E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(2022)06-1898-20

论文引用格式:Gao L C, Li Y B, Du L, Zhang X P, Zhu Z Y, Lu N, Jin L W, Huang Y S and Tang Z. 2022. A survey on table recognition technology. Journal of Image and Graphics,27(06):1898-1917(高良才,李一博,都林,张新鹏,朱子仪,卢宁,金连文,黄永帅,汤帜. 2022. 表格识别技术研究进展. 中国图象图形学报,27(06):1898-1917)[DOI:10.11834/jig.220152]

表格识别技术研究进展

高良才[1],李一博[1],都林[2],张新鹏[1],朱子仪[1],卢宁[2],金连文[3],黄永帅[2],汤帜[1*]

1. 北京大学王选计算机研究所,北京　100871;2. 华为技术有限公司 AI 应用研究中心,北京　100085;
3. 华南理工大学电子与信息学院,广州　510640

摘　要:表格广泛存在于科技文献、财务报表、报纸杂志等各类文档中,用于紧凑地存储和展现数据,蕴含着大量有用信息。表格识别是表格信息再利用的基础,具有重要的应用价值,也一直是模式识别领域的研究热点之一。随着深度学习的发展,针对表格识别的新研究和新方法纷纷涌现。然而,由于表格应用场景广泛、样式众多、图像质量参差不齐等因素,表格识别领域仍然存在着大量问题亟需解决。为了更好地总结前人工作,为后续研究提供支持,本文围绕表格区域检测、结构识别和内容识别等 3 个表格识别子任务,从传统方法、深度学习方法等方面,综述该领域国内外的发展历史和最新进展。梳理了表格识别相关数据集及评测标准,并基于主流数据集和标准,分别对表格区域检测、结构识别、表格信息抽取的典型方法进行了性能比较。然后,对比分析了国内相对于国外,在表格识别方面的研究进展与水平。最后,结合表格识别领域目前面临的主要困难与挑战,对未来的研究趋势和技术发展目标进行了展望。

关键词:表格区域检测;表格结构识别;表格内容识别;深度学习;单元格识别;表格信息抽取

A survey on table recognition technology

Gao Liangcai[1], Li Yibo[1], Du Lin[2], Zhang Xinpeng[1], Zhu Ziyi[1], Lu Ning[2],
Jin Lianwen[3], Huang Yongshuai[2], Tang Zhi[1*]

1. Wangxuan Computer Institute, Peking University, Beijing　100871, China; 2. Huawei AI Application Research Center, Huawei Technology Co., Ltd., Beijing　100085, China; 3. School of Electronics and Information Engineering, South China University of Technology, Guangzhou　510640, China

Abstract: Optimal data access and massive data derived information extraction has become an essential technology nowadays. Table-related paradigm is a kind of efficient structure for the clustered data designation, display and analysis. It has been widely used on Internet and vertical fields due to its simplicity and intuitiveness. Computer based tables, pictures or portable document format(PDF) files as the carrier will cause structural information loss. It is challenged to trace the original tables back. Inefficient manual based input has more errors. Therefore, two decadal researches have focused on the computer automatic recognition of tables issues originated from documents or PDF files and multiple tasks loop. To obtain the table structure and content and extract specific information, table recognition aims to detect the table via the image or PDF and other electronic files automatically. It is composed of three tasks recognition types like table area detection, table structure recognition and table content recognition. There are two types of existed table recognition methods in common.

收稿日期:2022-02-26;**修回日期**:2022-03-23;**预印本日期**:2022-03-30

*通信作者:汤帜　tangzhi@pku.edu.cn

基金项目:国家重点研发计划资助(2019YFB1406303)

Supported by: National Key R&D Program of China (2019YFB1406303)

One is based on optical character recognition (OCR) technology to recognize the characters in the table directly, and then analyze and identify the position of the characters. The other one is to obtain the key intersections and the positions of each frameline of the table through digital image processing to analyze the relationship between cells in the table. However, most of these methods are only applicable to a single field and have poor generalization ability. At the same time, it is constrained of some experience-based threshold design. Thanks to the development of deep learning technology, semantic segmentation algorithm, object detection algorithm, text sequence generation algorithm, pre training model and related technologies facilitates technical problem solving for table recognition. Most deep learning algorithms have carried out adaptive transformation according to the characteristics of tables, which can improve the effect of table recognition. It uses object detection algorithm for table detection task. Object detection and text sequence generation algorithms are mainly used for table structure recognition. Most pre training models have played a good effect on the aspect of table content recognition. But many table structure recognition algorithms still cannot handle these well for wireless tables and less line tables. On the aspects of table images of natural scenes, the relevant algorithms have challenged to achieve the annotation in practice due to the influence of brightness and inclination. A large number of datasets provide sufficient data support for the training of table recognition model and improve the effect of the model currently. However, there are some challenging issues between these datasets multiple annotation formats and different evaluation indicators. Some datasets provide the hyper text markup language (HTML) code of the structure only in the field of table structure recognition and some datasets provide the location of cells in the table and the corresponding row and column attributes. Some datasets are based on the position of cells or the content of cells in accordance with evaluation indicators. Some datasets are based on the adjacent relationship between cells or the editing distance between HTML codes for the recognition of table structure. Our research critically reviews the research situation of three sub tasks like table detection, structure recognition and content recognition and try to predict future research direction further.

Key words: table area detection; table structure recognition; table content recognition; deep learning; table cell recognition; table information extraction

0 引　言

在大数据时代,高效地存取数据,以及从海量数据中提取有效信息是各行各业都亟需利用的重要技术。表格作为数据的一种重要载体,具有信息精炼集中、方便体现数据关系等特点,已经在各个行业得到了广泛应用。在教育领域中,表格常常会出现在各类试卷、题目中;在金融领域,表格用来展示和分析数据;在科学领域,表格用来记录各类实验配置以及结果;在现实生活中也常常在幻灯片、车站时刻牌上看到表格。因此对表格进行区域检测、结构识别乃至对其中信息进行识别理解都有着广阔的应用前景。

表格在生成或存储过程中往往以图片或 PDF (portable document format)文件的形式存在,会丢失易于计算机理解的原有结构信息。若是采用人工手段对表格进行重新处理录入,会面临效率低下、数据量大导致出错等问题。因此,如何让计算机从文档或图像中自动识别表格、提取信息,成为文档识别领域一个重要的研究问题。

早期对于表格的识别大多是针对较为简单或模板化的表格。从表格的布局结构出发,抽取表格线条或抽取文本块,然后使用规则方法进行分析,但这些方法往往泛化能力较差,且难以处理复杂表格。随着深度学习的发展,无论是机器视觉方面还是自然语言处理方面都获得了巨大的进展,各种表格识别的方案相继提出,并有研究者开始尝试对自然场景下的表格进行处理。

本文将围绕表格的区域检测、表格结构识别和表格内容识别 3 个表格识别子任务,从传统方法、深度学习方法等方面,综述该领域国内国外的发展历史和最新进展,同时对国内国外的研究进行对比,对未来的趋势和技术发展目标进行展望。

1 国内外研究现状

1.1 表格识别相关数据集及评测标准

针对表格识别的不同子任务、表格格式、数据量和文档类型等,本文对该领域的相关数据集总结如表 1 所示。

表 1　表格识别数据集

Table 1　The collection of table recognition datasets

数据集	检测	结构识别	内容识别	内容	数量/幅	格式
UW-Ⅲ(University of Washington English/technical document image database Ⅲ)(Chen 等,1996)	√	×	×	文档页面图像	120	图片
UNLV(Shahab 等,2010)	√	√	×	扫描页面图像	427	图片
Marmot(Fang 等,2012b)	√	×	×	文档页面图像	2 000	图片
DeepFigures(Siegel 等,2018)	√	×	×	生成文档图像	5.5 M	图片
ICDAR2013(Göbel 等,2013)	√	√	√	电子文档	156	PDF
ICDAR2019(Gao 等,2019)	√	√	×	文档页面图像	2 539*	图片
FUNSD(form understanding in noisy scanned documents)(Jaume 等,2019)	√	√	√	扫描页面图像	199	图片
Tablebank(Li 等,2020)	√	√	×	文档页面图像	417 k**	图片
SciTSR(Chi 等,2019)	×	√	×	表格图像、PDF	15k	PDF/图片
TNCR(table net detection and classification dataset)(Abdallah 等,2022)	√	×	×	表格图像	6 621	图片
WTW(wired table in the wild)(Long 等,2021)	√	√	×	文档页面图像	14 581	图片
PubTabNet(Zhong 等,2020)	×	√	√	文档页面图像	568 k +	图片
DECO(dresden enron corpus)(Koci 等,2019)	×	√	√	电子表格	1 165	图片
SROIE(scanned receipts OCR and key Information Extraction)(Huang 等,2019b)	×	√	√	表格图像	1 000	图片
TabLex(Desai 等,2021)	×	√	√	表格图像	1 M +	图片
ICDAR2017-POD(Gao 等,2017)	√	×	×	文档页面图像	2 417	图片
Publaynet(Zhong 等,2019)	√	×	×	文档页面图像	113 k	PDF
NTable(Zhu 等,2021)	√	×	×	表格图像	39 k	图片
TABLE2LATEX(Deng 等,2019)	×	√	√	表格图像	450 k	图片
FinTab(Li 等,2021c)	√	√	√	金融文档页面图像	110 k +	PDF
PubTables-1M(Smock 等,2021)	√	√	√	文档页面图像	948 k	PDF
TableGraph-350K(Xue 等,2021)	×	√	√	表格图像	350 k	图片
Tal ocr_table(好未来表格识别技术挑战赛)	√	√	√	文档页面图像	16 k	图片

注:* ICDAR2019 数据集(cTDaR)分成现代表格和历史表格,两者都能进行表格检测任务,但是只有历史表格(训练 600 + 测试 150)包含结构信息的标注;** Tablebank 数据集有两部分,其中有表格检测标注的数据是 417 k,表格结构识别标注的数据是 145 k;×表示该数据集不涉及该任务,√表示该数据集涉及该任务。

表格区域检测目前通常采用给定 IoU(intersection over union)的 F1 进行评测,IoU 表示的是预测框和真实框的交并比。对于图像中的表格,会选择 IoU 值超过阈值且具有最大 IoU 值的预测框作为正确预测。据此可以计算出正确预测、错误预测和未被召回的表格的数量,从而计算召回率和准确率,得到 F1 值。

表格结构识别的评测标准从早期到现在出现了多种形式,分别有单元格对的 F1 值、行列的预测准确性、序列化标注出现之后的 BLEU(bilingual evaulation understudy)和 TEDS(tree edition distance similiarity)等。单元格对的 F1 值的评测标准首先在 ICDAR2013(International Conference on Document Analysis and Recognition)比赛中提出,这种方法将在结构上处于同一行或同一列的单元格组成一个单元格对,从而将表格分解成多个单元格对,之后计算

