武漢工程大学

Wuhan Institute Of Technology

硕士学位论文

真实场景下的楚系简帛古汉字识别研究

类别/领域： 计算机视觉

研 究 生： 李胜楠

指导教师： 王凯丽 讲师

培养单位： 计算机科学与工程学院

二〇二四年六月

中国知网 https://www.cnki.net

分类号：TP391 **学校代号：10490**
学　号：22107010013 **密　　级：公开**

武汉工程大学硕士学位论文

真实场景下的楚系简帛古汉字识别研究

作者姓名：李胜楠　　指导教师姓名、职称：王凯丽讲师

申请学位类别：学硕　　专业学位类别名称：计算机应用技术

研究方向：计算机视觉

论文提交日期：2024 年 4 月 1 日　论文答辩日期：2024 年 5 月 25 日

学位授予单位：武汉工程大学　　学位授予日期：2024 年 6 月 15 日

答辩委员会主席：邹华

中国知网　https://www.cnki.net

Research on the Intelligent Recognition of Chu Bamboo and Silk ancient Chinese Characters in the Real Scene

A Thesis Submitted for the Degree of Master

Major： **Computer Vision**

Candidate： **Li ShengNan**

Supervisor： **Lect. Wang Kaili**

Wuhan Institute of Technology

Wuhan, Hubei 430073, P. R. China

中国知网 https://www.cnki.net

摘 要

楚系简帛古汉字是战国时期楚国人书写于竹简木牍或丝织品上使用的文字，距今已经有2000余年历史，是介于象形-甲骨文与秦统一符号化文字-篆书之间的过渡性地区通用文字，识别该文字对于研究湖北地区历史风貌、文明发展具有重要历史意义。楚系简帛古汉字与其他书体文字差异大，且手写风格、变体众多，具有较大的类内差异与类间相似性。另外，长时间保存不善与出土过程中的自然腐蚀、风化导致简帛古汉字残缺、被污染，严重影响了楚系简帛古汉字的可读性。目前甲骨文的智能化识别研究已经获得了较大的关注，并取得了初步的进展，但对于楚系简帛古汉字的识别研究还较为少见，仍依赖于专家的人工考释，识别过程耗时耗力。计算机视觉技术的快速发展为楚系简帛古汉字的智能化识别提供了可能。本文在数据分析基础上，结合楚系简帛古汉字的文字结构特性与数据集分布特点，从以下几方面展开研究：

（1）因目前没有公开的可用于智能识别的真实场景楚系简帛古汉字数据集，本文构建了一个包含38894张楚系简帛古汉字图像的数据集-Chu730，其涵盖730类字符，占据目前已出土楚系简帛古汉字通用字形的29.2%，并对数据分布特征等进行分析，用于后续的识别。

（2）针对真实场景下的楚系简帛古汉字图像存在严重噪声干扰、文字残缺退化等问题，本文提出一种小波注意力网络。利用小波变换的多尺度特征提取与多分辨率特性，将卷积神经网络的特征学习空间扩展到小波域，结合小波低频和高频子带构建注意力模型，更好提取“网格化”楚系简帛古汉字中具有辨别力的关键结构特征，其在识别退化和模糊文本图像时表现更优，Top-1识别准确率提升1.82%。

（3）通过数据集分析，本文发现真实场景下的楚系简帛古汉字数据集具有长尾分布的特征，且含标签楚系简帛古汉字图像数据量有限。为了解决因文字使用频率不一导致的极端数据不平衡问题，本文提出了一种混合损失函数。该损失函数根据困难样本和尾部小样本特点，通过在网络训练过程中自适应调整权重，来提升模型对于困难样本与样本不足文字的特征提取与分类能力。实验结果表明，该方法相比于基础方法，Top-1到Top-5的提升增量分别为1.66%，1.45%，1.31%，1.17%，

中国知网 https://www.cnki.net

1.02%，说明了本文方法的有效性。

关键词：楚系简帛古汉字；文本识别；离散小波变换；小波注意力；长尾分布

中国知网 https://www.cnki.net

Abstract

Chu Bamboo and Silk ancient Chinese characters (CBSC) is a script used by the people of the Chu state during the Warring States period, written on bamboo slips, wooden tablets, or silk fabrics. With a history of over 2000 years, it serves as a transitional script between pictograms or oracle bone inscriptions and the standardized script of the Qin dynasty, known as seal script. Recognizing this script is of significant historical importance for studying the historical landscape and civilization development of the Hubei region. CBSC differs significantly from other script styles, with numerous handwriting styles and variations, exhibiting substantial intra-class differences and inter-class similarities. Moreover, poor preservation over time and natural deterioration and weathering during excavation have led to damage and contamination of ancient Chinese characters on bamboo slips and silk fabrics, severely affecting the readability of CBSC. While there has been considerable attention and initial progress in the intelligent recognition of oracle bone inscriptions, research on the recognition of CBSC remains limited, relying heavily on expert manual interpretation, which is time-consuming and labor-intensive. The rapid development of computer vision technology offers the possibility for the intelligent recognition of CBSC. Based on data analysis, this thesis investigates the intelligent recognition of CBSC from the following aspects, considering the structural characteristics of the script and the distribution patterns of datasets:

(1) Due to the absence of publicly available datasets suitable for intelligent recognition of CBSC in real-world scenarios, this thesis constructs a dataset named Chu730, comprising 38,894 images of CBSC characters. The dataset covers 730 character classes, representing 29.2% of the currently excavated commonly used character forms in CBSC. Analysis of data distribution characteristics and other relevant factors is conducted to facilitate subsequent recognition tasks.

(2) In response to the significant noise interference and text degradation present in real-world CBSC images, this thesis proposes a Wavelet Attention Network (WA-Net). Leveraging the multiscale feature extraction and multiresolution properties of wavelet transform, the network extends the feature learning space of convolutional neural networks

中国知网 https://www.cnki.net

to the wavelet domain. By combining wavelet low-frequency and high-frequency subbands to construct an attention model, it effectively extracts discriminative structural features from the "grid-like" CBSC. This approach demonstrates superior performance in recognizing degraded and blurry text images, resulting in a 1.82% improvement in Top-1 recognition accuracy.

(3) Through dataset analysis, it was determined that Chu730 exhibits a long-tail distribution, and there is a limited number of well-preserved images of CBSC. To address the issue of extreme data imbalance arising from variations in character usage frequency, this thesis proposes a hybrid loss function. This loss function dynamically adjusts the weight values of challenging classification samples and tail-end small samples during network training to enhance the classification accuracy of the dataset. Experimental results demonstrate that compared to the baseline method ResNet34, the proposed method achieves incremental improvements from Top-1 to Top-5 of 1.66%, 1.45%, 1.31%, 1.17%, and 1.02%, respectively, thereby confirming the effectiveness of the method presented in this thesis.

Keywords: Chu Bamboo and Silk ancient Chinese characters (CBSC); Text Recognition; Discrete wavelet transform; Wavelet attention; Long-tailed distribution; Hard-sample construction

中国知网 https://www.cnki.net

目 录

第 1 章 绪论.....1
1.1 研究背景及意义.....1
1.1.1 研究背景.....1
1.1.2 研究意义.....3
1.2 研究现状.....4
1.3 研究难点与贡献.....6
1.3.1 本文研究难点.....6
1.3.2 本文贡献.....7
1.4 本文组织结构.....8
第 2 章 古汉字识别相关理论.....11
2.1 文本识别方法.....11
2.1.1 文本识别.....11
2.1.2 汉字识别方法.....15
2.2 卷积神经网络.....17
2.3 小波变换.....20
2.4 注意力机制.....22
2.4.1 通道注意力.....24
2.4.2 空间注意力.....25
2.4.3 自注意力.....26
2.5 长尾分布.....26
2.6 本章小结.....27
第 3 章 楚系简帛古汉字数据集.....29
3.1 Chu730 数据集简介.....29
3.2 数据集统计与分析.....30
3.3 实验配置.....32
3.3.1 实验环境及参数设置.....32
3.3.2 评价指标.....33
3.4 基准实验结果与分析.....34
3.4.1 基准实验.....34
3.4.2 实验结果与分析.....37
3.5 本章小结.....37
第 4 章 基于离散小波变换的楚系简帛古汉字识别方法.....39
4.1 研究背景.....39
4.2 相关工作.....40

中国知网 https://www.cnki.net

4.2.1 注意力机制.....40
4.2.2 小波变换.....41
4.3 基于离散小波变换的楚系简帛古汉字识别方法.....42
4.3.1 网络结构.....44
4.3.2 离散小波变换.....44
4.3.3 W-bneck 模块.....46
4.3.4 小波注意力.....48
4.4 实验结果与分析.....51
4.4.1 对比实验.....51
4.4.2 消融实验.....55
4.5 本章小结.....56
第 5 章 基于长尾分布的楚系简帛古汉字的特征平衡损失.....57
5.1 研究背景.....57
5.2 相关工作.....58
5.2.1 长尾分布.....58
5.2.2 损失函数.....58
5.3 基于长尾分布的楚系简帛古汉字识别方法.....59
5.3.1 交叉熵损失函数.....59
5.3.2 困难样本重加权损失.....60
5.3.3 权重重平衡损失.....60
5.3.4 图像标签损失.....61
5.4 实验结果.....61
5.4.1 对比实验.....61
5.4.2 消融实验.....63
5.5 本章小结.....64
第 6 章 总结与展望.....67
6.1 研究总结.....67
6.2 研究展望.....68
参考文献.....71

中国知网 https://www.cnki.net

第 1 章 绪论

1.1 研究背景及意义

本章主要介绍了真实场景下的楚系简帛古汉字智能识别研究任务，介绍内容包括研究背景、研究意义、楚系简帛古汉字特点和识别任务的研究难点以及相关方向的研究现状。简述了本文的研究内容和文章结构。

1.1.1 研究背景

文字识别[1-4]（Optical Character Recognition，OCR）是一种将印刷文本或手写文本转换为数字文本的技术，因其在数字化图书馆、电子商务、自然语言处理等领域的广泛应用而受到广泛关注。文字识别的发展大致可以分为传统的文本识别方法和基于深度学习的文本识别方法[5]。传统的文本识别方法[6]依赖于传统图像处理技术[7]和统计机器学习方法[8, 9]实现，可以分为图像预处理、文字识别和后处理三个阶段。基于深度学习的文本识别方法[10-12]利用卷积神经网络（Convolutional Neural Networks，CNN）或循环神经网络（Recurrent Neural Network，RNN）等网络结构的特征表达能力来实现，目前主流的方案有两种，分别是文本行检测与文字识别的两阶段方案[13, 14]和端到端的文字识别方案[3, 15]。



