文章编号：1005-1228（2024）06-0001-05

基于深度学习的马王堆汉墓简帛文字识别研究

盛　威，彭　欢，卢彦杰，刘　伟

（湖南中医药大学 信息科学与工程学院，湖南 长沙　410208）

摘　要：通过构建马王堆简帛文字识别模型，可以对同时期出土的简帛进行文字识别，提高简帛研究人员的文字识别效率，为构建古文字手写识别模型提供研究思路和技术路线。使用 BAGAN-GP 生成对抗网络，结合传统数据增强方法对原始数据集进行数据增强，使用 5 种图像分类网络开展马王堆简帛手写文字识别的对比实验。ResNet 网络在扩充后的平衡数据集上训练的模型识别准确率达 98.589%，比原始数据集训练的模型准确率提高了 9.181%。对比实验中，EfficientNet V2 small 表现最优，准确率达 99.048%。BGAN-GP 生成模型与传统数据增强方法结合的方式能够很好地适用于马王堆简帛手写文字数据集的扩充。扩充后的平衡数据集在不同的图像分类网络上都可以取得很高的识别准确率。结合迁移学习方式，导入预训练权重，模型的训练可以更快地收敛，准确率也相应提升。

关键词：深度学习；BAGAN-GP；ResNet；数据增强；迁移学习

中图分类号：TP391.1　　**文献标志码：**A

DOI:10.19414/j.cnki.1005-1228.2024.06.003

Research on Recognition of Bamboo and Silk Characters in Mawangdui Han Tomb Based on Deep Learning

SHENG Wei, PENG Huan, LU Yan-Jie, LIU Wei

（School of Informatics, Hunan University of Chinese Medicine, Changsha 410208, China）

Abstract: By constructing a model for text recognition of the Mawangdui bamboo slips, it can be used to recognize the text of the bamboo slips unearthed at the same time, improving the efficiency of text recognition for researchers of bamboo slips. In addition, it also provides research ideas and technical routes for building ancient handwritten character recognition. The original dataset was enhanced by using the BAGAN-GP generative adversarial network combined with the traditional data augmentation method, and then Five image classification networks were used to carry out a comparative experiment on Mawangdui handwriting text recognition. The recognition accuracy of the model trained on the enriched balanced dataset is 98.589%, which is 9.181% higher than that of the model trained on the original dataset. In the comparison experiment, EfficientNet V2 small performed the best, with an accuracy rate of 99.048%. The combination of BGAN-GP generative model and traditional data augmentation method can be well applied to the expansion of Mawangdui simple handwritten text dataset. The enriched balanced dataset can achieve high recognition accuracy on different image classification networks. Combined with the transfer learning method and the introduction of pre-training weights, the training of the model can converge faster, and the accuracy can also be improved.

Key words: deep learning; BAGAN-GP; ResNet; data augmentation; transfer learning

1973 年出土的马王堆汉墓帛书是我国考古史上的一次重大发现，为古文字学的研究提供了珍贵的材料。简帛上承甲骨金文，下启魏晋以后纸的发明，作为战国至魏晋时期的文字载体，具有丰富的史料价值，对于历史学、考古学、古文献学等各个学科的研究都有十分重要的意义[1]。但是，由于马王堆出土简帛的残损状况较为严重，书体为篆隶，存在大量的异体字[2]，内容又涉及先秦时期的黄老学说、医学巫术，

收稿日期：2023-10-18

基金项目：湖南省中医药科研项目（项目编号：B2023039）；长沙市自然科学基金项目（项目编号：kq2202260）；湖南省自然科学基金项目（项目编号：2022JJ30438）；湖南中医药大学研究生创新课题项目（项目编号：2022CX120）；湖南中医药大学研究生创新课题项目（项目编号：2022CX121）

作者简介：盛威（1998—），男，安徽铜陵人，硕士生，主要研究方向为深度学习；（通信作者）刘伟（1982—），男，湖南娄底人，教授、硕士生导师，主要研究方向为深度学习。

非专业的古汉语研究人员，难以学习理解书中的知识。《马王堆汉墓简帛文字全编》（节选）[3] 如图 1 所示，该书整理归纳了马王堆汉墓出土简帛的所有单字图片，为本研究提供了详细的数据，极大降低了数据集构建的难度。

