

结合 OCR 和百度 AI 的智能健康饮食系统设计

谢香艳¹ 黄 逊¹
XIE Xiangyan HUANG Xun

摘 要

随着生活水平的提高,人们对健康饮食的关注度日益提升。合理的膳食搭配和营养摄入,有助于增强免疫系统功能,提高身体抵抗力,同时对心理健康也有积极影响。然而,在实际生活中,人们在饮食选择和营养分析方面面临诸多挑战。传统的饮食选择方式往往依赖个人经验和有限的信息,难以满足个性化的营养需求。同时,对食物营养成分的分析也较为困难,需要专业知识和复杂的检测手段。在此背景下,文章致力于设计并实现一款融合 OCR (optical character recognition) 与百度 AI 技术的智能健康饮食系统,其主要功能包括菜品随机选、菜品识别以及食物成分分析,在菜品随机选功能中,系统通过高效的抽取方式,为用户生成饮食建议,解决日常饮食选择困难的问题。简而言之,菜品识别功能依托百度 AI 的图像识别能力,可快速、准确地识别出菜品名称;食物成分分析功能则采用表格识别算法,通过融合 DB 分割模型、OCR 文本检测算法与 SLANet 表格定位模型,实现对食物成分表格的精准解析。实验结果表明,在海量数据场景下,采用“rand 函数与子函数相结合”的查询方法,较传统方法查询速度提升 30.981 s;百度 AI 菜品识别的置信度达 0.687 4,可较好完成菜品内容识别;同时,系统在食物成分表的表格识别上也达到了良好精度,整体满足稳定高效的算法设计要求。

关键词

DB 分割模型; OCR 文本检测算法; SLANet 表格定位模型

doi: 10.3969/j.issn.1672-9528.2025.10.020

0 引言

随着社会节奏的加快,大部分人群因上班、上学等因素,忽略了健康饮食的重要性。但随着国民生活水平的不断提高,人们对健康饮食的关注度也日益增长,作为维持身体健康的基石,其重要性不言而喻。

合理的饮食结构不仅有助于预防如高血压、糖尿病等各类慢性疾病,还能提升身体免疫力,改善精神状态。然而,尽管健康饮食的理念深入人心,但在实际生活中,人们面对食物,出现了“选择困难”。

一方面,“吃什么”成为许多人每日纠结的问题。面对琳琅满目的食材和各式各样的菜谱,常常陷入选择困难。传统的饮食决策方式,如参考家人朋友的建议、翻阅纸质食谱等,不仅效率低,且难以满足个性化的营养需求。另一方面,人们对所摄入食物的成分和营养价值缺乏清晰认知。多数时候,人们仅凭借经验或食物的外观来判断其健康程度,却无法准确了解食物中热量、蛋白质、脂肪、维生素等关键营养

成分的具体含量,这使得在追求健康饮食的道路上犹如“盲人摸象”,难以制定科学合理的饮食规划。

随着信息技术的飞速发展,人工智能和计算机视觉技术为解决饮食难题提供了新的途径。在健康饮食方面,有人将人工智能应用到菜品的自动结算中^[1],刘宇豪等人^[2]设计了健康饮食在微信小程序上的应用;吴思源等人^[3]也通过推荐算法给用户推荐了饮食,但仍有多个维度值得深入优化。光学字符识别(OCR)技术能够将图片中的文字信息转化为可编辑的文本,为信息提取和处理带来了极大便利。其中,张守涵^[4]将 OCR 算法应用于失物招领的智能识别中;左涛^[5]将其应用于不动产登记中得到了应用;张少宇^[6]也对此做出了贡献。百度 AI 平台凭借其强大的图像识别、数据分析等能力,在众多领域得到广泛应用,其中张牧琢等人^[7]和杨民峰等人^[8]分别在图像识别和植物识别方向进行了应用。基于这些计算机技术,本文设计并实现了一款智能健康饮食系统,系统的设计符合国家卫生健康委于 2025 年提出的“体重管理年”的计划,在强调科学普及的同时,提升公众体重管理意识。

该系统集成了多项实用功能。“吃什么”功能从海量的数据库中随机抽取菜品,数据集来源于下厨房食谱语料库(XiaChuFang Recipe Corpus),下厨房食谱语料库属于公开

1. 广州工商学院 广东佛山 528138

[基金项目] 广州工商学院实践教学项目“Web 开发技术(Java)实验”(SYKC2024041)