图 1-1 常见的文本识别研究对象

中国知网 https://www.cnki.net

最初的文字识别对象是由英文字母组合构成的英文单词，目前对英文的识别技术已商业化。在对数字、英文、符号识别研究的基础上，部分研究者开始对印刷体汉字的识别进行研究。随着科技的发展，对宋体、黑体、楷体等印刷体汉字的识别率高达 99.5%以上，印刷体的汉字识别已应用于表格的自动识别，名片识别，金融票据识别等方面。但由于真实场景下的文字存在模糊、笔划粘连、断笔等原因，使得真实场景下的汉字识别准确率较低。目前，手写文本识别和自然场景文本识别是近年来研究的热点，自然场景下的汉字因其分布广泛，场景多样，种类繁多而具有较高的识别难度。手写体汉字因书写方式随意、汉字字符复杂、连笔墨点等不同于印刷体的原因，而具有挑战性。目前常见的文本识别研究对象如图 1-1 所示。古汉字是自然场景下的手写体汉字，包含了从殷商时期至今三千多年的汉字字形变体，记录了不同时期的历史进程。古汉字的智能识别是文化遗产数字化保护研究领域的一项极具挑战性的技术[16-18]，对于追溯中国古代历史、经济、文化和科学技术的发展具有重要意义。

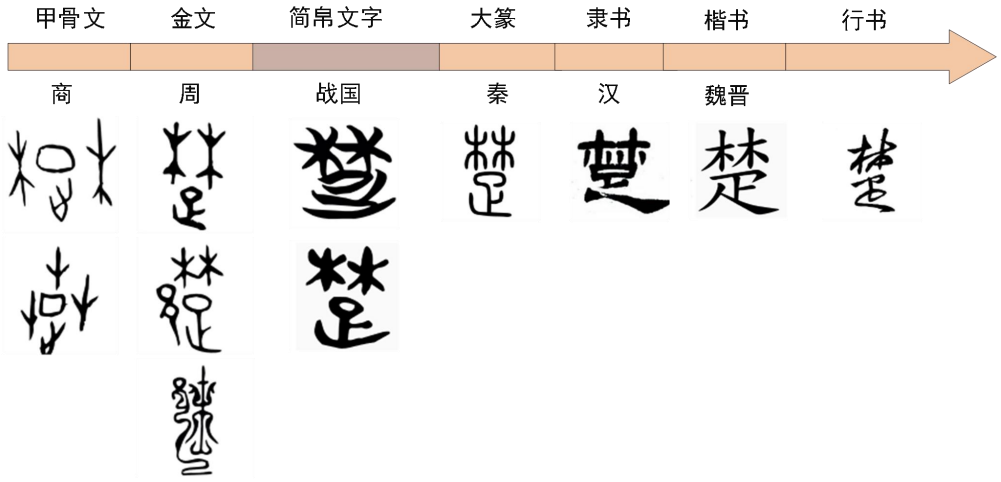

图 1-2 楚系简帛古汉字历史定位

楚系简帛古汉字是当今湖北地区在春秋战国时期使用的通用性文字，反映了当时该地域的文化、历史现象。其书写形式与一般的古汉字差别较大，具有手写文本的多变性和自然场景文字的背景复杂性[19, 20]，楚系简帛古汉字在历史上的定位如图 1-2 所示。楚系简帛古汉字继承了商周

中国知网 https://www.cnki.net

的文字结构，也是文字从图像化到符号化转变的过程，受书写载体和工具的影响，楚系简帛古汉字字形独特，横画形抹，竖画窄长，线条较粗、笔画轻重变化明显，字形衔接紧密，交叉处互相贯通，连接处无明显分界线，呈现出连续、流畅的书写风格。竹简和帛书作为楚系简帛古汉字的重要研究载体，由于保存环境较差，载体材质较为脆弱，经过数千年的时间和自然环境的侵蚀，大量包含楚系简帛古汉字的文物表面出现损伤、变形、缺失等现象，导致书写内容的完整性、清晰度和辨识度都大大降低，部分文字图像模糊甚至出现退化残缺。拓本图像噪声影响严重，背景干扰、文字残缺等情况普遍存在，文字识别难度相对较高，楚系简帛古汉字在载体上的分布如图 1-3 所示。目前专门针对楚系简帛古汉字的研究较少，主要识别难点包括文字自身结构较为复杂，并且存在较大的类内差异。其次，这些文字是人工刻在竹简上的，因此同一文字之间可能存在形体差异。同时，类间相似度较高，文字数量较多，且书写风格多样，部分文字甚至存在退化残缺。另外，由于数据量较少且单个样本数据量不均衡，都给研究带来了巨大的挑战。

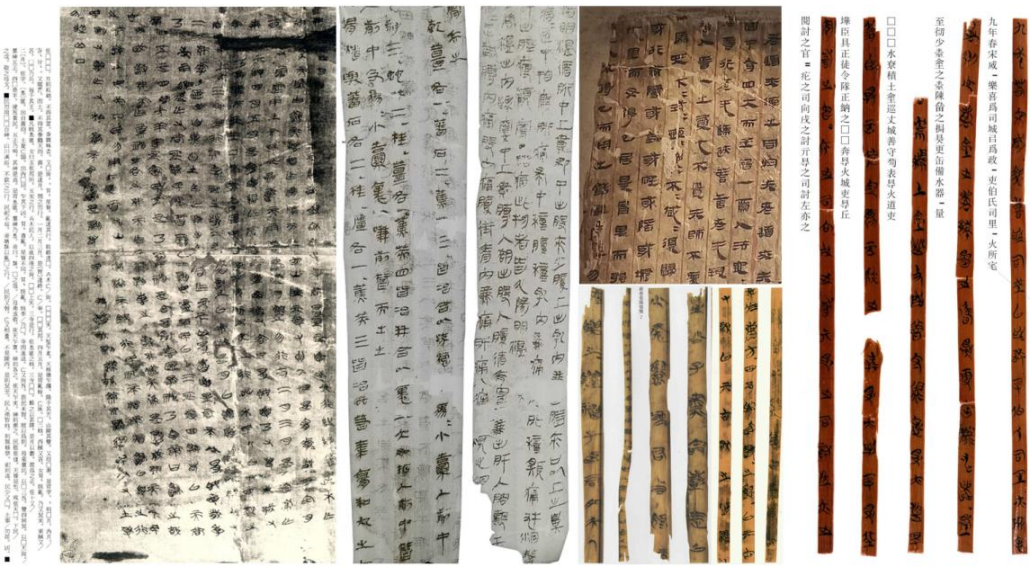

图 1-3 竹简和帛书上的楚系简帛古汉字

1.1.2 研究意义

楚系简帛古汉字的智能识别研究具有重要的实用价值和研究价值，

中国知网 https://www.cnki.net

实用价值为：

（1）由于楚系简帛古汉字存在历史悠久，对于没有接受过专业训练的人来说，阅读和理解这些古汉字十分困难。研发相关技术对于在人们日常生活中提供古汉字阅读的辅助具有重要意义。

（2）本文研究内容有助于建立楚系简帛古汉字的数字化图书馆，实现史诗资料的数字化存储和管理。这使得从人工识别转变为智能识别，可以使楚系简帛古汉字的处理流程更加便捷化。

（3）目前，国内建立许多简帛汉字博物馆，本文的研究内容旨在助力出土简帛文献的识别工作，并为相关文字文物的研究提供有力的辅助。有助于传播中华古代文化之美。

研究价值为：

（1）针对楚系简帛古汉字智能识别的专门研究较为罕见，为了有效的开展实验研究，首要任务是完成数据集的构建，并设计基准方法以及改进方法的研究和创新，这些工作不仅对于楚系简帛古汉字识别任务具有重要意义，同时也对于具有相似特点的其他文字识别任务具有借鉴价值。

（2）由于楚系简帛古汉字为出土文字，其在真实场景中的文字图像存在数千年，导致图像本身出现了退化，残缺等现象。因此，针对这些古文字的识别任务，需要一种有效的识别方式进行特征提取，并确保在面对超多分类的任务时具有较高的识别精度。在这方面，智能识别方法对于处理多分类识别任务具有重要的借鉴意义。

基于以上分析，楚系简帛古汉字的智能识别研究具有重要的研究意义和实用价值。

1.2 研究现状

专门针对楚系简帛古汉字智能识别的研究相对较少，但与此相关的有多个领域的研究直接联系，包括古汉字识别，书法文本识别，手写文本识别和自然场景文本的识别[5, 21, 22]等。传统的文本识别方法需要手动设计并提取文本的特征，如斑点、边缘等，然后通过分类算法进行识别，

中国知网 https://www.cnki.net

识别的结果受到图片的质量、图片内文本的字体等影响较大。基于深度学习的模型能够直接从原始的输入数据中学习到更高层次的抽象特征，具有强大的表征能力和泛化能力，不再依赖于手工设计的特征。

传统的文本识别方法包括马尔科夫模型、贝叶斯推断方法和支持向量机方法。Hu 等人[23]基于隐马尔科夫模型（Hidden Markov Model，HMM）法，结合笔画切线与弯曲度的不变性，将预处理和不变性特征提取相结合，利用点导向和笔画导向，有效地提高了识别效果。统计模式识别方法在匹配算法上更易于实现，贝叶斯网络分类器是经典的基于统计方法的分类模型，任义[24]提出随机贝叶斯森林（random Bayesian forest，简写为 RBF）算法能够捕捉和利用大数据中所包含的额外细节，有效地解决大规模数据集上的分类问题。支持向量机（Support Vector Machines，SVM）方法是一种最小代价判别式分类器，樊等人[25]在 SVM 方法的基础上提出了主动学习支持向量机（Active learning SVM，ASVM），该方法解决了 SVM 分类器受样本孤立点影响较大的问题。

