

车载振动环境下基于字根拆解的在线手写文字识别*

付强^{1,2}, 黄进¹, 康文惠¹, 李洋¹, 周超¹, 田丰^{1,3}, 戴国忠¹

¹(中国科学院 软件研究所 人机交互与智能信息处理实验室, 北京 100190)

²(中国科学院大学 计算机科学与技术学院, 北京 100049)

³(中国科学院大学 人工智能学院, 北京 100049)

通信作者: 黄进, E-mail: huangjin@iscas.ac.cn



摘要: 随着触控技术和自动驾驶系统的发展, 手写文字输入作为触屏设备重要的交互方式之一, 已广泛应用于车载智能终端. 然而, 车载环境下的车身晃动和瞬时冲击会导致手写输入出现连笔增多、部件缺失等情况, 严重影响了手写文字识别的精度. 针对这个问题, 提出一种基于字根拆解的在线识别方法 CHORRD. 该方法首先采用连笔风格数据增强技术对原始数据进行扩充, 用于提升基础模型对连笔书写风格的识别鲁棒性. 然后, 提取手写文字结构和字根序列, 并根据所设计的首尾确认表意文字描述序列 HTCIDS 进行距离计算, 从而确定完整的文字. 实验结果表明, 该方法在车载振动模拟环境中能够有效地实现字根拆解和文字识别, 显著提高了在线手写识别性能, 识别准确率达到 90% 以上.

关键词: 人机交互; 振动环境; 字根拆解; 在线手写文字识别

中图法分类号: TP391

中文引用格式: 付强, 黄进, 康文惠, 李洋, 周超, 田丰, 戴国忠. 车载振动环境下基于字根拆解的在线手写文字识别. 软件学报. <http://www.jos.org.cn/1000-9825/7650.htm>

英文引用格式: Fu Q, Huang J, Kang WH, Li Y, Zhou C, Tian F, Dai GZ. Online Handwritten Chinese Character Recognition Based on Radical Decomposition in In-vehicle Vibration Environment. Ruan Jian Xue Bao/Journal of Software (in Chinese). <http://www.jos.org.cn/1000-9825/7650.htm>

Online Handwritten Chinese Character Recognition Based on Radical Decomposition in In-vehicle Vibration Environment

FU Qiang^{1,2}, HUANG Jin¹, KANG Wen-Hui¹, LI Yang¹, ZHOU Chao¹, TIAN Feng^{1,3}, DAI Guo-Zhong¹

¹(Intelligence Engineering Laboratory, Institute of Software, Chinese Academy of Sciences, Beijing 100190, China)

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

³(School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: With the advancement of touch-control technology and autonomous driving systems, handwriting input, as one of the key interaction methods for touchscreen devices, has been widely adopted in in-vehicle intelligent terminals. However, vehicle vibrations and sudden impacts in automotive environments often lead to more cursive strokes and missing components in handwriting input, significantly reducing the accuracy of handwritten text recognition. To address this issue, this study proposes a Chinese handwriting online recognition method with radical decomposition (CHORRD). This approach first employs cursive-style data augmentation to expand the original dataset, enhancing the base model robustness in recognizing cursive writing styles. Subsequently, the structure and radical sequence of the handwritten text are extracted, and distances are calculated based on the proposed head- and tail-confirmation ideographic description sequence (HTCIDS) to determine the complete character. Experimental results demonstrate that the proposed method effectively achieves radical decomposition and character recognition in simulated vehicle vibration environments, significantly improving online handwriting recognition performance, with a recognition rate exceeding 90%.

* 基金项目: 国家自然科学基金重点项目 (62332017)

收稿时间: 2025-10-16; 修改时间: 2025-12-09, 2026-01-23; 采用时间: 2026-02-11; jos 在线出版时间: 2026-08-19

Key words: human-computer interaction; vibration environment; radical decomposition; online handwriting Chinese character recognition

相较于传统的键盘、鼠标等输入设备,触控技术实现用户与显示界面的直接交互,可支持用户通过手指或专用触控笔在显示界面上直接执行手写输入、滑动浏览及点选等交互操作,显著提升了人机交互的便捷性和操作直观性^[1].在消费电子、工业自动化、医疗设备等诸多领域获得了广泛应用.近年来,随着车载信息娱乐系统和智能座舱的普及,汽车驾驶舱也广泛采用触控设备作为交互界面.然而,汽车在行驶过程中通常处于复杂的振动环境,如路面不平、发动机运转及车辆加速与减速,持续的振动不仅会影响车辆的机械性能和乘坐舒适性,还会导致手指或触控笔的输入轨迹偏移从而干扰驾驶员或乘客对触摸屏操作的稳定性,进而降低输入或触控操作的识别准确率和操作效率.如单轴与多轴高斯随机振动环境下,采取两种坐姿的受试者在草图绘制任务中感知到的干扰程度,结果表明随着振动强度增加,图形扭曲度和绘图难度均呈上升趋势^[2].因此,研究车载振动环境下对触控设备交互性能的影响,对于优化车载人机界面设计、提升行车安全性和用户体验具有重要意义.

手写文字输入作为触控设备的核心交互任务,通过模拟自然书写行为,为用户提供了符合认知习惯的文本输入方式.原始的触控设备文字输入数据仅包含时空坐标序列等底层特征表征,需要通过手写文字识别算法进行特征提取和语义解析,才能实现计算机可理解的文本转换,此过程通常响应时间较长,并存在理解困难的问题.因此,研究高精度的识别方法对于提升手写交互的实用性至关重要.根据数据处理的方法,现有的手写文字识别方法主要可以分为离线识别和在线识别两大类^[3].其中,离线识别方法主要针对已完成的静态文字图像进行识别.在线识别方法通过捕捉用户在书写过程中的动态笔迹数据,并依据这些动态特征对实时手写过程数据进行识别.尽管离线识别图像数据容易获取且技术成熟,但其无法利用丰富的动态信息,识别精度通常不如在线识别.振动环境会降低手写输入的准确性^[4],增加手写输入的难度^[5],因此利用在线手写输入过程数据效率,对手写文字的部件进行拆分^[6],以基本部件作为识别单位,更有利于提高在振动环境下整体识别的鲁棒性.目前手写文字识别技术在静态和动态环境中均已取得了显著进展并得到了广泛应用,但缺乏振动环境下手写文字识别的研究,且面临以下两方面问题:(1)现有的数据集通常是静态环境下采集的手写数据,振动环境的复杂性和多变性显著增加了数据采集的难度,导致缺乏振动环境下的手写数据.(2)动态振动环境影响书写条件的稳定性,手写文字通常表现出较为复杂的特征模式(例如,连笔增多和部件缺失),从而降低了识别的准确率.因此,振动环境的手写文字识别还需要进一步深入研究.

针对以上问题,本文提出了一种车载振动环境下基于字根拆解的在线手写汉字识别方法(chinese handwriting online recognition with radical decomposition, CHORRD).该方法利用平滑连笔增加了连笔数据的训练量,提高了对连笔字的识别能力,平滑处理使扭曲文字更加平滑自然,符合正常书写的特征,更容易被识别,解决了笔画不连贯难以识别的问题,另一方面手写文字部件的缺失导致系统误判,本文提出的字根拆解方法增加了判别依据,有效提高部件缺少情况下的识别准确率.具体的,在基础文字识别模块中,将输入的手写数据改写为连笔风格并输入到基础手写识别器中,利用GRU编码器中对序列的动态特征进行编码以及聚合模块聚合前向和后向依赖关系,最后通过分类器输出概率分布;在字根识别模块中,对输入的手写数据分别进行结构分析和字根识别得到手写汉字的结构和字根,将两者合成为首尾确认表意文字描述序列(ideographic description sequence, IDS).最后,通过融合模块将概率分布和首尾确认表意文字描述序列进行融合得出最终的识别结果.为了验证本文所提出方法的有效性,我们在静止、低频和高频振动环境下分别进行了对比实验.一方面对比了本文方法与国内外方法的识别准确率;另一方面通过用户评估获取用户对本文提出的方法的主观评价.结果表明,本文所提方法在低频和高频振动两种情况下的准确率分别比最优基准方法提高了20.5%和30.1%;在使用体验上虽然识别时间略高于基准方法,但被试在准确率和效率等方面均给予了肯定,与定量结果相一致.本文的贡献总结如下.

- 针对车载振动环境导致对手写文字识别准确率下降的问题,提出一种基于字根拆解的手写识别方法,提高了手写文本识别方法的鲁棒性,使用户能够在振动环境下实现准确的手写文字输入.

- 针对车载振动导致的用户手写连笔增多的问题,基于连笔风格生成,通过端到端方法使模型能够有效学习连笔间的特征,增强基础识别模型对连笔字的识别能力.

- 针对车载振动导致的手写文字部件扭曲或缺失问题, 提出基于字根拆解的手写文字识别方法. 该方法从手写汉字的结构和字根入手, 并结合所提首尾确认表意文字描述序列 (HTC-IDS), 显著提升了对于部件缺失问题的应对能力.

1 相关工作

为了系统阐述振动环境下在线手写文字识别的研究背景与挑战, 本节首先梳理手写识别方法的基本范式与发展, 进而聚焦于振动环境对书写输入行为的影响研究, 最后总结在线手写汉字识别的关键方法. 这 3 个方面分别对应交互场景、环境干扰与识别技术这 3 个维度, 揭示当前研究在动态振动环境下的手写识别方面存在的空白.

1.1 手写识别方法

手写识别是模式识别和人工智能领域中一个重要的研究方向, 其目的是将手写内容转换为机器可以读取的字符代码. 在解决手写识别方法中, 按照处理数据的类型可以分为两个大类, 分别是离线手写识别 (offline handwriting recognition) 和在线手写识别 (online handwriting recognition). 其中笔式交互 (pen-based interaction) 是手写识别中重要的交互方式, 笔式交互是一种以数字笔 (如触控笔、手写笔等) 为核心输入工具的人机交互范式. 用户通过笔的点击、书写、绘图及手势操作^[7], 与各类电子设备 (包括平板电脑^[8,9]、智能手机^[10,11]及交互式白板^[12,13]等) 进行高效、自然的交互. 笔式交互作为最符合人类自然书写习惯的交互方式之一, 凭借其直观性和表现力, 在图形学和人机交互领域受到了学术界和工业界的广泛关注^[14]. 例如, Perteneder 等人^[15]提出大型交互表面上手绘草图的智能聚类的方法, 探究了大型交互式手绘草图进行结构化和重新排列的问题. Riche 等人^[16]探究了传统钢笔与数字钢笔在使用体验上的差异和不一致, 以此更好地提高数字墨水的使用体验. 全青山等人^[17]提出了基于混合输入的场景设计工具, 来更好地设计笔式用户界面. Zhang 等人^[18]提出姿势感知的手写笔+触摸技术, 感知并适应以用户为中心的姿势. 全青山等人^[19]提出一种面向指-笔混合输入的交互模型, 系统地考虑并行协作机制和不同通道输入之间的意图关联. 这些研究有效推动了笔交互技术的发展, 促使其应用场景日益丰富与广泛.

在手写识别应用中, 数字笔迹是不可或缺的关键元素, 而数字笔迹通常由一系列连续的数据点构成, 但这些数据点本身并不具备明确的语义, 因此需要借助识别技术对其进行解析与理解. 现有的识别方法大致可以分为手写文本识别^[20]和手写结构化符号^[21]和图形识别^[22]. 在手写文本识别方面, Aberdam 等人^[23]提出了一种用于视觉表示的序列到序列对比学习框架, 该框架同时考虑了序列到序列的结构以及文本视觉特征的提取来实现文本的识别. 此外, Ao 等人^[24]提出了一种跨模态原型学习方法来实现零样本识别, 该方法以印刷字符为原型, 利用特定模态嵌入网络对手写数据进行分类. 在手写结构化符号和图形识别方面, 康文惠等人^[25]提出了一种混合人机智能的手写数学表达式识别方法, 该方法借助手写数学公式识别算法、知识库和用户反馈, 增强机器对用户输入的数学表达式的理解, 从而提升手写数学表达式的编辑速度和准确率. Yun 等人^[26]提出了一种基于图神经网络的在线手写图标识别方法, 该方法将符号分割和符号识别表述为笔画图上的节点聚类和节点分类问题, 并在统一的学习框架下使用 GNN 模型共同解决复杂上下文关系的建模问题. Riba 等人^[27]提出一种基于结构信息的表格检测方法, 该方法将文档的底层结构建模为一个图, 并利用图神经网络从属于表格的节点中发现局部结构. Li 等人^[28]提出了一种表格结构识别网络 TSRNet, 用于联合检测和识别各种表格和表单的结构. 虽然现有的手写识别方法已经得到了广泛的研究, 但多数研究仍聚焦于静态环境下的笔交互, 缺乏动态环境下的研究.

