

基于 CRAFT 和 OCR 技术的药品名称识别方法

许莹 厉小明 于丰豪

湖南大学信息科学与工程学院 长沙 410082

(hnxxy@hnu.edu.cn)

摘要 在智能化药房的运作中,为实现药品的高效与精准挑选,机器人准确识别药品并完成取药至关重要。聚焦药品名称识别方法,提出一种融合 CRAFT 算法与 OCR 技术的 CRAFT-OCR 算法,以实现药品名称的高效识别。其中,CRAFT 算法用于检测药盒文本区域,为提升识别准确率,设计一种基于排序规则的药名区域定位方法来确定药名区域,最终借助先进的 OCR 技术完成文字识别。在采集的药盒图片数据集上开展的药名识别实验显示,CRAFT-OCR 方法检测药名区域的准确率为 96.43%,文字识别准确率为 96.00%,性能优于现有算法,为智能化药名识别提供了有效的解决方案。

关键词: 深度学习;图像处理;文本检测;文字识别;药名识别

中图分类号 TP317

Drug Name Recognition Method Based on CRAFT and OCR Technology

XU Ying, LI Xiaoming and YU Fenghao

College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China

Abstract In the operation of intelligent pharmacies, it is crucial for robots to accurately identify and retrieve drugs in order to achieve efficient and precise drug selection tasks. This study focuses on drug name recognition methods and proposes a CRAFT-OCR algorithm that integrates CRAFT algorithm and OCR technology to achieve efficient recognition of drug names. Among them, the CRAFT algorithm is used to detect the text area of the medicine box. To improve recognition accuracy, a drug name area localization method based on sorting rules is designed to determine the drug name area, and advanced OCR technology is finally used to complete text recognition. The drug name recognition experiments conduct on the collected dataset of medicine box images show that the accuracy of the CRAFT-OCR method in detecting drug name areas is 96.43%, and the accuracy of text recognition is 96.00%. The performance is better than existing algorithms in the literature, providing an effective solution for intelligent drug name recognition.

Keywords Deep learning, Image processing, Text detection, Text recognition, Drug name recognition

1 引言

随着自动化、信息化与人工智能技术的蓬勃发展,智能化药房系统的研究与产品开发成为医疗行业的热点。智能化药房系统集成药品存取机器人与先进的药品信息化管理系统,革新了传统药品发放模式^[1]。在智能化药房中,机器人取药的核心在于精准识别并抓取目标药品,这依赖于对药品存放位置与名称的精确认知。传统条形码扫描技术虽能快速识别药品,但受药品包装多样性影响,条形码位置不固定,限制了其应用范围。

近期,在药名识别或类似文本识别领域涌现出诸多研究成果。Liu 等^[2]将 CRAFT(Character Region Awareness for Text Detection)^[3]和 Tesseract^[4]技术相结合,分别用于药品文本区域检测和药名文字识别。Liu 等^[5]实现了一个基于 B/S 架构的药盒识别检测系统,采用 DBNet^[6-7]作为文本检测框架,并使用卷积递归神经网络(CRNN)^[8]作为文本识别框架,效果优于传统的 CTPN^[9]+CRNN。然而,这些方法在应对复杂的药品包装场景时,仍存在识别准确率不高、算法通用性受限等问题。另外,部分学者采用新的深度学习模型组合,如

将 Transformer^[10]架构与传统卷积神经网络相结合^[11],试图提升文本识别准确率,但在处理药盒复杂背景和不规则字体时仍存在不足。还有学者改进文本检测与识别方法,提出基于注意力机制的文本检测网络^[12],一定程度上提高了对特定场景文本的检测能力,但在药名识别的通用性方面有待增强。

为解决上述问题,本研究将 CRAFT 算法与光学字符识别(Optical Character Recognition, OCR)技术相结合,提出 CRAFT-OCR 算法。该算法创新性地将两种技术应用于药品识别领域,并针对药名区域定位设计基于排序规则的独特方法,以有效解决条形码位置不固定导致的识别难题,提升药名识别准确率。本研究主要工作如下:

1)提出了一种融合 CRAFT 和 OCR 的 CRAFT-OCR 算法,用于高效检测药名区域和识别药名。利用 CRAFT 检测文本区域,基于药名特征设计排序方法定位药名区域,借助 OCR 技术进行文字识别,提高药名识别准确率。

2)采集 1260 个药店常见药盒图片,构建药品图片数据集用于训练和测试。通过实验验证了基于排序规则的药名区域定位方法的有效性,并将 CRAFT-OCR 算法与文献中已有药名识别方法进行对比。

2 CRAFT-OCR 算法

药盒上药名的文字识别主要涉及文本检测和文字识别两项核心技术。本文采用的是 CRAFT 算法,该算法基于字符和字符间关系的感知,通过卷积神经网络生成字符中心点和字符间连接的概率图,实现对文本区域的精确检测。

