

赵志鹏 Zhipeng Zhao

多模态理解 · 跨模态对齐 · 多模态推理与评测 · 可信 AI

邮箱 zhaozhipeng@stu.ouc.edu.cn · 电话 17806246256 · GitHub · 知乎



求职意向：多模态算法工程师 | 2027 届博士，预计 2027.06 毕业

教育背景

博士	中国海洋大学 · 人工智能（信息科学与工程学部）	2023.08 – 2027.06（预计）
硕士	中国海洋大学 · 计算机应用技术	2020.08 – 2023.06
本科	青岛大学 · 测控技术与仪器	2015.09 – 2019.06

荣誉与学术服务：中国海洋大学优秀研究生；参与国家重点研发计划等 5 项课题；担任 AAAI / NeurIPS / ACL / EMNLP / IEEE TMM 审稿人；山东省“互联网+”大学生创新创业大赛铜奖（2021）。

核心优势

- 多模态架构与跨模态对齐：以因果中介分析对视觉编码器、多模态连接器（projector / Q-Former）与语言解码器做组件级激活干预，覆盖 6 个模型、3 类架构族，逐层产出架构级“机制签名”。
- 多模态理解与幻觉抑制：提出 token 级视觉依赖度量的免训练对比解码框架，在 MMHalBench 上超越全部同类解码基线，并揭示视觉依赖随生成衰减、幻觉时间聚集的规律。
- 数据与工程落地：主导 1.02 亿图像-文本对数据集构建与 8 人质检；完成 LLaVA-1.5 的 LoRA/QLoRA 微调并在工业质检场景落地；独立完成开源模型向华为昇腾 910 的迁移适配并开源。

项目经历

多模态模型机制级评测：跨组件因果激活替换诊断 2024.01 – 2025.01 · 第一作者 EMNLP 2026 · GitHub

- 以因果中介分析对视觉编码器、多模态连接器（projector / Q-Former）与语言解码器逐层做激活替换干预，将性别线索效应分解为特定层激活的受控间接效应（AIE）与其余路径的直接效应，产出架构级“机制签名”。
- 覆盖 6 个模型、3 类架构（LLaVA-1.5 / LLaVA-NeXT / InstructBLIP 7B-13B / MiniCPM-V / InternVL）、32 语言层与 24 视觉层：语言层敏感度最高，LLaVA-NeXT 视觉层敏感度升高，InstructBLIP 的 Q-Former 则将敏感度集中到语言层。

多模态幻觉诊断与解码优化 2025.01 – 2026.01 · 第一作者 EMNLP 2026 · GitHub

- 提出免训练对比解码框架：对比“有视觉 / 视觉消融”双路生成的 token 级 log-prob 差，得到 token 级视觉定位分数，用于定位视觉未落地的 token 并在解码时施加自适应惩罚。
- 在 LLaVA-1.5 / InstructBLIP / Shikra 与 HallusionBench / MMHalBench / CHAIR / POPE / MMMU 五个基准上验证，MMHalBench 显著超越全部基线；量化约 2x 推理开销并给出双路共享 KV-cache 的优化路径。

跨模态注意力干预与视觉-语言对齐 2023.08 – 2024.08 · 第一作者 EMNLP 2026 · GitHub

- 为 LVLM 构建结构因果模型并刻画视觉→语言、语言→视觉等 5 种跨模态注意力模式，提出基于后门调整的注意力掩码干预，强化视觉主导、抑制语言先验，免微调即可嵌入现有推理流程。
- FACET / MS COCO / FairFace 跨基准验证：单属性偏见降低 14.7%（2 / 3 属性 11.9% / 10.3%），LLaVA 在 MS COCO 上由 0.6708 升至 0.9978；POPE / MMMU 推理能力基本无损，1,000 样本端到端推理延迟开销 < 1%。

工业场景多模态落地（发明专利 · 第一发明人） 2023.08 – 2024.12

- 基于合作工厂监控数据主导构建 1.02 亿图像-文本对数据集，覆盖安全装备检测、6 类拆解动作识别与危险行为预警，带 8 人团队完成数据质检。
- 基于 LLaVA-1.5 设计 LoRA/QLoRA(4-bit) + FlashAttention-2 微调方案，单卡 RTX 4090 完成训练；落地后作业效率 +52%、事故率 -34%。

学术论文

[1] ViD: Vision-Dominant Gender Bias Mitigation for Large Vision-Language Models. EMNLP 2026（一作，CCF-B）. [arXiv:2609.16647](#)

[2] Mechanism-Level Evaluation for Vision-Language Models: Controlled Activation-Replacement Diagnosis of Gender Bias. EMNLP 2026（一作，CCF-B）. [arXiv:2609.16651](#)

[3] What Do Hallucinations Reveal About Multimodal Reasoning? Diagnosing Visual Grounding Failures via Contrastive Decoding Probes. EMNLP 2026（一作，CCF-B）. [arXiv:2609.16646](#)

另有 EI 收录会议论文 5 篇（一作 3 篇）、在投一作论文 6 篇（IEEE TMM × 4、AAAI × 2）。

技术栈

多模态模型 LLaVA-1.5 / NeXT · InstructBLIP · Shikra · MiniCPM-V · InternVL · Qwen-VL · DeepSeek-VL

跨模态技术 跨模态注意力分析与干预 · projector / Q-Former 连接器层分析 · token 级视觉定位 · 对比解码

训练 / 推理 LoRA/QLoRA · SFT · DPO · FlashAttention-2 · Token Merging · DeepSpeed · PyTorch；vLLM · KV-cache 优化 · CUDA · 昇腾 910（CANN）