

基于深度学习算法的碑文提取与修复系统

李超¹ 李思樵¹ 张靖熙¹ 杨柳¹ 邹豪¹ 张攀¹

LI Chao LI Siqiao ZHANG Jingxi YANG Liu ZOU Hao ZHANG Pan

摘要

当前碑文及其拓本等非印刷字体文字的数字化录入仍需手动进行。据此，首先使用 YOLOv5 模型对整体图像中的文字进行精确检测和定位；随后基于 UNet++ 模型实现单一文字图片中前景的提取，并使用 Rewrite2 模型对可能残缺的文字图片进行修复；最后基于 PaddlePaddle OCR 模型完成图片中文字的识别，获得可编辑状态汉字。在数据集中的实验结果表明，系统能很好地完成文字的检测定位、前景提取、残缺修复、识别等过程，全流程整体的准确率达到 90.6%。

关键词

碑文数字化录入；文字检测定位；文字前景提取；残缺文字修复；文字识别

doi: 10.3969/j.issn.1672-9528.2024.10.047

0 引言

随着经济社会的不断发展，文化软实力正日益成为综合国力的重要组成部分，以传统碑文、木雕等为载体的文献、

书法研究等越来越多地进入大众的视野。文字作为其核心，是中华文化的瑰宝，相关数字化保护和传承对于弘扬中华文化具有重要意义。因此，有必要对古籍、碑文等进行数字化处理，方便存储检索和共享，从而辅助科学研究，加快传播效率。

然而，对于其中存在的大量非印刷字体文字，现有文本识别软件仍然无法有效处理，其数字化采集仍需手动进行，

1. 内江师范学院人工智能学院 四川内江 641100

[基金项目] 2023 年度内江市基础研究与应用基础研究项目
“隆昌石牌坊碑文数字化提取与残缺修复研究”

诸多挑战。随着电力网络结构的日益复杂和新能源的迅速崛起，火电厂需要不断调整其运行策略以应对多变的市场环境和电力需求。为了克服这些挑战，需要继续深入研究大数据分析技术，并探索其在火电厂调频调峰优化中的更多应用。需要提高数据分析的准确性和效率，以便更快速地响应电力市场的变化。同时，还需要关注新技术的发展，如人工智能和机器学习，以进一步优化本文的分析模型。此外，还需要与电力系统运营商和可再生能源生产商进行紧密合作，共同制定和实施调频调峰策略，以便更好地理解电力市场的需求和变化，从而制定出更有效的优化方案。综上所述，基于大数据分析的火电厂调频调峰优化研究是一个复杂而重要的课题。需要持续投入研究力量，以应对未来的挑战，并推动电力行业的可持续发展。

参考文献：

- [1] 李军徽, 张靖祥, 穆钢, 等. 辅助服务市场下独立储能调峰调频协同优化调度[J/OL]. 中国电机工程学报, 1-15[2024-03-08]. <http://kns.cnki.net/kcms/detail/11.2107.TM.20240517.1333.010.html>.
- [2] 李扬, 杨鹏, 吕瑞瑞. 基于数据驱动优化的火电发电机组

调峰调频方法研究[J]. 科技创新, 2023(26):9-12.

- [3] 李项. 微电网背景下的燃气轮机和光储调频调峰一体化研究[J]. 电气时代, 2023(10):58-61.
- [4] 高奎, 孔庆龙, 谭祥帅, 等. 基于现货市场下的燃气—蒸汽联合循环机组辅助调频协调控制系统优化研究与应用[J]. 电气技术与经济, 2023(1):36-44.
- [5] 宁剑, 吴继平, 江长明, 等. 考虑资源运行特性的可调节负荷调峰调频优化控制策略[J]. 电力系统自动化, 2022, 46(15): 11-19.
- [6] 李秀慧, 崔炎. 考虑调峰调频需求的新能源电网储能优化配置[J]. 储能科学与技术, 2022, 11(11):3594-3602.
- [7] 王亚莉, 叶泽, 黄际元, 等. 基于动态峰谷时段划分的储能调峰调频经济调度研究[J]. 中国电力, 2022, 55(8):64-72.
- [8] 陈静鹏, 张鸿轩, 张勇, 等. 面向区域调频辅助服务市场的南方电网统一调频控制区设计与策略优化[J]. 电网技术, 2022, 46(7): 2657-2664.

