

基于结构理解的手绘草图表格识别

盖鑫^{1,2)}, 黄进^{1)*}, 王丹琳^{1,2)}, 刘杰¹⁾, 田丰¹⁾, 戴国忠¹⁾

¹⁾ (中国科学院软件研究所人机交互北京市重点实验室 北京 100190)

²⁾ (中国科学院大学计算机科学与技术学院 北京 100049)
(huangjing@iscas.ac.cn)

摘要: 基于笔交互的手绘草图表格比基于 WIMP 界面范式的传统电子表格更易于促进用户思维交流和创造性交互工作, 同时可以避免用户学习成本高和交互界面烦琐等问题. 然而, 由于缺乏专门用于手绘草图表格识别的开源数据集, 以及草图本身具有的模糊性、抽象性和用户在绘制草图时的随意性, 手绘草图表格的识别仍然面临着巨大的挑战. 为了解决存在复杂单元格、结构框线与内容重叠、笔迹重描补笔等具体挑战, 提出了一种基于结构理解的手绘草图表格识别算法, 通过一个支持向量机(support vector machines, SVM)对笔划信息进行划分, 再使用多峰值检测算法与真实交点找寻和单元格特征属性检测结合, 实现复杂草图表格结构的识别. 客观评估结果显示, 与光学字符识别(optical character recognition, OCR)基线算法相比, 基于结构理解的手绘草图表格识别算法在面向草图表格结构识别任务中的树编辑距离相似度(tree-edit-distance-based similarity, TEDS)指标提升了 13%以上; 专家评估结果显示, 基于结构理解的手绘草图表格识别算法在表格结构识别和将内容匹配到对应单元格 2 个评价维度中的识别效果均优于 OCR 基线算法的识别效果.

关键词: 手绘草图; 草图结构理解; 表格识别算法; 人机交互

中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2024.20130

Freehand Sketch Table Recognition Based on Structure Understanding

Ge Xin^{1,2)}, Huang Jin^{1)*}, Wang Danlin^{1,2)}, Liu Jie¹⁾, Tian Feng¹⁾, and Dai Guozhong¹⁾

¹⁾ (Beijing Key Laboratory of Human-Computer Interaction, Institute of Software, Chinese Academy of Sciences, Beijing 100190)

²⁾ (School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049)

Abstract: Pen-based sketching tables are more likely to facilitate users' communication and creative interaction, avoiding the problems of high learning cost and cumbersome interaction interface that exist in traditional spreadsheet systems, compared to traditional spreadsheets systems based on the WIMP interface paradigm. However, the recognition of hand-drawn sketch tables still faces great challenges due to the lack of open source datasets dedicated to the recognition of hand-drawn sketch tables, as well as the ambiguity and abstraction of the sketches themselves and the arbitrariness of the users in drawing them. In order to solve the specific challenges such as the existence of complex cells, the overlapping of structural frame lines and content, and the redrawing and filling of handwriting, propose an algorithm for freehand sketch table recognition based on table structure understanding. A support vector machines (SVM) is used to classify the stroke information, and then the multi-peak detection algorithm is combined with the real intersection finding and cell feature attribute detection con-

收稿日期: 2023-01-17; 修回日期: 2023-07-25. 基金项目: 国家自然科学基金(62172397); 中国科学院青年创新促进会项目(2020113). 盖鑫(1998—), 女, 硕士研究生, 主要研究方向为笔交互、草图理解; 黄进(1985—), 男, 博士, 助理研究员, CCF 会员, 论文通信作者, 主要研究方向为人机交互、图形图像处理; 王丹琳(2000—), 女, 硕士研究生, 主要研究方向为笔交互、草图理解; 刘杰(1981—), 男, 博士, 高级工程师, CCF 会员, 主要研究方向为人机交互、计算机图形学; 田丰(1976—), 男, 博士, 研究员, 博士生导师, CCF 会员, 主要研究方向为人机交互、虚拟现实; 戴国忠(1944—), 男, 学士, 研究员, 博士生导师, CCF 会员, 主要研究方向为人机交互、计算机图形学.

structured in this paper to achieve the recognition of complex sketch table structure. The objective evaluation results show that the tree-edit-distance-based similarity (TEDS) of this algorithm in the sketch-oriented table structure recognition task is improved by more than 13% compared with the OCR baseline algorithm; the expert evaluation results show that this algorithm outperforms the OCR baseline algorithm in the evaluation dimensions of table structure recognition and matching the content to the corresponding cells.

Key words: freehand sketch; sketch structure understanding; table recognition algorithm; human-computer interaction

近年来,随着计算机技术的快速发展与计算机应用在各行各业的逐渐深入,以键盘和鼠标作为输入设备的 WIMP(window, icon, menu, pointer)界面范式的精确交互方式已经难以适应不断变化的应用场景.相比之下,基于笔交互的笔式用户界面具有更自然、更灵活和更符合人类习惯的特点,随着移动设备、大型电子白板和平板电脑等触控设备的普及,笔式用户界面已经成为了主流的自然交互方式之一,被广泛应用于各个领域^[1].在表格创作中引入笔交互可以在融合 WIMP 界面范式优点的同时,为用户提供草图自由形式输入、发挥笔迹视觉优势和帮助用户集中注意力等,相比于基于 WIMP 界面范式的传统电子表格系统,基于笔交互的手绘草图表格更容易促进用户的思维交流和创造性交互工作,可以避免基于 WIMP 界面范式的传统电子表格系统存在的用户学习成本高和交互界面烦琐等问题^[1-3].

目前更多的研究工作集中在印刷体表格的光学字符识别(optical character recognition, OCR),对于草图表格,尤其是考虑草图输入模糊、随意性的研究较少.以文献[4]为代表的印刷体表格 OCR 方法主要使用深度学习算法,对带标注信息的表格数据依赖较大.在草图文档分析的研究中,存在一些与草图表格相关的检测和内容分割工作^[5-6],然而因为它们的研究目标以文档分析为主,并非一般意义上的表格识别方法研究^[7]——文中不包含表格中每一个单元格结构的标签.此外,当前业界公开的表格识别开源数据集也以印刷体表格为主^[8],缺乏专门用于手绘草图表格识别的数据集,因此,使用现存的端到端算法难以有效地解决手绘草图表格的识别任务.同时,由于草图本身具有的模糊性、抽象性和用户在绘制草图时的随意性,手绘草图表格的识别面临着巨大的挑战.面临的挑战具体表现为:准确地识别表格中的复杂单元格、有效地分离表格结构与内部承载的内容信息以及如何克

服用户在手绘草图时伴随的随意性特点.

针对上述挑战,本文提出一种基于结构理解的手绘草图表格识别算法,以数字笔采集到的笔划信息源作为输入,输出相应的标准电子表格.为了验证本文算法,召集了 36 名被试进行用户实验,通过 4 个表格输入任务,将本文算法的识别结果与 2 种基线算法^[4,9]分别进行客观评估与专家评估对比,其中第 1 种基线算法选用由腾讯公司开发的目前在工业界中比较成熟且落地使用的研究工作^[9],第 2 种基线算法选用斩获 ICDAR 2021 表格识别赛道冠军的一项研究工作^[4].客观评估结果显示,本文算法在面向草图表格结构识别任务中的树编辑距离相似度(tree-edit-distance-based similarity, TEDS)^[10]比第 1 种和第 2 种基线算法分别提升了 16.94%和 13.66%;专家评估结果表明,本文算法在表格结构识别和将内容匹配到对应单元格 2 个评价维度中的识别效果均优于基线算法的识别效果.最后,对本文工作过程中的发现和经验进行了讨论.

1 相关研究

1.1 草图识别技术应用

草图被认为是笔式用户界面中一种重要的信息获取手段和信息载体,是一种以捕捉和理解用户的输入意图为目的的表现形式^[11],手绘草图能够以最直接、最自然且最简单快速的方式将构思意图具体化.草图识别与重建技术是用来提取用户输入时的所有构思意图的关键技术.草图识别与重建技术通常使用图形学、机器学习和深度学习技术,来识别手绘草图中的各种几何图形、符号和文本等元素,如字符识别、箭头识别、方框识别、电路图的图元识别等,或将二维或三维的手绘草图进行三维建模等.随着草图识别与重建技术的快速发展,它被广泛应用到计算机辅助设计(computer-aided design, CAD)、图像合成、三维建模等多个应用场景.

例如, Li 等^[12]提出了一种名为 Free2CAD 的系统, 它可以将手绘草图解析成 CAD 命令, 从而实现快速建模; 其采用强大的 Transformers 神经网络架构, 训练了一个端到端的神经网络模型, 以将用户绘制的手绘草图转化为 CAD 命令. 该研究对于加速 CAD 建模, 特别是对于非专业用户的快速建模具有重要的意义. 此外, 草图识别技术也应用于图像合成场景中, Xian 等^[13]提出了一种通过控制草图结构与纹理色块来合成人们所期望图像的算法, 用户在草图的任意位置上放置一个任意比例的纹理色块, 算法会输出同时结合草图与纹理色块的合成图像. Chen 等^[14]提出了基于生成对抗网络 (generative adversarial networks, GAN)^[15] 的 SketchyGAN 技术, 旨在实现从草图到图像的合成, 从多个视觉角度重构手绘草图, 在保留物体的局部特征的同时, 生成更逼真性与多样性的合成图像; 该研究展示了一种完全自动化的草图数据扩充技术, 这对于计算机视觉领域中的扩充图像数据集具有重要意义. 另外, 三维建模也是草图识别技术的热点应用场景, Lun 等^[16]提出一种从线图形式的二维草图中重构三维物体的方法, 其中多视图卷积神经网络 (multi-view convolutional networks, MVCNN) 将单张草图或多张草图作为输入, 输出代表输入草图三维重建后的密集点云. 纵观草图识别与重建技术在以上各应用场景中的相关研究工作, 可以了解到草图作为一种借助视觉信息来自然、快速地捕捉并描述用户意图的常见表达形式, 目前已经应用到各种各样的应用场景中, 草图对于笔式用户界面是不可或缺的一种输入方式.

1.2 印刷体表格 OCR 算法

表格作为一种有效的结构化数据组织和展现形式被广泛应用在不同应用场景中, 它不仅可以用在草图表格中作为思考交流、灵感梳理和将构思具体化的最自然且直观的媒介形式, 同样表格因为其自身结构化的特点, 也成为各类文档中常见的数据展示与数据统计形式之一, 它可以用于紧凑地存储和展现数据的科技文献、财务报表、报纸杂志等各类电子文档中^[17]. 而表格在具体领域中的应用和其种类、样式是极其复杂多样的, 例如印刷体表格 (即文档中的表格) 依附的文档既包括现代的、电子的文档, 也有历史的、扫描的文档, 且这些数量激增的文档在现实应用场景中常常会处于不同的光线环境中, 以及表格中有不同的行列合并形式等. 随着机器学习方法与深度学习方法逐渐应用到不同的专业领域, 印刷体表格 OCR 已

逐步成为国内外研究者关注的热点研究问题之一. 纵观本领域的研究, 可将相关研究工作分为 2 类研究任务: (1) 探究从电子文档印刷体表格中提取有效的结构化数据; (2) 探究从扫描的纸质文档表格中提取有效的结构化数据.

