

专业实践总结和教学案例

2021.9 广东轻工职业技术学院 罗佳

广州市勤思网络科技有限公司成立于 2013 年，一直专注于商业智能，大数据及人工智能的技术研发及行业应用。公司团队 100 多人，其中本科及以上学历占比 80%，拥有多个自有知识产权的大数据和人工智能产品，并已经成功于金融，政府等多个行业。作为以技术为本的科技型企业，勤思科技不断在人才培养，技术研发上投入，立志成为中国 AI 技术的引领者。通过技术与行业的深度结合，目前勤思科技拥有保险，海事，智能制造等行业的全方位智能化解决方案；通过人工智能技术的赋能，为客户实现高效、低成本的无人化办公场景及科学化的决策管理。

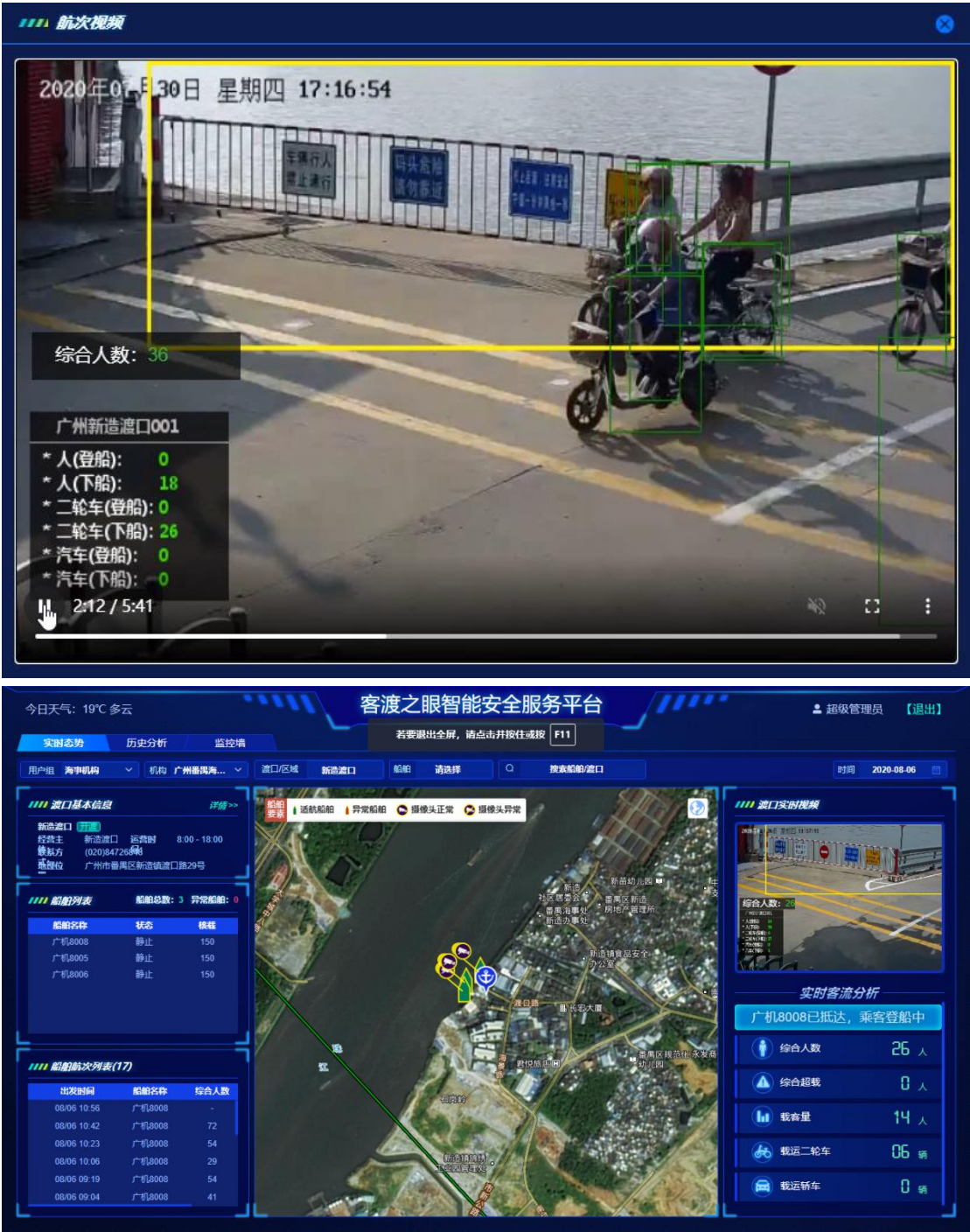
勤思科技作为广东轻工职业技术学院校外实践基地，致力于人工智能和大数据领域的人才培养和科研合作，每年举办的“勤思杯”创新设计大赛成为广东轻工职业技术学院的品牌赛项，在人工智能研发领域，与我校教师团队合作带领学生完成了智慧海事、智慧校园、智能制造等多个项目的技术攻关。在工业互联网领域，勤思科技将凭借其机器视觉领域的技术研发优势，与创新团队一起协同促进企业的数字化转型升级。2021 年，勤思科技成为我校省级校外实践基地的重点培育对象。

