

教育信息化背景下改进Bi-LSTM的英语手写识别理论研究

张晏濒

上海外国语大学贤达经济人文学院, 数据科学学院, 上海

收稿日期: 2026年7月25日; 录用日期: 2026年9月7日; 发布日期: 2026年9月15日

摘要

教育信息化背景下, 大规模纸笔考试英语填空主观题依赖人工阅卷, 存在效率低、评阅主观偏差大等问题。离线手写文本识别(HTR)是搭建智能阅卷系统的核心技术, CNN-BiLSTM-CTC为英文手写识别主流基础架构, 但单层Bi-LSTM难以捕捉长单词远距离字符依赖, 且无法自适应区分涂改、连笔等卷面噪声。本文以中学英语统考阅卷为应用场景, 提出理论优化方案, 设计双层堆叠Bi-LSTM搭配后置注意力机制的改进网络架构, 可为基础教育智慧英语测评体系建设提供理论参考。

关键词

教育信息化, 智慧阅卷, 英语手写识别, Bi-LSTM, 注意力机制

Research on Improving English Handwriting Recognition Using Bi-LSTM in the Context of Educational Informatization

Yanbin Zhang

School of Data Science, Xianda College of Economics and Humanities, Shanghai International Studies University, Shanghai

Received: July 25, 2026; accepted: September 7, 2026; published: September 15, 2026

Abstract

In the context of educational informatization, large-scale paper-based English cloze tests and subjective questions rely on manual grading, which suffers from low efficiency and significant subjective bias. Offline Handwritten Text Recognition (HTR) is the core technology for building intelligent grading

systems. CNN-BiLSTM-CTC is the mainstream architecture for English handwritten recognition, but a single-layer Bi-LSTM struggles to capture long-distance character dependencies in long words and cannot adaptively distinguish between corrections, connected strokes, and other form noise. Taking middle school English standardized exam grading as the application scenario, this paper proposes a theoretical optimization scheme and designs an improved network architecture with a double-layer stacked Bi-LSTM combined with a post-attention mechanism, which can provide a theoretical reference for the construction of an intelligent English assessment system in basic education.

Keywords

Educational Informatization, Intelligent Grading, English Handwritten Recognition, Bi-LSTM, Attention Mechanism

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

教育信息化持续推进背景下，各类大型纸笔考试的手写主观答卷评阅工作，始终是制约测评效率的突出瓶颈。当前国内统考广泛采用标准化答题卡作答模式，客观选择题可依托光学读卡设备完成自动化批阅，效率与准确率均有保障；但英语填空、短文写作等手写主观试题，仍需教师人工逐份核对批改，人力成本高、评阅周期长。如何在保障阅卷评判准确度的基础上，大幅提升主观题批改效率，是智慧教育测评领域亟待解决的现实问题。

手写文本识别(Handwritten Text Recognition, HTR)技术可将纸质手写字符转化为机器可解析的数字化文本，多年来始终是计算机视觉与模式识别的研究热点，在手稿数字化、纸质文档电子化归档等场景均具备成熟落地价值。面向英语试卷主观题的智能自动评分，核心前提便是实现手写词汇、语句的高精度识别。

伴随深度学习技术快速迭代，国内外学者已从字符、单词、文本行、完整文档等多维度开展手写识别相关研究。其中融合卷积神经网络、循环神经网络与连接时序分类损失的端到端 CNN-RNN-CTC 架构(Convolutional Neural Network-Recurrent Neural Network-Connectionist Temporal Classification)，凭借优异的序列建模能力，现已成为离线手写文本识别任务的主流基础框架。现有研究普遍采用 Bi-LSTM 完成图像特征序列的双向时序建模，但标准 Bi-LSTM 对特征序列内全部时间步采用均等权重处理，无法自动突出图像中核心字符区域、弱化涂改与纸面噪声等无效干扰信息；同时国内中学生手写存在连笔杂乱、字符倾斜、单词间距不均等多样风格特征，进一步加大单词识别难度。

