基于深度学习的智能OCR识别关键技术及应用研究

Research on Key Technology and Application of Intelligent OCR Recognition Based on Deep Learning

王日花(中国传媒大学,北京 100024)
Wang Rihua(Communication University of China,Beijing 100024,China)

摘　要:
智能OCR技术作为人工智能领域的重要原子能力之一,在行业转型过程中发挥作用。对传统的OCR和基于深度学习的智能OCR技术进行对比,着重分析智能OCR技术的关键技术和部署优势;深度学习的融入实现了OCR对复杂文本的识别。其他信息化手段的综合应用,使智能OCR具备移动端适配、多任务检测、整行识别、图像分割定位和分类等功能,应用场景更加广泛,在图书情报领域中的图书文本和卡证识别会更高效。最后对智能OCR的技术发展和赋能做了展望。

关键词:
图书馆;智能服务;OCR识别;深度学习
doi:10.12045/j.issn.1007-3043.2021.08.005
文章编号:1007-3043(2021)08-0020-05
中图分类号:G250
文献标识码:A
开放科学(资源服务)标识码(OSID):

Abstract:
As one of the important atomic capabilities in the field of artificial intelligence, intelligent OCR technology plays a role in the transformation of the industry. It compares traditional OCR and intelligent OCR technology based on deep learning, focusing on the analysis of the key technologies and deployment advantages of intelligent OCR technology. The integration of deep learning realizes the recognition of complex text by OCR. The comprehensive application of other information methods enables intelligent OCR to have functions such as mobile terminal adaptation, multi-task detection, whole line recognition, image segmentation positioning and classification, etc., and the application scenarios are more extensive. Book text and card identification in the field of library information will be more efficient. Finally, the technical development and empowerment of smart OCR are prospected.

Keywords:
Library; Intelligent service; OCR recognition; Deep learning

引用格式:王日花. 基于深度学习的智能OCR识别关键技术及应用研究[J]. 邮电设计技术,2021(8):20-24.

0 引言

近年来,移动互联、大数据等新技术飞速发展,倒逼传统行业向智能化、移动化的方向转型[1-2]。随着运营集约化、数字化的逐渐铺开,尤其是以OCR识别、数据挖掘等为代表的人工智能技术逐渐深入业务场景,为用户带来持续的经济效益和品牌效应。图书情报领域作为提升公共服务的一个窗口,面临着新技术带来的冲击,必须加强管理创新,积极打造智能化的图书情报服务平台[3-5],满足读者的个性化需求。无论是高校图书馆还是公共图书馆,都需加强人工智能基础能力的建设,并与图书馆内部的信息化系统打通,优化图书馆传统的服务模式,提升读者的借阅体验。

影像分类和录入纸质材料是图书馆的常态生产需求,比如:拍照的图书文本和借阅证件信息的分类与录入,会消耗大量人力、物力和时间成本,影响业务流程的效率和用户体验。人工录入的效率和准确性低,且易受馆员情绪影响。长期从事繁琐机械的录入

收稿日期:2021-06-16

工作，对于馆员是极大的心理负担。智能OCR利用机器24 h连续工作，不受时间限制，可解决上述图书馆业务的痛点，提高影像处理效率。

1 传统OCR识别技术介绍

光学字符识别（Optical Character Recognition，OCR）指自动识别图像中的文字内容，属于人工智能机器视觉领域的一个重要的分支[6-8]，即把文本、卡证等载体上的文字通过光学等技术手段转化为计算机认识的电子化数据。传统OCR识别采用统计模式，处理流程较长，包括图像的预处理、二值化、连通域分析、版面分析、行切分、字切分、单字符识别和后处理等步骤。典型的传统OCR识别流程如图1所示。