早期对印刷体表格识别算法研究主要是基于启发式规则的方法^[18-20], 即预先根据先验知识指定一组规则来进行决策, 以便识别出满足具有独特特征的表格结构. 此类方法既可以将文档图像作为输入源, 又可以将 PDF 文档作为输入源. 对于文档图像作为输入源时, Kieninger 等^[18]提出了 T-Rect 算法, 算法首先使用自底向上的方法对文档图像进行连通区域分析, 然后按照预先定义的规则对不同的连通分支区域进行合并, 得到表格中的逻辑结构. Koci 等^[19]将文档页面图像中的版面布局区域以图的形式为表示, 使用遗传算法将表格作为一个子图识别出来. 而以 PDF 文档作为输入源的研究中, Yildiz 等^[20]提出了一个将 PDF 文档作为输入源的 pdf2table 算法, 它利用了 PDF 文件中与文档图像相比外特有的信息来辅助完成表格识别任务, 例如, 文本线条和绘制路径等. 然而, 基于启发式规则的表格识别方法涉及大量的超参数, 而这些超参数与不同表格独有的特征呈强相关, 例如, 表格的种类、结构、复杂程度和显示样式等. 因此泛化性较差, 在不同场景的表格识别中难以获得较高且较稳定的准确率.

尽管本领域的研究主要分为 2 类研究任务, 但其中大都使用文档图像作为识别算法的输入源对象, 因此二者有共通的识别算法, 继而国内外的研究者使用传统的图像处理方法对表格识别算法进行了大量研究. 一种被广泛应用的解决思路为首先通过 Hough 直线检测、腐蚀膨胀等形态学操作或 Harris 角点检测算法来对表格结构提取、划分行列区域, 然后将表格的逻辑结构信息与单元格内的文本内容相结合重构为表格对象^[21-22]. 然而, 传统方法中的特征提取主要依赖人工设计的提取器, 需要具备专业领域知识, 并且还会伴随复杂的调整超参数的过程, 泛化能力及鲁棒性较差.

近年来, 随着深度学习技术迅猛发展, 涌现出许多基于深度学习的方法. 研究者将目标检测、图像语义分割和图神经网络等深度学习方法应用到表格识别任务中. 由于表格本身具有结构化的特点, 因此既可以将行、列作为目标检测的对象, 也可以将行之间和列之间的分隔符作为目标检测的对象. 同样, 这些对象也可以用图像语义分割方法

进行分割^[23]. Siddiqui 等^[24]将表格结构识别任务定义为图像语义分割问题, 使用全卷积神经网络 (fully convolutional networks, FCN) 网络结构进行端到端的识别, 对表格的行、列结构分别进行检测. Tensmeyer 等^[25]提出了一种两阶段表格结构识别模型 SPLERGE, 其由分割模型和合并模型组成, 分割模型又分为行分割模型和列分割模型, 分别用于预测表格结构中的行、列分割; 在合并模型中对分割模型输出的每对邻接网格依次进行检测, 判断它们是否需要合并, 进而得到最终的表格结构识别输出. 而近年来提出的 TSRFormer 模型^[26]遵循的范式也是如此, TSRFormer 先由拆分模块预测出所有行、列的表格分割线, 对所有行、列

线一一求交点后, 生成 $N \times M$ 个单元格, 再由合并模块预测相邻单元格是否需要合并从而恢复出跨多行、多列的单元格. 虽然使用深度学习技术提出的模型能在具体的测试集上得到较好的识别效果, 但是由于深度学习方法是由数据驱动的, 需要大量带标注信息的数据集对模型进行训练, 而当前业界公开的表格识别开源数据集也以印刷体表格为主^[8], 如表 1 所示. 若将深度学习技术应用到手绘草图表格的识别任务中, 则需要采集和人工标注大量手绘草图表格数据作为数据集, 毫无疑问, 这需要巨大的人力、财力成本. 因此, 目前应用于印刷体表格识别任务的相关算法不能直接迁移应用于手绘草图表格识别任务.

表 1 用于表格识别的公开数据集^[8]

数据集	来源	内容	数量	格式	语言
Marmot	e-Books and Citeseer website	tree structure of all document layouts	2 000	bmp, xml	中文、英文
UW3	skew-corrected document	images of page, manually edited bounding boxes of page frame, text and non-text zones, textlines, and words, type of each zone (text, math, table, half-tone, ...)	120	png, xml	英文
UNLV	magazines, news papers, business letter, annual Report, etc.	scanned images of page, bounding boxes for rows, columns and cells, OCR recognized words within the whole page	427	png, xml	英文
ICDAR 2013	European Union and US Government websites	PDF documents, bounding boxes of table, textual content, structure labels, bounding boxes for each cell	150	pdf, xml	英文
ICDAR 2019	modern documents and archival ones with various formats	840 jpgs and xmls including bounding boxes of table in modern documents, 1 149 jigs and xmls including structure labels, bounding boxes of table and bounding boxes for each cell in archival documents	≈200	jpg, xml	英文
PubTabNet	scientific publications included in the PubMed Central Open Access Subset	images of tabular data, textual content, table structure labels in HTML	>568 000	png, json	英文
SciTSR	LaTeX source files	PDF documents, images, textual content, structure labels, bounding boxes for each cell	15 000	png, json	英文
TableBank	Word and Latex documents on the internet such as business documents, official fillings, research papers, etc.	Table Detection: bounding boxes of table	417 234	jpg, HTML	英文
		Table Structure Recognition: table structure labels	145 463		
FinTab	Annual and semi-annual reports, debt financing, bond financing, collection of medium-term notes, short-term financing, prospectus	textual ground truth, structure information and the unit of the table	1 685	pdf, json	中文

此外, 一些研究中还包含对于草图文档中的表格进行检测与内容分割的工作, 它们旨在实现文档中草图表格的定位、内容提取与表格生成任务. 例如, Zhang 等^[5]提出了一种基于矩阵模型的方法, 将包含各种对象, 如文本、表格、图形、流程图等的草图墨迹文档, 将其建模为矩阵模型, 并基于此进行表格定位提取与内容分割, 从而生成电子表格. 然而, 文中提取对象为仅由文字笔迹构成的表格, 对于解决框线与文字内容共存的表格

还存在欠缺, 尤其是框线与内容存在重叠遮挡的表格. Liu 等^[6]提出了一种基于图注意力网络的多任务学习方法, 用于在草图在线文档中实现表格定位检测与内容分割. 需要注意的是, 这些研究主要关注对于草图墨迹文档的分析理解, 完成表格的定位检测与内容分割, 从而实现表格的提取与生成任务. 它们不包含构成表格的所有单元格属性结构的标签, 因此, 不属于一般意义上的表格识别任务^[7].

1.3 面向特殊应用场景的草图结构识别算法

草图识别算法面临的首要核心问题在于对草图内复杂图形结构的理解,这需要在对图形识别的基础上应用一定的背景信息,如领域知识、用户信息等,或借助其他手段来赋予草图真实的含义或结构,将用户输入的承载具体化构思的手绘草图图形映射为某个应用领域内具有特定含义的对象,从而达到捕捉用户绘制意图的目的^[11]。当前大部分研究者都集中在某个实际的应用场景,结合领域知识以达到理解草图对象的目的。因此,对面向的不同应用场景的草图表现形式,研究者提出了众多基于结构理解的草图识别算法。

例如,面向在线或离线的手写公式应用场景的草图识别正在受到国内外研究者的广泛关注。Zhelezniakov 等^[27]在一篇综述性文章中介绍了手写数学公式(handwritten mathematical expressions, HME)识别的研究现状和发展趋势,对手写数学公式识别的现状进行了全面的梳理和总结。首先介绍了在线或离线手写数学表达式识别面临的挑战和各种应用场景;其次从笔迹预处理、符号分割与分类、结构分析、端到端识别等方面概述了现有的技术和方法,并介绍了在线手写数学公式识别的经典工作流程。Zhao 等^[28]提出的手写数学公式识别模型双向 Transformer,基于 Transformer 模型和双向 LSTM 模型,以手写数学公式的图片作为输入源形式,将其转换为一系列语义标记,并利用双向训练的 Transformer(bidirectionally trained Transformer, BTT)模型将这些标记转换为数学公式的 LaTeX 表达式。

此外,对于手绘结构网状图的草图识别也是一大研究热点,不限于电路图、组织结构图、数据流图和算法流程图。Rachala 等^[29]提出了一种基于目标检测和节点识别的手绘电路图识别方法。首先,利用 YOLOv5 模型进行电路元器件目标物体检测,得到电路图中各个元器件的位置和大小信息;然后,使用节点识别方法—Hough 变换和 k -means 聚类方法来检测电路元器件间的连接情况,并将节点信息转化为电路拓扑图;最后,利用电路拓扑图进行电路分析和仿真,实现手绘电路图的自动化识别和分析。

可见,研究者主要研究面向特殊应用场景的草图结构识别理解算法,如手写数学公式等公式结构图应用场景、模拟电路图和流程图等结构网状图应用场景。而目前关于草图表格这一应用场景的相关研究还有待进一步探索,并且值得关注

的是,尽管通过笔式用户界面采集到的草图既有笔迹信息,又有图像信息,但大多数研究工作选择笔迹信息作为算法的输入源对象。因为在结构理解的场景下,笔迹信息比图像信息更丰富,它可以集成更多用户在绘制草图时产生的交互信息,如笔迹的时序、采样点的空间坐标和采样密度等^[30]。因此,在本文中,笔迹信息也同样被选择为算法的输入源对象之一。

2 本文算法

针对上文中提到的手绘草图识别存在的 3 个挑战,本文的草图识别算法需要实现以下 3 个目标:

(1) 面对不同的输入信息,手绘草图表格的种类复杂多样,表格中可能存在不同的行列合并形式,例如存在复杂单元格,即跨行跨列的单元格等,如图 1a 所示。

(2) 手绘草图表格的结构与承载的内容信息可能存在重叠区域,如图 1b 所示。

(3) 手绘草图表格的过程是随意且模糊的,主要体现在用户在绘制过程中存在重描、补笔等现象,如图 1c 所示。

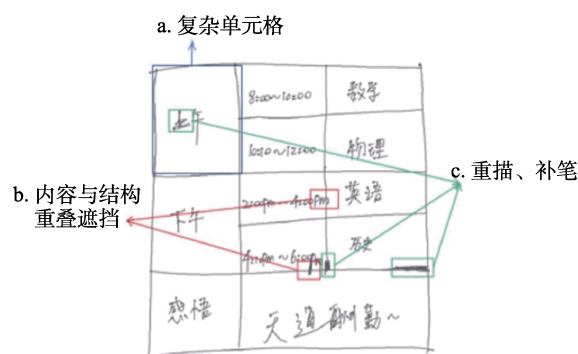


图 1 用户手绘的草图表格

本文基于结构理解的手绘草图表格识别算法主要步骤如下:

