BAI, HAOLI

HOMEPAGE: haolibai.github.io
(+852) 5531 8737 / (+86) 177 1003 0395 ⋄ haolibai@gmail.com
Rm 101A, SHB, CUHK, Hong Kong.

RESEARCH INTEREST

My research interest majorly lies in **efficient deep learning** (e.g., **network quantization**, **pruning**, **knowledge distillation**) for computer vision and natural languages. We recently focus on **accelerating LLMs** to reduce the latency and cost of deployment.

EDUCATIION

The Chinese University of Hong Kong Aug. 2017 - Oct. 2021
PhD in Computer Science and Engineering
Supervisors: Michael Lyu and Irwin King

University of Electronic Science and Technology of China Sep. 2013 - Jun. 2017
BEng in Computer Science, Yingcai Honor's College
Supervisor: Zenglin Xu

EXPERIENCES

Huawei Noah's Ark Lab, Speech and Semantic Group, *Senior Researcher* Dec. 2021 - Now

Amazon Device, Halo Health Technology, *Internship* Jul. 2021 - Oct. 2021
Huawei Noah's Ark Lab, Speech and Semantic Group, *Internship* Jul. 2020 - Jun. 2021
Tencent AI Lab, Machine Learning Group, *Internship* Jun. 2018 - Jun. 2020

SELECTED PUBLICATIONS

(*****: Equal contribution. Click here for the full publication list.)

1. Yingtao Zhang, **Haoli Bai**, Haokun Lin, Jialin Zhao, Lu Hou, Carlo Vittorio Cannistraci, Plug-and-Play: An Efficient Post-training Pruning Method for Large Language Models, ICLR 2024.
2. **Haoli Bai***, Zhiguang Liu*, Xiaojun Meng*, Wentao Li, Shuang Liu, Nian Xie, Rongfu Zheng, Liangwei Wang, Lu Hou, Jiansheng Wei, Xin Jiang, Qun Liu, Wukong-Reader: Multi-modal Pre-training for Fine-grained Visual Document Understanding, ACL 2023.
3. Chaofan Tao, Lu Hou, **Haoli Bai**, Jiansheng Wei, Xin Jiang, Qun Liu, Ping Luo, Ngai Wong, Structured Pruning for Efficient Generative Pre-trained Language Models, Findings of ACL 2023.
4. **Haoli Bai**, Lu Hou, Lifeng Shang, Xin Jiang, Qun Liu, Irwin King, Michael Lyu. Towards Efficient Post-training Quantization of Pre-trained Language Models. NeurIPS 2022.
5. **Haoli Bai**, Hongda Mao, Dinesh Nair, Dynamically Pruning Segformer for Efficient Semantic Segmentation, ICASSP 2022.
6. **Haoli Bai**, Wei Zhang, Lu Hou, Lifeng Shang, Jin Jing, Xin Jiang, Qun Liu, Michael Lyu, Irwin King. BinaryBERT: Pushing the Limit of BERT Quantization, ACL, 2021. (Rating: **5, 5, 4**)
7. **Haoli Bai***, Jiaxing Wang*, Jiaxiang Wu, Xupeng Shi, Junzhou Huang, Irwin King, Michael Lyu, and Jian Cheng. Revisiting Parameter Sharing for Automatic Channel Number Search, NeurlPS, 2020.

8. **Haoli Bai**, Jiaxiang Wu, Irwin King, Michale Lyu. Few Shot Network Compression via Cross Distillation, AAAI, 2020.
9. Jiaxing Wang, Jiaxiang Wu, **Haoli Bai**, Jian Cheng. MetaNAS: Meta Neural Architecture Search, AAAI, 2020.
10. Yuhang Li, Xin Dong, Saiqian Zhang, **Haoli Bai**, Yuanpeng Chen, Wei Wang. RTN: Reparameterized Ternary Network, AAAI, 2020.
11. **Haoli Bai**, Zhuangbin Chen, Michael Lyu, Irwin King and Zenglin Xu. Neural Relational Topic Models for Scientific Articles, CIKM, 2018.
12. **Haoli Bai**, Zenglin Xu, Bin Liu and Yingming Li. Hierarchical Probabilistic Matrix Factorization with Network Topology for Multi-relational Social Network, ACML, 2016. **Best Student Paper Runner-up**.

SERVICES

Senior PC Member: IJCAI-21

PC Member: ICML 21-23, NeurIPS 20-23, AAAI 19-22, IJCAI 20, ICLR 21-23

Journal Reviewer: Cognitive Computation, Neural Networks, Neurocomputing

SELECTED AWARDS

Outstanding Intern at Huawei Noah's Ark Lab, 2021.

Student Travel Grant of CIKM 2018, AAAI 2020.

Postgraduate Studentship of the Chinese University of Hong Kong, 2017-2021.

Best Student Paper Runner-up of Asian Conference on Machine Learning, 2016.

National Scholarship (Top 2%), 2015

Meritorious Winner of the American Mathematical Contest in Modeling, 2016.

TECHNICAL SKILLS

Programming	PyTorch, Tensorflow, Python, MATLAB
Developping Tools	Git, Vim, Linux
TOEFL	100 (R:26, L:25, S:23, W:26)