本人于 2021 年 1 月开始，在勤思大数据与人工智能事业部进行企业实践活动。该事业部主要进行研究性项目的研发工作，为公司在 AI 和大数据领域做技术储备，因此其项目方向有很多，如计算机视觉技术在航运渡口码头等领域的应用、计算机视觉技术在智能制造中的应用、智慧课室、AI 在保险领域中的应用等等。本人在勤思实践过程中，结合自身的业务领域，在以下几个方面进行了探索和研究：

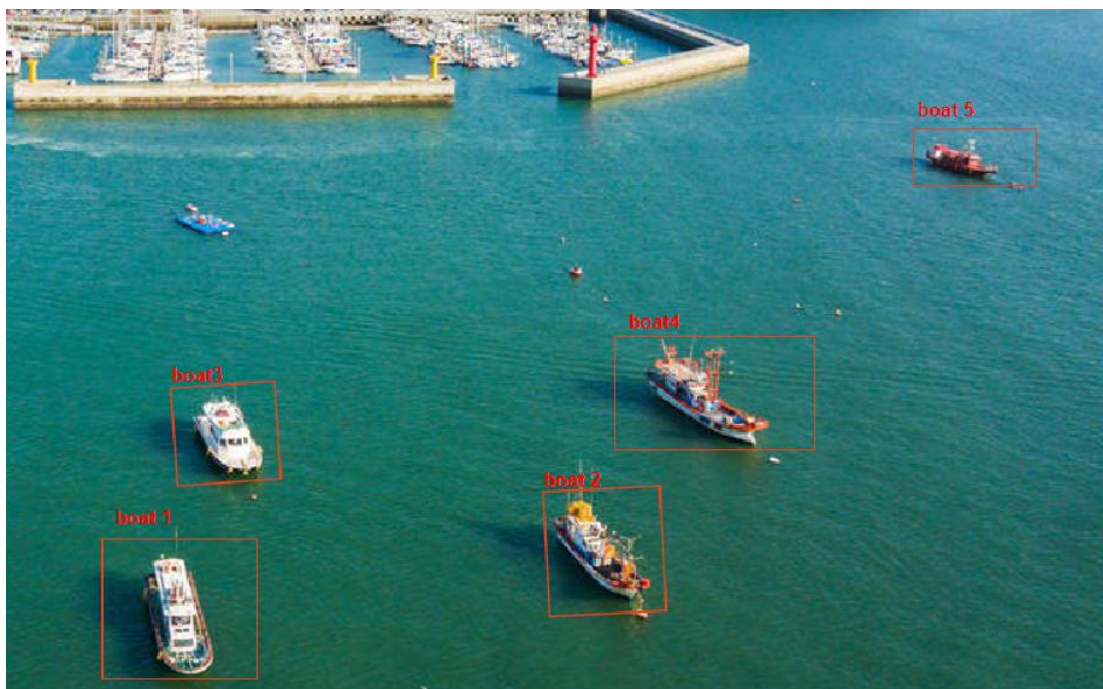
一、参与了其船舶航运领域的计算机视觉方面的研发工作

目前，在涉及船舶航运领域中的 AI 技术应用还比较缺乏，而勤思在利用信息技术提升船舶航运领域的管理运行水平领域已经深耕多年，尤其是将 AI 和大数据技术应用到航运管理方面有很多核心技术，如