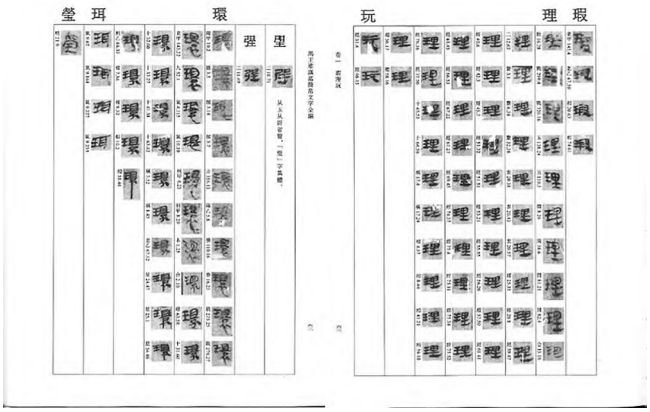

图 1　《马王堆汉墓简帛文字全编》（节选）

近年来，深度学习 [4] 技术快速兴起，在图像处理领域取得了显著进展。文字识别问题本质上是一个分类问题，十分契合图像处理领域中的图像分类问题，已有一系列学者在公开数据集上对图像分类算法进行了研究，但是在马王堆简帛领域暂未出现相关的识别模型。因此，基于深度学习技术，构建马王堆简帛文字自动识别模型，既可以提高相关研究人员阅读简帛的效率，助力古籍的理解与学习，又可以降低文字识别的错误率，推动马王堆简帛的数字化发展，助力优秀文化的传承。

1　理论基础

1.1　YOLOv5

YOLO（You Only Look Once）是一种单阶段的目标检测算法，与传统两阶段检测器相比，具有更快的速度，且其检测精度仍然相当高 [5]。YOLO 能实现端到端的训练，因此可以在大规模数据集上训练，也能自适应不同任务与场景，其架构简单，功能性强，因此得到了广泛的应用。YOLOv5[6] 模型构建方便，易于训练，本文将 YOLOv5 模型用作对原始简帛文字图片的自动检测和分割，以提升数据集构建的效率。

1.2　数据增强

数据增强是为了防止因为数据集较小而出现模型过拟合或欠拟合问题。在制作数据集时工作量巨大，大部分任务的数据量都较小。利用数据增强技术则能大幅度扩充数据集，提高模型的泛化能力。常用的数据增强技术有旋转、缩放、平移、噪声等。

传统的数据增强方法虽然使用方便、操作简单，但是生成的样本数量有限。目前生成对抗网络（Generative Adversarial Network，GAN）成功应用于许多研究领域，如手写体生成、动画人物生成等，可以实现批量的样本生成 [7]。基础的生成对抗网络训练需要样本足够且类间平衡的数据集作为支撑，而本文的原始马王堆单字数据集极不平衡，因此不能使用基础生成对抗网络来进行样本生成。BAGAN-GP[8]（BAGAN with Gradient Penalty）在 2021 年被提出，其在平衡生成对抗网络（Balancing GAN，BAGAN）[9] 的基础上，引入基于中间嵌入模型的自动编码器以及梯度惩罚，在不平衡数据集的样本生成方面取得不错的效果。

1.3　图像分类

卷积神经网络（Convolutional Neural Network，CNN）[10] 在图像分类和分割领域备受推崇，但其仍存在特征提取不完整、样本训练过拟合等问题，由此在图像分类领域衍生出了 AlexNet[11]、VGG[12]、Google Net[13]、ResNet[14]、DenseNet[15]、MobileNet[16]、ShuffleNet[17] 和 EfficientNet[18] 等一系列优秀的图像分类网络。在 ResNet 网络出现之前，研究人员通过不断增加神经网络的层数来得到更强的表达能力。但随着层数的增加，神经网络的表达能力不仅没有提高，反而出现了退化，网络损失值不断上升，准确率也不断降低，即出现梯度消失或者梯度爆炸等问题。ResNet 则利用残差块，使得信息直接跨过几层从输入传递直接到达输出，从而保留了更多信息。这种跨层连接将网络中原有的恒等映射加到当前的变换中，在一定程度上解决了梯度消失问题，并提高了网络的精度。ResNet 还可以通过堆叠残差块，在不影响网络性能的情况下构建更深的网络。本文将 ResNet、DenseNet、MobileNet、ShuffleNet 和 EfficientNet 用作文本识别模型的对比实验对象。