Step1. 对采集到的笔划数据进行多笔划的二分类,将笔划划分为表格框架结构类别和内容类别,如图 2a 所示。

Step2. 使用表格旋转矫正算法对整体倾斜角度较大的表格进行旋转矫正,如图 2b 所示。

Step3. 对笔划分类为表格框架结构的笔划进行距离均匀的重采样预处理,如图 2c 所示。

Step4. 使用多峰值检测、真实交点找寻和单元格特征属性检测确定表格结构,如图 2d 所示。

Step5. 对表格内容进行区域定位和内容识别,最

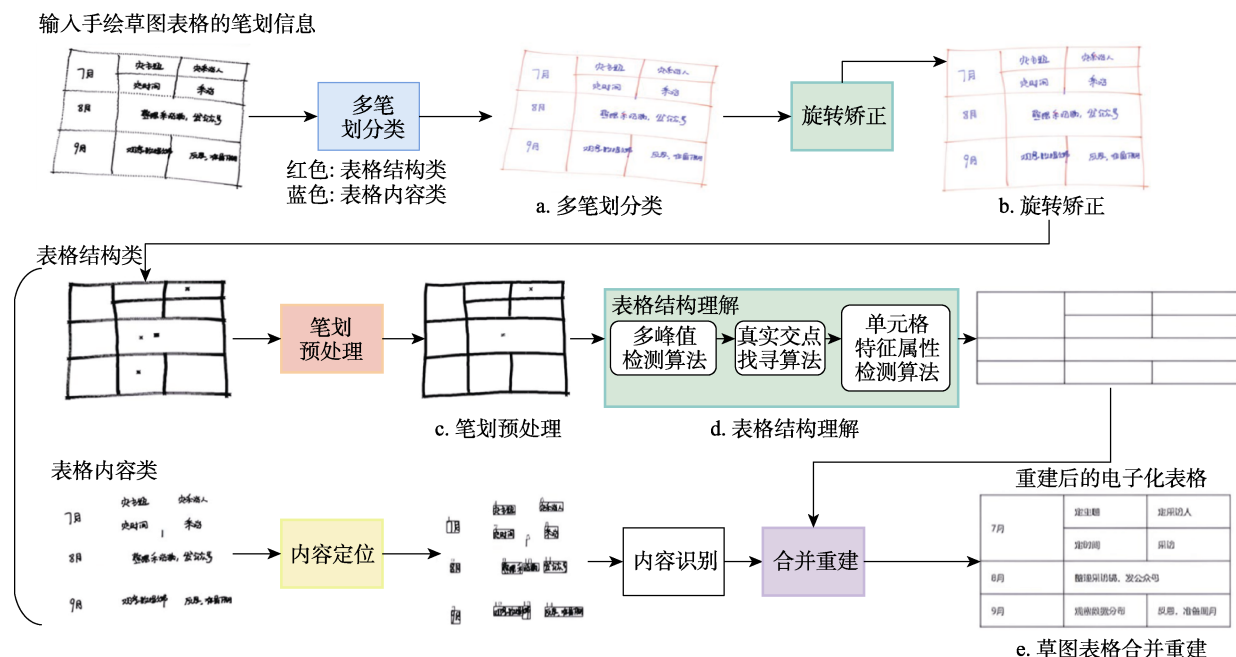


图 2 基于结构理解的手绘草图表格识别算法流程

后将识别内容合并、重建到表格的结构中,如图 2e 所示。

通过 Step1, 本文算法利用分类准确率^[31]为 94%的支持向量机(support vector machines, SVM)分类模型对多笔划分类,以有效地区分草图绘制时得到的结构笔划组合与内容笔划组合,可以实现目标(2)的要求。通过 Step2~Step4, 本文算法首先利用表格旋转矫正算法、笔划重采样预处理操作为后续确定表格结构提供有效且不额外引入噪声的笔划输入源数据;其次利用多峰值检测算法确定表格的横、竖线数目与位置;最后结合真实交点找寻和单元格特征属性检测以确定表格逻辑结构,即表格的行列数、表格中各单元格的属性(各单元格的所处相对位置、各单元格是否为复杂单元格以及分别跨越几行几列等),可以实现目标(1)的要求。通过 Step1~Step5, 本文算法利用多笔划分类、旋转矫正与笔划重采样预处理、表格结构理解确定、内容区域定位、内容识别与将识别内容合并重建到表格结构中,完成草图表格到电子表格的重建,可以实现目标(3)的要求。

Step4 是本文算法的关键所在。由于表格结构识别的关键在于确定单元格的跨行跨列属性,而每个单元格是由 4 个顶点构成,它们分别为 2 对横、竖线的交点。因此, Step4 中首先通过多峰值检测算法确定表格的横、竖线数目与相对位置,确定表格结构中由横、竖线相交得到的每一个二维交点的特征属性;其次,通过真实交点找寻算法从“交点候选集”中筛选出真正作为某个单元格顶点的交

点,确定真实交点的属性;最后,通过单元格特征属性检测算法确定表格逻辑结构,包括表格的行列数目、表格中各单元格的属性,其中单元格属性包括构成单元格的横、竖线、单元格跨越的行列数以及单元格的类型(仅跨越一行一列的简单单元格或跨越多行多列的复杂单元格)。

2.1 多笔划分类

当用户在笔式用户界面中手绘草图表格时,由于草图绘制过程中伴随的非顺序性、重描或多笔等特点,导致采集到的手绘草图表格的笔划顺序是不确定的。具体表现为: (1) 用户在手绘草图表格时,可能没有严格的先后顺序来依次绘制草图表格结构与内容; (2) 尽管有用户习惯先绘制表格结构,再绘制表格内容,但同样存在当内容绘制结束后,再去对表格结构进行“添补”笔划的现象。所以本文提出的草图表格识别算法首先根据用户绘制的多笔划特征来实现多笔划二分类。值得一提的是,如果用户绘制的草图表格结构笔划与内容笔划产生了一定程度的重叠,采用笔划作为输入源对象进行处理可以有效地避免后续识别中的干扰。相比之下,现有以图像作为输入源的研究工作在面对用户绘制的草图表格结构笔划和内容笔划重叠时,分离二者是一个极具挑战性的问题。但是,本文以笔划作为输入源对象时,这个问题便迎刃而解了。

本文使用 SVM 分类模型^[32-34]这一机器学习分类算法中的典型算法,对多笔划进行二分类,

以区分草图绘制时得到的结构笔划组合与内容笔划组合。

具体来说, 本文用于训练 SVM 二分类模型的草图笔划数据集构造过程分为数据采集、数据预处理、特征提取与数据标注步骤。

首先, 在草图笔划数据集的采集过程中, 本文使用支持草图绘制的笔式用户界面作为交互载体, 以数字笔作为输入工具, 以每条笔划轨迹(Stroke)作为数据采集的单位。其中, 每条笔划轨迹是指一条手写笔划的运动轨迹, 有其起始点和终止点, 它们通常代表手写中用户在抬起输入设备和落下输入设备之间一个连续的运动过程。一条笔划通常由多个采样点组成, 它们记录了数字笔的二维坐标位置、时间戳和笔的压力等信息。而根据采样方式的不同分为以时间均匀采样的笔划轨迹和以距离均匀采样的笔划轨迹^[35]。本文共采集了 3 046 个以时间均匀采样的笔划轨迹数据样本, 其中表格框线笔划轨迹的数据样本数量为 1 422, 表格承载的内容笔划轨迹的数据样本数量为 1 642, 这 2 类数据样本的比例为 1 : 1.15。

其次, 在从笔划轨迹数据中提取特征之前, 应

经过一些轨迹预处理的过程, 使得提取的特征对于分类任务更加有效。在数据预处理的过程中, 本文是针对数据集中的每个笔划轨迹数据样本进行的。预处理过程涉及 2 个操作, 分别是对笔划轨迹按距离均匀进行重采样和将笔划轨迹起始点平移到原点。平移到原点操作在本文中采取的做法是将笔划轨迹的起点移动到坐标系原点, 使得笔划轨迹的起始点在(0,0)处。

接下来, 特征提取过程需要选择合适的特征属性子集, 特征提取的过程是机器学习和模式识别领域分类任务中的必要阶段。其主要目的为筛选有助于分类的特征属性与排除不相关的特征属性, 降低特征属性的维度, 并尽量提高决策分类的准确性。本文对采集的数据集中每条笔划轨迹数据提取的特征分为 2 大类, 分别为笔划轨迹数据组成的采样点特征(二维坐标位置、时间戳、速度、加速度)的统计学特征(求和 sum、均值 mean、标准差 std、最大值 max、最小值 min、中位数 median、下四分位数 Q_1 、上四分位数 Q_3)和笔划轨迹的长度, 如表 2 所示。最终, 组成数据集的每条笔划轨迹数据经过特征提取阶段后的特征总维度为 57。

表 2 草图笔划数据集中的特征提取

特征	特征维度	基于特征的描述
Point_x(sum, mean, std, min, max, median, Q_1 , Q_3)	8	二维坐标位置的统计学特征
Point_y(sum, mean, std, min, max, median, Q_1 , Q_3)	8	二维坐标位置 y 的统计学特征
Timestamp(sum, mean, std, min, max, median, Q_1 , Q_3)	8	时间戳的统计学特征
Speed_x(sum, mean, std, min, max, median, Q_1 , Q_3)	8	速度 x 的统计学特征
Acceleration_x(sum, mean, std, min, max, median, Q_1 , Q_3)	8	加速度 x 的统计学特征
Speed_y(sum, mean, std, min, max, median, Q_1 , Q_3)	8	速度 y 的统计学特征
Acceleration_y(sum, mean, std, min, max, median, Q_1 , Q_3)	8	加速度 y 的统计学特征
PathLength	1	笔划轨迹的长度特征

最后, 本文对采集的数据集进行人工标注, 其策略是基于笔划轨迹所属的类别(表格框线或表格承载的内容)进行标记: 属于表格框线类别的所有笔划轨迹都标记为类别 0, 属于表格承载内容类别的所有笔划轨迹都标记为类别 1。随后, 采用一种称为交叉验证^[36]的统计方法来将本文构造的数据集中的数据分为 2 部分, 80%用于模型训练, 20%用于模型测试。

基于 SVM 模型在本文构造的草图笔划训练集中训练的结果, 在测试集中测试得到的准确率(Accuracy)^[31]为 94%。图 3 所示为对手绘草图表格应用多笔划二分类子算法模块后的可视化效果。

