

基于卷积神经网络的音像与手写体档案智能识别及语义化归档系统

李 彤 邹德芳

沈阳建筑大学 机械工程学院 辽宁沈阳 110168

摘要: 本文研究基于卷积神经网络的音像档案与手写体档案智能识别及语义化归档系统, 解决数字化转型中海量历史档案处理的挑战。在语音识别模块中, 采用改进的ResNet-50结合双向LSTM与自注意力机制, 提升语音转写准确性, 并通过对抗训练增强方言鲁棒性。在手写体OCR模块中, 利用MobileNetV3和CycleGAN扩增小样本数据, 提高识别精度。为确保档案安全, 引入Hyperledger Fabric构建分层哈希链, 实现高效存证与篡改检测。实验表明, 该系统在字符错误率、词错误率及篡改检测准确率等关键指标上均优于传统方法, 推动档案管理向自动化、知识化与可信化转型, 为未来研究提供了重要参考和技术支持, 具有广泛的应用前景。

关键词: 卷积神经网络; 语音识别; 手写体OCR; 区块链存证; 多模态融合

引言

随着数字化转型加速推进, 传统纸质和音像档案逐渐向数字形式转变, 不仅极大地提升了信息存储的效率, 也为档案的智能化管理和利用提供了可能。然而, 如何高效、准确地处理和解析海量的历史档案资料, 尤其是那些包含复杂手写体以及多语种、多方言的音像资料, 成为了档案管理领域亟待解决的问题。传统方法依赖于人工识别与分类, 耗时费力且难以避免主观误差。因此, 探索新的技术手段以实现自动化、智能化的档案处理显得尤为重要。近年来, 深度学习技术, 特别是卷积神经网络(CNN)的发展为这一难题带来了曙光。将AI技术应用于档案管理中, 可以大幅提高档案处理的速度和准确性, 还能挖掘出更多隐藏在档案背后的价值信息。

基于此背景, 研究基于卷积神经网络的音像档案与手写体档案智能识别及语义化归档系统, 聚焦两大核心问题: 一是音像档案中语音与视频内容的高效特征提取与语义分析; 二是手写体档案的高精度识别与上下文关联分析。为此, 提出多项创新方案: 采用改进的ResNet-50模型提取音频频谱图与视频关键帧特征, 结

合双向LSTM与自注意力机制实现语音分段与场景标注; 利用MobileNetV3构建轻量化模型, 并通过CycleGAN扩增小样本数据, 提升手写体识别准确率。同时, 引入区块链技术, 基于Hyperledger Fabric构建分层哈希链, 实现档案的可信存证与安全溯源。该系统显著提升档案处理效率与质量, 为档案管理的智能化发展提供技术支撑。

一、音像与手写体档案智能处理技术的现状与趋势

随着信息技术的飞速发展, 档案管理领域正经历着深刻的变革。传统的纸质档案和音像资料逐渐被数字化替代, 在这过程中, 如何高效、准确地处理海量的历史档案资料, 成为了研究者们关注的重点。近年来, 深度学习技术, 特别是卷积神经网络的发展为解决这些问题带来了新的希望。

(一) 国内外音像档案智能处理技术进展

在国内音像档案的智能化处理已取得显著进展。科大讯飞与安徽省档案局开发的“档案超脑”系统, 利用卷积神经网络实现语音转写准确率92%, 并支持人脸检测与检索, 但在多模态融合方面存在局限。东北财经大学通过DeepSeek R1模型对700余小时声像档案进行自动化转写和语义分段, 解决了非结构化数据治理难题, 但大规模异构数据标准化仍待突破。江苏省太仓市智慧档案馆将AI用于库房安防, 提升档案实体管理的安全性。

国际上音像档案智能处理技术聚焦于多模态融合与复杂场景泛化。2016年He等人的ResNet-50模型缓解

基金项目: 2025年辽宁省档案科技项目: 基于卷积神经网络的音像档案与手写体档案智能识别及语义化归档系统研究(2025-X-110); 沈阳建筑大学第十一批教育科学研究立项课题“校企合作虚实融合的机械类新工科实验实践教学体系研究”。

了深层网络的梯度消失问题，显著提升图像识别性能。2017年Vaswani的Transformer架构增强了序列建模效果，成为语音识别热门选择。2020年Park的Conformer模型结合CNN和Transformer优点，简化端到端语音识别训练流程，为音像档案智能化处理提供有力支持。

