

周殿松

☎ 19523465817 ✉ zhouds1918@163.com



🎓 教育背景

- 中国科学技术大学，通信工程，在读硕士 2022年9月 – 2025年7月
- 研究方向: 多模态大模型、视频理解 导师: 何向南
- 吉林大学，电子信息科学与技术，学士 2018年9月 – 2022年7月

📄 学术成果

- PosMLP-Video: Spatial and Temporal Relative Position Encoding for Efficient Video Recognition**
IJCV, CCF-A类期刊, 与老师共一, 发表软著 2022年10月 – 2023年6月
- 问题背景: 纯MLP网络由于其 $O(N^2)$ 的复杂度, 在处理视频时往往面临着参数量过多的问题。本文探索如何利用参数复杂度 $O(N)$ 的可学习相对位置编码替代MLP中的全连接层来缓解这个问题。
- 研究内容: 提出了一种轻量且强大的MLP视频识别骨干网络: PosMLP-Video。使用高效的分组可学习相对位置编码(GLRPE)取代全连接层来建模时空token关系。与传统纯MLP网络相比, 参数复杂度由 $O(N^2)$ 下降到了 $O(N)$ 。对比同为MLP网络的 MorphMLP, 在K400上的准确率由79.5提升到80.3, 参数量由67.8M下降到19.1M, GFLOPs由98下降到59。
- Adapting CLIP with Select-First-Feed-Then for Efficient Video Action Recognition**
AAAI, CCF-A类会议, 第一作者, 在投 2023年12月 – 2024年4月
- 问题背景: 将CLIP适配到视频领域, 大量工作往往关注如何减少可训练的参数, 忽略了CLIP本身带来的庞大计算量。本文探索如何从数据输入的角度来减轻CLIP处理视频时的计算负担。
- 研究内容: 提出了一种先选择再输入的机制SFFT, 适配CLIP用于高效的视频识别。使用轻量级的视频理解网络作为帧选择器, 根据视频语义减少输入视频帧, 将计算量减少近乎一半, 并不损失精度。对比ST-Adapter, SFFT在K400上的准确率由82.7提升到84.7, GFLOPs由1821下降到1101。并且相较于ST-Adapter, 由于SFFT采用了对比学习的方式训练, 所以具有zero-shot能力。

💼 实习经历

- 美团, 自动车配送部感知组, 多模态大模型算法实习生 2024年4月 – 2024年7月
- 项目背景: 自动驾驶任务具有大量的长尾问题。车端小模型泛化能力有限, 不具备开集检测能力。多模态大模型具有强大的场景理解和泛化能力。本项目构建高危场景多模态大模型, 用于在云端辅助自动驾驶。
- 工作内容: 发现LLaVA对于当前任务的不足之处-分辨率过低。调研高分辨率多模态大模型, 使用S2模型解决分辨率问题, 在多个任务上的平均precision和recall提升了4.5%和6.76%。其中施工区域任务突破了90%的性能瓶颈, 达到了95%。并根据当前任务特性, 修改S2模型, 使推理速度相较于原版S2模型提升46%。使用Lora微调多模态大模型。梳理任务逻辑, 制定数据标注规则, 使用GUI编写数据标注脚本, 构造prompt。
- 数据空间研究院, 认知推荐团队, 计算机视觉算法实习生 2023年8月 – 2023年12月
- 项目背景: 负责面部特征分析任务, 难点在于没有数据集提供。(由于项目需要保密, 所以只简要介绍)
- 工作内容: 基于没有数据集进行训练的难点, 选择在成熟的人脸属性分析开源算法上进行逻辑封装的方案。

🔧 掌握技能

- 编程技能: 掌握Python编程, Pytorch深度学习框架, 熟悉linux系统操作, Java编程。
- 语言技能: 英语CET-6: 478分。

🏆 个人荣誉

中国科学技术大学学业二等奖学金, 吉林大学三等奖学金