

基于机器视觉的二简字档案识别检测算法研究

李 冉¹ 刘 峰² 董近东² 郭 嘉² 王晓露²
LI Ran LIU Feng DONG Jindong GUO Jia WANG Xiaolu

摘 要

针对历史档案数字化进程中“第二次汉字简化方案”字符（以下简称“二简字”）识别鲁棒性不足的问题，文章提出一种融合全局感知能力的深度学习识别框架。首先通过系统表征二简字的形态演变规律与视觉特征，构建了涵盖2038例手写样本与2946例印刷样本的专项标准数据集，并实施了像素级标注与图像归一化预处理。在检测算法层面，对YOLOv12架构进行了针对性改进：通过集成Bottleneck Transformers模块，建立了特征图的全局上下文依赖关系，弥补了传统卷积算子在处理复杂异形字结构时感受野受限的缺陷；同时引入锐度感知最小化（SAM）优化算法，通过平滑损失景观提升模型在非标准书写场景下的泛化性能。实验结果显示，改进模型在二简字检测任务中的平均精度（AP50）达到94.8%，较基准模型在AP@50:95指标上提升了4.1%。不仅为特定历史文献的智能化解析提供了高精度的技术支撑，也为非标准字符集的自动机识别提供了理论参考。

关键词

二简字；YOLO；Transformer；CPCA；SAM

doi: 10.3969/j.issn.1672-9528.2026.04.042

0 引言

在档案资源数字化进程中，一批形成于特定历史时期的非常规汉字对现有识别技术构成了现实挑战。这类字形既不同于传统繁体字，亦有别于当前通用的规范简化字，结构介于偏旁与别字之间，广泛见于20世纪六七十年代的手写与印刷文献。经考证，此类字形对应于《第二次汉字简化方案（草案）》中颁布的“二简字”。

“二简字”是汉字简化历程中一次具有广泛社会影响的简化实践，该方案于1977年正式试行，旨在通过进一步减少笔画提高书写效率。然而，由于部分字形简化过度，在实际应用中引发了识读与使用上的混乱，国务院于1986年正式废止该方案，使其成为特定历史阶段的文字遗存。

该类非标准字形在档案文献中大量存在，已对当前数字化过程中的OCR识别、内容理解及深层利用造成实际障碍。因此，开展面向二简字的机器识别研究，实现其准确检测与辨识，已成为突破该时期档案全文数字化瓶颈的关键技术问题之一。本文研究响应《“十四五”全国档案事业发展规划》中关于提升档案数字化智能识别能力的要求，探索计算机视

觉技术与档案学方法的融合路径，旨在为历史文献的抢救性保护与智能化开发提供有效支撑，助力国家记忆工程的技术体系建设。

1 背景

随着历史档案数字化进程的深入推进，二简字作为一类已废止的非标准字符，其机器识别问题日益突出。当前主流光学字符识别方法在规范汉字识别方面已趋于成熟，但由于二简字未被纳入通用字符编码体系，现有OCR系统普遍缺乏相应的字符映射与识别能力，导致在该类文字上的识别准确率显著偏低。

学术界的相关研究呈现明显的“人文”与“技术”两条发展线路。一方面，语言学与历史学领域对二简字的关注较多，主要集中于其规范探讨^[1]、社会应用^[2]及形体演变^[3]等方面，积累了宝贵的本体认识^[4-6]。另一方面，在计算机视觉与信息处理领域，尽管OCR技术本身发展迅猛，研究热点涵盖复杂版式分析^[7]，多语言文档理解^[8-9]，乃至针对特定文字的识别优化^[10]，但迄今尚未研究直接面向“二简字档案”这一特定对象的智能识别问题。

这一空白主要源于两大挑战：一是缺乏公开、系统的二简字样本库，制约了数据驱动模型的应用；二是其字形简化规则与手写变体复杂，导致通用模型难以适配。因此，亟须开展以构建专项数据集和开发针对性识别模型为核心的交叉研究。

