

王家伟

✉ jarvisustc@gmail.com · ☎ (+86) 17398385512 · in wangjiawei

🎓 教育背景

- 中国科学技术大学, 安徽, 合肥 2021 – 至今
微软亚洲研究院联合培养博士研究生 信息科学与技术学院
研究方向: 智能文档处理、多模态学习
GPA: 3.8/4.3
- 中国科学技术大学, 安徽, 合肥 2017 – 2021
学士 少年班学院 计算机科学与技术 省优秀毕业生
GPA: 3.8/4.3

📄 论文发表

大语言模型

- Yichun Feng, **Jiawei Wang**, Lu Zhou, and Yixue Li. “A Scalable Retrieval-Augmented Reasoning Framework Based on Large Language Models for Knowledge Mining in Biomedical Literature”. In: *Submitted to Nature Communication* (2025), **Nature 子刊**.
- Jiawei Wang**, Yushen Zuo, Yuanjun Chai, Zhendong Liu, Yicheng Fu, Yichun Feng, and Kin-man Lam. “Safeguarding Vision-Language Models: Mitigating Vulnerabilities to Gaussian Noise in Perturbation-based Attacks”. In: *Submitted to CVPR 2025* (2025), **CCF-A**.
- DeepSeek-AI. “DeepSeek-V3 Technical Report”. In: *arXiv preprint arXiv:2412.19437* (2024), **Cutting-Edge Project**.
- DeepSeek-AI. “DeepSeek-VL2: Mixture-of-Experts Vision-Language Models for Advanced Multimodal Understanding”. In: *arXiv preprint arXiv:2412.10302* (2024), **Cutting-Edge Project**.

文档智能

- Jiawei Wang**, Kai Hu, and Qiang Huo. “UniHDSA: A Unified Relation Prediction Approach for Hierarchical Document Structure Analysis”. In: *Submitted to Pattern Recognition* (2025), **中科院一区**.
- Kai Hu, **Jiawei Wang**, Weihong Lin, Zhuoyao Zhong, Lei Sun, and Qiang Huo. “UniVIE: A Unified Label Space Approach to Visual Information Extraction from Form-like Documents”. In: *International Conference on Document Analysis and Recognition*. **Oral Presentation**, 2024.
- Jiawei Wang**, Kai Hu, and Qiang Huo. “DLAFormer: An End-to-End Transformer for Document Layout Analysis”. In: *International Conference on Document Analysis and Recognition*. **Oral Presentation**, 2024.
- Jiawei Wang**, Kai Hu, Zhuoyao Zhong, Lei Sun, and Qiang Huo. “Detect-Order-Construct: A Tree Construction based Approach for Hierarchical Document Structure Analysis”. In: *Pattern Recognition* (2024), **中科院一区**.
- Jiawei Wang**, Shunchi Zhang, Kai Hu, Chixiang Ma, Zhuoyao Zhong, Lei Sun, and Qiang Huo. “Dynamic Relation Transformer for Contextual Text Block Detection”. In: *International Conference on Document Analysis and Recognition*. **Oral Presentation**, 2024.
- Jiawei Wang**, Weihong Lin, Chixiang Ma, Mingze Li, Zheng Sun, Lei Sun, and Qiang Huo. “Robust Table Structure Recognition with Dynamic Queries Enhanced Detection Transformer”. In: *Pattern Recognition* (2023), **中科院一区**.
- Chixiang Ma, Lei Sun, **Jiawei Wang**, and Qiang Huo. “DQ-DETR: Dynamic Queries Enhanced Detection Transformer for Arbitrary Shape Text Detection”. In: *International Conference on Document Analysis and Recognition*. **Oral Presentation**, 2023.
- Zhuoyao Zhong, **Jiawei Wang**, Haiqing Sun, Kai Hu, Erhan Zhang, Lei Sun, and Qiang Huo. “A Hybrid Approach to Document Layout Analysis for Heterogeneous Document Images”. In: *International Conference on Document Analysis and Recognition*. 2023.

- [9] Weihong Lin, Zheng Sun, Chixiang Ma, Mingze Li, **Jiawei Wang**, Lei Sun, and Qiang Huo. "TSRFormer: Table Structure Recognition with Transformers". In: *Proceedings of the 30th ACM International Conference on Multimedia*. CCF-A, 2022.

👤 实习经历

华为 2012 实验室, 中央媒体技术院, 杭州

2019 年 12 月 – 2020 年 9 月

实习 导师: 张丽萍博士

- 主导研究照片对比度迁移, 探索深度神经网络在高动态范围 (HDR) 图像生成中的应用
- 设计并实现了高效的图像风格迁移模型, 在计算效率和视觉质量上取得显著提升

微软亚洲研究院, 多模态交互组, 北京

2020 年 9 月 – 至今

实习 导师: 霍强教授

- 研究复杂表格检测与识别算法, 提出将 Transformer 引入旋转目标检测任务, 开发高性能旋转表格检测器, 并创新性地采用渐进式分隔线回归方法, 高效解析拍照场景中的弯曲复杂表格
- 提出基于树构建的层级文档结构分析方法, 覆盖检测、排序与构建三阶段, 实现端到端的文档结构解析, 全面提升文档解析的准确性和效率
- 针对级联错误与任务分离问题, 提出基于统一关系预测的文档解析框架, 通过共享标注空间与统一预测模块, 显著提升系统的鲁棒性与扩展性