与其他主流文本检测算法,如 SwinTextSpotter^[13] 和 DiZNet^[14] 等相比,CRAFT 算法具有显著优势。SwinTextSpotter 在处理大规模场景文本时表现较好,但对于药盒上复杂排列、尺寸较小的文本检测效果欠佳;DiZNet 专注于端到端的文本检测与识别,但在检测弯曲文本时存在局限性。CRAFT 算法基于单字和字间亲和度的检测方式,能够更精准地捕捉字符间的细微联系,即使面对药盒上方向各异、曲线排列或发生畸变的文本,也能准确勾勒出完整文本行。其采用 VGG-16^[15] 作为主干网络,运用批规范化(Batch Normalization, BN)^[16] 方式提取深层次语义特征,并结合类似 U-Net^[17] 的结构。这种结构通过自顶向下的特征聚合,将浅层结构特征和深层语义特征有效融合,增强了模型对细节的捕捉

能力,在复杂场景文本检测中展现出强大的鲁棒性和准确性。

OCR 技术的核心是将图像中的文字信息转化为计算机可识别的字符编码,其基本原理涵盖图像预处理、文字区域检测、字符分割、特征提取与字符识别等步骤^[18-19]。在文本识别领域,常见方法包括利用 CRNN 模型,以 CNN 特征作为输入,通过双向 LSTM 进行序列处理,提升文字识别效率和模型泛化能力;引入注意力机制的方法,通过注意力模型计算 RNN 状态和上一状态的注意力权重,实现更精准的识别。

本研究提出的 CRAFT-OCR 算法流程主要包括图像预处理、文本区域检测和文字识别 3 个模块,如图 1 所示。在文本区域检测阶段,CRAFT 算法通过训练深度神经网络预测字符区域以及字符之间的亲和力,采用弱监督方式,降低了对标注数据的依赖,提高了模型的泛化能力。检测到所有可能的文本区域后,算法根据药品名称的特有文本特征(如字体、大小、位置等),运用基于排序规则的药名区域定位方法,进一步精准定位药名所在区域。在文字识别阶段,选用合适的 OCR 技术对药名区域内的文字进行识别,确保药品名称识别的高效性与准确性。

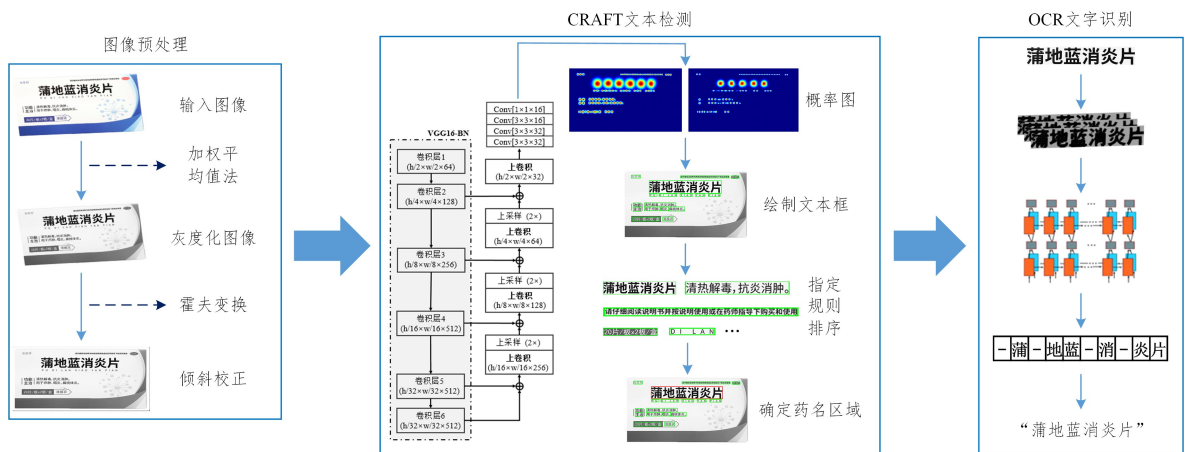


图 1 CRAFT-OCR 算法的药品名称识别流程

Fig. 1 Drug name recognition process of CRAFT-OCR algorithm

3 图像预处理

3.1 图像灰度化处理

在图像处理中,图像预处理是后续分析、特征提取及识别的重要基础,其中图像灰度化处理对提升处理效率和质量意义重大。相机捕获图像时,基于红(R)、绿(G)、蓝(B) 3 个颜色通道记录每个像素点的颜色信息,每个通道有 255 种取值,可组合出 16 581 375 种颜色(即 255^3)。丰富的色彩虽提升了图像表现力,但在图像处理时,庞大的颜色空间会导致计算量剧增,影响处理速度^[20]。为解决这一问题,研究人员提出灰度化方法^[21],将像素点的 R,G,B 3 个分量值统一处理为单一灰度值,实现从彩色图像三通道数据到单通道数据的转换。这不仅减少了计算量,还为后续处理提供了更高效的输入。