1.2 振动环境下手写输入

在日常生活中, 许多场景都伴随着明显的振动^[29], 例如乘坐火车、地铁或汽车等交通工具. 然而, 在这种振动环境下, 人们仍然需要使用数字笔进行书写记录、绘图创作等交互行为. 为此, 相关研究针对振动环境对书写、绘画等行为的影响展开了深入探讨^[2]. 例如, Bhiwapurkari 等人^[4]研究了振动对书写能力的影响. 在这项研究中, 受试者以两种姿势坐在振动环境中完成素描任务, 并使用主观评价和两种客观方法调查了受试者在单轴和双轴随机振动下的素描难度. 结果表明, 随着振动幅度的增加, 变形百分比和感知难度也随之增加. 此外, Bhiwapurkari 等人^[5]探究了火车在行使过程中的振动水平、振动方向和坐姿对乘客书写性能的影响. 具体来说, 他们在模拟轨道车辆

上进行了一项实验研究,以检测在 1–20 Hz、振幅分别为 0.4 m/s^2 、 0.8 m/s^2 和 1.2 m/s^2 的低频范围内,单轴、双轴和多轴随机振动对书写任务的干扰。结果显示,书写性能的下降会随着振动刺激的增加而增加,横向和纵向的下降幅度最大。同样,多轴振动的影响与双轴振动相似,大于单轴振动。在 X 方向上,在膝上工作比在桌子上工作的性能下降幅度更大,但在桌子上工作时,所有其他单轴和多轴振动的影响都最大。

1.3 在线手写汉字识别方法

在线手写汉字识别主要是对实时书写过程进行识别,通常通过触控屏、数字笔或压力感应板采集数据,记录笔迹的动态时序信息(如笔画顺序、书写速度、压力变化等)^[30]。早期的工作主要基于分割方法,利用有限类的字符分类器和路径搜索算法进行文本识别。例如,Wang 等人^[31]提出了一种整合字符分类、语言上下文和几何上下文的动态维护的分割-识别候选网格来实时识别中文手写句子,在实现过程中则利用动态文本行分割、字符过度分割、路径评估和搜索等方法。Wu 等人^[32]将前馈神经网络语言模型和循环神经网络语言模型与后退 n 元语言模型相结合,构建了混合语言模型用于手写文本识别。Hu 等人^[33]提出了一种基于检索的手写中文文本识别语言模型,首先采用通用语言模型识别文档输出初始文本,并利用该文本从大型文本语料库中检索相关内容,最后根据检索到的内容建立自适应语言模型,并与普通语言模型相结合,从而获得最终的文本。

虽然显式分割步骤使识别过程更透明,便于错误溯源,但是中文字符间结构相似度高且空中书写缺乏触觉反馈会导致笔画变形,因此非分割方法得到了广泛的研究。例如,Sun 等人^[34]提出了一种基于递归神经网络的新型端到端识别方法,该方法采用深度双向长短期记忆(bidirectional long short-term memory, BiLSTM)层和前馈子采样层的混合架构来编码长上下文历史轨迹。此外,在解码过程中,设计了一种改进的连接时序分类(connectionist temporal classification, CTC)束搜索算法来整合语言约束。Gan 等人^[35]提出了一种用于空中手写英文单词识别的统一卷积神经网络(convolutional neural network, CNN)和循环神经网络(recurrent neural network, RNN)方法,该方法采用编码器-解码器框架,其中编码器为深度 CNN 用于高效处理输入的时序特征,解码器为 RNN 用于准确生成目标字符序列。Gan 等人^[36]提出了一种时空卷积循环网络架构,通过 1 维卷积提取低级轨迹的局部上下文特征,然后利用循环网络捕捉高级输出中的长期依赖关系,不仅避免了特征图像提取中的领域特定知识依赖,还通过更紧凑的模型实现了更高的训练效率。Qu 等人^[37]提出了一种端到端多尺度注意力卷积递归网络,直接从原始时序坐标中提取多尺度局部上下文特征,然后利用双向长短期记忆(BiLSTM)捕捉多尺度局部上下文特征间的相关性,最后利用焦点连接时序分类(focal connectionist temporal classification, FCTC)进行预测。

目前,大多数手写文本识别研究均基于静态环境展开,然而在现实生活中,动态环境(如交通工具行驶时的振动)广泛存在。在此类振动条件下,由于书写过程不稳定,手写文字往往呈现出更加复杂的特征模式。具体而言:(1)连笔现象显著增加,导致字符间边界模糊;(2)字形结构完整性受损,关键笔画容易缺失或发生形变。这些振动环境特有的干扰因素不仅降低了传统识别算法的准确性,也对系统的鲁棒性提出了更高要求。因此,设计一种适应振动环境的手写文本识别方法,对提升系统在实际应用中的性能和可靠性具有重要意义。

2 手写识别方法设计分析

本文所提 CHORRD 方法主要基于手写汉字的连笔平滑处理与字根拆解,因此从手写汉字的连笔平滑处理与字根拆解两个核心维度展开分析。

2.1 连笔平滑处理设计分析

基础手写文字识别模块主要解决平滑扭曲和连笔字的问题。振动环境干扰了正常的书写过程,导致书写连笔增多且文字扭曲。为了解决该问题,本模块对原始笔迹轨迹在保持笔画连续性的前提下,实施平滑、重采样、降噪及曲线拟合等处理,以消除抖动和振动干扰,提高书写稳定性,从而得到更平滑、结构更合理的笔迹表示。对于在线手写文字识别,以往的方法通常先将手写序列数据生成手写图像,然后通过卷积神经网络(CNN)模型对图像数据进行处理,或者使用循环神经网络(RNN)模型采用端到端的方式直接处理手写序列数据。考虑到简单高效且能充分利用原始序列数据包含的丰富动态信息,并且验证了 RNN 作为判别模型来完成识别任务的有效性^[38]。因此,

本文选择 RNN 作为基本框架处理手写文字识别模块处理序列数据. 此外, 提升数据质量与数据增广均可有效提高文字识别的准确率. 在保证基本预处理的前提下, 提出针对失真与扭曲文字的贝塞尔曲线平滑处理方法, 以实现曲线形状可控并保证笔迹平滑度. 同时, 通过生成大量相关有效数据扩充训练集, 以增强模型的识别能力.

2.2 字根拆解设计分析

车辆的晃动使得手写笔难以按照预定轨迹稳定书写, 出现非预期的落笔动作, 或在应当抬笔时未能及时离纸并且笔速控制也会受到影响, 从而导致手写汉字的部件缺失, 引起识别系统的错误判断. 为了解决这个问题, 需要充分分析汉字的结构以及字根拆解方式^[39], 制定合理的表意文字描述序列^[40].

按照汉字的结构层次, 汉字可分为 3 个级别: 整字级别、部件级别和笔画级别. 整字即我们平时所使用的汉字, 在整字级别上, 汉字又可分为两种类型: 一种是独体字, 独体字方面 (如图 1(a1) 所示) 为单一结构, 是由不可分割的独立部件构成; 另一种是合体字 (如图 1(a2)–(a7) 所示), 分为左右结构、左中右结构、上下结构、上中下结构、左上结构、左下结构、右上结构、上左下结构、左上右结构、左下右结构、全包围结构和镶嵌结构.



图 1 汉字基本结构和拆解方式图

如图 1(b) 所示列举了几个常见的汉字部件, 每个汉字由基本部件构成. 据统计, 构成简体字的基本部件约为 540 个, 加上构成繁体字的基本部件大约 560 个.

笔画是组成汉字的基本单位, 如图 1(c) 所示, 基本汉字笔画包括点、横、竖、撇、捺、提、折、钩这 8 个笔画. 每一个汉字都是使用这 8 个笔画组成并按照一定顺序书写完成, 在整字层面上, 依据汉字结构, 一般遵循“先上后下、先左后右、先外后内、先外后内再封口”的顺序. 在笔画层面上, 则通常遵循“先横后竖、先撇后捺”的书写顺序, 所以每个字都可按照笔画拆解. 图 1(d) 为“永”字的笔画级别拆解 (拆解方式参照百度教育汉语系列). 同样按照汉字的基本结构和基本部件, 汉字也可以按照部件进行拆解, 图 1(e) 为“远”字的结构部件拆解示意图.

每个字都有自己的结构和书写特征, 根据这些汉字结构和特征可以帮助我们有效地推测出所写汉字内容, 提高手写汉字识别的准确率.

3 手写文字识别方法

如图 2 所示, 本文所提方法 CHORRD 的框架主要由基础手写文字识别、字根识别以及首尾确认表意文字描述序列 (head and tail confirmation ideographic description sequence, HTC-IDS) 生成 3 个部分组成. 首先, 系统采集动态的在线手写数据, 并依据采集数据选取相应的参考风格 (如连笔风格), 通过风格转换将输入手写数据适配至目标场景. 基础手写识别器采用基于双向 GRU 的编码器, 对输入序列的动态特征进行编码, 有效捕获手写数据中前后文之间的依赖关系, 从而为分类层提供融合历史与当前信息的动态特征向量. 分类器基于该向量输出对应字

符的概率分布. 此外, 系统对输入手写汉字进行结构分析, 提取其结构表示 (Str), 并输入至 HTC-IDS 生成器. 同时, 字根识别器处理输入手写数据, 通过编码器生成上下文向量, 再经解码器输出构成该汉字的字根序列, 该过程通过交叉熵损失进行优化. HTC-IDS 生成器将汉字结构与识别出的字根生成为 HTC-IDS 序列. 最后, 融合模块将基础识别器输出的概率分布与根据 HTC-IDS 计算出的结果进行融合, 生成最终的识别结果.

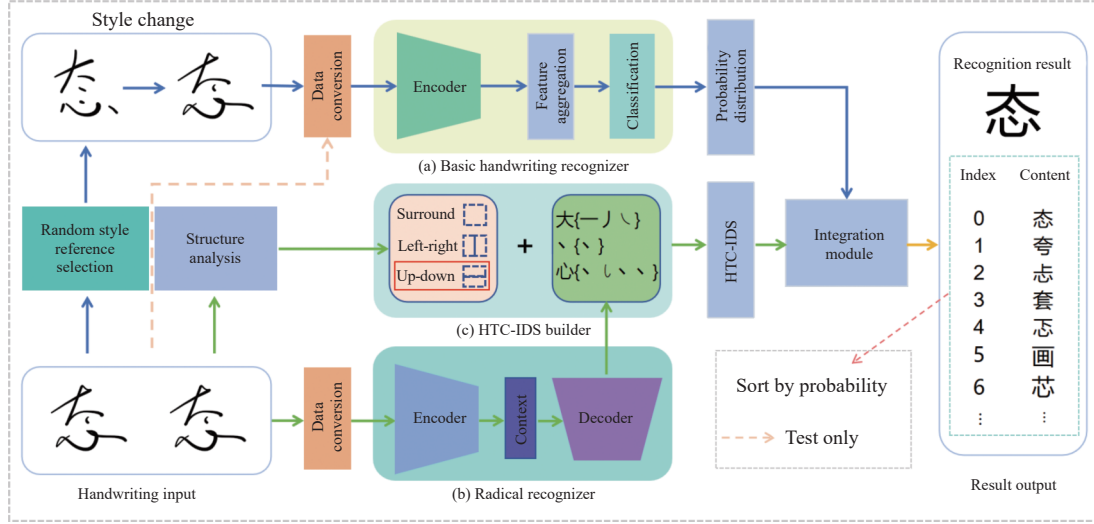


图2 车载振动环境下基于字根拆解的在线手写文字识别方法框架图

3.1 数据结构转换

在线手写汉字可以收集大量书写过程中的动态信息, 因此本文利用组成手写笔画的点坐标信息作为输入, 将一个手写汉字的数据结构表示为不同的概念. 如图3所示, **Strokes** 表示书写一个汉字的笔画数, i 代表该汉字由 i 个笔画组成; **Number** 表示笔画编号, S_1 表示第1笔; **Points** 表示该笔画由多少个点组成, n 表示该笔画共有 n 个点; $[x_1, y_1]$ 和 $[x_n, y_n]$ 分别表示第1个点和第 n 个点的坐标. $Position(x_1, y_1)$ 表示第1个点的位置, $Pressure$ 代表当前点的压力值, $TiltX$ 和 $TiltY$ 分别表示笔在手写平面绕 X 轴和 Y 轴旋转的角度, $Timestamp$ 代表当前的时间戳信息. 然后, 将提取的数据结构信息转换为向量 X 并作为编码器的输入. 该部分可以表示为:

$$\begin{cases} X_i = [x_i, y_i, \Delta x_i, \Delta y_i, f_1(s_i, s_{i+1}), f_2(s_i, s_{i+1})] \\ f_1(s_i, s_{i+1}) = \begin{cases} 0, & s_i \neq s_{i+1} \\ 1, & s_i = s_{i+1} \end{cases} \\ f_2(s_i, s_{i+1}) = \begin{cases} 1, & s_i \neq s_{i+1} \\ 0, & s_i = s_{i+1} \end{cases} \end{cases} \quad (1)$$

其中, $\Delta x_i = x_{i+1} - x_i$, $\Delta y_i = y_{i+1} - y_i$. $[f_1, f_2] = [1, 0]$ 代表落笔状态, $[f_1, f_2] = [0, 1]$ 代表抬笔状态. 图4展示了手写轨迹图像及其输入参数表示.