其中图中红色表示预测为表格结构的笔划组合, 蓝色表示预测为表格内容的笔划组合。

2.2 草图表格旋转矫正

用户在手绘草图时具有随意性的特点, 这导致用户绘制的草图表格整体可能存在一定程度的倾斜, 倾斜角度或大或小, 受用户手绘草图时的习惯、绘制时的周围环境等因素影响, 因此需要采用旋转矫正算法对整体倾斜角度较大的草图表格进行整体的旋转矫正, 以减少倾斜角度过大对后续表格结构和内容识别产生的噪声干扰, 进而提升本文算法的识别效果。

为了矫正用户绘制草图表格时存在的整体倾

7月	灾主题	灾采访人
8月	灾时间	采访
9月	整理采访稿, 发公众号	
	观察数据分布	反思, 准备汇报

图 3 多笔划二分类可视化图

斜, 本文通过计算多条表格框线结构笔划的倾斜角度得到表格整体倾斜角度。需要注意的是, 文中提及的倾斜是一个矢量, 既包含倾斜角度, 又包含倾斜方向。以倾斜角度的形式化表述为例, 如果手写笔划作为竖线向逆时针方向倾斜了 15° , 则其倾斜角度的形式化表述值为 -15° , 经过应用旋转矫正算法后, 组成此竖线的每个坐标点会沿顺时针方向旋转 15° , 如图 4a 所示; 如果手写笔划作为横线向逆时针倾斜了 15° , 则其倾斜角度的形式化表述值为 -15° , 经过应用旋转矫正算法后, 组成此横线的每个坐标点会沿顺时针方向旋转 15° , 如图 4b 所示。

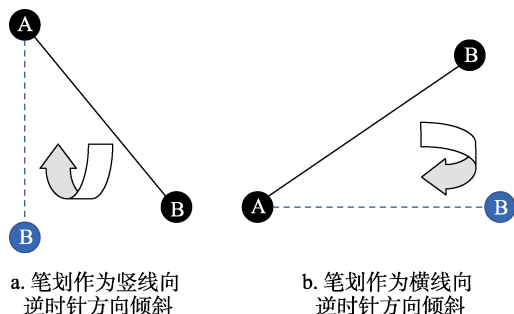


图 4 笔迹倾斜与旋转过程举例

基于以上假设, 本文提出的草图表格旋转矫正算法包含 2 个步骤。

Step1. 需要根据表格框线结构的每条笔划的倾斜角度计算得到表格所有结构笔迹的倾斜角度, 以此代表表格整体的倾斜角度。

Step2. 需要对表格的所有笔划中的每个轨迹点绕以某一选定中心点进行旋转, 旋转方向为计算得到的反倾斜方向, 旋转值为计算得到的倾斜角度值。

对于旋转矫正算法的 **Step1**, 需要计算表格框线结构中每条笔划的倾斜角度, 并统计所有结构笔划的倾斜角度均值, 将其作为整个草图表格的

倾斜角度 α 。由于用户在绘制每条笔划时存在随意性和不确定性, 每条笔划是否发生倾斜以及倾斜的程度不能仅由几个二维轨迹点(如笔划轨迹的起点和终点)反映。因此, 本文通过计算每条笔划中相邻的二维轨迹点之间的倾斜角度来得到每条笔划的倾斜角度。为了减少离群点对倾斜角度的值产生负面干扰, 使用四分位点内距(interquartile range, IQR)来排除异常值, 并对剩余的相邻轨迹点对的倾斜角度取平均值来得到此笔划的倾斜角度。最后, 计算所有结构笔划倾斜角度的均值, 以此代表此草图表格整体的倾斜角度。

对于旋转矫正算法的 **Step2**, 要对表格的所有笔划中的每个轨迹点绕以某一选定旋转中心点进行旋转, 旋转方向为 α 的反方向, 旋转值为 $|\alpha|$ 。依次旋转后得到的所有笔划点组成的表格即为整体旋转矫正后的草图表格, 将作为本文后续结构和内容识别算法模块的输入。

2.3 笔划重采样

为了使得草图结构理解模块的实现效果达到最大程度的高鲁棒性, 需要先对从笔式用户界面中采集到的各原始笔划进行重采样的预处理工作。因为在用户手绘草图表格时采集到的每条原始的笔划数据信息是基于时间均匀采样的, 其适用于还原速度、加速度、急动度等运动学场景中的属性; 而后续表格结构理解模块的首要步骤是准确地找到目标表格中每行每列的特征点, 这在很大程度上依赖空间位置结构信息。所以对采集到的每条原始笔划在保证原始结构信息不丢失的前提下, 通过指定点与点之间的均等距离间隔与使用二维线性插值方法进行了距离均匀的重采样^[35], 为后续表格结构理解识别算法提供了有效且不额外引入噪声的输入源信息数据。

对于结构笔划组合应用基于距离均匀的重采样子算法模块前后的可视化效果如图 5 所示, 本文中设定的均等距离间隔为 1 像素, 可以明显看到重采样后的各笔划的采样点之间的距离都是均等的。

2.4 草图表格结构理解

在表格结构理解与识别的子算法模块中, 输入为预处理后的结构笔划组合, 即对多笔迹分类得到的各结构笔划应用距离均匀的重采样子算法, 输出为识别得到的表格逻辑结构, 即表格的行列数、行列之间的逻辑关系、行列相交得到的各单元格的属性, 其中包括各单元格的所处相对位置、各单元格是否为复杂单元格以及分别跨越几行几列等。

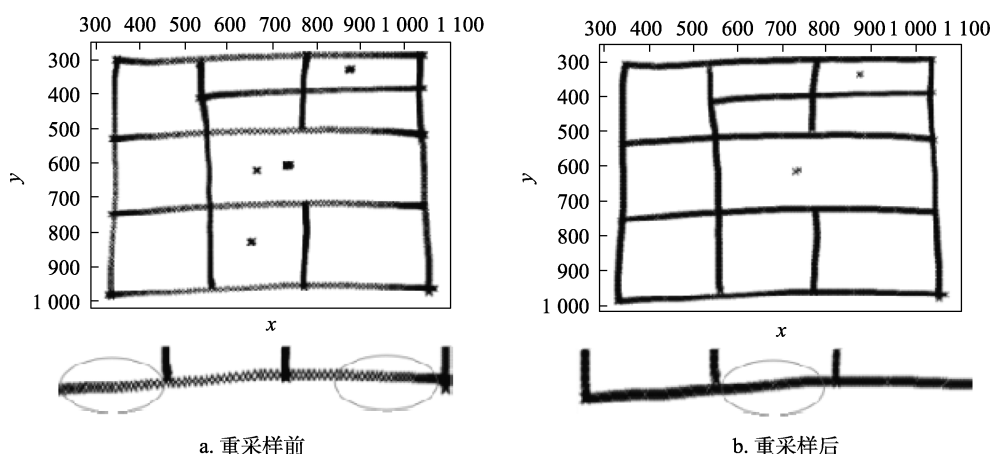


图 5 重采样前后的可视化图

假设用户绘制的草图表格是标准的全线表格, 具体表现为表格外边框与内部的框线均存在、线与线之间预期存在的交点也均存在. 基于此假设, 得到以下基本结论: (1) 表格是由一个或多个单元格组成. (2) 每个单元格是由 2 对横、竖线对应相交产生的, 表现为每个单元格有 4 个顶点(不管是否为复杂单元格), 即分别为 2 对横、竖线的交点. 因此对于表格结构理解子算法模块的核心思路为确定每个单元格的 4 个真正顶点, 通过检查 4 个顶点的相对位置得到对应单元格的属性, 即是由哪 2 对横、竖线组合产生的、对应单元格的所处相对位置, 以及是否为复杂单元格等属性.

首先, 确定表格结构中由横、竖线相交得到的每一个二维交点的特征属性, 如每个二维交点的二维位置坐标以及分别是由哪 2 条横、竖线相交而产生的. 具体来说, 对预处理后的各结构笔划中的每个二维点分别投影到 x 轴和 y 轴的, 以生成 2 个频数直方图. 接着, 通过多峰值检测算法可以在 x 轴和 y 轴的频数直方图中找到每个轴上的多个一维特征值, 此过程的可视化效果如图 6 所示, 其中图 6a 所示为用户手绘的草图表格结构, 图 6b 所示为将每条结构笔划中的所有二维点向 x 轴投影后生成的频数直方图, 图 6c 所示为将每条结构笔划中的所有二维点向 y 轴投影后生成的频数直方图.

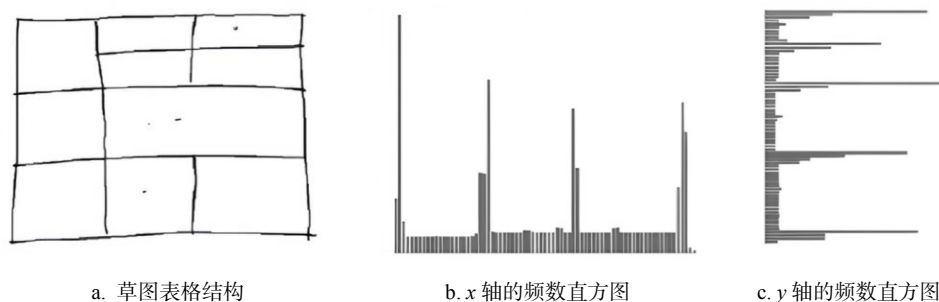


图 6 找表格结构中所有交点

多峰值检测算法(以 x 轴为例)步骤如下:

输入. x 轴上的频数直方图 x_hists .

输出. x 轴上的多个一维特征值对应的下标 $x_features$.

$mean_hist$ = 频数直方图的均值.

Step1. 对直方图的各项频数做均值滤波以达到数据降噪目的:

$filt_length = 5$ # 滤波窗口取值为 5.

$x_hists = ava_filter(x_hists, filt_length)$.

Step2. 找多个局部极大值下标处:

FOR $index$ IN x_hists THEN

CASE $index=0$:

IF $x_hists[0] > x_hists[1]$ and
 $x_hists[0] > mean_hist$ THEN
 $x_features.append(0)$.

CASE $index=[1, x_hists.len-2]$:

IF $x_hists[index] > x_hists[index-1]$ and
 $x_hists[index] > x_hists[index+1]$ and
 $x_hists[index] > mean_hist$ THEN
 $x_features.append(index)$.

CASE $index=x_hists.len-1$:

IF $x_hists[index] > x_hists[x_hists.len-2]$
and $x_hists[index] > mean_hist$ THEN
 $x_features.append(x_hists.len-1)$.