灰度化处理方法多样,包括最大值法、平均值法和加权平均值法等。最大值法将灰度化后的像素值取为 R,G,B 三者中的最大值,操作简单,但会使图像整体亮度过高,丢失大量细节;平均值法计算 R,G,B 3 个分量的平均值作为灰度值,虽简便,但易忽略通道间的色彩细节,影响色彩还原度。相比之下,加权平均值法更为精细有效,它依据人眼对不同颜色敏感度的差异,为 R,G,B 3 个分量分配不同权重,通常设置为:

$$Gray = 0.144 \times B + 0.587 \times G + 0.299 \times R \quad (1)$$

这一方法在灰度化过程中能更好地保留图像原始色彩信息和细节特征。因此,本研究选用加权平均值法进行图像灰度化处理,处理结果如图 2 所示。



图 2 图像灰度化处理结果

Fig. 2 Image grayscale processing result

3.2 图像倾斜校正

在药盒图像采集,药盒倾斜常导致其上文字倾斜,给后续文本检测与识别带来挑战。为改善这一状况,本研究采用基于霍夫变换的倾斜校正方法^[22]。霍夫变换的基本原理是:图像上的每条直线在霍夫空间中都对应一条正弦曲线,图像上相交于同一点的所有直线,在霍夫空间中对应的曲线会相交于同一点,该点坐标表示图像中直线的参数(即极角 θ)。

在图像处理中,直线可通过笛卡尔坐标系(直角坐标系)和极坐标系表示。笛卡尔坐标系中,通过点 (x, y) 的直线表

示为 $y=kx+b$,其中 k 是斜率, b 是截距。而在处理图像倾斜校正时,极坐标系更为便捷。在极坐标系下,直线由极径 r 和极角 θ 描述,具体形式为:

$$r = x \cos \theta + y \sin \theta \quad (2)$$

这一转换(如图 3 所示)将图像空间中的直线检测问题转化为霍夫空间(参数空间)中的点检测问题。

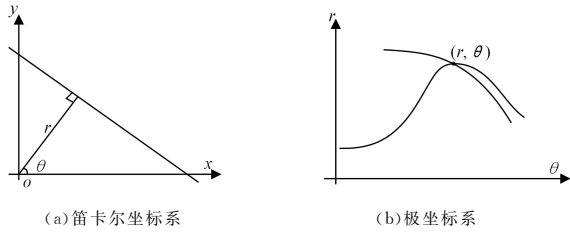


图 3 直角坐标系到极坐标系的转换^[23]

Fig. 3 Conversion from Cartesian coordinate system to polar coordinate system^[23]

具体来说,边缘检测完成后,在霍夫空间中相交于同一点的曲线数量最多的点对应的坐标 θ 即为所需的倾斜角度。

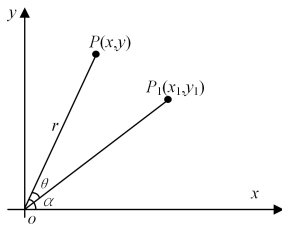


图 4 倾斜图像校正示意图^[23]

Fig. 4 Schematic diagram of tilted image correction^[23]

为直观理解,设倾斜图像中任意一点 P 的原始坐标为 (x, y) ,该点到原点 O 的距离为 r , OP 与 x 轴的夹角为 α 。根据极坐标与直角坐标的转换关系, P 点坐标表达为:

$$\begin{cases} x = r \cos \alpha \\ y = r \sin \alpha \end{cases} \quad (3)$$

为校正图像倾斜,需将图像旋转 $-\theta$ 度(因为 θ 是倾斜角,校正时需反方向旋转)。校正后,点 P 的新坐标 $P_1(x_1, y_1)$ 通过旋转矩阵计算得出,即:

$$\begin{cases} x_1 = r \cos(\alpha - \theta) \\ y_1 = r \sin(\alpha - \theta) \end{cases} \quad (4)$$

进一步利用三角函数的和差公式进行化简,可得校正前后坐标的直接关系:

$$\begin{cases} x_1 = r \cos \alpha \cos \theta + r \sin \alpha \sin \theta = x \cos \theta + y \sin \theta \\ y_1 = r \sin \alpha \cos \theta - r \cos \alpha \sin \theta = -x \sin \theta + y \cos \theta \end{cases} \quad (5)$$