随着深度学习技术的兴起和发展，人们开始探索基于深度学习的文本识别方法。相对于支持向量机只能通过手工设计特征来提取特征的浅层学习方法，深度学习是一种层级特征提取算法，通过逐层特征映射将原始数据转换到更高维度的特征空间，从而实现自动学习层次化的特征表示。LeCun 等人[26]提出了一种卷积神经网络模型框架 LeNet-5，该网络在手写数字识别方向取得了非常显著的效果。Shi 等人[27]提出了一种具有自动纠正功能的鲁棒文本识别器 RARE。Bahdanau 等人[28]首次提出使用注意力机制来处理机器翻译问题，在文本识别中使用注意力机制对齐识别结果。Liao 等人[29]通过从二维的角度分析文本图像特征，提出 CA-FC 替换常规神经网络为全卷积网络（Fully Convolutional Networks，FCN），该方法解决了输入图像尺寸的限制问题，结合注意力机制做像素级分类，实现任意形状文本的分类。

汉字的复杂结构、总类达到三千多种、不同汉字之间的相似度较高以及字体风格的多样化等问题，给汉字识别研究带来了巨大挑战。前期的古汉字识别研究使用的多是人工生成特征，Lin 等人[30]使用 GIST 描述

中国知网 https://www.cnki.net

子来处理书法文本图像，并采用局部感知哈希 LSH 来寻找相似的文本图像，以实现书法文本的识别。Xu 等人[31]介绍了一个用于汉字识别研究的中国古代手写体数据库，以解决手写形态多样且异体字较多等问题。Tian 等人[32]提出了古文图书排版图像分析和汉字图像分割算法，用于提高古书排版图像分割的准确率。Zhang 等人[33]提出了基于远程监督和深度学习的高级神经网络中文古籍人名自动识别方法。对古汉字的计算机智能化识别需要深入研究其结构和字形。Jia 等人[34]尝试通过探索字符和笔画的特征来识别实体。Chen 等人[10]研究了基于图像分析的汉字识别算法，提出了一种书法字符识别算法，效率更高。Miao 等人[11]提出了一种基于汉字书写标准模型的古文字图像去噪方法，以应对古汉字结构的复杂性。

1.3 研究难点与贡献

1.3.1 本文研究难点

目前的汉字文本识别研究主要包括书法文本、场景文字以及标准简体字表单表格等方面。历史上的古汉字多以石刻、铜铸、手写等形式产生，这些多变的形态与异体字等现象提升了古汉字的识别难度。在古汉字的识别领域，主要的成果集中在甲骨文智能识别方面，而对于楚系简帛古汉字的智能识别研究相对较少，其识别过程面临着一些独特的问题：

（1）数据集：目前暂无公开的用于计算机识别的楚系简帛古汉字数据集。楚系简帛古汉字因其使用区域集中在当今湖北一带而保存较少，是出土量有限的古汉字之一。由于单一文字数据量较小，且文字的使用频率不同，导致文字样本的类别间存在较大差异。

（2）文字：楚系简帛古汉字是甲骨文到大篆的过渡文字，与商周文字同属一系，字体颀长秀美。由于古代社会交通不发达，相同文字可能呈现不同字体，同一文字之间存在形体差异，更易出现类间相似性和类内差异性等现象。

（3）载体：楚系简帛古汉字源自战国时期的大墓出土文献。经过长时间的保存，竹简和帛书的材质差异以及保存过程中的损伤导致部分

中国知网 https://www.cnki.net

文字出现了残缺和退化现象。因此，楚系简帛古汉字的文字图像普遍存在画质较差的情况。

（4）识别研究：目前对古汉字的智能识别研究集中在甲骨文字符，对于文字演化过程的其他文字研究较少。对楚系简帛古汉字的研究主要停留在对文字的字形和字意考究，对其计算机识别的研究较少。

1.3.2 本文贡献

楚系简帛古汉字文本识别任务面临诸多挑战，其中包括文字自身结构的复杂性、类内差异性和类间相似性等问题，以及文字载体保存过程中可能出现的退化和残缺。此外，由于文字样本数据量的不平衡，研究面临更大的困难。然而深度学习方法具有强大的特征提取能力，为本文任务的实现提供了可能性。为了应对上述挑战，本文的研究内容和贡献如下所述：

（1）为了实现基于深度学习的楚系简帛古汉字研究，本文创建了一个适用于深度学习的真实场景下的楚系简帛古汉字数据集，该数据集涵盖了楚系简帛古汉字大部分常用的文字字形，并且符合真实场景数据集的分布特点。

（2）为了验证深度学习方法在楚系简帛古汉字识别任务中的适用性和有效性，在前文所建立的数据集上进行了经典深度学习网络的初步实验。考虑到小波变换具有强大的结构特征提取和多尺度特征提取能力，以及楚系简帛古汉字具有较强的结构性，提出了一种增加小波卷积以保留全局信息的方法。由于原始残差连接下采样操作可能导致特征信息的丢失，采用了平均池化来保留更多的特征图像信息，并通过设计小波注意力机制来重点关注图像的特征信息。

（3）考虑到真实场景中的楚系简帛古汉字数据集呈现出长尾分布的特性，本文提出了一种基于图像和标签的感知损失函数，该损失函数同时考虑了图像层面和标签层面的情况。首先，在图像层面，由于不同图像之间存在差异，其分类难度也各不相同，因此提出了一种困难样本重加权损失函数，以提高对困难样本的分类准确性。其次，在标签层面，

中国知网 https://www.cnki.net

鉴于数据集存在严重的长尾分布问题，我们采用了重加权的方法，对尾部标签赋予更高的权重，以缓解这一问题。实验结果表明，我们提出的方法在收集的数据集上表现出明显的有效性，与传统方法相比具有显著的改进。

1.4 本文组织结构

考虑到目前尚无公开的楚系简帛古汉字数据集可供使用，本研究创建了真实场景下的楚系简帛古汉字数据集。考虑到楚系简帛古汉字的结构性特征，我们特别关注小波变换对于图像结构特征的提取能力，并将其引入卷积神经网络，以进一步优化文本图像特征的提取过程。对数据集的分析表明，真实场景中的文本数据呈现出文字使用频率不均的特点，导致文字图像样本数量的长尾分布，因此我们提出了一种基于图片标签的感知损失函数，以提升判断准确性。本文的整体结构共分为六章，各章节的安排如下所述：

第一章为绪论部分，本章首先对楚系简帛古汉字的识别背景进行了详细介绍。其次，对与楚系简帛古汉字相关的研究方向进行了深入分析，系统总结了当前研究现状，以便为本文的研究框架提供有力的理论支撑。进一步剖析了本研究任务所面临的诸多难点，并强调了本研究将带来的重要贡献，研究的科学意义和实践价值。最后介绍了本文的研究内容。为本文的研究方向和主要思路提供了指引。

第二章为相关工作介绍，首先介绍了与楚系简帛古汉字识别研究相关的领域的研究现状和详细技术。其次，详细介绍了卷积神经网络的定义以及其各个层次结构的构成，以便读者能够对此关键技术有清晰的认识。同时，还对后续研究所用到的小波变换和长尾分布等相关工作进行了介绍和分析，以为读者理解后续研究内容奠定基础。

第三章详细介绍了本研究所收集整理的楚系简帛古汉字数据集以及基准实验的相关内容，着重阐述了楚系简帛古汉字数据集的背景。其次，探讨了实验配置和评价指标，并对基准实验结果进行了分析。

第四章基于前文建立的数据集，本研究进行了结合传统方法和深度

中国知网 https://www.cnki.net

学习方法的楚系简帛古汉字识别方法的研究。通过引入 W-bneck 模块、小波通道注意力和小波卷积注意力的设计，提出了一种改进型网络结构—WA-Net，该网络结构基于离散小波变换。随后，对实验结果进行了详尽的分析和评估。

第五章基于前文建立的数据集，针对真实场景下的数据集呈现出的长尾分布特征，本研究提出了一种基于图像和类别的感知损失函数，以进一步推动对楚系简帛古汉字在真实场景下的识别研究。

第六章为总结与展望。总结本文的主要内容，并对未来研究方向进行展望。

中国知网 https://www.cnki.net

中国知网 https://www.cnki.net

第 2 章 古汉字识别相关理论

本章主要介绍文本识别的相关理论知识，包括传统文本识别和深度学习的文本识别所使用的方法和技术，并介绍了卷积神经网络相关知识以及本文后续用到的小波变换、注意力机制和长尾分布等相关基础。

2.1 文本识别方法

文本识别方法经历了从传统的光学字符识别到现代基于深度学习的方法的转变。自上世纪 90 年代以来，卷积神经网络在文本识别领域取得了显著的进展，2014 年是传统方法和深度学习方法的分水岭。在此之前，传统方法主导了文本识别领域，传统的文本识别方法通常包括多个步骤，如预处理、特征提取、特征降维、分类器设计和后处理等。传统方法包括基于图模型、条件随机场、贝叶斯推断以及马尔可夫蒙特卡洛模型等方法。然而，这些方法通常依赖于手工设计的特征和复杂的数学模型，限制了其在复杂场景下的性能和泛化能力。相比之下，深度学习方法具有较强的表征能力和泛化能力，能够直接从原始数据中学习到更高层次的抽象特征，并且在大规模数据集上表现出色，使得文本识别方法发生了根本性的变化，深度学习方法通过端到端的训练方式，使得模型能够直接从原始数据中学习到最终的文本识别结果。因此，深度学习方法已经成为了当前文本识别领域的主流方法，取代了传统方法中复杂的手工设计特征的过程，并且在不断地推动着该领域的发展和进步。

2.1.1 文本识别

（1）传统文本识别

OCR 技术在 20 世纪初首次问世，主要用于识别印刷体字符。传统的 OCR 技术通过扫描文本资料，对图像文件进行分析处理，以提取文字和版面信息，该方法需要检测和定位单个字符的区域，然后对定位后的字符区域进行识别，并通过语言模型进行修正。传统 OCR 技术已在文档数字化、票据处理、笔迹录入、智能交通、文档检索与信息抽取等

中国知网 https://www.cnki.net

领域得到广泛应用，为手写体识别的发展奠定了基础。

OCR 的主要流程包括图像预处理、版面处理、图像切分、特征提取、匹配及模型训练、识别后处理，如图 2-1 所示。在图像预处理阶段，通常会采用去噪、灰度化、二值化等操作以提高字符的辨识度。版面处理阶段主要是对文档的版面结构进行分析和识别，包括文本行和段落的划分，以及文字与图片等元素的区分。图像切分阶段将文本区域切分成单个字符或单词的区域，为后续的识别提供准确的输入。特征提取阶段则是从每个字符的图像中提取出代表其形态和结构的特征，常用的特征包括形状、边缘、投影等。在匹配及模型训练阶段，会利用这些特征进行模式匹配或者训练 OCR 模型，以实现对字符的准确识别。在识别后处理阶段，对识别结果进行校正和修正，以提高整体识别的准确性和稳定性。