传统OCR识别方法存在诸多弊端，汇总如下：

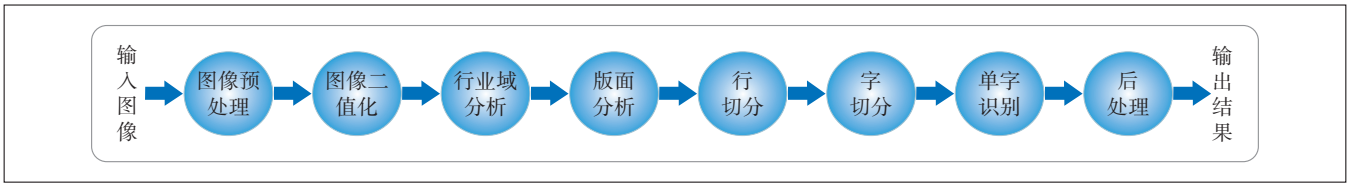

图1 传统OCR识别技术流程

a）在进行版面分析时，使用大量的规则，导致程序维护成本很高。

b）行业域分析完全依靠图像二值化得到的二值图，对于扫描文档效果尚可，面对手机拍摄和高拍仪取图时，难取得效果良好的二值化图，造成二值化过程中大量信息的丢失。

c）传统OCR技术包含8个模块，如图1所示，其中任何一个模块的不完善都会产生误差，误差的累积将导致识别率大幅下降。

d）传统OCR识别灵活性差，对于自然场景下拍摄的复杂样本基本无法处理，没有修改提升空间，可用性不高。

e）传统的方法将OCR系统割裂成过多的环节，倚重人工规则，需要在每个环节上引入人工干预并根据场景设定方法参数，难做到端到端的训练。

深度学习算法可以有效地规避传统OCR识别的不足，通过组合低层特征形成更加抽象的高层表示属性类别或特征，挖掘数据的分布式特征表示。借助神经网络来模拟人脑进行分析、学习和训练，即模仿人脑机制来分析图像、声音和文本等数据，被广泛应用于人工智能的模型构建和处理中。

2 基于深度学习的智能OCR识别概述

随着2012年Imagenet竞赛采用深度学习技术的AlexNet夺得冠军，深度学习算法开始应用于图像视频领域。基于深度学习的智能OCR技术是一次跨越式的升级[9-12]，深度学习算法实现整行识别，提升了OCR的识别率和识别速度，人工需要几分钟才能录入的文本，智能OCR技术可以秒速进行精准识别。智能OCR识别技术对识别流程进行了优化，优化后的识别流程包括检测、识别和后处理3个主要步骤，如图2所示。

基于深度学习的OCR定位与识别通过卷积神经网络CNN、循环神经网络RNN、长短期记忆网络LSTM技术实现，可在灰度图像上实现文字区域的自动定位和整行文字的识别，解决了传统OCR技术中单字识别无法借助上下文来判断形似字的问题。此外，智能OCR识别技术在低质量图片的容忍能力和识别准确率方面得到了显著的提升，可在印刷体低分辨率与模糊字符识别、印刷体复杂或者非均匀背景识别、印刷体多语言混合识别、印刷体艺术字体识别、手写小写数字识别、手写大写金额识别、手写通用文本识别等场景下实现高效的识别和分类。基于深度学习的智能OCR识别技术[13-15]支持移动设备拍摄的图像识别，可适用于对焦不准、高噪声、低分辨率、强光影等复杂背景。

除了在卡证识别、票据识别、表单识别、文档识别，智能OCR可应用于互联网广告推荐系统、UCG图片视频过滤、医学影像识别、街景路牌识别等。智能OCR识别属于多类分类问题，场景复杂、挑战性大；尤其是中文识别，字符集达到20 000类，而英文数字加字母只有62类。影响OCR识别效果的因素较多，比如背景的复杂度、字体的种类、分辨率的高低、多语言混合度、字体的排列、变形和透视情况等。