这些单元格对的准确率、召回率和 F1 值。ICDAR2019 比赛采取了相类似的方法，但是使用了 IoU 来确认单元格是否被检测到，将超过阈值的单元格组成单元格对计算 F1 值。行列预测准确性的评测标准由 Shahab 等人(2010)提出，其将检测的结果分为正确检测、部分检测、过分割、分割不完全、丢失以及错误检测等 6 类来评估检测的效果。Li 等人(2019)在使用序列标注表格结构的同时借鉴了自然语言处理中的 BLEU 来评测表格结构识别的效果。Zhong 等人(2020)认为基于单元格对的评测标准无法评估由于空白单元格和非直接邻接的单元格未对齐对表格识别结果的影响，同时单元格对的评测标准是精准匹配，因此无法衡量每个单元格的识别效果。据此，其提出了 TEDS(树编辑距离相似度)，将表格的 HTML 代码看成一棵树，HTML 代码的每个标签即为树中的节点，计算树之间的编辑距离和树长度的比值作为错误的比例。即

$$TEDS(T_a,T_b) = 1 - \frac{Edit(T_a,T_b)}{\max(|T_a|,|T_b|)} \quad (1)$$

式中，T_a 代表预测的 HTML 代码，T_b 代表真实的 HTML 代码，$Edit(T_a,T_b)$ 代表两种代码序列的标记距离，$|T_*|$ 代表的是代码的长度。

Smock 等人(2021)提出了 GriTS(grid table similarity)，GriTS 将表格的拓扑结构表示为 2 维网格或矩阵，并分为单元格内容相似度、单元格位置相似度和单元格拓扑结构相似度 3 类来计算。对于单元格内容相似度，使用最长子串来计算；对于单元格位置相似度，使用 IoU 来计算；对于单元格的拓扑结构相似度，则使用跨行跨列来计算开始行开始列，并使用类似 IoU 的方式来计算。

1.2 表格区域检测相关研究

表格区域检测指的是从页面中框出对应的表格区域位置。在早期的研究中，检测目标多集中于扫描文档图片和 PDF 文档。随着图像采集技术水平的提升，以及表格应用领域的扩展，还出现了自然场景表格的检测任务。

1.2.1 传统的表格区域检测方法

国外的表格区域检测研究起步较早，这些早期方法大多数基于启发式规则或者简单的机器学习算法，依赖于图像预处理和文档分析获得的线条、文本块等视觉信息，或者依赖于 PDF 编码中自带的一些文字信息。

Watanabe 等人(1993a)、Hirayama(1995)通过对扫描文档进行图像处理，获取文档中的文本块以及水平线和垂直线来定位表格。Ramel 等人(2003)只尝试寻找表格区域顶部的第 1 条水平线，该表格的其他区域则通过匹配 9 种框线相交情况中的 4 种“T”字形模板来寻找；Watanabe 等人(1993b)使用水平垂直线等特征的同时，在具体的检测策略上更注重用单元格的左上角作为基准点来确定表格位置(Watanabe 和 Luo，1996)；Wang 等人(2001)提出在初步定位表格时不用表格本身的特征，而是利用表格上下在水平方向上贯穿文档的空白区域得到待定表格区域，再计算该区域内的空白比例、单元格坐标差异等信息进行二次确认；Kieninger 和 Dengel(2001)认为空白、框线等都不是表格必须具备的特征，而表格中的文本区域具有和其他普通文本区域不一样的特性——不同列上的文本区域在 x 轴上投影基本不相交，并以此检测表格区域。

国内的表格区域检测研究起步较晚，启发式方法较少。其中，具有代表性的是 Fang 等人(2011)提出的基于表格结构特征和视觉分隔符的方法。该方法以 PDF 文档为输入，分 4 步进行表格检测：PDF 解析、页面布局分析、线条检测和页面分隔符检测以及表格检测。在最后的表格检测部分中，通过对上一步检测出的线条和页面分隔符进行分析得到表格位置。

1.2.2 基于深度学习的表格区域检测方法

随着计算机硬件水平的提高，深度学习在计算机视觉的语义分割和目标检测等任务上取得了优异表现。作为语义分割或目标检测领域的一个具体应用，国际上提出了诸多方法来解决表格区域检测问题。一些具有代表性的方法在 ICDAR2013 和 ICDAR2017 竞赛上的结果如表 2 和表 3 所示。

Schreiber 等人(2017)采用了 Faster R-CNN(region convolutional neural network)(Ren 等，2015)作为表格检测的模型网络，来获取每个表格的区域。Gilani 等人(2017)在采用相同的目标检测网络的同时，还使用了 3 种距离变换来增强页面图像特征。He 等人(2017a)将表格检测作为文档分割的子任务，使用 FCN(fully convolutional networks)(Long 等，2015)作为基础模型，考虑了多尺度特征，同时进行表格、段落以及图像的边缘检测和分类，最后通过连通体分析、条件随机场等获得表格区域。Kavasidis

表 2　ICDAR2013 表格检测结果比较

Table 2　Comparison results of table detection on ICDAR2013

方法	F1/%	备注
Schreiber 等人(2017)(deep learning for detection and structure recognition of tables, DeepDeSRT)	96.77	
Kavasidis 等人(2019)	98.1	IoU = 0.5
Siddiqui 等人(2018)(deep deformable CNN for table detection, DeCNT)	99.6	IoU = 0.5
Melinda 和 Bhagvati(2019)	92.64	
Huang 等人(2019a)	86.8	
Paliwal 等人(2019)(TableNet)	96.62	在 Marmot 训练
Zheng 等人(2020)(global table extractor, GTE)	99.31	
Prasad 等人(2020)(CascadeTabNet)	100.0	

表 3　ICDAR2017 表格检测结果比较

Table 3　Comparison results of table detection on ICDAR2017

方法	F1/%
Siddiqui 等人(2018)(DeCNT)	95.2
Saha 等人(2019)(graphical object detection, GOD)	96.8
Huang 等人(2019a)	97.1
Sun 等人(2019)	94.9
Li 等人(2019)	90.3

等人(2019)同样使用了一个典型的语义分割架构,使用 VGG(Visual Geometry Group)(Simonyan 和 Zisserman,2014)作为骨干网络,同时使用了空洞卷积(Yu 和 Koltun,2015)来扩大感受野,之后再通过上采样和反卷积将特征放缩为原图尺寸,以获得每个像素的分类。使用条件随机场来平滑表格边缘,得到更加准确的候选区域,并对每个区域使用 Inception(Szegedy 等,2015)网络来进行最终的表格分类。Siddiqui 等人(2018)提出的 DeCNT(deep deformable CNN for table detection)网络将形变卷积(Dai 等,2017)应用在目标检测网络中,使用了 ResNet-101(He 等,2016)作为特征提取网络,使用了特征金字塔(Lin 等,2017)来抽取更全面的特征。Saha 等人(2019)将表格检测作为文档检测中图形类目标检测的子任务,尝试了 Faster R-CNN 和 Mask R-CNN(He 等,2017b)网络,并证明了预训练模型在表格检测中的效果。Riba 等人(2019)将图神经网络(graph neural network, GNN)(Scarselli 等,2009)应用到了表格检测中,他们先检测出文档的文本区域和图像区域,以这些区域为顶点构建一个图,然后送入图网络进行特征交互,对点和边进行分类,判断每个区域是否属于表格,以及相邻的两个区域是否需要合并,从而获得最终的表格区域。Melinda 和 Bhagvati(2019)将表格分为封闭表格和开放表格。其中封闭表包含表格线条,可以直接得到表格区域。对于开放表则通过使用混合高斯模型和 EM(expectation maximization)算法对所有文本块进行分类,判断其是否属于表格区域,然后将属于表格区域的单元格进行合并得到表格的区域。Zheng 等人(2020)将单元格检测和表格检测放在同一个检测网络中,使用单元格的位置来调整表格检测的结果。此外,还有一些同时对表格进行检测和结构识别的研究,将在表格结构识别算法中进行介绍。

近年来,国内也涌现出了许多基于深度学习的表格区域检测算法。Huang 等人(2019a)对 Yolov3(you only look once)网络的锚进行了适应性调整,同时在后处理时去除了检测框的空白区域,过滤掉了噪声对象,使得检测的表格更加准确。Sun 等人(2019)提出,基于锚的表格检测方法比较依赖于锚的设置,而锚的设置很难包含所有情况,因此借鉴 CornerNet(Law 和 Deng,2018)的思想,在检测表格的同时回归表格的 4 个角的点的位置,最后再用 4 个点来矫正表格检测的结果,提高了检测的精度。Li 等人(2019)则关注了少线表和无线表,使用对抗生成网络来使生成器重点抽取到表格的布局特征,并将此特征和检测网络的骨干网络抽取的特征进行融合,在无线表检测上取得了更好的效果。Zhang 等人(2021)提出了 VSR(vision, semantics and rela-

tions)网络,融合了文档的视觉和语义信息。文档以图像(视觉)和文本嵌入映射(字符级和句子级语义)的形式输入 VSR。然后,通过一个双流网络提取对应模态的视觉和语义特征,这些特征随后被有效地组合到一个多尺度自适应聚合模块中。最后,结合基于 GNN 的关系模块,对候选组件之间的关系进行建模,并生成最终结果。

1.3 表格结构识别相关研究

表格结构识别是表格区域检测之后的任务,其目标是识别出表格的布局结构、层次结构等,将表格视觉信息转换成可重建表格的结构描述信息。这些表格结构描述信息包括:单元格的具体位置、单元格之间的关系和单元格的行列位置等。在当前的研究中,表格结构信息主要包括以下两类描述形式:1)单元格的列表(包含每个单元格的位置、单元格的行列信息和单元格的内容);2)HTML 代码或 Latex 代码(包含单元格的位置信息,有些也会包含单元格的内容)。

1.3.1 传统的表格结构识别方法

与表格区域检测任务类似,在早期的表格结构识别方法中,研究者们通常会根据数据集特点,设计启发式算法或者使用机器学习方法来完成表格结构识别任务。

Itonori(1993)根据表格中单元格的 2 维布局的规律性,使用连通体分析抽取其中的文本块,然后对每个文本块进行扩展对齐形成单元格,从而得到每个单元格的物理坐标和行列位置。Rahgozar 等人(1994)则根据行列来进行表格结构的识别,其先识别出图片中的文本块,然后按照文本块的位置以及两个单元格中间的空白区域做行的聚类和列的聚类,之后通过行和列的交叉得到每个单元格的位置和表格的结构。Hirayama(1995)则从表格线出发,通过平行、垂直等几何分析得到表格的行和列,并使用动态规划匹配的方法对各个内容块进行逻辑关系识别,来恢复表格的结构。Zuyev(1997)使用视觉特征进行表格的识别,使用行线和列线以及空白区域进行单元格分割。该算法已经应用到 FineReader OCR 产品之中。Kieninger(1998)提出了 T-Recs(table recognition system)系统,以词语区域的框作为输入,并通过聚类和列分解等启发式方法,输出各个文本框对应的信息,恢复表格的结构。随后,其又在此基础上提出了 T-Recs ++ 系统(Kieninger 和 Dengel,2001),进一步提升了识别效果。Amano 等人(2001)创新性地引入了文本的语义信息,首先将文档分解为一组框,并将它们半自动地分为 4 种类型:空白、插入、指示和解释。然后根据文档结构语法中定义的语义和几何知识,分析表示框与其关联条目之间的框关系。Wang 等人(2004)将表格结构定义为一棵树,提出了一种基于优化方法设计的表结构理解算法。该算法通过对训练集中的几何分布进行学习来优化参数,得到表格的结构。同样使用树结构定义表格结构的还有 Ishitani 等人(2005),其使用了 DOM(document object model)树来表示表格,从表格的输入图像中提取单元格特征。然后对每个单元格进行分类,识别出不规则的表格,并对其进行修改以形成规则的单元格排布。Hassan 和 Baumgartner(2007)、Shigarov 等人(2016)则以 PDF 文档为表格识别的载体,从 PDF 文档中反解出表格视觉信息。后者还提出了一种可配置的启发式方法框架。

