

吴泳良

15558561706 | yongliangwu@seu.edu.cn

教育经历

东南大学 - 计算机科学与技术 硕士

2023.09 - 2026.06

东南大学 - 人工智能 本科

2019.08 - 2023.06

• GPA: 3.87/4.8 (排名:8/91)

期待岗位: 计算机视觉算法工程师; 自然语言处理算法工程师

项目经历

探索视觉语言大模型的少样本上下文学习配置

2023.03 - 2023.05

• 在发现语言模型 (LMs) 可以成为良好的上下文少样本学习器后, 过去一些工作提出了许多优化上下文序列配置的策略。最近, 视觉-语言模型 (VLMs) 领域的研究人员也开发了他们的少样本学习模型及策略, 然而他们只使用了最简单的方式, 即随机抽样, 来配置上下文图像-文本对。为了探索不同配置对于视觉-语言上下文学习的影响, 我们设计了四种图像选择策略和四种标题分配策略, 用于配置图像-标题对进行图像字幕生成。我们的综合实验得出了两个违反直觉但有价值的观点, 突显了多模态协同作用下视觉-语言上下文学习的独特特征, 同时我们最优的策略配置相比于随机抽取的基线有着20.9的 CIDEr scores 提升。本项目相关论文已经被机器学习领域顶会NeurIPS2023录用。论文链接 [2305.14800.pdf](https://arxiv.org/abs/2305.14800) (arxiv.org)。

基于音频驱动的BlendShape生成项目

2023.01 - 2023.05

• 本项目为上海人工智能实验室、美国[Hyper Online](https://www.hyperonline.com/)公司合作开发项目, 目的为仅使用音频对人脸的BlendShape进行推理生成。本项目使用一个MOCAP模型在一个大规模的音频-视觉数据集MEAD上提取对应的人脸BlendShape参数后进行训练, 避免了真实采集数据的过程。该模型主要应用在一个自定义的ChatBot上, 用户进行输入后由后台大模型产生回复文本, 经由TTS合成语音后生成对应的面部表情返回。目前第一版本已经上线至[Hyper: AI Companions & VTubers](https://www.hyperonline.com/), 目前正在进行更细致的情感表达改进, 目的为根据返回文本的内容产生不同情绪的面部表情。演示视频已上传至[B站 BV1qH4y1s7i4](https://www.bilibili.com/video/BV1qH4y1s7i4)。

基于多模态对比学习的RGB-D稠密预测任务研究

2021.11 - 2022.07

• 稠密预测任务是计算机视觉领域的一项基础课题, 其中主流有语义分割和显著物体检测两种。基于 RGB-D 数据集的稠密预测任务通过引入深度图片这一额外信息, 大大提高了系统性能。但由于其存在标注困难的问题, 目前的工作依赖于 ImageNet 有监督分类网络进行初始化, 且需要设计复杂的下游融合网络对两个模态间的互补性和冗余性进行取舍, 由于存在领域和模态的沟壑, 性能往往不佳。考虑到 RGB-D 数据提供了天然的配对样本, 我们将多模态对比学习应用到其中, 针对两者的空间完全匹配特性设计了多尺度配对模块, 使模型在不依赖标签的前提下学习到跨模态稠密特征以实现在减少对标签和模型初始化的依赖程度的同时提高稠密预测任务的性能。本项目已有一篇相关论文撰写完成, 已经投稿至TIM期刊, 并在申请一项专利。

实习经历

OPUS Clip - 算法实习生

2023.10 - 2024.02

视频自动剪辑项目 本项目主要针对视频的长剪短课题, 涉及到合适片段挑选, Reframe, B-roll 等子任务。为此我们设计了一个 LLM Agent, 集成了包括 ASR, Active Speaker Detection, Face Recognition 等基础模型, 用以实现自动的长剪短任务。同样的基于该方法所得到的数据, 我们提出了一个benchmark, 用以训练端对端的视频长剪短模型, 相关论文已经投稿至ICML2024。

梅特勒-托利多测量技术有限公司 - 计算机视觉算法工程师

2022.07 - 2023.01

商用收银称果蔬识别项目 本项目基于商超收银称进行开发, 目标为实现果蔬的自动识别。通过采集相关数据训练分类器, 实现在训练、测试集类别存在不重叠现象的果蔬零食 One-Shot 识别功能。本人负责相关网络特征提取框架及自监督算法研究, 最终模型能够在相关测试集下达到92.8%的Top-1和99.3%的Top-5准确率。目前该模型已加装在一款上市销售产品上——[FreshBase Plus AI智能识别秤](https://www.freshbase.com/)

科研论文

- Xu Yang, **Yongliang Wu**, Mingzhuo Yang, Haokun Chen, Xin Geng. Exploring Diverse In-Context Configurations for Image Captioning. In: Advances in Neural Information Processing Systems (NeurIPS'23), 2023. (CCF-A, 导师一作)
- **Yongliang Wu**, Xu Yang. A Glance at In-Context Learning. In: Frontiers of Computer Science, 2024. (CCF-B)