（二）手写体档案智能处理技术进展

在国内的手写体档案智能识别方面，清华大学的研究团队提出了基于PP-YOLOv2的轻量化模型PaddleOCR，采用MobileNetV3和深度可分离卷积技术，显著减少模型参数并提高推理速度，适合处理大量手写体数据。团队还利用CycleGAN进行数据扩增，增强小样本数据的鲁棒性和泛化能力，减少了人工标注成本，并提高了连笔字及中/满/藏文混合场景的识别率，达到93%以上的精度。该方法为历史文献的高效数字化提供了可靠技术路径，在实际应用中展现出良好的适应性与推广价值。

国际上，2009年Scarselli等人提出的图神经网络则为多模态数据的融合提供了新思路，通过构建知识图谱，实现了人物、事件、时空之间的三维关联模型，大大提升了档案检索的精准度和用户体验。2014年Goodfellow等人提出的生成对抗网络也被用于增强手写体数据集，通过生成逼真的新样本来扩增训练集，进而提高模型的鲁棒性和泛化能力。2021年Radford等人提出的CLIP模型通过自然语言监督学习，实现了视觉模型的跨域迁移，展示了强大的泛化能力。

（三）多模态融合与区块链存证的前沿探索

面对日益增长的多模态档案数据，如何实现高效、准确的语义化归档成为了一个重要的研究方向。邹德芳等人采用异构图神经网络融合音像、手写体档案与元数据的多模态特征，建立“人物-事件-时空”三维关联模型，实现了“以人搜档”、“以事件搜档”的智能检索。此外，基于Hyperledger Fabric的分层哈希链技术，不仅保障了档案的安全性和完整性，还实现了自动化的合规校验与权限审计，篡改检测准确率高达99.99%。

在国际上，多模态融合与区块链存证研究取得显著进展。Google Brain的M4C模型通过自注意力机制处理图像和文本信息，实现跨模态特征对齐并生成描述性文字，适用于历史图片或手写文档转换为结构化文本，便于检索和知识挖掘。IBM Research开发的Crypto Anchor Verifier结合区块链和AI算法，用于验证物品的真实性和来源，其原理同样适用于电子档案的安全存储与验证。

通过将档案的SHA-256哈希值记录在区块链上，任何篡改行为都会被立即发现，因为修改后的文件会产生不同的哈希值，无法匹配区块链上的记录，有效保障了电子档案的安全性和完整性。

二、方法与技术实现

（一）数据收集与预处理

数据的获取和处理是至关重要的第一步，为了构建一个全面且多样化的训练集，从多个合作单位收集了大约1000小时的历史音像档案，包括方言、模糊录音以及20万页的手写体档案，覆盖汉、满、藏文及草书场景。这些档案包含了丰富的信息，还代表了不同历史时期、文化背景下的书写习惯与语言特点，为模型提供了广泛的样本空间。

对于音像档案的数据预处理，采用多种方法提升音视频质量。音频方面使用谱减法降噪，并结合语音活动检测分割有效语音段，降低背景噪声影响。视频方面通过光流法结合ResNet-50特征相似度筛选关键帧，生成时序对齐的频谱图序列，便于后续分析。手写体档案利用CycleGAN进行风格迁移以扩增数据，并应用随机仿射变换增强泛化能力。为应对实际中常见的褪色、污损等问题，设计融合谱减法与空间-时序注意力机制的处理方法，有效抑制噪声干扰，提升模型对低质量输入的适应性，保障训练的稳定性与鲁棒性。

（二）模型架构设计

为提升音像与手写体档案的识别精度及上下文关联性，设计了涵盖语音识别、OCR与语义检索的深度学习模型。语音识别基于DeepSpeech2框架，采用改进的ResNet-50并嵌入SE注意力模块，结合双向LSTM和CTC损失函数，优化语音转写效果，并引入对抗训练增强方言鲁棒性，CER和WER优于传统HMM模型。手写体OCR以PP-OCRv5为基础，结合MobileNetV3构建轻量化CNN模型，通过动态笔画分解模块解析连笔结构，并引入Transformer上下文网络，利用清代奏折数据训练，F1值评估与A/B测试显示“王/玉”等字符误判率降低62%，验证了上下文建模的有效性。