国内的表格结构识别研究起步较晚,因此传统的启发式方法和机器学习方法较少。在早期,Liu 等人(1995)提出了表格框线模板方法,使用表格的框架线构成框架模板,可以从拓扑上或几何上反映表格的结构。然后提出相应的项遍历算法来定位和标记表格中的项。之后 Li 等人(2012)使用 OCR(optical character recognition)引擎抽取表单中的文本内容和文本位置,使用关键词来定位表头,然后将表头信息和表的投影信息结合起来,得到列分隔符和行分隔符,从而得到表格结构。

总体来说,表格结构识别的传统方法可以归纳为以下 4 种:基于行和列的分割与后处理,基于文本的检测、扩展与后处理,基于文本块的分类和后处理,以及几类方法的融合。

1.3.2 基于深度学习的表格结构识别方法

在传统的表格结构识别算法基础之上,基于深度学习的表格结构识别算法可以分为:自底向上的方法、自顶向下的方法和图像文本生成的方法。其中,自底向上的方法主要特点是先进行表格单元格和文本块的检测,再进行单元格关系的分类;自顶向下的方法则先进行表格行列的分割,之后对单元格进行合并等操作;图像文本生成方法是指基于表格图像直接生成表格结构所对应的序列文本(HTML、Latex 等)。

针对近年来的一些具有代表性的方法及代表性数据集(ICDAR2013, PubTabNet),其效果总结如表4和表5所示。由于此类方法所采用的评测标准各有不同,因此在备注一栏进行具体阐述。

表4 ICDAR2013表格结构识别结果比较

Table 4 Comparison results of table structure recognition on ICDAR2013

方法	评测标准	结果/%	备注
Shigarov等人(2016)	F1	93.64	
Schreiber等人(2017)(DeepDeSRT)	F1	91.44	
Siddiqui等人(2019)(DeepTabStR, deep learning based table structure recognition)	F1	92.98	
Tensmeyer等人(2019)	F1	95.26	
Paliwal等人(2019)(TableNet)	F1	91.51	在Marmot上训练
Khan等人(2019)	F1	93.39	
Chi等人(2019)(GraphTSR)	F1	87.2	
Xue等人(2019)(Res2TIM)	F1	74.04	
Zheng等人(2020)(GTE)	F1	93.5	
Raja等人(2020)(tabstruct-net)	F1	90.6	在SciTSR上训练
Qiao等人(2021)(LGPMA, local and global pyramid mask alignment)	F1	97.9	在SciTSR上训练
Long等人(2021)	F1	98.0	
Li等人(2021b)	行/列分割准确率	57.78/90.63	

表5 PubTabNet表格结构识别结果比较

Table 5 Comparison results of table structure recognition on PubTabNet

方法	结果/%	评测标准
Zhong等人(2020)(EDD, encoder-dual-decoder)	88.3	TEDS
Raja等人(2020)(tabstruct-net)	90.1	TEDS, 在SciTSR上训练
Qiao等人(2021)(LGPMA)	94.6	TEDS
Li等人(2021b)	95.32/97.39	行/列分割准确率
He等人(2021)(TableMaster)	96.84	TEDS

自底向上的基于单元格检测和单元格关系分类的深度学习算法的基本框架如图1所示(Qasim等,2019),图中前半部分为单元格检测阶段,后半部分为单元格关系判断阶段。

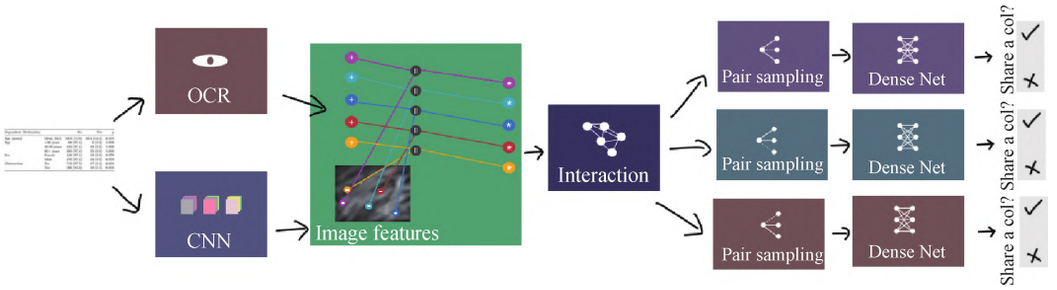

图1 自底向上的表格结构识别深度学习算法框架(Qasim等,2019)

Fig. 1 The framework of bottom-up algorithm for table structure recognition(Qasim et al., 2019)

Prasad 等人(2020)主要在前半部分的单元格检测阶段进行研究,提出了 CascadeTabNet,一种基于级联掩膜区域的 CNN 高分辨率网络,同时检测单元格和表格。在检测表格位置的同时,将表格分类为有线的表格和无线的表格。对于有线的表格直接使用常规的行列检测算法,并使用行列交点来确定单元格;对于无线的表格则使用检测到的单元格来预估缺失的线,进而恢复表格结构。Siddiqui 等人(2019)提出的 DeepTabStR 网络将可变型卷积应用于目标检测网络中,同时对行、列和单元格进行检测,并根据单元格的位置特点恢复表格。还有一些研究是专注于后半部分的表格关系判断阶段,即给出单元格或文本区域,使用深度网络模型来判断单元格之间的关系。Clinchant 等人(2018)在历史文档的表格识别中尝试了条件随机场和图卷积网络的作用。Qasim 等人(2019)提出使用图网络来解决单元格之间的关系判断问题,首先使用 OCR 引擎获取图片中文本的位置和内容,之后使用卷积神经网络获取单元格的视觉特征,并以单元格位置作为位置特征,以文本的长度作为文本特征,3 种特征相融合为每个文本块的特征。随后将这些文本块作为顶点构建全连接的无向图,并进行图卷积,卷积得到的特征送入 DenseNet,然后判断两个文本块是否处于同一行或同一列,以及是否需要合并,最后通过启发式方法获得表格结构。另外,在训练中使用基于蒙特卡洛的采样方法,解决正负样本不均衡和单元格对内存占用过大的问题。

自顶向下的行列分割和单元格合并的基本流程如图 2 所示(Tensmeyer 等,2019)。其基本思路是先检测单元格的行和列分隔符,将表格划分为最基本的单元,然后再使用规则类方法或深度学习方法将这些基本单元进行合并,以避开难度较大的单元格检测环节。最早期的行列分割方法忽略了单元格的跨行跨列问题,直接进行行和列的检测,而不进行后续的行列合并等操作。

图 2　自顶向下的表格结构识别深度学习方法流程(Tensmeyer 等,2019)

Fig. 2　The framework of top-down algorithm for table structure recognition(Tensmeyer et al. , 2019)

Siddiqui 等人(2019)将表格识别定义为一个语义分割问题,并使用了类似于编码器—解码器的架构, 编码阶段通过卷积和池化来获取表格特征,解码阶段则通过反卷积和上采样还原出和原图相同大小的特征图,并对每个像素进行分类,再通过后处理获得表格结构识别结果。Schreiber 等人(2017)在其提出的 DeepDeSRT 系统中,以 FCN(Long 等,2015)为基础架构,进行行和列的语义分割。此外,由于行与行之间的间隔相对较小,在进行行检测时,此方法还会对图片的高度进行拉伸。Paliwal 等人(2019)提出了 TableNet, 同样使用语义分割框架,将表格检测和结构识别放在一个框架下进行处理,同时进行表格检测和行列检测。此方法针对表格检测和行列检测的不同,分别提取骨干网络中不同尺度的特征进行融合。之后又制定启发式规则对表格的行进行分割,得到表格的结构。Khan 等人(2019)则认为,卷积网络受限于感受野无法获取更广的特征,同时忽略了行列(行—空白或线—行)的排布规律,会降低行列检测的准确率。因此使用了两个双向的循环神经网络进行像素级别的行列分隔符的识别。Tensmeyer 等人(2019)对表格的行列分割和分割后的合并都进行了详细的讨论,提出了一个合并网络,将表格分割为最细粒度的基本单元,然后进行合并得到真正的表格结构。

Raja 等人(2020)把自顶向下和自底向上的处理流程进行了融合,一方面使用检测网络来检测单元格,另一方面对检测出来的单元格进行特征抽取,对文本块对进行同行和同列的判断,从而获得表格的完整结构。

得益于 Table2Latex(Deng 等,2019)、TableBank(Li 等,2019)等给定 HTML 或 Latex 代码的表格数据集,图片文本生成的方法逐渐兴起。其基本架构如图 3 所示。

Deng 等人(2019)使用了经典的 IM2LATEX 模型(Deng 等,2017), 此方法使用 CNN 抽取特征,并使用带有注意力机制的长短期记忆网络(long short

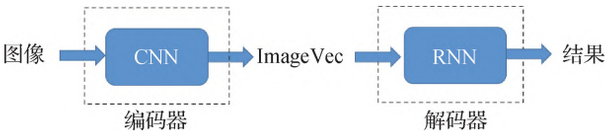

图 3　基于图片文本生成的表格结构识别方法框架

Fig. 3　The framework of image to text algorithm for table structure recognition

term memory, LSTM)(Hochreiter 和 Schmidhuber, 1997)来生成对应的 Latex 代码。Zhong 等人(2020)提出的 PubTabNet 数据集不仅提供了表格结构的 HTML 代码,同时也提供了每个单元格的文本内容。因此,他们提出了一种编码器—双解码器的模型 EDD,在编码阶段单独使用了一个卷积神经网络,而在解码阶段则使用两个循环神经网络,其中一个负责解码出表格结构标签,另一个负责解码出具体的文本结果。值得注意的是,该网络设置只有结构解码器解码出“〈td〉”标签时,文本解码器才会被激活。在训练时,需要先对结构解码器单独进行训练,之后再将两个解码器联合训练。

随着深度学习的发展,以及工业界对表格识别需求的日益增长,国内的表格结构识别研究迅速发展,并产生了一批有影响力的研究成果。

在自底向上的表格结构识别研究中,Chi 等人(2019)提出了 GraphTSR 模型,通过对 PDF 文档的解析,得到单元格内容以及相应的边界框。将每个单元格视做一个顶点,构建出全连接图,并根据单元格大小、位置设计了相应的特征。对于每个边,通过点与点的距离计算权重,得到一个完整的图。之后使用基于注意力机制的图网络来对每条边进行分类,判断 K 邻近的单元格对是否在同一行或同一列。Xue 等人(2019)提出了 Res2TIM 系统,在使用检测网络获得各个文本区域后,将区域原图特征和经过卷积网络的特征相融合,并构建单元格对来判断两个单元格的上下左右关系,最终达到重建表格的目的。Qiao 等人(2021)则将重心放在单元格检测上,提出了 LGPMA 网络。该网络从局部和全局角度考虑了视觉特征,充分利用了局部和全局特征的信息,通过提出的掩码重评分策略,获得更可靠的对齐单元格区域,并使用软标签的方式,巧妙解决了空白单元格对检测模型的干扰。Li 等人(2021a)使用多任务的语义分割网络同时进行前景单元格和背景的表格线分割。为了消除表格尺度不一致的影响,设计了一种基于每个文档图像中平均单元格大小和划线密度的自适应图像缩放方法。Long 等人(2021)在提出一个自然场景的表格识别数据集 WTW 的同时,还提出了 Cycle-CenterNet 的表格结构识别方法。他们认为过去的文档表格识别针对的都是非常规整的表格图片,而在自然场景中由于表格存在扭曲,行和列之间没有非常完备的对齐关系。他们以 CenterNet(Duan 等,2019)为基础,同时检测单元格的中心以及 4 个单元格的交汇点,这样在单元格检测完成之后就可以直接对表格结构进行恢复。

