

一、专业实践训练整体情况

实践单位名称	杭州观图科技有限公司	
实践单位地点	杭州市西湖区三墩镇西园八路3号智汇众创中心E2-308	
实践岗位名称	算法工程师	
专业实践训练时间	分段进行	2021年09月16日开始 至 2022年04月16日结束
		专业实践训练累计 212 天（单位考核前），其中项目研究天数 120 天（单位考核前）
<p>（1）基本概况（含实践单位简介、实习实践内容等）</p> <p>单位简介： 杭州观图科技有限公司成立于2018年10月17日，位于浙江省杭州市西湖区三墩镇西园八路3号智汇众创中心E2幢301室，由杭州市西湖区市场监督管理局批准成立，是一家具有独立法人资格的软件开发公司。本公司主要业务为数字技术、数据处理技术、网络信息技术、计算机软件的技术开发、技术咨询、技术服务、成果转让，数字化服务，网页设计，计算机软硬件等。公司注册资本100万元，设施完善，设备先进。公司主要业务是图像视频文字检测识别、目标检测识别、软硬件协同技术，以及数据标注服务等解决方案。</p> <p>实践内容： 针对照片、复印件等，实现高精度、快速的卡证的检测和分割算法</p>		
<p>（2）项目研究概述（含项目名称、项目来源、项目经费、主要研究目标和技术难点等）</p> <p>项目名称：多卡证分类识别系统 项目来源：杭州观图科技有限公司 项目经费：20万元 主要研究目标：实现高精度、快速的多卡证的检测和分割算法 技术难点：卡证在图片上的位置检测、分类，卡证上的文本检测、识别</p>		

(3) 项目开展情况（含项目研究内容、研究方案及技术路线，研究团队分工、本人承担任务及完成情况，存在问题与改进建议等，不少于 500 字。）

研究内容：

在日常生活中办理各种业务时经常需要提供各种证件的照片、复印件等，收件单位需要提取这些证件信息，该项目的主要研究内容是利用 CV 技术，自动提取图片、复印件上的卡证、文本信息，并探索轻量级模型以部署于边缘设备

技术方案：

采用 YOLOX 模型作为目标检测框架

采用轻量级的骨干网络，网络宽度和深度比率分别设为 0.25 与 0.33

采用解耦的检测头，并进一步进行扩展，增加一个旋转角度检测头

采用无锚框的方案，对图片网格直接预测

输入图片尺寸设为 416*416 像素

采用 mosaic 图像增强，不采用 mixup 增强，不采用翻转图片等影响文字方向的增强方式

采用 cos 的 warmup 策略

训练结束前 10 个轮次关闭数据增强

团队分工：

校外导师高海东博士负责提供理论、技术指导与标注好的训练数据，周波负责代码编写、模型训练

项目总结：

首先，基于 yolox 开源代码进行了原论文的复现与参数微调；其次，针对已标注的卡证复印件图片，编写预处理、增强、训练等代码，并进行模型训练，实现了较为准确并且快速的卡证、文本的目标检测；第三，修改模型结构，在原模型基础上增加角度检测，并对图片增强方式进行了调整，进行多次试验后，获得了阶段性的成功。

该模型，约 0.91M 的参数量，可以实现比较优秀的目标检测性能，推理速度在 3090 显卡上约为 5-20 毫秒，在 nVidia Nano 级别的边缘芯片上基于 TensorRT 进行移植的话预计可以在 40ms 以内，基本上可以实现实时检测；但是，针对目标角度的准确回归相对而言还有待提高。

二、专业实践训练收获

(一) 围绕考核评价指标体系，举例说明以下收获（不少于 800 字）

1. 系统性学习了计算机视觉领域目标检测任务的理论知识

目标检测是计算机视觉领域中一个重要的分支，是自动驾驶等应用的核心技术。在深度学习时代，目标检测算法主要可分为两类，一类是基于 Region Proposal 的 R-CNN 系列模型（R-CNN[1]，Fast R-CNN[2]，Faster R-CNN[3]等），此类模型的特点是先预测出 Region，再进行分类与回归，训练过程相对繁琐，速度较慢，精度较好；另一类则是单阶段的算法，以 YOLO[4]-[9]系列为代表，此类算法可实现端对端的解决方案，训练、预测都相对简单，速度快，目前处于主流位置。

本项目采用了 YOLO 家族中最新，也是目前处于 SOTA 地位的 YOLOX。

1) YOLO: 算法实现了端到端的目标检测，大体流程为：将输入图片缩放到 448×448 ，送入 CNN，再对预测结果执行非极大值抑制（NMS），之后获得检测结果。具体的，首先将输入图片分割成 $S \times S$ 的网格，之后再每个单元格检测中心点落在该单元格内的目标。每个单元格预测 B 个边界框与相应的置信度，并给出单元格属于每个类别的置信度，再利用非极大值抑制，输出最终的预测结果。