1. 江苏联合职业技术学院徐州财经分院 江苏徐州 221008

2. 徐州市档案馆 江苏徐州 221016

[基金项目] 江苏省档案科技项目“基于机器视觉的二简字智能识别技术在档案全文数据建设中的应用研究”阶段性研究成果（2024-16）

本研究旨在回应上述现实需求,通过填补二简字智能识别技术空白,为历史档案的完整、准确数字化提供关键技术支撑,兼具理论意义与应用价值。

2 理论与技术概述

为实现档案中二简字的精准辨识,本研究确立了以机器视觉为核心的技术路线。通过图像获取、预处理、特征提取到分类决策的完整流程,为将纸质档案中的非标准字形转化为可处理的文本数据提供了基础框架。然而,二简字自身特性给通用识别框架带来三重主要挑战:一是其形体系统性地偏离繁体与现行规范简体,易导致基于标准字库的模型误判;二是简笔化设计削弱了汉字固有的区分性特征,如笔画合并与省略,增加了稳定特征提取的难度;三是其在短暂流通期内形成的印刷与多样手写变体,要求模型具备应对字形不确定性的泛化能力。

为应对这些挑战,本研究引入深度学习进行自适应特征学习与模式识别。卷积神经网络(CNN)能自动提取从笔画到部件的层次化结构特征,而注意力机制可动态聚焦于关键区域并建模长距离依赖,两者协同有助于在字形失真或背景复杂时仍实现可靠判读。因此,融合机器视觉的系统流程与深度学习的表征能力,构成了实现二简字高效准确识别的可行路径。通过构建针对性数据集并设计适配模型,本研究旨在突破档案全文数据化的关键瓶颈。

3 研究方法

3.1 数据采集与预处理

本研究以 20 世纪六七十年代留存的历史档案为研究对象,针对其中的二简字构建专用数据集。原始数据来源于徐州市档案馆,共计收录文书档案页面中 2 038 例手写体与 2 946 例印刷体二简字样本。所有档案页面均通过专业平板扫描仪进行数字化采集,图像分辨率满足后续模型输入要求。标注环节由具备档案学背景的专业人员使用 Labelme 工具对图像中的二简字区域进行精确标注,确保标注质量。为支撑模型训练与性能评估,将标注完成的数据按 4:1 的比例随机划分为训练集与测试集。在模型输入前,对图像进行标准化预处理:统一缩放至 $512\text{ px}\times 512\text{ px}$,并将像素值归一化至 $[0, 1]$ 区间。此外,采用水平翻转、随机缩放及亮度调节等数据增强策略,以扩充训练样本的多样性,降低模型对成像条件变化的敏感性,提升其在真实档案图像中的泛化能力。

3.2 模型结构

为精准捕获档案图像中二简字的微观结构特征,本研究对 YOLOv12^[11] 检测框架进行了深度定制化改良。尽管原生 YOLOv12 利用区域注意力机制在计算负载与感知野之间取得了权衡,但在针对档案扫描件中常见的小尺度、形态畸变目标时,其深层特征的语义判别力仍显不足。