在自顶向下的行列分割方法中,Li 等人(2021b)考虑到表格行和列的分类结果遵从“行—分隔符—行—分隔符—行”的规律,将行列检测视为一个逐像素的序列标注问题。先用卷积神经网络获取图像特征并视做一个行或者列的序列,随后使用序列标注网络对行和列的每个像素进行分类,得到行和列的检测效果,从而识别表格。

在图片到文本序列的方法中,He 等人(2021)提出的 TableMaster 模型以文字识别模型 Master(Lu 等,2021)为基础,先对表格 HTML 代码进行划分,之后在解码器部分增加了一个单元格检测分支,使得单元格检测和 HTML 代码的生成一一对应,同步进行。同时为了解决此模型单元格检测效果相对较差的问题,该方法又使用了 PSENet(Wang 等,2019b)对文本块进行检测,对 TableMaster 的单元格检测进行矫正。

在实际场景应用中,表格结构识别的流程比以上的研究领域复杂,需要同时进行表格检测和结构识别,还需要对每个单元格的文本进行识别和信息抽取。为了提高最终效果,会采用多模型的融合,其对表格识别的研究也有重要的借鉴意义。

好未来于 2021 年 6 月—9 月举办了一个表格识别技术挑战赛。本次比赛提供了 2 万幅包含表格的图像。这些图像来源于教育场景下学生的作业、试卷以及部分的扫描合同表。其中 16 000 幅图像作为训练集,提供了详细的 HTML 代码、单元格框以及单元格内容标注;2 000 幅含有内容标注的图像作为验证集,完全没有标注的 2 000 幅作为测试集。比赛以 TEDS 作为评测指标对各个队伍的结果进行打分。

总体而言,前 3 名的技术方案基本思路相同,都

包含表格检测、表格结构识别、文本识别和 HTML 代码恢复等阶段，但在各个阶段采用的模型存在差异。第 1 名的方法采用了 Cascade R-CNN（Cai 和 Vasconcelos，2018）对表格进行检测，并使用了后处理以提高表格检测的准确率；随后其使用 TableMaster 模型（He 等，2021）来预测 HTML 序列和单元格区域；之后使用 BDN（barcode detection network）（Jia 等，2020）来检测文本行，并使用 CRNN（convalutional recurrent neural network）和 CTC（connectionist temporal classification）（Shi 等，2017）来进行文本行的识别，最后将所有结果合并起来得到最终结果。第 2 名采用了 CDeC-Net（Agarwal 等，2020）来进行表格和单元格的检测；考虑到自然场景下表格会出现的扭曲、褶皱问题，在表格检测结束之后，采用了 TPS（thin plate spline）变换和仿射变换来对图像进行矫正，矫正结果保证了每个单元格行和列的对齐，之后根据坐标来直接还原出表格结构，最后采用 DBNet（Liao 等，2020）检测文本行，并使用 CRNN 和 CTC 的方法来识别文字。第 3 名在表格结构检测上使用多任务的方式，同时分割单元格和表格线并检测单元格；之后在文本识别过程中对任务进行了细化，对单元格内容进行分类，判断手写单元格、空单元格以及插图单元格。

1.4 表格内容识别相关研究

表格内容识别的研究包含两个方面，一方面是对于单元格内的文本进行识别，一般在获得单元格区域之后，使用较为鲁棒的光学字符识别方法（OCR）进行解决，这方面不属于表格识别的研究范畴，不做详细介绍；另一方面是根据整个表格内容进行的表格分类、单元格分类以及表格信息抽取等任务，这是当前表格识别研究的热点之一。

1.4.1 表格分类与单元格分类相关研究

表格分类指的是根据表格的结构或内容，对表格进行分类。Wang 和 Hu（2002）根据表格包含的内容将表格分为正品表（genuine table）和非正品表（non-genuine table）。其中非正品表是指仅仅使用 HTML 表格标签来进行网页布局的内容，并不是用来展示表格数据。他们对 HTML 中包含〈table〉标签的部分抽取了一系列特征，提出了一个可训练的机器学习方法对表格进行分类。Crestan 和 Pantel（2011）将表格分为两大类别：关系型知识表格以及不包含知识仅仅用于布局的表格。其中前者又细分为列表型、属性/值、矩阵型、枚举型和填空型；后者细分为导航型、格式化型。他们从表格中抽取了表格的布局特征和内容特征，进而使用有监督机器学习算法对表格进行分类。

单元格分类指的是将单元格分成表头、数据单元格等类别。Fang 等人（2012a）比较了简单的启发式算法和基于机器学习的分类算法之间的效果。其中启发式算法假设表格的行列表头分别存在于表格的左边和上边，计算表格中连续的两行/列的相似性，并以从上到下/从左到右出现的第 1 个局部最小值当做表头和数据的分隔，从而得到表头；基于机器学习的分类算法利用了一系列能够区分表头的特征，并使用支持向量机分类器、逻辑回归和随机森林来将单元格行或列分类为表头或数据单元格。Seth 和 Nagy（2013）将单元格分成 5 种不同类型：行表头、列表头、数据、存根表头（stub head）和额外信息，利用表格中“每个数据单元格可以被行列表头路径的文本序列唯一确定”这一特性，来识别表格中的每个单元格的类型。Koci 等人（2016）将表格中的单元格分为 5 个类别：元数据、表头、属性、数据和派生数据，抽取表格单元格的内容特征、单元格风格特征、字体特征、引用特征和空间特征，然后将这些特征以及标注结果输入到常用的机器学习分类器中进行学习，得到单元格类别。Gol 等人（2019）综合考虑了单元格中的文本和风格特征，通过一个表格预训练模型得到每个单元格的向量，利用单元格向量将其归类到 6 个类别中。

国内在表格分类和单元格分类领域的研究相对较少。其中具有代表性的是北京航空航天大学和微软亚洲研究院的 Dong 等人（2019）的研究，他们利用 BERT（bidirectional encoder representation from Transformers）（Devlin 等，2018）提取表格中的文本语义特征，并与其他手工特征一同输入到 FCNN（fully CNN）骨干网络中，然后以 3 个分支网络将表格信息抽取任务、表格区域检测和单元格分类 3 个任务融入到这一多任务提取框架中。

1.4.2 表格信息抽取相关研究

基于表格的信息抽取任务是从表格或者包含表格的文档中提取给定的关键信息字段，并对其进行归纳、分析。从实际应用的角度来看，自动地从表格、票据和合同等文档中收集个人信息、重要日期、地址和金额等关键字段具有很高的应用价值。

其中具有代表性的 ICDAR2019 举办的表格信息提取竞赛的 SROIE 数据集的一些代表性结果如表 6 所示。

表 6　SROIE 数据集信息抽取结果比较

Table 6　Comparison results of table information extraction on SROIE

方法	F1/%
Liu 等人(2019a)	95.10
RoBERTa(robustly optimized BERT approach)(Liu 等,2019b)	95.60
LayoutLM(Xu 等,2020)	95.24
TRIE(text reading and information extraction)(Zhang 等,2020)	96.18
PICK(Yu 等,2020)	96.12
LAMBERT(Garncarek 等,2020)	96.93
VIES(visual information extraction system)(Wang 等,2021a)	96.12
LayoutLM v2(Xu 等,2022)	96.61
TILT(text-image-layout Transformer)(Powalski 等,2021)	98.10
MatchVIE(Tang 等,2021)	96.57
TCPN(Tag · Copy or predict network)(Wang 等,2021b)	96.54
ViBERTgrid(Lin 等,2021)	96.40
StrucTexT(Li 等,2021d)	96.88

近年来,随着自然语言处理技术的发展,一部分研究者的研究兴趣从传统的序列文本逐渐转向表格等(半)结构化文档上来,并将序列文本上先进的语言模型,例如 LSTM(Hochreiter 和 Schmidhuber,1997),Transformer(Vaswani 等,2017),GPT(generative pre-training)(Radford 和 Narasimhan,2018),BERT(Devlin 等,2018)及 LayoutLM(Xu 等,2020)等应用于表格等(半)结构化文档上,取得了良好的效果,说明这些模型在自然语言处理任务中具有良好的普适性和可迁移性。其中 BERT 及其变体 RoBERTa(Liu 等,2019b)、LayoutLM 及其变体 LayoutLMv2(Xu 等,2022)在表格信息抽取理解的各类任务中都取得了较为稳定且高效的性能,成为该领域中的基线方法。

目前常用的表格等(半)结构化文档信息抽取的公开数据集有 SROIE(Huang 等,2019b),FUNSD(Jaume 等,2019),CORD(consolidated receipt dataset)(Park 等,2019),Kleister(Graliński 等,2020)等。另外,近年来的研究中很多研究者在提出信息抽取方法时,也会建立一套特定应用场景的数据集,例如中文增值税发票(Liu 等,2019a)、火车票、医学处方(Yu 等,2020)、出租车收据(Zhang 等,2020)和试卷标题(Wang 等,2021a)等。

表格信息抽取是表格内容识别中的一项基础任务,根据对表格文档表示形式的不同,可以分为基于序列、基于图和基于 2 维特征网格等信息抽取方法。

基于序列的方法与典型的自然语言处理方法类似,需要将表格文档首先序列化为 1 维文本序列,然后使用现有的序列标记模型(如 LSTM-CNN(Chiu 和 Nichols,2016)、Bi-LSTM-CRF(Ma 和 Hovy,2016)、BERT(Devlin 等,2018)、RoBERTa(Liu 等,2019b)等)提取字段值。较为新颖的方法(如 LayoutLM,LAMBERT(Garncarek 等,2020)等)则会在序列文本信息之外加入表格的布局信息和结构信息,通过融合不同模态的信息、联合训练不同模态特征等方式来提高精度。

基于图的方法将每个文档页面建模为一个图,其中文本片段(单词或文本行)表示为节点。每个节点的初始表示可以结合其对应文本段的视觉、文本和位置特征。然后利用图神经网络或自注意力机制(Vaswani 等,2017)在图中相邻节点之间传播信息,得到每个节点的更丰富的表示,随后将这些图节点的特征输入到分类器模型(如 PICK(Yu 等,2020)),或与文本特征共同输入到序列标记模型中获得所需的字段(如 GraphIE(Qian 等,2019)、Liu 等人(2019a)、Wei 等人(2020)、TRIE(Zhang 等,2020)和 VIES(Wang 等,2021a)等)。

基于 2 维特征网格(2D grid)的方法将文档表示为一个包含字符特征的 2D 网格,然后使用标准实例分割模型从 2D 网格中提取字段值。这一类方法首先由 Katti 等人(2018)在 Chargrid 中提出。Chargrid 引入了 2D 网格作为新的文本表示类型,通过将每个文档页面编码为 2 维字符网格,可以保留文档的 2 维布局,并提出一个用于结构化文档的通用文档理解处理流程,利用完全卷积的编码器—解码器网络来预测分割掩码和边界框。2D 网格表示保留了文档的文本和布局信息,但忽略了图像信息,为此,VisualWordGrid(Kerroumi 等,2021)将这些网

