

# 付临

☎ 19857109661 · ✉ 748544794@qq.com · 📍 杭州

当前状态: 在校生

## 教育背景

浙江大学 硕士, 电子信息工程, 集成电路学院, 研究方向: 多模态大模型 2024.09 – 2027.06  
浙江大学 本科, 信息与计算科学, 数学科学学院, GPA: 3.52 / 4.00 (保研) 2020.09 – 2024.07

## 实习经历

阿里巴巴 达摩院, 研究型实习 2026.02 – 至今

- **GUI Agent World Model:** 面向移动端智能体构建基于视觉生成的界面状态预测模型, 将移动端交互过程形式化为“current UI + action → next UI”学习任务, 整合多源数据集, 对 I2I 模型与 Bagel 模型进行微调训练, 以支持后续开展 action conditioned 的界面生成与下一状态预测实验。目前项目正在进行中, 预期投递 NeurIPS 2026。

华为技术有限公司 终端 BG, 算法工程师 2025.04 – 2025.08

- **AIGC 辅助配图工具 Demo:** 基于大模型/生成式能力实现图片扩图、重绘、擦除、自动抠图、色彩增强、超分等编辑能力, 完成从前端交互(上传、选区、提示词/参数配置)到后端推理调用与结果回传的闭环; 对不同模型的输入输出格式、mask 表达、分辨率与色彩空间等进行适配封装, 沉淀可复用的调用接口与 Demo 流程。
- **自动配图流程优化(封面裁剪 badcase 治理):** 围绕封面生成中“关键帧不佳/主体被裁剪”等核心问题, 整理 badcase 类型与触发场景并制定策略优化路径; 将大模型引入为决策模块, 自动判断“是否需要抠图”“是否适合裁剪/可裁剪幅度”等关键决策点, 并与现有规则/检测模块结合完成封面裁剪与主体保留策略升级, 提升封面主体完整性, 减少线上 badcase。

## 研究经历

**VideoKR: 知识与推理密集型视频理解数据与后训练** 2025.07 – 2026.02

- **研究动机:** 针对现有视频理解数据缺少结构化领域知识与高质量推理监督、难以支撑科学场景后训练的问题, 提出 VideoKR, 构建知识驱动的视频理解数据体系。
- **数据与实验:** 构建覆盖 82 个学科、6K 个知识点的课程级知识库, 知识点驱动检索收集 126K 个 YouTube(CC) 视频; 扩展生成 VIDR / KNOWVID / KNOWVIDR 三类样本, 得到 260K SFT 数据与 170K RL 数据, 并基于 Qwen2.5-VL-7B、Qwen3-VL-8B 开展 SFT+GRPO 训练。
- **评估与分析:** 在 Video-MME、MVBench、VideoMMU、SciVideoBench 等基准上, SFT+RL 相比仅 SFT / 仅 RL 带来一致增益, 模型能够更好融合视觉证据与领域知识完成推理。
- **成果:** Lin Fu\*, Zheyuan Yang\*, Yang Wang, Tingyu Song, Arman Cohan, Yilun Zhao. *VideoKR: Towards Knowledge- and Reasoning-Intensive Video Understanding*. Submitted to ICML 2026.

**TEXOCR: 面向可编译 Page-to-LaTeX 重建的文档 OCR Benchmark 与模型训练** 2025.09 – 2026.01

- **研究动机:** 针对现有文档 OCR 主要聚焦 PDF 到纯文本或 Markdown 转换、缺少可编译 LaTeX 页面级重建评测与训练方法的问题, 提出 TEXOCR。
- **数据与实验:** 构建包含 3,000 篇专家标注文档的 TEXOCR-Bench, 以及基于 arXiv 2022–2025 年论文整理的 TEXOCR-Train (52K 篇论文、384K 组 page image-LaTeX 对齐样本); 基于 Qwen3-VL-2B 采用两阶段 SFT+RLVR, 并利用文本、公式、表格、引用与可编译性等 verifiable rewards 进行优化。
- **评估与分析:** 系统评测 23 个 OCR/MLLM 模型, 发现现有方法在公式转写、表格恢复、引用一致性和编译成功率上仍存在明显不足; TEXOCR 在开源模型中取得领先表现, 且 RLVR 相比仅 SFT 带来稳定增益。
- **成果:** Chengye Wang\*, Lin Fu\*, Zexi Kuang, Yilun Zhao. *TEXOCR: Benchmarking and Advancing Document OCR Models for Compilable Page-to-LaTeX Reconstruction*. ACL 26.

**Sci-VBench: 面向科学推理的视频生成评测基准** 2025.12 – 2026.04

- **研究动机:** 针对现有视频生成评测偏重视觉质量与文本对齐、缺少对科学知识约束、多步因果推理与时序一致性衡量的问题, 提出 Sci-VBench。
- **数据与实验:** 构建覆盖自然科学、医疗健康、人文社科、工程四大领域、78 个学科的 1500 条专家标注样本, 并为每个样本配套 high-level reference guide 与 1–5 分 anchored rubric, 系统评测 11 个前沿闭源/开源文本到视频模型。
- **评估与分析:** 提出结合自动感知指标与 rubric-conditioned MLLM-as-Judge 的评测框架, 验证引入 evaluation specification 后非专家与 MLLM 与专家评分的一致性提升; 结果表明当前模型在科学因果正确性与时空一致性上仍存在明显不足。
- **成果:** Diandian Zhang\*, Tingyu Song\*, Lin Fu\*, Zheyuan Yang, Yilun Zhao. *Sci-VBench: Benchmarking Knowledge-Grounded Scientific Reasoning in Video Generation*. Submitted to COLM 2026.

## 专业技能

编程与框架: Python, PyTorch, LLaMA-Factory, Verl

训练与微调: 熟悉多模态大模型的 SFT、LoRA、QLoRA 等微调方法

后训练与对齐: 熟悉 DPO、RLHF、RLVR 等常见后训练与对齐范式

工程与数据: 熟悉机器学习、深度学习基础与常用训练流程, 具有大规模数据集合成经验