2　研究方法

2.1　数据集构建

本文使用 LabelImg 对《马王堆汉墓简帛文字全编》中的手写单字图片进行标注，训练了一个 YOLOv5 文字检测模型，准确检测并分割出了《马王堆汉墓简帛文字全编》中的所有单字图片。对分割下来的所有单字图片进行整理与分类，共计 3 339 个字，每个字作为一类，共 3 339 类，整理马王堆手写汉字图片共 93 841 张，马王堆简帛初始数据集示例如图 2 所示。

图 2　马王堆简帛初始数据集示例

马王堆初始数据集包括 3 339 类，然而初始数据集样本数据极其不平衡，其中最多的“之”字有 3 086 个样本，而最少的字仅有 1 个样本。数据不平衡的问题将导致模型对小样本类别文字识别的欠拟合，或对大样本类别文字识别的过拟合，从而影响整体的预测能力和性能。因此，本文在原始数据集基础上训练了一个基于 BAGAN-GP 的图像生成模型，对原始数据集进行扩充，BAGAN-GP 生成样本示例如图 3 所示。通过 GAN 的样本生成，数据集的不平衡性在一定程度上得到了缓解。每一类补充 500 个样本，样本数量上的差距仍然存在。因此，本文通过随机旋转、随机缩放、加入噪声等方式来进一步平衡数据集。最终，数据集中每一类别的图片数量约为 3 000 张，样本图片总计为 10 203 022 张。考虑最终得到的马王堆手写单字数据集样本过多，因此本文采用欠采样技术和过采样技术相结合，最终选择每个类别约 200 个样本作为模型训练的数据集，共计包含 665 129 张样本图片。

图 3　BAGAN-GP 生成样本示例

对原始数据集和平衡数据集以 8:1:1 的比例划分训练集、验证集与测试集，并确保各划分数据集没有相同样本，其中训练集用于完全训练模型，验证集用于检验训练过程中模型是否得到有效的训练效果，测试集则用于检验模型训练的最终效果。

2.2　马王堆简帛文字识别模型

马王堆简帛的手写文字识别问题本质上是一个图像分类的问题，关键点在于构建一个 3 339 类的分类模型，提升模型的泛化能力，解决其在常见字中的过拟合和非常见字中的欠拟合问题。

为了研究使用 GAN 和传统数据增强方法平衡数据集对文字识别模型准确率提升的效果，本文基于图像分类领域常见数据集 ImageNet 上各个网络模型的表现，对于多个基础图像分类网络进行研究，最终选择 ResNet 作为基础网络模型，在原始数据集和平衡数据集上进行马王堆汉墓简帛文字的识别对比实验。

此外，为了研究预训练权重对模型泛化能力提升的影响，本文对 ResNet、DenseNet、MobileNet、ShuffleNet 和 EfficientNet 这 5 个模型分别进行从头训练以及迁移学习训练，比较两种不同方式下训练的分类模型的准确率和收敛速度。

2.3　实验环境及超参数设置

马王堆古简帛文字的识别模型的训练环境均为 Ubuntu 操作系统，CPU 为 Intel(R)Xeon(R) Platinum 8255C，内存 40GB，GPU 为 NVIDIA V100，显存大小为 32G。

由于文字识别是一个多分类问题，因此选择交叉熵作为损失函数。优化器选择 Adam，其能够按照目前的真实情况给出最适合的学习率，从而加快模型收敛，并且能够很好地处理大规模数据集。ResNet 超参数设置见表 1。

表 1　ResNet 超参数设置

参数	原始数据集	平衡数据集
学习率	0.001	0.001
优化器	Adam	Adam
迭代次数	20	20
批大小	64	64
损失函数	交叉熵误差	交叉熵误差

迁移学习对比实验用于对比从头训练与迁移学习之间的效果，迁移学习实验超参数设置见表 2。

表 2　迁移学习实验超参数设置

训练方式	学习率	优化器	迭代次数	批大小	损失函数
从头训练	0.001	Adam	20	64	交叉熵误差
迁移学习	0.001	Adam	20	64	交叉熵误差

3　结果与分析

3.1　评价指标

二分类常用的评价指标有准确率（Accuracy）、精确率（Precision）、召回率（Recall）、F1-score。多分类的评价指标有微平均准确率、微平均精确率和微平均 F1-score。

设类别数为 n，TP、FN、FP、TN 表示 True Positive、False Negative、False Positive、True Negative 的数量。

二分类的准确率计算式为：

$$\text{Accuracy}=\frac{\sum_{i=1}^{n}\text{TP}_i}{\sum_{i=1}^{n}(\text{TP}_i+\text{FN}_i)} \tag{1}$$

