

魏红陈

手机 132-0180-7991

邮箱 hc_wei@whu.edu.cn

更新日期 2025-03-21

教育背景

西安石油大学, 材料科学与工程学院, 学士	2016.09 - 2020.06
南京理工大学, 计算机科学与工程学院, 硕士,	2020.09 - 2023.04
武汉大学, 遥感信息与工程学院, 博士	2023.04 - 预计 2026.06

- 博士研究组 & 导师: 人工智能与机器感知实验室 (IIP); 陈震中教授
- 研究方向: 多模态大语言模型; 视觉内容理解

开源项目

- 基于单模态语言推理模型 QwQ-32B 和多模态非推理模型 InternVL-38B, 构建了第一个原始的 R1-like 多模态推理模型——FRANK-ZERO 38B (<https://huggingface.co/hcwei/FRANK-ZERO-38B>)。该模型在 MathVista test-mini 基准测试中取得了 74.1% 的准确率, 超越了闭源的 OpenAI o1 模型 (73.9% vs. 74.1%), 并展现出卓越的 long CoT 思维推理和反思能力。

科研工作

成果概览: 围绕半监督学习、图像/视频内容描述、视频问答、时空视频定位等发表论文8篇 学术论文 (含投稿中论文), 具体如下:

- S2OSC: A Holistic Semi-Supervised Approach for Open Set Classification. TKDD'22. [CCF-B 期刊, 学生一作]
 - 提出了一种新的开放集分类算法, 通过半监督学习结合类外实例过滤和模型重训练来解决嵌入混淆问题, 并在增量学习场景下扩展为 I-S2OSC, 实验证明其在多种 OSC 任务中表现优异。
- Exploiting Cross-Modal Prediction and Relation Consistency for Semi-Supervised Image Captioning. TCYB'23. [CCF-B 期刊, 学生一作]
 - 提出了一种半监督图像描述方法, 通过跨模态预测和关系一致性, 利用未标注图像约束生成句子的语义空间, 从而在有限标注数据的情况下提升描述生成效果。
- Improving Generalization of Image Captioning with Unsupervised Prompt Learning. TOMM'24. [CCF-B 期刊, 一作]
 - 提出了一种无监督提示学习方法, 通过对齐视觉和语言模态并优化领域特定提示向量, 在不依赖标注数据的情况下, 提升预训练视觉-语言模型在图像描述任务上的泛化能力。
- Visual Context Window Extension: A New Perspective for Long Video Understanding. [在投, 一作]
 - 提出了用于长视频理解的视觉上下文窗口扩展方法, 它能够将预训练的 MLLMs 轻松地扩展到 1024 帧, 并显著减少内存使用, 使 7B 模型在 MLVU 基准上超越了 GPT4o。
- Training-Free Extension from Images to Videos for Real-World Spatio-Temporal Video Grounding. [在投, 一作]
 - 提出了一种无需训练的灵活时空视频定位方法, 基于 MLLMs, 通过问答形式适应现实世界的时空视频定位, 并设计了时空解耦模块和查询引导的视觉标记过滤器来增强模型对时空信息的理解。
- LongCaptioning: Unlocking the Power of Long Caption Generation in Large Multimodal Models. [在投, 一作]
 - 研究了 MLLMs 在长视频描述生成中的输出长度限制问题, 发现训练数据中长描述样本的稀缺是主要原因, 并提出了 LongCaption-Agent 框架来合成长描述数据, 构建了 LongCaption-10K 数据集和 LongCaption-Bench 评估基准, 通过在 LongCaption-10K 上训练, 使 MLLMs 能够生成超过 1000 个单词的长描述, 并保证了生成质量。
- TDSAgent: A Task-Driven Sampling Agent for Long Video Question Answering. [在投, 其他]
 - 提出了一种用于长视频问答的 Agent 模型, 该模型通过任务驱动的采样代理系统, 首先利用查询引导的检索过程来判断视频片段的相关性, 然后根据相关性自适应地分配采样粒度, 在最大化保留相关信息的同时减少内存消耗。
- Remote Sensing Semantic Segmentation Quality Assessment based on Vision Language Model. [在投, 其他]
 - 提出基于预训练视觉语言模型 CLIP-RS 的无监督遥感影像语义分割质量评价评估框架 RS-SQA, 通过融合语义特征与分割中间特征构建高效评估方法, 并结合新构建的 RS-SQED 数据集验证其显著优于现有模型。

专利产出

- 基于注意力机制多视图深度学习的球鞋真伪鉴定方法. CN114186613A. [学生一作]

竞赛获奖

- 第四届中国高校计算机大赛: 人工智能创意赛, 队长, 华东赛区 top 1% 2021 年 4 月

- 项目名称: 基于深度神经网络的球鞋真伪鉴定
- 第四届百度飞桨论文复现赛: 多模态赛道, 队长, 第一名 2021 年 7 月
- 第三届 DIGIX 全球校园人工智能算法精英大赛, 队长, 一等奖 2021 年 9 月
- 赛题名称: 基于多目标多视图的用户留存周期预测
- 百度认知 AI 创意赛: 创意开发组, 队长, 二等奖 2022 年 4 月
- 项目名称: 基于 ERNIE3.0 大模型的自动家装描述
- 全国人工智能大赛: AI+ 视觉特征编码, 队员, 8/2839 2023 年 12 月

相关链接

- 个人主页: <https://hcwei13.github.io/>