数据集，其提供了 100 多万条菜谱数据，能为用户提供丰富多样的饮食选择，满足不同个体在不同场景下的饮食需求。菜品识别功能借助百度 AI 的先进技术，根据用户拍摄的菜品图片，快速准确地识别菜品名称以及营养成分。而食物成分表格识别算法融合了 DB 分割的文本检测算法、CRNN 模型以及 SLANet 模型，能够精准解析食物成分表格，让用户清晰了解食物的营养构成。通过这些功能的有机结合，该系统旨在为用户打造一站式的智能健康饮食解决方案，帮助人们轻松实现科学、健康的饮食生活。

1 功能需求分析

为打造智能健康饮食系统，本系统功能模块图如图 1 所示。系统分为客户端、服务端以及第三方接口，客户端运用微信小程序技术，借助微信庞大的用户基础和便捷的使用特性，为用户提供了一个易于操作、交互性强的前端界面。借助百度 AI 平台这一第三方接口，系统可通过拍摄的菜品图片快速准确识别菜品名称。服务端采用 SpringBoot + MySQL + Python 技术。其中，SpringBoot 框架负责搭建稳定高效的后端服务架构；MySQL 数据库用于存储海量的食谱、食物成分等数据；Python 则用于实现复杂的数据处理逻辑，包括运行食物成分表格识别算法。

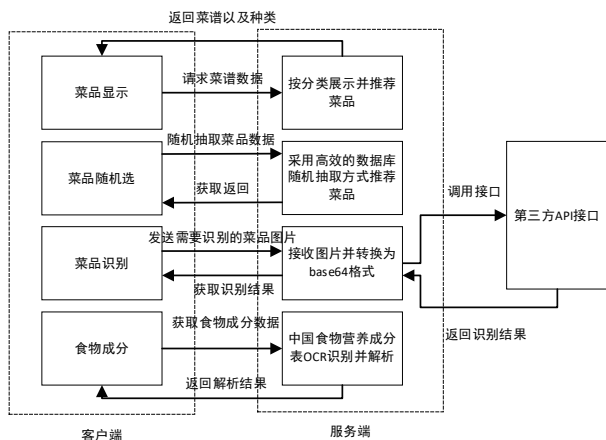


图 1 功能模块图

2 关键技术

2.1 菜品随机选

在智能健康饮食系统中，“吃什么”功能通过从数据库随机抽取菜品来解决这一难题。值得注意的是，在数据库中实现随机查询 1 条数据的方法众多，以下 3 种方法为例：

(1) `SELECT * FROM recipe ORDER BY RAND() LIMIT 1;`

当数据库数据量较小时，可以使用 `ORDER BY RAND()` 方法，这是较为简单直接的方法，但在数据量较大时，运行速度较慢。

(2) `SELECT * FROM recipe AS t1 JOIN (SELECT`

`ROUND(RAND()*(SELECT MAX(id) FROM recipe)) AS id) AS t2 WHERE t1.id>=t2.id ORDER BY t1.id LIMIT 1;`

该方法适用于数据量较大的数据库，结合使用 `JOIN` 和 `RAND()` 函数，并按 `RAND()` 排序后限制结果为 1 条。此方法通过利用主要范围来实现随机获取记录，避免了全表扫描，大幅提升了查询效率。

(3) `SELECT * FROM recipe WHERE id>= (RAND() * (SELECT MAX(id) FROM recipe)) LIMIT 1;`

当数据量超过百万条时，可以使用子函数和 `RAND()` 函数相结合的查询方法，能进一步优化查询性能。

以上 3 种方法在本系统中的查询速度如表 1 所示，总数据量为 1 550 842 条，根据查询速度比较，显然方法三的运行速度要明显优于方法一和方法二，因而本次采用的查询方法为方法三。

表 1 查询方法运行速度表

方法	运行速度 /s
方法一	30.985
方法二	0.015
方法三	0.004

2.2 菜品识别

菜品识别作为本系统的一大亮点，本文对比了不同菜品在百度 AI 以及阿里云 AI 的识别置信度，如表 2 所示，结果显示百度 AI 图像识别更适合本系统。

表 2 两种菜品识别接口置信度比较

菜品	阿里云 AI	百度 AI
清炒娃娃菜	0.51	0.68
鱼香肉丝	0.03	0.28

在使用百度 AI 的菜谱识别服务时，首先需要获取令牌来进行访问鉴权，这是保障系统安全调用服务的关键步骤。用户选择菜品图片提交后，服务端接收到图片并将信息传送给第三方服务器，第三方服务器接收图片，有两种图片输入方式。