从现有优化方向来看，多数 Bi-LSTM 改进方案聚焦中文汉字、阿拉伯文等复杂表意文字系统，针对中学英语手写词汇、适配校内自动阅卷场景的注意力机制优化研究仍较为匮乏，难以满足基础教育批量试卷识别的实际需求。针对上述研究短板，本文在经典 CNN-BiLSTM-CTC 架构基础上，于双层 Bi-LSTM 时序模块后嵌入注意力机制，通过自适应加权重组特征序列，强化有效字符特征权重、抑制卷面涂改与字迹畸变带来的噪声干扰，提升英语手写词汇识别鲁棒性。

2. 相关研究基础

2.1. 手写文本识别技术发展

离线手写文本识别(HTR)的核心目标是将无时序的纸质手写图像转化为可编辑字符序列，历经传统

机器学习、基础深度学习、端到端多模型融合三个发展阶段,手写样本数据扩充方案也随技术同步迭代。传统手写识别高度依赖人工设计笔画特征,泛化能力受字迹风格约束极强。深度学习普及后,CNN-RNN-CTC 端到端框架成为离线手写识别的主流基线架构,依靠卷积网络提取空间笔画特征、循环网络建模字符时序关系,依托 CTC 损失实现无对齐字符解码,大幅降低手写分割难度。Mahadevkar 等人搭建通用 CNN-BiLSTM-CTC 混合架构,以 IAM 公开英文手写集完成验证,证实该框架对离线手写单词具备稳定适配性,成为多语种手写识别的通用底层方案[1]。多语种实践进一步验证框架通用性,Duwal 等将 CNN-BiLSTM-CTC 用于尼泊尔文手写检测识别[2],Ghosh 完成梵文、孟加拉文在线手写单词识别,证明 CNN-RNN 时序架构适配各类文字手写任务[3]。

手写识别的精确度受手写样本数量、字迹多样性制约,数据增强成为图像识别前必备流程。国内学者沈强针对手写小样本困境,研究多策略混合图像增强方案,通过几何变换、噪声扰动扩充手写样本,缓解小数据集训练过拟合问题[4]。面向手写字迹域偏移问题,国内外学者提出自动增强、大语言模型驱动的跨模态数据增强方法,Chen 等实现手写字符精细化自动扩充,依托像素级扰动丰富字迹细节[5];Si 等人构建大模型驱动的数据增强统一框架,借助大语言模型语义能力扩充手写文本对应的图像样本,解决手写文本语义与图像不匹配的问题[6]。同时 GAN 等生成式方案被广泛用于手写字迹仿真,Diaz 等人系统性梳理 2019~2024 年手写合成全领域研究,总结各类手写生成方案的优劣,为贴合本土书写风格的仿真数据集构建提供完整理论参考[7]。除此之外,Transformer、TransRNN 等新型时序模型不断丰富技术路线,Devi 等人采用改进 TransRNN 完成手写光学字符识别[8];部分研究使用 CNN-Transformer 混合网络适配繁体中文手写识别,为时序建模提供 Bi-LSTM 以外的备选路径。EGO-LM 等专用语言模型也被嵌入识别流程,利用语言先验修正手写识别错漏字符,从语义层面补充视觉识别缺陷。

2.2. Bi-LSTM 在手写识别中的应用

单向 LSTM 仅能沿单一方向读取字符时序,无法同时利用单词前文、后文的双向上下文信息,在连笔、字迹倾斜、字符间距不均的手写场景容错率偏低,Bi-LSTM 通过双向结构同时利用前后文信息,显著优于单向 LSTM。现阶段绝大多数拼音文字手写识别以 CNN-BiLSTM-CTC 为基础架构,Mahadevkar 团队完整验证 CNN-BiLSTM-CTC 在通用英文手写文本识别的有效性,依托 IAM 数据集完成全流程测试,证实双向时序网络相比单向循环网络,对不规则手写单词的识别稳定性更强[1]。针对小语种手写,Duwal 搭建 CNN-BiLSTM 搭配 CTC 解码器完成尼泊尔文离线手写识别,验证 Bi-LSTM 对小众拼音文字的适配能力[2]。Imane 等人面向阿拉伯手写单词搭建 CNN-BiLSTM-CTC 模型,后续引入注意力与自适应增强做二次优化,也反向证明 Bi-LSTM 是多语种手写单词识别不可替代的基础时序模块[9]。