即 x 坐标沿 θ 方向进行了旋转变换, y 坐标也相应变换但方向相反。

图 5 展示了图像倾斜校正前后的对比效果,凸显了校正对改善图像质量和后续处理的重要性。



图 5 倾斜校正前后的对比

Fig. 5 Comparison before and after tilt correction

4 文本检测

4.1 文本区域检测

文字识别首先要精准检测图片中的文字区域,当前主流方法包括传统文本检测技术和基于深度学习的文本检测方法。

传统文本检测方法在处理背景单一、清晰的图像时,能较好地定位文本。然而,面对药盒等背景图案复杂、文字排列多样的图像,其检测效果不佳。近年来,基于深度学习的文本检测技术逐渐兴起,成为处理复杂图像文本检测问题的首选方案,如 SwinTextSpotter^[13], DiZNet^[14] 等,以及一些多模块组成的文本检测器^[24]。这类方法借助强大的数据拟合能力,通过海量数据集优化模型,显著提升了泛化性能,在复杂场景文本检测中表现出色,尤其在药盒文字识别领域优势明显。

本研究采用 Baek 提出的 CRAFT 算法^[3],该算法基于单字与字间亲和度进行文本检测,将文本检测任务转化为对单个字符及其相互连接关系的识别。CRAFT 算法的网络结构见图 6,它以 VGG-16^[15] 作为主干网络,运用批规范化(BN)提取深层次语义特征,并采用类似 U-Net^[17] 的结构。

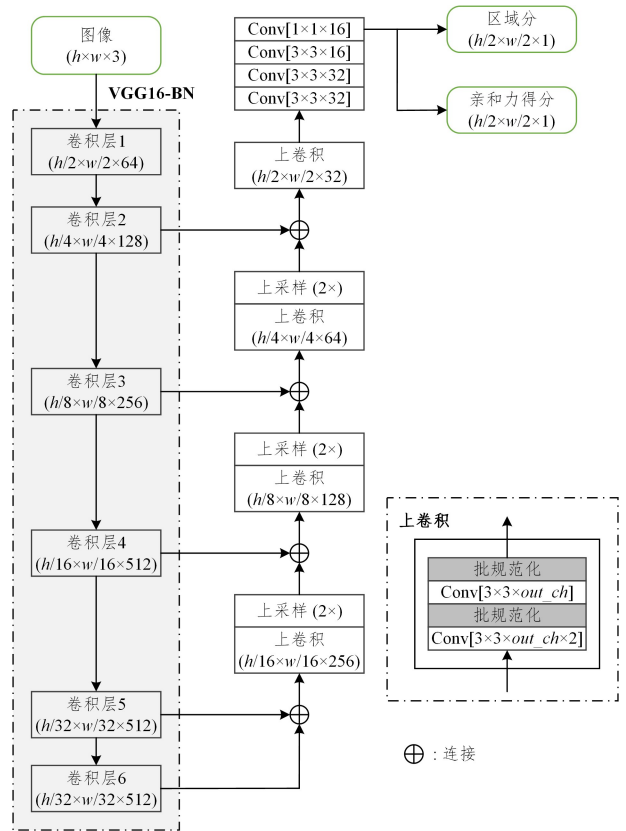


图 6 CRAFT 网络结构图

Fig. 6 Architecture diagram of CRAFT network

在 U-Net 结构中,CRAFT 通过自顶向下的特征聚合方式,结合上采样和浅层特征进行特征拼接,增强了对细节的捕捉能力。网络添加一系列卷积层,最终的 1×1 卷积层使用两个卷积核输出两个分支结果(两个通道的特征图),分别代表区域分数(Region Score)和亲和力分数(Affinity Score),用于表示每个像素点属于单字符中心区域的概率和相邻字符区域中心的概率。获得输出概率图后,通过二值化阈值计算、连通组件标记(CCL)以及使用 OpenCV 的 Rectangle 函数等精细化后处理步骤,将概率图转化为精确的文本边框,实现对图片

中文字区域的完整捕捉与标识。

4.2 确定药名区域

药品名称在药盒表面的位置不固定,无法直接依据识别出的文字区域确定药名所在区域。为此,本研究通过分析药品名称特性对文字区域进行排序,以确定药名区域。通过观察常见药盒图片,药品名称具有以下特点。1)字体较大:为突出药品种类,方便顾客识别,药盒上药品名称字体通常明显大于其他文字,使得包含药品名称的矩形区域高度大于其他文字区域。2)位置偏上:药品名称一般位于药盒上方,可根据矩

形区域中心点位置排序,优先识别上方文字区域。3)面积较大:由于字体较大,药品名称占用面积比例相对较大。4)字数为5到7个字:多数药品名称由5到7个字组成。

基于上述药名区域特性,本研究提出一种基于排序规则的药名区域定位方法。具体排序规则为:首先按矩形区域面积排序,选取前三名;再依据药名字体较大的特性,按矩形区域高度排序,选取前两名;最后根据药名一般位于药盒上方的特点,用矩形区域面积除以矩形区域中心点的y坐标值,最大者即为药名区域。排序过程如图7所示。

