self-supervised 3D learning without pose or geometry labels.

2025 LVSM: A Large View Synthesis Model with Minimal 3D Inductive Bias
ICLR 2025 Oral. Transformer-based view synthesis with minimal explicit 3D structure.

2024 LRM: Large Reconstruction Model for Single Image to 3D
Last author | ICLR 2024 Oral. Introduced a scalable feed-forward large reconstruction model.

2024 DMV3D and PF-LRM
ICLR 2024 Spotlights. Multi-view diffusion and pose-free joint pose/shape reconstruction.

2024 GS-LRM: Large Reconstruction Model for 3D Gaussian Splatting
Co-first author | ECCV 2024. High-resolution Gaussian splats from sparse posed images.

2023 Scaling Data Generation in Vision-and-Language Navigation
ICCV 2023 Oral. Generated 4.9M instruction-trajectory pairs across 1,200+ environments, reaching 80% R2R test success.

2020 Vokenization: Improving Language Understanding with Contextualized, Visually-Grounded Supervision
First author | EMNLP 2020. Visual supervision for language representations.

2019 LXMERT: Learning Cross-Modality Encoder Representations from Transformers
First author | EMNLP 2019 Oral. A foundational vision-language pretraining framework.

EARLIER RESEARCH EXPERIENCE

Research internships: Hugging Face (2020 - 2021, vision-language pretraining); Bloomberg AI (2020, multimodal chart summarization); Google Research (2019, vision-language navigation); Adobe Research (2018, image-editing language); Microsoft Research Asia (2015 - 2016, video-based hair reconstruction).

HONORS AND RESEARCH LEADERSHIP

- **Bloomberg Data Science Ph.D. Fellowship**, 2019 - 2021; selected among the top 5% of applicants.
- **Benchmark leadership**: Rank 1 on VizWiz, GQA, NLVR2, and Room-to-Room leaderboards, 2018 - 2019; Top 3 in the 2019 VQA Challenge.
- **Senior Area Chair**: ACL, EMNLP (Outstanding Senior Area Chair), and NAACL. **Area Chair**: AAAI, ACL, EMNLP, CoNLL, and ICLR.
- **Workshop Organizer**: 3D Foundation Models at CVPR 2024; SPLU-RoboNLP at EMNLP 2021.
- Reviewer for ICLR, NeurIPS, TMLR, CVPR, ICCV, ACL, EMNLP, NAACL, AACL, AAAI, IJCAI, and ICRA.