

王少博

✉ shaobowang1009@sjtu.edu.cn · ☎ (+86)-15000937315 · 🌐 <https://gszfwsb.github.io>

🎓 教育背景

上海交通大学，人工智能，博士	2024.09 – 至今
导师：张林峰助理教授，研究方向：高效深度学习，深度学习理论	
上海交通大学，计算机科学与技术，硕士	2022.09 – 2024.09
导师：严骏驰教授，研究方向：深度学习理论，可解释性机器学习	
GPA: 3.98/4.0, 排名: 2/81	
哈尔滨工业大学，软件工程，本科	2018.09 – 2022.06
GPA: 3.96/4.0, 排名: 1/181	

📄 代表性成果 (* 表示同等贡献)

Direction 1: (Theory-Inspired) Application of Deep Learning

Not All Samples should be Utilized Equally: Towards Understanding and Improving Dataset Distillation R01

under submission to ICLR 2025

Shaobo Wang, Yantai Yang, Qilong Wang, Kaixin Li, Linfeng Zhang, Junchi Yan.

Gnothi Seauton: Empowering Faithful Self-Interpretability in Black-Box Models R02

under submission to ICLR 2025

Shaobo Wang, Hongxuan Tang, Mingyang Wang, Hongrui Zhang, Xuyang Liu, Xuming Hu, Linfeng Zhang.

DRUPI: Dataset Reduction Using Privileged Information R03

under submission to ICLR 2025

Shaobo Wang, Yantai Yang, Shuaiyu Zhang, Xuming Hu, Linfeng Zhang.

Unified Batch Normalization: Identifying and Alleviating the Feature Condensation in Batch Normalization and a Unified Framework R04

arxiv preprint, 2023

Shaobo Wang, Xiangdong Zhang, Dongrui Liu, Junchi Yan.

Direction 2: (Theoretical) Understanding of Deep Learning

Unveiling Induction Heads: Provable Training Dynamics and Feature Learning in Transformers R05

The Thirty-Eighth Annual Conference on Neural Information Processing Systems (NeurIPS 2024)

Siyu Chen, Heejune Sheen, Shaobo Wang, Tianhao Wang, Zhuoran Yang.

Extracting Consistent Attribution Values *w.r.t.* Different Partitions of Input Variables R06

Shaobo Wang*, Huiqi Deng*, Hanying Zhang, Quanshi Zhang.

Trap of Feature Diversity in the Learning of MLPs R07

arxiv preprint, 2022

Dongrui Liu*, Shaobo Wang*, Jie Ren, Kangrui Wang, Sheng Yin, Huiqi Deng, Quanshi Zhang.

Visualizing the Emergence of Intermediate Visual Patterns in DNNs R08

The Thirty-Five Annual Conference on Neural Information Processing Systems (NeurIPS 2021)

Mingjie Li, Shaobo Wang, Quanshi Zhang.

Other Directions

Think2Drive: Brick by Brick to Build a Model-based RL Agent for Quasi-Realistic Autonomous Driving (in CARLA-v2) R09

The 18th European Conference on Computer Vision ECCV 2024 (ECCV, 2024)

Qifeng Li, Xiaosong Jia, Shaobo Wang, Junchi Yan.

科研实习经历

中国工商银行，大数据与人工智能部门 2024.09 – 至今
科研实习生（金融大模型）

上海交通大学 EPIC lab 2024.03 – 至今

科研助理（高效深度学习、可解释深度学习），师从张林峰助理教授

在张林峰教授指导下，参与机器学习、数据蒸馏等方向研究。主要包括

- 首次实现了较大黑盒模型的自解释能力。基于旁路网络和高效微调构建了模型解释旁路，实现了对于黑盒模型的高效自解释。该方法降低了 98%+ 以上的训练参数量，解释准确率为领域内 SOTA，解释成本远低于 FastSHAP、ViT-Shapley 等模型相关方法，成果投稿于 ICLR 2025[R02]。
- 为蒸馏数据集引入“特权信息”，提高数据集的鲁棒性，提升各种数据集蒸馏方法的性能，成果投稿于 ICLR 2025[R03]。

上海交通大学 ReThinklab 2023.08 – 至今

硕士研究生（深度学习理论、机器学习、计算机视觉），师从严骏驰教授

在严骏驰教授指导下，参与机器学习、计算机视觉研究。包括但不限于

- 从神经扩展原则的角度理论理解数据集蒸馏任务，提出样本困难度矫正方案，在基于匹配的数据集蒸馏方法上取得最佳效果，成果投稿于 CVPR 2025[R01]。
- 发现批归一化操作中的“特征坍塌”现象，提出了一个自适应的基于特征相似度的归一化矫正方案，在利用归一化操作的神经网络中取得最佳效果，成果投稿于 CVPR 2025[R04]。
- 参与基于世界模型的端到端自动驾驶模型，作为首次解决 CARLA v2 仿真环境上的方案。成果收录于 ECCV 2024[R09]。

耶鲁大学统计与数据科学系 2024.03 – 至今

科研实习（大模型理论），师从杨卓然助理教授

- 从理论角度推导了 Transformer 在上下文学习中学习 n-gram 语言规则的学习动力学和最优收敛性，实验角度验证了 Transformer 可以捕捉不同形式数据的因果关系。成果收录于 NeurIPS 2024[R05]。

主要获奖情况

国家奖学金 2019

哈尔滨工业大学，中国教育部

哈尔滨工业大学优秀毕业生 2022

哈尔滨工业大学

哈尔滨工业大学（威海）最具影响力毕业生 2022

哈尔滨工业大学（威海）

山东省优秀毕业生 2022

哈尔滨工业大学，山东省教育部