
Education

University of California, Merced

Ph.D. student, Computer Science

08/2023 - Present

Advisor: Meng Tang

University of Southern California

M.Sc., Electrical Engineering

08/2021 - 05/2023

GPA: 3.83/4.0

Chongqing University of Posts and Telecommunications

B.E., Digital Media Technology

09/2016 - 06/2020

GPA: 3.85/4.0

Professional Skills

Programming Language:

Python, MATLAB, C/C++, JavaScript

Deep Learning Framework:

PyTorch, TensorFlow, Keras

Parallel and Distributed Computation:

CUDA C, PyCuda

Documentation Formatting:

Latex

Research Experience

Ph.D. Research | University of California, Merced

08/2023 – Present

Advisor: Meng Tang

- **Project 1: Contrastive Conditional–Unconditional Alignment for Long-tailed Diffusion Model [1]**

- We propose Contrastive Conditional-Unconditional Alignment for Diffusion Model (CCUA) with imbalanced data. Our proposed contrastive loss with conditional-unconditional alignment are easy to implement with standard DDPM and SBDM pipeline using both UNet-based architecture and Diffusion Transformer(DiT).
- Our first loss, Unsupervised Contrastive Loss (UCL), employs unsupervised contrastive learning loss with negative samples only, enhancing within-class diversity.
- Our second Alignment Loss (AL) aligns unconditional generation and conditional generation for the initial steps in the denoising process, facilitating knowledge sharing between head and tail classes.
- We improved the diversity and fidelity of tail class images for conditional generation while maintaining the quality of head class images for multiple datasets and various resolutions, in particular ImageNet-LT with 256x256 resolution.

- **Project 2: ReDistill: Residual Encoded Distillation for Peak Memory Reduction [2]**

- We propose a distillation framework tailored for **reducing the peak memory** of convolutional neural networks, which allows **aggressive downsampling** of feature maps via pooling layers, while incurring a **negligible accuracy drop**.
- We propose a **residual encoded distillation (RED)** block to align features between high-peak-memory teacher networks and low-peak-memory student networks, based on a **multiplicative gating mechanism** and **additive residual learning**.
- For image classification tasks, our method **yields about $2\times \sim 3.2\times$ reduction in measured peak memory** with a slight decrease in the classification accuracies for CNN based models. Additionally, our method **improves the accuracy of compact ViT based models**, when distilled from large CNNs.
- We also show the versatility of our distillation method for image generation. For a U-Net based denoising diffusion probabilistic method, our method **reduces the theoretical peak memory by $4\times$** while maintaining the fidelity and the diversity of synthesized images.

Research Assistant | Energy Efficient Secure Sustainable Computing Group of USC

12/2021 – 05/2023

Advisor: Peter A. Bearel

- **Project 1: Self-Attentive Pooling for Efficient Deep Learning [4]**

- Proposed a non-local self-attentive pooling method to enable aggressive down-sampling with minimal accuracy loss.
- Hypothesized that combining local and non-local information during pooling reduces accuracy loss in early layers.

- Validated across STL10, VWW, ImageNet, and COCO with MobileNetV2/V3, ResNet-18, ResNeXt-18; outperformed SOTA pooling by 1.2% on ImageNet and achieved 1.43% higher accuracy under 22x memory reduction.

Research Assistant | Key Laboratory of Signal and Information Processing of Chongqing

03/2019 - 06/2021

Advisor: Chenqiang Gao

• Project 1: Local Patch Network for Infrared Small Target Detection [5]

- A **local patch network with global attention** was proposed to eliminate the **extreme class-imbalance**, that the main challenge of small target detection, between sparse small target pixels and low-rank background pixels, through **leveraging global and local features** of infrared small targets.
- Proposed an **attention module** to **suppress** most irrelevant **background pixels** from the **global view**, and a **local patch network (LPNet)** to **capture small targets** by viewing the attended feature maps patch by patch from the **local view**.
- The proposed method outperformed the state-of-the-art methods on two widely used public datasets and one of our private datasets under **probability of detection** ($\sim +3\%$), **AUC** ($\sim +7\%$) and **f1-measure** ($\sim +3\%$) metrics.