针对英语手写场景,已有研究搭建专属 CNN-BiLSTM 英文手写识别模型,专门适配英文大小写字母、短词汇的书写规律,优化 Bi-LSTM 的时序步长匹配英文单词字符长度,但现有方案大多依托国外 IAM 手写数据集训练,面向国内中学生本土化手写字迹适配性研究偏少。多数研究仅采用单层 Bi-LSTM 堆叠,对长单词远距离字符依赖捕捉能力不足;且 Bi-LSTM 对所有时序特征赋予均等权重,无法自动弱化涂改、纸面污渍等无效噪声特征,难以适配考试答卷带有涂改、划痕的复杂手写环境。多数 Bi-LSTM 相关研究聚焦阿拉伯文、尼泊尔文、中文等语种,专门面向国内中学英语填空手写词汇、阅卷场景的 Bi-LSTM 定制化优化研究较为零散。

2.3. 注意力机制在手写识别中的进展

手写图像普遍存在笔画重叠、污渍干扰、局部字迹模糊问题,常规 CNN-BiLSTM 对全部时序特征无差别处理,容易被噪声干扰识别判断。注意力机制依靠特征自适应加权,强化有效字符特征权重、抑制

无关干扰信息，近年成为手写识别重要优化方向。

注意力机制使模型能够动态关注输入序列中的重要部分，主要分为面向单词整体的全局注意力、面向笔画路径的局部分层注意力。*Rastogi* 等人提出 *SCRAT-Net*，构建基于注意力的手写单词深度学习架构，依靠注意力层定位单词核心字符区域，降低手写边缘冗余像素的干扰，大幅提升多连笔手写单词的识别准确率，证明注意力可以有效区分手写有效特征与背景噪声[10]。*Imane* 团队针对阿拉伯手写单词，在 *CNN-BiLSTM-CTC* 完整架构中嵌入注意力机制，让 *Bi-LSTM* 时序建模聚焦真实手写字符路径，为涂改类手写样本优化提供成熟参考模式[9]。

海外现阶段已经衍生出多种成熟的注意力优化架构。*Kass* 与 *Vats* 搭建基于编码器-解码器结构的 *AttentionHTR* 网络，将字符位置信息融入注意力权重计算，适配规整表单手写文字识别[11]；部分研究开展字符粒度的手写转录工作，依靠字符级注意力网络对独立字符区域分配权重，适合字符边界清晰的样本[12]；*Wang* 等研发解耦注意力网络，将对齐模块和解码模块相互分离，解除特征对齐和文字预测之间的互相干扰，适配通用场景下的文本识别任务[13]；*Coquenot* 等人提出垂直注意力网络，借助纵向注意力权重区分各行文本，专门用来解决多段落手写文本识别难题[14]。

整体来看，多数注意力优化工作围绕阿拉伯文、中文、尼泊尔文开展，面向国内中学英语手写词汇的注意力定制化设计不足。注意力机制已在多语言手写识别中验证有效，但针对国内中学英语手写词汇的识别，尤其在自动阅卷这一具体应用场景下的系统研究仍较少；且现有注意力大多仅简单接入网络末端，缺少针对双层堆叠 *Bi-LSTM* 时序层的分层注意力适配方案。本文聚焦于此，关注注意力机制对英语手写词汇 *Bi-LSTM* 识别模型的改进优化方案。

3. 改进 CNN-BiLSTM-Attention-CTC 网络架构理论设计

本文以面向中学考场英语手写词汇识别为目标，在经典 *CNN-BiLSTM-CTC* 端到端手写识别框架基础上完成理论优化，针对单层 *Bi-LSTM* 时序建模能力有限、手写涂改噪声无法被有效抑制等问题，在时序层后端引入自注意力模块，构建 *CNN-BiLSTM-Attention-CTC* 改进架构。