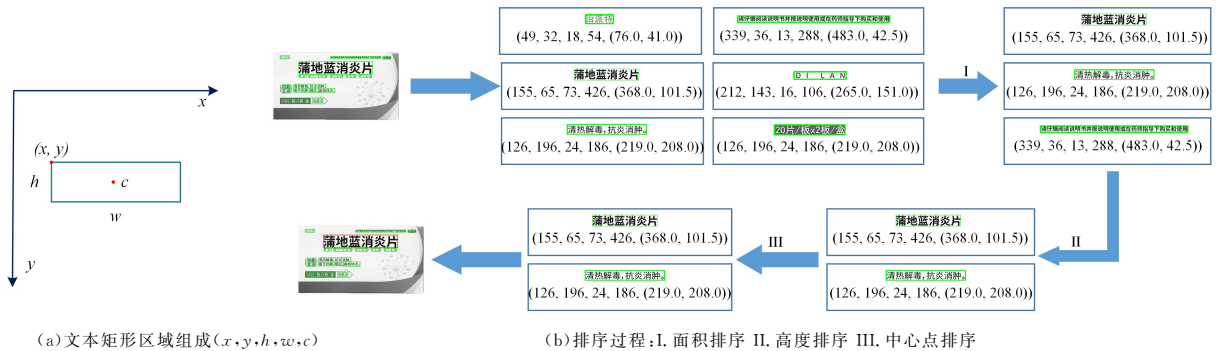


图7 排序方法原理图

Fig. 7 Schematic diagram of sorting method

采用该方法进行药名区域检测的结果如图8所示。识别字符后,利用 OpenCV 的 rectangle 函数绘制文本区域矩形框(红色表示药名区域,绿色表示其余文本区域)。对140个药盒进行检测,所提方法对药名区域的检测准确率达96.43%,识别效果良好。

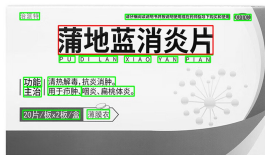


图8 药盒文字区域识别结果

Fig. 8 Recognition results of medication box text area

5 文字识别

为选择适用于本研究药名识别的技术,选取3种具有代表性的OCR技术分别与CRAFT结合进行研究。

5.1 Tesseract 技术

Tesseract^[4]是惠普公司开发的开源OCR引擎,其文本识别能力备受认可。Tesseract 5.0版本引入全新的神经网络模型LSTM(Long Short Term Memory Network)^[25](见图9),相较于传统的Tesseract 3,在功能和性能上有显著提升。

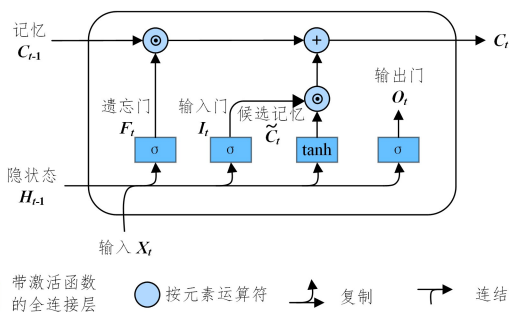


图9 LSTM网络结构图^[26]

Fig. 9 Architecture diagram of LSTM network^[26]

LSTM是对循环神经网络(Recurrent Neural Network, RNN)的优化改进,旨在解决传统RNN处理长距离文本信息时面临的梯度消失或梯度爆炸问题。LSTM通过独特的网络结构,包含输入门、遗忘门和输出门,协同控制信息在细胞状态中的保留、遗忘和输出。同时,其隐藏状态的细胞用于记录和文本额外信息,有效捕捉文本长期依赖关系,且计算代价较低。

5.2 CRNN 算法

CRNN(Convolutional Recurrent Neural Network)^[8]是融合CNN^[27]和RNN^[28]的深度学习模型,用于端到端的文本识别任务。CRNN利用CNN提取图像特征,将其输入RNN进行序列建模,实现对不定长度文本序列的端到端识别。CRNN网络结构及参数如图10和表1所示。

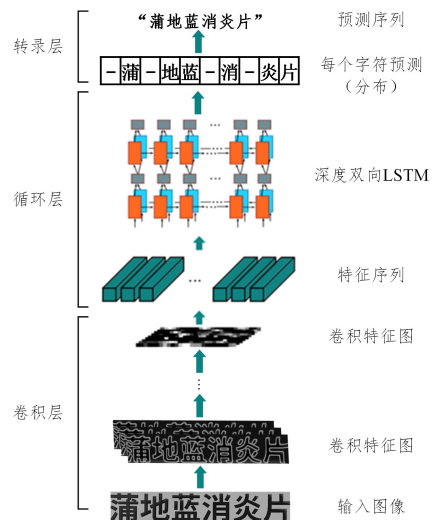


图10 CRNN结构图

Fig. 10 Structure diagram of CRNN

CRNN主要由卷积层、循环层和转录层组成。卷积层负责提取图像特征,通过多个卷积层和池化层构建深层CNN

网络,逐步减小特征图大小和维度;循环层通常采用双向 LSTM 循环单元,对 CNN 提取的特征序列进行序列建模,捕捉输入序列中的时序信息;转录层使用 CTC 损失,将循环层获取的标签分布转换为最终标签序列。