多分类问题的微平均准确率计算式为：

$$\text{Precision}_{\text{micro}}=\frac{\sum_{i=1}^{n}\text{TP}_i}{\sum_{i=1}^{n}(\text{TP}_i+\text{FN}_i)} \tag{2}$$

微平均召回率的计算式为：

$$\text{Recall}_{\text{micro}}=\frac{\sum_{i=1}^{n}\text{TP}_i}{\sum_{i=1}^{n}(\text{TP}_i+\text{FN}_i)} \tag{3}$$

微平均 F1-score 的计算式为：

$$\text{F1}-\text{score}_{\text{micro}}=\frac{2\text{Precision}_{\text{micro}}\times\text{Recall}_{\text{micro}}}{\text{Precision}_{\text{micro}}\times\text{Recall}_{\text{micro}}}=\text{Accuracy} \tag{4}$$

马王堆简帛文字的识别任务是多分类问题，而 $\text{Precision}_{\text{micro}}$ 与 Accuracy 一样，因此本文选择 Accuracy 作为多分类的评价指标。

3.2　ResNet 对比实验

在经过 BAGAN-GP 和数据增强补充样本后的平衡数据集上训练的 ResNet 分类模型整体性能更高。生成图片实验准确率对比见表 3，基于平衡数据集训练的 ResNet 模型在测试集和验证集上的表现都要优于基于原始数据集训练的 ResNet 模型。此外，原始数据集中由于样本量小，必定会识别错误的字，如“刖”字，可以在平衡数据集上被准确识别。因此，数据集的不平衡性确实会影响分类模型的整体准确率，而本文的 BAGAN-GP 和传统数据增强方法结合优化数据集的方式可以用于解决马王堆古简帛数据集的不平衡问题，生成的样本对模型的训练可以起到积极作用。

表 3　生成图片实验准确率对比

模型	测试数据集	检验数据集
原始数据集训练的 ResNet	89.408%	72.770%
平衡数据集训练的 ResNet	98.589%	98.328%

3.3　迁移学习对比实验

各模型测试集准确率和 loss 对比如图 4、图 5 所示，图例中“_pre”表示以迁移学习方式训练的模型，“_unpre”表示以从头训练方式训练的模型，虽然迁移学习和从头训练两种方式得到的分类模型在测试集的准确率差距不大，但迁移学习方式训练的模型在训练过程中在测试集上的准确率和损失值都能更快地提升并且达到收敛，且整体的准确率都要略高于从头训练的模型。从头训练和迁移学习准确率对比见表 4，根据模型大小以及测试集准确率，本文选择 EfficientNet V2 small 作为最终的马王堆汉墓简帛文字的识别分类模型。

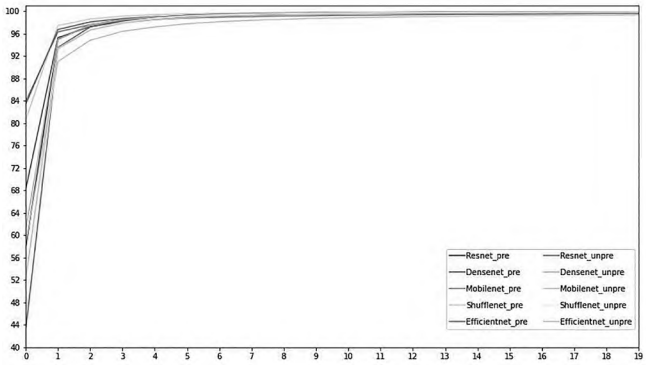

图 4　各模型测试集准确率对比

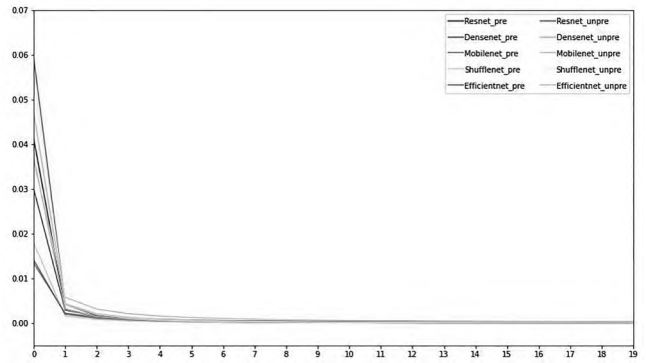

图 5　各模型测试集 loss 对比

表 4　从头训练和迁移学习准确率对比

分类模型	从头训练		迁移学习	
	测试集准确率	模型大小	测试集准确率	模型大小
ResNet152	98.589%	248.8 M	98.673%	248.8 M
DenseNet201	98.743%	94.8 M	98.598%	94.8 M
MobileNet V3 small	98.005%	19.0 M	98.405%	19.0 M
ShuffleNet V2 ×2.0	98.677%	46.7 M	98.871%	46.7 M
EfficientNet V2 small	**98.874%**	**94.2 M**	**99.048%**	**94.1 M**