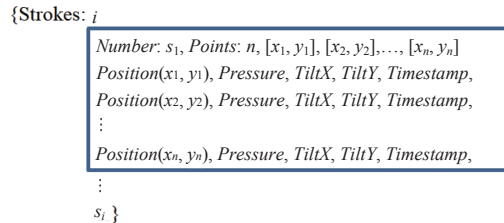


图3 手写汉字的数据结构表示图

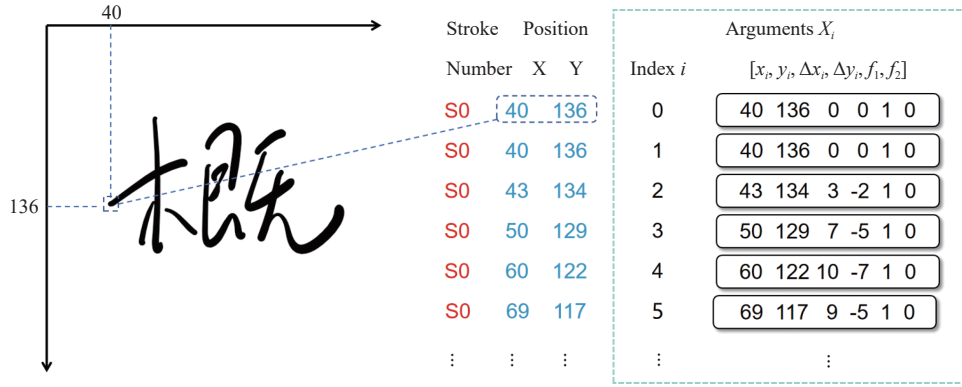


图4 手写轨迹图像及其输入参数表示图

3.2 基础手写文字识别

3.2.1 随机风格选取

在采样过程中, 由于冗余点的存在以及采样不统一会影响识别率, 因此本文首先对原始轨迹点进行去冗余和坐标归一化处理^[40]. 图5(a)到图5(b)为原始轨迹点去冗余点和坐标归一化处理过程. 此外, 针对振动环境下手写数据采集难、训练样本少的问题, 本文提出了一种随机风格选取的数据增强方法. 该方法将振动环境下现有的手写数据作为样本, 并随机选取样本的风格特征进行组合, 然后根据选取的风格对正常环境下采集的数据进行风格化处理, 从而增加振动环境下的手写文字的数据量. 由于本文主要研究振动环境下的连笔问题, 因此在随机风格选取时选择了连笔的风格特征.

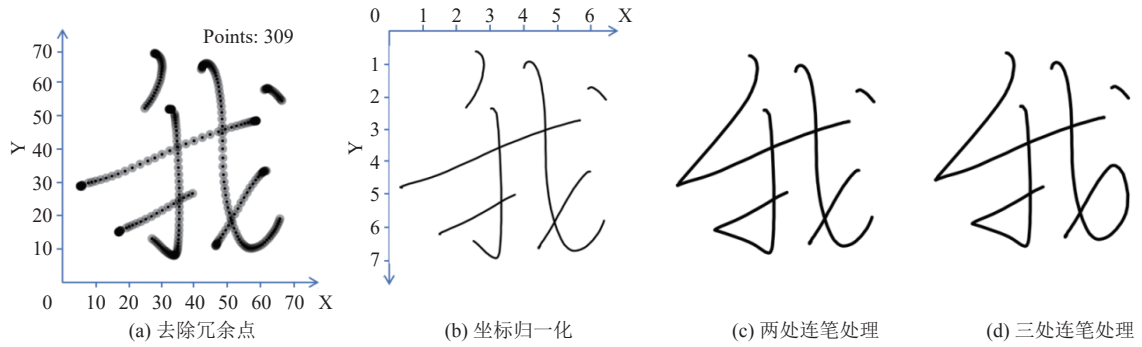


图5 随机风格选取和平滑连笔处理图

3.2.2 平滑连笔处理

与静态环境下的手写文字相比, 振动环境下的手写输入会产生更多的连笔字, 其典型特征是相邻笔画间出现“藕断丝连”的现象, 导致了手写识别软件的误识别. 因此, 为了增强模型对连笔字的识别能力, 本文提出了一种平滑连笔处理方法.

按照自然书写的规律和连近不连远的连笔规则, 两笔之间的距离通过当前笔画最后一点 (x_1, y_1) 与下一笔的第一个点 (x_2, y_2) 的距离计算, 具体表示为:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

两笔画之间是否存在连笔则由两笔的距离和对角线联合决定, 阈值定义为:

$$c_i = d - \frac{1}{5}b,$$

其中, $b = \sqrt{b_h^2 + b_v^2}$ 为对角线距离, b_h 和 b_v 分别代表两个笔画包围盒 (bounding box) 的水平距离和垂直距离. 当 c_i 大于 0 时不进行连笔操作, 笔画点之间的连接通过三阶贝塞尔曲线 (Bézier curve) 进行平滑处理, 连笔处理可视化如图 5(c) 和 (d) 所示.

3.2.3 手写文字识别器

在线手写数据属于时序数据, 虽然可采用 RNN 进行处理, 但 RNN 对于长序列的记忆能力有限且易出现梯度消失或爆炸问题. 同时, 双向循环神经网络通过融合序列的正向与反向信息, 能更全面地捕捉其上下文依赖模式^[40]. 因此, 本文选择改进的双向 GRU (gated recurrent unit) 作为基本文字识别的框架. 如图 6 所示, GRU 由重置门 R_n 和更新门 Z_n 组成, 重置门用于决定当前隐藏状态与之前隐藏状态的相关程度, 而更新门的作用是决定是否记住之前的状态信息. 重置门的计算公式如下:

$$R_n = \sigma(X_n W_{xr} + H_{n-1} W_{hr} + b_r),$$

其中, R_n 表示重置门的输出, σ 表示 Sigmoid 函数, X_n 表示当前输入, H_n 表示上一个时间步的隐藏状态, W_{xr} 、 W_{hr} 和 b_r 是可学习的权重参数.

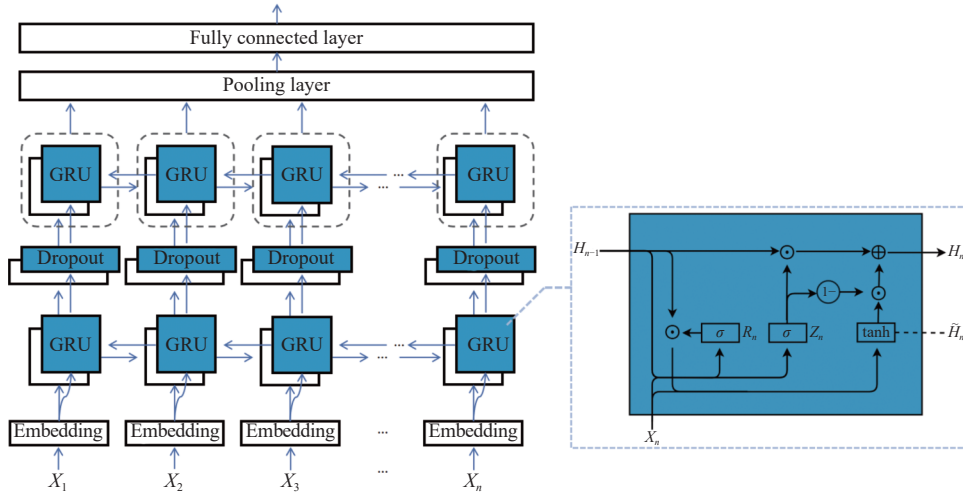


图 6 GRU 框架图

更新门用于决定是否记住之前的状态信息, 同样使用 Sigmoid 函数计算, 得到 0-1 之间的数值, 计算公式如下:

$$Z_n = \sigma(X_n W_{xz} + H_{n-1} W_{hz} + b_z),$$

其中, Z_n 表示更新门的输出, X_n 表示当前输入, H_{n-1} 表示上一个时间步的隐藏状态, W_{xz} 、 W_{hz} 和 b_z 是可学习的权重参数. 然后, 将重置门获得的向量 R_n 与上一时间步的隐藏状态 H_{n-1} 逐元素相乘得到 $R_n \odot H_{n-1}$, 在与当前时间步的输入 x_n 一起作为 $\tanh$ 激活函数相加, 可以得到候选隐藏状态 $\tilde{H}_n$. 计算公式如下:

$$\tilde{H}_n = \tanh(X_n W_{xh} + (R_n \odot H_{n-1}) W_{hh} + b_h),$$

其中, W_{xh} 和 W_{hh} 是权重参数, b 是偏置项, $\odot$ 表示逐元素相乘. 通过计算获得候选隐藏状态后, 仍需要结合更新门 Z_n 向量, 确定最终隐藏状态来自上一时间步的隐藏状态和候选隐藏状态的程度. 最终隐藏状态的计算公式如下:

$$H_n = Z_n \odot H_{n-1} + (1 - Z_n) \odot \tilde{H}_n.$$

最后, 通过特征汇聚层将所有时间步的隐藏状态整合为一个固定长度的特征向量. 为了均衡地提取全局上下文信息, 本文采用对全部隐藏状态计算平均值的方法. 随后, 分类输出层基于整合的特征向量预测类别概率.

3.3 首尾确认表意文字描述序列学习

在振动环境中, 由于手写汉字可能出现部件遗漏或笔画模糊等情况, 导致输出的序列不完整或存在错误, 难以与标准 IDS 字典正确匹配, 进而影响识别效果. 此外, IDS 序列越长所含部件越多, 完全匹配的难度也越大, 适当地

缩短序列有助于降低匹配复杂度、增强系统容错能力,从而提高识别率,但缩短序列时必须确保每个汉字所对应的描述序列仍具有唯一性. 针对以上问题,受五笔输入法中“字根序列+确认码”编码方式的启发,本文提出一种首尾确认表意文字描述序列 (HTC-IDS) 来整合结构类型、字根类型以及汉字的首笔与尾笔特征.

3.3.1 结构识别

汉字结构是构成汉字的基本空间组织方式,常见结构包括左右结构、上下结构、上下包围结构、左右包围结构、全包围结构、半包围结构、独体结构等. 与传统离线图像处理方法主要依赖于静态空间分布不同,在线手写数据采集记录了手写汉字的形成过程. 因此,根据笔画的位置和形成的先后顺序,我们提出了一种基于规则的手写汉字结构识别方法. 该方法根据笔画的空间分布、起止位置、笔画顺序、包围轨迹等特征,通过预定义的规则将汉字归类为若干种结构. 但汉字结构类别过多,不利于准确分类,因此我们将 13 种结构缩减为 3 种结构: 左右结构、上下结构、包围结构. 具体判断结构类型的规则设计如下.