3.1. CNN-BiLSTM-CTC 基础模型框架

CNN-BiLSTM-CTC 是当前离线手写文本识别领域应用最广泛的端到端基础框架，整套架构分为 *CNN* 特征提取、*Bi-LSTM* 时序建模、*CTC* 时序解码三大基础模块，可以完成手写图像到字符文本的完整转换。

CNN 负责手写图像二维空间特征挖掘，输入统一尺寸为 128×64 像素的手写单词灰度图，依托多层卷积、批量归一化与池化操作逐层提取笔画边缘、字母轮廓、书写纹理等视觉信息。

Bi-LSTM 承担时序建模工作，接收 *CNN* 输出的特征序列，依靠正向、反向两条 *LSTM* 链路双向遍历时序，同步利用单词左侧前文、右侧后文的字符关联信息，完整捕捉手写单词内部字符的先后依赖关系。相较于单向 *RNN*、单向 *LSTM*，*Bi-LSTM* 可以更好应对手写字母间距不均、局部字迹模糊问题，依靠双向上下文信息弥补局部特征缺陷。

CTC 解码部分可以帮助实现免分割训练。手写单词不存在天然的字符边界，无法精准实现时序步与字母的逐一对齐，而 *CTC* 可自动计算网络输出概率序列与真实单词标签的匹配程度，训练阶段无需人工标注字符分割位置；同时依靠空白占位符、重复字符合并规则完成解码，极大降低手写样本的标注门槛，适配大批量手写试卷的识别落地需求。

3.2. 改进模型架构设计

传统 *CNN-BiLSTM-CTC* 对 *Bi-LSTM* 输出的全部时序特征赋予均等权重，无法区分有效手写字母特

征与涂改墨点、划痕、纸面污渍等噪声特征。参考 SCRAT-Net 手写单词识别的注意力设计思路,本文在 Bi-LSTM 输出端、CTC 解码输入端之间嵌入自注意力(Self-Attention)模块,实现时序特征自适应加权筛选。

设 Bi-LSTM 输出全部时序隐藏状态集合为 $H=[h_1, h_2, \dots, h_T]$, T 代表时序总步长,每一项 h_i 对应第 i 个时间步的时序特征,主要分为三步完成注意力的权重计算。

1) 计算原始注意力得分 e_i 。通过可训练权重矩阵 W 、偏置项 b 作线性变换,配合 $\tanh$ 激活函数完成特征映射,得到每个时序步的初始关注度得分。

$$e_i = \tanh(W \cdot h_i + b)$$

2) 应用 Softmax 函数归一化权重 α_i ,即利用指数归一化将所有得分转为总和为 1 的权重系数,权重大小代表该时序步的有效程度。

$$\alpha_i = \frac{\exp(e_i)}{\sum \exp(e_i)}$$

3) 加权聚合上下文信息,使用归一化权重对全部隐藏状态加权求和,最终得到融合重点特征的上下文向量。

$$c = \sum \alpha_i \cdot h_i$$

本文整合预处理流程、CNN 特征层、双层 Bi-LSTM 时序层、自注意力模块、CTC 解码层,形成完整改进 CNN-BiLSTM-Attention-CTC 理论架构,整体流程如图 1 所示。

