Zangwei Zheng

✉ zangwei@comp.nus.edu.sg · ⌂ zhengzangw.github.io

EDUCATION

National University of Singapore	Aug. 2021 – Jul. 2025 (expected)
Ph.D. majored in Computer Science, supervised by Prof. Yang You	*Singapore, Singapore*
Nanjing University	Sept. 2017 – June 2021
B.S. in Computer Science and Technology, National Elite Program of Computer Science	*Jiangsu, China*

- **GPA:** 4.61/5.00 (92.2/100, top 2%)

RESEARCH INTEREST

Machine Learning: Efficient Foundation model, Efficient Training and Inference
High Performance Computing: Optimizer for Large-scale Deep Learning, Large Batch Training
ML Application: Efficient Learning in NLP, CV and Recommendation System

ACADEMIC RESEARCH EXPERIENCE

National University of Singapore (HPC-AI)	May 2019 – Present
Ph.D. student, supervised by Prof. Yang You	*Singapore, Singapore*

- Large language model inference acceleration by predicting response length and sequence scheduling.
- Continual learning of vision-language model to prevent zero-shot performance degradation.
- Acceleration of recommendation system training by large batch training.
- Introduce prompt learning for domain generalization with vision transformer.

University of California, Berkeley (iCyPhy, DOP Center)	Apr. 2020 – May. 2021
Research intern, supervised by Prof. Alberto Sangiovanni-Vincentelli & Dr. Xiangyu Yue	*(remote) CA, US*

- Few-shot Domain Adaptation via Self-supervised Learning with Clustering
- Proposed scene-aware learning with better backbones and data augmentations for radar object detection.

INDUSTRY RESEARCH EXPERIENCE

ByteDance Inc.	June 2021 – present
Research intern, in charge of large batch training for click-through rate prediction model	*Singapore, Singapore*

- Transformed the asynchronous CTR training model into the large-scale synchronous training framework.
- Deployed CowClip algorithm with batch size 512k and improved the AUC of CTR prediction (**AAAI 2023**).

PUBLICATIONS

* denotes equal contribution.

1. **A Study on Transformer Configuration and Training Objective**
   Fuzhao Xue, Jianghai Chen, Aixin Sun, Xiaozhe Ren, **Zangwei Zheng**, Xiaoxin He, Yongming Chen, Xin Jiang, Yang You
   ICML 2023

2. **CAME: Confidence-guided Adaptive Memory Efficient Optimization**
   Yang Luo, Xiaozhe Ren, **Zangwei Zheng**, Xin Jiang, Zhuo Jiang, Yang You
   ACL 2023

3. **CowClip: Reducing CTR Prediction Model Training Time from 12 hours to 10 minutes on 1 GPU**
   Zangwei Zheng, Pengtai Xu, Xuan Zou, Da Tang, Zhen Li, Chenguang Xi, Peng Wu, Leqi Zou, Yijie Zhu, Ming Chen, Xiangzhuo Ding, Fuzhao Xue, Ziheng Qing, Youlong Cheng, Yang You
   Distinguished Paper Award (0.1%), AAAI 2023

4. **Prototypical Cross-domain Self-supervised Learning for Few-shot Unsupervised Domain Adaptation**
   Xiangyu Yue*, **Zangwei Zheng***, Shanghang Zhang, Yang Gao, Trevor Darrell, Kurt Keutzer, Alberto Sangiovanni-Vincentelli
   CVPR 2021

5. **Scene-aware Learning Network for Radar Object Detection**
   Zangwei Zheng, Xiangyu Yue, Kurt Keutzer, Alberto Sangiovanni Vincentelli
   ICMR-W 2021

PREPRINTS

1. **InfoBatch: Lossless Training Speed Up by Unbiased Dynamic Data Pruning** Ziheng Qin, Kai Wang, **Zangwei Zheng**, Jianyang Gu, Xiangyu Peng, Daquan Zhou, Yang You **arXiv:2303.04947**
2. **Preventing Zero-Shot Transfer Degradation in Continual Learning of Vision-Language Models** **Zangwei Zheng**, Mingyuan Ma, Kai Wang, Ziheng Qin, Xiangyu Yue, Yang You **arXiv:2303.06628**
3. **Prompt vision transformer for domain generalization** **Zangwei Zheng**, Xiangyu Yue, Kai Wang, Yang You **arXiv:2208.08914**
4. **Sparse-MLP: A Fully-MLP Architecture with Conditional Computation** Yuxuan Lou, Fuzhao Xue, **Zangwei Zheng**, Yang You **arXiv:2109.02008**
5. **Multi-source Few-shot Domain Adaptation** Xiangyu Yue, **Zangwei Zheng**, Hari Prasanna Das, Kurt Keutzer, Alberto Sangiovanni Vincentelli **arXiv:2109.12391**

Skills

Languages	Python, C, C++, LaTeX
Frameworks	PyTorch, TensorFlow, Huggingface, OpenCV, Scikit-learn, NumPy