Fang Chen

4368 Briggs Ln, Merced, CA 95348
(213)-284-1488
fchen20@ucmerced.edu

Education

University of California, Merced *08/2023 - Present*
Ph.D. student, Computer Science *Advisor: Meng Tang*

University of Southern California *08/2021 - 05/2023*
M.Sc., Electrical Engineering *GPA: **3.83**/4.0*

Chongqing University of Posts and Telecommunications *09/2016 - 06/2020*
B.E., Digital Media Technology *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

Work Experience

Research Intern | Intel AI Labs US
Supervisor: *Anthony Sarah* *06/2022 - 01/2023*

- A **post-training Mixed Precision Quantization and Neural Architecture jointly aware Search (DyQ-NAS) method** was proposed for the deployment of extensive fully trained deep learning architectures on resource-constrained devices.
- Proposed a **feasible quantization policy search method** to reduce the search space size of DyQ-NAS, and developed the **mixed precision quantization module** that enables the quantization policy to be customized and jointly searched by NAS methods.
- Based on the observation and analysis to the performance of accuracy / latency predictors, **found and located a vital issue** that a set of different configures might generate the same subnetworks due to nonactivated parameters; and proposed a **masked encoding algorithm** for configuration parsing to address this issue.
- The dynamic quantization module has been merged into main branch of **Intel Dynamic Neural Architecture Search Toolkit (DyNAS-T)**. Based on ImageNet dataset, the quantized subnetworks averagely achieve **75%** reduction of model size, **90%** reduction of inference time with only **3.75%** reduction of accuracy.

Research Experience

Directed Research | Energy Efficient Secure Sustainable Computing Group of USC
Advisor: *Peter A. Beerel* *12/2021 – 05/2023*

- ***Project 1: Residual Encoded Distillation For Efficient Image Recognition***
  - To reduce the **activation peak memory**, a novel **knowledge distillation method** is proposed for aggressive feature aggregation, which enables the initial down-sampling layers to be assigned with large kernel sizes and strides with mitigated low-level representation loss.
  - Existing distillation methods mainly focus on the feature space transfer of convolutional layers and fully connected layers between the student and the teacher architectures, which might be inappropriate for pooling layers in which the input and output activations are expected to fit a similar distribution. A **residual encoded module** is proposed to adjust the feature distribution so that enables the knowledge distillation methods to be specified for the pooling layers.
  - Extensive experiments demonstrate that our method achieves the peak activation memory reduction by $16\times$ with only $\sim 2.5\%$ on average accuracy degradation on widely used image recognition datasets.
- ***Project 2: Self-Attentive Pooling for Efficient Deep Learning [1]***
  - A **non-local self-attentive pooling method** was proposed to address the issue that current pooling methods perform poorly in **aggressive feature aggregation**, of which the main purpose was to assign the pooling methods **with large pooling strides** but **without too much accuracy loss.**

- Based on the analysis to activations, we hypothesized that the accuracy loss typically associated with aggressive down-sampling could be minimized by **considering both local and non-local information** during down-sampling.
- Extensive experiments on standard **image recognition** (STL10, VWW, ImageNet) and **object detection** (Microsoft COCO) datasets with various backbone networks (MobileNetV2, MobileNetV3, ResNet-18, ResNeXt-18) demonstrated the superiority of our proposed mechanism over the state-of-the-art (SOTA) pooling techniques. For instance, we surpassed the **test accuracy** of existing methods on different variants of MobileNet-V2 on ImageNet by an average of ∼**1.2%**. With the aggressive down-sampling of the activations in the initial layers (providing up to **22x reduction** in **memory consumption**), our approach achieved **1.43%** higher **test accuracy** compared to SOTA techniques with iso-memory footprints.

Research Assistant | Key Laboratory of Signal and Information Processing of Chongqing
Advisor: *Chenqiang Gao* *03/2019 - 06/2021*

- ***Project 1: Local Patch Network for Infrared Small Target Detection [2]***
  - 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** (∼ **+3%**), **AUC** (∼ **+7%**) and **f1-measure** (∼ **+3%**) metrics.
- ***Project 2: Infrared Small-Dim Target Detection under Complex Backgrounds [3]***
  - 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** (∼ **+3%**), **AUC** (∼ **+8%**) and **f1-measure** (∼ **+2%**) metrics, and was especially effective on **cross-scene generalization** and **anti-noise performance**.
- ***Project 3: Face Anti-spoofing Based on Multi-layer Domain Adaptation [4]***
  - A **face anti-spoofing detection algorithm** based on **domain adaptation** was proposed to address the issue that the state-of-the-art methods might performed poorly in **cross scenes**.
  - The fusion between **mid-level** and **high-level** features of CNN was adapted to improve the model performance; and a **multi-layer maximum mean discrepancy loss** was utilized to distill the **cross scenes information** from the source domain to the target domain.
  - The proposed method had the lowest **EER** (reduced from **40.0%** to **30.0%**) and **HTER** (reduced from **80.0%** to **60.0%**) compared to the state-of-the-arts methods on the Replay-Attack dataset. Under **cross scenes**, the proposed model trained on the Replay-Attack dataset and validated on the CASIA Face Anti-Spoofing dataset had the best **EER** (reduced from **34.5%** to **34.3%**).

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%)
11/2019	**Annual Progress Scholarship** in 2018-2019 Academic Year (in top 0.1%)
07/2019	**Silver Award** (**Rank 2** out of 300+ teams) and **Best Report** in **IEEE ISI World Cup 2019 (IWC 2019)**
11/2017	**Second Award** of Chongqing Division in China Undergraduate Mathematical Contest in Modeling

Publications

[1] **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.*

[2] **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.

[3] 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.

[4] 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.