第一种通过 URL 地址输入，这种方式要求提供远程端地址，局限性在于无法直接使用本地图片。在实际应用场景中，如果图片存储在服务器等远程位置，可采用此方式进行输入。

第二种则是将图片转换为 Base64 格式，此过程需要特别注意，图片编码时不能包含“`data:image/jpeg;base64,`”否则会导致识别错误。完成格式转换后，再将图片用 `URLEncoder.encode` 进行编码，以确保数据在网络传输中的准确性和完整性。

在数据返回方面，百度 AI 菜谱识别服务会按照特定的格式返回参数。其参数如图 2 所示。

```

{
  "result": {
    {
      "probability": "0.687498",
      "has_calorie": true,
      "calorie": "8",
      "name": "新鲜娃娃菜"
    },
    Object{...},
    Object{...},
    Object{...},
    Object{...}
  },
  "result_num": 5,
  "log_id": 1912337392048589003
}

```

图 2 返回参数信息

返回参数中包含了菜品识别的关键信息，如识别出的菜品名称、置信度（用于表示识别结果的可靠程度）以及菜品可能对应的营养信息等，参数具体说明如表 3 所示。

表 3 返回参数说明表

参数	类型	说明
log_id	uint64	唯一的 log id
result_num	unit32	返回结果数目
result	probability	置信度
	has_calorie	是否含有卡路里
	name	菜品名
	calorie	卡路里

返回数据以后，需要用索引和模糊匹配方法查找数据库中所有关于该菜品的做法以及步骤。

2.3 食物成分

智能健康饮食系统的食物成分表来源于《中国食物成分表》第六版^[9]。由于该表中的数据丰富且按照表格分类，需要识别的表格数据量较大。在这种情况下，相较于调用开放接口，利用现有的开源模型进行识别能够有效节省成本。《中国食物成分表》采用图片格式存储，为将其转换为便于系统处理的格式，需要把图片转为 JSON 数据格式。

本系统在比较几种表格识别技术以后^[10-11]，采用现有的开源工具 PaddleOCR 进行表格识别工作^[12]。其中，PPStructure 是 PaddleOCR 团队推出的一款功能强大的文档结构分析工具，专注于对扫描文档或图片中的复杂版面进行自动化解析。在食物成分表识别中，PPStructure 表格识别包含了 3 种核心算法模型^[13]。

（1）文本检测算法 -DB：负责在图片中检测出文本的位置区域，能够精准定位表格中的文字所在位置，为后续，文本识别提供基础。

（2）文本识别算法 -CRNN：对检测到的文本区域进行

字符识别，将图片中的文字内容转换为可编辑的文本信息，是实现表格数据提取的关键步骤。

（3）表格结构和 Cell 坐标预测算法 -SLANet：用于预测表格的结构以及每个单元格（Cell）的坐标位置，从而准确地解析出表格的行列结构和数据分布，使得系统能够清晰地理解表格的内容组织。

具体的识别流程如图 3 所示，首先使用 PaddleOCR 的文本检测算法 -DB 对《中国食物成分表》的图片进行处理，定位其中的文本区域；接着使用文本识别算法 - CRNN 对检测到的文本区域进行识别，将图片中的文字转换为文本；最后通过表格结构和 Cell 坐标预测算法 - SLANet 对表格的结构和单元格坐标进行预测，将识别出的文本按照表格的行列结构进行整理和排列，生成 Excel 文件，最后再将 Excel 转为 Json 格式的数据，以便系统进一步分析和利用，为用户提供准确的食物成分信息，助力科学健康饮食。

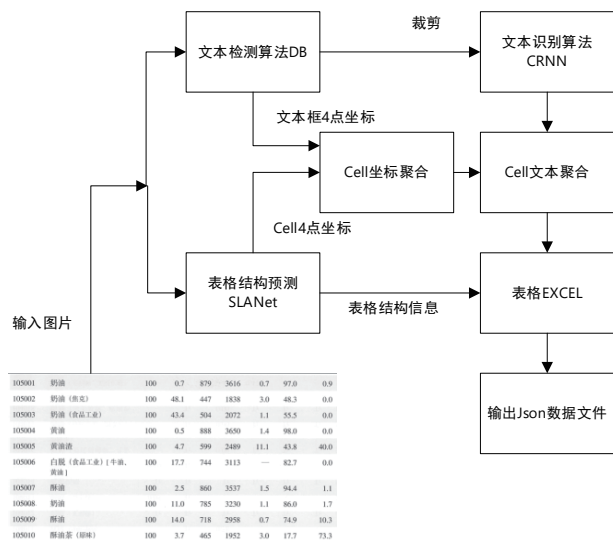


图 3 表格识别流程

2.3.1 文字分割

DB 是一种基于分割的文本检测算法^[14]，其优点是算法比较简单，同时在精度与准确度上表现良好，实现原理是设定一个自适应阈值用来分隔文字区域与非文字区域，DB 模型的函数为：