END FOR

Step3. 将相邻的局部极大值下标进行合并:

titl_value = 5 # 假设一条直线的倾斜度最多相隔 5 个柱子.

```
FOR i IN x_features THEN
  FOR j IN x_features THEN
    IF j - i <= 5 THEN
      x_features.remove(j).
    END IF
  END FOR
END FOR
```

但值得关注的是, 由于复杂单元格的存在, 表格结构中真正的交点数目并非一定等于分别向 *x* 轴和 *y* 轴投影的 2 个频数直方图中得到的特征值数目(即频数直方图中峰值的个数)的乘积. 因此, 需要通过真实交点找寻算法来寻找可以作为某个单元格顶点的真实交点. 具体来说, 将通过 2 个频数直方图中得到的多个一维特征值依次组合形成的二维点集称为交点候选集. 此外, 由于用户在手绘草图表格时草图本身的特点和用户绘制的随意性, 表格结构的框线(横线或竖线)是持一定倾斜度的, 因此本文认为, 组成草图表格框线结构的每条框线笔划都可以等效于一个矩形(即线条轨迹的最小外接矩形). 接下来, 通过判断 2 个矩形是否存在相交区域来从交点候选集中筛选出真正可以作为某个单元格顶点的交点, 并记录每个真实交点的属性信息.

真实交点找寻算法步骤如下:

输入. 交点候选集 *point_candidates*.

输出. 真实交点(能作为某单元格顶点)的集合 *points*.

- *getRecLine()* 的输入为某二维点, 输出为承载它的横线与竖线(横竖线均用矩形框表示, 并且因为并非真实的交点, 因此横线或竖线可能不存在).

- *rec_isCross()* 的输入为 2 个矩形, 输出为 *true/false*, 表示这 2 个矩形是否相交.

```
FOR i IN point_candidates THEN
  recHorLine, recVerLine = getRecLine(i).
  IF recHorLine is NULL or recVerLine is NULL
  or !rec_isCross(recHorLine, recVerLine) THEN
    point_candidates.remove(i).
  END IF
END FOR
```

最后, 遍历所有交点, 通过在其邻域范围内是否能找到某个单元格的 4 个顶点的标准来判断真正的单元格, 并确定每个单元格的属性, 如是由哪 2 对横向和竖向相交形成的、所处位置以及是否为复杂单元格. 正因为表格中可能存在复杂单元格, 并非在一定邻域范围内首次找到的交点就能作为某个单元格的 4 个顶点之一. 如图 7 所示, 其中通

过表格结构理解子模块识别得到的第 1 个复杂单元格区域内相关交点有 5 个, 但是只有 4 个交点是组成此单元格的真正顶点. 因此, 还需要使用单元格特征属性检测算法来进一步识别每个单元格.

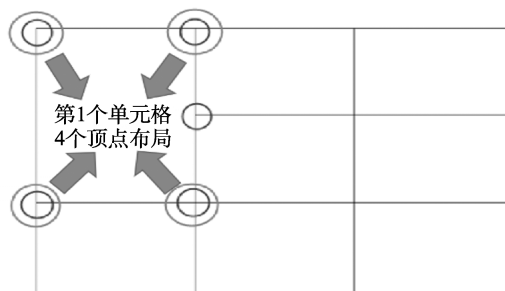


图 7 交点与组成单元格顶点示例

单元格特征属性检测算法步骤如下:

输入. 以行为分组依据的所有交点组合列表 *row_points*, 以列为分组依据的所有交点组合列表 *col_points*.

输出. 各单元格属性, 每个单元格的属性表示为 $[x_1, y_1, w, h, [n_1, n_2]]$.

- 输出的各元素含义: x_1 与 y_1 为单元格左上角顶点的坐标值; w 和 h 为单元格的宽度和高度; $[n_1, n_2]$ 中 n_1 表示此单元格跨越的行数, n_2 表示跨越的列数.

- 某单元格的真正左上角顶点表示为 A , 某单元格的右上角顶点表示为 B .

- *getNextPointInSameCol()* 的输入为交点 M , 输出为 X 点所在列相邻的下一个点 N .

- *getRowLoc()* 的输入为交点 M , 输出为 M 点所在行的位序.

- *getColLoc()* 的输入为交点 M , 输出为 M 点所在列的位序.

```
FOR A IN row_points THEN
   $A' = \text{getNextPointInSameCol}(A)$ .
  FOR B IN row_points THEN
     $B' = \text{getNextPointInSameCol}(B)$ .
    IF  $B'$  is not NULL THEN
       $x_1, y_1 = A.x, A.y$ ,
       $w = B.x - A.x$ ,
       $h = \text{Max}(A'.y, B'.y) - A.y$ ,
       $n_1 = \text{Max}(\text{getRowLoc}(A'), \text{getRowLoc}(B')) - \text{getRowLoc}(A)$ ,
       $n_2 = \text{getColLoc}(B) - \text{getColLoc}(B')$ .
    ELSE
      CONTINUE
    END IF
  END FOR
END FOR
```

2.5 草图表格内容定位和识别

本文中草图表格文本内容定位和识别模块主要包括 2 个步骤: (1) 文本内容区域块检测定位: 对文本内容出现的区域位置进行定位. (2) 文本内

容识别: 基于定位好的文本内容区域进行识别, 将图像中的手写文字内容转化为识别后的文本信息. 文本内容定位检测算法的输入源为经过多笔划分类算法和旋转矫正算法后的内容笔迹, 输出为识别到的多文本矩形区域位置和对多区域进行切割得到的多图像, 作为第 2 步手写文本识别任务的输入源数据. 文本内容检测算法参考 Jiang 等^[37]提出的结构理解方法, 它采用动态规划方法来提取连线笔迹块内容, 以检测到所有的文本内容区域块. 文本内容区域块检测定位的效果可视化如图 8 所示.

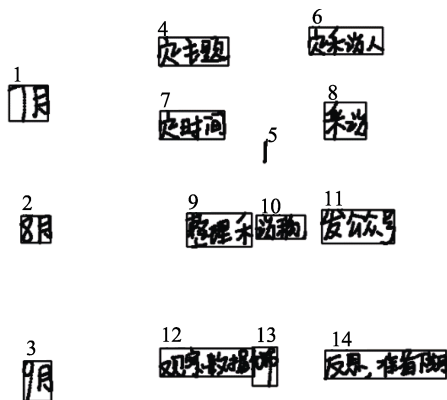


图 8 文本内容区域块检测定位结果可视化图

需要说明的是, 本文的研究重点在于草图表格的结构理解, 为了使得草图表格识别整个流程更加完整, 对于文本内容识别模块, 本文使用了目前已有的比较成熟且落地使用的研究工作——科大讯飞提供的基于卷积循环神经网络(convolutional recurrent neural network, CRNN)模型的公共 API 接口^①, 它可以将图片中的手写字体转化为计算机可编码的文字, 用于端到端地对不定长的文本行序列进行识别, 接口对中英文均提供支持.

2.6 草图表格合并重建与可视化

在表格结构识别和内容识别完成后, 将识别得到的各文本内容区域块内的文本内容合并到表格结构理解子算法模块输出的数据结构对应位置中, 即需要将文本内容区域块与单元格匹配映射. 本文实现文本内容区域块与单元格匹配的核心思路为判断文本内容区域块的中心点是否在结构识别得到的某单元格内. 当表格内所有的文本内容区域块均匹配到对应单元格时, 草图表格合并重建完成.

最终需要向终端用户进行识别重建结果的输出反馈, 使得识别后的表格对终端用户可见. 可以选用 Excel 或超文本标记语言(hypertext markup language, HTML)等常见可视化形式对带有复杂单元格的表格进行可视化展示. 本文使用 HTML 对表格进行可视化展示. 因此对于本文提出的基于结构理解的手绘草图表格识别算法的最后一个子模块, 以已被填充好的表格数据结构对象作为输入, 自动向脚本中写入 HTML 语言, 以经过浏览器渲染后的 HTML 表格作为输出. 举例草图表格识别后的可视化效果如图 9 所示.

7月	定主题	定采访人
	定时间	采访
8月	整理采访锦, 发公众号	
9月	观察数据分布	反思, 准备周月

图 9 草图表格识别后的可视化图

3 用户研究和实验评估

3.1 用户研究实验

为了验证本文提出的基于结构理解的手绘草图表格识别算法的效果, 本文发起了一项用户研究实验. 招募了 36 名受试者, 其中包括 24 名男性和 12 名女性, 年龄在 20~40 岁, 受试者具有长期使用笔的经验, 熟悉草图绘制和习惯使用表格草图来整理思路.

在用户研究实验正式开始前, 受试者的任务为熟悉手写采集数字笔设备与用于草图采集的笔式用户界面, 以便完成后续的实验采集任务. 在用户研究实验中, 每名受试者被要求完成 2 类实验任务(共 4 个实验任务). 第 1 类(任务 1~3)为指定的实验任务, 旨在尽可能引导受试者绘制多种类型的表格, 其中包含 3 个具体的场景任务描述, 每个任务都给出一段在特定场景中需要转化为草图表格的文字描述信息, 每个受试者在充分理解每个任务描述后, 在用于草图采集的笔式用户界面中绘制出一个承载有具体的任务描述数据元素的草图表格. 需要说明的是, 本文并不要求不同的受试者在面对同一个任务场景时绘制结构完全相同的草图表格, 这给了受试者很大的自主性与灵活性. 第 2 类(任务 4)为非指定的实验任务, 旨在给受试者最高程度的自主性和灵活性, 每个受试者可以随

① <https://www.xfyun.cn/doc/words/wordRecg/API.html>

心所欲地绘制任何想要表达的草图表格。综上, 在用户研究实验阶段, 每名受试者分别完成 4 个实验任务, 每个实验任务输出一个绘制的草图表格, 每个受试者的实验过程将持续 20 min。本文在实验阶段共收集了 144 个具有不同类型、不同样式、反映受试者不同绘制习惯的草图表格。

在第 1 类指定的实验任务中, 按照依次递进的关系对 3 个具体的场景任务描述进行设计定义, 以引导受试者绘制多种类型的表格, 例如, 简单结构的表格(表格由仅跨越一行或一列的不可再分割的单元格组成)、存在跨列单元格的复杂结构表格。其中, 任务 1 为假设受试者要组织一场交友活动, 初步思考与记录参加的人员信息, 请受试者按照提供的文字描述信息将其绘制成表格的形式, 而给出的文字描述信息中的各属性间不存在交叉关系, 因此用简单结构的表格可以对其进行充分的表示。任务 2 为假设受试者作为一名学生, 根据给出的文字描述中一天内的时间安排为自己绘制一张课程表表格, 本任务中给出的文字描述信息属性间存在交叉关系, 例如, 上午包含 2 门课程, 因此需要用存在跨行或跨列复杂单元格的复杂结构表格来表示。任务 3 为假设受试者是一名计划将自己的一项工作发表论文的学生, 请受试者对论文投稿前 3 个月的各项任务进行自由安排并绘制论文投稿冲刺时间计划表, 本任务给了受试者在绘制表格时更大的灵活性, 可以选择用简单结构表格来表示, 也可以选择用复杂结构表格对其表示。