【作者简介】

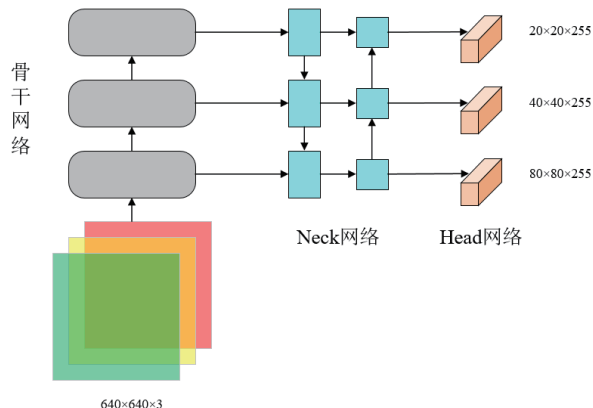
夏稀渊(1972—), 男, 浙江上虞人, 本科, 高级工程师, 研究方向: 电力技术。

(收稿日期: 2024-07-04)

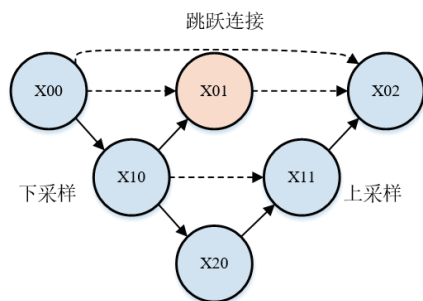
耗时长,成本高。因此,本文基于最新深度学习,设计了一款能够支持文字的修复与识别功能的软件系统,提高了相关文字的数字化录入效率。该系统主要实现包括文字的有效检测与定位、前景提取、残缺修复、识别为可编辑文本等功能,其中,检测与定位过程采用 YOLO v5^[1] 算法模型,前景提取过程采用 UNet++^[2] 算法模型,残缺修复过程采用 Rewrite2^[3] 算法模型,识别为可编辑文本过程采用百度的 PaddlePaddle OCR^[4] 算法模型,最终在构建的古迹与碑文文字数据集中的测试实验验证了系统功能的有效性。

1 相关原理

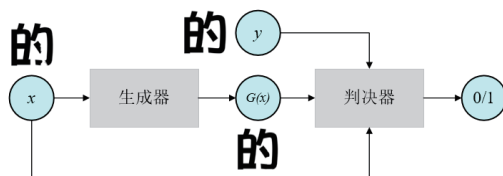
YOLO 系列模型是当前学术界进行目标检测研究的焦点,也是工程落地最成功的算法之一。如图 1 (a) 所示, YOLOv5 模型的核心结构主要由卷积运算搭建的骨干网络、包含尺度缩放的 neck 网络、三个分支的 head 网络构成,其主要在 head 网络的每个格点对目标的分类、位置、置信度等进行直接预测。后续系列其他模型都针对多尺度目标检测问题而改进,考虑到碑文及其拓本中的文字通常包含多种尺寸,因此选择 YOLOv5 模型。UNet++ 模型的核心结构如图 1 (b) 所示,其主要在多个尺度进行特征压缩与解压,并在压缩与解压的多个尺度上进行特征的拼接融合,从而极大提高了图像分割的准确率。GAN 网络是生成式算法使用的重要模型之一,如图 1 (c) 所示,其主要由生成器、判决器构成,在两者的对抗训练中提升生成器性能。本文参考 Chang 等人基于 GAN 原理提出的 Rewrite2 模型而改进,主要将其汉字风格迁移数据集替换为残缺文字修复数据集,从而实现残缺碑文的修复。基于深度学习的 OCR 模型框架通常需要经过图像预处理、文字检测、文本识别等过程,本文采用 PaddlePaddle OCR 算法提出的 StrucTexT 多模态预训练模型而实现,支持字符级提取、字段级提取、字段关系预测等功能,其核心框架如图 1 (d) 所示。



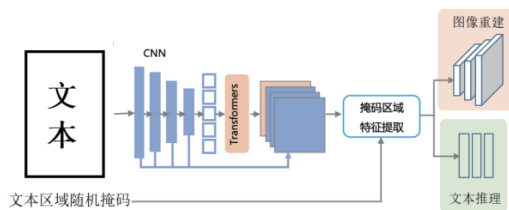
(a) YOLOv5 核心框架



(b) UNet++ 核心结构



(c) GAN 模型原理



(d) OCR 核心框架

图 1 模型原理

2 系统方案

系统整体方案如图 2 所示,待检测的碑文拓本图像输入到 YOLOv5 模型中实现文字的检测定位;基于位置信息切分后的每个文字图片被输入到 UNet++ 模型中,实现文字前景提取与背景分离;拓印噪声、残缺文本等在拓本文字中经常出现,为此每个分割出的碑文文字将输入到 Rewrite2 模型中,从而去除噪声并修复残缺;为了进一步对文本进行数字化存储,相应有效文字图片将输入到 PaddlePaddle OCR 模型中实现可编辑文本的输出,从而间接实现识别错误文字的人工纠正。