$$f(x) = \frac{1}{1 + e^{-kx}} \quad (1)$$

式中： $x = P_{ij} - T_{ij}$ ，其中 P 表示概率图， T 表示阈值图； (i, j) 表示地图中的坐标点； k 表示增益因子。

概率图是由传统的分割算法得出，而阈值图是用标准的二值化方法处理以后得出来的图。可以观察出，由于 k 增益因子的出现，使得放大效果更为显著，从而可以优化模型，产生更加精确的效果。

使用 DB 文字分割结果如图 4 所示。

105001	奶油	100	0.7	879	3616	0.7	97.0	0.9
105002	奶油 (鱼克)	100	48.1	447	1838	3.0	48.3	0.0
105003	奶油 (食品工业)	100	43.4	504	2072	1.1	55.5	0.0
105004	黄油	100	0.5	888	3650	1.4	98.0	0.0
105005	黄油渣	100	4.7	599	2489	11.1	43.8	40.0
105006	白脱 (食品工业) [牛油、黄油]	100	17.7	744	3113	—	82.7	0.0
105007	酥油	100	2.5	860	3537	1.5	94.4	1.1
105008	奶油	100	11.0	785	3230	1.1	86.0	1.7
105009	酥油	100	14.0	718	2958	0.7	74.9	10.3
105010	酥油茶 (原味)	100	3.7	465	1952	3.0	17.7	73.3

图 4 DB 文字分割结果

2.3.2 文字识别

OCR 的核心技术为卷积神经网络 CRNN^[15], CRNN 是一种结合了卷积神经网络 (CNN) 和循环神经网络 (RNN) 的深度学习模型, 其中 CNN 用于特征的提取, RNN 用于预测文字字符的位置, CRNN 主要用于序列识别任务, 特别是场景文本识别表现较为出色, CRNN 模型由 3 部分组成, 分为卷积层、递归层、转录层。其网络结构如图 5 所示。

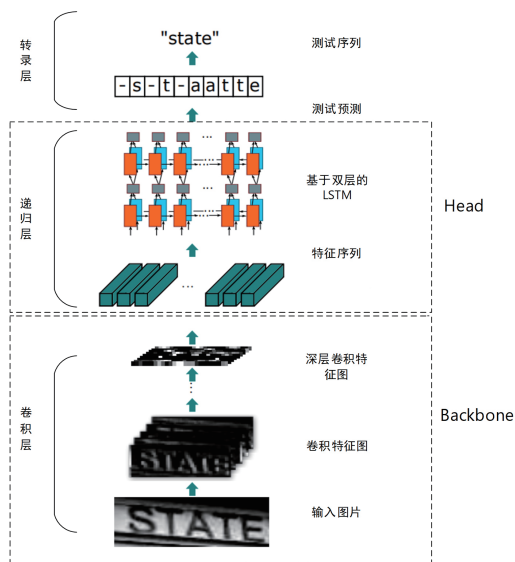


图 5 CRNN 模型网络结构图

CRNN 卷积层是由多个卷积层、池化层和批量归一化层组成, CNN 需要将输入图片缩放处理, 达到同一高度, 通常为 32 像素的文本图像, 输出特征图的每一列对应原始图像的一个水平区域, 然后再进行卷积运算。接着再通过特征序列, 提取特征值。递归层通常使用双向 LSTM (长短时记忆网络) 处理卷积层输出的特征序列, 从而捕捉序列中的上下文信息。在转录层处理上通过使用 CTC (connectionist temporal classification) 有效获取输入序列上下文的依赖关系, 从而可以精确字符位置标注^[16-17]。其测试识别结果如图 6 所示。