格表示与文档图像的 2D 特征图相结合，生成更强大的多模态 2D 文档表示，它可以同时保存文档的视觉、文本和布局信息。BERTgrid（Denk 和 Reiss-wig，2019）对 Chargrid 进行了改良，将文档表示为上下文词块特征向量的网格，在网络结构中加入了 BERT 网络，对来自目标领域的大量未标记文档进行预训练，为文档中的每个词块计算上下文特征向量。与其他基于 2 维网格的方法相比，虽然 BERT-grid 在网格表示中加入了语言模型 BERT，但在模型训练时，预训练的 BERT 参数是固定的，没有充分发挥语言模型的作用。

此外，在框架构建方面，Clova AI 的 Hwang 等人（2021）提出了一个信息抽取框架 SPADE（spatial dependency parser），将信息抽取任务表述为一个空间依赖解析问题。它以端到端方式在文档中建模高度复杂的空间关系和任意数量的信息层。BROS（BERT relying on spatiality）（Hong 等，2021）通过提出一种新的位置编码方法和基于区域掩蔽的训练，进一步改进了 SPADE，在大规模半结构化文档上使用新的区域掩蔽策略进行预训练，同时有效地包含了输入文档的空间布局信息。Applica. ai 的 Powalski 等人（2021）提出了一种同时学习布局信息、视觉特征和文本语义的神经网络架构 TILT，以预训练的 Transformer 为骨干网络，将布局信息表示为注意力机制中的偏差项，并使用 U-Net（Ronneberger 等，2015）提取上下文的视觉特征加入到模型的输入中。

国内的研究者近年来在表格信息抽取领域取得了丰硕的成果，尤其是在基于图的信息抽取方法研究中取得了领先地位，在基础模型的构建方面也颇有建树。

在基于序列的表格信息抽取方法中，由于顺序文本上的语言模型（如 Transformer、BERT 等）难以捕捉表格文档的结构信息，哈尔滨工业大学和北京航空航天大学的 Xu 等人（2020）提出了 LayoutLM 模型，现在已经成为表格内容理解领域中众多研究方向的基线模型。LayoutLM 模型相对于传统的序列语言模型有了明显的革新，将文档的结构信息也输入到了模型中，丰富了结构化文档的特征表示。哈尔滨工业大学和微软亚洲研究院的 Xu 等人（2022）随后对 LayoutLM 进行了优化，提出了性能更强的 LayoutLMv2。阿里巴巴公司的 Wang 等人（2020）提出了 StructBERT，将语言结构融入到预训练中，结合词结构目标和句子结构目标，利用语境表征中的语言结构来扩展 BERT。这使得 StructBERT 能够通过强制重建单词和句子的正确顺序进行预测，从而显式地对语言结构进行建模。

在基于图的方法方面，阿里巴巴集团的 Liu 等人（2019a）提出了一种基于图卷积的模型，以结合富信息视觉文档中呈现的文本和视觉信息。将表格数据转化为图特征，经过训练以总结文档中文本段的上下文，并进一步与文本特征相结合以进行实体提取。徐州医科大学和平安科技（深圳）有限公司的 Yu 等人（2020）提出了 PICK，充分而有效地利用文档的特性（包括文本、位置、布局和图像）来获得更丰富的语义表示，并结合图学习与图卷积，将图学习模块引入到现有的图架构中，没有人为预先定义图的边缘类型，而是学习一个软邻接矩阵，表示任务节点之间的关系。利用图卷积的方法，在输入信息中加入了文档的文本、图像、位置等特征，提供了更加丰富的表格表示。学习到更丰富的表示，并用于解码器，以辅助进行字符级别的序列标记。华南理工大学的 Tang 等人（2021）提出的 MatchVIE，首次将键值匹配模型用于视觉信息抽取任务中，集成了实体的语义、位置和视觉信息，通过图网络中边的关系来评价实体的相关性，证明了对键值关系进行建模可以有效地提取视觉信息，为表格信息抽取任务提供了一个新的视角。

在基于 2 维特征网格的方法中，Lin 等人（2021）提出了 ViBERTgrid 方法，拼接 BERTGrid 特征图到 CNN 中间层得到的多模态主干网络，并对参数进行联合训练，显著提高了模型的语言标识能力，将基于 2 维特征网格的方法与多模态融合、联合训练以及大规模预训练等方法相结合，相较于之前的同类方法有了大幅提升。

针对当下普遍流行的基于 OCR 结果进行表格文档信息提取所带来的高标注成本和标签歧义等弊端，华南理工大学的 Wang 等人（2021b）还提出了一种统一的弱监督学习框架 TCPN，在编码阶段引入了一种高效的 2D 文档表示方法，对 2 维 OCR 结果中的语义和布局信息进行建模，在解码阶段进行 OCR 纠错和快速推理，同时仅使用关键信息序列作为监督，极大地节省了标注成本并避免了标签歧义。这一方法对于如何缓解对完整标注的过度依赖，以

及如何减轻 OCR 错误带来的负面影响具有启发性。

2 国内外研究进展比较

2.1 表格检测

从总体上看,早期在表格检测识别研究上投入比较大的是美国、德国和日本等;后来随着深度学习的发展,表格检测和结构识别研究呈现了百花齐放的状态。其中比较突出的有印度的研究,在 IBM 公司支持下的澳大利亚、美国的一些研究,以及国内大学和互联网公司的一系列研究。目前,工业界也涌现了一大批表格检测和识别的服务。国外的一些大型云服务商已经在他们的平台上提供了表格检测和识别的功能,比如亚马逊的 Textact 服务、微软的 Azure 服务等。而在国内,既有一些提供表格检测和识别等云端基础服务的互联网公司,例如百度、阿里巴巴、腾讯、华为和网易等,也有一些深耕于相关领域多年的专业服务提供商,例如庖丁科技、好未来等。

2.2 表格结构识别

从表格结构识别的效果上看,国内目前已经处于世界较为领先的水平。2020 年末和 2021 年初由 IBM 公司发起举办了 ICDAR2021 科学文档解析比赛(Jimeno-Yepes 等,2021),其中的任务二——表格识别任务,吸引了来自国内外的多个公司、学校参加。国内许多公司都参与了这场比赛,其中海康威视提出的 LGPMA 模型和平安科技提出的 TableMaster 模型分别取得了比赛的第 1、2 名。由此可见,在表格检测和结构识别的研究领域,尤其是在应用方面,国内的研究者已经取得了国际领先的地位。

从数据集上看,国外的数据集主要为类 PDF 文档,其中的表格结构比较规整,不存在扭曲、阴影等问题,例如 SciTSR、PubTabNet 等。而国内除了规整文档的表格数据集 Tablebank 之外,已经开始出现自然场景表格的数据集,例如 WTW、NTable、TAL_OCR_TABLE 比赛等数据集,这些数据集中应用场景更丰富,也对表格识别方法提出了进一步的挑战。

2.3 表格内容识别

在表格内容识别的各个领域,国内外研究者研究方向和方法选择上呈现出了不同的偏好。在语言模型构建方面,由于目前表格内容识别领域常用的模型仍以序列语言模型的改进为主,国外起步较早,技术积累更为丰富,LSTM、Transformer、BERT 等一系列经典模型在表格内容识别任务中均取得了较好的效果。但国内近年来出现了 LayoutLM、StructBERT 等先进的文档表征模型,这些模型专门针对表格等(半)结构化文档进行设计,并成为相关领域常用的基线模型之一,在基础模型构建的方面呈现出了较好的发展势头。

具体而言,在表格信息抽取方面,国内的研究者在基于图和基于 2 维特征网格的方法上居于世界领先地位,PICK、MatchVIE 和 ViBERTGrid 等方法在各类信息抽取任务榜单中居于前列;国外的研究者在基于序列的方法上较为突出,提出了 LAMBERT、TILT 等一系列表现优异的模型,这与国外积累已久的语言模型发展经验密不可分,在基于 2 维特征网格的方法上国外起步更早,提出了 Chargrid 和 BERTgrid 等经典模型,而对于基于图的方法研究较少。总体而言,近年来国内外研究者对表格内容识别均有很高的研究热情,这一领域的方法也呈现出多样化发展的趋势。

3 发展趋势与展望

对于表格区域检测,其准确率已经达到了比较高的水平。而检测作为识别的一部分,两者逐渐一体化,单独的检测逐渐弱化。如何让检测和结构识别的结果相互促进将是以后研究的方向和重点。

由于表格应用场景较为广泛,表格形式多种多样,文档图像质量参差不齐,表格结构识别仍存在着较大的挑战。具体表现为:1)跨页表格对结构识别带来的识别困难;2)表格线未对齐带来的行列判定困难;3)表格嵌套(某些小表格是大表格的单元格)带来的识别困难;4)一些非常规的表格线标注形式;5)现实场景带来的扭曲、褶皱和光照等问题。

对于表格结构识别,现阶段主流的方案包括两种:1)单元格检测 + 单元格关系判断;2)编码解码器同时生成 HTML 或 Latex 代码以及相对应的单元格位置。方案 1)主要关注如何检测出更准确的单元格,在后续研究中可尝试使用表格文本的语义信息来提高;方案 2)主要关注生成的代码过长时,准确率的降低以及回归的单元格框漂移等问题,可尝试由目标检测网络提出单元格候选框来改善。未来随着表格应用场景的增加,表格数据集的丰富,现实

场景的表格识别以及表格识别的预训练模型都是值得深入挖掘的方向。

对于表格内容识别与理解，总体来说，随着自然语言模型的成熟和发展，自然语言处理的方法所能处理的信息形式已经不仅仅局限于 1 维的顺序文本，研究者们对于表格、票据等（半）结构化文档信息提取的研究热情日益增长。然而，由于表格形式复杂多样，并涉及各个行业的专业知识，目前研究者们面临着两大挑战：一方面是表格信息的表示方式难以统一，不同形式的表格有着不同形式的结构关系，很难构建出从表格信息到机器表征的通用识别框架，目前的大部分研究还处于针对某类特定的表格数据进行性能优化的阶段；另一方面，对于表格的查询、问答和文本生成等以内容为主导的任务，由于表格数据通常具有一定的专业性且表格中表达的语义不唯一，数据的标注难度很大且成本高昂，训练出的模型迁移能力较差。

随着深度学习技术的发展，大规模预训练模型已经成为自然语言处理领域中广泛认可的有效方法，表格内容的识别及理解在近年来快速发展，但在这一领域中目前并没有出现具有关键影响力的大规模预训练表格理解和表格生成模型。目前常用的方案大多都是对已有的语言模型进行改进，尽管这类方法针对某类具体问题可能是行之有效的，但往往不能很好地应用于其他表格内容识别相关的任务中。因此，寻找并构建出针对表格结构的大规模预训练模型，或是构建出在顺序文本、结构化文本和场景文本等多种形式的文档结构中都有良好表现的预训练语言模型，也是该领域目前面临的一大挑战和重要研究方向。

就整体趋势而言，一方面表格内容识别的任务具有具象化的特征，新的任务和新的应用场景纷纷出现，体现出了很高的应用价值，相关的任务类型和涵盖的领域也趋于具体，出现了很多专门针对具体问题的方法和模型；另一方面，表格内容识别也具有理论意义，研究者们对于基础模型的构建具有很高的研究兴趣，一些与表格内容识别相关的方法已经体现出了很高的泛化能力，能适用于序列文本、结构化文本和场景文本等不同类型的对象。在抽象层次，力图构建泛化性更好的基于文档的表征模型，寻找更加具有普适性的方法来描述、理解和处理表格信息，也是未来的研究热点之一。

致　谢　本文由中国图象图形学学会文档图像分析与识别专业委员会组织撰写，该专委会更多详情请见链接：http://www.csig.org.cn/detail/2551。