在语义化归档与检索层面，提出基于Neo4j图数据库和异构图神经网络的智能档案管理系统，通过HGNN深度整合多模态数据，构建“人物-事件-时空”三维知识图谱，支持“以人搜档”“以事件搜档”等智能检索模式。在数据安全方面，采用Hyperledger Fabric区块链技术，通过主链存哈希、侧链存数据的分层架构，确保不

可篡改。性能测试显示,系统在200并发下响应时间低于0.5秒,篡改检测准确率超99.99%,验证了系统的高效性与可靠性。该方案为多模态档案管理提供了高性能、高安全的技术路径。

(三) 实验设计

为了验证上述技术和模型的有效性,我们在实验设计阶段进行了详细的规划。设置了实验环境设置,所有实验均在配备了NVIDIA A100 GPU集群和PB级存储资源的实验室环境中进行,以确保算法训练和大规模数据管理的需求得到满足。实验平台选用了PyTorch和TensorFlow作为主要的深度学习框架,并利用PaddleOCR、Hyperledger Fabric等开源工具链支持模型开发和区块链部署。

针对不同任务设定相应性能指标,语音识别以CER和WER为主,手写体OCR侧重F1值与上下文修正效果,区块链存证可靠性通过篡改检测准确率衡量,并辅以用户满意度调查收集反馈。实验分四阶段推进:第一阶段完成数据采集与清洗,构建多模态训练集;第二阶段优化模型架构,训练语音识别与OCR模块;第三阶段集成区块链存证,开发语义检索API并开展系统联调;第四阶段在三家省级档案馆试点部署,优化体验并输出技术标准。各阶段均进行评估与调整,确保系统性能与应用效果达到预期目标。

三、结果分析与讨论

(一) 主要发现

本研究在语音识别、手写体OCR与区块链存证三方面均取得显著成效。语音识别模块通过改进ResNet-50与双向LSTM结合CTC损失优化,普通话转写准确率超95%,较传统HMM提升约3个百分点,对抗训练使吴语和粤语误识率分别降低15%和20%。手写体OCR采用MobileNetV3轻量化模型,结合动态笔画分解与Transformer上下文网络,规范文本识别率达93%以上,处理少数民族文字与草书时整体错误率下降62%,"王/玉"等易混字误判显著减少。区块链存证基于Hyperledger Fabric构建分层哈希链,支持200路并发,响应时间低于0.5秒,篡改检测准确率达99.99%。系统在准确性、鲁棒性与安全性方面表现优异,为档案智能化

管理提供了高效可靠的技术路径。

(二) 对比分析

在多模态信息融合方面,突破传统单一模态处理局限,通过改进的ResNet-50提取音频频谱图与视频关键帧特征,结合双向LSTM与自注意力机制,实现音、视、文信息的深度融合,显著提升复杂场景识别准确率。针对历史手写体档案数据稀缺、风格多样问题,采用轻量化MobileNetV3结合深度可分离卷积降低计算开销,并引入CycleGAN进行风格迁移与数据扩增,有效缓解小样本训练难题,模型参数量仅为传统OCR的1/5,兼顾精度与部署效率。在档案安全与知识组织方面,创新引入Hyperledger Fabric构建分层哈希链,实现链上链下协同存证,增强数据完整性与可追溯性;同时,利用异构图神经网络构建“人物-事件-时空”三维知识图谱,支持跨模态语义映射与智能检索,突破关键词检索局限,为档案的深度挖掘与关联分析提供有力支撑。

结论

本研究构建了一套基于卷积神经网络的音像与手写体档案智能识别系统,通过融合深度学习与多模态分析技术,显著提升了历史档案数字化水平。改进的ResNet-50与双向LSTM模型提高了语音识别准确率,MobileNetV3与CycleGAN框架解决了手写体OCR的小样本训练难题。系统通过异构图神经网络构建“人物-事件-时空”三维知识图谱,实现跨模态语义关联与智能检索。同时,引入Hyperledger Fabric与IPFS分层存证机制,篡改检测准确率提高,为电子档案安全保存提供保障。该系统为档案馆等机构提供了可落地的技术方案,未来有望在司法文书管理、历史研究等领域发挥更大价值。

参考文献

- [1]Radford, A., Kim, J. W., Hallacy, C., et al. (2021). Learning transferable visual models from natural language supervision. Proceedings of the 38th International Conference on Machine Learning, 8748-8763.
- [2]杨文刚.卷积神经网络在图像档案管理中的应用研究[J].中国档案,2024,(03):46-47.