

ZONGLIN DI

Ph.D. Candidate in Computer Science & Engineering

+1 858-568-1160 ◊ zdi@ucsc.edu

EDUCATION

University of California, Santa Cruz
Ph.D. in Computer Science and Engineering

2022 – 2026 (*expected*)
Santa Cruz, CA

- Advisor: Prof. Yang Liu
- Research area: **vision language models (VLM)**, **LLM post-training & agents**, weakly / self-supervised learning, trustworthy AI
- Finished **6** projects & published **5** papers in **ECCV, ICCV, ICLR and ACL**

University of California, San Diego
M.Sc. in Electrical and Computer Engineering

2020 – 2022
La Jolla, CA

- Advisor: Prof. Xiaolong Wang
- Research area: weakly / self-supervised learning in computer vision
- Published **1** paper in **NeurIPS**

Tongji University
B.S. in Electrical and Computer Engineering

2014 – 2019
Shanghai, China

- Advisor: Prof. Yin Wang
- Research area: computer vision in satellite imaging
- Published **1** paper in **CVPR**

INTERNSHIPS

Research Intern, Adobe Research | San Jose, CA

June 2026 –

- Distilled a cloud-scale unified multimodal model (image and text understanding and generation) into an on-device student, benchmarking six backbone architectures (dense, modality-split FFN, two-tower, MoE routing, mixture-of-blocks) at iso-compute to find the most distillation-friendly design
- Developed a cloud-to-device weight-inheritance scheme that initializes the student from the teacher
- Explored memory-efficient tokenization and encode/decode together with few-step diffusion distillation
- Advisor: Dr. Ohiremen Dibua, Dr. Yan Fang

Research Intern, Apple | Sunnyvale, CA

June 2025 – Sept. 2025

- Gather data to support the development of **proactive AI**, human comprehension, and models of user behavior
- Establish a pipeline for proactive AI development and conduct a performance comparison between open-source and proprietary **vision language models**
- Conduct research about **streaming online video language understanding models**
- Advisor: Mr. Yikang Liao (joined OpenAI during my internship)

Research Intern, Adobe Research | San Jose, CA

June 2024 – Nov. 2024

- Conduct research about reinforcement learning with human feedback (RLHF) by developing a new pipeline based on DPO, SimPO, SPIN, Online RLHF, etc.
- Investigate how to improve the generalization of the large language model (LLM) using limited high-quality data
- Advisor: Dr. Jing Shi, Dr. Ruiyi Zhang

- Conduct research about **vision-language representation learning**, more specifically how to improve image difference captioning performance
- Finetune main-stream vision-language models on 100+ A100 GPUs including Qwen, InternVL, LLaVA, etc.
- Propose a large-scale, comprehensive and high-quality image difference captioning dataset for finetuning vision language models
- 1 paper is accepted by **ICCV 2025**
- Advisor: Dr. Jing Shi

RESEARCH EXPERIENCE

Research Assistant, UC Santa Cruz | Santa Cruz, CA

June 2022 – Present

- Propose a **new method with theoretical proof** to solve openset noise learning in federated learning and published **1** paper in **ECCV 2024**
- Propose a new method to improve machine unlearning from a **Stackelberg game (adversarial) perspective** and published **1** paper in **ICLR 2025**
- Propose a new method to improve machine unlearning from a **label smoothing perspective** and published **1** paper in **ICLR 2026**
- Advisor: Prof. Yang Liu

Research Assistant, UBC | Remote

June 2021 – June 2022

- Conduct medical image registration work using self-supervised methods and reinforcement learning
- Study algorithms about un/weakly-supervised learning in federated learning
- 1 paper is accepted by **ACM MM 2024**
- Advisor: Prof. Xiaoxiao Li

Research Assistant, UC San Diego | San Diego, CA

Sept. 2020 – June 2022

- Improve the performance of **pose estimation** using **test-time training and self-supervised learning**; the proposed methods improve performance by reconstructing another image of the same subject at test-time without annotations
- 1 paper is accepted by **NeurIPS 2021**
- Advisor: Prof. Xiaolong Wang

Research Assistant, Tongji University | Shanghai, China

June 2018 – Sept. 2020

- Perform large-scale roadmap extraction training on 40+ GPUs (distributed training) and rough labeling to study the limitation of **weakly-supervised learning in semantic segmentation**
- Improve road extraction using deep learning and GPS data, allowing a 5% increase in accuracy and 40% boost in generalization ability for new-area prediction
- 1 paper is accepted by **CVPR 2019**
- Advisor: Prof. Yin Wang

