

Education

Northeastern University

M.S. in Computer Science (GPA: 4.0/4.0)

Boston, MA

University of Colorado at Denver

B.A. in Mathematics and B.A. in Economics (GPA: 3.5/4.0)

Sept. 2023 – Present

Denver, Co

Sept. 2019 - May 2023

Research Interests

Diffusion Models, Vision-Language Models, Model Generalization, Parameter-Efficient Fine-Tuning

Research Experience

MMLAB at Shenzhen Institutes of Advanced Technology

Research Assistant

Shenzhen, China

Nov. 2024 - Aug. 2025

Generalizable VLM Fine-tuning Methods

Developed GLAD, a generalizable fine-tuning framework for vision-language models (VLMs) in few-shot scenarios. By integrating parameter-efficient LoRA modules with a lightweight alignment network and gradient-based regularization, the method improves cross-modal representation learning and alleviates overfitting, achieving state-of-the-art generalization across multiple benchmarks.

- Investigated the limitations of CLIP prompt-tuning methods in few-shot settings, identifying their tendency to overfit and rely on heavy regularization for generalization.
- Proposed a LoRA-based fine-tuning strategy that updates only a small number of parameters in both the image and text encoders, providing efficient adaptation to new tasks.
- Designed a lightweight AlignNet module to enhance text embeddings with image-aware signals, strengthening vision-language alignment and improving transferability.
- Introduced a SAM-inspired gradient regularization term that encourages robustness to worst-case neighborhood perturbations, guides parameter updates toward flatter optima, leading to flatter minima and improved generalization under limited data.
- Conducted extensive evaluations on 15 datasets spanning base-to-novel generalization, domain generalization, and cross-dataset transfer, demonstrating that GLAD consistently outperforms existing methods such as CoOp, CoCoOp, MaPLe, and PromptSRC.
- Performed ablation studies confirming the complementary benefits of AlignNet and gradient regularization, and validated efficiency through reduced computational overhead.
- Led the project as first author, including conceptualization, method design, implementation, large-scale benchmarking, and manuscript writing.

Personalized Image Editing via Diffusion Models

Developed a novel framework (TARA) that enables diffusion models to generate high-fidelity images involving multiple personalized concepts simultaneously. By introducing token-level masking and alignment strategies, the method prevents mutual interference among LoRA modules and ensures spatially consistent concept representation, achieving training-free multi-concept composition directly at inference time.

- Conducted a systematic analysis of existing LoRA-based personalization methods, identifying token interference and spatial misalignment as the key causes of degraded multi-concept generation.
- Designed Token Focus Masking (TFM) to constrain each LoRA module's effect to its associated rare token, preventing conflicts across modules in cross-attention layers.
- Proposed Token Alignment Loss (TAL) to align rare-token attention with class-token regions, ensuring accurate spatial localization and reducing feature leakage.
- Implemented the full TARA framework on Stable Diffusion V1.5 and SDXL, enabling training-free, prompt-driven multi-concept generation without additional training or external conditions.
- Carried out extensive experiments on the DreamBooth dataset, benchmarking performance with CLIP-T, CLIP-I, and DINO metrics, and demonstrated superior identity preservation compared to DB-LoRA, Mix-of-Show, and other baselines.
- Performed detailed ablation studies to isolate the effects of TFM and TAL, confirming their complementary contributions to robust multi-concept composition.

- Led the project as first author from conceptualization, algorithm design, and implementation to experimental validation and manuscript writing, with source code and models released publicly.

Course Projects

Data-Efficient Subset Selection for Image Classification

Northeastern University, 2024

Developed a data-centric framework for efficient training by selecting informative subsets from large datasets. The method leverages submodular optimization to identify representative and diverse samples within each class, enabling substantial reductions in training data while maintaining accuracy.

- Investigated the challenge of reducing training cost without sacrificing accuracy, analyzing limitations of dataset distillation and random subset selection.
- Proposed a submodular-based selection strategy that iteratively identifies samples maximizing diversity and informativeness, ensuring each class is well represented.
- Employed ResNet-18 as feature extractor in the selection process, guiding the identification of critical samples through submodular gain evaluation.
- Constructed training subsets at different retention ratios (10%–50%) and evaluated performance on ResNet-18, ResNet-50, ResNet-101, and Vision Transformer (ViT) architectures.
- Demonstrated that the proposed subset selection consistently outperforms random selection across all settings; at 50% data retention, ResNet-18 trained on selected subsets achieved 89.16% accuracy vs. 82.17% with random subsets, approaching full-data accuracy (92.14%).

Work Experience

MIVII Technology

Research Intern

Beijing, China

May 2024 - Sept. 2024

- Annotated 3D point cloud motion data aligned with synchronized RGB video frames.
- Investigated NeRF-based mesh extraction methods and evaluated their performance on internal datasets.

Publications

- **Y. Peng**, P. Wang, J. Liu, S. Chen. “GLAD: Generalizable Tuning for Vision-Language Models.” *ICCV Workshops*, 2025. [PDF]
- **Y. Peng**, L. Zheng, Y. Yang, Y. Huang, M. Yan, J. Liu, S. Chen. “TARA: Token-Aware LoRA for Composable Personalization in Diffusion Models.” 2025. (Submitted to *AAAI*) [PDF]

Honors & Awards

- First Place - “Yuan Geng Cup” Shenzhen Badminton Tournament, 2025
- First Place - USA National Collegiate Team Badminton Championships, 2024
- Third Place - “Li-Ning Cup” Badminton Tournament (Beijing University Division, Men’s Singles), 2018

Technical Skills

- **Programming:** Python, Java, C, C++, R, Swift
- **Deep Learning Frameworks:** PyTorch