智慧海事---客渡之眼产品：



智慧海事---江河之眼：



分割



识别

浙普渔68979

智慧渔政：



同时，勤思在航运安全监控等方面持续发力，目标是研发出一款能在航船上部署的 AI 盒子，该盒子能够接入航船监控视频、航船重要设备、北斗等多路数据，对输入数据进行处理和识别，对航船的安全运行保驾护航。



本人在实践过程中全过程学习了公司在 AI 领域的研究开发流程，其开发平台架构对我院的人工智能、大数据等专业的实训教学条件的建设有很大的借鉴作

用。实践期间，本人也与公司工程师交流了平台架构的软硬件的配置问题，工程师对本人今年所主持建设的《工业大数据智能应用产教融合平台》也提出了很多宝贵的建议，并对该平台进行了试用，并给予了很高的评价，希望后期我们能够基于该平台合作开发更多的优秀产品。

勤思模型池							
Kinth Model Pool							
自研模型+经典模型							
算法层 Kinth Engine	基础设施		训练		边缘计算化		
	cfg可视化配置	动态图	分布式	异构	GPU/TPU	X86/ARM	tx2树莓派
算力层 Kinth Cloud	基础设施				资源运营		
	集群管理	高速网络	异构存储		管理权限	算力调度	GPU/TPU优化
数据层 Kinth Data	数据工厂				数据管理		
	采集	标注	检查清洗	增广	图片	视频	云存储 格式互转

二、 校企联合指导学生参加“中国软件杯”赛项

一年一度的中国软件杯于 2021 年春拉开序幕，大赛秉承“政府指导，企业出题，高校参与，专家评审，育才选才”方针，以“催生多重效应，引领产业创新”为宗旨，在深化产教融合、产融对接，创新人才培养、激励人才创新，推动科研创新、成果转化等方面搭建了重要平台和有效途径。大赛举办七年来，已成为全国软件行业规格最高、最具影响力的顶级赛事。

顺应信息技术发展的态势，本届赛事更多的加入了人工智能和大数据的赛题，如“在线数据可视化分析平台”、“有限边缘算力下的电路板故障检测”、“辅助桌面读屏软件”、“高并发条件下消息队列的设计与实现”等等。其中，我学院软件专业的同学组队报名参加了“有限边缘算力下的电路板故障检测”和“高并发条件下消息队列的设计与实现”。

两个赛题的核心技术难点一是计算机视觉技术的应用，二是高并发下的 web

应用程序开发，而这两个课题的技术点，也正是勤思科技工程师所涉及的领域。因此，本人在实践过程中，积极协调企业资源，利用周末等闲暇时间，邀请企业工程师共同参与到指导学生的行列中。



通过几个月时间的努力，我们指导的学生分别获得了中国软件杯国赛二等奖和三等奖的成绩，这是校企合作的一种新的尝试，为后续开展更多更灵活的校企合作提供了有益借鉴。

三、联合申报《工业大数据智能分析与应用创新团队》

工业大数据是工业互联网的核心。工业互联网和科学技术的高速发展，推动着大数据时代的来临，各行各业每天都在产生数量巨大的数据碎片，数据计量单位已从 Byte、KB、MB、GB、TB 发展到 PB、EB、ZB、YB 甚至 BB、NB、DB 来衡量。大数据时代数据的采集不再是技术问题，有各种各样的传感器收集多类型数据，也有功能丰富的信息系统产生多意义数据，而寻找到海量图像、文本、数值等数据的内在规律，不仅能帮助工业生产者更高效地管理事务，更能进一步促进工业场景的技术革新升级。因此，工业大数据是智能制造与工业互联网的核心，其本质是通过促进数据的自动流动解决控制和业务问题，减少决策过程带来的不确定性，并尽量克服人工决策的缺点。

