

陈文熙

(+86) 189-9837-4728 · 1029713857@sjtu.edu.cn · 微信号: worst_chan · GitHub: cw-x-worst-one

个人总结

本人在校成绩优秀、乐观向上，工作负责、自我驱动力强、热爱尝试新事物。在校期间主要从事音频自监督学习、音频分类、复杂音频理解、音频-语言大模型应用等相关研究，对深度学习在音频领域的发展趋势以及大语言模型与音频领域的结合有浓厚兴趣。

教育背景

上海交通大学 | 本科在读 2021.9 - 2025.6

计算机科学与技术 (IEEE 试点班) | 电子信息与电气工程学院

- **GPA: 4.07/4.30; 核心学积分: 93.64/100; 排名: 5/130**
- **专业课程:** 计算导论 (96); 程序设计原理与方法 (96); 数据结构 (97); 算法设计与分析 (93); 计算机网络 (96); 程序语言与编译原理 (95); 操作系统 (94); 自然语言处理 (97) 等
- **荣誉奖项:** 上海交通大学本科生 C 等优秀奖学金 (2021,2022,2023)
- **英语成绩:** CET4: 608; CET6: 515
- **专业技能:** 掌握 python 和 c++ 语言, 熟悉 pytorch 框架, 熟悉 huggingface、fairseq 等开源项目
- **研究兴趣与方向:** 音频自监督学习、音频分类、复杂音频理解、音频-语言大模型应用等

科研项目

EAT: Self-Supervised Pre-Training with Efficient Audio Transformer

上海交通大学计算机科学与工程系 (跨媒体语言智能实验室) 导师: 陈谐 2023.8-2023.12


- **科研内容:** 在该科研项目中, 我们提出了一个基于掩码式自蒸馏学习框架的音频自监督模型 EAT, 通过同时利用帧 (frame) 级别以及整体音频 (utterance) 级别的目标损失来挖掘模型对音频片段的学习能力。此外, EAT 通过高掩码率 (80%) 以及轻量级的 CNN 解码器很好地提升了预训练效率。实验中创新性地发现了逆向的块状掩码以及对全局音频的学习极大地提升了模型的表征学习性能。在该框架与设定的预训练任务下, 模型可以更有效率地学习音频表征, 并在相应下游任务上表现优异的性能。
- **科研贡献:** 课题的主要研究者, 提出并改进 EAT 框架, 负责大部分实验内容以及最终论文撰写。
- **科研结果:** EAT 模型在 AudioSet (AS-2M, AS-20K), ESC-50, SPC-2 等常见的音频/语音分类任务上实现了 SOTA 的性能, 同时与过去的 BEATs, Audio-MAE 模型相比在预训练总时间上取得了 10 倍以上训练速度提升。该研究成果的论文已被 **IJCAI 2024 (CCF-A)** 接收。(一作)

IEEE ICME 2024 Grand Challenge: Semi-supervised Acoustic Scene Classification under Domain Shift

上海交通大学计算机科学与工程系 (跨媒体语言智能实验室) 导师: 陈谐 2024.2-2024.3

- **科研内容:** 在主题为“域转移下的半监督声学场景分类”的 ICME 2024 挑战比赛中, 我们通过结合自监督模型 EAT, 自学习的半监督技术以及 test-time adaptation 的推理策略来实现更好的域迁移情境下的声学场景分类 (ASC)。通过在给定数据集下按权重预训练 EAT 模型并在标签数据集下进行微调-打伪标签的自学习迭代, 模型在预测和训练集差别较大的评测数据标签上有明显优于 baseline 的性能。
- **科研贡献:** 参赛队伍队长, 主要负责训练框架的搭建和优化, 和组员共同完成模型训练, 微调, 推理任务, 共同撰写最后的技术报告和 ICME workshop 论文。
- **科研结果:** 最后评测结果中的声学场景分类准确性 (Macro-accuracy) 为 75.2%, 超过 baseline (60.0%) 15.2%, 在全球所有参赛队伍中排名第二, 与第一名系统 (科大讯飞团队) 仅有 0.6% 的微弱差距。比赛中合作的两篇论文均被 **IEEE ICME 2024 (CCF-B) Workshop** 接收。

论文发表

- **Chen, W., Liang, Y., Ma, Z., Zheng, Z., & Chen, X.** (2024). EAT: Self-Supervised Pre-Training with Efficient Audio Transformer. arXiv preprint arXiv:2401.03497. (accepted by IJCAI 2024) 
- **Liang, Y., Chen, W., et al.** Improving Acoustic Scene Classification via Self-supervised and Semi-supervised Learning with Efficient Audio Transformer. (accepted by ICME 2024 Workshop)
- **Huang, W., Jiang, A., Han, B., Zheng, X., Qiu, Y., Chen, W., et al.** Semi-supervised Acoustic Scene Classification with Test-time Adaptation. (accepted by ICME 2024 Workshop)

- Ma Z., Chen M., Zhang H., Zheng Z., **Chen W.**, et al. EmoBox: Multilingual Multi-corpus Speech Emotion Recognition Toolkit and Benchmark. (accepted by Interspeech 2024) 