参考文献（References）

Abdallah A, Berendeyev A, Nuradin I and Nurseitov D. 2022. TNCR: table net detection and classification dataset. Neurocomputing, 473: 79-97 [DOI: 10.1016/j.neucom.2021.11.101]

Agarwal M, Mondal A and Jawahar C V. 2020. CDeC-Net: composite deformable cascade network for table detection in document images//Proceedings of 2020 25th International Conference on Pattern Recognition (ICPR). Milan, Italy: IEEE: 9491-9498 [DOI: 10.1109/ICPR48806.2021.9411922]

Amano A, Asada N, Motoyama T, Sumiyoshi T and Suzuki K. 2001. Table form document synthesis by grammar-based structure analysis//Proceedings of the 6th International Conference on Document Analysis and Recognition. Seattle, USA: IEEE: 533-537 [DOI: 10.1109/ICDAR.2001.953846]

Cai Z W and Vasconcelos N. 2018. Cascade R-CNN: delving into high quality object detection//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 6154-6162 [DOI: 10.1109/CVPR.2018.00644]

Chen S, Jaisimha M Y, Ha J, Haralick R M and Phillips I T. 1996. User's Reference Manual for the UW English. Washington: Seattle University

Chi Z W, Huang H Y, Xu H D, Yu H J, Yin W X and Mao X L. 2019. Complicated table structure recognition [EB/OL]. [2022-01-25]. https://arxiv.org/pdf/1908.04729.pdf

Chiu J P C and Nichols E. 2016. Named entity recognition with bidirectional LSTM-CNNs [EB/OL]. [2022-01-25]. https://arxiv.org/pdf/1511.08308.pdf

Clinchant S, Déjean H, Meunier J L, Lang E M and Kleber F. 2018. Comparing machine learning approaches for table recognition in historical register books//Proceedings of the 13th IAPR International Workshop on Document Analysis Systems (DAS). Vienna, Austria: IEEE: 133-138 [DOI: 10.1109/DAS.2018.44]

Crestan E and Pantel P. 2011. Web-scale table census and classification//Proceedings of the 4th ACM international conference on Web search and data mining (WSDM11). New York, USA: Association for Computing Machinery: 545-554 [DOI: 10.1145/1935826.1935904]

Dai J F, Qi H Z, Xiong Y W, Li Y, Zhang G D, Hu H and Wei Y C. 2017. Deformable convolutional networks//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE: 764-773 [DOI: 10.1109/ICCV.2017.89]

Deng Y, Kanervisto A, Ling J and Rush A M. 2017. Image-to-markup generation with coarse-to-fine attention [EB/OL]. [2022-01-25].

https://arxiv.org/pdf/1609.04938.pdf

Deng Y T, Rosenberg D and Mann G. 2019. Challenges in end-to-end neural scientific table recognition//Proceedings of 2019 International Conference on Document Analysis and Recognition (ICDAR). Sydney, Australia: IEEE: 894-901 [DOI: 10.1109/ICDAR.2019.00148]

Denk T I and Reisswig C. 2019. BERTgrid: contextualized embedding for 2D document representation and understanding [EB/OL]. [2022-01-25]. https://arxiv.org/pdf/1909.04948.pdf

Desai H, Kayal P and Singh M. 2021. TabLeX: a benchmark dataset for structure and content information extraction from scientific tables//Proceedings of the 16th International Conference on Document Analysis and Recognition—ICDAR 2021. Lausanne, Switzerland: Springer: 554-569 [DOI: 10.1007/978-3-030-86331-9_36]

Devlin J, Chang M W, Lee K and Toutanova K. 2018. BERT: pre-training of deep bidirectional transformers for language understanding//Proceedings of 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis, Minnesota: Association for Computational Linguistics: 4171-4186 [DOI: 10.18653/v1/N19-1423]

Dong H Y, Liu S J, Fu Z Y, Han S and Zhang D M. 2019. Semantic structure extraction for spreadsheet tables with a multi-task learning architecture//Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS). Vancouver, Canada: [s. n.]

Duan K W, Bai S, Xie L X, Qi H G, Huang Q M and Tian Q. 2019. CenterNet: keypoint triplets for object detection//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 6568-6577 [DOI: 10.1109/ICCV.2019.00667]

Fang J, Gao L C, Bai K, Qiu R H, Tao X and Tang Z. 2011. A table detection method for multipage PDF documents via visual seperators and tabular structures//Proceedings of 2011 International Conference on Document Analysis and Recognition. Beijing, China: IEEE: 779-783 [DOI: 10.1109/ICDAR.2011.304]

Fang J, Mitra P, Tang Z and Giles C L. 2012a. Table header detection and classification//Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI'12). Toronto Ontario, Canada: AAAI Press: 599-605 [DOI: 10.5555/2900728.2900814]

Fang J, Tao X, Tang Z, Qiu R H and Liu Y. 2012b. Dataset, ground-truth and performance metrics for table detection evaluation//Proceedings of the 10th IAPR International Workshop on Document Analysis Systems. Gold Coast, Australia: IEEE: 445-449 [DOI: 10.1109/DAS.2012.29]

Gao L C, Huang Y L, Dejean H, Meunier J L, Yan Q Q, Fang Y, Kleber F and Lang E. 2019. ICDAR 2019 competition on table detection and recognition (cTDaR)//Proceedings of 2019 International Conference on Document Analysis and Recognition (ICDAR). Sydney, Australia: IEEE: 1510-1515 [DOI: 10.1109/ICDAR.2019.00243]

Gao L C, Yi X H, Jiang Z R, Hao L P and Tang Z. 2017. ICDAR2017 competition on page object detection//Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Kyoto, Japan: IEEE: 1417-1422 [DOI: 10.1109/ICDAR.2017.231]

Garncarek Ł, Powalski R, Stanisławek T, Topolski B, Halama P, Turski M and Graliński F. 2020. LAMBERT: layout-aware (language) Modeling for information extraction [EB/OL]. [2022-01-25]. https://arxiv.org/pdf/2002.08087.pdf

Gilani A, Qasim S R, Malik I and Shafait F. 2017. Table detection using deep learning//Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Kyoto, Japan: IEEE: 771-776 [DOI: 10.1109/ICDAR.2017.131]

Göbel M, Hassan T, Oro E and Orsi G. 2013. ICDAR 2013 table competition//Proceedings of the 12th International Conference on Document Analysis and Recognition. Washington, USA: IEEE: 1449-1453 [DOI: 10.1109/ICDAR.2013.292]

Gol M G, Pujara J and Szekely P. 2019. Tabular cell classification using pre-trained cell embeddings//Proceedings of 2019 IEEE International Conference on Data Mining (ICDM). Beijing, China: IEEE: 230-239 [DOI: 10.1109/ICDM.2019.00033]

Graliński F, Stanislawek T, Wróblewska A, Lipiński D, Kaliska A, Rosalska P, Topolski B and Biecek P. 2020. Kleister: a novel task for information extraction involving long documents with complex layout [EB/OL]. [2022-01-25]. https://arxiv.org/pdf/2003.02356.pdf

Hassan T and Baumgartner R. 2007. Table recognition and understanding from PDF files//Proceedings of 9th International Conference on Document Analysis and Recognition. Curitiba, Brazil: IEEE: 1143-1147 [DOI: 10.1109/ICDAR.2007.4377094]

He D F, Cohen S, Price B, Kifer D and Giles C L. 2017a. Multi-scale multi-task FCN for semantic page segmentation and table detection//Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Kyoto, Japan: IEEE: 254-261 [DOI: 10.1109/ICDAR.2017.50]

He K M, Gkioxari G, Dollár P and Girshick R. 2017b. Mask R-CNN//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE: 2980-2988 [DOI: 10.1109/ICCV.2017.322]

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

He Y L, Qi X B, Ye J Q, Gao P, Chen Y H, Li B C, Tang X and Xiao R. 2021. PingAn-VCGroup's solution for ICDAR 2021 competition on scientific table image recognition to latex [EB/OL]. [2022-01-26]. https://arxiv.org/pdf/2105.01846.pdf

Hirayama Y. 1995. A method for table structure analysis using DP matc-

hing//Proceedings of the 3rd International Conference on Document Analysis and Recognition. Montreal, Canada: IEEE: 583-586 [DOI: 10.1109/ICDAR.1995.601964]

Hochreiter S and Schmidhuber J. 1997. Long short-term memory. Neural Computation, 9(8): 1735-1780 [DOI: 10.1162/neco.1997.9.8.1735]

Hong T, Kim D, Ji M G, Hwang W, Nam D and Park S. 2021. BROS: a pre-trained language model focusing on text and layout for better key information extraction from documents [EB/OL]. [2022-01-26]. https://arxiv.org/pdf/2108.04539v4.pdf

Huang Y L, Yan Q Q, Li Y B, Chen Y F, Wang X, Gao L C and Tang Z. 2019a. A YOLO-based table detection method//Proceedings of 2019 International Conference on Document Analysis and Recognition (ICDAR). Sydney, Australia: IEEE: 813-818 [DOI: 10.1109/ICDAR.2019.00135]

Huang Z, Chen K, He J H, Bai X, Karatzas D, Lu S J and Jawahar C V. 2019b. ICDAR2019 Competition on scanned receipt OCR and information extraction//Proceedings of 2019 International Conference on Document Analysis and Recognition (ICDAR). Sydney, Australia: IEEE: 1516-1520 [DOI: 10.1109/ICDAR.2019.00244]

Hwang W, Yim J, Park S, Yang S and Seo M. 2021. Spatial dependency parsing for semi-structured document information extraction//Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. [s.l.]: Association for Computational Linguistics: 330-343 [DOI: 10.18653/v1/2021.findings-acl.28]

Ishitani Y, Fume K and Sumita K. 2005. Table structure analysis based on cell classification and cell modification for XML document transformation//Proceedings of the 8th International Conference on Document Analysis and Recognition (ICDAR'05). Seoul, Korea (South): IEEE: 1247-1252 [DOI: 10.1109/ICDAR.2005.225]

Itonori K. 1993. Table structure recognition based on textblock arrangement and ruled line position//Proceedings of the 2nd International Conference on Document Analysis and Recognition (ICDAR'93). Tsukuba, Japan: IEEE: 765-768 [DOI: 10.1109/ICDAR.1993.395625]

Jaume G, Ekenel K H and Thiran J P. 2019. FUNSD: a dataset for form understanding in noisy scanned documents//Proceedings of 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW). Sydney, Australia: IEEE: 1-6 [DOI: 10.1109/ICDARW.2019.10029]

Jia J, Zhai G T, Ren P, Zhang J H, Gao Z P, Min X K and Yang X K. 2020. Tiny-BDN: an efficient and compact barcode detection network. IEEE Journal of Selected Topics in Signal Processing, 14(4): 688-699 [DOI: 10.1109/JSTSP.2020.2976566]

Jimeno-Yepes A, Zhong P and Burdick D. 2021. ICDAR 2021 competition on scientific literature parsing//Proceedings of the 16th International Conference on Document Analysis and Recognition (ICDAR). Lausanne, Switzerland: Springer: 605-617 [DOI: 10.1007/978-3-030-86337-1_40]

Katti A R, Reisswig C, Guder C, Brarda S, Bickel S, Höhne J and Faddoul J B. 2018. Chargrid: towards understanding 2D documents//Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics: 4459-4469 [DOI: 10.18653/v1/D18-1476]