(1) 左右结构: 将所有笔画的重心投影到 X 轴后,按平均 X 坐标排序,若左半部分的所有笔画结束后再开始右半部分,且左右笔画的 X 轴重心明显偏离,则判断为左右结构.

(2) 上下结构: 与左右类似,但使用 Y 轴投影判断. 若上部笔画全部结束后开始下部笔画,且上下重心分布差异显著,则为上下结构.

(3) 包围结构: 有连续 3、4 笔形成矩形闭环,末笔与首笔连接,中间区域含主要笔画部分.

结构识别作为在线汉字识别的重要环节,能够有效分解复杂字形,提升整体识别准确率,尤其在基于字根拆解的识别方法中具有关键作用.

3.3.2 字根识别

接收经过预处理的输入序列,进而按时间步更新内部隐藏状态,最终输出一个融合了输入序列全部语义信息的上下文向量 (context vector). 隐藏状态更新如下计算公式:

$$H_t = GRU(H_{t-1}, x_t),$$

其中, H_t 为 t 时间步的隐藏状态, H_{t-1} 是上一时间步的隐藏状态, x_t 为 t 时间步的输入.

为了解码器能够聚焦于输入序列中的关键信息,解码器将观察输入的所有隐藏状态,并为每一个隐藏状态打分,计算得到注意力得分 Score. 然后,将得到的 Score 进行 Softmax 归一化处理,得到每个输入对应的注意力权重 (attention weight). 假设解码器时刻为 t' ,注意力得分计算公式如下:

$$e_{t',i} = v^T \tanh(W_s s_{t'} + W_h h_i),$$

其中, $e_{t',i}$ 表示注意力的分, $s_{t'}$ 当前解码器状态预测, h_i 代表隐状态, W_s 、 W_h 和 v^T 是可学习的参数. 获得注意力得分后,下一步需要对注意力得分进行加权计算,公式如下:

$$\alpha_{t',i} = \frac{\exp(e_{t',i})}{\sum_{j=1}^T \exp(e_{t',j})},$$

其中, $\alpha_{t',i}$ 是注意力权重. 得到注意力权重后,将注意力权重与对应编码器输出的隐藏状态进行加权求和得到上下文向量 $c_{t'}$. 计算公式如下:

$$c_{t'} = \sum_{i=1}^T \alpha_{t',i} h_i.$$

假设解码器的任意时刻为 t' ,解码器输出 $y_{t'}$ 的概率 $p(y_{t'} | y_{t'}, \dots, y_{t'-1}, c)$ 取决于之前的输出序列和上下文变量. 循环神经网络将上一个时间步的输出 $y_{t'-1}$ 、上下文变量 $c_{t'}$ 以及 $s_{t'-1}$ 作为输入,转换成隐藏状态 $s_{t'}$. 隐藏状态更新计算公式如下:

$$s_{t'} = f_{\text{dec}}(y_{t'-1}, c_{t'}, s_{t'-1}).$$

在获得 $s_{t'}$ 后,可以通过 Softmax 获得输出 $y_{t'}$ 的条件概率分布 $p(y_{t'} | y_{t'}, \dots, y_{t'-1}, c)$. 输出计算公式如下:

$$p(y_{t'} | y_1, \dots, y_{t'-1}, c) = g(y_{t'-1}, c_{t'}, s_{t'}).$$

最后, 使用交叉熵损失函数作为损失函数. 计算公式如下:

$$L = - \sum_{t=1}^T \log p(y_t^* | y_1, \dots, y_{t-1}, c),$$

其中, y_t^* 是目标序列的第 t 个真实标签. 在每个时间步, 解码器预测了输出的概率分布, 并通过计算交叉熵损失函数来进行优化. 字根识别框架如图 7 所示.

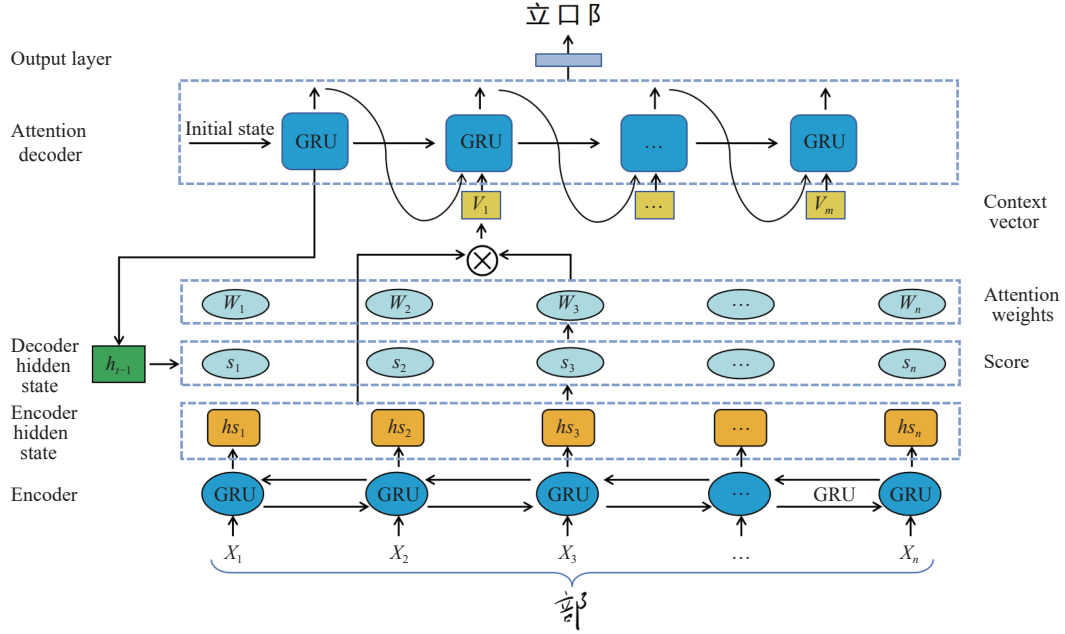


图 7 字根识别框架图

3.3.3 表意文字描述序列生成器

得到结构和字根识别结果之后, 利用表意文字描述序列生成器学习汉字的 HTC-IDS. 具体实现步骤为: (1) 确认首笔种类. 汉字按笔画顺序书写, 每一笔都不外乎点、横、竖、撇、捺、提、折、钩, 为了减少笔画分类, 将以上 8 种笔画分为点、横、竖、撇、捺、折这 6 种, 根据第 1 笔的特征, 确定首笔种类; (2) 确认尾笔种类. 根据最后一笔的特征, 确定尾笔种类; (3) 确认汉字结构. 根据第 3.3.1 节设计的结构类型的判断规则将结构分为 $\square$ (左右结构)、 $\boxplus$ (上下结构) 和 $\boxminus$ (包围结构); (4) 确认组成汉字的偏旁. 对汉字进行拆分, 按照书写的先后顺序, 将一个汉字所包含的偏旁种类进行罗列, 得出这个汉字所包含的所有偏旁.

根据以上步骤, 一个汉字的 HTC-IDS 可表示为: {首笔, 尾笔, 结构类型符{偏旁 1, 偏旁 2, ...}}. 其中, 偏旁数量依据具体汉字而定且最小数量为 1.

3.4 识别结果整合

IDS 匹配器将生成的 HTC-IDS 与字典进行匹配, 以确定输入汉字的分类结果. 匹配方法采用编辑距离 (Levenshtein distance), 计算两个序列之间转换所需的最小操作步骤的次数, 次数越少则两个序列相似度越大. 当部件识别算法得出的前两个或多个结果的权重值相同时, 按顺序遍历基本识别结果, 分别与部件识别结果对比, 第 1 个相等的汉字为最终识别结果. 当部件识别算法输出的前两个或多个候选结果权重相同时, 按顺序遍历基本识别结果, 并依次与部件识别结果比对, 第 1 个匹配的汉字作为最终识别结果. 若部件识别算法输出的前两个候选结果权重存在差异, 则按照以下规则确定最终结果; 若部件识别结果与基本识别结果一致, 则直接输出该结果; 若部件识别结果的权重高于基本识别结果, 则以部件识别结果为最终结果; 反之, 则以基本识别结果为最终结果.

4 实验和分析

4.1 实验设计

实验基于两因素 (3×4) 被试内设计实验, 自变量为振动强度和识别方法. 其中振动强度包含 3 个水平: 静止状态、低频振动以及高频振动, 具体参数依据“GT/T4857.23-2021”^[41]标准进行设定, 如表 1 所示. 4 种识别方法分别为搜狗输入法、讯飞输入法、百度输入法以及本文提出的手写汉字识别方法. 选取理由如下: 当前主流商业手写识别软件的识别性能已均历经大规模用户检验, 普遍适用于日常使用场景. 国内厂商针对中文手写习惯进行了深度优化, 在中文识别表现通常优于国外产品. 根据文献 [42] 统计数据, 2024 年中国输入法市场份额前 3 为搜狗输入法 (54.64%)、讯飞输入法 (26.04%) 以及百度输入法 (14.60%). 三者合计超过 95%. 因此, 针对本文选择以上 3 种主流输入法作为对比方法, 以评估本文所提出的手写汉字识别方法的性能.

表 1 振动参数设置

自由度	静态环境		低频环境		高频环境	
	振幅	周期 (s)	振幅	周期 (s)	振幅	周期 (s)
横荡	±0 mm	0	±35 mm	8	±50 mm	8
纵荡	±0 mm	0	±35 mm	4	±50 mm	4
垂荡	±0 mm	0	±35 mm	4	±50 mm	4
纵摇	±0°	0	±2.5°	4	±5°	4
横摇	±0°	0	±2.5°	4	±5°	4
艏摇	±0°	0	±2.5°	4	±5°	4

用于评估振动环境下识别方法的因变量包括绩效指标和感知指标, 绩效指标为识别准确率, 指任务中被正确识别的汉字的百分比. 感知指标包括准确性、效率、实时性、难易度、流畅度和满意度^[26].

4.2 被 试

在本次实验中, 通过发布被试招募公告, 共召集了 20 名被试者及 1 名主试者自愿参与. 被试者中包含 12 名男性和 8 名女性, 年龄范围为 21–40 岁 (平均值 $M=28.5$, 标准差 $SD=5.40$). 所有被试均无肢体功能障碍、无运动障碍且没有晕车症状. 所有被试均来自研究团队外部, 与团队无任何社会关联. 背景调查结果显示, 所有被试均为右利手且具备触控设备与笔输入设备的操作经验, 并熟悉相关手写输入方法, 每位参与者都签署了知情同意书.

4.3 设备与材料

自由度平台是用于飞行器、运动器 (如飞机、车辆) 模拟训练的动感模拟装置 (如图 8 所示), 是一种并联运动机构, 通过改变 6 个可以伸缩的动作筒来实现平台的空问六自由度运动 (横荡、纵荡、垂荡、纵摇、横摇、艏摇), 即 X、Y、Z 方向的平移、绕 X、Y、Z 轴的旋转运动以及这些自由度的复合运动. 该设备尺寸为 1100 mm×1100 mm×1900 mm, 行程 100–600 mm, 负重重量小于 40 kg, 速度小于 80 km/h, 下坡速度小于 110 km/h, 角度小于 27°, 系统响应频率 0–3 Hz, 系统功率 2 kW, 4 kW, 电压 220 V, 工作温度 0–40°C, 具体实验振动参数如表 1 所示.

用户手写数据采集设备使用微软 Surface Pro 9, 屏幕尺寸为 27.5 cm×18.3 cm, 分辨率为 2880×1920, 支持 120 Hz 高刷新率, 处理器 12th Gen Intel(R) Core(TM) i5-1245U (2.50 GHz), 内存 RAM 8.00 GB, 64 位操作系统, 基于 x64 的处理器, 操作系统版本为 Windows 11 专业版, 版本号 24H2, 笔和触控为 10 触摸点. 该设备兼具平板电脑和笔记本电脑的双重功能, 具有优异的操作系统兼容性和开发环境适应性, 支持手指触控和 Surface Pen 输入, 其轻薄便携的特性为实验操作提供了极大便利.