3 智能OCR的关键技术和创新应用

3.1 移动端适配和图像质量判断

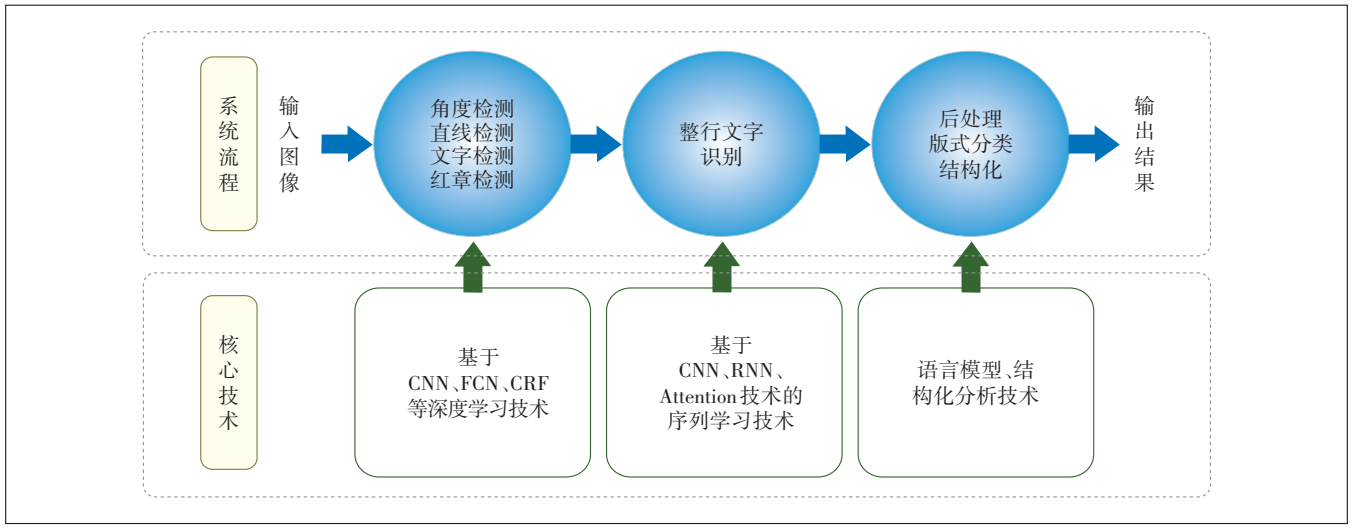

图2 智能OCR识别技术流程

图书馆生产需求更多的发生在移动端，用户更喜欢用手机拍照后即可识别，智能OCR技术综合已有的信息化技术，可在各种移动端实现适配。首先，基于轻量级深度学习技术，实现移动端的取图功能；其次，融合视频流识别技术，即从视频中识别出图书馆卡证的有效信息。深度学习网络可高效地学习到边缘情况，通过边缘的检测，得到物体的边缘轮廓，然后通过边缘跟踪合并，保障识别效果。移动端适配网络计算量很小，大多数的移动端设备均支持，即使透视变换很严重的图像也能很好地校正，保证移动端识别的准确率。

移动端图像的采集受光照强弱、拍摄抖动、对焦方式等条件影响，有时会导致采集的原始图像非常模糊，最终使得图像无法被有效地识别。基于此，需要将模糊的图像阻挡在识别之前，使得系统资源被合理的利用。基于深度学习的图像质量判断，提供一种图像质量判断能力，通过CNN学习得到输入图像质量的分类，给出判断的可信度。

3.2 多任务目标检测

角度检测和文本检测是文本识别的前提，可在杂乱无序、千奇百怪的复杂场景中准确定位出角度、直线、图章、文字等区域。由于图像可能带有一定角度，有的甚至有可能是90°以上倾斜或者倒立图像，需要检测出图像的主方向角度；处理的图像可能存在表格线，图章等，都需要检测出来；对于图像中的文字行区域，需按照文本行检测出每一块的外接四边形。传统的方法是功能模块分开，各自采用不同的网络进行定位，所需的网络规模巨大，串行效率较低。为解决此问题，可采用基于多任务（MultiTask）的FCN检测网络，将角度检测、直线检测、图章检测、文字检测融合在一个检测网络中，从输出的特征图中预测出需要检测结果。

3.3 整行识别的核心技术

文字图像是按照一定的规则和顺序排列的，OCR可看成是一种与语音识别类似的序列识别问题。基于与语音识别问题类似，OCR技术可视为时序依赖的词汇或短语识别问题。利用CNN+LSTM+Attention+CTC网络实现端到端的整行文字识别，精度和效率均有较大提升，下面介绍2种常见的整行识别算法。