ON-GOING PROJECTS

Resource2Skill

- Build a framework that auto-distills human-created multimodal resources (tutorial videos, code repos, articles, reference artifacts) into 3,800+ executable, retrievable agent skills organized as a hierarchical multimodal wiki across seven authoring domains (Web, PowerPoint, Excel, Blender, audio, UE5, CAD), each bundling procedural text, runnable code, visual evidence, and provenance
- Design a two-stage hierarchy-then-LM retrieval (BM25 + LM selection) and an MCP-based execution runtime that lets agents compose skills and operate real tools to produce artifacts; lift average task quality by +11.9 points and outperform Claude Code / Codex baselines on 26 of 28 model-domain cells

SkillsBench

- Evaluate both the capability and safety of LLM productivity agents across five high-fidelity, production-conformant mock services (Gmail, Slack, Calendar, Docs, Drive), with deterministic state-based scoring that penalizes irreversible harmful actions
- Contribute to the mock-environment infrastructure and the release of 7,800+ scored agent trajectories; show that agent capability and safety do not track together, with top models reaching 39–64% task success but 7–33% unsafe-action rates

Dynamic MCTS Tree and Decoupled Role in LLM Agent

- Decouple the LLM agent into “Planner-Actor-Critic” with a dynamic MCTS tree to address the limitation of sparse reward and inaccurate process reward in long-horizon GUI tasks

Rethinking the LLM Evaluation

- Analyze **limitations of ELO-based evaluation methods** for large language models and **develop a robust evaluation framework** resilient to dataset-size variations and evaluation noise across diverse problem formats including chat, reasoning, and agent tasks

Improving the Spatial Reasoning Ability of the VLM

- Identify critical limitations in **vision-language models’ spatial reasoning abilities** and develop novel methods to enhance occlusion understanding (**ECCV 2026 submission**) and video-based spatial reasoning capabilities

Improving GRPO Sparse Reward with an Auxiliary Reward

- Solve the sparse reward in GRPO by using an auxiliary reward, formalized in an unbiased semi-parametric way

PRE-PRINTS

1. Jinrui Yang*, **Zonglin Di***, Ohi Dibia, Qing Liu, Seun Adekunle, Daniil Pakhomov, Darshan Ganesh Prasad, Cihang Xie, Yuyin Zhou. “Improving the Capability of Visual Language Models with Occlusion Reasoning”. *In submission*. [VLM]
2. Chris Liu*, **Zonglin Di***, Jeffrey Flanigan, Yang Liu. “Advancing Machine Unlearning Evaluation Requires Rethinking Retraining”. *In submission*. [Machine Unlearning]
3. Minghao Liu*, **Zonglin Di***, etc. “Auto-Dataset-Construction (ADC) in the Era of Large Language Models”. *In submission*.

PUBLICATIONS

1. Xia Su, Ruiqi Chen, Benlin Liu, Jingwei Ma, **Zonglin Di**, Ranjay Krishna, Jon E. Froehlich. “CapNav: Benchmarking Vision Language Models on Capability-conditioned Indoor Navigation”. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2026)*. [VLM]
2. **Zonglin Di**, Zhaowei Zhu, Jinghan Jia*, Jiancheng Liu*, Bo Jiang, Zafar Takhirov, Yuanshun Yao, Sijia Liu, Yang Liu. “Label Smoothing Improves Machine Unlearning”. *The 14th International Conference on Learning Representations (ICLR 2026)*. [Trustworthy AI, Machine Unlearning]
3. Kaizhi Zheng*, Kaiwen Zhou*, Jing Gu*, Yue Fan*, Jialu Wang*, **Zonglin Di**, Xuehai He, and Xin Eric Wang. “JARVIS: A Neuro-Symbolic Commonsense Reasoning Framework for Conversational Embodied Agents”. *19th International Conference on Neurosymbolic Learning and Reasoning (NeSy 2025)*. [VLM, Agentic AI]
4. **Zonglin Di**, Jing Shi, Yifei Fan, Hao Tan, Alexander Black, John Collomosse and Yang Liu. “DiffTell: A Comprehensive Dataset for Image Difference Captioning”. *International Conference on Computer Vision (ICCV 2025)*. [VLM]
5. **Zonglin Di***, Sixie Yu*, Yevgeniy Vorobeychik, Yang Liu. “Adversarial Machine Unlearning”. *The 13th International Conference on Learning Representations (ICLR 2025)*. [Trustworthy AI, Machine Unlearning]