核心手写输入设备采用微软 Surface Slim Pen 2, 尺寸为 136.8 mm×11.3 mm×6.3 mm, 重量 13 g, 支持 4096 级压感和零压墨迹. 实验过程中, 受试者需完成随机呈现的手写文字任务, 任务呈现与数据采集均通过自主研发的手写数据采集软件实现, 其用户界面具体包含 5 个功能区域: 用户常用手选择框、输入目标字符的目标框、手写输入的区域、实时状态信息显示区域, 实现包括清除、加载、回放、手动保存、自动保存、手指触控以及擦除等功能.



图8 实验场景

为贴近日常汉字的使用习惯,本文用字的选取以汉字的使用频率为依据.根据《现代汉语语料库汉字频率表》^[43],前1028个高频汉字的累计覆盖率达90%,可覆盖绝大多数日常使用场景,此后汉字频率增长区域平缓(如图9所示),继续增加数字对覆盖率提升作用有限,反而会显著增加被试的输入负担.因此,本研究以累计覆盖率达90%作为汉字选取的基准抽取实验用字.

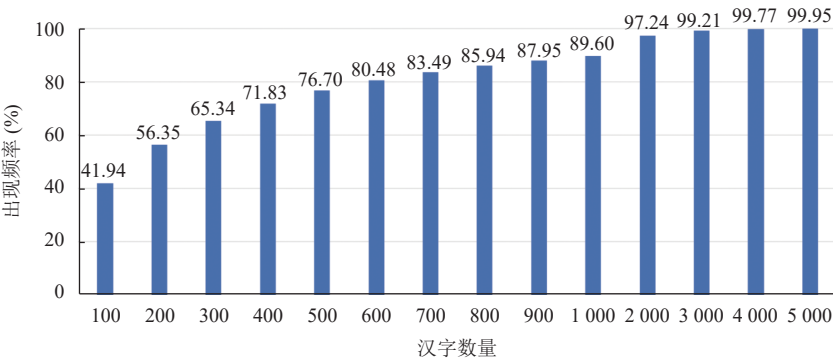


图9 现代汉语语料库汉字频率

此外,与静态环境相比,振动条件会显著增加被试的认知负荷与操作难度.前期调研结果显示,被试者单次持续书写任务的工作量以约200个汉字为宜.基于此,实验共设置4组任务,每组安排5名被试者共同完成.每位被试分别在静止、低频和高频这3种环境下各书写206个汉字.最终,共采集20名被试者在每种振动环境下各4120个汉字的书写样本,3种环境下总计12360个样本,累计覆盖率达90.03%.

为了获得用户主观评价,被试在指导下熟悉本文方法与其他方法的使用方式,并用几种方法完成汉字输入任务,任务完成后被试需要填写相关调查问卷,并进行访谈.问卷调查表如表2所示,该表主要用于收集被试在准确性、效率、实时性、难易度、满意度和偏好等方面的主观评价,访谈当前手写文字输入方法在哪些方面存在不足、给出低分的原因、是否需要为识别算法增加更多功能并列举相关功能、几种方法在使用过程中存在的问题,以及本文方法其他需要改进的地方为主要访谈问题.

表2 调查问卷(4种方法对应的问卷分别填写)

编号	您是否认同以下结论? 评分结果中1表示非常不认同, 5表示非常认同
Q1	您觉得通过搜狗/百度/讯飞/CHORRD能够准确获得手写文字
Q2	您觉得搜狗/百度/讯飞/CHORRD在手写文字识别方面的效率非常高
Q3	您觉得搜狗/百度/讯飞/CHORRD实时性更好
Q4	您觉得搜狗/百度/讯飞/CHORRD更不容易操作
Q5	您觉得搜狗/百度/讯飞/CHORRD手写笔画更流畅
Q6	您对搜狗/百度/讯飞/CHORRD很满意

4.4 实验流程

在正式实验开始前,研究人员向受试者详细介绍了实验的目的、流程及注意事项,并收集其基本信息.随后,

受试者被引导至实验区域, 自选舒适姿势坐在安装于振动平台上的可调节座椅上. 为确保实验操作的规范性, 正式实验安排在隔音良好的独立房间内进行. 在正式开始前, 设置了练习环节, 以帮助受试者熟悉实验流程并适应环境.

在正式实验时, 受试者被分成 5 人一组, 每组要求共同完成 1030 个汉字的手写输入任务, 每人随机分配 206 个汉字书写任务. 每位受试者需要随机在静止、低频振动和高频振动这 3 种环境下, 使用惯用手通过笔尽可能快速、准确地书写屏幕上呈现的 206 个随机分配的汉字, 每次书写约 10 min 休息 1–2 min, 直至完成全部任务, 受试者在完成全部任务后需填写问卷.

4.5 测量指标

在本研究中, 参与者需要完成一系列的手写汉字输入任务, 在任务执行过程中, 系统实时采集被试者的在线手写汉字数据作为实验原始数据, 并将其与多种不同手写输入识别方法进行对比分析. 具体比较内容包括以下两个方面: (1) 识别准确率: 准确率是评估手写输入识别方法有效性的核心指标, 直接反映系统正确识别用户手写字的能力. 本实验以正确识别汉字的百分比作为准确率的量化表示; (2) 用户主观评价: 通过收集用户的主观使用反馈, 获取其对不同方法的实际感受, 从而揭示在真实使用场景中可能存在的体验问题, 为后续方法优化提供参考依据.

4.6 实验结果

4.6.1 识别准确率

图 10 列出了在静止状态、低频振动以及高频振动这 3 种不同的识别环境下, 本文所提 CHORRD 方法和对比方法的识别准确率 (accuracy of recognition, AR). 从图 10 观察到的主要结果如下.

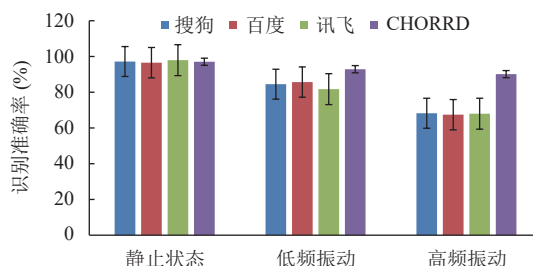


图 10 本文提出的方法和对比方法在 3 种不同的识别环境下的识别准确率

(1) 与所有对比方法相比, 本文提出的方法在振动环境下取得了最佳的识别准确率. 例如本文提出的 CHORRD 方法在高频和低频振动环境下分别取得了 90.15% 和 92.89% 的识别准确率, 而次优的识别方法在高频振动环境下取得的识别准确率为 68.26%, CHORRD 方法的识别准确率比次优方法最高可以获得 21.89% 的性能优势. 可能的原因是, 一方面是本文提出的方法基于连笔风格生成, 通过端到端方法使模型能更有效地学习连笔之间的特征, 增强基础识别模型对连笔字识别能力; 另一方面是从手写汉字的结构和字根方面着手, 并结合本文提出的首尾确认表意文字描述序列, 提高了对部件缺失的问题的应对能力.

(2) 所有方法的识别准确率随着振动幅度的增加而降低, 其中对比方法的识别准确率降幅 30% 左右. 这表明, 随着振动幅度的增加, 书写字体的感知难度显著上升, 字体呈现出更加复杂的特征模式. 然而, 本文提出的方法从静止状态到高频振动环境的识别准确率仅下降 6.93%, 远低于对比方法的下降幅度, 体现出 CHORRD 具有更强的鲁棒性. 另外还可以发现, 本文提出的方法在高频振动环境下获得了比对比方法在低频振动环境下更好的识别准确率. 例如在高频振动环境下, 本文提出的方法得到的识别结果为 90.15%, 高于对比方法在低频环境下的识别准确率, 可能原因在于本文提出的方法利用字根拆解有效地辨别复杂汉字, 从而在高频振动环境下, 依然具有强大的汉字识别能力.

4.6.2 用户主观评价

为了获得更真实的用户使用体验, 我们还进行了用户研究, 分别在输入准确率、效率、实时性、难易度、流

畅度和满意度这 6 个方面来获取用户的主观评价, 结果如图 11 所示. 结果显示, 本文的方法在输入准确率、效率、流畅度和满意度方面都优于其他方法. 在实时性方面略低于其他方法, 且主观评价得分都较高, 能够满足用户的实时输入的需求. 在难易度方面本文方法比其他方法难度略高, 说明操作略复杂, 且整体难度较低.

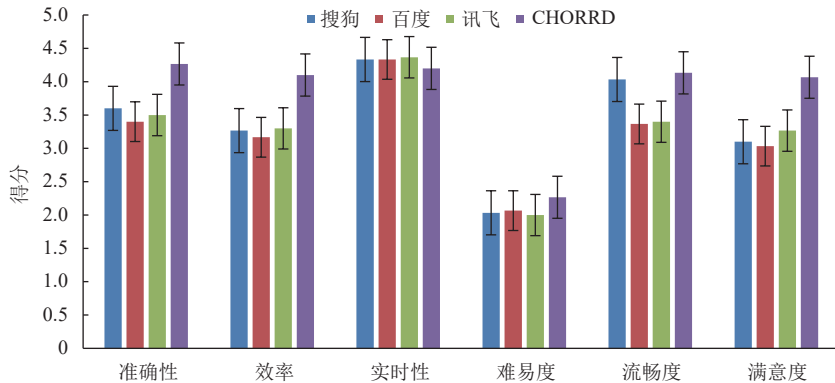


图 11 用户主观评价

由于整体数据不完全服从正态分布, 因此本节进行了威尔科克森符号秩检验进一步验证的 CHORRD 性能.

CHORRD 方法和搜狗方法在准确性 ($z=-3.0685, p<0.05$)、效率 ($z=-5.123, p<0.05$) 和满意度 ($z=-5.397, p<0.05$) 上存在显著性差异, 但在实时性 ($z=1.298, p=0.194$)、难易度 ($z=0.980, p=0.327$) 和流畅度 ($z=0.165, p=0.869$) 上不存在显著性差异; CHORRD 方法和百度方法在准确性 ($z=-4.0468, p<0.05$)、效率 ($z=-6.31, p<0.05$)、流畅度 ($z=-3.792, p<0.05$) 和满意度 ($z=-5.75, p<0.05$) 上存在显著性差异, 但在实时性 ($z=1.774, p=0.076$) 和难易度 ($z=-0.850, p=0.395$) 上不存在显著性差异; CHORRD 方法和讯飞方法在准确性 ($z=-3.8025, p<0.05$)、效率 ($z=-4.316, p<0.05$)、流畅度 ($z=-3.478, p<0.05$)、满意度 ($z=-5.00, p<0.05$) 和实时性 ($z=2.238, p=0.0252$) 上存在显著性差异, 但在难易度 ($z=-1.475, p=0.140$) 上不存在显著性差异. 由此可知, 本文提出的 CHORRD 方法在准确性、效率、流畅度和满意度方面均优于其他方法, 具备更好的综合性能表现. 然而, 在实时性方面略低于讯飞, 与其他方法在实时性和难易度方面的差异并不显著, 表明本文方法在提升性能的同时对实时响应与操作便捷性造成的负面影响较低.

在半结构化访谈中, 被试对 CHORRD 方法的使用体验普遍持积极评价, 认为该方法不仅能满足日常使用需求, 在书写流畅性和字形美观度方面也更易被接受. 例如, 被试 6 表示: “相比其他方法, CHORRD 在使用手写笔时书写流畅顺滑, 且书写效果更美观”; 被试 9 指出: “除搜狗方法和本文方法支持压感外, 其他方法的笔画粗细无法随压力变化, 显得生硬”; 被试 19 反馈: “CHORRD 方法基本能实现笔迹实时跟随笔尖移动, 而其他方法存在延迟甚至笔画丢失”. 然而, 也有被试指出了 CHORRD 方法目前存在的不足, 如被试 7 提到: “对倾斜文字的识别率较低, 且过于依赖笔顺”; 被试 3 认为: “整个书写过程只能连续输入, 无法对已书写文字进行单笔画修改, 交互功能不足”; 被试 6 补充: “所有方法在修改错误时仅支持重写或回退, 修改方式单一”.