3.2 基线算法的选取

由于目前在草图结构理解领域暂无对草图表格这一应用场景的研究, 因此在验证算法效果时, 主要对比的基线为目前被众多研究者青睐、利用深度学习技术应用于印刷体表格的 OCR 算法。在印刷体表格识别任务中, Hu 等^[7]将此任务分为表格检测和表格识别 2 个子任务, 其中表格检测任务是从一个页面中检测出表格所在的区域, 也称为版面分析。表格识别任务是在表格检测任务检测到的表格区域基础上, 进一步地对表格的结构与内容进行识别, 得到表格的行列和单元格等内部逻辑结构信息, 并通过 OCR 技术对单元格中的内容进行字符识别, 最终用不同的可视化形式进行表格的重构。因此, 结合本文的研究内容与现有相关研究, 本文把验证的重点放在表格识别任务中。

首先, 本文选用工业界目前已有的比较成熟且落地使用的研究工作——由腾讯公司开放的表格识别 API 接口^[9]作为本文的第 1 个基线算法, 后

文将此基线算法称为“基线算法-腾讯 OCR”。该算法的输入支持中英文图片和 PDF, 输出为每个单元格的文字内容以及将表格重建结果的 HTML 可视化展示。该算法首先对表格图片应用深度学习进行图像分割, 进而对分割图分别做几何分析得到框线, 根据框线计算出每个单元格在图中 4 点坐标的位置(top_left, top_right, bottom_left, bottom_right); 其次通过对表格图片调用 OCR 以识别其中的文本内容, 以及每个字符的坐标; 最后将单元格位置与字符坐标进行匹配, 决定每个字符在哪个单元格中。

其次, 本文选用斩获 ICDAR 2021 表格识别赛道冠军的研究工作^[4]作为本文的第 2 个基线算法。其提出一种名为 LGPMA 的表格结构识别框架, 其中采用了本地和全局金字塔掩模学习方法, 并引入了一种统一的表格结构识别流程, 包括单元格匹配、空单元格搜索以及空单元格合并等步骤。这些步骤可以有效地定位和分割非空单元格和空单元格。该研究在多个基准测试上的表现具有竞争力甚至领先水平的结果。后文将此基线算法称为“基线算法-LGPMA”。

值得注意的是, 本文算法以表格绘制过程中采集到的笔划作为输入源, 而选用的基线算法以表格图像作为输入源。因此, 本文将在用户研究中使用采集程序得到的所有笔划轨迹进行可视化, 输出保存为图片以适应基线算法的输入。

3.3 客观评估

为了对比本文提出的基于结构理解的手绘草图表格识别算法和目前已有的研究工作中较优基线算法对草图表格的识别效果, 以 TEDS^[10]作为客观评价指标, 而由于本文的研究重点在于对草图表格的结构识别, 因此本文使用的客观评价指标计算步骤为: (1) 将表格结构以 HTML 标记形式的树状结构表示; (2) 通过 TEDS 公式计算 2 个表格结构之间的树编辑距离, 该距离表示将一个表格结构转换为另一个表格结构所需的最少操作数(插入、删除和替换节点)。

本文基于用户研究实验阶段采集得到的 144 个草图表格数据, 对不同算法基于 TEDS 指标进行定量对比评估。结果表明, 与基线算法-腾讯 OCR^[9]相比, 本文算法在面向草图表格结构识别任务中的 TEDS 指标提升了 16.94%; 与基线算法-LGPMA^[4]相比, 本文算法在面向草图表格结构识别任务中的 TEDS 指标提升了 13.66%。具体来说, 本文算法在所有实验任务上对于表格结构识别任务的平均相

似度为 0.967 6(SD=0.102 2), 基线算法-腾讯 OCR 在所有实验任务上对于表格结构识别任务的平均相似度为 0.827 4(SD=0.197 5), 基线算法-LGPMA 在所有实验任务上对于表格结构识别任务的平均相似度为 0.851 3(SD=0.149 6)。

接下来, 本文以用户研究实验中的 4 个实验任务作为自变量, 分别统计本文算法和 2 种基线算法在每个实验任务中对表格结构识别的 TEDS, 具体指标如图 10 所示。在引导用户绘制简单结构草图表格的任务 1 中, 本文算法对表格结构识别任务的平均相似度为 0.995 6(SD=0.026 7), 基线算法-腾讯 OCR 对表格结构识别任务的平均相似度为 0.840 9(SD=0.186 0), 基线算法-LGPMA 对表格结构识别任务的平均相似度为 0.951 7(SD=0.090 1); 在引导用户绘制复杂结构草图表格的任务 2 中, 本文算法对表格结构识别任务的平均相似度为 0.965 3(SD=0.078 6), 基线算法-腾讯 OCR 对表格结构识别任务的平均相似度为 0.811 8(SD=0.158 9), 基线算法-LGPMA 对表格结构识别任务的平均相似度为 0.782 8(SD=0.128 9); 在引导用户绘制简单或复杂结构草图表格的任务 3 中, 本文算法对表格结构识别任务的平均相似度为 0.952 5(SD=0.160 1), 基线算法-腾讯 OCR 对表格结构识别任务的平均相似度为 0.799 6(SD=0.218 3), 基线算法-LGPMA 对表格结构识别任务的平均相似度为 0.765 5(SD=0.154 2); 在用户可以随心所欲地绘制任何想要表达的草图表格结构的任务 4 中, 本文算法对表格结构识别任务的平均相似度为 0.957 1(SD=0.096 5), 基线算法-腾讯 OCR 对表格结构识别任务的平均相似度为 0.857 2(SD=0.225 6), 基线算法-LGPMA 对表格结构识别任务的平均相似度为 0.905 4(SD=0.133 4)。

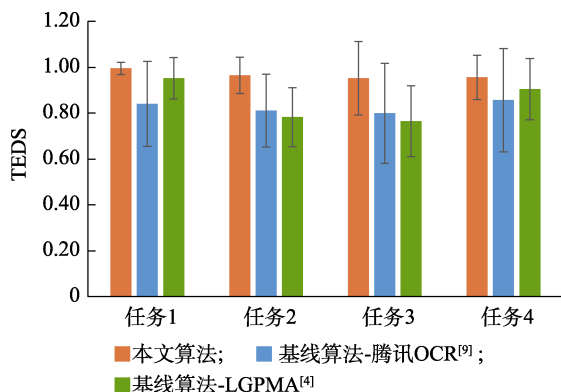


图 10 不同算法对结构识别的 TEDS 客观指标对比

3.4 专家评估

本节通过专家的定性评估, 以更全面地比较

本文提出的基于结构理解的手绘草图表格识别算法和已有研究中表现较优的基线算法对于草图表格的识别效果。邀请了 4 名研究笔交互、图像处理的科研工作者作为专家来对 3 种算法的识别结果进行对比评价, 并将评价结果反映在问卷调查表的相应问题中。

本文的研究重点在于对草图表格的结构理解, 而文本内容识别模块并非本文重点, 因此, 在专家评估时本文不对比不同算法的文本内容识别效果。进一步地, 为了避免文本内容的干扰, 本文将对 144 个草图表格识别结果中的文本内容进行脱敏和字符替换处理。

根据对本文算法以及基线算法的识别过程的拆分, 将主观评价指标分为 E_1 和 E_2 这 2 个维度, 分别为对当前算法识别表格结构和将内容匹配到对应单元格的效果评价, 并将它们映射到问卷调查表。问卷调查表中的 2 个问题如表 3 所示, 其中每个问题采用 5 分制的李克特量表^[38]进行计分, 数值越大, 代表识别效果越准确。

表 3 问卷中的评估项映射

评估项	评估项描述
E_1	对当前算法识别表格结构的效果评价
E_2	对当前算法将内容匹配到对应单元格的效果评价

在对专家给出的定性评价进行统计时, 本文以实验任务作为自变量, 因此对于每个实验任务, 分别得到了一组 3 种算法在 2 个评价指标维度中的效果评价, 如图 11 所示, 本文算法在表格结构识别和将内容匹配到对应单元格 2 个评价维度中的识别效果均优于基线算法-腾讯 OCR 与基线算法-LGPMA 的识别效果。具体地, 对于任务 1 本文算法识别效果优于基线算法-腾讯 OCR 与基线算法-LGPMA 的识别效果, 3 种算法识别表格结构的平均分分别为 4.701 4 (SD=0.555 7), 3.729 2 (SD=1.252 8)和 4.652 8 (SD=0.813 5), 将内容匹配到对应单元格的平均分分别为 4.118 1 (SD=0.927 4), 3.736 1 (SD=1.211 7)和 3.854 2 (SD=1.205 9); 对于任务 2, 本文算法的识别效果优于基线算法-腾讯 OCR 与基线算法-LGPMA 的识别效果, 3 种算法识别表格结构的平均分分别为 4.645 8 (SD=0.642 0), 3.645 8 (SD=1.173 5)和 3.201 4 (SD=1.232 3), 将内容匹配到对应单元格的平均分分别为 4.090 3 (SD=1.009 8), 3.722 2 (SD=1.167 4)和 3.715 3 (SD=1.113 6); 对于任务 3, 本文算法的识别效果优于基线算法-腾讯 OCR 与

基线算法-LGPMA 的识别效果, 3 种算法识别表格结构的平均分分别为 4.465 3 (SD=0.90), 3.666 7 (SD=1.257 1)和 3.194 4 (SD=1.129 9), 将内容匹配到对应单元格的平均分分别为 4.090 3 (SD=1.133 8), 3.638 9 (SD=1.282 5)和 3.486 1 (SD=1.176 6); 对于任务 4, 本文算法的识别效果优于基线算法-腾讯 OCR 与基线算法-LGPMA 的识别效果, 3 种算法识别表格结构的平均分分别为 4.312 5 (SD=0.956 9), 3.819 4 (SD=1.255 2)和 4.291 7 (SD=0.945 2), 将内容匹配到对应单元格的平均分分别为 3.972 2 (SD=1.102 7), 3.944 4 (SD=1.261 5)和 3.965 3 (SD=1.167 2)。

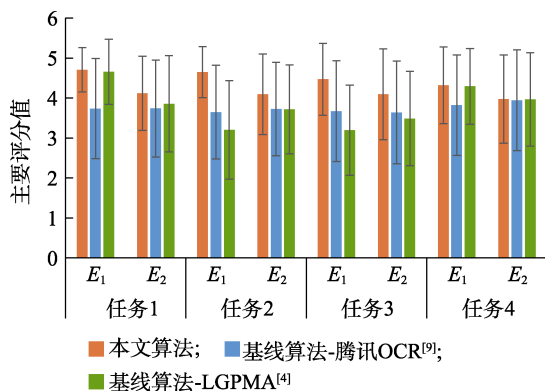


图 11 算法在 4 个实验任务中的专家评估结果对比

4 讨 论

通过观察在用户研究实验阶段采集到的 144 个草图表格、分析本文提出的算法和基线算法对于草图表格识别结果的客观指标以及对专家评估统计结果的分析, 可以得到一些有意义的发现结果, 在本节将逐一讨论。