Kavasidis I, Pino C, Palazzo S, Rundo F, Giordano D, Messina P and Spampinato C. 2019. A saliency-based convolutional neural network for table and chart detection in digitized documents//Ricci E, Rota Bulò S, Snoek C, Lanz O, Messelodi S and Sebe N, eds. Image Analysis and Processing—ICIAP. Cham: Springer: 292-302 [DOI: 10.1007/978-3-030-30645-8_27]

Kerroumi M, Sayem O and Shabou A. 2021. VisualWordGrid: information extraction from scanned documents using a multimodal approach//Barney Smith E H, Pal U, eds. Document Analysis and Recognition – ICDAR 2021 Workshops. Cham: Springer: 389-402 [DOI: 10.1007/978-3-030-86159-9_28]

Khan S A, Khalid S M D, Shahzad M A and Shafait F. 2019. Table structure extraction with bi-directional gated recurrent unit networks//Proceedings of 2019 International Conference on Document Analysis and Recognition (ICDAR). Sydney, Australia: IEEE: 1366-1371 [DOI: 10.1109/ICDAR.2019.00220]

Kieninger T and Dengel A. 2001. Applying the T-Recs table recognition system to the business letter domain//Proceedings of the 6th International Conference on Document Analysis and Recognition. Seattle, USA: IEEE: 518-522 [DOI: 10.1109/ICDAR.2001.953843]

Kieninger T G. 1998. Table structure recognition based on robust block segmentation//Proceedings of SPIE 3305, Document Recognition V. San Jose, USA: SPIE [DOI: 10.1117/12.304642]

Koci E, Thiele M, Romero O and Lehner W. 2016. A machine learning approach for layout inference in spreadsheets//Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2016). Lda, Setubal, Portugal: SCITEPRESS-Science and Technology Publications: 77-88 [DOI: 10.5220/0006052200770088]

Koci E, Thiele M, Rehak J, Romero O and Lehner W. 2019. DECO: a dataset of annotated spreadsheets for layout and table recognition//Proceedings of 2019 International Conference on Document Analysis and Recognition (ICDAR). Sydney, Australia: IEEE: 1280-1285 [DOI: 10.1109/ICDAR.2019.00207]

Law H and Deng J. 2018. CornerNet: detecting objects as paired keypoints//Proceedings of the 15th European Conference on Computer Vision-ECCV 2018. Munich, Germany: Springer: 765-781 [DOI: 10.1007/978-3-030-01264-9_45]

Li J F, Wang K, Hao S Q and Wang Q R. 2012. Location and recognition of free tables in form//Zhang W, ed. Software Engineering and Knowledge Engineering: Theory and Practice. Berlin, Heidelberg: Springer: 685-692 [DOI: 10.1007/978-3-642-29455-6_94]

Li M H, Cui L, Huang S H, Wei F R, Zhou M and Li Z J. 2020.

Tablebank: table benchmark for image-based table detection and recognition//Proceedings of the 12th Language Resources and Evaluation Conference (LREC). Marseille, France: European Language Resources Association: 1918-1925

Li X H, Yin F, Zhang X Y and Liu C L. 2021a. Adaptive scaling for archival table structure recognition//Proceedings of the 16th International Conference on Document Analysis and Recognition (ICDAR). Lausanne, Switzerland: Springer: 80-95 [DOI: 10.1007/978-3-030-86549-8_6]

Li Y B, Gao L C, Tang Z, Yan Q Q and Huang Y L. 2019. A GAN-based feature generator for table detection//Proceedings of 2019 International Conference on Document Analysis and Recognition (ICDAR). Sydney, Australia: IEEE: 763-768 [DOI: 10.1109/ICDAR.2019.00127]

Li Y B, Huang Y L, Zhu Z Y, Pan L M, Huang Y S, Du L, Tang Z and Gao L C. 2021b. Rethinking table structure recognition using sequence labeling methods//Proceedings of the 16th International Conference on Document Analysis and Recognition (ICDAR). Lausanne, Switzerland: Springer: 541-553 [DOI: 10.1007/978-3-030-86331-9_35]

Li Y L, Qian Y X, Yu Y C, Qin X M, Zhang C Q, Liu Y, Yao K, Han J Y, Liu J T and Ding E R. 2021d. StrucTexT: structured text understanding with multi-modal transformers//Proceedings of the 29th ACM International Conference on Multimedia. [s. l.]: Association for Computing Machinery: 1912-1920 [DOI: 10.1145/3474085.3475345]

Li Y R, Huang Z, Yan J C, Zhou Y, Ye F and Liu X H. 2021c. GFTE: graph-based financial table extraction//Del Bimbo A, Cucchiara R, Sclaroff S, Farinella G M, Mei T, Bertini M, Escalante H J and Vezzani R, eds. International Conference on Pattern Recognition (ICPR). Cham: Springer: 644-658 [DOI: 10.1007/978-3-030-68790-8_50]

Liao M H, Wan Z Y, Yao C, Chen K and Bai X. 2020. Real-time scene text detection with differentiable binarization. Proceedings of the AAAI Conference on Artificial Intelligence, 34(7): 11474-11481 [DOI: 10.1609/aaai.v34i07.6812]

Lin T Y, Dollár P, Girshick R, He K M, Hariharan B and Belongie S. 2017. Feature pyramid networks for object detection//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 936-944 [DOI: 10.1109/CVPR.2017.106]

Lin W H, Gao Q F, Sun L, Zhong Z Y, Hu K, Ren Q and Huo Q. 2021. VIBERTgrid: a jointly trained multi-modal 2 d document representation for key information extraction from documents//Proceedings of the 16th International Conference on Document Analysis and Recognition (ICDAR). Lausanne, Switzerland: Springer: 548-563 [DOI: 10.1007/978-3-030-86549-8_35]

Liu J H, Ding X Q and Wu Y S. 1995. Description and recognition of form and automated form data entry//Proceedings of the 3rd International Conference on Document Analysis and Recognition. Montreal, Canada: IEEE [DOI: 10.1109/ICDAR.1995.601963]

Liu X J, Gao F Y, Zhang Q and Zhao H S. 2019a. Graph convolution for multimodal information extraction from visually rich documents//Proceedings of 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Industry Papers). Minneapolis, Minnesota, USA: Association for Computational Linguistics: 32-39 [DOI: 10.18653/v1/N19-2005]

Liu Y H, Ott M, Goyal N, Du J F, Joshi M, Chen D Q, Levy O, Lewis M, Zettlemoyer L and Stoyanov V. 2019b. RoBERTa: a robustly optimized BERT pretraining approach [EB/OL]. [2022-01-26]. https://arxiv.org/pdf/1907.11692.pdf

Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]

Long R J, Wang W, Xue N, Gao F Y, Yang Z B, Wang Y P and Xia G S. 2021. Parsing table structures in the wild//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 924-932 [DOI: 10.1109/ICCV48922.2021.00098]

Lu N, Yu W W, Qi X B, Chen Y H, Gong P, Xiao R and Bai X. 2021. MASTER: multi-aspect non-local network for scene text recognition. Pattern Recognition, 117: #107980 [DOI: 10.1016/j.patcog.2021.107980]

Ma X Z and Hovy E. 2016. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin, Germany: Association for Computational Linguistics: 1064-1074 [DOI: 10.18653/v1/P16-1101]

Melinda L and Bhagvati C. 2019. Parameter-free table detection method//Proceedings of 2019 International Conference on Document Analysis and Recognition (ICDAR). Sydney, Australia: IEEE: 454-460 [DOI: 10.1109/ICDAR.2019.00079]

Paliwal S S, Vishwanath D, Rahul R, Sharma M and Vig L. 2019. TableNet: deep learning model for end-to-end table detection and tabular data extraction from scanned document images//Proceedings of 2019 International Conference on Document Analysis and Recognition (ICDAR). Sydney, Australia: IEEE: 128-133 [DOI: 10.1109/ICDAR.2019.00029]

Park S, Shin S, Lee B, Lee J, Surh J, Seo M and Lee H. 2019. CORD: a consolidated receipt dataset for post-OCR parsing//Proceedings of the 33rd Conference on Neural Information Processing Systems. Vancouver, Canada: [s. n.]

Powalski R, Borchmann Ł, Jurkiewicz D, Dwojak T, Pietruszka M and Pałka G. 2021. Going full-TILT boogie on document understanding with text-image-layout transformer//Proceedings of the 16th International Conference on Document Analysis and Recognition—ICDAR

2021. Lausanne, Switzerland: Springer: 732-747 [DOI: 10.1007/978-3-030-86331-9_47]

Prasad D, Gadpal A, Kapadni K, Visave M and Sultanpure K. 2020. CascadeTabNet: an approach for end to end table detection and structure recognition from image-based documents//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, USA: IEEE: 2439-2447 [DOI: 10.1109/CVPRW50498.2020.00294]

Qasim S R, Mahmood H and Shafait F. 2019. Rethinking table recognition using graph neural networks//Proceedings of 2019 International Conference on Document Analysis and Recognition (ICDAR). Sydney, Australia: IEEE: 142-147 [DOI: 10.1109/ICDAR.2019.00031]

Qian Y J, Santus E, Jin Z J, Guo J and Barzilay R. 2019. GraphIE: a graph-based framework for information extraction//Proceedings of 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis, Minnesota: Association for Computational Linguistics: 751-761 [DOI: 10.18653/v1/N19-1082]

Qiao L, Li Z S, Cheng Z Z, Zhang P, Pu S L, Niu Y, Ren W Q, Tan W M and Wu F. 2021. LGPMA: complicated table structure recognition with local and global pyramid mask alignment//Proceedings of the 16th International Conference on Document Analysis and Recognition (ICDAR). Lausanne, Switzerland: Springer: 99-114 [DOI: 10.1007/978-3-030-86549-8_7]

Radford A and Narasimhan K. 2018. Improving language understanding by generative pre-training [EB/OL]. [2022-01-26]. http://www.nlpir.org/wordpress/wp-content/uploads/2019/06/Improving-language-understanding-by-generative-pre-training.pdf

Rahgozar M A, Fan Z G and Rainero E V. 1994. Tabular document recognition//Proceedings of SPIE 2181, Document Recognition. San Jose, USA: SPIE: #171096 [DOI: 10.1117/12.171096]

Raja S, Mondal A and Jawahar C V. 2020. Table structure recognition using top-down and bottom-up cues [EB/OL]. [2022-01-25]. https://arxiv.org/pdf/2010.04565.pdf

Ramel J Y, Crucianu M, Vincent N and Faure C. 2003. Detection, extraction and representation of tables//Proceedings of the 7th International Conference on Document Analysis and Recognition, 2003. Proceedings. Edinburgh, UK: IEEE: 374-378 [DOI: 10.1109/ICDAR.2003.1227692]

Ren S Q, He K M, Girshick R and Sun J. 2015. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]

Riba P, Dutta A, Goldmann L, Fornés A, Ramos O and Lladós J. 2019. Table detection in invoice documents by graph neural networks//Proceedings of 2019 International Conference on Document Analysis and Recognition (ICDAR). Sydney, Australia: IEEE: 122-127 [DOI: 10.1109/ICDAR.2019.00028]

Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//Navab N, Hornegger J, Wells W and Frangi A, eds. Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Cham: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]

Saha R, Mondal A and Jawahar C V. 2019. Graphical object detection in document images//Proceedings of 2019 International Conference on Document Analysis and Recognition (ICDAR). Sydney, Australia: IEEE: 51-58 [DOI: 10.1109/ICDAR.2019.00018]

Scarselli F, Gori M, Tsoi A C, Hagenbuchner M and Monfardini G. 2009. The graph neural network model. IEEE Transactions on Neural Networks, 20(1): 61-80 [DOI: 10.1109/TNN.2008.2005605]