表 1 CRNN 配置参数

Table 1 Configuration parameters of CRNN

类型	参数配置
双向 LSTM	隐藏节点数 256
双向 LSTM	隐藏节点数 256
卷积层	输出通道数 512,卷积核大小 2×2,步长 1,边界填充 0
最大池化层	池化窗口 1×2,步长 2
卷积层	输出通道数 512,卷积核大小 3×3,步长 1,边界填充 1
卷积层	输出通道数 512,卷积核大小 3×3,步长 1,边界填充 1
最大池化层	池化窗口 1×2,步长 2
卷积层	输出通道数 256,卷积核大小 3×3,步长 1,边界填充 1
卷积层	输出通道数 256,卷积核大小 3×3,步长 1,边界填充 1
最大池化层	池化窗口 2×2,步长 2
卷积层	输出通道数 128,卷积核大小 3×3,步长 1,边界填充 1
最大池化层	池化窗口 2×2,步长 2
卷积层	输出通道数 64,卷积核大小 3×3,步长 1,边界填充 1
输入层	宽 W×32 的灰度图

5.3 百度 OCR 算法

百度 OCR 算法是百度在人工智能领域的重要技术成果,依托其强大的深度学习平台和大数据处理能力,具备高度的准确性与强大的识别能力,在众多应用场景中得到广泛应用,如文档电子化、办公自动化、财税报销、远程身份认证以及教育阅卷等。百度为用户提供了丰富多样的场景化文字识别解决方案,涵盖通用场景文字识别、卡证文字识别、交通场景文字识别等^[29]。在本研究中,选用通用场景文字识别。用户需先在百度云 AI 开放平台注册并申请 OCR 应用,随后在程序里调用 API 接口实现文字识别功能。

在使用百度 OCR 技术进行药名识别测试时,为确保其与 Tesseract 和 CRNN 在同一层次、同一尺度下进行公平对比,对所有参与测试的图像进行统一的预处理操作,包括前文所述的图像灰度化、倾斜校正等。经查阅百度 OCR 技术文档及相关实验验证,百度 API 接口算法内部包含图像增强等预处理步骤。为保证实验的公正性和可比性,在对比实验中,对 Tesseract 和 CRNN 处理的图像也进行类似的增强处理,使得 3 种 OCR 技术在处理相同质量图像数据的基础上进行性能对比。

6 实验与分析

6.1 实验环境配置

本实验基于 Windows 系统开展,借助 Miniconda 创建虚拟环境。Miniconda 作为 Anaconda 的轻量级替代方案,具有占用空间小、运行速度快的优势,便于用户构建所需的 Python 环境。虚拟环境配置如下:Python 版本为 3.9.5,PyTorch 版本为 1.12.1+cu116,Tesseract 版本为 5.3.0,pytesseract 版本为 0.3.10。采用 PyTorch 框架搭建 CRNN 网络结构,运行环境参数如表 2 所列。

利用采集的药品图片数据集对 CRNN 进行训练。首先运用 CRAFT 算法获取所有药品图片的药名区域图片,并剔除错误样本,然后按照 7:2:1 的比例将数据集划分为训练集、验证集和测试集。训练过程中设置相关参数,具体包括:训练轮数 *epochs* 为 200 轮,批次大小 *batch_size* 为 8,学习率 *lr* 为

0.001,选用 Adam 优化器。每完成一轮训练便进行一次验证,保存准确率最高和最后一轮的模型权重。

表 2 CRNN 运行环境参数

Table 2 Operating environment parameters of CRNN

名称	版本
操作系统	Windows 10
GPU 运行环境	NVIDIA GeForce RTX 2060,CUDA 11.6
编程语言	Python 3.9.5
深度学习库	PyTorch 1.12.1

6.2 对比实验及结果分析

为验证基于排序规则的药名区域定位方法的有效性,开展对比实验,比较 CRAFT 网络在不同排序规则下定位药名的准确率。选取 140 个药品样本进行测试,结果如表 3 所列。其中,CRAFT-3 采用本文提出的排序规则,CRAFT-1 以文本区域高度最大作为药名区域判断标准,CRAFT-2 则以文本区域面积最大为判断依据。

表 3 CRAFT 网络在不同排序规则下的药名区域识别结果对比
Table 3 Comparison of drug name region recognition results of CRAFT network under different sorting rules