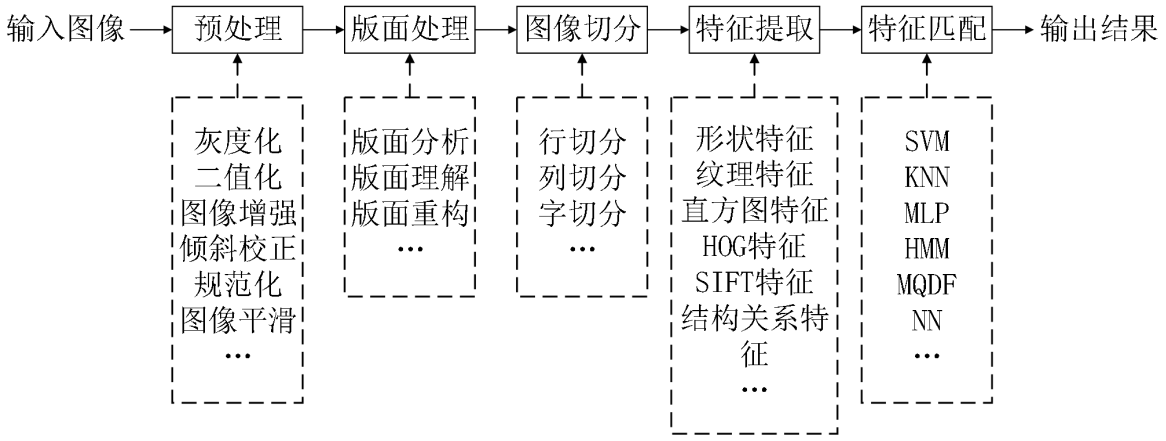

图 2-1 传统文本识别流程

图像预处理：图像预处理是在文本通过扫描仪输入计算机时进行的一项重要工作。其主要目的是应对由于纸张厚度、光洁度和印刷质量的变化而引起的文字图像的畸变问题，包括断笔、粘连和污点等。这些问题可能会导致文字图像的不清晰和变形，从而影响后续的文字识别效果。通过图像预处理，可以有效地降低噪声干扰，改善文字图像的质量，使得后续的文字识别算法能够更加准确地识别文字内容，从而提高整个文字识别系统的性能。这种处理工作在文字识别之前进行，因此被称为预处理。

中国知网 https://www.cnki.net

特征提取：特征提取直接影响着识别性能的稳定性和有效性。针对单个字符图像，通常会采用统计特征和结构特征进行描述。统计特征的提取方法通常可以参考统计模式识别的技术。这些特征可以捕获字符图像的全局统计信息，如像素分布、灰度直方图等，从而帮助识别系统理解字符的整体形态和特征。另一方面，结构特征的选择取决于具体的识别基元。对于不同的字符形态，可能需要使用不同的结构特征来描述其形状和结构。这些结构特征可以包括边缘特征、变换特征、穿透特征、网格特征、特征点特征等。

特征匹配：特征匹配是从已有的特征库中找出与待识别文字最相似文字的过程。特征库中存储了字符集所有文字的统计特征和结构特征，这些特征需要与待识别文字的特征进行对比，以确定最相似的字符。统计特征可以捕获字符的全局信息，而结构特征则可以描述字符的形态和结构特征。特征匹配的方法多种多样，包括欧式空间比对法、松弛比对法、动态程序比对法以及 HMM 法等。在神经网络兴起之前和之后的很长一段时间里，汉字 OCR 领域一直主要采用模板匹配方法。虽然模板匹配方法在一定程度上能够满足 OCR 的需求，但其受限于特征的表达能力和匹配的准确性，随着深度学习等新技术的发展，特征匹配方法也在不断演进和改进。

在传统文字识别方法中，研究人员主要关注从图像中提取有效的特征来表示文字，包括形状、纹理、边缘和投影等。传统的特征提取方法如 SIFT、HOG 和 SURF 等被广泛应用于文字识别中，并使用适当的特征分类器如支持向量机和随机森林等进行特征匹配。然而，传统文字识别方法需要大量标注样本来训练和评估模型，这需要人工参与，是一个耗时且费力的过程。另外，传统方法对于文字图像的变化和噪声较为敏感，识别准确性和鲁棒性有一定的局限性。

（2）基于深度学习的文本识别

深度学习作为一种新兴的机器学习方法，在模拟人脑神经网络进行分析学习方面展现了巨大的潜力。通过其多层网络的分层结构，深度学习能够有效地处理各种类型的数据，包括图像、声音和文本等。在这个

中国知网 https://www.cnki.net

过程中，输入层接收原始数据，隐藏层提取数据的抽象特征，输出层根据学习目标产生相应的结果。这种分层结构使得深度学习能够自动地发现数据中的规律和模式，从而实现了强大的学习能力。

通过神经网络的训练，深度学习模型能够识别图像中的对象、场景和特征，从而实现图像分类、目标检测、图像生成等任务。随着深度学习技术的蓬勃发展，特别是 CNN 和 RNN 的广泛应用，文本识别领域迎来了革命性的变革。深度学习方法在文本识别方面展现出了强大的能力，能够直接从原始数据中学习到更高层次的抽象特征，并具有强大的表征能力和泛化能力，不再依赖于手工设计的特征，这些方法主要分为字符级别和字符序列级别的文本识别两大类。

基于时序特征分类的方法：该方法结合了卷积网络和循环神经网络，通过卷积网络将图像转换为特征序列，再采用循环神经网络将图像特征序列识别为字符概率预测序列，并且通过引入连接时序分（Connectionist Temporal Classification，CTC）算法来解决可能因无法对齐导致模型无法计算损失函数和训练。识别流程如图 2-2 所示，过程如下：

（a）利用基于滑动窗口的卷积神经网络提取图像的一组序列特征或直接分割卷积神经网络提取的图像特征，得到一组序列特征；

（b）序列预测结果与真值文本无法对齐，借助 CTC 算法对齐预测结果和真值。

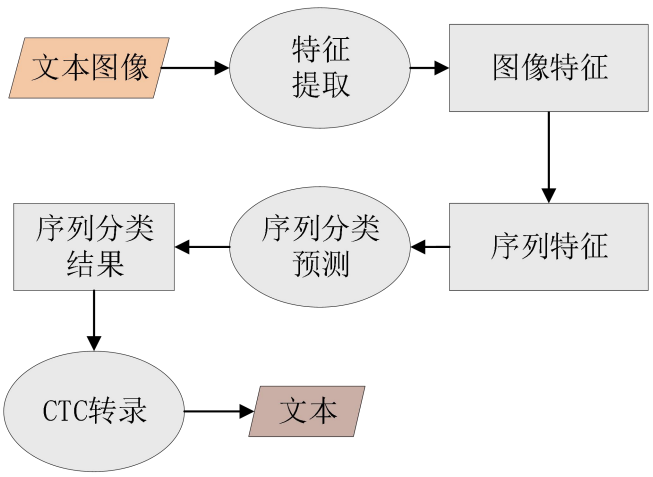

图 2-2 基于时序特征分类的方法

中国知网 https://www.cnki.net

基于编码器和解码器的方法：该方法是将一种序列转换为另一序列的算法框架，通过编码器将图像特征转换为中间语义编码特征，再由解码器将中间语义编码特征解码为文本序列。这种方法可以避免时序特征分类识别方法中的序列对齐问题，但存在信息丢失问题。为了解决这些问题，引入注意力机制来使得中间特征向量的权重不同，在解码器中时刻调整对编码器中不同部分的关注程度，以便于更准确地生成目标序列，提高模型的性能和准确性。识别流程如图 2-3 所示，过程如下：

（a）该方法包括编码器和解码器两个模块；

（b）编码器通过网络得到文本图像的中间特征编码向量；

（c）解码器利用中间特征编码向量进行循环解码，直到输出结束标记。

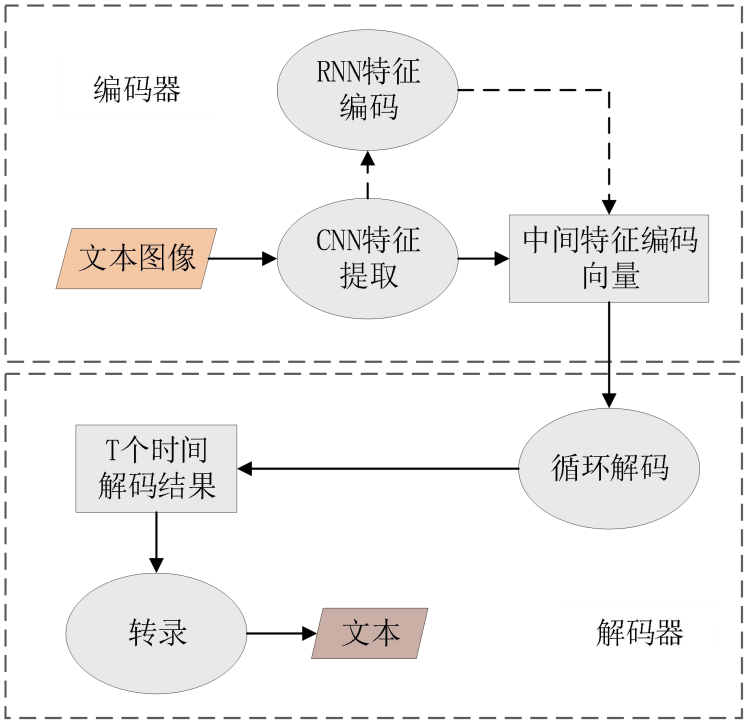

图 2-3 基于编码器和解码器的方法

2.1.2 汉字识别方法

作为世界上使用最广泛的语言之一，大约有 16%的人以汉语为母语。汉字识别（Chinese Character Recognition，CCR）作为计算机视觉的重要组成部分，在过去的几十年里受到了广泛的研究兴趣。汉字单字符拥

中国知网 https://www.cnki.net

有较为完整的语义信息，古代文字单字符拥有更多的信息，且词组间没有明显的区分，在识别过程中没有单词级别的识别过程。历史上的古汉字以石刻、铜铸、手写等形式产生，多变的形态与异体字等提升了古汉字的识别难度。汉字识别从类型角度来说，主要分为两个方面，一个是印刷汉字识别，另一个则是手写字体识别，如图 2-4 所示。