105001	奶油	100	0.7	879	3616	0.7	97.0	0.9	奶油	100	0.7	879	3616	0.7	97.0	0.9
105002	奶油 (鱼克)	100	48.1	447	1838	3.0	48.3	0.0	奶油 (鱼克)	100	48.1	447	1838	3.0	48.3	0.0
105003	奶油 (食品工业)	100	43.4	504	2072	1.1	55.5	0.0	奶油 (食品工业)	100	43.4	504	2072	1.1	55.5	0.0
105004	黄油	100	0.5	888	3650	1.4	98.0	0.0	黄油	100	0.5	888	3650	1.4	98.0	0.0
105005	黄油渣	100	4.7	599	2489	11.1	43.8	40.0	黄油渣	100	4.7	599	2489	11.1	43.8	40.0
105006	白脱 (食品工业) [牛油、黄油]	100	17.7	744	3113	—	82.7	0.0	白脱 (食品工业) [牛油、黄油]	100	17.7	744	3113	—	82.7	0.0
105007	酥油	100	2.5	860	3537	1.5	94.4	1.1	酥油	100	2.5	860	3537	1.5	94.4	1.1
105008	奶油	100	11.0	785	3230	1.1	86.0	1.7	奶油	100	11.0	785	3230	1.1	86.0	1.7
105009	酥油	100	14.0	718	2958	0.7	74.9	10.3	酥油	100	14.0	718	2958	0.7	74.9	10.3
105010	酥油茶 (原味)	100	3.7	465	1952	3.0	17.7	73.3	酥油茶 (原味)	100	3.7	465	1952	3.0	17.7	73.3

图 6 使用 OCR 识别表格文字图

2.3.3 表格定位

SLANet 是 PaddleOCR 于 2022 年 8 月发布的轻量级表格识别模型^[18], 将表格结构识别任务建模为序列标注任务, 通过自回归方式预测表格的 HTML 序列和单元格位置。模型由三部分组成: Backbone 采用轻量级 PP-LCNet 提取多尺度特征。Neck 使用 CSP-PAN 融合多尺度特征。Head 基于 GRU 和 Attention 机制递归生成 HTML 序列和单元格位置。其模型结构如图 7 所示。

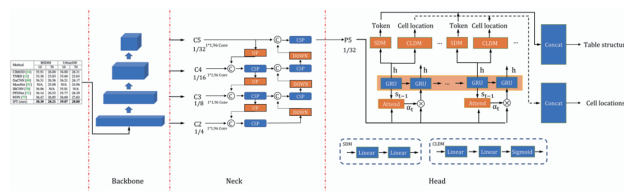
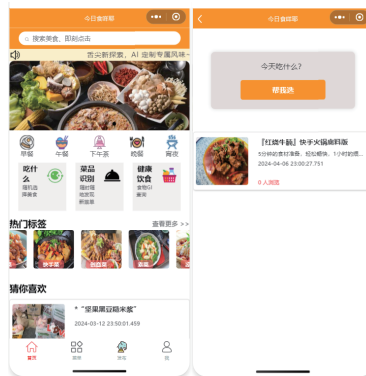


图 7 SLANet 模型结构图

表格使用 TEDS (tree-edit-distance-based similarity) 作为模型的评估指标。在食物成分表识别中 TEDS 测试值为 76.31%, 具有较高的应用价值。

3 核心功能实现

智能健康饮食系统采用微信小程序作为客户端与用户进行交互, 这一设计充分利用了微信庞大的用户基数和便捷的使用特性。微信小程序实现用户交互的具体方式如下: 用户打开微信小程序后, 可以在简洁直观的界面上进行操作。首页功能如图 8 (a) 所示, 对于“吃什么”功能, 用户点击相应按钮, 系统会从数据库中随机抽取菜品并展示在界面上, 同时提供菜品的制作方法, 如所需要的食材、制作步骤等, 方便用户做出选择, 其实现结果如图 8 (b) 所示。



(a) 首页图 (b) 吃什么

图 8 首页图与“吃什么”展示图

在菜品识别方面, 用户可以通过小程序的拍照功能拍摄菜品图片发送给服务端, 服务端发送数据给百度 AI 平台进行识别。识别结果返回后, 小程序会在界面上清晰地展示菜品名称、卡路里以及该系统下关于该菜品的制作方法, 用户可以直观地了解所拍摄菜品的详细情况, 其实现结果如图 9 (a)

所示。在制作方法上,会有食材的组成,通过食材可以查询该食材的营养成分信息,如热量、蛋白质、脂肪、维生素含量等,帮助用户科学地了解食物的营养价值,其实现结果如图9(b)所示。通过这种全方位的交互设计,使系统更加贴近用户需求,为用户提供了便捷、高效的智能健康饮食服务。