• Project 2: Infrared Small-Dim Target Detection under Complex Backgrounds [6]

- Based on the idea widely used in traditional methods that treating the **small target** as the **noise item**, the challenge was to **distinguish** the small target from the ground-truth **noise distribution** of background.
- Due to the ability of capturing **long-rang dependencies** of multi-head attention mechanism, a **Transformer and U-Net-like** skipped connection framework was proposed to capture the discriminative **differences** between **small target** and **global noise distribution** from complex backgrounds.
- The proposed method outperformed the state-of-the-art methods on two widely used public datasets under **probability of detection** ($\sim +3\%$), **AUC** ($\sim +8\%$) and **f1-measure** ($\sim +2\%$) metrics, and was especially effective on **cross-scene generalization** and **anti-noise performance**.

Work Experience

Graduate Technical Intern | Intel AI Lab, U.S.

06/2022 - 01/2023

Supervisor: Anthony Sarah

• Project 1: Simq-nas: Simultaneous quantization policy and neural architecture search [3]

- Developed SimQ-NAS, a **post-training Mixed Precision Quantization and Neural Architecture jointly aware Search method** for deploying deep learning models on resource-constrained devices.
- Designed a quantization policy search to reduce SimQ-NAS search space and built a module enabling joint NAS integration.
- Identified a key issue with **non-activated parameters causing redundant subnetworks**; resolved it with a **masked encoding algorithm**.
- Integrated the dynamic quantization module into Intel's DyNAS-T. On ImageNet, achieved **75%** model size and **90%** inference time reduction with only **3.75%** accuracy drop.

Honors and Awards

10/2021

Best Masters Poster Award of the 11th Annual Research Festival by **USC Ming Hsieh Institute**

06/2020

Outstanding Graduate of Chongqing (Provincial Level, in top 0.1%)

Publications

- [1] **Fang Chen**, Alex Villa, Gongbo Liang, Xiaoyi Lu, Meng Tang. Contrastive Conditional-Unconditional Alignment for Long-tailed Diffusion Model. *Submitted to International Conference on Learning Representations (ICLR), 2026*
- [2] **Fang Chen**, Gourav Datta, Mujahid Al Rafi, Hyeran Jeon, Meng Tang. ReDistill: Residual Encoded Distillation for Peak Memory Reductio. *In Transactions on Machine Learning Research, April 2025.*
- [3] Sharath Nittur Sridhar, Maciej Szankin, **Fang Chen**, Sairam Sundaresan, Anthony Sarah. SimQ-NAS: Simultaneous Quantization Policy and Neural Architecture Search. *Accepted by AAAI Edge Intelligence Workshop, 2024.*
- [4] **Fang Chen**, Gourav Datta, Souvik Kundu, and Peter Beerel. Self-attentive pooling for efficient deep learning. *In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3974-3983, 2023.*
- [5] **Fang Chen**, Chenqiang Gao, Fangcen Liu, Yue Zhao, Yuxi Zhou, Deyu Meng, and Wangmeng Zuo. Local patch network with global attention for infrared small target detection. *In IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 5, pp. 3979-3991, 2022.
- [6] Fangcen Liu, Chenqiang Gao, **Fang Chen**, Deyu Meng, Wangmeng Zuo, and Xinbo Gao. Infrared small-dim target detection with transformer under complex backgrounds. *In IEEE Transactions on Image Processing*, vol. 32, pp. 5921-5932, 2023.
- [7] Fengshun Zhou, Chenqiang Gao, **Fang Chen**, Chaoyu Li, Xindou Li, Feng Yang, and Yue Zhao. Face anti-spoofing based on multi-layer domain adaptation. *In IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, pp. 192-197, 2019.