此外, 被试者还提出了多项有价值的改进建议. 例如, 有用户指出“应增强与输入文字之间的交互功能, 而非仅提供清除和后退等基础按钮”; 另有建议提出“引入按笔画删除功能或橡皮擦工具, 以减少重复书写次数, 从而提升输入效率”. 基于这些反馈, 我们计划在后续工作中引入更先进的交互技术 (如人机混合智能) 以优化手写输入体验. 同时, 还将新增一系列旨在提升输入效率的实用工具, 包括可调节尺寸的橡皮擦、笔画级修改功能以及手势操作等.

4.6.3 消融实验

为了更全面地评估 CHORRD 方法的性能, 本节在 3 种振动环境下按照不同笔画复杂度 (将笔画数不超过 8 的划分为简单字, 笔画数大于 8 的划分为复杂字) 进行了消融实验. (1) 为了验证基本文字识别对字根拆解手写文字识别性能的影响, 在 CHORRD 的基础上去掉了基本文字识别模块, 直接利用字根识别进行手写文字识别, 得到

变体 CHORRD-base. (2) 为了验证平滑连笔对系统整体性能的影响, 在 CHORRD 的基础上删除了平滑连笔功能, 直接对连笔数据进行识别, 得到变体 CHORRD-cursive. (3) 为了评估字根拆分策略的效果, 在 CHORRD 基础上删除了字根拆分模块, 直接利用基本文字识别方法, 得到变体 CHORRD-radical. (4) 为了验证注意力机制的引入对系统整体效果的影响, 删除了注意力机制模块, 保留其他模块, 得到了 CHORRD-attention. (5) 为了验证预处理的有效性, 在没有进行预处理的前提下直接进行手写文字识别, 得到变体 CHORRD-preprocessing. CHORRD 方法及其变体的识别准确率 (AR) 值如表 3 所示. 从表 3 中可以观察到以下结论.

表 3 CHORRD 及其变体在 3 种振动环境下的识别准确率 (%)

振动环境	方法	简单字	复杂字	整体
静止状态	去除基本文字识别	91.39	85.14	88.79
	去除平滑连笔	95.42	81.64	89.68
	去除字根拆分	93.43	79.84	87.77
	去除注意力机制	96.67	81.24	90.24
	去除预处理	97.30	83.10	91.38
	本文方法	98.38	95.28	97.09
低频状态	去除基本文字识别	95.97	76.82	87.57
	去除平滑连笔	95.17	81.41	89.44
	去除字根拆分	93.09	79.66	87.50
	去除注意力机制	96.13	81.88	90.19
	去除预处理	94.59	86.07	91.04
	本文方法	96.09	88.40	92.89
高频状态	去除基本文字识别	94.22	72.55	85.19
	去除平滑连笔	91.93	78.55	86.36
	去除字根拆分	89.93	76.98	84.54
	去除注意力机制	93.47	78.50	87.23
	去除预处理	93.93	80.59	88.37
	本文方法	95.92	82.05	90.15

(1) CHORRD 方法在 3 种振动环境下的所有手写汉字 AR 结果均优于 CHORRD 的几种变体, 这表明本文提出的改进在提高振动环境下的手写汉字识别准确率上的有效性. 并且在两种不同笔画复杂度的分类情况下, 手写汉字识别的准确率都有所提高, 这说明本文提出的改进既对简单字有效同样适用复杂字.

(2) 基本手写文字识别模块是整个识别系统中重要的组成部分, 在静止状态下, CHORRD 方法的识别率为 97.09%, 而去除基本文字识别模块后的识别率为 88.79%, 识别率下降了 8.3 个百分点; 在低频振动状态下, CHORRD 方法的识别率为 92.89%, 而去除基本文字识别模块后的识别率为 87.57%, 识别率下降了 5.32 个百分点; 在高频振动状态下, CHORRD 方法的识别率为 90.15%, 而去除基本文字识别模块后的识别率为 85.19%, 识别率下降了 4.96 个百分点. 实验结果证明, 当去除基本文字识别模块后, 手写识别率在静止、低频、高频振动环境下的识别准确率都明显降低, 说明基本手写文字识别模块能有效提高识别准确率.

(3) CHORRD 方法对原始序列数据应用了平滑连笔处理. 为了检验其有效性, 我们在不进行平滑连笔操作的情况下重新训练了基本识别模块, 在静止状态下识别准确率由 97.09% 下降至 89.68%, 下降了 7.41 个百分点, 低频状态下准确率由 92.89% 下降到 89.44%, 下降了 3.45 个百分点, 高频环境下准确率由 90.15% 下降到 86.36%, 下降了 3.79 个百分点. 这表明平滑连笔操作对于模型来说能更好地处理连笔问题, 进而提高识别准确率.

(4) 在实验设计中, 我们进一步探究了字根拆分对识别效果的影响. 通过对比使用字根拆分与未使用字根拆分的模型性能, 可以发现, 在未引入字根拆分的情况下, 静止环境下准确率下降到 87.77%, 降低了 9.32 个百分点, 低频环境准确率下降到 87.50%, 降低了 5.39 个百分点, 高频环境下准确率下降到 84.54%, 下降了 5.61 个百分点. 这一结果表明, 字根拆分有助于提升模型对手写汉字结构的解析能力, 特别是在处理复杂结构汉字时, 其作用更为显

著.此外,我们还观察到,字根拆分能够在一定程度上缓解模型对完整书写顺序的依赖,提高识别鲁棒性.因此,我们认为字根拆分是提升手写识别性能的关键因素之一.

(5) 注意力机制的加入能够提升模型对关键特征的捕捉能力.实验表明,在去除注意力模块情况下,静止状态识别结果的准确率为 90.24%,下降了 6.85 个百分点,低频环境下准确率 90.19%,下降了 2.7 个百分点,高频环境下准确率 87.23%,下降了 2.92 个百分点.从表 3 中可以看出,在处理笔画密集或结构复杂的汉字时,模型表现出更强的适应性与稳定性.这是因为注意力机制通过动态调整不同区域的权重,使模型能够更聚焦于识别关键部分,有效减少了冗余信息的干扰.

(6) 在第 3 节中我们对原始数据进行了预处理操作,为了检验其有效性,我们将原始未处理数据作为训练数据对模型进行训练,对比使用过预处理数据的模型,整体识别率在静止状态下下降了 5.71 个百分点,在低频环境下下降了 1.85 个百分点,高频环境下下降 1.78 个百分点,这说明预处理对识别准确率有一定的影响,有利于模型更好地捕捉长期依赖关系.

4.7 分 析

4.7.1 汉字复杂度对识别率的影响

在进行振动对文字识别的实验时,我们发现汉字的复杂程度对识别正确率有一定影响,因此,为了进一步分析汉字复杂度对在线识别的影响,我们按照笔画数将采集的数据平均分成两个类,简单字和复杂字,并统计两类字的识别准确率,实验结果如图 12 所示.实验结果表明,在所有方法中,复杂字比简单字对识别准确率的影响更大,可能的原因是复杂字相比简单字的笔画数更多,复杂字书写完成所花的时间更多,受到的振动影响也就随之增多.此外可以发现,本文所提出的 CHORRD 方法在简单字和复杂字方面都要优于其他对比方法,尤其在复杂字方面比其他方法更有优势,说明 CHORRD 方法对振动影响有更强的适应性.

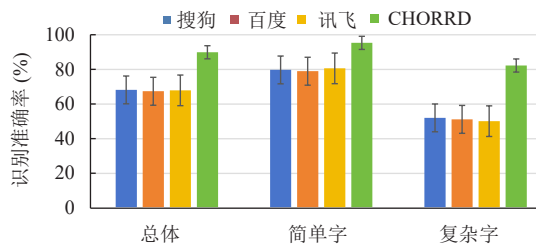


图 12 在不同汉字复杂度下的识别准确率

4.7.2 连笔作用

在振动环境下,手写文字中频繁出现的连笔现象显著增加了在线识别的难度.针对这一问题,本文提出的连笔风格增强方法能够有效应对连笔数量增多的情况.后文图 13 列出了基准识别方法识别错误而 CHORRD 方法识别正确的案例.

图 13(a)–(d) 表示基准识别方法的真实识别结果为第 2 顺位,表明基准方法虽然无法准确给出识别结果,但是该方法仍可为用户提供备选方案,使用户能够快速选择正确结果.然而,在图 13(e) 中,真实值被排在第 3 位;图 13(f) 中真实值位列第 5;图 13(h) 中更是位列第 9,随着排名靠后,用户选择正确结果的难度增加,耗时增长,对手写输入任务的完成效率产生明显负面影响.因此,即使不能实现完全准确识别,将真实值置于靠前位置也能有效降低因识别错误带来的时间损耗.值得注意的是,在图 13(h) 中,真实值“部”未出现在任何识别结果中,面对此类情况用户只能重新书写,这无疑增加了手写输入的总成本.

4.7.3 字根作用

在整个手写汉字识别过程中,字根识别发挥着关键作用,尤其对于书写过程中出现笔画缺失的汉字而言,其重要性更为突出.图 14 展示了不同字体下存在字根缺失的典型示例.

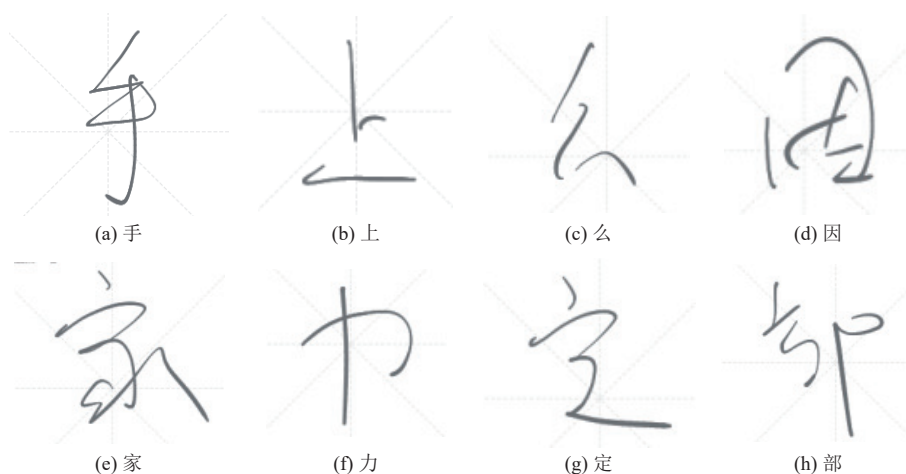


图 13 文字连笔图



图 14 字根缺失图

如图 14(a) 所示, 手写内容为“放”字, 但受环境的影响导致缺少笔画, 系统采集到的该字的 HTC-IDS 序列为: 万文{一\□{万{一┘□{一┘┘}}文{┘\□{┘┘┘}}}, 基准方法的识别结果为“故”, “故”的序列为: 故{一\□{古{一一□{一一┘┘}}文{┘\□{┘┘┘}}}, 第 2 顺位的识别结果为: 攷{一\□{万{一┘□{一┘┘}}文{┘\□{┘┘┘}}}, 第 1 顺位的识别结果需要修改 6 次才能得到目标序列, 第 2 顺位需要修改 5 次, 而目标文字“放”的序列为{、\□{方{、┘□{、一┘┘}}文{┘\□{┘┘┘}}}, 只需要修改 4 次. 因此, 通过首尾确认表意文字描述序列法可以确认手写文字为“放”.

图 14(b) 的手写内容为“题”, 第 1 顺位“颀”需要修改 14 次, 第 2 顺位“题”需要修改 33 次, 而目标文字“题”需要修改 13 次. 图 14(c) 第 1 顺位“父”需要修改 3 次, 第 2 顺位“父”需要修改 4 次, 目标文字“处”只需修改 1 次, 因此“处”为最后识别结果. 图 14(d) 的手写内容为“育”, 第 1 顺位“昏”修改 11 次, 第 2 顺位“育”修改 9 次, 目标文字只需修改 4 次. 图 14(e) 的手写内容为“将”, 第 1 顺位“壮”需要修改 14 次, 第 2 顺位“疔”需要修改 16 次, 而目标文字只需修改 4 次, 因此“将”为最后识别结果. 图 14(f) 第 1 顺位“余”需要修改 16 次, 第 2 顺位“余”需要修改 18 次, 而手写文字“金”只需要修改 1 次, 因此可以确定识别结果为“金”. 具体拆分方法见附录 A.