👤 项目经历

多模态大语言模型鲁棒性与安全性, 多机构合作项目

2024 年 8 月 – 2024 年 11 月

项目主导 / 论文一作

- 分析当前多模态视觉语言模型 (VLM) 训练中缺乏噪声增强的问题, 揭示其对高斯噪声等简单扰动的脆弱性
- 提出 DiffPure-VLM 防御框架, 将扩散模型与安全微调的 VLM 相结合, 有效将扰动噪声转化为高斯噪声, 增强模型的防御能力, 并支持与其他防御策略的兼容
- 构建了视觉语言安全性数据集 Robust-VLGuard, 在增强模型抗高斯噪声能力的同时保持实用性, 为针对性微调提供了高效工具
- 通过统计分析揭示 DiffPure 在对抗性噪声攻击下的局限性, 发现残余噪声往往保留并接近高斯分布, 进一步优化防御机制

通用文档解析器, Microsoft

2022 年 9 月 – 2024 年 9 月

核心开发者

- 专注于鲁棒文档结构分析算法的设计与实现, 其中基于树构建的文档结构分析算法已成功集成至 Azure AI Layout API, 为高效文档解析提供了强有力的支持
- 针对当前缺乏端到端解决方案的局限性, 提出了一种基于树构建的层级文档结构分析方法。该方法通过检测、排序和构建三个阶段直观建模文档的层级关系, 全面涵盖关键子任务, 实现了端到端的层级文档结构分析
- 为克服现有方法中的级联错误和任务分离问题, 创新性地提出了基于统一关系预测的层级文档结构分析方法。该方法通过共享标注空间和统一预测模块同时处理多个任务, 大幅提升了系统的效率、鲁棒性和扩展性

通用表格解析器, Microsoft

2020 年 9 月 – 2022 年 9 月

核心开发者

- 研究各类场景 (如扫描和拍照) 下文档图像中复杂表格的检测与识别算法的设计与实现。其中, 表格结构识别算法已成功集成至 Azure AI Layout API, 实现了高效且鲁棒的表格解析能力
- 针对旋转表格检测问题, 创新性地引入 Transformer 结构引入旋转目标检测任务, 并深入研究了 Detection Transformer (DETR) 在旋转目标检测中的独特鲁棒性。基于该研究成果, 成功开发出一款在旋转表格检测中表现卓越的表格检测器

- 针对复杂表格结构识别问题，创新性地将表格线检测重新定义为回归问题而非传统的分割问题，并首次将 **Transformer** 引入表格结构识别领域。在此基础上，提出了一种基于渐进式分隔线回归的表格结构识别算法，能够高效应对拍照场景中弯曲复杂表格的解析需求

粤港澳大湾区国际算法算例大赛 **VPT 赛道第二名**, 广州

2023 年 8 月 – 2023 年 11 月

团队队长 奖金: 20 万元人民币

- 通过集成大语言模型 (LLM)，有效增强了 CLIP 在多标签图像分类任务中的表现，显著提升了模型的准确性和鲁棒性。相关算法已开源至 **GitHub**
- 提出了创新的自反馈数据生成框架，利用 LLM 创建详细的场景描述数据，增强了数据集的多样性和质量
- 开发了先进的滑动窗口策略，结合多尺度、多块和多形状技术，在多标签分类任务中捕捉不同尺度和极端尺寸比率下对象的细粒度局部细节

🔧 IT 技能

- 编程语言: Python > C/C++ > SQL > HTML/CSS
- 平台: Linux, Shell (Bash/Zsh)
- 工具: PyTorch, NumPy, Scikit-learn, L^AT_EX(Overleaf/Markdown), Git
- 软技能: 时间管理, 团队合作, 问题解决, 文档撰写, 演讲

♡ 获奖情况

优秀贡献奖, 微软 Azure AI 文档智能核心贡献者	2021–2024
第二名, CVPR 2023 HIT Workshop 视觉提示调优挑战赛 (奖金 20 万元)	2023
第二名, ECCV 2022 SenseHuman Workshop 全景场景图生成挑战赛 (奖金 10 万元)	2022
唐仲英德育奖学金, 授予在学业和社会公益方面表现优异的大学生	2017–2021
省优秀毕业生奖, 授予优秀毕业生	2021
优秀学生奖学金金奖, 授予学业成绩优异的学生	2020
唐立新奖学金, 每年资助 1 万元人民币直至博士毕业, 授予学术、商业或政治潜力卓越的学生	2019
苏州育才奖学金, 每年授予 10 名综合能力突出的本科生	2019
优秀学生奖学金金奖, 授予学业成绩优异的学生	2018
新生研讨会论文一等奖, 授予学术潜力优秀的新生	2018

📄 其他

- 技术博客: <http://home.ustc.edu.cn/wangjiawei>
- GitHub: <https://github.com/JarvisUSTC>
- 语言: 英语 - 熟练 (CET-6: 550)