当前业界公开的表格识别的开源数据集以印刷体表格为主^[8], 缺乏专门用于手绘草图表格识别数据集。而本文为了验证评估提出的表格识别算法表现, 构建了一个包括框线内容重叠和复杂单元格的草图表格数据集。本文招募了 36 名被试参与本文的用户实验与数据集构建, 通过 4 个实验任务, 共收集到 144 个用户绘制的草图表格。在本文的用户实验与构建数据集阶段, 用户在完成实验任务 4(任务 4 要求用户随心所欲地绘制任何想要表达的草图表格结构)的过程中得到了一些有意义的发现。用户在绘制草图表格时主要分为 2 类场景: (1) 规划安排类和知识汇总类。在规划安排类场景中, 用户绘制了各种结构、承载各种内容信息的表格, 包括某一日三餐的菜谱安排、周末的景点

游玩安排、访谈计划表、某年度论文投稿周期表、一周的吃饭计划、一周的开销回顾、一天的学习时间分配以及赛车比赛日历等。(2) 在知识汇总类场景中, 用户则绘制了某领域文献调研后的归类、球队与所在赛区的归类等。这些观察结果为本文提供了深入理解用户绘制草图表格的动机和场景, 可以看出, 草图表格作为一种自然、灵活的信息表达形式, 在辅助用户梳理计划安排和汇总知识等多个场景中可以具有广泛的应用。

此前, 研究者较少尝试使用笔迹结构理解算法识别包含框线内容重叠和复杂单元格的草图表格, 本文工作是对这一方面研究的重要补充。本文通过客观、主观评估对比本文算法和 2 种基线算法对于草图表格识别结果, 可以发现, 本文算法不仅可以有效地解决手绘草图表格识别面临的复杂单元格、手绘草图表格的结构与内容可能存在重叠区域的挑战, 更重要的是, 本文算法可以有效地克服用户绘制草图表格时伴随的随意性特点。具体来说, 通过对用户实验的观察可以得出随意性的 3 种具体表现形式: (1) 并列的结构框线长度不等, 有的长于其他线, 有的短于其他线; (2) 结构框线存在不同程度的倾斜; (3) 结构框线和内容存在不同程度的重叠遮挡。接下来会对以上的 3 种情况依次举例, 同样, 以上绘制草图的 3 种随意性表现形式也是算法在识别与重建表格时的噪声。

对于并列的结构框线长度不等的情况, 如图 12 所示, 可以看到用户绘制的结构框线有超出多余, 并非在框线交点处及时停笔。3 种算法对这个草图表格的识别结果如图 13 所示。可以从图 13b 和图 13c 中看到, 原图中结构框线的超出导致 2 种基线算法识别出的表格结构多了最右侧的一列和最底部的一行。但是对于草图表格, 随意性是其天然的特点, 因此对于 2 种基线算法, 面对此情境下的草图表格, 其结构识别方面仍有较大的提升空间。

时间(月)	序号	安排
1	1	周计划
	2	跑数据
2	3	算数据
	4	解数据
3	5	写数据
	6	回数据

图 12 随意性表现 1: 并列的结构框线长度不等

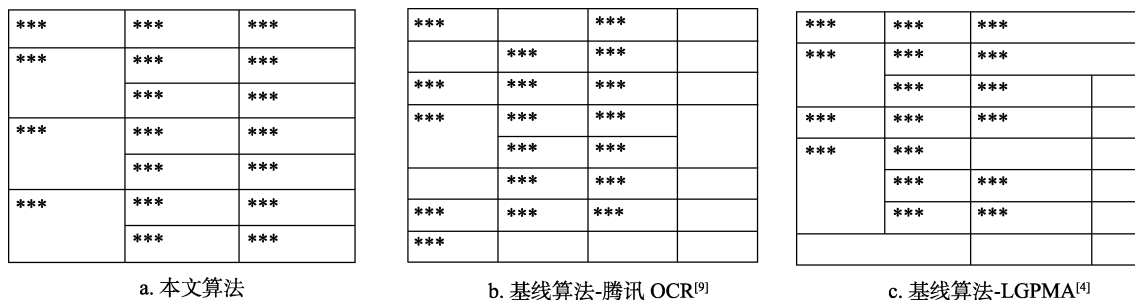


图 13 3 种算法对图 12 草图表格识别结果可视化

对于结构框线倾斜的情况, 如图 14 所示, 用户绘制的多条结构框线中存在不同程度的倾斜. 3 种算法对此表格的识别结果如图 15 所示. 可以看到, 当待识别的草图表格存在结构框线倾斜的噪声时, 本文算法与基线算法-LGPMA 对于表格结构的识别与重建是符合预期的, 与之相比, 基线算法-腾讯 OCR 的结构识别与重建仍有提升空间.

姓名	性别	籍贯
赵四	女	山东
王五六	男	北京
汪杨	女	山东
向天	男	深圳

图 14 随意性表现 2: 结构框线存在倾斜



图 15 3 种算法对图 14 草图表格识别结果可视化

对于结构框线和内容存在不同程度的重叠遮挡的情况, 此类噪声导致的错误类型主要表现在表格结构识别效果较差、定位出待识别的内容区域中用户绘制的草图表格真实内容区域变少. 通过观察发现, 用户绘制的草图中结构和内容存在的重叠遮挡区域越多, 其噪声越大, 基线算法识别的效果越差. 举例来看, 图 16 是其中一个受试者在任务 1 中绘制的草图表格, 图 17 是 3 种算法对图 16 的识别结果. 可以看到, 基线算法-腾讯 OCR 对于表格结构的识别较不理想, 相比之下, 本文算法与基线算法-LGPMA 识别的表格结构是完全符合预期的; 而基线算法-LGPMA 定位的表格内容区域数量少于用户绘制的草图表格中真实的内容区域数量, 导致该算法对内容与对应表格结构单元格的匹配映射结果较不理想. 究其原因, 基线算法以图像作为输入源时, 去除表格结构与内容的遮挡是具有挑战性的, 而正因为本文算法的输入源是笔迹, 依据多笔迹分类来对结构和内容进行分离比对图像中的结构和内容分离更简单且有效,

这也进一步印证了笔迹作为草图表格识别算法的输入源有其先天的优势. 此类噪声更多的例子详见附录 B.

赵四	女	山东
王五六	男	北京
汪杨	女	山东
向天	男	深圳

图 16 随意性表现 3: 结构和内容存在重叠遮挡

此外, 对于文本内容识别, 当用户手写的内容较简单、规整时, 3 种算法的识别效果均表现良好. 但是, 一旦手写内容存在连笔、字与字之间存在重叠等现象时, 3 种算法对内容区域块的识别效果对比如表 4 所示, 可以看出, 3 种算法的文字识别效果均需要进一步优化. 因此, 提升表格内容的识别准确率是后续的研究工作之一.

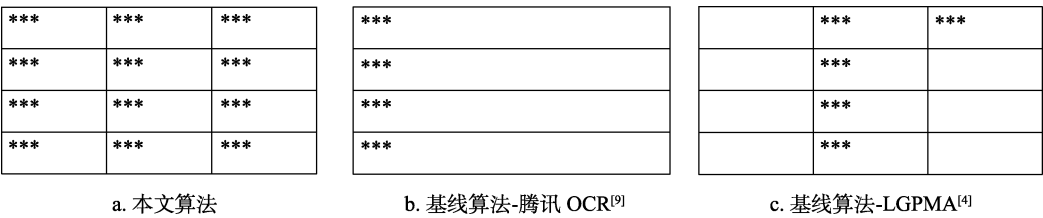


图 17 3 种算法对图 17 草图表格识别结果可视化

表 4 3 种算法对内容区域块的识别效果对比

定位的内容 区域块	本文算法	基线算法- 腾讯 OCR ^[9]	基线算法- LGPMA ^[4]
下午	下午	下午	下午
汪杨	汪杨	汪杨	汪杨
大学物理	大学物理	灼物理	大学物理
向伦	风纪	瓦	向伦
高数	高数	高粮	高数
语文	语文	谣改	設文
1C于	1C于	午	3午
汪杨	汪杨	汪楠	汪构
解释模型	解释模型	解译模型	解释模型
解释模型	解释接型	解释喉型	解科模型
马记文	145记文	丐政	写论文

5 总结与展望

近年来,随着移动设备、大型电子白板和平板电脑等触控设备的普及,笔式用户界面成为了主流的自然交互方式之一,被广泛应用于各个领域.基于笔交互的手绘草图表格可以凭借笔交互更自然、更灵活和更符合人类习惯的特点来拉近人类思维空间与物理空间的距离,更容易促进用户的思维交流和创造性交互工作,可以避免传统电子表格系统存在的用户学习成本高和交互界面烦琐等问题.然而,由于草图本身具有的模糊性、抽象性和用户在绘制草图时的随意性,对于存在复杂单元格、结构框线与内容重叠、笔迹重描补笔等情况

的手绘草图表格的识别面临着巨大的挑战.为此,本文提出一种基于结构理解的手绘草图表格识别算法,其输入为基于笔式用户界面采集的手绘草图表格,输入源对象形式为手绘草图表格过程中采集到的笔划数据信息,算法的输出为重建后的可视化表格.经过基于用户研究实验的客观评估和专家评估验证,在客观定量化的评估结果中表明,与基线算法相比,本文算法在面向草图表格结构识别任务中的 TEDS 指标提升了 13% 以上;在专家评估的统计结果中表明,本文提出的算法在表格结构识别和将内容匹配到对应单元格 2 个评价维度中的识别效果均优于基线算法的识别效果.

但本文仍然存在一些问题需要后续进一步地研究.在多笔划分类任务中,优化数据集质量,提升样本多样化,以进一步提升分类结果的准确率,尽最大可能地减少在多笔划分类阶段引入的噪声;在表格旋转矫正中,对于表格内容笔迹进行单独的旋转矫正以应对更多类型的草图表格也是下一步的工作目标;在草图表格结构识别任务中,提升结构识别算法,以解决非全线框草图的表格结构理解;在草图表格内容识别任务中,对笔划作为输入源对象来提高此任务的准确率与可解释性也是下一步的工作目标.另外,可以利用字词之间的上下文信息作为辅助信息进行内容识别的后处理工作,以进一步提高识别准确率.