采用不同排序策略的 CRAFT	样本数	定位药名的准确率/%
CRAFT-1	140	85.71
CRAFT-2	140	93.57
CRAFT-3	140	96.43

从表 3 数据可知,以文本区域高度最大作为药名区域时,CRAFT 网络的定位准确率为 85.71%;以面积最大作为判断标准时,定位准确率提升至 93.57%,显著高于高度最大策略;而采用本文提出的排序规则时,CRAFT 网络的定位准确率进一步提高到 96.43%。实验结果充分验证了基于排序规则的药名区域定位方法的有效性。

第二组实验将 CRAFT 技术分别与上述 3 种 OCR 技术相结合,测试算法性能。为增强实验结果的广泛性和说服力,从数据集中随机抽取 50 个具有代表性的药品图片进行测试,涵盖不同类型、不同包装的药品,包括西药、中药,以及药盒颜色、文字排版各异的样本。实验结果如表 4 所列。

表 4 采用不同 OCR 技术的药名识别结果对比示例

Table 4 Comparison examples of drug name recognition results using different OCR technologies

药品样本	CRAFT-Tesseract	CRAFT-CRNN	CRAFT-OCR
	维 C 银起片	维 C 银翘片	维 C 银翘片
	阿莫西林胶费	阿莫西林胶囊	阿莫西林胶囊
	复方氨酚烷胺片	复方氨酚烧胺片	复方氨酚烷胺片

实验结果显示,采用百度 OCR 的 CRAFT-OCR 算法在识别准确率上优于采用其他 OCR 技术的算法。百度 OCR 技术基于深度学习算法,不仅具备高精度的识别能力,还支持多语种文字识别,能准确地将图片中的文字信息转换为可编辑、可搜索的文本格式。通过大量数据训练,其识别精度和泛化能力不断提升,能够有效识别包括中文、英文、法语、德语、日语、韩语等多种语言的文字信息。此外,百度 OCR 提供了稳

定易用的在线 API、离线 SDK、软件部署包、一体机等多种服务形式,用户可根据实际需求灵活选择,这使其在药盒文字识别中展现出独特优势,因此本研究选用百度 OCR 技术作为药盒文字识别的方法。

第三组实验对比 CRAFT-OCR 与文献中已有算法 CRAFT-Tesseract^[2]和 DBNet-CRNN^[5],传统文字识别方法 CTPN^[9]-CRNN,以及 CRAFT 采用基于 CRNN 的 OCR 技术在药名识别中的表现。使用 140 个药品图片进行测试,部分药品样本识别结果与实验结果对比如表 5 所列。

表 5 CRAFT-OCR 与其他算法的识别准确率对比

Table 5 Comparison of recognition accuracy between CRAFT-OCR and other algorithms

算法名称	样本数	药名识别准确率/%
CRAFT-Tesseract ^[2]	140	84.47
CRAFT-CRNN	140	86.07
CTPN ^[9] -CRNN	140	89.29
DBNet-CRNN ^[5]	140	92.86
CRAFT-OCR	140	96.00

由表 5 可知,CRAFT-Tesseract 算法的识别准确率为 84.47%,最低。这表明 Tesseract 作为开源 OCR 引擎,虽然使用方便,但在处理复杂背景和不规则文本时存在一定局限性。CRAFT-CRNN 和 CTPN-CRNN 算法的识别准确率分别为 86.07%和 89.29%,略高于 CRAFT-Tesseract 算法,显示出这两种算法在文字识别方面具有一定优势,但仍有提升空间。DBNet-CRNN 算法识别准确率为 92.86%,在对比算法中排名第二,这得益于 DBNet 算法在文本检测方面的较高准确性,从而提高了整体识别准确率。CRAFT-OCR 算法的识别准确率达到 96.00%,在所有算法中表现最佳,充分证明了 CRAFT-OCR 算法在文字识别方面的卓越性能。与其他算法相比,CRAFT-OCR 算法在处理复杂背景和不规则字体时具有更强的鲁棒性和准确性。

结束语 本研究深入开展药品名称识别方法的理论分析与实验验证,将 CRAFT 算法与 OCR 技术有机结合,提出 CRAFT-OCR 算法,并设计基于排序规则的药名区域定位方法,有效提高了药名识别准确率。通过采集 1260 个药盒样本构建数据集进行药名识别实验,验证了基于排序规则的药名区域定位方法的有效性。在对比多种 OCR 技术与 CRAFT 结合的文字识别准确率后,选定百度 OCR 作为药品文字识别技术。与文献中已有算法相比,CRAFT-OCR 算法取得了最优的药名识别准确率。在未来研究中,可进一步探索和改进文字识别算法。一方面,增加与药品相关的训练样本,涵盖更多不同类型、不同包装的药品,丰富数据多样性;另一方面,扩大药品名称字体库,纳入更多字体样式和变体,以适应复杂多变的药盒文字情况,从而进一步提高药名识别的准确率和效率,推动智能化药房技术的发展。

参考文献

[1] XIE J L, CHEN M Y, XIE Y Y. Application of Hospital Logistics Robots in Intelligent Pharmacy[J]. China Health Standard Management, 2020, 11(11): 20-22.