本研究引入(bottleneck transformers, BoT)^[12-13] 设计范

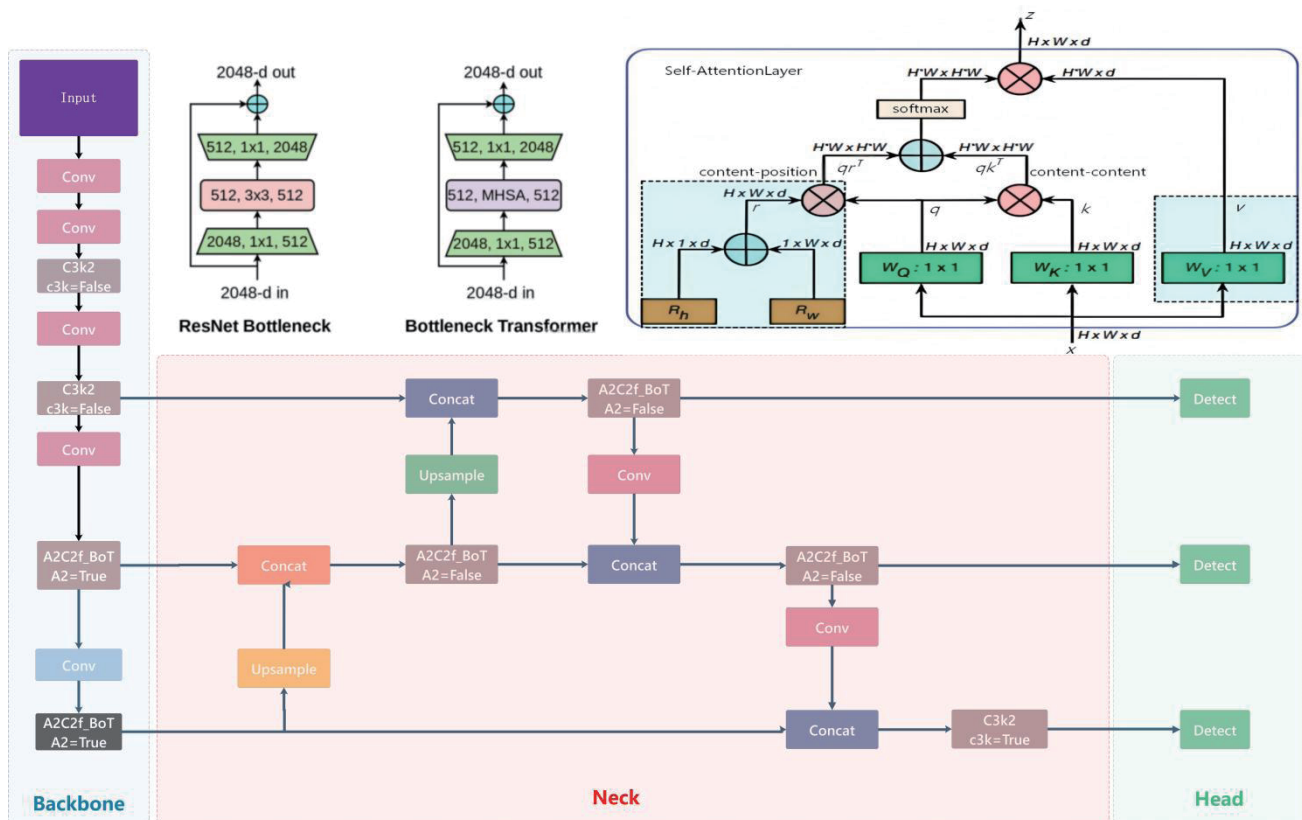


图 1 YOLOv12+Bottleneck Transformers 网络结构图

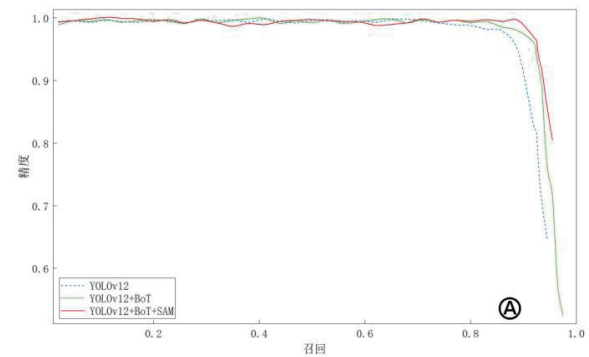
式，构建了一种卷积与自注意力机制耦合的混合骨干网络。具体而言，在模型的高级特征图阶段，利用多头自注意力（multi-head self-attention, MHSA）模块替代传统的空间卷积层。该设计的逻辑在于：底层卷积算子能够通过局部归纳偏置（Inductive Bias）高效提取纹理与边缘等初级特征，而顶层 MHSA 则通过全局感受野建模长距离空间依赖，有效聚合二简字与其上下文的语义关联。

此外，为补偿自注意力机制对空间拓扑结构的感知缺失，集成了二维相对位置编码（2D Relative Position Encoding），以增强模型对复杂笔画构成的细粒度解析能力。改进后的架构实现了从单一卷积模式向“局部感知 - 全局整合”协同范式的转变，在保持实时推理效率的同时，显著提升了模型对低分辨率目标的判别阈值。具体拓扑结构如图 1 所示。