工业大数据技术的应用，核心目标是全方位采集各个环节的数据，并将这些数据汇聚起来进行深度分析，利用数据分析结果指导各个环节的控制与管理决策，

并通过效果监测的反馈闭环，实现决策控制持续优化。如果将工业互联网的网络比做神经系统，那工业大数据的汇聚与分析就是工业互联网的大脑，是工业互联网的智能中枢。

工业大数据分析应用、基于机器视觉的工业智能和基于虚拟仿真技术的工业互联网数字孪生是目前工业互联网领域相对比较成熟的应用领域。三者之间相互独立又互相融合，工业大数据分析需要结合人工智能的算法支持；以机器视觉为代表的人工智能技术正是有了大数据的加持才有了今天的迅速发展；而工业互联网数字孪生技术更是虚拟仿真技术结合大数据和人工智能技术的经典应用。本团队分别吸收三个领域的技术人才，定位工业互联网技术应用领域，在工业大数据智能分析与应用、机器视觉和数字孪生三个方面融合创新，为工业互联网产业发展提供原创动力，为当地产业信息化智能改造提供技术支持，为工业互联网产业人才培养贡献力量。

勤思科技在智能制造领域，特别是基于计算机视觉的工业产品质检领域有长期的积累，因此公司积极支持我校的创新团队申报工作，并表示将在后期持续支持创新团队的建设。

8	田霖	中科院计算所南京研究院	计算机技术与通信应用	兼职	
9	邓健爽	广州市勤思网络科技有限公司	人工智能创新应用	兼职	

依托单位和合作单位承诺

已按填报说明对申请人的资格和申请书内容进行了审核。本单位保证对研究计划实施所需要的人力、物力和工作时间等条件给予保障，严格遵守广东省教育厅有关规定，督促负责人和主要成员以及本单位科研管理部门按照广东省教育厅的规定及时报送有关材料。

	依托单位	合作单位 1	合作单位 2
名称	广东轻工职业技术学院 (公章)	中科院计算所南京研究院 (公章)	广州市勤思网络科技有限公司 (公章)
承诺经费	340 (万元)	0 (万元)	0 (万元)
日期:	年 月 日	年 月 日	年 月 日

四、 联合申报《智能制造产业转型升级场景社会实验》项目

为贯彻落实《新一代人工智能发展规划》、《国家新一代人工智能创新发展试验区建设工作指引》等文件精神，按照科技部关于支持广州市建设国家新一代人工智能创新发展试验区（以下简称广州国家试验区）的工作要求，根据《广州市重点领域研发计划揭榜挂帅制技术攻关项目试点工作方案（试行）》，通过公开发榜形式，围绕全市人工智能重点应用领域，组织开展人工智能社会实验，探索智能社会治理新模式。

勤思科技基于其在智能制造领域的技术优势，结合我校在工业互联网技术方面人才培养的优势，强强联合，联合申报《智能制造产业转型升级场景社会实验》项目，研究项目以企业、院校、政府调研为基础，开展智能制造产业转型升级过程中对于企业、社会人才需求、政策方面的影响的分析，最终形成调研报告和相关标准。

项目阶段与进度表

项目阶段	起止时间	主要工作内容
第一阶段	2022.4.1 至 2022.9.30	1、整理需求，调研智能制造产业，确定研究的目标企业对象；文献调研智能制造技术现状，梳理确定调研的技术指标；调研职业院校，确定调研的目标职业院校及专业
第二阶段	2022.10.1 至 2023.3.31	1、深入目标企业，进行实地走访，收集数据 2、深入目标职业院校，进行问卷调查，收集数据 3、通过互联网爬取，获取数据
第三阶段	2023.4.1 至 2023.9.30	1、进行数据整理，数据处理和数据建模
第四阶段	2023.10.1 至 2024.3.31	1、进行模型调整与验证 2、撰写面向广州市智能制造企业的综合分析报告1份
第五阶段	2024.4.1 至 2024.9.30	1、持续进行数据的更新 2、撰写智能制造环境下的职业教育政策建议1份
第六阶段	2024.10.1 至 2025.3.31	1、撰写人工智能技术在智能制造领域应用的技术标准1份 2、准备项目验收

