姜宝伟 多模态内容理解 / 视觉算法

WeChat Phone 15040215187 | **j_baowei@163.com** | 北京 | jang2333.github.io

个人简介

4 年视觉与多模态算法经验，现任快手可灵 AI 事业部内容理解算法工程师。负责把业务判断转成可执行的标签、数据和评价口径，再完成模型训练、服务接入与线上验证；目前带领 3 名正式员工、8 名实习生。已接收 4 篇视觉与多媒体顶会论文，另有 4 项工作在审，其中 2 项为通讯作者。

核心技能

多模态内容理解、UMM、4K 视频抠像、图文检索、图像篡改定位、图像与视频质量评估、内容安全、三维重建。
Python / C++，PyTorch，视觉语言模型（VLM），SigLIP / CLIP，Sapiens，LoRA，ONNX，Kafka。

工作经历

快手 · 可灵 AI 事业部 | 多模态内容理解算法工程师

2025.09 - 至今

- **AIGC 内容理解：**

  1）主要工作：把提示词、参考内容和生成内容联合输入一个模型，统一处理主题、元素、风格、意图、风险、质量和版权 7 类标签任务；牵头定义标签体系，建设 25 万训练数据，训练引入标签相似度约束，改善标签幻觉。

  2）上线效果：平均准确率约 90%；复用一次视觉计算并行完成多项任务，推理提速 2.4 倍、单条约 3 秒，3 张 A10 支持日均约 2 万条实时处理。App/Web 端 Tag 页供给提升 2～6 倍，圣诞、春节和海外棒球等节日/热点标签可在 2 小时内完成全站圈选；项目目前支持全量用户创作意图监控和作品发布文案生成。

- **检索与推荐：**从零建设多模态 embed 模型，推荐打散已上线，跨模态检索正在接入； CTR +1.06%，人均时长 +1.98%、人均播放量 +2.04%。支持向量检索，且理解能力较基线 +1.1%。应用成果投稿中，本人为通讯。
- **社区与质量：**负责优化作品准入率，组合 9 个模型、2 个服务和 2 条规则，优质内容准入量翻倍，不依赖 API 调用；另带团队完成图像美学、视频故事性、视频色彩质感 3 套评分模型，并用于推荐/运营侧内容精选。
- **内容安全：**按名人、版权角色和整图风险拆成三类自研筛查。覆盖约 3.8k 人/ 2.8 w 向量；版权角色覆盖 1.8k ，整图模型覆盖 5 大类风险，第一期减少约 30% 第三方调用。
- **级联意图理解：**用词表规则和语义 embed 检索召回候选集，再用自研 vlm-classify 模型结合提示词和候选集进行二分类，仅将边际 case 交给 api 复判；覆盖 2.8k 版权和 3.7k 明星。200w 请求评估召回 84%，误伤率 0.26%。
- **绿幕抠图：**面向车载数字宠物和真实人物卡通化，负责 4K 视频抠像的方案设计、代码实现、论文写作和数据构造，模型以 API 支持两家头部新能源车企；相关工作被 CVPR 2026 接收，本人为共同一作。

快手 · 特效技术中心 | 特效理解算法工程师

2024.09 - 2025.09

- **特效标签与审核：**在复杂标签体系、数据有限的条件下，将特效名称、图标、视频和元信息联合输入模型；几千条数据微调后，准确率和召回率由约 70% 提升到约 95%，线上准确率 98%，日审核量翻倍。
- **特效优质分级：**牵头把优质特效评估标准拆成特效元素、构图布局、颜色、材质和图标五个维度；修正类别失衡和标注冲突后，微调 Qwen2.5-VL-7B，验证集准确率 98.27%，完成在线接入。
- **竞品视频拍同款（PLC）：**只看语义会把真实眼镜误认成眼镜特效，借鉴图像篡改定位领域思想，从短视频中提取纯特效特征并在线检索，引导拍同款；PLC 点击率 +0.72%、标签页曝光用户 +2.17%、特效作品 +0.35%。
- **特效去同质化**：建设特效图标特征库，减少推荐中的重复特效，特效作者数 +0.30%。

启元实验室 | 2D / 3D 视觉算法

2022.06 - 2024.09

- **人脸隐私保护**：常规高斯模糊会给人脸检测器引入伪特征。提出神经网络打码方法，在隐藏身份的同时保持后续检测效果；DartBlur 被 CVPR 2023 接收，本人为共同一作。
- **GigaVision**：连续 3 个赛季负责比赛数据、评分代码、复赛代码复现和决赛评审；参与十亿像素、红外可见光、无人机和多机位滑冰数据采集，以算法预标注配合人工勘误降低标注成本。

- **深度与三维重建**：用 AirSim + UE 构造无人机 RGBD 数据，3D Gaussian Splatting 完成大场景任意视角 RGBD 渲染。室内场景快速 instant-NGP 重建与渲染，发表 ICCV 2023，二作。

其他经历

希迪智驾（算法实习，2020.07 - 2020.10）：公交车环视场景旋转目标检测。清华大学自动化系（实习，2016.12 - 2017.03）：直播人像抠图、无穿戴数字人驱动，手部求解速度提升近 3 倍。

论文与专利

已接收 5 篇：αMatte4K & μMatting（4K 视频扣绿，CVPR 2026，共一）、JointTok（统一理解与重建的 tokenizer，ACM MM 2026，二作）、DartBlur（CVPR 2023，共一）、PARF（ICCV 2023，二作）、图优化三维重建（IEEE VR 2023，共一）。另有多模态检索、统一理解与生成、图像篡改定位、流式视频理解 4 项工作在审，其中 2 项通讯。

专利：自动剪枝量化方法及系统（CN202310308115.6）、三维对象重建方法（CN202110106935.8）、视频扣绿处理（CN122073039A）、三维场景重建（CN121392122A）和三维场景生成（CN121482763A）。

教育经历

首都师范大学 | 计算数学 | 硕士（双一流） 2019.09 - 2022.06

东北大学 | 信息与计算科学 | 本科（985） 2014.10 - 2018.06