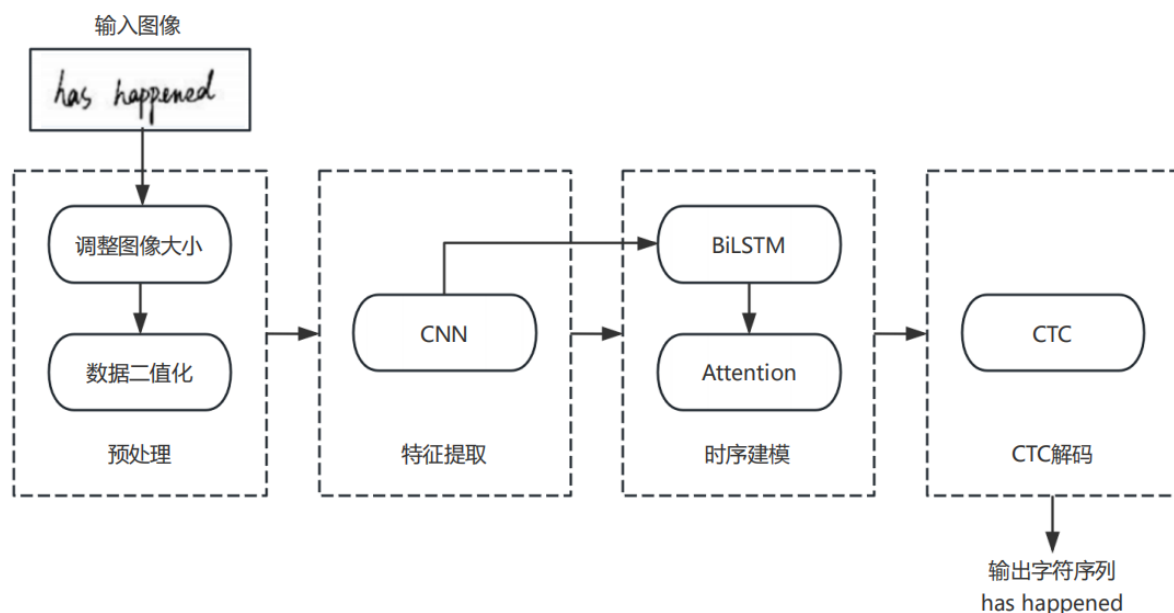


Figure 1. Improved model framework diagram

图 1. 改进模型框架图

第一阶段为图像预处理,包含尺寸统一调整、灰度图二值化两项操作。将任意尺寸手写单词图像规整为 128×64 规格,通过二值化简化画面色彩信息,去除多余灰度干扰,凸显手写笔画与背景的区分度,输出二维矩阵送入 CNN。

第二阶段为 CNN 特征提取，采用卷积搭配多层池化的经典结构，批量提取手写单词的空间笔画特征，将二维图像转换为一维时序特征，完成空间与时序维度的衔接。

第三阶段为时序建模，分为双层 Bi-LSTM 与时序注意力两部分。先由双层 Bi-LSTM 完成双向上下文时序挖掘，得到完整隐藏状态序列；再送入自注意力模块做自适应加权，过滤噪声时序特征，输出降噪后的融合时序向量。

第四阶段为 CTC 解码，接收加权时序特征，计算各时间步字符与空白符的预测概率，通过时序规则去重、剔除空白，最终解码输出完整英文单词文本。

4. 结论与展望

面向中学英语纸笔阅卷存在的人工成本高、评判主观性较强的现实问题，本文针对传统单层 CNN-BiLSTM-CTC 手写识别架构可能存在的问题，进行了理论层面的方案优化设计。传统单层 Bi-LSTM 难以充分捕捉长英语单词远距离字符的上下文关联，同时模型对涂改、墨迹、连笔等考场手写噪声存在无差异化处理问题。对此，本文构建融合注意力机制的改进 CNN-BiLSTM-CTC 理论架构，将时序模块优化为双层堆叠 Bi-LSTM 结构，在时序编码后增设注意力模块，依靠特征加权逻辑让架构主动聚焦有效字母笔画特征，弱化卷面涂改、污渍带来的无效特征干扰。方案设计可为中学英语智能阅卷提供完整的识别优化设计思路，为后续智慧测评落地提供相应的技术参考框架。

后续可扩大多源数据规模与提升模型泛化能力，落地本土手写样本采集与扩充工作。未来需通过与教育考试机构合作获取真实学生答卷样本，完善含涂改、多字迹的本土化手写资源，进一步提升算法方案面向真实考场环境的适配性与泛化潜力。同时，迁移学习与少样本学习策略可用于缓解目标场景数据稀缺的问题，使模型在有限本地样本条件下仍能实现较好的性能迁移。

基金项目

本论文得到了 2025 年度上海外国语大学贤达经济人文学院校级科研项目“基于深度学习的英语手写文本数据集构建及其自动阅卷应用研究”(A3106.25.1801.2523)资助。

