Feng Gao

fenggo@amazon.com • f.gao@ucla.edu • +1 (310) 569-4532 • Personal Page • Google Scholar • Last Update on 2025-06-25

CURRENT RESEARCH INTERESTS

Multimodal Understanding (LLM) & Generation

- Improve multimodal LLMs on **Image/Video Understanding/Reasoning**
  - Enhance *multimodal pre-training* with *scalable data quality filtering and RL training*
  - Improve *multimodal-LLM reasoning* with *visual Chain-of-thoughts (CoT) and implicit CoT reasoning*
- Unified LLM for **Any-to-Any Generation**
  - Build a unified LLM training paradigm with both *autoregressive and diffusion objectives*.
  - Extend the unified LLM capability for *text-to-image/video/3-D generation and VLA-based embodied AI*

EDUCATION

Ph.D. in Statistics, **University of California, Los Angeles (UCLA)** 09/2017 – 06/2022
- Committee Chair: Prof. Mark S. Handcock, Prof. Ying Nian Wu
- Research Advisor: Prof. Song-Chun Zhu (2017-2021)
- Ph.D. Thesis: Multi-Modal Robotic Learning, Reasoning and Planning

M.S. in Computer Science, **University of Southern California (USC)** 08/2015 – 05/2017

B.Eng. in Computer Eng., **Univ. of Electronic Sci. and Tech. of China (UESTC)** 09/2011 – 06/2015

INDUSTRY EXPERIENCE

Applied Scientist, **Amazon** Store Foundation AI 08/2022 – present

- **Build Rufus** [Ref-1, Ref-2], Amazon's LLM-powered Shopping Assistant
  - Launched multimodal-Rufus for image/video understanding.
  - Launched multimodal-Rufus-7B to improve Rufus-VQA accuracy and experience coverage rate.
  - Leading development of Rufus-MM 7B-100B+ base model, focus but not limit on:
    - ***anyres/S^2/RoPE2D*** etc. for general VQA/OCR performance improvement
    - ***token compression/nativeViT*** for training/inference efficiency
    - ***multimodal LLM training data/training recipe*** for general VL tasks improvements
    - Led/Built ***multimodal scaling laws*** for multimodal pre-training ***text/vision/interleaved/video data mixture***.
  - Led/Built Rufus-MM framework: ***VLM as a judge*** and ***hallucination*** benchmarks.
- **Multimodal** and **Embodied AI** Research
  - 10+ **multimodal LLM/generation** papers on top-tier conferences (CVPR, NeurIPS, ECCV etc.)
  - 4 **embodied AI** papers on top-tier conference and workshops (NeurIPS, SIGGRAPH, EMNLP, 3DV)

SELECTED PUBLICATIONS

MULTIMODAL UNDERSTANDING & REASONING

[1] M-LLM Based Video Frame Selection for Efficient Video Understanding
K. Hu, **F. Gao**, X. Nie, P. Zhou, S. Tran, T. Neiman, L. Wang, M. Shah, R. Hamid, B. Yin, T. Chilimbi
Conference on Computer Vision and Pattern Recognition 2025 (CVPR 2025)

[2] GIVL: Improving Geographical Inclusivity of Vision-and-Language Models with Pre-Training Methods
D. Yin, **F. Gao**, G. Thattai, M. Johnston, K.W. Chang
Conference on Computer Vision and Pattern Recognition 2023 (CVPR 2023)

[3] Transform-Retrieve-Generate: Natural Language-Centric Outside-Knowledge Visual Question Answering
F. Gao, Q. Ping, G. Thattai, A. Reganti, Y.N. Wu, P. Natarajan
Conference on Computer Vision and Pattern Recognition 2022 (CVPR 2022)

[4] Dark, Beyond Deep: A Paradigm Shift to Cognitive AI with Human-like Commonsense
Y. Zhu, T. Gao, L. Fan, S. Huang, M. Edmonds, H. Liu, **F. Gao**, C. Zhang, S. Qi, Y.N. Wu, J.B. Tenenbaum, S.-C. Zhu
Engineering, Special Issue on Artificial Intelligence, 2020 (Engineering)

[5] Learning Perceptual Inference by Contrasting
C. Zhang, B. Jia, **F. Gao**, Y. Zhu, H. Lu, S.-C. Zhu
33rd Conference on Neural Information Processing Systems (NeurIPS 2019, spotlight)

[6] RAVEN: A Dataset for Relational and Analogical Visual rEasoNing
C. Zhang*, **F. Gao***, B. Jia, Y. Zhu, S.-C. Zhu (* Joint First Authors)
Conference on Computer Vision and Pattern Recognition 2019 (CVPR 2019)