Schreiber S, Agne S, Wolf I, Dengel A and Ahmed S. 2017. DeepDeSRT: deep learning for detection and structure recognition of tables in document images//Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Kyoto, Japan: IEEE: 1162-1167 [DOI: 10.1109/ICDAR.2017.192]

Seth S and Nagy G. 2013. Segmenting tables via indexing of value cells by table headers//Proceedings of the 12th International Conference on Document Analysis and Recognition. Washington, USA: IEEE: 887-891 [DOI: 10.1109/ICDAR.2013.181]

Shahab A, Shafait F, Kieninger T and Dengel A. 2010. An open approach towards the benchmarking of table structure recognition systems//Proceedings of the 9th IAPR International Workshop on Document Analysis Systems. Boston, USA: ACM: 113-120 [DOI: 10.1145/1815330.1815345]

Shi B G, Bai X and Yao C. 2017. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(11): 2298-2304 [DOI: 10.1109/TPAMI.2016.2646371]

Shigarov A, Mikhailov A and Altaev A. 2016. Configurable table structure recognition in untagged PDF documents//Proceedings of 2016 ACM Symposium on Document Engineering (DocEng′16). New York, USA: Association for Computing Machinery: 119-122 [DOI: 10.1145/2960811.2967152]

Siddiqui S A, Fateh I A, Rizvi S T R, Dengel A and Ahmed S. 2019. DeepTabStR: deep learning based table structure recognition//Proceedings of 2019 International Conference on Document Analysis and Recognition (ICDAR). Sydney, Australia: IEEE: 1403-1409 [DOI: 10.1109/ICDAR.2019.00226]

Siddiqui S A, Malik M I, Agne S, Dengel A and Ahmed S. 2018. DeCNT: deep deformable CNN for table detection. IEEE Access, 6: 74151-74161 [DOI: 10.1109/ACCESS.2018.2880211]

Siegel N, Lourie N, Power R and Ammar W. 2018. Extracting scientific figures with distantly supervised neural networks//Proceedings of the

18th ACM/IEEE on Joint Conference on Digital Libraries. Fort Worth, USA: ACM: 223-232 [DOI: 10.1145/3197026.3197040]

Simonyan K and Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2022-01-25]. https://arxiv.org/pdf/1409.1556.pdf

Smock B, Pesala R and Abraham R. 2021. PubTables-1M: Towards a universal dataset and metrics for training and evaluating table extraction models [EB/OL]. [2022-01-25]. https://arxiv.org/pdf/2110.00061v2.pdf

Sun N N, Zhu Y P and Hu X M. 2019. Faster R-CNN based table detection combining corner locating//Proceedings of 2019 International Conference on Document Analysis and Recognition (ICDAR). Sydney, Australia: IEEE: 1314-1319 [DOI: 10.1109/ICDAR.2019.00212]

Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A. 2015. Going deeper with convolutions//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE: 1-9 [DOI: 10.1109/CVPR.2015.7298594]

Tang G Z, Xie L L, Jin L W, Wang J P, Chen J D, Xu Z, Wang Q Y, Wu Y Q and Li H. 2021. MatchVIE: exploiting match relevancy between entities for visual information extraction//Proceedings of the 30th International Joint Conference on Artificial Intelligence. [s. l.]: IJCAI: 1039-1045 [DOI: 10.24963/ijcai.2021/144]

Tensmeyer C, Morariu V I, Price B, Cohen S and Martinez T. 2019. Deep splitting and merging for table structure decomposition//Proceedings of 2019 International Conference on Document Analysis and Recognition (ICDAR). Sydney, Australia: IEEE: 114-121 [DOI: 10.1109/ICDAR.2019.00027]

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin I. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS17). Red Hook, USA: Curran Associates Inc.: 6000-6010[DOI: 10.5555/3295222.3295349]

Wang J P, Liu C Y, Jin L W, Tang G Z, Zhang J X, Zhang S T, Wang Q Y, Wu Y Q and Cai W X. 2021a. Towards robust visual information extraction in real world: new dataset and novel solution//Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI21). [s. l.]: AAAI Press, 2738-2745

Wang J P, Wang T W, Tang G Z, Jin L W, Ma W H, Ding K and Huang Y C. 2021b. Tag, copy or predict: a unified weakly-supervised learning framework for visual information extraction using sequences//Proceedings of the 30th International Joint Conference on Artificial Intelligence. [s. l.]: IJCAI: 1082-1090 [DOI: 10.24963/ijcai.2021/150]

Wang W, Bi B, Yan M, Wu C, Bao Z Y, Xia J N, Peng L W and Si L. 2020. StructBERT: incorporating language structures into pre-training for deep language understanding [EB/OL]. [2022-01-25]. https://arxiv.org/pdf/1908.04577.pdf

Wang W H, Xie E Z, Li X, Hou W B, Lu T, Yu G and Shao S. 2019b. Shape robust text detection with progressive scale expansion network//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 9328-9337 [DOI: 10.1109/CVPR.2019.00956]

Wang Y L and Hu J Y. 2002. Detecting tables in HTML documents//Lopresti D, Hu J Y and Kashi R, eds. Document Analysis Systems V. Berlin, Heidelberg: Springer: 249-260 [DOI: 10.1007/3-540-45869-7_29]

Wang Y L, Phillips I T and Haralick R M. 2004. Table structure understanding and its performance evaluation. Pattern Recognition, 37(7): 1479-1497 [DOI: 10.1016/j.patcog.2004.01.012]

Wangt Y, Phillipst I T and Haralick R. 2001. Automatic table ground truth generation and a background-analysis-based table structure extraction method//Proceedings of the 6th International Conference on Document Analysis and Recognition. Seattle, USA: IEEE: 528-532 [DOI: 10.1109/ICDAR.2001.953845]

Watanabe T and Luo Q. 1996 A multilayer recognition method for understanding table-form documents. International Journal of Imaging Systems and Technology, 7(4): 279-288

Watanabe T, Luo Q and Sugie N. 1993a. Structure recognition methods for various types of documents. Machine Vision and Applications, 6(2/3): 163-176 [DOI: 10.1007/BF01211939]

Watanabe T, Luo Q and Sugie N. 1993b. Toward a practical document understanding of table-form documents: its framework and knowledge representation//Proceedings of the 2nd International Conference on Document Analysis and Recognition (ICDAR93). Tsukuba Science City, Japan: IEEE: 510-515 [DOI: 10.1109/ICDAR.1993.395684]

Wei M X, He Y F and Zhang Q. 2020. Robust layout-aware IE for visually rich documents with pre-trained language models//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR20). New York, USA: Association for Computing Machinery: 2367-2376 [DOI: 10.1145/3397271.3401442]

Xu Y, Xu Y H, Lv T C, Cui L, Wei F R, Wang G X, Lu Y J, Florencio D, Zhang C, Che W X, Zhang M and Zhou L D. 2022. LayoutLMv2: multi-modal pre-training for visually-rich document understanding [EB/OL]. [2022-01-25]. https://arxiv.org/pdf/2012.14740v1.pdf

Xu Y H, Li M H, Cui L, Huang S H, Wei F R and Zhou M. 2020. LayoutLM: pre-training of text and layout for document image understanding//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD20). Virtual Event, USA: Association for Computing Machinery, 1192-1200 [DOI: 10.1145/3394486.3403172]

Xue W Y, Li Q Y and Tao D C. 2019. ReS2TIM: reconstruct syntactic structures from table images//Proceedings of 2019 International Conference on Document Analysis and Recognition (ICDAR). Syd-

ney, Australia: IEEE: 749-755 [DOI: 10.1109/ICDAR.2019.00125]

Xue W Y, Yu B S, Wang W, Tao D C and Li Q Y. 2021. TGRNet: a table graph reconstruction network for table structure recognition [EB/OL]. [2022-01-25]. https://arxiv.org/pdf/2106.10598.pdf

Yu F and Koltun V. 2015. Multi-scale context aggregation by dilated convolutions [EB/OL]. [2022-01-25]. https://arxiv.org/pdf/1511.07122.pdf

Yu W W, Lu N, Qi X B, Gong P and Xiao R. 2020. PICK: processing key information extraction from documents using improved graph learning-convolutional networks [EB/OL]. [2022-01-25]. https://arxiv.org/pdf/2004.07464.pdf

Zhang P, Li C, Qiao L, Cheng Z Z, Pu S L, Niu Y and Wu F. 2021. VSR: a unified framework for document layout analysis combining vision, semantics and relations [EB/OL]. [2022-01-25]. https://arxiv.org/pdf/2105.06220.pdf

Zhang P, Xu Y L, Cheng Z Z, Pu S L, Lu J, Qiao L, Niu Y and Wu F. 2020. TRIE: end-to-end text reading and information extraction for document understanding//Proceedings of the 28th ACM International Conference on Multimedia (MM20). Seattle, USA: ACM: 1413-1422 [DOI: 10.1145/3394171.3413900]

Zheng X Y, Burdick D, Popa L, Zhong X and Wang N X R. 2020. Global table extractor (GTE): a framework for joint table identification and cell structure recognition using visual context//Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa, USA: IEEE: 697-706 [DOI: 10.1109/WACV48630.2021.00074]

Zhong X, ShafieiBavani E and Jimeno Yepes A. 2020. Image-based table recognition: data, model, and evaluation//Vedaldi A, Bischof H, Brox T and Frahm J M, eds. Computer Vision-ECCV 2020. Cham: Springer: 564-580 [DOI: 10.1007/978-3-030-58589-1_34]

Zhong X, Tang J B and Yepes A J. 2019. PubLayNet: largest dataset ever for document layout analysis//Proceedings of 2019 International Conference on Document Analysis and Recognition (ICDAR). Sydney, Australia: IEEE: 1015-1022 [DOI: 10.1109/ICDAR.2019.00166]

Zhu Z Y, Gao L C, Li Y B, Huang Y L, Du L, Lu N and Wang X F. 2021. NTable: a dataset for camera-based table detection//Proceedings of the 16th International Conference on Document Analysis and Recognition (ICDAR). Lausanne, Switzerland: Springer: 117-129 [DOI: 10.1007/978-3-030-86331-9_8]

Zuyev K. 1997. Table image segmentation//Proceedings of the 4th International Conference on Document Analysis and Recognition. Ulm, Germany: IEEE: 705-708 [DOI: 10.1109/ICDAR.1997.620599]

作者简介

高良才，1980 年生，男，副教授，主要研究方向为模式识别。
E-mail: glc@pku.edu.cn

汤帜，通信作者，男，研究员，主要研究方向为人工智能、文档处理技术。
E-mail: tangzhi@pku.edu.cn

李一博，男，硕士研究生，主要研究方向为表格识别。
E-mail: yiboli@pku.edu.cn

都林，男，工程师，主要研究方向为机器学习、文字识别。
E-mail: dulin09@huawei.com

张新鹏，男，硕士研究生，主要研究方向为表格理解。
E-mail: zhangxinpeng@pku.edu.cn

朱子仪，女，本科生，主要研究方向为表格识别。
E-mail: zhuziyi@pku.edu.cn

卢宁，男，博士，主要研究方向为计算机视觉。
E-mail: luning12@huawei.com

金连文，男，教授，主要研究方向为计算机视觉、文字识别。
E-mail: eelwjin@scut.edu.cn

黄永帅，男，工程师，主要研究方向为计算机视觉。
E-mail: huangyongshuai1@huawei.com