个人简历

个人信息

姓名：	陆鑫益	性别：	男
出生日期：	1994 年 11 月 1 日	工作年限：	3 年
手机：	18606224388	邮箱：	645096158@qq.com

专业技能

语言及框架：熟悉 Python、SQL、Pytorch、Spark，掌握 C++、Go

领域知识：NLP、机器视觉、数据分析

工作经历

2021.11-至今　腾讯微信事业群 WXG 搜索应用部　算法工程师　T8

1. 负责垂类 box 意图识别及效果优化

微信搜索垂搜 box 对标百度阿拉丁，目前共 140+类 box，占大盘流量 20%+。

a) 搭建和维护大规模特征数据流　Spark 开发

利用海量微信基础数据构建数十亿级别的数据宽表，包括 query 的曝光点击数据、query 切换数据、doc 基础信息数据等，每天稳定例行更新，支撑组内包括数据分析、意图识别、相关性计算、排序等团队对基础数据的需求。

b) 构建深度意图模型触发框架　Bert 预训练、下游任务 finetune

主导构建深度意图模型触发框架，包括微信垂搜场景基座模型预训练、意图样本自动打标、模型特征生成、模型结构优化等。针对垂搜特定场景，在微信搜索语料的基础上，加入垂搜图谱在预训练阶段注入更多的领域知识；基于大规模特征数据流，构建不同基础算子实现意图 finetune 阶段的样本自动打标；针对 query 短文本语义信息缺乏的问题，利用曝光点击 doc 实现语义增强；针对类目多且复杂的问题，基于预训练模型采用 MMoE 结构实现行业下多类目统一触发，大幅降低上线成本。

c) 探索使用语义大模型来解决意图问题　LLM 大模型实践

基于现有的开源基座大模型（LLaMA、GLM 等），使用微信搜索场景语料构造数据来对基座模型进行 SFT 指令学习，将原来独立的 Bert 意图分类和实体抽取任务转化为统一的开放场景生成任务，目前已经在大部分垂搜 box 上已经取得超越 BERT 的效果。

2. 探索表情图片生成

表情搜索是微信生态的特色内容，希望通过表情生成来扩充生态内的资源供给，同时探索文生图、图生图等能力的不同应用场景。

a) 基于 stable diffusion，采用 PEFT 方式如 Lora 来快速训练不同主题、风格、人物形象的 lora 模型，通过 prompt 调用和融合不同的 lora 模型来扩充当前的表情数据库。

b) 结合 ControlNet、OpenPose 等模型，探索不同的应用场景，包括人物表情生成、动作姿态生成、一键换装以及特定风格迁移等，为新的内置功能做技术积累。

2020.7-2021.11　腾讯 PCG 电商搜索算法团队　校招入职

1. 负责商品搜索特征体系、精排模型的搭建和迭代　搜索排序

从零到一搭建搜索特征体系以及迭代精排模型，模型迭代路径包括 DeepFM、DIN、TransFormer、ESMM 等，线上 CTR 以及 CVR 均取得明显收益。

2. 参与 QU 模型的优化，包括搜索意图识别、Term 实体识别、类目预测等模块

QU 模型采用 Bert 结构，通过多任务方式集成搜索意图识别和 Term 实体识别两个关键任务，有效提升各个任务的精度，其中 Bert 模型产出的 Embedding 还可以用于商品的向量召回。

3. 负责搜索数据的分析下钻、相关指标看板的建设和维护

统计分析用户行为数据，比如高频 q-d 对、用户点击偏好等来丰富商品、人群的画像内容，辅助模型个性化效果的提升。

实习经历

无感知考勤项目　　视源股份 CVTE 中央研究院实习

1. 构建新型损失函数对人脸特征空间进行进一步的优化。在损失函数中直接添加类代理向量夹角惩罚项，可以更高效的优化类代理向量在空间中的分布，并提升网络对于人脸角度变化、光照变化和遮挡的鲁棒性。

2. 通过与球机摄像头的交互实现非配合场景下的无感知考勤。实际场景中对于人脸识别影响较为严重的是远距离小人脸的识别，通过集成人头检测算法并与球机摄像头交互，实现对人脸的自动定位与对焦。

3. 在最终决策阶段，集成人脸聚类算法，对部分不确定人脸做进一步的筛查，并形成最终考勤结果。整个系统目前已在部分教 室投入实际测试，反馈较好。

基于机器视觉的施工安全检测系统　　企业横向项目

为有效避免施工人员出现作业安全隐患，并降低人力成本，搭建包含安全帽佩戴检测、工作服穿着检测与安全区域越界检测三个模块的施工安全检测系统。

1. 安全帽佩戴检测基于 RetinaNet 网络结构。

2. 工作服穿着检测的难点在于工作服样式和颜色深浅存在较大差异以及工作场景的多样性带来的背景干扰。针对第一个问题，在数据有限的情况下，使用数据增强方式能够有效缓解；针对背景干扰，采用 Mask-RCNN 进行像素级的人体分割，使得后续的分类网络能够更好的捕捉到工人着装的纹理信息。

3. 安全区域识别主要基于 HSV 空间的颜色检测算法，结合点检测以及凸包检测等，最终确定安全区域范围。

教育经历

2017.9-2020.7

华南理工大学　控制工程专业　硕士

2013.9-2017.7

华南理工大学　自动化专业 / 金融专业 双学位学士

自我评价

3 年搜索相关技术经验，同时具备机器视觉的经历，对新技术抱有足够的好奇心。

技术栈较广，熟悉自然语言处理，了解图像相关知识，同时对于基本的大规模数据处理和分析具备一定的经验。

业余时间喜欢尝试量化模型以及读各类书籍。