

# 王勇惠

联系方式: 178-6071-8319 邮箱: wyh1998@mail.ustc.edu.cn

中科大多媒体计算与通信教育部-微软联合实验室

导师: 李厚强教授 (杰青, 长江, IEEE Fellow) 周文昱教授 (优青)

求职方向: 计算机视觉算法相关研发岗位



## 教育经历

2021.09 - 至今	中国科学技术大学	信息与通信工程 (博士在读)	研究方向: 多模态理解
2017.09 - 2021.06	中国海洋大学	电子信息工程 (学士学位)	排名: 3/78

## 科研经历

- **研究方向:** 主要围绕阴影、文档图像恢复和多模态文档图像理解开展研究, 相关技术涉及阴影检测、阴影去除、视频阴影检测、文档图像光照恢复、多模态大模型的文档图像理解等。
- **研究成果:** 研究生期间, 一作/共一顶会顶刊论文共计发表 3 篇; 一作/共一顶会顶刊论文 2 篇在审。任 ACM MM, TMM 等会议期刊审稿人。
- **已发表论文:**
  - [1] **Yonghui Wang**, Wengang Zhou, Zhenbo Lu, and Houqiang Li, "UDoc-GAN: Unpaired Document Illumination Correction with Background Light Prior," Accepted to **ACM MM**, 2022.
  - [2] **Yonghui Wang**, Wengang Zhou, Hao Feng, Li Li, and Houqiang Li, "Progressive Recurrent Network for shadow removal," Accepted to **CVIU**, 2023.
  - [3] **Yonghui Wang**, Wengang Zhou, Yunyao Mao, and Houqiang Li, "Detect Any Shadow: Segment Anything for Video Shadow Detection," Accepted to **TCSVT**, 2023.
  - [4] **Yonghui Wang**, Shaokai Liu, Li Li, Wengang Zhou, and Houqiang Li, "SwinShadow: Shifted Window for Ambiguous Adjacent Shadow Detection," Accepted to **TOMM**, 2024.
- **在投论文:**
  - [5] **Yonghui Wang**, Wengang Zhou, Hao Feng, Keyi Zhou, and Houqiang Li, "Towards Improving Document Understanding: An Exploration on Text-Grounding via MLLMs," Submitted to **ACL**, 2024.
  - [6] **Yonghui Wang**, Wengang Zhou, Hao Feng, and Houqiang Li, "AdaptVision: Dynamic Input Scaling in MLLMs for Versatile Scene Understanding," Submitted to **TCSVT**, 2024.
  - [7] **Yonghui Wang**, Shi-Yong Chen, Zhenxing Zhou, Siyi Li, Haoran Li, Wengang Zhou, and Houqiang Li, "ROOT: VLM based System for Indoor Scene Understanding and Beyond," Submitted to **CVPR**, 2025
  - [8] Bozhi Luan, Hao Feng, Hong Chen, **Yonghui Wang**, Wengang Zhou, Hao Feng, and Houqiang Li, "TextCoT: ZoomIn for Enhanced Multimodal Text-Rich Image Understanding," Submitted to **TIP**, 2024.
  - [9] Keyi Zhou, Li Li, Wengang Zhou, **Yonghui Wang**, Hao Feng, and Houqiang Li, "LaneTCA: Enhancing Video Lane Detection with Temporal Context Aggregation," Submitted to **TCSVT**, 2024.

## 项目经历

- 文档图像智能理解系统 (负责人, 2023.03-2023.09) 企业合作: 华为

**项目简介：**提升多模态大模型在文档场景中问答的可解释性问题，输出答案的同时提供其依据。

**项目内容：**负责项目申请调研和全部代码撰写。1. 收集文档相关的图片数据，利用 OCR 和 GPT-4 提取并设计带有文本位置信息的对话数据，微调多模态模型。2. 实验发现输出文本的位置信息能够提升多模态大模型的问答能力。3. 该方法及其实验结果现已整理成论文，投至顶刊 TIP。

● **通用缺陷识别和诊断系统（负责人，2023.09-2024.3）** 企业合作：华为

**项目简介：**设计缺陷识别系统，在未见过的产品上达到通用的缺陷识别能力，同时结合多模态大语言模型，实现多模态缺陷诊断能力。

**项目内容：**负责项目申请调研和科学研究。1. 利用少样本学习实现通用的缺陷检测识别能力，达到很强的泛化效果。2. 根据用户指令，垂直领域多模态大模型准确识别用户指定类型、指定关键位置和大小图像缺陷、输出相关的诊断文字并给出后果及处理方式。

● **面试官投资意向判别系统（参与者，2024.01-至今）** 企业合作：招联金融

**项目简介：**利用视频多模态大模型辅助面试官判别面试人的投资意向。

**项目内容：**负责项目科学研究。1. 设计视频多模态大模型根据面试人的面部表情实时输出其投资意向。2. 面试官根据模型的输出来辅助判断投资人的投资倾向，并给出相应回复。

## 实习经历

● **智能摆放—基于 VLM 的室内场景物品关系预测（2023.05-至今）** 腾讯 IEG-研发效能部

**项目简介：**针对 3D 世界的感知、理解和交互问题，高质量、开源的 3D 数据较少，因此需自建 3D 场景数据处理流程。项目通过利用海量 2D 数据和基础模型，预测室内场景物品的垂直位置关系，进而辅助 3D 数据的构造，弥补 3D 数据不足，提升模型在 3D 环境中的表现。

**主要工作：**1. 调研并下载了合适的 2D 室内数据集，进行了图片去重、信息获取等处理。2. 对目标图片进行了精细化的检测、分割、深度、场景和物体的描述等处理，获取物体场景信息。3. 利用 VLM 预测物品间的垂直位置关系，并生成物体关系的 json 文件，发标人工审核形成指令微调数据对 InternVL-26B、混元、QwenVL2-7B 等进行微调供 3D 场景生成 pipeline 使用。

## 个人技能

- 英语技能：CET-6，有较强的科技论文阅读和撰写能力，以及基本的交流能力
- 编程技能：熟悉 Python，能使用 MATLAB、C\C++
- 框架工具：熟悉 PyTorch、Docker
- 沟通交流：目前担任一年课题组多模态图文理解小组组长，在老师的指导下协调和交流科研工作

## 荣誉奖项

- 研究生国家奖学金
- 研究生龙湖奖学金
- 本科学习优秀一等奖学金
- 山东省数学竞赛一等奖

## 其它

- 个人主页：<https://github.com/harrytea>
- 博客主页：<https://wangyh.co/>