然而, 本文提出的首尾确认方法存在修改次数相同的情况. 如图 14(g) 所示, 手写内容为“利”, 系统采集到的

HTC-IDS 序列为{一丿□□木{一丿□□一丨丿\}} 可{丨丨□□{丨丨}}}, 基准识别第 1 顺位是“朴”, 序列为{一、□□木{一丿□□一丨丿\}} 卜{丨、□□{丨、}}}, 第 2 顺位是“刺”, 序列为{一丿□□{一{一一□□{一一}}□□{丨丿}}小{丨、□□{丨丿、}} 可{丨丨□□{丨丨}}}, 而手写文字“利”的序列为{丨丿□□{禾{丨丿□□{丨一丨丿\}} 可{丨丨□□{丨丨}}}, 在修改次数方面, 第 1 顺位“朴”需修改 4 次, 第 2 顺位“刺”需修改 31 次, 而目标文字“利”同样仅需 4 次. 由于第 1 顺位与目标文字的修改次数相同, 若仅依据此标准确定识别结果, 则“朴”与“利”均符合条件, 导致识别结果无法保证唯一性. 此外, 本文方法还可能因修改次数较少而产生误识别, 如图 14(h) 所示, 手写内容为“总”, 系统采集到的 HTC-IDS 序列为{丨、□□□□{丨一□□{丨一}}心{、\□□{、\、\、}}}, 基准识别方法的第 1 顺位是“忌”, 序列为{丨、□□{己{丨一□□{一一}}心{、\□□{、\、\、}}}, 第 2 顺位是“思”, 序列为{丨、□□{田{丨一□□{丨一}}心{、\□□{、\、\、}}}. 在修改次数方面, 第 1 顺位需修改 7 次, 第 2 顺位仅需 3 次, 而目标文字需修改 11 次. 根据本文方法, 应选择修改次数最少的第 2 顺位“思”作为识别结果, 然而实际手写内容为“总”字, 从而导致识别错误.

本文提出的首尾确认表意文字描述法, 能够在部分笔画缺失的情况下有效辅助手写内容的正确识别, 如图 14(a)–(f) 所示. 然而, 该方法在特定情况下仍存在识别结果不确定性的问题. 例如, 在图 14(g) 中, 系统判定“朴”与“利”两个候选字的修改次数相同均可作为最终结果, 导致无法输出唯一识别结果. 通过人工观察可发现, “利”字仅比采集序列多一笔“丨”, 更接近目标字形, 因而易于做出正确选择, 但系统未能有效区分此类细微差异. 在图 14(h) 中, 系统依据规则将“思”字判定为最佳识别结果, 而实际书写内容为“总”字. 人工比对发现, “总”与“思”二字与系统采集序列的差异均仅为两笔, 均具备作为候选字的合理性. 然而, 系统未将二者同时提供为用户选项, 导致识别结果未能反映实际书写意图. 综上所述, 尽管 CHORRD 方法在部件缺失情况下能有效提升识别率, 但仍存在识别结果模糊或错误的情况. 后续研究需针对此类特殊情形进一步优化识别策略, 以增强系统的鲁棒性与实用性.

4.7.4 错误识别示例

虽然 CHORRD 方法能够有效提升振动环境下的手写文字识别准确率, 但在实际应用中仍存在若干容易导致识别错误的情况, 图 15 中展示了一些识别错误的案例.

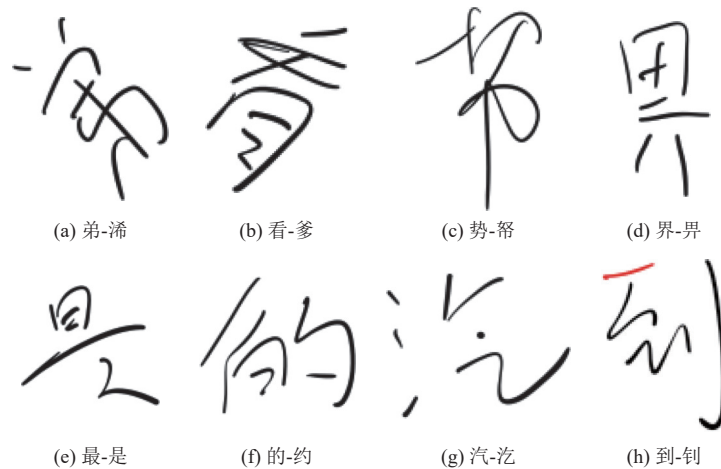


图 15 错误识别案例

图 15(a) 和 15(b) 的真实值分别为“弟”和“看”, 而识别的结果则分别为“涕”和“爹”, 其原因为手写文字的倾斜, 例如“弟”字向左倾斜, “看”字向右倾斜. 这一现象主要因为字根拆解对倾斜角度较敏感, 且未在数据处理方面加入矫正倾斜的相关操作. 而在振动等复杂的环境中, 手写文字倾斜是很正常的现象, 因此对于发生倾斜的手写文字将在后续工作中展开研究. 图 15(c) 的真实值是“势”, 识别结果却为“帑”, 其原因是手写文字过于夸张, 笔画顺序难以

匹配真实值的笔画序列。图 15(d) 的真实值为“界”, 识别结果为“界”, 该错误源于用户书写偏差, 而错误序列恰好与“界”字完全匹配, 导致系统自然选取了最吻合的错误候选。图 15(e) 的真实值为“最”, 识别结果为“是”, 导致识别错误的主要原因是手写过程中丢失了大量的笔画, 且系统采集到的手写序列与“是”字更加匹配。图 15(f) 的真实值为“的”, 识别结果为“约”, 识别错误的原因是受原本手写序列以外的笔画干扰, 且影响较大导致系统判别错误。图 15(g) 的真实值为“汽”, 识别结果为“沏”, “汽”和“沏”两个字过于相似, 两字只差一个笔画“一”, 而且手写过程中这一笔写的不标准更像一个点, 因此系统误判为“沏”。对于手写过程中的补笔操作也容易导致识别错误, 原因在于在线识别对手写笔画的顺序很敏感, 补笔相当于打乱了手写文字的笔画顺序, 因此依靠序列的方法将受到很大的影响。

5 总 结

本文提出了一种基于字根拆解的在线手写汉字识别方法 CHORRD, 以提高振动环境下的识别准确率。针对部件缺失问题, CHORRD 构建了字根拆解框架, 并结合所提出的字根结构分类策略, 有效解决了部分字根或文字结构易被识别但整体识别结果不理想的情况。在整体文字识别方面, CHORRD 对在线手写数据集中的手写文字进行连笔处理, 从而有效应对振动环境下连笔字增多的问题。对比实验结果显示, CHORRD 在振动环境下的手写汉字识别准确率均优于国内外常用商业方法。用户研究结果显示, 虽然识别时间略高于基准, 但并未影响实时输入效果, 反应时间与基准方法无明显差异, 未出现可感知延迟。

在后续工作中, 将进一步优化系统架构。针对结构多样性与复杂字形, 充分利用用户书写过程中的动态信息, 以进一步提升振动环境下手写汉字输入的准确率与效率, 同时降低系统响应时间。此外, 将根据用户需求引入更多自然交互技术, 进一步改善用户体验。

References

- [1] Aksan E, Pece F, Hilliges O. DeepWriting: Making digital ink editable via deep generative modeling. In: Proc. of the 2018 CHI Conf. on Human Factors in Computing Systems. Montreal: ACM, 2018. 205. [doi: 10.1145/3173574.3173779]
- [2] Bhiwapurkar MK, Saran VH, Harsha SP. Effect of multi axis vibration and subject postures on sketching distortion. Int'l Journal of Engineering, Science and Technology, 2010, 2(12): 14–24. [doi: 10.4314/ijest.v2i12.64564]
- [3] Zhong ZY, Jin LW, Xie ZC. High performance offline handwritten Chinese character recognition using GoogLeNet and directional feature maps. In: Proc. of the 13th Int'l Conf. on Document Analysis and Recognition. Tunis: IEEE, 2015. 846–850. [doi: 10.1109/ICDAR.2015.7333881]
- [4] Bhiwapurkar MK, Saran VH, Harsha SP, Goel VK, Berg M. Effect of magnitudes and directions (mono-axis and multi-axis) of whole body-vibration exposures and subjects postures on the sketching performance. Proc. of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 2011, 225(1): 71–83. [doi: 10.1243/09544097JRR336]
- [5] Bhiwapurkar MK, Saran VH, Harsha SP. Quantitative evaluation of distortion in sketching under mono and dual axes whole body vibration. Industrial Health, 2011, 49(4): 410–420. [doi: 10.2486/indhealth.MS1234]
- [6] Zhang JS, Zhu YX, Du J, Dai LR. Trajectory-based radical analysis network for online handwritten Chinese character recognition. In: Proc. of the 24th Int'l Conf. on Pattern Recognition. Beijing: IEEE, 2018. 3681–3686. [doi: 10.1109/ICPR.2018.8546074]
- [7] Li Y, Kang WH, Huang J, Yuan XF, Chen XJ, Yu Q, Wu HY, Tian F, Dai GZ. Human-computer hybrid framework based on pen-based interaction for CAD conceptual design. Computer Integrated Manufacturing Systems, 2022, 28(7): 2050–2063 (in Chinese with English abstract). [doi: 10.13196/j.cims.2022.07.011]
- [8] Hinckley K, Pahud M, Benko H, Irani P, Guimbretière F, Gavrilu M, Chen XA, Matulic F, Buxton W, Wilson A. Sensing techniques for tablet+stylus interaction. In: Proc. of the 27th Annual ACM Symp. on User Interface Software and Technology. Honolulu: ACM, 2014. 605–614. [doi: 10.1145/2642918.2647379]
- [9] Pfeuffer K, Hinckley K, Pahud M, Buxton B. Thumb + Pen interaction on tablets. In: Proc. of the 2017 CHI Conf. on Human Factors in Computing Systems. Denver: ACM, 2017. 3254–3266. [doi: 10.1145/3025453.3025567]
- [10] Kienle W, Hinckley K. Writing handwritten messages on a small touchscreen. In: Proc. of the 15th Int'l Conf. on Human-computer Interaction with Mobile Devices and Services. Munich: ACM, 2013. 179–182. [doi: 10.1145/2493190.2493200]
- [11] Yoon D, Chen N, Guimbretière F. TextTearing: Opening white space for digital ink annotation. In: Proc. of the 26th Annual ACM Symp.

