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: Computer Vision (generative models, image classification, object detection/tracking), AI Safety (adversarial attack & defense, machine unlearning)
Optimization: sparse optimization 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***, 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", *Under Review.*

[2] 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*

[3] Y. Zhang*, **Y. Zhang***, A. Chen*, J. Jia, J. Liu, G. Liu, M. Hong, S. Chang, S. Liu, "Selectivity Drives Productivity: Efficient Dataset Pruning for Enhanced Transfer Learning", *NeurIPS'23*

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

[5] **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*

[6] **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

Evaluation Framework for unlearned DMs. [Research Intern@Intel] 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, making it as simple for DMs as it is for image classification attacks.
- Our research explores the (worst-case) robustness of unlearned DMs in eradicating unwanted concepts, styles, and objects, assessed by the generation of adversarial prompts.
- **Publications**: [1]

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.
- Develop the methods of feature reuse and forward parallelization to advance the practical implementations of ZO training.
- **Publications**: [2]

Dataset Pruning for Transfer Learning Oct. 2022 - May. 2023

Supervisor: Sijia Liu (MSU)

- Propose two new dataset pruning (DP) methods, label mapping and feature mapping, for supervised and self-supervised pretraining settings respectively, by revisiting the DP problem through the lens of source-target domain mapping.
- **Publications**: [3]

Efficient 2D Temporal Video Grounding (TVG) [**Research Intern@Intel**] 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
- Achieve empirical success of our proposal to boost the performance of 2D TVG on Charades-STA and ActivityNet Captions datasets
- **Publications**: [4]

Model Compression for Object Tracking [**DARPA IP2 Program**] Sept. 2021 - May. 2022

Supervisor: Sijia Liu (MSU)

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

- Reinforcement learning-based lightweight design for temporal data reduction.
- 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**: [5]

Robustification of Black-Box ML Models by Zeroth-Order Optimization Jan.2021-Oct.2021

Supervisor: Sijia Liu (MSU) Collaborator: Jinfeng Yi(JD AI), Mingyi Hong(UMN), Shiyu Chang(UCSB)

- Formulate black-box defense problem through the lens of zeroth-order (ZO) optimization
- Propose scalable ZO optimization method to tackle defense challenge in high dimension
- Achieve state-of-the-art certified robustness on CIFAR-10 and STL-10
- Extend black-box defense from image classification to image reconstruction
- **Publications**: [6]

SERVICE

- Reviewer for CVPR, ICLR, ICML, NeurIPS, KDD, ICASSP, ACMMM, TPAMI
- Student chair for ICML'22 workshop - AdvMl Frontier