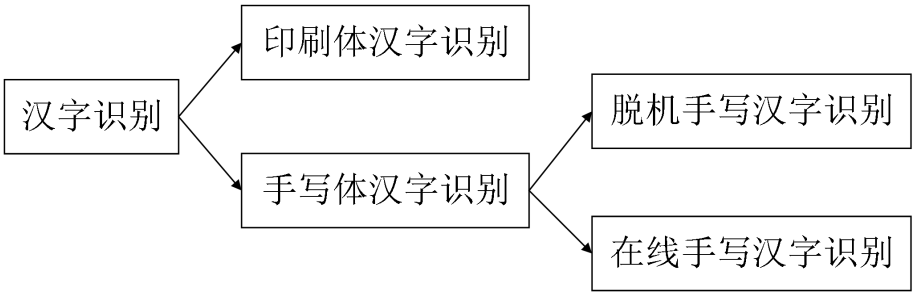

图 2-4 汉字识别类型

汉字识别方法的发展经历了多个阶段，从最初的手工特征设计到后来的深度学习模型。传统的 CCR 方法通常依赖于手工制作的特征以像素方式匹配汉字，需要手动提取文本的特征，然后通过分类算法进行识别。目前的 CCR 研究对象主要是宋体楷体等较为规范的印刷体文本，但是在较为潦草的手写汉字、较为多变的字体和复杂场景下的文字等具有挑战性的情况下实现效果较差。

传统的 OCR 识别方法集中在对特征的判别分类上，大体上可以被分为四类：基于文本结构、统计模式、神经网络和多分类器结合的方法。在文本识别的前期研究中，特征提取方法主要被分类为两类：结构特征提取和统计特征提取。基于结构的特征提取方法是针对汉字的轮廓、构造，如字的骨架信息、笔画特点等进行文本的区分，对存在笔画粘连、背景复杂、扭曲的文本鲁棒性较差。随着深度学习技术的兴起和发展，人们开始探索基于深度学习的古汉字识别方法，利用卷积神经网络提取更具有鲁棒性的识别方法。根据汉字的分解程度，这些基于深度学习的方法大致可以分为基于汉字、基于根式和基于笔划的方法。这些 CCR 方法在具有挑战性的手写文本和场景文本数据集准确性有较大提高。

传统的文字识别方法需要人工进行特征提取，虽然准确率较高，但是需要消耗大量的时间和精力，深度学习的文本识别可以自动提取特征，

中国知网 https://www.cnki.net

但是需要大量的数据进行支撑才能获得较好的识别精度。总体而言，汉字识别方法经历了从传统方法到深度学习的演进，取得了显著的进步，使得对汉字进行自动化识别变得更为准确和高效。

2.2 卷积神经网络

卷积神经网络是一种深度学习神经网络架构，专门用于处理和分析具有网格状结构的数据，尤其在图像识别领域表现出色。CNN 由输入层、隐藏层和输出层三层组成，其中输入层用来输入原始信息，隐藏层一般由卷积层、池化层和全连接层组成，输出层是 CNN 处理后的输出信息。CNN 通过卷积层、池化层和全连接层等组件实现对图像的高效特征提取和识别，网络的基本结构如图 2-5 所示。

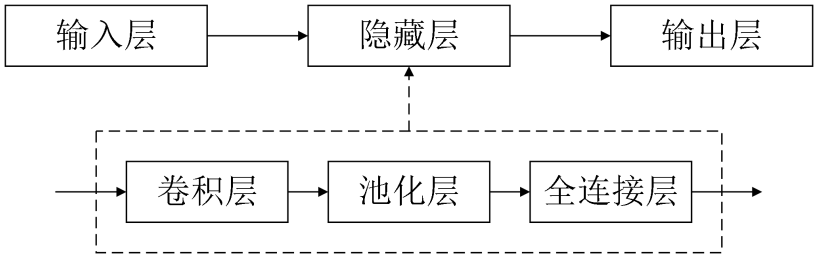

图 2-5 CNN 网络的基本结构

卷积层是 CNN 的核心组件之一，利用滑动卷积核在输入图像上提取局部特征，并保留空间信息。这种局部连接的方式使得网络能够有效地捕获图像中的局部模式和结构。池化层通过池化操作降低特征维度，增强模型的平移不变性，同时减少了参数数量和计算量。全连接层将特征映射转换为最终输出结果，并通过激活函数进行非线性映射，从而使网络能够对复杂模式进行建模和识别。优化算法通过调整网络参数使损失函数最小化，以提高模型的性能。

（1）卷积层：在图像处理中，卷积是一种重要的特征提取方法，在卷积神经网络中扮演关键角色。卷积层通过卷积运算提取输入图像的特征信息，其中关键参数包括滑动窗口步长、卷积核尺寸、边缘填充和卷积核个数。

滑动窗口步长指的是卷积核在进行卷积运算时每次移动的距离，影

中国知网 https://www.cnki.net

响输出特征图的尺寸。卷积核尺寸决定了卷积操作涉及的局部区域大小，影响特征的提取粒度。卷积核个数决定了输出特征图的深度，即提取的特征数量。边缘填充是在输入图像边缘周围填充额外像素值，以保持输出特征图的尺寸不变，常用于处理边缘信息。

浅层网络中的卷积层提取图像的低级特征，如颜色、纹理和边缘，而深层网络中的卷积层则提取更复杂、更抽象的高级特征，如形状和结构。卷积参数共享是指用同一组卷积核对图像的每个区域进行特征提取，有助于减少参数数量和提高模型效率。通过共享参数，卷积层可以学习到更加通用和抽象的特征，从而提高了模型的泛化能力和鲁棒性。卷积核的工作原理如图 2-6 所示：

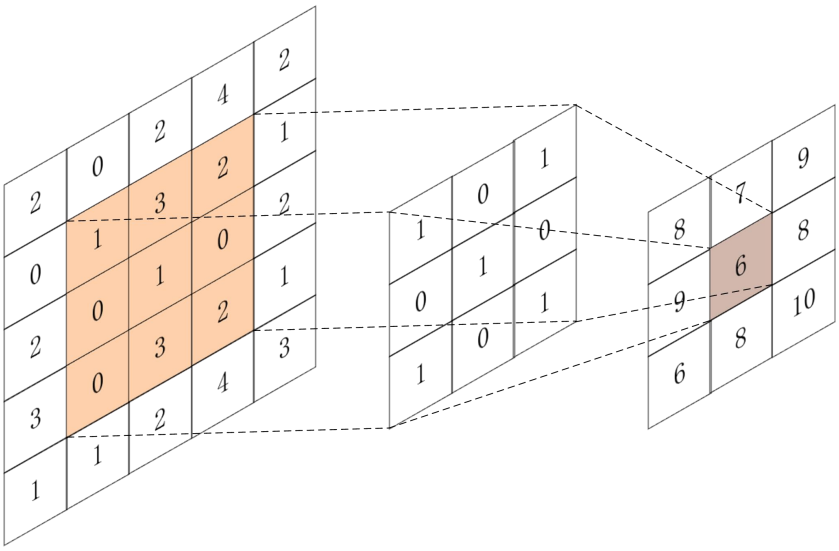

图 2-6 卷积操作流程

（2）池化层：实现了输入图像的局部连接，减少了网络模型中的参数量。因为相邻像素经过卷积操作后的特征往往具有较大的相似性，导致输出特征中存在大量重复信息。因此，为了减少参数量和降低计算量，通常会在卷积层之间插入池化层，通过池化操作对卷积层提取的特征信息进行维度降低。

常见的池化操作包括平均池化（Average Pooling）、最大池化（Max Pooling）、中值池化（Median Pooling）、随机池化（Stochastic Pooling）以及组合池化等。其中，均值池化选择每个区域的平均值作为

中国知网 https://www.cnki.net

输出值，最大池化选择特征图中每个区域的最大值作为输出值，中值池化选择每个区域的中值作为输出值，随机池化选择每个区域的随机值作为输出值，组合池化则采用不同的池化操作进行组合。

池化层的压缩过程是在卷积层后对特征图进行筛选，通过滑动窗口的方式，对特征图中的每个区域进行池化操作，从而得到池化后的输出特征图。最大池化和平均池化的过程如图 2-7 所示：

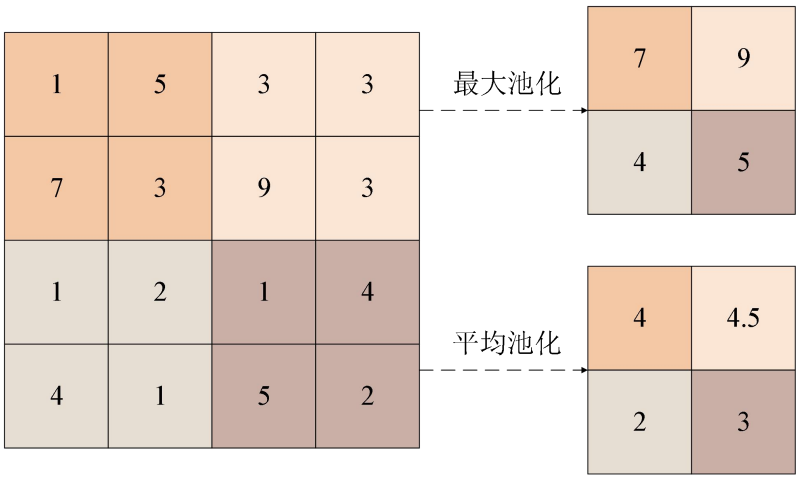

图 2-7 最大池化和平均池化

（3）全连接层：全连接层在卷积神经网络中起到了整合特征和进行分类的作用，是建立在卷积层基础上的一种神经网络结构，用于重新组合通过卷积层提取的局部特征，形成完整的图像表示，如图 2-8 所示。

全连接层类似于分类器，它利用所有的局部特征进行图像分类。在卷积神经网络中，归一化指数函数（Softmax）是一种常见的用于解决多分类问题的方法，它将二分类回归应用到多分类问题上。Softmax 将输入数据的线性组合映射到 0~1 之间，并进行归一化，得到每个类别的概率分布，模型选择具有最高概率的类别作为预测结果。