[2] LIU D Y, ZHANG F S, MENG T, et al. Research on Drug Name Recognition Technology for Medication Boxes Based on Deep

Learning[J]. Journal of Qingdao University (Natural Science Edition), 2021.

[3] BAEK Y, LEE B, HAN D, et al. Character region awareness for text detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 9365-9374.

[4] SMITH R. An overview of the Tesseract OCR engine[C]// Ninth International Conference on Document Analysis and Recognition (ICDAR 2007). IEEE, 2007, 2: 629-633.

[5] LIU Q X, HAN Y W, MING Z. Pill Box Text Identification Using DBNet-CRNN[J]. International Journal of Environmental Research and Public Health, 2023, 20(5): 3881.

[6] LIAO M, WAN Z, YAO C, et al. Real-time scene text detection with differentiable binarization[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020: 11474-11481.

[7] LIAO M H, ZOU Z S, WAN Z Y, et al. Real-time scene text detection with differentiable binarization and adaptive scale fusion [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(1): 919-931.

[8] SHI B, BAI X, YAO C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(11): 2298-2304.

[9] TIAN Z, HUANG W, HE T, et al. Detecting text in natural image with connectionist text proposal network[C]// Computer Vision - ECCV 2016: 14th European Conference, Amsterdam, the Netherlands. Springer International Publishing, 2016: 56-72.

[10] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017: 6000-6010.

[11] GUO J, HAN K, WU H, et al. Cmt: Convolutional neural networks meet vision transformers[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 12175-12185.

[12] SONG P P, ZENG X J, ZHENG A Y, et al. Natural Scene Text Detection Based on Attention Mechanism[J]. Electronic Measurement Technology, 2021, 44(4): 6.

[13] HUANG M, LIU Y, PENG Z, et al. Swintextspotter: Scene text spotting via better synergy between text detection and text recognition[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 4593-4603.

[14] ZHOU D, ZHANG J, LI C. DiZNet: An end-to-end text detection and recognition algorithm with detail in text zone[J]. Journal of Visual Communication and Image Representation, 2024, 104: 104261.

[15] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409. 1556, 2014.

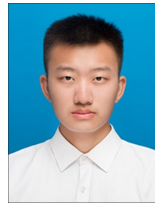
[16] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]// International Conference on Machine Learning. PMLR, 2015: 448-456.

[17] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// 18th

- International Conference Medical Image Computing and Computer-assisted Intervention (MICCAI 2015). Springer International Publishing, 2015; 234-241.
- [18] HUANG X, SHEN T, WANG R, et al. Text detection and recognition in natural scene images[C]// 2015 International Conference on Estimation, Detection and Information Fusion (ICEDIF). IEEE, 2015; 44-49.
- [19] WANG J X, WANG Z Y, TIAN X. Review of natural scene text detection and recognition based on deep learning[J]. Journal of Software, 2020, 31(5): 1465-1496.
- [20] ZHANG G H, FENG Y B, LU W D. Grayscale processing of images and acquisition of feature regions[J]. Journal of Qiqihar University; Natural Science Edition, 2007, 23(4): 49-52.
- [21] DE QUEIROZ R L, BRAUN K M. Color to gray and back: color embedding into textured gray images[J]. IEEE transactions on image processing, 2006, 15(6): 1464-1470.
- [22] LI Y H, HUANG Z H, XU X, et al. Tilt image correction technique based on Hough transform[J]. Journal of Hunan University of Engineering (Natural Science Edition), 2019, 29(3): 30-32.
- [23] LIU D Y. Research on Machine Vision based Medicine Box Detection and Drug Inventory Management System[D]. Qingdao: Qingdao University, 2021.
- [24] HAN X, GAO J, YANG C, et al. Focus entirety and perceive environment for arbitrary-shaped text detection[J]. arXiv: 2409.16827, 2024.
- [25] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [26] ASTON. Hands on Deep Learning [M]. People's Posts and Telecommunications Press, 2019.
- [27] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [28] ZAREMBA W, SUTSKEVER I, VINYALS O. Recurrent neural network regularization[J]. arXiv: 1409.2329, 2014.
- [29] LI Y. Building Atlas Retrieval System Based on OCR Recognition Technology[D]. Shijiazhuang: Hebei Normal University, 2020.



XU Ying, born in 1978, Ph.D, associate professor, is a member of CCF (No. 21672D). Her main research interests include multi-objective optimization, evolutionary optimization and intelligent computing.



LI Xiaoming, born in 2001, postgraduate. His main research interests include robot perception and understanding, and robot navigation.