3.3.1 基于CRNN的整行识别技术（CNN+LSTM+CTC）

基于联结时序分类CTC（Connectionist Temporal Classification）训练RNN的算法，在语音识别领域中相对于传统算法具有显著优势，所以尝试在OCR识别中借鉴CTC损失函数。CRNN就是其中代表性算法，CRNN算法输入100×32归一化高度的词条图像，基于7层CNN提取特征图，把特征图按列切分（Map-to-Sequence），每一列包含512个维度特征，输入到两层双向LSTM神经网络（每层包含256个单元格）进行分类。在训练过程中，通过CTC损失函数的指导，实现字符位置与类标的近似软对齐。CRNN借鉴语音识别中的LSTM+CTC的建模方法，不同点是输入的LSTM特征，从语音领域的声学特征（MFCC），替换为CNN网络提取的图像特征向量。CRNN算法把CNN做图像特征工程的潜力与LSTM做序列化识别的潜力结合，既提取了鲁棒特征，又通过序列识别避免了传统算法中难度

极高的单字符切分与单字符识别等问题，同时序列化识别也嵌入时序依赖（隐含利用语料）。

智能OCR识别技术通过改进LSTM+CTC算法，在CNN一侧，通过在卷积层采取类似VGG网络的结构，减少CNN卷积核数量的同时增加卷积层深度，既保证精度又降低时耗，同时加入BatchNorm机制。在RNN一侧，针对LSTM有对语料和图像背景过拟合的倾向，在双向LSTM单元层实现Dropout。在训练阶段，针对CTC loss对初始化敏感和收敛速度慢的问题，采用样本由易到难、分阶段训练的策略。在测试阶段，针对字符拉伸导致识别率降低的问题，保持输入图像尺寸比例，根据卷积特征图的尺寸动态决定LSTM时序长度。

3.3.2 联合CTC和Attention机制的整行识别

近年来，注意力机制广泛应用于语音识别、图像描述、自然语言处理等领域。就其在OCR的应用而言，注意力机制能够实现特征向量与原图字符区域的近似对齐，聚焦词条图像特征向量的ROI，优化深度网络Encoder-Decoder模型的准确率。相比于CNN+LSTM+CTC模型，注意力模型更显式的把当前时刻待分类字符与原图位置对齐，也更显式的利用前一时刻语料；注意力模型配合自回归连接，除了精度提升，收敛速度也加快了。

联合训练方案的精度更优，且收敛速度与CTC相当，注意力机制就是采用基于内容和历史相结合的方法。基于内容的方法利用上一步预测的字符向量和预测该向量的加权特征向量作为联合特征，LSTM的输入也来源于联合特征向量，并生成注意力机制的查询向量。基于历史的方法借助上一步的注意力，并利用CNN模型提取上一步注意力的特征，生成注意力机制索引向量的部分内容。除此，还在训练数据与技巧等方面做多处改进，如引入图像随机填补、依据每个batch内样本动态填补图像长度等。

3.4 多文档图像分割定位和智能分类

对于识别的各种票据、单据图像，如果一次只能上传识别一张，且需要指定图像必须正立的，会大大影响用户体验。多目标分割定位技术，可同时对一张图像上的不同目标进行分割定位，实现多种票据的同时识别。算法支持任意角度和任意方向的文档，分割得到最佳拟合文档的多边形，做到最大限度的所见即所得，有利于后面的图像校正和识别。

多图像的智能分类运用了分层特征融合方法，从图像分割开始就支持图像的大类分割分类，然后基于图像特征和OCR文本特征进行图像类别的精分类。图3是一种可注册的图像分类流程。

3.5 识别结果结构化

在各种场景中，要求不但要定位识别出图像中文字，还需要将图像分类到之前定义的版式中，方便图像归类和识别结果入库。在版式分类模块中，通过工具配置模板，然后利用模板信息对输入图像进行匹配打分，提取最大的匹配分数；当分数大于预定值时，则匹配成功，否则匹配不成功。整个版式匹配的算法流程图如图4所示。版式匹配分3个步骤。