（4）激活函数：激活函数使得神经网络具有非线性建模能力。传统的模型仅由线性组合构成，只能表示线性关系。引入了激活函数后，神经网络得以引入非线性关系，从而使得网络能够更灵活地逼近各种复杂的非线性函数，扩展了其应用范围到众多非线性模型中。

激活函数的引入使得神经网络可以处理更为复杂的数据模式和任务，

中国知网 https://www.cnki.net

如自然语言处理、图像识别等。常见的激活函数包括 Sigmoid 函数、ReLU 函数、Tanh 函数等，它们在不同情况下具有不同的表现，但都能有效地引入非线性因素，丰富了神经网络的表达能力。

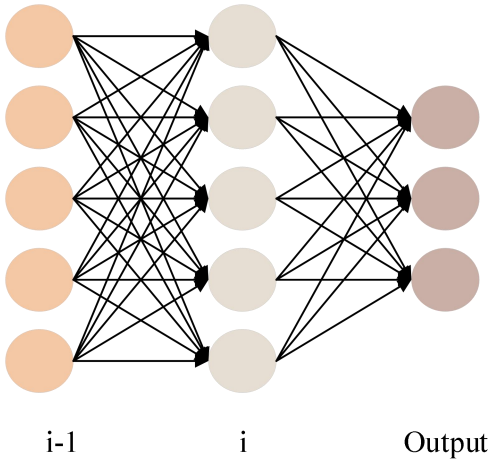

图 2-8 全连接层

（5）损失函数：损失函数用于衡量模型预测值与真实值之间的差异，是一个非负实值函数，通常表示为$L(Y,f(x))$。选择适当的损失函数直接影响模型的性能表现，不同类型的模型通常会采用不同的损失函数。在训练过程中，模型的目标是最小化损失函数，以使预测值尽可能地接近真实值。为了达到这个目标，通常会使用梯度下降等优化算法来调整模型的参数。

2.3 小波变换

小波变换是 1974 年法国信号处理工程师 J.Morlet[35]提出的一种常用于信号处理、图像处理领域的数学分析工具。小波变换的前身是傅里叶变换，由于传统的傅里叶变换只能处理时域空间的平稳信号，当面对非平稳信号时只能获得信号的频率组成，而无法提供信号在各个频率所对应的时刻。因此，研究者开始基于傅里叶算法进行改进，小波变换便是其中之一。小波变换是一种时间频率的局部化分析方法，能够自适应时频信号分析的需求，弥补傅里叶变换的不足，是信号分析领域的一项重大突破。

小波变换能够将函数、算子或数据分解为不同频率成分，并在多尺

中国知网 https://www.cnki.net

度上进行细化。在图像处理中，它能够将图像分解为具有不同方向特征的子带信号，进而进行频域和时域分析，以实现图像的融合和处理。通过小波变换能够更好地理解信号的局部特性和频率结构，从而实现对信号的有效分析和处理。小波变换具有多尺度性质，这意味着它能够同时捕获信号的整体和局部信息，使得我们可以在不同尺度下对信号进行分析。此外，小波变换还具有变聚焦的能力，通过逐步调整时域步长，可以将注意力集中在需要分析的特定细节上，从而实现对信号的精细描述。在小波变换理论框架下，存在两种典型的变换方式：连续小波变换（Continuous Wavelet Transform，CWT）和离散小波变换（Discrete Wavelet Transform，DWT）。连续小波变换通过在时域和频域上同时调整小波函数的尺度和平移来实现对信号的分解，而离散小波变换则是在有限尺度和离散时间下进行分解，使得其在实际应用中更为高效。

离散小波变换是一种在信号处理、图像处理和数据压缩等领域广泛应用的数学工具。与连续的信号不同，图像是以二维数据的形式存在的，二维离散小波变换可以将图像分解成不同尺度和方向的频域信息，从而对图像进行更加细致和全面的分析，通过小波基函数对图像进行分解和重构。在希尔伯特（Hilbert）空间中，小波函数族可以被分解为尺度函数$\phi(\cdot)$和小波函数$\psi(\cdot)$，使其得到一组正交且完备的小波族。具体而言，尺度函数$\phi(\cdot)$对应于原始函数的低频部分，而小波函数$\psi(\cdot)$对应于其高频部分。数学表达式如下：

$$W_{\varphi}(a,b)=\frac{1}{\sqrt{a}}\sum_{n}\varphi^{*}(\frac{n-b}{a}) \tag{2-1}$$

其中a,b分别表示尺度参数、平移参数。$\psi(\cdot)$是基本小波函数，$\psi^{*}(\cdot)$是$\psi(\cdot)$的共轭函数。由于小波函数族有尺度平移不变性和可调性，因此可用于提取信号的高低频成分，并生成一组小波系数，这些系数反映了不同尺度和频率信息，如峰值和边缘。

在图像处理中，给定一幅图像I，可以使用$2D-DWT$和四个卷积滤波器$(f_{LL},f_{HL},f_{LH},f_{HH})$将图像$I$分解为具有不同频率特征的低频子带$(I_{LL})$、垂直高频子带$(I_{HL})$、水平高频子带$(I_{LH})$和对角高频子带$(I_{HH})$。

中国知网 https://www.cnki.net

低频子带(I_{LL})表示图像I的近似图像，该图像包含了原始图像I的大部分平滑信息和图像结构，也可以理解为对原始图像进行压缩得到的有损失的图像。垂直高频子带(I_{HL})表示图像I的纵向的边缘特征、水平高频子带(I_{LH})表示图像I的横向的边缘特征、对角高频子带(I_{HH})表示图像I的对角特征。这些高频子带包含图像中的细节、边缘和纹理等关键信息，可以用于检测和提取图像的重要特征。由于高频子带方向和频率不同，它们有助于有效地反映图像的多维特性。本文将二维 Haar 离散小波变换的卷积滤波器定义为:

$$f_{LL}=\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} \quad f_{HL}=\begin{bmatrix}1 & -1\\ 1 & -1\end{bmatrix} \quad f_{LH}=\begin{bmatrix}1 & 1\\ -1 & -1\end{bmatrix} \quad f_{HH}=\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix} \tag{2-2}$$

在DWT分解过程中，滤波器被视为 4 个2×2卷积核，权值固定，步长为 2。得到$I_{LL},I_{HL},I_{LH},I_{HH}$经过$DWT$后各$(i,j)$的像素值为:

$$\begin{cases} I_{LL}(i,j)=I(2i-1,2j-1)+I(2i,2j-1)+I(2i-1,2j)+I(2i,2j) \\ I_{HL}(i,j)=I(2i-1,2j-1)-I(2i,2j-1)+I(2i-1,2j)-I(2i,2j) \\ I_{LH}(i,j)=I(2i-1,2j-1)+I(2i,2j-1)-I(2i-1,2j)-I(2i,2j) \\ I_{HH}(i,j)=I(2i-1,2j-1)-I(2i,2j-1)-I(2i-1,2j)+I(2i,2j) \end{cases} \tag{2-3}$$

在实际应用中，小波变换在医学图像处理和通信领域等方面有着成功的应用案例。近年来，研究者们开始探讨深度学习与小波变换的结合，以提高对信号和图像的表示能力。这种结合可以充分利用深度学习的强大特征学习能力和小波变换的多尺度分解能力，实现对信号和图像更加准确和高效的分析和处理。

2.4 注意力机制

注意力机制（Attention Mechanism）最早由 Bahdanau 等人[28]于 2014 年提出，是一种在计算机科学和机器学习中常用的技术，其目的是解决 RNN 中只支持固定长度输入的问题，可以使模型在处理序列数据时更加准确和有效。在自然语言处理中，注意力机制常常被用于机器翻译、语音识别、文本摘要等任务中。在传统的神经网络中，每个神经元

中国知网 https://www.cnki.net

的输出只依赖于前一层的所有神经元的输出，而在注意力机制中，每个神经元的输出不仅仅取决于前一层的所有神经元的输出，还可以根据输入数据的不同部分进行加权，即对不同部分赋予不同的权重。这样可以使模型更加关注输入序列中的关键信息，从而提高模型的精度和效率。

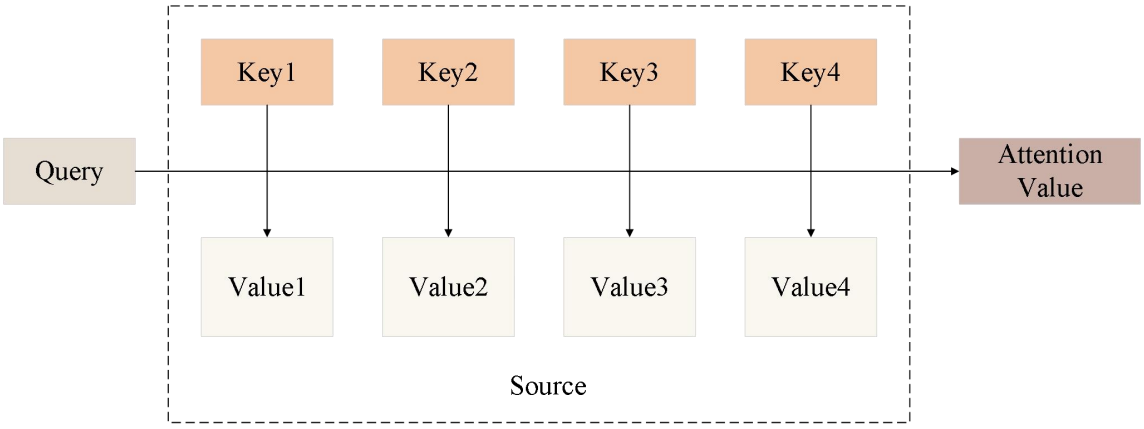

图 2-9 注意力计算模型

注意力机制通过给定一个查询特征向量（*Query*）和被对比的键（*key*），通过计算 *Query* 与 *Key* 的相关性，实现对物体本身的特征向量值（*Value*）的注意力权重分配，生成最终的输出结果，如图 2-9 所示。注意力机制的计算公式如下公式 2-4：

$$Attention(Query, Source) = \sum_{i=1}^{Lx} Similarity(Query, Key_i) * Value_i \tag{2-4}$$