参考文献(References):

[1] Dai Guozhong, Tian Feng. Pen-based user interface[M]. 2nd ed. Hefei: University of Science and Technology of China Press, 2014(in Chinese)
(戴国忠, 田丰. 笔式用户界面[M]. 合肥: 中国科学技术大学出版社, 2014)

[2] Saariluoma P, Sajaniemi J. Visual information chunking in spreadsheet calculation[J]. International Journal of Man-Machine Studies, 1989, 30(5): 475-488

[3] Hendry D G, Green T R G. Creating, comprehending and explaining spreadsheets: a cognitive interpretation of what discretionary users think of the spreadsheet model[J]. International Journal of Human-Computer Studies, 1994, 40(6): 1033-1065

- [4] Qiao L, Li Z S, Cheng Z Z, *et al.* LGPMA: complicated table structure recognition with local and global pyramid mask alignment[C] //Proceedings of the 16th International Conference on Document Analysis and Recognition. Heidelberg: Springer, 2021: 99-114
- [5] Zhang X W, Lyu M R, Dai G Z. Extraction and segmentation of tables from Chinese ink documents based on a matrix model[J]. Pattern Recognition, 2007, 40(7): 1855-1867
- [6] Liu Y, Zhang H, Yun X L, *et al.* Table detection and cell segmentation in online handwritten documents with graph attention networks[C] //Proceedings of the 2nd ACM International Conference on Multimedia in Asia. New York: ACM Press, 2021: Article No.42
- [7] Hu J, Kashi R S, Lopresti D, *et al.* Evaluating the performance of table processing algorithms[J]. International Journal on Document Analysis and Recognition, 2002, 4(3): 140-153
- [8] Li Y R, Huang Z, Yan J C, *et al.* GFTE: graph-based financial table extraction[C] //Proceedings of the 25th International Conference on Pattern Recognition. Heidelberg: Springer, 2021: 644-658
- [9] 腾讯云. 表格识别[OL]. [2023-01-17]. <https://cloud.tencent.com/act/event/ocrdemo>
- [10] Zhong X, ShafieiBavani E, Jimeno Yepes A. Image-based table recognition: data, model, and evaluation[C] //Computer Vision. Heidelberg: Springer, 2020: 564-580
- [11] Sun Zhengxing, Xu Xiaogang, Sun Jianyong, *et al.* Sketch-based graphic input tool for conceptual design[J]. Journal of Computer-Aided Design & Computer Graphics, 2003, 15(9): 1145-1152+1159(in Chinese)
(孙正兴, 徐晓刚, 孙建勇, 等. 支持方案设计的手绘图形输入工具[J]. 计算机辅助设计与图形学学报, 2003, 15(9): 1145-1152+1159)
- [12] Li C J, Pan H, Bousseau A, *et al.* Free2CAD: parsing freehand drawings into CAD commands[J]. ACM Transactions on Graphics (TOG), 2022, 41(4): Article No.93
- [13] Xian W Q, Sangkloy P, Agrawal V, *et al.* TextureGAN: controlling deep image synthesis with texture patches[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 8456-8465
- [14] Chen W L, Hays J. SketchyGAN: towards diverse and realistic sketch to image synthesis[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 9416-9425
- [15] 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. Red Hook: Curran Associates Inc., 2017: 6000-6010
- [16] Lun Z L, Gadelha M, Kalogerakis E, *et al.* 3D shape reconstruction from sketches via multi-view convolutional networks[C] //Proceedings of the International Conference on 3D Vision. Los Alamitos: IEEE Computer Society Press, 2017: 67-77
- [17] Gao Liangcai, Li Yibo, Du Lin, *et al.* A survey on table recognition technology[J]. Journal of Image and Graphics, 2022, 27(6): 1898-1917(in Chinese)
(高良才, 李一博, 都林, 等. 表格识别技术研究进展[J]. 中国图象图形学报, 2022, 27(6): 1898-1917)
- [18] Kieninger T, Dengel A. Applying the T-RECS table recognition system to the business letter domain[C] //Proceedings of the 6th International Conference on Document Analysis and Recognition. Los Alamitos: IEEE Computer Society Press, 2001: 518-522
- [19] Koci E, Thiele M, Romero O, *et al.* A genetic-based search for adaptive table recognition in spreadsheets[C] //Proceedings of the International Conference on Document Analysis and Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 1274-1279
- [20] Yildiz B, Kaiser K, Miksch S. pdf2table: a method to extract table information from PDF files[OL]. [2023-01-17]. https://www.researchgate.net/publication/220887997_pdf2table_A_Method_to_Extract_Table_Information_from_PDF_Files
- [21] Abdullah H S, Jasim A H. A new approach for detection and extraction of tables in scanned document image using improved Hough transform[J]. Engineering and Technology Journal, 2016, 34(5): 738-753
- [22] Xu Bochen. Analysis of complex layout analysis techniques based on machine learning[D]. Beijing: North China University of Technology, 2018(in Chinese)
(许博臣. 基于机器学习的复杂版面分析技术研究[D]. 北京: 北方工业大学, 2018)
- [23] Goodfellow I, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144
- [24] Siddiqui S A, Khan P I, Dengel A, *et al.* Rethinking semantic segmentation for table structure recognition in documents[C] //Proceedings of the International Conference on Document Analysis and Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 1397-1402
- [25] Tensmeyer C, Morariu V I, Price B, *et al.* Deep splitting and merging for table structure decomposition[C] //Proceedings of the International Conference on Document Analysis and Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 114-121
- [26] Lin W H, Sun Z, Ma C X, *et al.* TSRFormer: table structure recognition with transformers[C] //Proceedings of the 30th ACM International Conference on Multimedia. New York: ACM Press, 2022: 6473-6482
- [27] Zhelezniakov D, Zaytsev V, Radyvonenko O. Online handwritten mathematical expression recognition and applications: a survey[J]. IEEE Access, 2021, 9: 38352-38373
- [28] Zhao W Q, Gao L C, Yan Z Y, *et al.* Handwritten mathematical expression recognition with bidirectionally trained transformer[C] //Proceedings of the 16th International Conference on Document Analysis and Recognition. Heidelberg: Springer, 2021: 570-584
- [29] Rachala R R, Panicker M R. Hand-drawn electrical circuit recognition using object detection and node recognition[J]. SN Computer Science, 2022, 3(3): Article No.244
- [30] Jiang Yingying. Research on online sketch understanding technology[D]. Beijing: Institute of Software Chinese Academy of Sciences, 2009(in Chinese)
(姜映映. 在线草图理解技术研究[D]. 北京: 中国科学院软件研究所, 2009)

[31]

Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks[J]. Information Processing & Management, 2009, 45(4): 427-437

[32]

Sánchez A V D. Advanced support vector machines and kernel methods[J]. Neurocomputing, 2003, 55(1-2): 5-20

[33]

Chen P H, Lin C J, Schölkopf B. A tutorial on v-support vector machines[J]. Applied Stochastic Models in Business and Industry, 2005, 21(2): 111-136

[34]

Müller K R, Mika S, Tsuda K, *et al.* An introduction to kernel-based learning algorithms[M] //Hu Y H, Hwang J N. Handbook of Neural Network Signal Processing. Boca Raton: CRC Press, 2018: 181-201

[35]

Wobbrock J O, Wilson A D, Li Y. Gestures without libraries, toolkits or training: a \$! recognizer for user interface prototypes[C] //Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology. New York: ACM Press, 2007: 159-168

[36]

Berrar D. Cross-validation[J]. Encyclopedia of Bioinformatics and Computational Biology, 2019, 1: 542-545

[37]

Jiang Y Y, Tian F, Wang H A, *et al.* Intelligent understanding of handwritten geometry theorem proving[C] //Proceedings of the 15th International Conference on Intelligent User Interfaces. New York: ACM Press, 2010: 119-128

[38]

Jebb A T, Ng V, Tay L. A review of key Likert scale development advances: 1995-2019[J]. Frontiers in Psychology, 2021, 12: Article No.637547

附录 A 用户研究实验中的 3 个指定型任务

任务 1. 假设你要组织一场交友活动, 初步思考与记录参加的人员信息, 请你按照如下的文字描述信息将其绘制成表格的形式. 要邀请的人一共有 4 位. 第 1 位名字叫“赵四”, 女, 来自山东; 第 2 位名字叫“王五六”, 男, 来自北京; 第 3 位名字叫“汪杨”, 女, 来自山东; 第 4 位名字叫“向天”, 男, 来自深圳.

任务 2. 假设你是一名大二的学生, 请根据如下文字描述的一天内的时间安排为自己绘制一张课程表表格. 一天中分为上午、下午和晚上 3 个时间段, 其中上午的时间安排了高数和大学物理两

门课程, 授课老师分别为刘本和汪杨; 下午有英语和语文两门课程, 授课老师分别为王数基和向任; 晚上的时间安排为写作业.

任务 3. 假设你是一名研三的学生, 计划将前段时期的一项工作发表论文, 请你根据论文投稿前 3 个月的各项任务安排来绘制论文投稿冲刺时间计划表. 所有待完成任务分别为做用户实验、跑数据、分析算法指标、讨论模型、写论文, 请自行安排到 3 个月内完成.

附录 B 绘制草图表格随意性表现形式 3: 结构框线和内容存在重叠遮挡

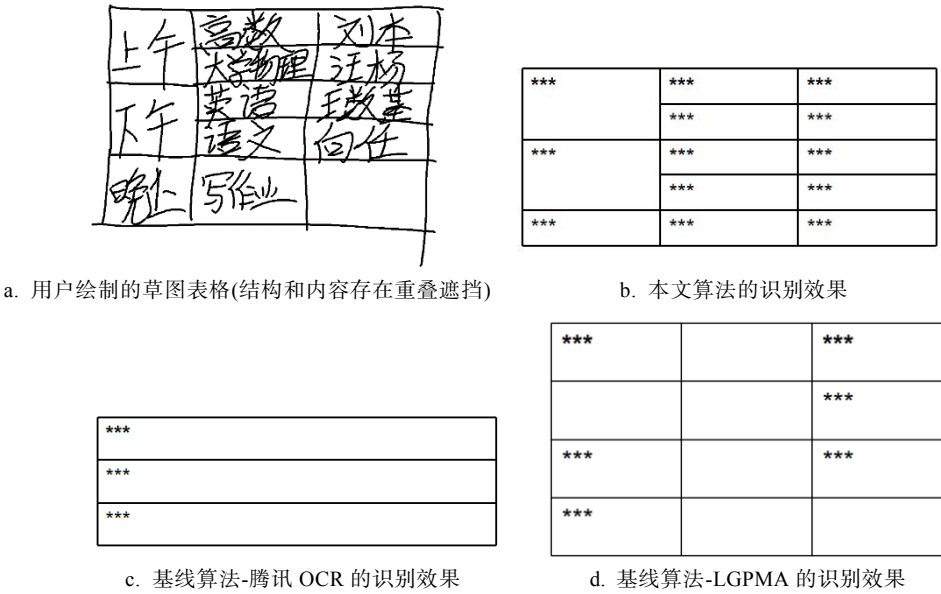


图 B1 随意性表现形式 3: 结构框线和内容存在重叠遮挡