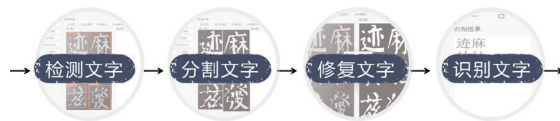


图 2 系统方案

按照系统方案实现的文字提取与修复软件系统如图 3 所示。整体界面分为四个部分,包括主菜单栏、输入图片显示栏、输出图片显示栏、OCR 识别为可编辑文本显示栏。主菜单中可对系统方案的四个阶段分阶段运行或全部运行,从而在输入图片显示栏、输出图片显示栏展现出不同的输入信息与输出信息。



图3 软件系统界面

3 实验

3.1 数据集及评价指标

为了验证选取四种模型构建的方案的有效性,系统单独对每个模块及整体进行了相关实验。实验构建的数据集共包含 4315 个汉字图片,分别完成了检测定位、分割、修复等数据标签标注,并按照 6:2:2 划分为训练集、验证集、测试集。对于文字检测定位模块主要评价其精确率 P ,其公式为:

$$P = \frac{TP}{TP+FP} \quad (1)$$

式中: TP 表示正样本被预测为正的数、FP 表示负样本被预测为正的数。

对于文字前景分割与残缺文字修复,主要计算结构相似度 (structural similarity index measure, SSIM),其公式为:

$$SSIM(y, x) = \frac{(2u_y u_x + c_1)(2\delta_{yx} + c_2)}{(u_y^2 + u_x^2 + c_1)(\delta_y^2 + \delta_x^2 + c_2)} \quad (2)$$

式中: x 表示提取的碑文前景图像; y 表示真实标签图像; u_a 为图像 y 的像素均值; u_b 为图像 x 的像素均值; δ_a^2 为 y 的标准差; δ_{ab} 为 y 与 x 的协方差; c_1 和 c_2 为常量,分别取 6.502 5、58.522 5。

对于 OCR 文字识别,主要计算识别准确率 Acc,其公式为:

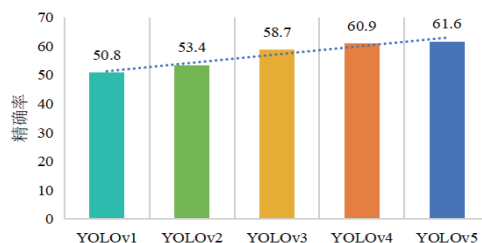
$$Acc = \frac{RR}{RR+RE+RU} \quad (3)$$

式中: RR 表示正确识别文字数量, RE 表示错误识别文字数量, RU 表示未能识别出文字数量。

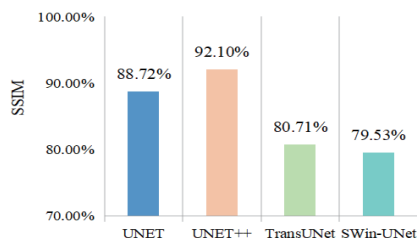
3.2 各模块性能测试

对于检测文字模块,采用同样的数据集,并在相同的硬件设备下,本文分别针对 YOLO 系列中的 Anchor Base 模型,即 YOLOv1^[5]、YOLOv2^[6]、YOLOv3^[7]、YOLOv4^[8]、YOLOv5 等模型进行了训练与测试,获得精确率的量化结果如图 4(a)所示。对比分析可知,YOLOv1 和 YOLOv2 模型对多尺度文字检测整体较弱。YOLOv3、YOLOv4、YOLOv5 等模型效

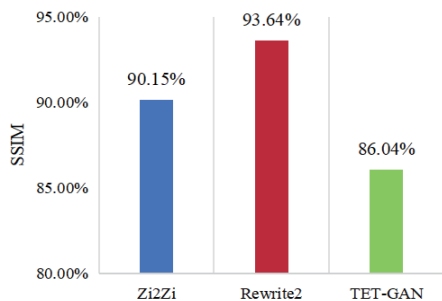
果相当,考虑碑文拓本文字可能会出现极端的很大和很小尺寸,因此系统方案最终选择 YOLOv5 进行文字的检测与定位。针对分割文字模块,选择 UNet^[9] 及其改进形成的 UNet++、TransUNet^[10]、Swin-UNet^[11] 等四个模型进行对比。实验获得的量化对比结果如图 4(b)所示。分析可知,UNet++ 模型性能最佳,该模型相比 UNet 增加的短连接对于需要分割获取细节边缘信息的文字提取效果提升明显。尽管 TransUNet 和 Swin-UNet 两个模型都使用了性能强大的 Transformer 模块,但以其为基础构建的模型往往需要大量的数据进行训练,在使用本文构建的有限数量数据集训练后,效果相对 UNet++ 的提升并不明显,因此最终系统方案选择 UNet++ 进行文字的前景提取分割。对于修复与识别文字模块,基于当前中文汉字风格迁移研究进展,本文进行了基于 GAN 模型的 3 种拓展模型的对比实验。三种模型的实验量化结果如图 4(c)所示。zi2zi^[12] 网络融入 pix2pix 而改进,将数据充足情况下的汉字风格迁移性能提升明显,但其添加的字体分类损失函数会使得在样本量较少的字体风格迁移中的生成对抗能力弱化,最终在残缺修复实验中表现出测试集效果不佳。Rewrite2 模型针对 zi2zi 模型的问题进行了优化,效果进一步提升。TET-GAN^[13] 基于最新的 Transformer 模块实现,天然需要大量数据进行训练,尽管已有各种字体文件,但这些印刷字体与本文采集的碑文拓本等非印刷字体仍然具有一定区别,因此提升效果不明显。文字识别模块输出结果严重依赖于文字修复模块产生的修复图像的输入,且百度对中文 OCR 技术已经研究多年,因此本文仅对 Rewrite2 模块输出的文字图片使用 PaddlePaddle OCR 模块进行测试。该模块整体识别准确率高达 95.3%,分析其余识别失效案例,发现主要问题在于部分碑刻文字修复效果不佳、文字本身与印刷体差别较大,其分别占到 41.8%、39.2%。



