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. Zhi, J. Liu, S. Sang, L. Jiang, Q. Yan, S. Liu, L. Luo, "ID-Patch: Robust ID Association for Group Photo Personalization", *CVPR'25*

[2] **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*

[3] **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*

[4] J. Jia, Y. Zhang, **Y. Zhang**, J. Liu, B. Runwal, J. Diffenderfer, B. Kailkhura, S. Liu, "SOUL: Unlocking the Power of Second-Order Optimization for LLM Unlearning", *EMNLP'24*

[5] Y. Zhang*, P. Li*, J. Hong*, J. Li*, **Y. Zhang**, W. Zheng, P.-Y. Chen, J. D. Lee, W. Yin, M. Hong, Z. Wang, S. Liu, T. Chen, "Revisiting Zeroth-Order Optimization for Memory-Efficient LLM Fine-Tuning: A Benchmark", *ICML'24*

[6] 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*

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

[8] **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*

[9] **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

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**: [1]

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**: [2]

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**: [3]

Machine Unlearning for LLM Feb.- Sep. 2024

Supervisor: Sijia Liu (MSU)

- Develop a second-order unlearning framework, termed SOUL, built upon the second-order clipped stochastic optimization (Sophia)-based LLM training method.
- **Publications**: [4]

Memory-Efficient LLM Fine-Tuning Oct. 2023 - April. 2024

Supervisor: Tianlong Chen (UNC)

- Expands the exploration to a wider array of ZO optimization techniques, through a comprehensive benchmarking study across five LLM families (Roberta, OPT, LLaMA, Vicuna, Mistral)
- **Publications**: [5]

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**: [6]

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**: [7]

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**: [8]