Computer Science and Engineering Department
Michigan State University
East Lansing, MI 48824, USA

Mobile: (+1)-347-574-5875
Email: zhan1853@msu.edu
Website: https://damon-demon.github.io

RESEARCH FOCUSES

Deep learning: Foundation Models, Computer Vision (generative models, image classification, object detection/tracking), AI Safety (adversarial attack & defense, machine unlearning)
Optimization: Sparsity learning for model/dataset compression, Black-box optimization

EDUCATION

Ph.D. Candidate in Computer Science, Michigan State University Jan. 2021– Present.
M.S. in Electrical Engineering, Columbia University Aug. 2018– Dec. 2019
B.Eng in Electronic and Electrical Engineering, University of Sheffield Sep. 2015– July 2018

SELECTED PUBLICATIONS

Google Scholar (* represents equal contribution)

[1] **Y. Zhang**, T. Wang, J. Gesi, Z. Wang, Y. Lu, J. Lin, S. Zhan, V. Gao, R. Jiao, J. Liu, K. Qian, Y. Tang, R. Xue, H. Zhang, Q. Cui, Y. Guo, D. Wang, "Shop-R1: Rewarding LLMs to Simulate Human Behavior in Online Shopping via Reinforcement Learning", Under Review.

[2] **Y. Zhang**, T. Zhi, J. Liu, S. Sang, L. Jiang, Q. Yan, S. Liu, L. Luo, "ID-Patch: Robust ID Association for Group Photo Personalization", *CVPR'25*

[3] **Y. Zhang**, X. Chen, J. Jia, Y. Zhang, C. Fan, J. Liu, M. Hong, K. Ding, S. Liu, "Defensive Unlearning with Adversarial Training for Robust Concept Erasure in Diffusion Models", *NeurIPS'24*

[4] **Y. Zhang***, J. Jia*, X. Chen, A. Chen, Y. Zhang, J. Liu, K. Ding, S. Liu, "To Generate or Not? Safety-Driven Unlearned Diffusion Models Are Still Easy To Generate Unsafe Images . . . For Now", *ECCV'24*

[5] A. Chen*, **Y. Zhang***, J. Jia, J. Diffenderfer, J. Liu, K. Parasyris, Y. Zhang, Z. Zhang, B. Kailkhura, S. Liu, "DeepZero: Scaling up Zeroth-Order Optimization for Deep Model Training", *ICLR'24*

[6] **Y. Zhang**, X. Chen, J. Jia, S. Jia, K. Ding "Text-Visual Prompting for Efficient 2D Temporal Video Grounding", *CVPR'23*

[7] **Y. Zhang***, A.K. Kamath*, Q. Wu*, Z. Fan*, W. Chen, Z. Wang, S. Chang, C. Hao, S. Liu, "Data-Model-Circuit Tri-Design for Ultra-light Video Intelligence on Edge Devices", *ASP-DAC'23*

[8] **Y. Zhang**, Y. Yao, J. Jia, J. Yi, M. Hong, S. Chang, S. Liu, "How to Robustify Black-Box ML Models? A Zeroth-Order Optimization Perspective", International Conference on Learning Representation (***ICLR'22 - Spotlight, acceptance rate 5%***)

RESEARCH EXPERIENCE

Human Online Shopping Behavior Simulation via RL June. 2025 - July. 2025

Supervisor: Dakuo Wang (NEU), Jiri Gesi (Amazon)

- Introduce RL into a simulation-oriented human online shopping behavior modeling task.
- Develop a reinforcement-learning framework with a hybrid reward design. It integrates a self-certainty signal for rationale generation with a hierarchical reward scheme for action prediction.
- **Publications**: [1]

Multi-ID Consistency for Personalized Diffusion Model — May. 2024 - Nov. 2024

Supervisor: Tiancheng Zhi (ByteDance)

- Explore how to link face ID features with their corresponding locations using visual patches in conditioning images, ensuring better resemblance and accurate position control without ID leakage.
- Removal of the reliance on auxiliary segmentation models, requiring only a single point for ID position control, as opposed to segmented masks or head bounding boxes.
- **Publications**: [2]

Adversarial Unlearning for Diffusion Model — Nov. 2023 - May. 2024

Supervisor: Sijia Liu (MSU)

- Explore the integration of AT with concept erasing (or machine unlearning) in DMs.
- Design a utility-retaining regularization using curated external retain prompt data to balance the trade-off between effective unlearning and high-quality image generation.
- **Publications**: [3]

Robustness Evaluation for Unlearned Diffusion Models — May. 2023 - Oct. 2023

Supervisor: Sijia Liu (MSU), Xin Chen (Intel)

- Propose an evaluation framework built upon adversarial attacks (also referred to as adversarial prompts), in order to discern the trustworthiness of these safety-driven unlearned DMs.
- Develop a novel adversarial learning approach called UnlearnDiff that leverages the inherent classification capabilities of DMs to streamline the generation of adversarial prompts.
- **Publications**: [4]

Scalable Model Training without Backpropogation — Jan. 2023 - May. 2023

Supervisor: Sijia Liu (MSU)

- Propose a sparsity-induced ZO training protocol that extends the model pruning methodology using only finite differences to explore and exploit the sparse DL prior in CGE.
- **Publications**: [5]

Efficient 2D Temporal Video Grounding (TVG) — May.- Dec. 2022

Supervisor: Xin Chen (Intel)

- Propose an effective and efficient framework to train 2D TVG models, in which we leverage text-visual prompting (TVP) to improve the utility of sparse 2D visual features
- **Publications**: [6]

Model Compression for Object Tracking — Sept. 2021 - May. 2022

Supervisor: Sijia Liu (MSU)

Collaborator: Callie Hao(Georgia Tech), Shiyu Chang(UCSB), Zhangyang Wang(UT Austin)

- Saliency-guided spatial data reduction method is devised to eliminate uninformative pixels from both the input frames as well as the intermediate feature maps
- Utilizing kernel-wise pattern-aware model sparsity to achieve hardware-friendly model compression.
- **Publications**: [7]