其中，L_x 代表 Source 的长度。根据 *Query* 和 *Key* 计算两者之间的相关性或相似性，得到注意力得分，常见的方法包括点积、余弦相似度，MLP 网络：

$$\text{点积：}\ Similarity(Query, Key_i) = Query \cdot Key_i \tag{2-5}$$

$$\text{余弦相似性：}\ Similarity(Query, Key_i) = \frac{Query \cdot Key_i}{\|Query\| \cdot \|Key_i\|} \tag{2-6}$$

$$\text{MLP 网络：}\ Similarity(Query, Key_i) = MLP(Query, Key_i) \tag{2-7}$$

对注意力得分进行缩放，再进行归一化，将原始计算分值整理成所有元素权重之和为 1 的概率分布；另一方面也可以通过 *Softmax* 函数的

中国知网 https://www.cnki.net

内在机制更加突出重要元素的权重。一般采用公式 2-8 进行计算：

$$a_i = Softmax(Sim_i) = \frac{e^{Sim_i}}{\sum_{j=1}^{L_x} e^{Sim_j}} \tag{2-8}$$

根据权重系数对 *Value* 值进行加权求和，得到注意力权值，此时的 *Value* 值具有一定的注意力信息，关注更重要的信息，忽视不重要的信息：

$$Attention(Query, Source) = \sum_{i=1}^{L_x} a_i \cdot Value_i \tag{2-9}$$

以上三个阶段计算流程图如图 2-10 所示：

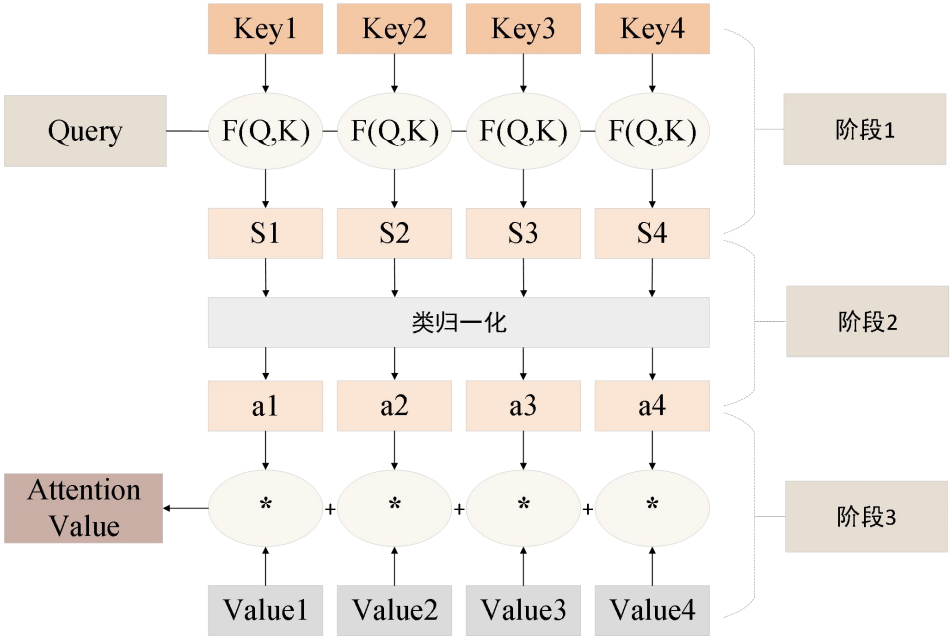

图 2-10 注意力计算流程图

目前常见的注意力机制包括通道注意力机制、空间注意力机制、自注意力机制。

2.4.1 通道注意力

对于处理二维图像的卷积神经网络来说，一个重要的维度是图像的尺度空间，另一个维度则是通道。为了更好地利用通道信息，通道注意力机制被引入，核心思想是通过网络学习的方式自动获取每个特征通道

中国知网 https://www.cnki.net

的重要程度，并为每个通道赋予不同的权重系数，从而强化重要的特征并抑制非重要的特征。通道注意力机制能够使输入图像更有意义，通过网络计算出输入图像各个通道的权重，重点关注包含关键信息的通道，减少对重要信息较少的通道的关注，从而提高特征表示能力。

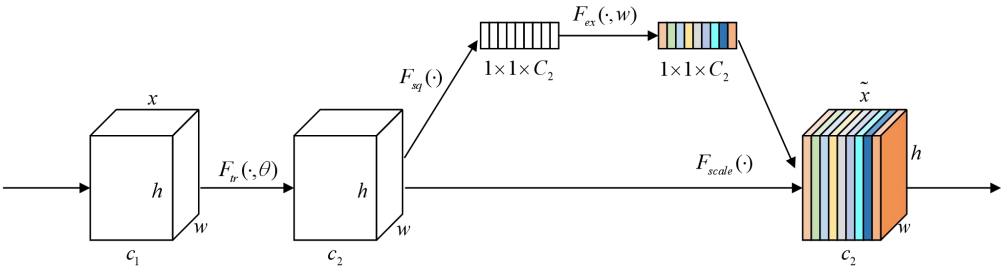

图 2-11 SE 通道注意力模型

常见的通道注意力机制包括 SE 注意力机制[36]、ECA 注意力机制[37]、CBAM 注意力机制[38]，其中最为经典的通道注意力机制是 SE 注意力机制，其核心操作包括挤压和激励，主要目的是在通道维度上引入注意力机制。如图 2-11 所示，在 SE 注意力机制应用前，特征图中的通道具有相等的重要性。然而，经过 SE 注意力机制处理后，每个特征通道的重要性被区分，使得神经网络能够集中关注权重值较高的通道，从而增强了对于当前任务相关特征图通道的关注度，同时抑制了对当前任务无关特征通道的影响。

2.4.2 空间注意力

在图像的空间维度中，只有与任务相关的区域才具有重要性，例如文本图像中的文字区域。为此，引入了空间注意力，通过对特征空间内部的关系来产生空间注意力特征图，以便于聚焦特征图上的有效信息。

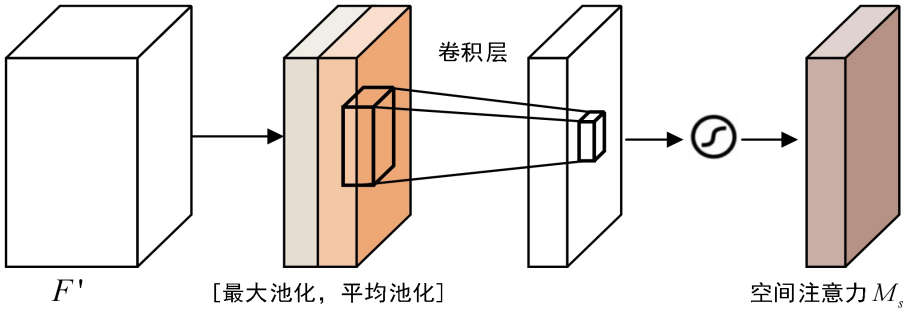

图 2-12 空间注意力

中国知网 https://www.cnki.net

空间注意力的目标在于增强目标区域的特征表达能力，其本质是通过空间转换将原始图像中的空间特征信息映射到另一个空间，并保留关键信息。该模块为每个位置生成权重掩膜，并对输入特征信息进行加权输出，从而增强感兴趣的目标区域，同时减弱不相关的背景区域，如图2-12所示。

2.4.3 自注意力

自注意力机制通过筛选重要信息、过滤不重要信息的原理，降低了对外部信息的依赖，能够更有效地捕捉特征的内部相关性。其解决的核心问题在于神经网络所接收的输入通常是多个大小不同的向量，并且这些向量之间存在一定的关系。然而，在传统的全连接神经网络中，很难有效地利用这些关系信息，导致模型在训练过程中表现不佳。为了解决这一问题，自注意力机制被引入，以便让模型更好地理解输入数据中不同部分之间的关联性。

自注意力机制允许模型根据输入数据的内部关系来动态地调整权重，使得模型可以更加关注与任务相关的部分。自注意力机制虽然能够有效地处理序列数据，但在处理图像等具有平移不变性和特征局部性的数据时，其性能可能受到限制。此外，自注意力机制通常需要大量数据进行训练，以便有效地学习数据之间的关系。因此，在小样本数据情况下，自注意力机制可能表现不及卷积神经网络。另外，自注意力机制在设计上是考虑了所有输入向量之间的关系，但并不包含关于向量位置的信息。

2.5 长尾分布

真实场景下的数据通常呈现为长尾分布特征，少数头部类别具有大量的样本，而大量的尾部类仅有少数样本，基于深度学习的分类算法应用于长尾分布数据集时，识别效果不好，长尾分布特征如图2-13所示。在卷积网络训练过程中，由于尾部类别的数据量较少，模型学习到的特征信息不够充分，导致模型对尾部类别的学习效果较差。目前解决长尾分布的两个主要思路是：重采样和重加权，二者可以单独使用，也可以

中国知网 https://www.cnki.net

结合使用，以提高模型对于不平衡数据的处理效果。

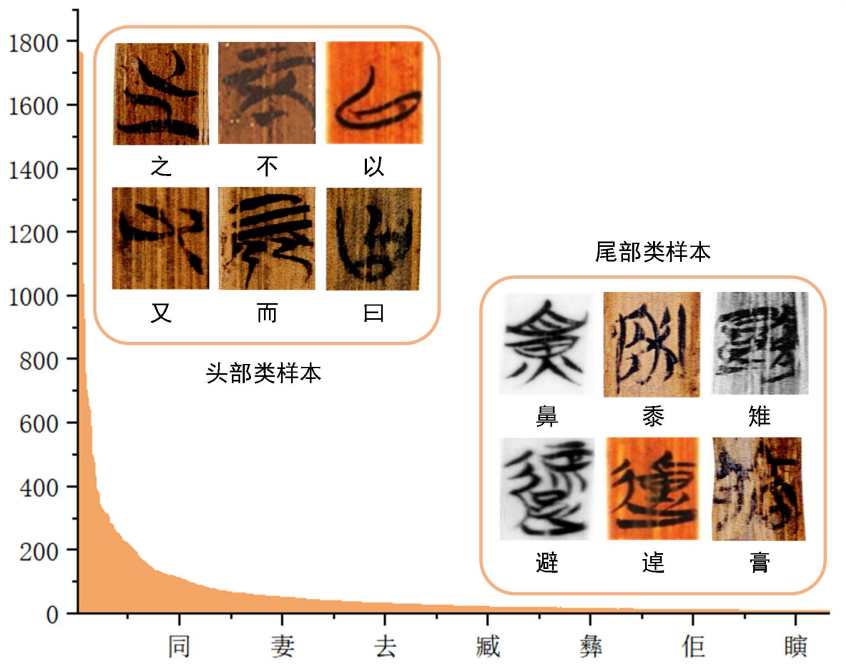

图 2-13 长尾分布

（1）重采样：该方法通过对数据集进行对头部类别样本的欠采样和尾部类别样本的过采样。过采样操作通过增加少数类别的样本数量，使得不同类别之间的样本数量更为接近，从而缓解样本不平衡问题，但过采样容易在尾部类别过拟合。欠采样操作通过减少头部类别的样本数量，来达到样本数量平衡的状态，但由于样本获取难度大，为了不损失过多的头部类别信息，重采样过程一般会选择过采样的方式。为避免丢失大量信息，可采用数据增强来增加样本的多样性，减少过拟合的风险。

