

Derong Jin

djin77@umd.edu · LinkedIn · Website · GitHub

Multimodal Generative Models · Audio-Visual Learning · 3D Vision · Spatial Audio

EDUCATION

University of Maryland, College Park, MD, USA

2024 – May 2030 (Expected)

Ph.D. in Computer Science

GPA: 4.0/4.0

The Chinese University of Hong Kong, Shenzhen, China

2020 – 2024

B.Eng. in Computer Science and Engineering

- Accumulated GPA: 3.81/4.00, Major GPA: 3.98/4.00, rank in class: 4th/135
- **Outstanding Graduate Award** (Top 1.1% among 2024 Undergraduate Cohort)
- **Academic Performance Scholarship** and **Dean's List Award** in every year

PUBLICATIONS

(* indicates equal contribution.)

1. **Derong Jin***, Xiyi Chen*, Ming C. Lin, and Ruohan Gao, “SonoWorld4D: Generating a 4D Audio-Visual Scene from One Video via Diffusion Synchronization”, *Under Review*.
2. Zisen Shao*, Zihao Wei*, **Derong Jin**, and Ruohan Gao, “Objects as Audio-Visual Modal Sound Fields”, in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2026.
3. **Derong Jin***, Xiyi Chen*, Ming C. Lin, and Ruohan Gao, “SonoWorld: From One Image to a 3D Audio-Visual Scene”, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.
4. **Derong Jin** and Ruohan Gao, “Differentiable Room Acoustic Rendering with Multi-View Vision Priors”, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2025. (**Oral presentation**)
5. Kenkun Liu, **Derong Jin**, Ailing Zeng, Xiaoguang Han, and Lei Zhang, “A Comprehensive Benchmark for Neural Human Radiance Fields”, in *Advances in Neural Information Processing Systems, Datasets and Benchmarks Track (NeurIPS)*, 2023.
6. Zhangyang Xiong, Di Kang, **Derong Jin**, Weikai Chen, Linchao Bao, Shuguang Cui, and Xiaoguang Han, “Get3DHuman: Lifting StyleGAN-Human into a 3D Generative Model using Pixel-aligned Reconstruction Priors”, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2023.

RESEARCH EXPERIENCE

University of Maryland, College Park

Research Assistant, Multisensory Machine Intelligence Lab

Advisor: Prof. Ruohan Gao

SonoWorld4D: 4D Audio-Visual Scene Generation (*Under Review*)

May 2026 – Present

- Developed SonoWorld4D, the first framework to turn a **monocular video into an explorable 4D audio-visual scene**, enabling real-time free-viewpoint exploration with spatial audio synchronized with moving objects.
- Designed **foreground-background alignment** to ground independently generated moving objects and the background in a shared world frame, giving each sound source a physically consistent position and trajectory.
- Designed **dual-space diffusion synchronization** to couple video-conditioned and object-centric audio generation, producing per-object sounds that match the input video yet can be re-rendered along any listener trajectory.
- Collected a **54-scene benchmark** with ambisonics recorded at novel listener poses; improved spatial correlation by **45%** vs. Sonic4D, and was preferred in **66–94%** of user-study comparisons.

SonoWorld: 3D Audio-Visual Scene Generation (CVPR 2026)

May 2025 – Mar 2026

- Developed SonoWorld to lift a **single image into an explorable 3D scene** via panoramic semantic grounding, with differentiable ambisonics rendering for point sources, spatially extended emitters, and ambient sound.
- Built an **interactive real-time audio-visual viewer** with synchronized 3D visuals and pose-dependent audio.
- Curated SonoScene360, a calibrated 360° audio-visual benchmark; achieved at least **47%** lower angular DOA error and **239%** higher spatial correlation compared with spatial-audio baselines in 3D audio-visual scenes.

- AV-DAR: Differentiable Audio-Visual Acoustic Rendering (ICCV 2025, Oral)Sep 2024 – May 2025
- Developed AV-DAR in PyTorch using a physics-based differentiable room-acoustic rendering pipeline with end-to-end differentiable acoustic beam tracing to learn from sparse measured room impulse responses (RIRs).
 - Designed cross-attention to align the features from multiple camera views with a 3D scene representation and derive material-aware priors for acoustic reflection, enabling RIR prediction at unobserved listener locations.
 - Evaluated AV-DAR on **six real-world scenes**; achieved comparable performance on the Real Acoustic Field dataset using only **0.1%** of the training RIRs, versus baselines trained on **1%** of the same data (**10×** more).
 - On RAF-Empty at 1% training data, reduced the loudness error from **3.00 to 2.50 dB (16.6%)** vs. AVR and T60 error from **21.8% to 10.7% (50.9%)** vs. AV-NeRF; each comparison uses the best baseline for that metric.

The Chinese University of Hong Kong, Shenzhen

Research Assistant, GAP Lab

Advisor: Prof. Xiaoguang Han

- Neural Human Radiance Fields Benchmark (NeurIPS 2023)May 2023 – Aug 2023
- Reproduced **four human rendering baselines across three datasets** with standardized data splits and metrics.
 - Analyzed how estimated foreground masks and SMPL body parameters affect neural human rendering quality.

- Get3DHuman: 3D Human Generation (ICCV 2023)May 2022 – May 2023
- Redesigned the volumetric decoder with signed distance fields (SDFs), tri-plane features, and progressive multi-resolution training to improve the geometric fidelity and topological consistency of generated humans.
 - Reproduced and evaluated three state-of-the-art neural 3D human generation baselines on a lab-curated dataset.

TECHNICAL SKILLS

- Machine Learning:** Python, PyTorch, OpenCV; diffusion models, flow matching, multimodal learning.
- 3D & Simulation:** Gaussian splatting, differentiable rendering, spatial audio, Three.js, OpenGL, Blender.
- Programming & Tools:** C/C++, CUDA, JavaScript, Git, CMake, MPI, OpenMP, LaTeX.

TEACHING EXPERIENCE

- University of Maryland, College Park
- Aug 2025 – Dec 2025
- Graduate Teaching Assistant — CMSC498K: Introduction to Multimodal Deep Learning
- Designed coding assignments and midterm review materials for **51 students** in Multimodal Deep Learning; held weekly office hours, answered online course questions, and graded student assignments and examinations.

- The Chinese University of Hong Kong, Shenzhen
- Feb 2022 – May 2024
- Undergraduate Teaching Assistant — CSC4120, CSC3100, CSC1002, STA2001
- Led weekly tutorials and review sessions, preparing problem sets, solutions, and additional course materials.

SELECTED HONORS & AWARDS

Academic Performance Scholarship (top 3–7%)	2021, 2022, 2023
Dean’s List Award, every undergraduate year	2020 – 2024
School of Data Science TA/USTF Award	2023
Undergraduate Research Award	2022
ICPC Asia Regional Silver Medals: Kunming (2022), Shanghai (2021)	2021 – 2022
Meritorious Winner, Mathematical Contest in Modeling	2023

ACADEMIC SERVICE

Peer reviewer: ICLR 2027, AAAI 2026, CVPR 2026, NeurIPS 2026 (**Top Reviewer**); IEEE Transactions on Emerging Topics in Computing.