第1步就是利用提取的直线，分析出表格各个格

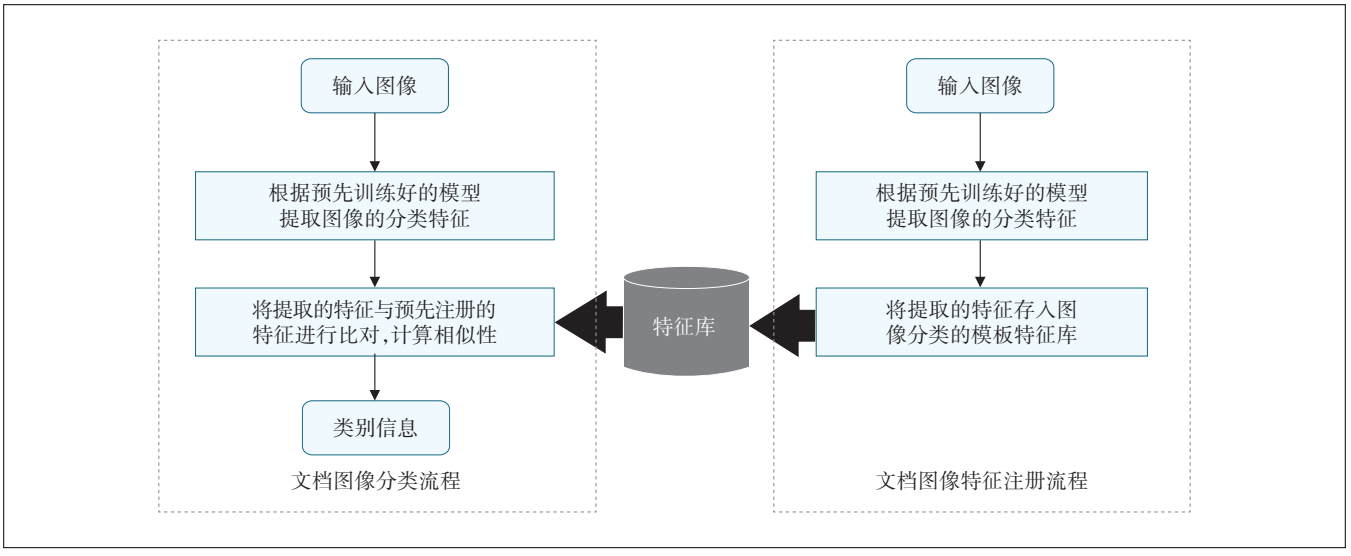

图3 智能OCR多文档图像智能分类

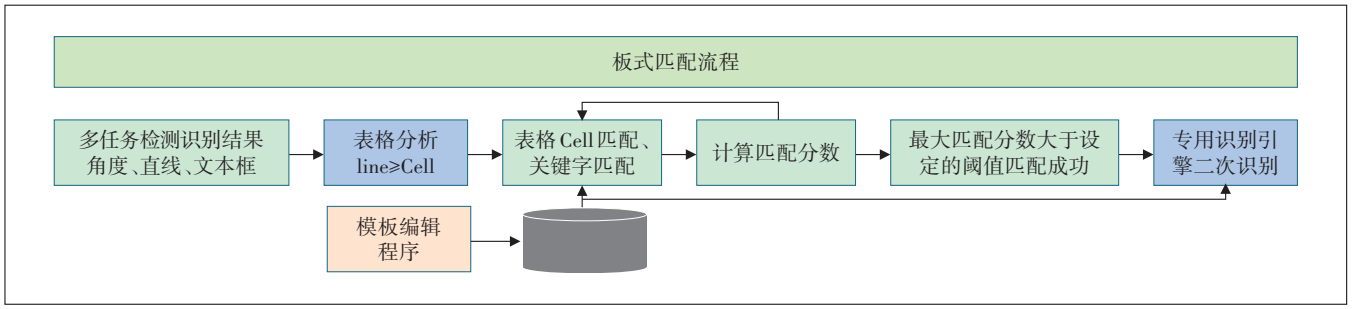

图4 智能OCR版式分类流程图

子(Cell)和表格的结构,将文字行纳入该Cell。

第2步,匹配表格结构、行列数量、表格Cell的相对尺寸、Cell占的行数和列数,特别是需要匹配表格Cell内部关键字。

第3步,计算线匹配分数和表格线匹配分数,计算关键字文本匹配分数并加权相加后得到最终的匹配分数。最后,计算所有的模板与识别结果的匹配分数,匹配分数最大者为表格分类结果,调用设定的多类识别核心,完成对应内容的二次识别。