- on User Interface Software and Technology. St. Andrews: ACM, 2013. 107–112. [doi: [10.1145/2501988.2502036](https://doi.org/10.1145/2501988.2502036)]
- [12] Mynatt ED, Igarashi T, Edwards WK, LaMarca A. Flatland: New dimensions in office whiteboards. In: Proc. of the 1999 CHI Conf. on Human Factors in Computing Systems. Pittsburgh: ACM, 1999. 346–353. [doi: [10.1145/302979.303108](https://doi.org/10.1145/302979.303108)]
 - [13] Xia HJ, Hinckley K, Pahud M, Tu X, Buxton B. WritLarge: Ink unleashed by unified scope, action, & zoom. In: Proc. of the 2017 CHI Conf. on Human Factors in Computing Systems. Denver: ACM, 2017. 3227–3240. [doi: [10.1145/3025453.3025664](https://doi.org/10.1145/3025453.3025664)]
 - [14] Aksan E, Hilliges O. Generative ink: Data-driven computational models for digital ink. In: Li Y, Hilliges O, eds. Artificial Intelligence for Human Computer Interaction: A Modern Approach. Cham: Springer, 2021. 417–461. [doi: [10.1007/978-3-030-82681-9_13](https://doi.org/10.1007/978-3-030-82681-9_13)]
 - [15] Perteneder F, Bresler M, Grossauer EM, Leong J, Haller M. cLuster: Smart clustering of free-hand sketches on large interactive surfaces. In: Proc. of the 28th Annual ACM Symp. on User Interface Software & Technology. Charlotte: ACM, 2015. 37–46. [doi: [10.1145/2807442.2807455](https://doi.org/10.1145/2807442.2807455)]
 - [16] Riche Y, Riche NH, Hinckley K, Panabaker S, Fuelling S, Williams S. As we may ink? Learning from everyday analog pen use to improve digital ink experiences. In: Proc. of the 2017 CHI Conf. on Human Factors in Computing Systems. Denver: ACM, 2017. 3241–3253. [doi: [10.1145/3025453.3025716](https://doi.org/10.1145/3025453.3025716)]
 - [17] Tong QS, Zhang ZQ, Huang J, Tian F, Liu J, Dai GZ. Scenario design tool based on hybrid input. Ruan Jian Xue Bao/Journal of Software, 2019, 30(S2): 48–61 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/19017.htm>
 - [18] Zhang Y, Pahud M, Holz C, Xia HJ, Laput G, McGuffin M, Tu X, Mittereder A, Su F, Buxton W, Hinckley K. Sensing posture-aware pen+touch interaction on tablets. In: Proc. of the 2019 CHI Conf. on Human Factors in Computing Systems. Glasgow: ACM, 2019. 55. [doi: [10.1145/3290605.3300285](https://doi.org/10.1145/3290605.3300285)]
 - [19] Tong QS, Kang WH, Fu Q, Huang J, Tian F, Dai GZ. Research on interaction primitives and interaction model for pen+touch inputs. Ruan Jian Xue Bao/Journal of Software, 2024, 35(4): 2022–2038 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6909.htm> [doi: [10.13328/j.cnki.jos.006909](https://doi.org/10.13328/j.cnki.jos.006909)]
 - [20] Julca-Aguilar FD, Hirata NST. Symbol detection in online handwritten graphics using faster R-CNN. In: Proc. of the 13th IAPR Int'l Workshop on Document Analysis Systems. Vienna: IEEE, 2018. 151–156. [doi: [10.1109/DAS.2018.79](https://doi.org/10.1109/DAS.2018.79)]
 - [21] Kang WH, Huang J, Tian F, Fan XM, Liu J, Dai GZ. Human-in-the-loop based online handwriting mathematical expressions recognition. Journal of Computer-aided Design & Computer Graphics, 2021, 33(11): 1773–1785, 1794 (in Chinese with English abstract). [doi: [10.3724/SP.J.1089.2021.18796](https://doi.org/10.3724/SP.J.1089.2021.18796)]
 - [22] Liu CL, Jin LW, Bai X, Li XH, Yin F. Frontiers of intelligent document analysis and recognition: Review and prospects. Journal of Image and Graphics, 2023, 28(8): 2223–2252 (in Chinese with English abstract). [doi: [10.11834/jig.221112](https://doi.org/10.11834/jig.221112)]
 - [23] Aberdam A, Litman R, Tsiper S, Anshel O, Slossberg R, Mazor S, Manmatha R, Perona P. Sequence-to-sequence contrastive learning for text recognition. In: Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 15297–15307. [doi: [10.1109/CVPR46437.2021.01505](https://doi.org/10.1109/CVPR46437.2021.01505)]
 - [24] Ao X, Zhang XY, Liu CL. Cross-modal prototype learning for zero-shot handwritten character recognition. Pattern Recognition, 2022, 131: 108859. [doi: [10.1016/j.patcog.2022.108859](https://doi.org/10.1016/j.patcog.2022.108859)]
 - [25] Kang WH, Huang J, Tong QS, Fu Q, Tian F, Dai GZ. HchMER: Hybrid human-machine intelligence method for handwritten mathematical expression recognition. Ruan Jian Xue Bao/Journal of Software, 2025, 36(2): 915–938 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/7169.htm> [doi: [10.13328/j.cnki.jos.007169](https://doi.org/10.13328/j.cnki.jos.007169)]
 - [26] Yun XL, Zhang YM, Yin F, Liu CL. Instance GNN: A learning framework for joint symbol segmentation and recognition in online handwritten diagrams. IEEE Trans. on Multimedia, 2022, 24: 2580–2594. [doi: [10.1109/TMM.2021.3087000](https://doi.org/10.1109/TMM.2021.3087000)]
 - [27] Riba P, Dutta A, Goldmann L, Fornés A, Ramos O, Lladós J. Table detection in invoice documents by graph neural networks. In: Proc. of the 2019 Int'l Conf. on Document Analysis and Recognition. Sydney: IEEE, 2019. 122–127. [doi: [10.1109/ICDAR.2019.00028](https://doi.org/10.1109/ICDAR.2019.00028)]
 - [28] Li XH, Yin F, Dai HS, Liu CL. Table structure recognition and form parsing by end-to-end object detection and relation parsing. Pattern Recognition, 2022, 132: 108946. [doi: [10.1016/j.patcog.2022.108946](https://doi.org/10.1016/j.patcog.2022.108946)]
 - [29] Bhiwapurkar MK, Saran VH, Harsha SP. Interference in writing performance under whole-body vibration exposure together with subject posture. Int'l Journal of Vehicle Noise and Vibration, 2016, 12(2): 182–198. [doi: [10.1504/ijvnnv.2016.079055](https://doi.org/10.1504/ijvnnv.2016.079055)]
 - [30] AlKendi W, Gechter F, Heyberger L, Guyeux C. Advancements and challenges in handwritten text recognition: A comprehensive survey. Journal of Imaging, 2024, 10(1): 18. [doi: [10.3390/jimaging10010018](https://doi.org/10.3390/jimaging10010018)]
 - [31] Wang DH, Liu CL, Zhou XD. An approach for real-time recognition of online Chinese handwritten sentences. Pattern Recognition, 2012, 45(10): 3661–3675. [doi: [10.1016/j.patcog.2012.04.020](https://doi.org/10.1016/j.patcog.2012.04.020)]
 - [32] Wu YC, Yin F, Liu CL. Improving handwritten Chinese text recognition using neural network language models and convolutional neural network shape models. Pattern Recognition, 2017, 65: 251–264. [doi: [10.1016/j.patcog.2016.12.026](https://doi.org/10.1016/j.patcog.2016.12.026)]

- [33] Hu SY, Wang QF, Huang KZ, Wen M, Coenen F. Retrieval-based language model adaptation for handwritten Chinese text recognition. *International Journal on Document Analysis and Recognition (IJDAR)*, 2023, 26(2): 109–119. [doi: [10.1007/s10032-022-00419-2](https://doi.org/10.1007/s10032-022-00419-2)]
- [34] Sun L, Su TH, Liu C, Wang RG. Deep LSTM networks for online Chinese handwriting recognition. In: *Proc. of the 15th Int'l Conf. on Frontiers in Handwriting Recognition*. Shenzhen: IEEE, 2016. 271–276. [doi: [10.1109/ICFHR.2016.0059](https://doi.org/10.1109/ICFHR.2016.0059)]
- [35] Gan J, Wang WQ, Lu K. A unified CNN-RNN approach for in-air handwritten English word recognition. In: *Proc. of the 2018 IEEE Int'l Conf. on Multimedia and Expo*. San Diego: IEEE, 2018. 1–6. [doi: [10.1109/ICME.2018.8486534](https://doi.org/10.1109/ICME.2018.8486534)]
- [36] Gan J, Wang WQ, Lu K. In-air handwritten Chinese text recognition with temporal convolutional recurrent network. *Pattern Recognition*, 2020, 97: 107025. [doi: [10.1016/j.patcog.2019.107025](https://doi.org/10.1016/j.patcog.2019.107025)]
- [37] Qu XW, Wu ZH. End-to-end multi-scale attention convolutional recurrent network for online handwritten Chinese text recognition. *Expert Systems with Applications*, 2025, 282: 127626. [doi: [10.1016/j.eswa.2025.127626](https://doi.org/10.1016/j.eswa.2025.127626)]
- [38] Ingle RR, Fujii Y, Deselaers T, Baccash J, Popat AC. A scalable handwritten text recognition system. In: *Proc. of the 2019 Int'l Conf. on Document Analysis and Recognition*. Sydney: IEEE, 2019. 17–24. [doi: [10.1109/ICDAR.2019.00013](https://doi.org/10.1109/ICDAR.2019.00013)]
- [39] Yu HY, Chen JY, Li B, Xue XY. Chinese character recognition with radical-structured stroke trees. *Machine Learning*, 2024, 113(6): 3807–3827. [doi: [10.1007/s10994-023-06450-6](https://doi.org/10.1007/s10994-023-06450-6)]
- [40] Zhang XY, Yin F, Zhang YM, Liu CL, Bengio Y. Drawing and recognizing Chinese characters with recurrent neural network. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 849–862. [doi: [10.1109/TPAMI.2017.2695539](https://doi.org/10.1109/TPAMI.2017.2695539)]
- [41] State Administration for Market Regulation, National Standardization Administration. Packaging—Basic Tests for Transport Packages, Part 23: Vertical Random Vibration Test Method: GB/T 4857.23-2021. Beijing: China Standards Press, 2021.
- [42] China Report Hall. 2024–2029 China input method industry project research and market prospect forecast and evaluation report. 2024. <https://www.chinabgao.com/report/14330738.html>
- [43] State Language Commission. Statistical Tables of Chinese Character Frequencies in Modern Chinese. Language and Culture Press, 1992.

附中文参考文献

- [7] 李洋, 康文惠, 黄进, 袁晓芳, 陈雪锦, 于茜, 武汇岳, 田丰, 戴国忠. 基于笔式交互的人机混合 CAD 概念设计. *计算机集成制造系统*, 2022, 28(7): 2050–2063. [doi: [10.13196/j.cims.2022.07.011](https://doi.org/10.13196/j.cims.2022.07.011)]
- [17] 全青山, 张宗琦, 黄进, 田丰, 刘杰, 戴国忠. 基于混合输入的场景设计工具. *软件学报*, 2019, 30(S2): 48–61. <http://www.jos.org.cn/1000-9825/19017.htm>
- [19] 全青山, 康文惠, 付强, 黄进, 田丰, 戴国忠. 面向指-笔混合输入的交互原语和交互模型研究. *软件学报*, 2024, 35(4): 2022–2038. <http://www.jos.org.cn/1000-9825/6909.htm> [doi: [10.13328/j.cnki.jos.006909](https://doi.org/10.13328/j.cnki.jos.006909)]
- [21] 康文惠, 黄进, 田丰, 范向民, 刘杰, 戴国忠. 人在回路的在线手写数学公式识别方法. *计算机辅助设计与图形学学报*, 2021, 33(11): 1773–1785, 1794. [doi: [10.3724/SP.J.1089.2021.18796](https://doi.org/10.3724/SP.J.1089.2021.18796)]
- [22] 刘成林, 金连文, 白翔, 李晓辉, 殷飞. 文档智能分析与识别前沿: 回顾与展望. *中国图象图形学报*, 2023, 28(8): 2223–2252. [doi: [10.11834/jig.221112](https://doi.org/10.11834/jig.221112)]
- [25] 康文惠, 黄进, 全青山, 付强, 田丰, 戴国忠. HchMER: 混合人工智能的手写数学表达式识别方法. *软件学报*, 2025, 36(2): 915–938. <http://www.jos.org.cn/1000-9825/7169.htm> [doi: [10.13328/j.cnki.jos.007169](https://doi.org/10.13328/j.cnki.jos.007169)]
- [41] 国家市场监督管理总局, 国家标准化管理委员会. 包装 运输包装件基本试验 第 23 部分: 垂直随机振动试验方法: GB/T 4857.23-2021. 北京: 中国标准出版社, 2021.
- [42] 中国报告大厅. 2024–2029 年中国输入法行业项目调研及市场前景预测评估报告. 2024. <https://www.chinabgao.com/report/14330738.html>
- [43] 国家语言文字工作委员会. 现代汉语字频统计表. 语文出版社, 1992.

作者简介

付强, 博士生, 主要研究领域为人机交互, 草图理解.

黄进, 博士, 研究员, CCF 专业会员, 主要研究领域为人机交互, 图形图像处理, 机器学习.

康文惠, 博士, 助理研究员, 主要研究领域为人机混合智能, 草图理解, 手写公式识别.

李洋, 工程师, CCF 专业会员, 主要研究领域为人机交互, 虚拟现实, 增强现实.

周超, 博士, 助理研究员, CCF 专业会员, 主要研究领域为认知控制, 人机认知耦合.

田丰, 博士, 研究员, 博士生导师, CCF 会士, 主要研究领域为人机交互, 虚拟现实.

戴国忠, 研究员, 博士生导师, CCF 杰出会员, 主要研究领域为人机交互, 计算机图形学.