

# LEI ZHANG

Email: lez023@ucsd.edu | Phone: (323)824-7872 | Homepage | Google Scholar

## EDUCATION

---

### University of California, San Diego

*Ph.D. in Computer Science*

Advisor: Prof. Julian McAuley

San Diego, USA

Sep. 2024 – Present

### Zhejiang University

*M.S. in Computer Science*

Hangzhou, China

Sep. 2021 – Jun. 2024

### South China University of Technology

*B.E. in Software Engineering*

Canton, China

Sep. 2017 – Jun. 2021

## PUBLICATIONS

---

- **Lei Zhang**, Junjiao Tian, Zhipeng Fan, Kunpeng Li, Jialiang Wang, Weifeng Chen, Markos Georgopoulos, Felix Juefei-Xu, YuXiang Bao, Julian McAuley, Manling Li, Zecheng He. “Think in Strokes, Not Pixels: Process-driven Image Generation via Interleaved Reasoning,” *Under Review*, 2025.
- Fangshun Shu\*, Yue Liao\*, **Lei Zhang\***, Le Zhuo, Chenning Xu, Guanghao Zhang, Haonan Shi, Weilong Dai, Tao Zhong, Zhelun Yu, Wanggui He, Siming Fu, Haoyuan Li, Hongsheng Li, Hao Jiang. “LLaVA-MOD: Making LLaVA Tiny Via MOE-Knowledge Distillation” *ICLR*, 2025. (\* core contribution) [PDF](#)
- **Lei Zhang**, Fangxun Shu, Tianyang Liu, Sucheng Ren., Hao Jiang, Cihang Xie, “Filter & Align: Leveraging Human Knowledge to Curate Image-Text Data,” *Technical Report*, 2024. [PDF](#)
- Fangxun Shu\*, **Lei Zhang\***, Hao Jiang, Cihang Xie, “Audio-visual LLM for Video Understanding,” *Technical Report*, 2024. (\* equal contribution) [PDF](#)
- **Lei Zhang**, Jie Zhang, Bowen Lei, Subhabrata Mukherjee, Xiang Pan, Bo Zhao, Caiwen Ding, Yao Li, Dongkuan Xu. “Accelerating Dataset Distillation via Model Augmentation,” *CVPR*, 2023 **Highlight**. [PDF](#)
- **Lei Zhang**, Zhibo Wang, Xiaowei Dong, Yunhe Feng, Xiaoyi Pang, Zhifei Zhang, Kui Ren, “Towards Fairness-aware Adversarial Network Pruning,” *ICCV*, 2023. [PDF](#)
- Wanggui He, Siming Fu, Mushui Liu, Xierui Wang, Wenyi Xiao, Fangxun Shu, Yi Wang, **Lei Zhang**, Zhelun Yu, Haoyuan Li, Ziwei Huang, Leilei Gan, Hao Jiang, “MARS: Mixture of Auto-Regressive Models for Fine-grained Text-to-image Synthesis,” *AAAI*, 2025. [PDF](#)
- Jie Zhang\*, **Lei Zhang\***, Gang Li, Chao Wu, “Adversarial Examples for Good: Adversarial Examples Guided Imbalanced Learning,” *ICIP*, 2022. (\* equal contribution) [PDF](#)
- Jie Zhang, Bo Li, Jianghe Xu, Shuang Wu, Shouhong Ding, **Lei Zhang**, Chao Wu, “Towards Efficient Data Free Black-box Adversarial Attack,” *CVPR*, 2022. [PDF](#)
- Wenqiao Zhang, Tianwei Lin, Jiang Liu, Fangxun Shu, Haoyuan Li, **Lei Zhang**, Wanggui He, Hao Zhou, Zheqi Lv, Hao Jiang, Juncheng Li, Siliang Tang, Yueting Zhuang, “HyperLLaVA: Dynamic Visual and Language Expert Tuning for Multimodal Large Language Models,” *Arxiv Preprint*, 2024. [PDF](#)

## EXPERIENCE

---

### Research Scientist Intern

*Meta MSL*

Jun. 2025 – Dec. 2025

Menlo Park, USA

- Supervisor: Dr. Zecheng He and Dr. JunJiao Tian
- Research focus: Interleaved Reasoning in Unified Multimodal Model

### Research Assistant

*University of California, San Diego.*

Sept. 2024 – Present

San Diego, USA

- Supervisor: Prof. Julian McAuley and Prof. Prithviraj Ammanabrolu
- Research focus: Multimodal Generation

- Research Intern** Apr. 2023 – Jun. 2024  
*University of California, Santa Cruz.* Remote
- Supervisor: Prof. Cihang Xie
  - Research focus: Vision-Language Learning
- Research Intern** May. 2023 – Jun. 2024  
*Alibaba Group.* Beijing, China
- Research focus: Multimodal Understanding
- Reserach Intern** Dec. 2022 – Feb. 2023  
*Microsoft Research Asia.* Remote
- Supervisor: Dr. Xun Guo
  - Research focus: Text-to-Image with Diffusion Model

## PROJECTS

---

- Process-driven Image Generation via Interleaved Reasoning*** Jun. 2025 – Dec. 2025
- Introduce process-driven image generation, a multi-step paradigm that decomposes synthesis into a co-evolving reasoning trajectory of textual plans and visual states, orchestrated through a recurring four-state process: plan, sketch, inspect, and refine.
  - The core challenges of process-driven generation stem from the ambiguity of intermediate states: (a) How to generate images in a step-by-step manner, ensuring semantic and spatial consistency. (b) How to evaluate the partially-completed image, preserving prior visual knowledge and correcting prompt-violated errors.
  - Our method, based on BAGEL, improves the GenEval benchmark by over 7% and the WISE benchmark by nearly 11%, achieving gains of nearly 15% on some challenging tasks.
- Efficient LLaVA via Mixture of Experts and Knowledge Distillation*** Mar. 2024 – Sept. 2024
- Integrate a sparse MoE architecture into the language model to strike a balance between computational efficiency and model expressiveness.
  - Propose progressive knowledge transfer strategy (a) Mimic Distillation on dense and sparse architecture separately to facilitate hierarchical knowledge transfer. (b) Preference Distillation to adjust probability distribution on preference data.
  - Ours-2B exceeds Qwen-VL-Chat-7B by an average of 8.8% on comprehension benchmarks with 0.3% training data and 23% training parameters.
  - Ours-2B matches RLHF-based methods with 7B and 13B parameters on hallucination benchmarks.
- Audio-Visual Large Language Model*** Jun. 2023 – Oct. 2023
- Propose modality-augmented training, integration of modality-specific tokens, for joint end-to-end training on different modalities.
  - Curate a high-quality instruction dataset ranging from multi-turn conversations and audio-visual narratives to complex reasoning tasks.
  - Our method outperforms non-LLM-based InterVideo by 6.6% and LLM-based Valley by 4.4% on video understanding benchmarks.

## TECHNICAL SKILLS

---

**Languages:** Python, C/C++, LaTeX

**Frameworks:** Pytorch

**Languages:** TOEFL 105, GRE 325+3.0