4 结束语

本文对OCR技术和应用进行了分析,相比于传统OCR,基于深度学习的智能OCR技术具有识别准确率更高、速度更快、无格式依赖、支持私有化快速部署等优势,深度学习算法和模型构建也是OCR应用的关键。随着智能OCR技术不断演进,需要细化业务需求,和已有的信息化系统相结合,打造智能OCR创新服务模型,解决实际生产中的痛点问题[16-18]。以图书和情报领域为例,其本身的信息化水平有待提升,以OCR为代表的智能化应用相对不足;下一步,要以智慧图书馆建设为目标,需要调研已有的OCR识别应用,强化更多识别模块,以技术突破作为优化图书馆业态的基础,促进管理模式的创新,不断打造读者满意的图书和知识服务。

参考文献:

[1] 杨福义,叶其松. 人工智能时代知识工程的初步探索[J]. 人工智能与机器人研究,2021,10(1):9-28.
[2] 沈良朵,巩玉芳,高郁. 人工智能时代图书馆的发展机遇与变革趋势[J]. 中文信息,2020(1):29-30.
[3] 张晓霞. 人工智能在图书馆的应用与发展[J]. 大学图书情报学刊,2020,38(2):40-43.
[4] 王世伟. 深化人工智能与图书馆更新的若干问题——再论人工智能与图书馆更新[J]. 图书与情报,2020(3):93-103.
[5] 王师爽. 基于人工智能的图书馆服务策略研究[J]. 图书馆学刊,2020,42(4):69-72.
[6] 杨俊叶,刘佳,王丽. 计算机视觉技术在工业领域中的应用[J]. 科技创新导报,2020,17(1):108-109.
[7] 钱锦浩,宋展仁,郭春超,等. 基于时空共现模式的视觉行人再识别[J]. 自动化学报,2021:1-12.
[8] 袁南星,魏文武,刘明洁. 基于计算机视觉的水稻杂株识别研究[J]. 农机化研究,2020,42(1):213-216.
[9] 何文琦. 基于OCR技术的高校财务报销新探索[J]. 商业会计,2020(10):79-81.
[10] 全文举. 电力企业基于RPA技术助力财务智能化应用实践[J]. 电信科学,2020,36(1):139-143.
[11] 王振,魏志强. 交通标识牌字符提取算法[J]. 计算机应用,2011,31(1):266-269.
[12] 夏勇,戴汝为,肖柏华,等. 基于OCR与词形状编码的英文扫描文档检索[J]. 模式识别与人工智能,2009,22(3):488-493.
[13] 杨晨,马瑞成,王雨石,等. 深度学习与工业互联网安全:应用与挑战[J]. 中国工程科学,2021,23(2):95-103.
[14] 刘云,薛盼盼,李辉,等. 基于深度学习的关节点行为识别综述[J]. 电子与信息学报,2021,43(6):1789-1802.
[15] 郭旦怀,张鸣珂,贾楠,等. 融合深度学习技术的用户兴趣点推荐研究综述[J]. 武汉大学学报(信息科学版),2020,45(12):1890-1902.
[16] 姚慧慧. OCR技术下的医保费用智能审核研究——以蚌埠市为例[J]. 行政事业资产与财务,2020(7):31-32.
[17] 鲍相宇. 基于OCR技术的智能辅助评标解决方案探讨[J]. 招标采购管理,2018(7):56-58.
[18] 白翔,庞彦伟,章国锋. 计算机视觉中的深度学习专题(2020)简介[J]. 中国科学(信息科学),2020,50(2):303-304.

作者简介:
王日花,毕业于南开大学,中国传媒大学图书馆资源建设部副主任,硕士,主要从事信息咨询、信息资源建设工作