五、开展横向课题的研究

基于本人在大数据领域的相关经验，本人在企业实践期间，积极开展校企合作项目的研发，与广州市崑坤室内设计有限公司合作研发《跨境电商数据分析系统》，为境外商家提供国内供应商的相关信息资源。

广东轻工职业技术学院横向项目合同书

项目名称：跨境电商数据分析系统

甲方：广州市崑坤室内设计有限公司

乙方：广东轻工职业技术学院

签订时间：2021年8月

签订地点：广州

有效期限：一年

广东轻工职业技术学院科技处制

六、总结

通过半年多的实践，本人对人工智能和大数据领域的研发有了进一步的认识，对本人以后的教学和科研过程有及重大的影响。

(1) 计算难以瞬间完成

完成任何一个人工智能解决方案的计算，都需要一定的时间，方案的响应速度，对商业应用的成功与否起到十分关键的作用。不能总是盲目假设任何算法在所有数据集上都一定能在规定时间内完成，你需要测试下算法的性能是否在可接受的应用范围内。

以搜索引擎为例，用户对结果返回的时长是有忍耐的限度的。如果用户等待的时间超过 10 秒，50%的用户会流失，如果等待时间超过 1 分钟，90%以上的用户会流失。在开发智能应用系统时，不能为了达到更好的算法精度而忽略系统运算和等待的时间，否则会导致整个产品的失败。

（2）数据规模非常重要

当我们考虑智能应用时，数据规模是很重要的因素。数据规模的影响可以分为两点来考察：第一点是规模会影响应用系统的响应速度，上一节我们刚提过；第二点是在很大的数据集上的挖掘出有价值结果的能力会受到考验。例如为 100 个用户开发的电影或音乐推荐系统可能效果很好，但是同样的算法移植到有着 100000 个用户的环境里，效果可能就不尽如人意了。

其次，使用更多的数据来训练的简单算法，比受制于维度诅咒(Dimension Curse)的复杂算法往往有好得多的效果。类似 Google 这样拥有海量数据的大型企业，优秀的应用效果不仅来自于精妙复杂的算法，也来自于其对海量训练数据的大规模分析挖掘。

（3）不同的算法具有不同的扩展能力

我们不能假设智能应用系统都可以通过简单增加服务器的方法来扩展性能。有些算法是有扩展性的，而另一些则不行。

例如如果我们要从数亿的文章标题里，找出标题相似的各个组的文章，注意并不是所有的聚类算法此时都能并行化运行的，你应该在设计系统的同时就考虑可扩展性。有些情况下你需要将数据切分成较小的集合，并能够让智能算法在各个集合上并行运行。设计系统时所选择的算法，往往需要有并行化的版本，而在一开始就需要将其纳入考虑，因为通常围绕着算法还会有很多相关联的商业逻辑

和体系结构需要一并考虑。

(4) 泛化能力

机器学习实践中最普遍存在的一个误区是陷入处理细节中而忘了最初的目标——通过调查来获得处理问题的普适的方法。测试阶段是验证某个方法是否具备泛化能力(generalization ability)的关键环节(通过交叉验证、外部数据验证等方法)，但是寻找合适的验证数据集不容易。如果在一个只有几百个样本的集合上去训练有数百万维特征的模型，试图想获得优秀的精度是很荒唐的。

(5) 考虑融入更多的可能性

你很可能听说过谚语“进来的是垃圾，出去的也是垃圾”(garbage in, garbage out)，在建立机器学习应用中这一点尤其重要。为了避免挖掘的效果失控，关键是要充分掌握问题所在的领域，通过调查数据来生成各种各样的特征，这样的做法会对提升分类的准确率和泛化能力有很大的帮助。仅靠把数据扔进分类器就想获得优秀结果的幻想是不可能实现的。