2) 针对上述 YOLO 模型的一些缺点以及目标检测领域的先进技术，原作者进行了一系列的改进，产生了 YOLOv2。作者对 YOLOv2 提出了更好（Better）、更快（Faster）、更强（Stronger）的要求，在这几个方面对 YOLO 进行了改进，取得了 SOTA 的效果。YOLOv2 在 YOLO 的基础上进行了一系列的改进，在快速的同时达到 state of the art。同时，YOLOv2 可以适应不同的输入尺寸，根据需要调整检测准确率和检测速度。基于 YOLO v2 架构的 YOLO9000 开创性的提出联合使用分类样本和检测样本的训练方法，使对象检测能够扩展到缺乏检测样本的对象。

3) YOLO v3[6]延续了 YOLO 前两版的主要思想，包括单元格划分、端到端训练、多尺度训练等，并进行了若干改进进一步提升了算法的性能。YOLO v3 是一个优秀的架构，在兼顾准确率的同时能够以非常快的速度运行，并且便于对主干网络进行缩放，非常适合工程使用。YOLO v3 是原作者 YOLO 系列的最后一个版本，其 Darknet-53 网络结构、anchor 锚框、FPN 等非常优秀的结构依然被后继者沿用至今。

4) Yolov4 本质上和 Yolov3 相差不大，在原有 YOLO 目标检测架构的基础上，采用了近些年 CNN 领域中先进的优化策略，从数据处理、主干网络、训练、激活函数、损失函数等各个方面都进行了不同程度的优化。虽没有理论上的创新，但作者对各种优化算法的尝试，非常值得学习，其效果达到了实现精度与速度上的一个新的平衡。

5) 相比 YOLO v4, v5 并无太多改变，不过 YOLO v5 提出四种不同规模的模型，适应不同场景的需求，在灵活性上有极强的优势。通过控制网络的深度和宽度，YOLO v5 的主干网络可以实现很大的伸缩性，对于工程应用，有很多可借鉴的亮点。

6) 在设计上，YOLOX 遵循了几个原则：1. 所有组件全平台可部署；2. 避免过拟合 COCO，在保持超参规整的前提下，适度调参；3. 不做或少做稳定涨点但缺乏新意的工

作（更大模型，更多的数据）YOLOX 采用了 YOLO v3 的主干网络，并像 YOLO v5 一样，可以方便的对网络深度、宽度进行配置，从超轻量级的 YOLOX Nano 到重量级的 YOLOX-x，工程师和研究人员可以根据需求选择相应的网络结构。YOLOX 是一个优秀的工作，在尝试多种 Trick 的同时，兼顾了工程应用，尽可能使模型架构简化，达到了相当好的稳定与平衡。

2. 了解、掌握了 pytorch 深度学习框架

1) PyTorch 是相当简洁且高效快速的框架，设计追求最少的封装，设计符合人类思维，它让用户尽可能地专注于实现自己的想法，与 google 的 Tensorflow 类似，FAIR 的支持足以确保 PyTorch 获得持续的开发更新，PyTorch 作者亲自维护的论坛 供用户交流和求教问题，入门简单。与 Tensorflow v1 相比，pytorch 采用了动态计算图的设计，使得研究人员可以方便地进行算法调试，实时监视模型的中间结果；而与同样支持动态图的 tensorflow v2 相比，其原生的动态图 API 设计相对 TensorFlow 的易用性更好，接口更加清晰、简洁。

2) 混合精度训练

混合精度训练是在尽可能减少精度损失的情况下利用半精度浮点数加速训练。它使用 FP16 即半精度浮点数存储权重和梯度。在减少占用内存的同时起到了加速训练的效果。

其中核心的几个问题包括：

- a) 舍入误差：将训练过程中的数据利用 FP16 进行存储，同时复制一份 FP32 的权重，模型更新的时候在 FP32 下进行。
- b) Loss Scale：由于在训练的后期梯度会过小，FP16 表示会出现下溢现象，因此需要动态对损失进行缩放，使得梯度平移到 FP16 能够表示的范围内。
- c) 计算精度：在矩阵运算的过程中，为了保证舍入误差，可以采用 FP16 进行乘法运算，利用 FP32 进行加法计算，最终采用 FP16 进行存储。

3) 分布式训练

若设备允许，多卡训练可以有效提升训练效率。Pytorch 多卡训练主要过程为：

指定主机节点；

主机节点划分数据，一个 batch 数据平均分到每个机器上；

模型从主机拷贝到各个机器；

每个机器进行前向传播；

每个机器计算 loss 损失；

主机收集所有 loss 结果，进行参数更新；

将更新后参数模型拷贝给各个机器。

(二) 取得成效

1. 首先，基于 yoloX 开源代码进行了原论文的复现与参数微调；其次，针对已标注的卡证复印件图片，编写预处理、增强、训练等代码，并进行模型训练，实现了较为准确并且快速的卡证、文本的目标检测；第三，修改模型结构，在原模型基础上增加角度检测，并对图片增强方式进行了调整，进行多次试验后，获得了阶段性的成功。该模型，约 0.91M 的参数量，可以实现比较优秀的目标检测性能，推理速度在 3090 显卡上约为 5-20 毫秒，在 nVidia Nano 级别的边缘芯片上基于 TensorRT 进行移植的话预计可以在 40ms 以内，基本上可以实现实时检测；但是，针对目标角度的准确回归相对而言还有待提高。