MULTIMODAL GENERATION

[7] ARM: Appearance Reconstruction Model for Relightable 3D Generation
X. Feng*, C. Yu*, Z. Bi*, Y. Shang*, **F. Gao**, H. Wu, K. Zhou, C. Jiang, Y. Yang
Conference on Computer Vision and Pattern Recognition 2025 (CVPR 2025)

[8] GarmentDreamer: 3DGS Guided Garment Synthesis with Diverse Geometry and Texture Details
B. Li*, X. Li*, Y. Jiang*, T. Xie, **F. Gao**, H. Wang, Y. Yang, C. Jiang
International Conference on 3D Vision 2025 (3DV 2025)

[9] Atlas3D: Physically Constrained Self-Supporting Text-to-3D for Simulation and Fabrication
Y. Chen*, T. Xie*, Z. Zong*, X. Li, **F. Gao**, Y. Yang, Y.N. Wu, C. Jiang
38th Annual Conference on Neural Information Processing Systems (NeurIPS 2024)

[10] Flow Priors for Linear Inverse Problems via Iterative Corrupted Trajectory Matching
Y. Zhang, P. Yu, Y. Zhu, Y. Chang, **F. Gao**, Y.N. Wu, O. Leong
38th Annual Conference on Neural Information Processing Systems (NeurIPS 2024)

[11] Skews in the Phenomenon Space Hinder Generalization in Text-to-Image Generation
Y. Chang, Y. Zhang, Z. Fang, Y.N. Wu, Y. Bisk, **F. Gao**
The 18th European Conference on Computer Vision (ECCV 2024)

[12] TPA-Net: Generate A Dataset for Text to Physics-based Animation
Y. Qiu, **F. Gao**, M. Li, G. Thattai, Y. Yang, C. Jiang
arXiv:2211.13887

EMBODIED AI & ROBOTICS

[13] Planning as In-Painting:
A Diffusion-Based Embodied Task Planning Framework for Environments under Uncertainty
C. Yang, T. Wu, X. Gao, K.W. Chang, **F. Gao**
38th Conference on Neural Information Processing Systems, OWA workshop (NeurIPS 2024 OWA)

[14] VR-GS: A Physical Dynamics-Aware Interactive Gaussian Splatting System in Virtual Reality
Y. Jiang, C. Yu, T. Xie, Y. Feng, H. Wang, M. Li, H. Lau, **F. Gao**, Y. Yang, C. Jiang
ACM SIGGRAPH 2024

[15] Learning non-Markovian Decision-Making from State-only Sequences
A. Qin, **F. Gao**, Q. Li, S.-C. Zhu, S. Xie
37th Conference on Neural Information Processing Systems (NeurIPS 2023)

[16] Masked Path Modeling for Vision-and-Language Navigation
Z. Dou, **F. Gao**, N. Peng
The 2023 Conference on Empirical Methods in Natural Language Processing 2023 (EMNLP 2023)

[17] A Tale of Two Explanations: Enhancing Human Trust by Explaining Robot Behavior
M. Edmonds*, **F. Gao***, H. Liu*, X. Xie*, S. Qi, B. Rothrock, Y. Zhu, Y.N. Wu, H. Lu, S.-C. Zhu
Science Robotics 18 Dec 2019: Vol. 4, Issue 37, eaay4663 (Science Robotics) (* Joint First Authors)

[18] Feeling the Force:
Integrating Force and Pose for Fluent Discovery through Imitation Learning to Open Medicine Bottles
M. Edmonds*, **F. Gao***, X. Xie, H. Liu, S. Qi, Y. Zhu, B. Rothrock, S.-C. Zhu (* Joint First Authors)
30th International Conference on Intelligent Robots and Systems (IROS 2017)

[19] A Glove-based System for Studying Hand-Object Manipulation via Pose and Force Sensing
H. Liu, X. Xie, M. Millar, M. Edmonds, **F. Gao**, Y. Zhu, V.J. Santos, B. Rothrock, S.-C. Zhu
30th International Conference on Intelligent Robots and Systems (IROS 2017)

PROFESSIONAL SERVICES

Conference Reviewer

- Reviewer, CVPR — 2019-2021, 2023-2024
- Reviewer, ICLR — 2022
- Reviewer, NeurIPS Dataset Track — 2021
- Reviewer, NeurIPS — 2020-2022
- Reviewer, ECCV — 2020
- Reviewer, AAAI — 2020, 2021
- Reviewer, ICCV — 2019
- Reviewer, ICRA — 2018
- Reviewer, IROS — 2024

PROFESSIONAL SKILLS

Programming Languages

- Python, C++, MATLAB, LaTeX

Deep Learning & LLM Infrastructure

- **Deep Learning Framework & Tools**
  - PyTorch, TensorFlow, Hugging Face
- **LLM Training Infrastructure**
  - NemoMegatron, Deepspeed