参考文献

- [1] Mahadevkar, S., Patil, S. and Kotecha, K. (2024) Enhancement of Handwritten Text Recognition Using AI-Based Hybrid Approach. *MethodsX*, **12**, Article 102654. <https://doi.org/10.1016/j.mex.2024.102654>
- [2] Duwal, U., Karki, S. and Khadka, A. (2025) Offline Nepali Handwriting Detection Using CNN-BiLSTM with CTC Decoder. 2025 *International Conference on Intelligent Computing, Information and Control Systems (ICOIICS)*, Lalitpur, 19-21 November 2025, 1637-1642. <https://doi.org/10.1109/icoiics67115.2025.11390446>
- [3] Ghosh, R., Vamshi, C. and Kumar, P. (2019) RNN Based Online Handwritten Word Recognition in Devanagari and Bengali Scripts Using Horizontal Zoning. *Pattern Recognition*, **92**, 203-218. <https://doi.org/10.1016/j.patcog.2019.03.030>
- [4] 沈强, 李辉, 张燕. 基于小规模手写体汉字数据集的数据增强方法[J]. 北京化工大学学报(自然科学版), 2021, 48(1): 58-65.
- [5] Chen, W., Su, X. and Hou, H. (2025) Fine-Grained Automatic Augmentation for Handwritten Character Recognition. *Pattern Recognition*, **159**, Article 111079. <https://doi.org/10.1016/j.patcog.2024.111079>
- [6] Si, L., Guo, C., Li, Z. and Yang, Y. (2025) A Unified Framework of Data Augmentation Using Large Language Models for Text-Based Cross-Modal Retrieval. *Pattern Recognition*, **167**, Article 111755. <https://doi.org/10.1016/j.patcog.2025.111755>
- [7] Diaz, M., Mendoza-García, A., Ferrer, M.A. and Sabourin, R. (2025) A Survey of Handwriting Synthesis from 2019 to 2024: A Comprehensive Review. *Pattern Recognition*, **162**, Article 111357. <https://doi.org/10.1016/j.patcog.2025.111357>
- [8] Devi, S.N. and Fatima, N.S. (2024) Handwritten Optical Character Recognition Using Transrnn Trained with Self

-
- Improved Flower Pollination Algorithm (SI-FPA). *Multimedia Tools and Applications*, **84**, 19947-19969. <https://doi.org/10.1007/s11042-024-19758-9>
- [9] Imane, B., Alae, A., Ghizlane, K. and Mrabti, M. (2025) Enhancing Arabic Handwritten Word Recognition: A CNN-BiLSTM-CTC Architecture with Attention Mechanism and Adaptive Augmentation. *Discover Applied Sciences*, **7**, Article No. 460. <https://doi.org/10.1007/s42452-025-06952-z>
 - [10] Rastogi, A., Tiwari, S., Shafat Ali, S. and Wani, M.A. (2026) SCRAT-Net: An Attention-Based Deep Learning Architecture for Handwritten Word Recognition. *IEEE Access*, **14**, 18622-18639. <https://doi.org/10.1109/access.2026.3657953>
 - [11] Kass, D. and Vats, E. (2022) Attentionhtr: Handwritten Text Recognition Based on Attention Encoder-Decoder Networks. In: Uchida, S., Barney, E. and Eglin, V., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 507-522. https://doi.org/10.1007/978-3-031-06555-2_34
 - [12] Poulos, J. and Valle, R. (2021) Character-Based Handwritten Text Transcription with Attention Networks. *Neural Computing and Applications*, **33**, 10563-10573. <https://doi.org/10.1007/s00521-021-05813-1>
 - [13] Wang, T., Zhu, Y., Jin, L., Luo, C., Chen, X., Wu, Y., *et al.* (2020) Decoupled Attention Network for Text Recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 12216-12224. <https://doi.org/10.1609/aaai.v34i07.6903>
 - [14] Coquenat, D., Chatelain, C. and Paquet, T. (2023) End-to-End Handwritten Paragraph Text Recognition Using a Vertical Attention Network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45**, 508-524. <https://doi.org/10.1109/tpami.2022.3144899>