人工智能技术的盈利主要集中在推荐、广告等领域。在探索人工智能技术的盈利模式的过程中，软硬件结合的边缘端人工智能设备展现出了巨大的市场潜力。边缘计算设备能够从物联网设备以及相关存储和分析中无缝收集数据。此外，它还支持本地实时决策。边缘的人工智能使用紧凑的架构，但是专门设计的神经计算加速芯片，可以高效的处理深度学习模型计算，驱动本地的数据信息进行决策。边缘设备越智能，成本就越高，但同时，它可以在本地处理和存储大量数据，从而减少在其他地方处理和存储数据的需要。一些通用的支持人工智能的边缘设备包括智能汽车传感器、消费者和商业机器人、无人机和安全摄像头、车牌识别、人脸识别等边缘设备，已经得到了广泛的应用。

在本项目中，采用的轻量级目标检测卷积神经网络，是具备边缘端应用的潜力的。在目标客户的实际需求中，一般会使用轻量级、易于移动的设备对卡证或者其复印件进行扫描，获得图像，并实时识别图像上的信息，提取准确范围的图像，有效减轻图像存储的负担；在后续的处理中，也可以有效减轻网络传输的负担；

2. 该实践项目属于人工智能领域中基于深度学习的研发，具体是计算机视觉技术，本人的学位论文为自然语言处理，两者不完全相符，但所采用的技术都基于深度学习，并且采用了相同的技术平台（pytorch），因此该实践项目中，本人学习了深度学习技术中混合精度训练、分布式训练等关键技术，为后续的学位论文研究打下了非常好的基础。

3. 在校期间主要研究成果【含产品与样机、专利（含申请）、著作、软件著作权、论文、标准、获奖、成果转化等】

成果名称	类别[含产品与样机、专利（含申请）、著作、软件著作权、论文、标准、获奖、成果转化等]	发表时间/授权或申请时间等	刊物名称/专利授权或申请号等	本人排名/总人数	学校排名/总参与单位数
------	--	---------------	----------------	----------	-------------

本人承诺

在专业实践训练及考核报告撰写过程中，如实提供材料，严守
学术道德、遵循学术规范。

签字：周波

2022年5月27日

三、考核评价

<p>校外合作 导师(或现 场导师) 评价</p>	<p>重点对研究生项目研究开展情况、职业素养、行业知识掌握、环境和岗位适应能力、工程实践能力、团队协作能力，以及通过技术创新、成果转化、解决工程实际问题等取得的经济和社会效益等方面的评价：</p> <p>该生动手能力强，思维敏捷，能够快速提出，解决问题的模型，以及模型之外的新思路等，有效地解决了实际中出现的问题，得到了很好的收益。</p> <p>校外合作导师（或现场导师）签字：高海东 2022年5月27日</p>
<p>校内导师 评价</p>	<p>重点对研究生科学素质、基础及专业知识掌握、技术创新能力、取得的研究成果、项目研究与学位论文撰写的相关程度等方面的评价：</p> <p>该生专业基础知识扎实，动手能力强，在专业实践期间努力学习掌握新知识，其解决工程问题的能力获得了校外合作导师的认可。</p> <p>校内导师签字：张寅 2022年5月27日</p>

实践单位 过程考核 意见	<p> 实际实践开始时间: 2021年9月16日 实际实践结束时间: 2022年4月16日 专业实践训练累计天数: 212 其中项目研究天数: 120 实践单位过程考核结果: <input checked="" type="checkbox"/>优秀 <input type="checkbox"/>良好 <input type="checkbox"/>合格 <input type="checkbox"/>不合格 审核签字并盖公章: 高海东 2022年5月2日 </p>
最终考核 结果审核 备案	<p> 考核总成绩 (由现场答辩考核成绩 90%+单位过程考核成绩 10%组成): 是否重修: <input type="checkbox"/>是 <input type="checkbox"/>否 教学管理部 (或相关分院) 审核签字 (公章): _____ 年 月 日 </p>

四、相关支撑材料

在校期间主要研究成果【含产品与样机、专利（含申请）、著作、软件著作权、论文、标准、获奖、成果转化等】证明材料原件扫描件，具体提交要求如下：

1. 产品与样机扫描件包含企业证明材料（含产品与样机功能及创新性介绍、社会经济效益、个人贡献说明及相关照片等）。

2. 授权专利扫描件包含专利证书授权页；未授权专利扫描件包含专利受理书扫描件和专利请求书扫描件。

3. 著作扫描件包含封面、封底和版权页。

4. 软件著作权扫描件包含著作权证书和登记申请表。

5. 论文扫描件包含封面、封底、目录和论文全文（含收录证明）。

6. 标准扫描件包含封面、版权页、发布公告、前言和目次。

7. 获奖扫描件包含显示单位和个人排名的获奖证书。

8. 成果转化扫描件包含企业证明材料（含成果技术说明、社会经济效益、个人贡献说明及相关照片等）。