

曹志杰

年龄: 31岁

籍贯: 浙江

联系电话: 18217765951

性别: 男

工作年限: 5年经验

联系邮箱: cao_zhijie@aliyun.com



教育背景

2018.09 - 2021.03

上海交通大学

控制工程 (硕士)

研究方向: 在自动化系的图像处理和模式识别研究所, 从事计算机视觉, 监控场景下的人群计数相关研究

2014.09 - 2018.06

同济大学

自动化 (本科)

研究方向: 嵌入式系统, 硬件开发, 控制理论, 机器学习, 计算机视觉

工作经验

2023.12 - 至今

字节跳动-tiktok-广告创意

广告算法工程师

信息流广告视频Refresh

- 视频Refresh类策略研发和维护: 上线了基于多模态模型的内容理解+LLM的文本生成+爆款视频的字体检索的, **Replyto策略**和**视频加卖点策略**。其中通过xid对比实验, Replyto策略的advv+12.83%, 视频加卖点的advv+13.56%。同时维护了加pov文案和加cta文案等历史遗留的Refresh类策略。
- 探索基于**self-improve Agent**的视频策略的优选: 搭建了后验收集+经验总结+经验检索的multi agent系统, 其中rules-generator-agent实现基于后验数据的经验总结, 用rules-manage-agent实现经验库的更新和管理, 以及decider-agent实现经验检索和策略优选。该方案应用在多策略优选上, 实现了策略选择准确率+13.33%, pvr+4.92%的效果。

搜索广告图片优化

- 电商广告封面图优选: 针对商品首图过于简单, 无法充分表达商品信息的问题, 通过商品卖点抽取+图片加字的方式产生新的图片, 结合在线EE优选实现了电商商城广告advv+1.57%, cvr+0.63%
- 基于query的广告图片生成: 对已有图片进行Refresh的方案: 从广告落地页和视频抽帧中, 用MLLM选出跟query相关图, 再用LLM结合query和广告信息生成图上文案, 最后通过规避人脸的加字方式, 渲染到图片上。结合在线关键词召回+相关性优选, 实现了advv+2.6%的效果

搜索广告标题优化

针对用户自填标题表达粗糙, 无法match用户搜索需求的问题

- 基于SFT提升相关性: 离线收集每个广告历史用户搜索的query, 在结合query和广告原始信息对标题进行改写。通过GPT4产生改写标题, 再通过qwen2.5+SFT的方式蒸馏GPT4, 在垂搜商城内的商品广告上实现了大盘advv+0.8%, ctr+0.13%, 综搜开环视频广告上实现大盘advv+3.5%, ctr+0.7%。
- 基于RLHF提升点击率: 在SFT的基础上, 训练点击率预估模型为reward model, 通过RLHF的方式, 实现面向点击率的标题改写, 在综搜开环上实现了大盘advv+0.6%, ctr+1.3的收益。
- 基于RRHF的蒸馏: 在RLHF基础上, 通过RRHF实现对多个RLHF模型的蒸馏, 离线验证集上reward胜出率从0.62提升到0.93, 模型上线后进一步获得advv+0.4%个点的收益

搜索广告创意链路建设

从0~1搭建了tiktok搜索广告创意的素材生成-管理和在线优选投放的整个创意链路

- 离线素材改写链路建设: 实现了daily的离线标题改写、封面图优化、视频转图文等离线素材生成链路
- 在线优选功能建设: 对在线投放链路进行改造, 在不增加链路耗时的情况下, 实现了综搜、电商场景的标题、视频封面图、商品封面图等的相关性优选以及EE优选功能

大模型影视剧剪辑

1. 基于大模型，成功搭建了一套自动化的影视剧讲解类剪辑的算法框架，其中涉及到基于 LLM 的影视剧剧本改写+Scenedet 场景分割+BLIP2 文本到视频检索+视频后处理。
2. 针对影视剧垂类检索不准的问题，通过收集垂类数据，拆解成该方案可以实现30分钟左右产出一集电视剧的自动剪辑。相比于人工效率大大提升。且投放到逛逛视频 tab 之后，人均停留时长达到20s+,超过了逛逛视频大盘的14s。

直播主播行为理解

定位出直播中主播的精彩动作（细节展示，试妆，背包），应用于直播推荐和直播剪辑

- 1.定义了覆盖8大电商类目，总计30+类别的主播行为标签体系。
2. 搭建了VideoSwin 的流式动作识别算法，mAP为0.43, 90%准确率下 recall 达到0.26。
3. 对流式分析方法做出改进，提出 label mask的 videoSwin+actionformer 的方案，mAP 提升至0.62，recall 达到0.42，最终又通过端到端的训练，将 mAP 提升至0.69，recall 提升至0.55。剪辑后的短视频相比于原始看点，在淘宝首页投放的人均消费 PV+6.7%，人均停留时长+3.2%，人均成交提+4.4%。
4. 搭建了基于多模态时序分析算法 UMT 的直播间的高光时刻识别算法，并且将高光时刻作为剪辑视频的前置片段，该项改造，进一步实现了单 pv 时长增长+1.7%，成交转化率(ipv_uv)+3.8%，有效播放率+3.7%。

直播间信息卡项目

为了提高直播间的成交转化率，提出了制作一张包含商品简要信息的信息卡，以提示主播正在讲解的商品。

- 1.第一阶段的方案是商品检测+OCR +人体检测+策略抠图的方案，该方案实现了基本的任意商品图抠出主体的能力，但是不同品类商品图差异大，抠图结果的 top 1可接受度为0.668。
- 2.后续第二阶段以商品图预分类+Visual Grounding 的方案，抛弃了复杂的策略方案，将 top 1的可接受度提升至0.773，同时算法显存需求也从4GB 降低至2GB。
3. 该项目主播的整体复用率很高，达到了40%。引导成交人均 GMV+9.18%,说明卡片对成交金额上有贡献。

直播内容理解端云一体化改造

- 1.搭建了端上小模型+云端大模型组合推理的端云一体化直播内容理解算法框架
2. 实现了端上稳定的算法覆盖率保障：在日均10w+的直播情况下我们实现了在 Android，Ios，PC 三端的人脸，人声，商品检测，商品检索，ASR 等算法的75%+的覆盖率
3. 降低了看点生产所需算力：创新地提出了端云结合的多模态看点生产算法，object embedding和 ASR 由端上生产，云端只做大模型推理，可以降低7/8的云端算力，节省算力成本523w。
4. 降低了直播标签生产所需要的算力：直播 faceid 生产，直播空镜识别，直播主题标签生产等业务，节省了云端30%左右的算力，一年大约节省1000w左右的成本。

- 1.负责人群计数模型的迭代
- 2.设计新的 loss function在 crowd counting 中的应用以及结合 detection和 keypoint 去做 crowd counting，投稿 ECCV

荣誉证书

论文：

1. 《GIGAPIXEL-LEVEL IMAGE CROWD COUNTING USING CSRNET》ICME2019 workshop
2. 《Synthetic guided domain adaptive and edge aware network for crowd counting》SCI期刊

专利：

1. 《基于深度学习的光场相机阵列行人计数技术 》
2. 《基于深度学习的光场相机多行人跟踪技术》
3. 《基于人头分块的损失函数人群计数优化技术》
4. 《基于生成图像的人群计数模型建立方法、计数方法及系统》