(a) 菜品识别图 (b) 食物成分图

图9 菜品识别与食物成分展示图

4 结论

本系统成功设计并实现了一个智能健康饮食系统,旨在有效解决人们在健康饮食方面面临的诸多难题。结果表明,在面对海量数据时,采用rand函数与子函数相结合方法的查询速度比传统方法的查询速度提高了30.981s,使用百度AI菜品识别的置信度在0.6874,较好地识别了菜品的内容,在食物成分表的表格识别方面,也达到了不错的精度。具备了作为系统所需求的稳定高效的算法要求。

该系统蕴含着巨大的发展潜力,功能还可以继续完善。尤其是个性化饮食定制模块、商家引入机制,以及食材智能生成菜谱功能,后续将不断迭代升级,为用户带来更为卓越、多元的服务体验,构建更为完善的饮食生态体系。

参考文献:

[1] 余毅, CANILLO A C, SEBIAL A. 一种基于 AI 视觉技术的菜品自动识别结算设备的设计 [J]. 物联网技术, 2024, 14(12): 56-58.

[2] 刘宇豪, 刘佳. 基于微信小程序的健康饮食软件设计与实现 [J]. 电脑编程技巧与维护, 2020(7):69-71.

[3] 吴思源, 曹湛慧, 黄志豪, 等. “轻享健康”微信小程序的设计与实现 [J]. 软件, 2023, 44(8):69-71.

[4] 张守涵. 基于 OCR 技术的智能招领系统研究与应用 [D]. 西安: 西安电子科技大学, 2022.

[5] 左涛. 基于 OCR 技术实现文字识别在不动产登记中的应用 [J]. 地理空间信息, 2023, 21(2):133-135.

[6] 张少宇. 基于人工智能机器学习的文字识别方法研究 [J].

电脑编程技巧与维护, 2022(9):154-156.

[7] 张牧琢, 张隽霏. 基于百度 AI 的物体图像识别软件的设计与实现 [J]. 电脑知识与技术, 2023, 19(33):11-13.

[8] 杨民峰, 孙洪迪. 基于百度 AI 开放平台的植物识别设计与实现 [J]. 北京工业职业技术学院学报, 2023, 22(4):20-24.

[9] 《中国食物成分表》标准版第 6 版第一二册出版 [J]. 营养学报, 2019, 41(5):426-426.

[10] 陈佩佩. 表格识别关键技术研究与应用 [D]. 长沙: 湖南大学, 2023.

[11] 黄玮. 基于深度学习的表格识别系统的研究与实现 [D]. 济南: 山东大学, 2023.

[12] DU Y N, LI C X, GUO R Y, et al. PP-OCR: a practical ultra lightweight OCR system[EB/OL]. (2020-10-15)[2025-05-15]. <https://doi.org/10.48550/arXiv.2009.09941>.

[13] ZHONG X, SHAFIEIBAVANI E, YEPES A J. Image-based table recognition: data, model, and evaluation[EB/OL]. (2020-04-03)[2025-01-19]. <https://doi.org/10.48550/arXiv.1911.10683>.

[14] LIAO M H, ZOU Z S, YAO C, et al. Real-time scene text detection with differentiable binarization[EB/OL]. (2019-12-03)[2025-06-26]. <https://doi.org/10.48550/arXiv.1911.08947>.

[15] SHI B G, 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, 2017, 39(11):2298 - 2304.

[16] PALIWAL S S, RAHUL R, SHARMAT M, et al. TableNet: deep learning model for end-to-end table detection and tabular data extraction from scanned document images[C]// 2019 International Conference on Document Analysis and Recognition (ICDAR). Piscataway: IEEE, 2019:128-133.

[17] NASSAR A, LIVATHINOS N, LYSAK M, et al. TableFormer: table structure understanding with transformers[EB/OL]. (2022-03-11)[2025-08-02]. <https://doi.org/10.48550/arXiv.2203.01017>.

[18] LI C X, GUO R Y, ZHOU J, et al. PP-StructureV2: a stronger document analysis system[EB/OL]. (2022-10-13)[2025-08-12]. <https://doi.org/10.48550/arXiv.2210.05391>.

【作者简介】

谢香艳(1991—), 女, 湖南永州人, 硕士研究生, 工程师, 研究方向: 软件应用与算法研究。

黄逊(1993—), 男, 湖南邵阳人, 硕士研究生, 工程师, 研究方向: 算法研究与网络安全。

(收稿日期: 2025-05-21 修回日期: 2025-10-14)