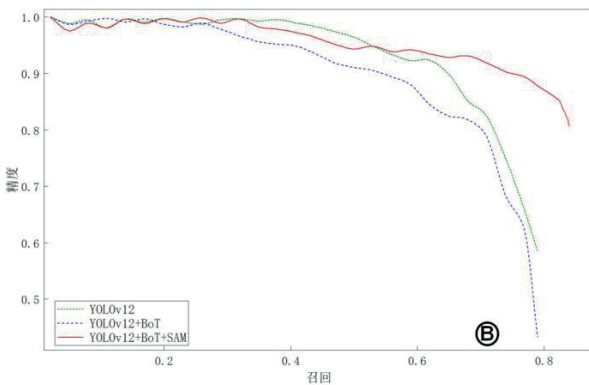
4 实验结果

为验证 BoT 模块与 YOLOv12 模型集成后对二简字检测性能的影响，本研究在训练过程中引入了锐度感知最小化（sharpness-aware minimization, SAM）优化器^[14]，以提升模型在测试集上的泛化能力并抑制过拟合风险。

图 2 展示了基准模型与各改进模型在不同交并比（IoU）阈值下的精确率 - 召回率曲线。



(A) PR 曲线 (AP50)



(B) PR 曲线 (AP75)

图 2 基础模型及不同改进模型的 PR 曲线图

当 IoU 阈值设为 0.5 时，所有模型均能有效检测；

其中，集成 BoT 模块后的模型在不同目标尺寸上的适应性增强，在召回率方面提升尤为明显，说明该模块有助于提高目标的检出率。在更为严格的 IoU 阈值 0.75 条件下，改进模型在目标边缘定位精度上表现出更优的性能，反映出其对局部细节特征保持能力更强，有利于实现高精度的目标边界定位。

综合引入 BoT 模块与 SAM 优化器后，模型在保持基准检测能力的基础上，其检测精度与召回性能均得到显著提升。定量结果验证了所提方法在二简字数据集上的有效性。需要指出的是，在面对极端尺度变化的目标时，SAM 优化器所带来的精度增益略低于预期，但该结果为本研究后续优化方向提供了明确的实证参考。

模型性能的定量评估结果如表 1~2 所示。相较于原始 YOLOv12，单独引入 BoT 模块使综合指标 AP@50:95 提升了 1.5%；在此基础上结合 SAM 优化器，模型整体性能再提升 0.5%，表明两项改进在二简字检测中具有正向叠加效应。在与多种主流检测模型的对比中，本文所提改进模型在 AP50 指标上达到 94.8%，显著优于对比模型，验证了该方法在二简字识别任务中的有效性。该性能提升主要归因于 BoT 模块增强的全局上下文建模能力及其与局部特征的有效协同，二者共同提升了模型对不同尺度目标的判别鲁棒性。

表 1 基础模型及不同改进模型的平均精度对比

单位：%			
Model	AP 50%	AP 75%	AP@50:95
YOLOv12	92.9	86.3	71.3
YOLOv12+BoT	94.3	88.1	73.1
YOLOv12+BoT+SAM	94.8	89.1	74.2

表 2 不同现有模型和改进模型的平均精度对比

单位：%			
Model	AP 50%	AP 75%	AP@50:95
YOLOv7	78.3	64.9	58.5
YOLOv8	83.3	69.6	61.3
YOLOv11	87.3	73.3	64.8
YOLOX	90.2	76.1	65.3

图 3 以可视化方式对比了逐次改进过程中模型检测效果的变化。第一组（图 3A、3B）显示，基准 YOLOv12 模型在部分样本中因特征提取不充分而导致目标漏检。第二组（图 3C、3D）引入特征金字塔网络（FPN）结构后，通过多尺度特征融合提升了检测的稳定性，漏检情况得到有效缓解。第三组（图 3E、3F）进一步集成 Bottleneck Transformers 模块，该模块在抑制背景干扰与形似字符误检方面发挥了明显作用，检测结果的定位精度与分类置信度均有所提高。第四组（图 3G、3H）结合锐度感知最小化（SAM）优化器后，