(a) 文字检测对比结果



(b) 文字提取对比结果



(c) 文字恢复对比结果

图4 算法对比量化结果

3.3 系统整体测试

本文从碑文拓本图片输入到文字识别输出进行了整体全流程测试,典型测试结果如图5所示,整体过程中四个步骤的分离性带来了潜在的失效问题,最终系统经过文字检测定位、文字前景提取、残缺文字修复后的识别文字准确率为90.6%。



图5 整体测试可视化效果

4 结论

针对碑文拓本数字化保存问题,本文实现了利用YOLOv5进行文字检测定位、UNet++进行文字前景提取、Rewrite2进行残缺文字修复、PaddlePaddle OCR进行文字识别等过程的碑文提取与修复系统,极大提升了相关工作效率。后续研究中,尤其将针对系统多个过程分离问题导致的整体性能瓶颈问题进行优化。

参考文献:

- [1] 吴迪,于正林,徐式达,等.基于改进YOLOv5的轴承表面缺陷检测[J].组合机床与自动化加工技术,2024(6):166-170.
- [2] 凌晓,王昕越,郭凯,等.基于UNet模型燃气管道高后果区分割方法研究[J].中国安全生产科学技术,2024,20(4):157-162.
- [3] BO C. Rewrite2[EB/OL].[2024-05-18]. <https://github.com/changebo/Rewrite2>.
- [4] LI Y, QIAN Y, YU Y, et al. StrucTexT: structured text understanding with multi-modal transformers[EB/OL].(2021-11-08)[2024-05-25]. <https://arxiv.org/abs/2108.02923v3>.
- [5] 梅健强,黄月草.改进YOLOv1的视频图像运动目标检测[J].天津职业技术师范大学学报,2022,32(2):29-35.
- [6] 张烈平,李智浩,唐玉良.基于迁移学习的轻量化YOLOv2口罩佩戴检测方法[J].电子测量技术,2022,45(10):112-117.
- [7] 马志艳,李辉,杨光友.基于YOLOv3算法的智能采茶机关键技术研究[J].中国农机化学报,2024,45(4):199-204+236.
- [8] 张斐,文灏淳,林浩彬,等.基于YOLOv4改进算法的智能化冰柜商品数量检测方法[J].机电工程技术,2024,53(4):121-124+151.
- [9] 朱立学,伍荣达,付根平,等.基于多尺度串联空洞卷积的轻量化UNet香蕉图像分割[J].农业工程学报,2022,38(13):194-201.
- [10] 侯作勋,鲁泓言,武伟超,等.基于深度监督的TransUNet可见光卫星图像分割[J].航天返回与遥感,2022,43(5):142-149.
- [11] 张越,王逊.基于改进Swin-Unet的遥感图像分割方法[J].无线电工程,2024,54(5):1217-1225.
- [12] EUPHORIA Y. zi2zi-pytorch[EB/OL].(2020-10-11)[2024-05-25]. <https://github.com/EuphoriaYan/zi2zi-pytorch>.
- [13] YANG S, LIU J, WANG W, et al. TET-GAN: Text Effects Transfer via Stylization and Destylization[EB/OL].(2018-12-27)[2024-05-25]. <https://arxiv.org/pdf/1812.06384>.

【作者简介】

李超(1991—),女,四川内江人,硕士,讲师,研究方向:交叉学科、优化技术。

张攀(1989—),通信作者(email:2358792637@qq.com),男,四川资阳人,硕士,讲师,研究方向:交叉学科、图像处理。

(收稿日期:2024-06-30)