6. Nan Wang, **Zonglin Di**, Houlin He, Qingchao Jiang, Xiaoxiao Li. “A Simple and Provable Approach for Learning on Noisy Labeled Multi-modal Medical Images”. *ACM International Conference on Multimedia (MM 2024)*. [Computer Vision, Trustworthy AI]
7. **Zonglin Di**, Zhaowei Zhu, Xiaoxiao Li, Yang Liu. “Federated Learning with Local Openset Noisy Labels”. *Proceedings of the European Conference on Computer Vision (ECCV 2024)*. [Computer Vision, Trustworthy AI]
8. Yunchao Zhang, **Zonglin Di**, Kaiwen Zhou, Cihang Xie and Eric Xin Wang. “Navigation as Attackers Wish? Towards Building Robust Embodied Agents under Federated Learning”. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2024)*. [Trustworthy AI]
9. Jialu Wang, Xinyue Gabby Liu, **Zonglin Di**, Yang Liu, Xin Eric Wang. “T2IAT: Measuring Valence and Stereotypical Biases in Text-to-Image Generation”. *Findings of the Association for Computational Linguistics: (ACL 2023)*. [VLM, Trustworthy AI]
10. Miao Hao*, Yizhuo Li*, **Zonglin Di***, Nitesh B. Gundavarapu and Xiaolong Wang. “Test-Time Personalization with a Transformer for Human Pose Estimation”. *Advances in Neural Information Processing Systems (NeurIPS 2021)*. [Computer Vision]
11. Tao Sun, **Zonglin Di**, Pengyu Che, Chun Liu and Yin Wang. “Leveraging Crowdsourced GPS Data for Road Extraction from Aerial Imagery”. *IEEE Conference on Computer Vision and Pattern Recognition, pp. 7509-7518, (CVPR 2019)*. [Computer Vision]
12. **Zonglin Di**, Qi Kang, Daogang Peng and Mengchu Zhou. “Density Peak based Pre-clustering Support Vector Machine for Multi-class Imbalanced Classification”. *IEEE Conference on System Man and Cybernetics, pp. 27-32, (SMC 2019)*.
13. **Zonglin Di**, Xiaoliang Gong, Jingyu Shi, Hosameldin O. A. Ahmed and Asoke K. Nandi. “Internet Addiction Disorder Detection of Chinese College Students Using Several Personality Questionnaire Data and Support Vector Machine”. *Addictive Behavior Reports, Elsevier, 2019, IF=3.299*.
14. Tao Sun, **Zonglin Di** and Yin Wang. “Combining Satellite Imagery and GPS Data for Road Extraction”. *2nd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery, pp. 29-32, (GeoAI 2018)*. [Computer Vision]
15. **Zonglin Di**, Siya Yao, Qi Kang and Mengchu Zhou. “AP-SVM: An Adaptive Clustering-based Support Vector Machine Algorithm for Imbalance Classification”. *IEEE International Conference on Systems, Man, and Cybernetics, pp. 681-686, (SMC 2018)*.
16. **Zonglin Di**, Xiaoliang Gong, Jingyu Shi, Hosameldin O. A. Ahmed and Asoke K. Nandi. “Detection of Internet Addiction Disorder Based on Personality Questionnaires of Chinese College Students and SVMs”. *10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, pp. 1-6, (ICSP-BMEI 2017)*.
17. Xiaoliang Gong, Bozhong Long, Kun Fang, **Zonglin Di**, Yichu Hou and Lei Cao. “A Prediction Based on Clustering and Personality Questionnaire Data for IGD Risk: A Preliminary Work”. *12th International Conference on Neural Computation, Fuzzy System and Knowledge Discovery, pp. 1699-1703, (ICNC-FSKD 2016)*.

AWARDS

- **Chancellor’s Fellowship**, UC Santa Cruz (2 / all applicants) Sept 2022
- **2nd Prize at the 30th National “Challenge Cup” Academic Science and Technology Competition**, Ministry of Education of the People’s Republic of China Nov 2019

SKILLS

Languages	C, C++, Python, PyTorch
Others	Git, Linux, L ^A T _E X

SERVICES

Reviewer for: ACM SIGSPATIAL 2019; WACV 22, 25, 26; EMNLP 22, 23; AAAI 23, 24, 25, 26; CVPR 23, 24, 25, 26; ACL 23; ICCV 23, 25; ECCV 24, 26; NeurIPS 23, 24, 26; ACCV 24; ICLR 25, 26; AISTATS 25, 26; ICML 25, 26