模型对样本中形态差异较大的罕见字符仍保持了稳定的检测性能,反映出其良好的泛化能力。



图3 不同优化方法的可视化效果图

5 讨论

本研究构建了一种融合 Bottleneck Transformers 与 YOLOv12 的检测框架,用于历史档案中二简字的识别任务。该框架中, YOLOv12 作为基础检测器提供高效的目标定位能力,而 BoT 模块通过其卷积-自注意力混合机制,增强了模型对全局上下文信息与长距离依赖关系的建模能力,从而提升了二简字在复杂背景和多尺度条件下的辨识与定位精度。

为进一步提升模型在大规模二简字数据集上的泛化能力,本研究在训练阶段引入锐度感知最小化(SAM)优化器。该优化器通过同时最小化损失值及其曲率锐度,引导模型收敛至更平坦的损失极小值区域,从而增强其在未见样本上的适应性与鲁棒性。

实验结果显示,本文提出的改进模型在二简字检测任务中取得了显著性能提升: AP50 指标由原始 YOLOv12 的 92.9% 提升至 94.3%, AP@50:95 指标整体提升 4.1%。相较于现有方法,本方法在检测精度上具有明显优势,验证了其作为二简字目标检测解决方案的有效性。本研究不仅为历史档案中非标准字符的识别提供了可行的技术路径,也为面向类似小尺寸、复杂结构字符的检测任务提供了具有参考价值的架构设计思路。

参考文献:

- [1] 刘佳. 重新认识“二简字”: 利用库普尔的八问方案探索第二次汉字简化[J]. 青年文学家, 2011(2):125-126.
- [2] 徐贵水, 毕志蓉. 二简字及民间简化字邮戳探析[J]. 上海集邮, 2022(1):24-29.
- [3] 王开扬.《规范汉字表》可以吸收少量“二简”字[J]. 现代语文, 2006(4):23-25.
- [4] 郑丽. 二简字探析[J]. 吉林广播电视大学学报, 2014(11): 145-147.
- [5] 朱玉金.《第二次汉字简化方案(草案)》研究[D]. 济南: 山东师范大学, 2005.

- [6] MEMON J, SAMI M, KHAN R A, et al. Handwritten optical character recognition (OCR): a comprehensive systematic literature review (SLR)[J]. IEEE access, 2020, 8:142642 - 142668.
- [7] LIU Y L, YANG B, LIU Q, et al. TextMonkey: an OCR-free large multimodal model for understanding document[J]. IEEE transactions on pattern analysis and machine intelligence, 2026, 48(5): 6008-6019.
- [8] MOSBAH L, MOALLA I, HAMDANI T M, et al. ADOCRNet: a deep learning OCR for arabic documents recognition[J]. IEEE access, 2024:55620 - 55631.
- [9] PATEL C I, PATEL A, PATEL D. Optical character recognition by open source OCR tool tesseract: a case study[J]. International journal of computer applications, 2012, 55(10): 50-56.
- [10] RABBY A S A, ALI H, ISLAM M M, et al. Enhancement of bengali OCR by specialized models and advanced techniques for diverse document types[EB/OL]. (2024-02-07)[2026-04-08]. <https://doi.org/10.48550/arXiv.2402.05158>.
- [11] BUMBÁLEK R, UFITIKIREZI J D D M, UMURUNGI S N, et al. Computer vision in precision livestock farming: benchmarking YOLOv9, YOLOv10, YOLOv11, and YOLOv12 for individual cattle identification[J]. Smart agricultural technology, 2025, 12:101208.
- [12] AHMAD S M K M A. Emotion recognition through multimodal bottleneck transformers: the integration of speech and facial data to improve precision[EB/OL]. (2024-01-06)[2026-04-09]. <https://inass.org/wp-content/uploads/2025/01/2025043011-2.pdf>.
- [13] JI B, XU J F, LIU Y, et al. Improved YOLOv8 for small traffic sign detection under complex environmental conditions[J]. Franklin open, 2024, 8:100167.
- [14] RATCHATORN T, TANAKA M. Improving sharpness-aware minimization using label smoothing and adaptive adversarial cross-entropy loss[J]. IEEE access, 2025, 13:100326 - 100337.

【作者简介】

李冉(1982—), 男, 江苏徐州人, 博士研究生, 教师、副教授, 研究方向: 人工智能技术、网络操作系统、网络通信技术、网络安全技术等。

(收稿日期: 2025-12-09 修回日期: 2026-04-14)