4　结束语

本文选取马王堆简帛文字为研究对象，利用深度学习技术，构建马王堆简帛文字识别网络，将BGAN-GP生成对抗网络与传统数据增强方法结合，解决古籍文字识别时经常出现的数据不平衡问题，并通过RestNet构建分类模型验证该方式的可行性，再结合迁移学习导入预训练权重，提升模型的泛化能力和收敛速度。

实验结果表明，BGAN-GP生成模型与传统数据增强方法结合的方式能够很好地适用于马王堆简帛手写文字数据集的扩充。扩充后的平衡数据集在不同的图像分类网络上都可以取得很高的识别准确率。此外，结合迁移学习方式，导入预训练权重，模型的训练可以更快地收敛，准确率也可以得到一定提升。

虽然目前出土古籍的数字化工作已经取得了很大的进展，但是由于不同朝代文字字形变化过大、古籍破损严重，部分文字样本过少等一系列问题，设计一款涵盖不同朝代、不同字体的古籍手写文字识别模型有一定难度。目前图像分类领域的深度学习技术已日趋成熟，但是将其应用于古籍文字的自动识别上，仍碍于数据集构建的难度。得益于《马王堆汉墓简帛文字全编》对马王堆简帛中出现文字的归纳整理，本文数据集的构建难度大大降低。因此，古籍文字识别领域亟须一个整理归纳了各个朝代、各种字体的数据库，以降低古籍文字识别领域数据集构建的难度，从而推动深度学习技术在古籍文字识别领域的应用，推动中华优秀古籍的数字化进程。

参考文献：

[1] 吴云燕 . 马王堆汉墓帛书通用字研究 [D]. 上海 : 华东师范大学 ,2006.

[2] 黄文杰 . 马王堆简帛异构字初探 [J]. 中山大学学报 (社会科学版),2009,49(4):66-79.

[3] 郑健飞 , 李霜洁 , 程少轩 , 等 . 马王堆汉墓简帛文字全编 [M]. 北京 : 中华书局 ,2020.

[4] 赵明敏 , 杨秋辉 , 洪玫 , 等 . 基于深度学习和信息反馈的智能合约模糊测试方法 [J]. 计算机科学 ,2023,50(9):117-122.

[5] 邓亚平 , 李迎江 .YOLO 算法及其在自动驾驶场景中目标检测综述 [J]. 计算机应用 ,2024,44(6):1949-1958.

[6] OMAR J, JOSH H, STUART S, et al. ResNet and YOLOv5-enabled non-invasive meat identification for high-accuracy box label verification[J]. Engineering Applications of Artificial Intelligence,2023:125.

[7] 张卓 , 雷晏 , 毛晓光 , 等 . 基于对抗生成网络的缺陷定位模型域数据增强方法 [J]. 软件学报 ,2024,35(5):2289-2306.

[8] HUANG G, JAFARI H A. Enhanced balancing GAN: minority-class image generation[J]. Neural Computing and Applications,2021,35(7):1-10.

[9] MARIANI G, SCHEIDEGGER F, ISTRATE R, et al. Bagan: Data augmentation with balancing gan[J]. arXiv preprint arXiv:1803.09655, 2018.

[10] 季长清 , 高志勇 , 秦静 , 等 . 基于卷积神经网络的图像分类算法综述 [J]. 计算机应用 ,2022,42(4):1044-1049.

[11] KRIZHEVSKY A, SUTSKEVER I, HINTON E G. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM,2017,60(6):84-90.

[12] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. CoRR, 2014,abs/1409.1556

[13] SZEGEDY C, 0015 L W, JIA Y, et al. Going deeper with convolutions[J]. CoRR, 2014,abs/1409.4842.

[14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[J]. CoRR,2015,abs/1512.03385.

[15] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.

[16] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.

[17] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6848-6856.

[18] TAN M, LE Q. EfficientNet: Rethinking model scaling for convolutional neural networks[C]//International conference on machine learning. PMLR, 2019: 6105-6114.