（2）重加权：该方法通过调整不同类别在损失函数中的占比来缓解长尾分布导致的梯度占比失衡问题，重加权操作主要体现在分类的损失函数中，因为损失函数设计过程更方便，在复杂的任务上，使用重加权损失函数来解决长尾分布问题的方法更为常见。

2.6 本章小结

本章首先对文本识别的传统方法和基于深度学习的方法进行了阐述，并简要介绍了卷积神经网络的各层结构。另外，还探讨了小波变换、注

中国知网 https://www.cnki.net

意力机制以及基于长尾分布的数据集改进方法，为后续章节对楚系简帛古汉字的识别方法奠定了基础。这些内容为后续的创新研究提供了启发和方向。

中国知网 https://www.cnki.net

第 3 章 楚系简帛古汉字数据集

本章介绍了本研究所收集整理的楚系简帛古汉字数据集以及基准实验的相关内容，着重阐述了楚系简帛古汉字数据集的背景。其次，探讨了实验配置和评价指标，并对基准实验结果进行了分析。

3.1 Chu730 数据集简介

古汉字是研究中国文化时不可或缺的重要对象[39-41]。在文字的历史演变过程中，简帛古汉字在春秋金文与秦汉篆隶之间扮演着关键的角色，楚系简帛古汉字是中国汉字发展史上风格独特的一种文字，是湖北地区在春秋战国时期所使用的文字。竹简和帛书是古文字的主要载体，是中国古代主要用于书写典籍、文书的材料。简帛被视为研究古文字形体发展与演变的重要载体之一，书写保存的文字记录了春秋战国时期的部分历史和文明。因此，认识和研究这些珍贵的文字，对中国考古学和文献学具有重要的意义[16-18]。尽管对古文字的形体、演变规律以及内容结构已经进行了大量研究，但是古文字识别与人工智能文字识别的交叉研究尚未得到充分展开。特别是在简帛古汉字领域，研究者仍然需要依赖于专业的古文字学者来进行文字识别，而每一个文字的识别都需要进行大量的检索工作。这种状况导致了人工识别工作量的显著增加，同时也给非专业人员学习古文字带来了一定的难度，阻碍了该领域的进一步研究。为了加快对中华传统文化的探索与传承，实现中华文化载体的永久性保存，利用计算机技术辅助古籍保存与识别的研究[42]日益受到重视。针对古汉字的智能识别研究已成为古文化研究的重要组成部分，对于追溯中国古代历史、经济、文化、科学技术的发展具有重要意义。

简帛古汉字是古人写在竹片或帛书上的文字，属于手写体的一种。与现代手写体识别相比，简帛手写体识别存在显著差异，主要表现在文字的载体特性、书写方式、保存状态以及时间等方面的影响上。竹简文字的识别受到多种因素的影响，包括载体材料、墨痕、残缺和修护等，而其字形也可能与现代文字不完全一致，这为竹简手写体识别带来了一

中国知网 https://www.cnki.net

定的挑战，如图 3-1 所示。作为古代的文字载体，竹简的保存状态和文字形态往往受到时间和环境等多种因素的影响，常常出现残缺或修护的情况，这给手写体识别增加了额外的困难。目前尚未有公开的楚系简帛古汉字识别数据集，本文在导师的指导下独立整理并准备了楚系简帛古汉字数据集—Chu Bamboo and Silk 730（Chu730），该数据集的文字图像源于楚系简帛字典清华篇和中国古代简帛字形、辞例数据库，共包含 38894 幅楚系简帛文字图像，覆盖了上海博物馆藏简、清华大学藏简、曾侯乙墓简、望山简、郭店 1 号墓简、包山 2 号墓简等已出土整理的文字资料。

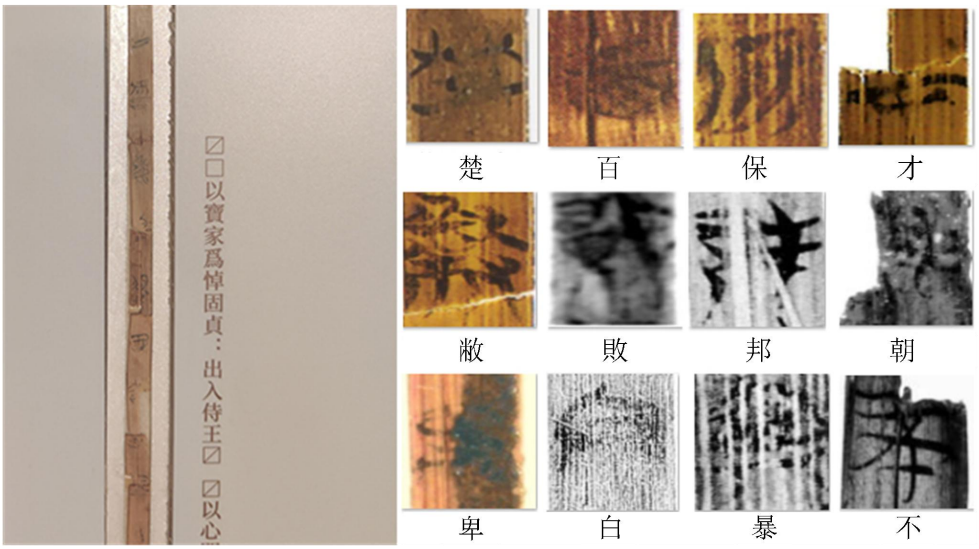

图 3-1 文字退化、模糊、干扰、残缺

3.2 数据集统计与分析

Chu730 数据集由 38894 张真实古汉字图像组成，文字按照简体汉字可分为 730 类，本文实验将数据集划分为训练图像 32918 张，测试图像 5976 张。采用随机裁剪和翻转等常用的数据增强方案。本文对该数据集的相关性信息进行统计，表 3-1 为样本数量，文字数量，训练集和测试集划分等信息。训练集和测试集中的单字符样本数量不是严格的一一对应关系。这种现象反映了实际样本的特点，即在训练和测试集中可能存在一定程度的重叠或不均衡。

中国知网 https://www.cnki.net

表 3-1 数据集分布

数据集名称	图像数	训练集/测试集	类别
Chu730	38894	32918/5976	730

数据集中训练集和测试集样本均为原始尺寸，总体上文本宽高比为 0.69。真实场景下的楚系简帛古汉字呈现出长尾分布的特征，与严格的正态分布不符合。在数据集中，尾部类别包含较多单个字体样本，这些类别样本数量占总样本数量的 0.23%，占字符总数量的 12.05%。这说明了在真实场景中，某些字符出现的频率要低于其他字符。图 3-2 列出了前五十个出现次数最多的字符以及字符的样本数量，“之”，“不”，“以”，“又”，“而”，“曰”，“人”，“子”，“王”，“君”，“可”等，反映了真实场景下楚系简帛古汉字的分布特点。这些字符可能在古代文献、铭文或文物中频繁出现，因此在数据集中出现的次数较多。

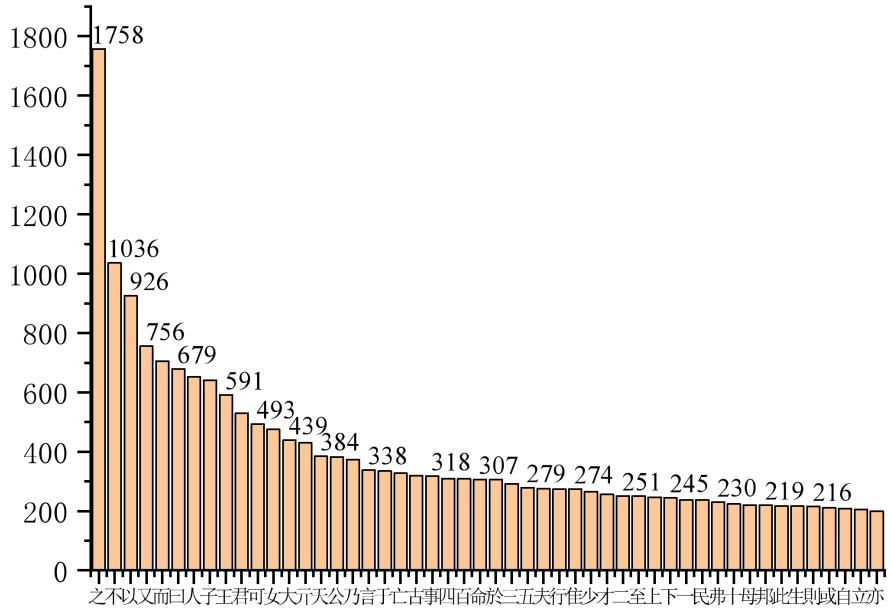

图 3-2 Chu730 前 50 类文字样本及数量

楚系简帛古汉字的特点是笔画多、构造繁复，且没有标准化的规范。受不同地域、时间、写作者以及书写工具的质量影响，笔画和字形在写作中会存在一定的变式和差异。例如，“美”与“媚”、“使”与“事”、“往”与“遊”这几个汉字的笔画分布相似，且笔画走势具有

中国知网 https://www.cnki.net

的方向性信息相似。“善”与“商”在所占面积相同的情况下，笔画数却不同。汉字的基础笔划数量有限，为了表示更多的意义，需要通过变形、组合和拼接形成大量的汉字。由于笔画和结构的微小变化会影响字的意义，因此在书写时容易出现相似的汉字和混淆。图 3-3（a）显示了楚系简帛古汉字的类间相似性。同时，楚系简帛古汉字是由甲骨文演变而来的，形象和功能上相近的字在书写时较容易混淆。图 3-3（b）显示了同一汉字的类内差异性。

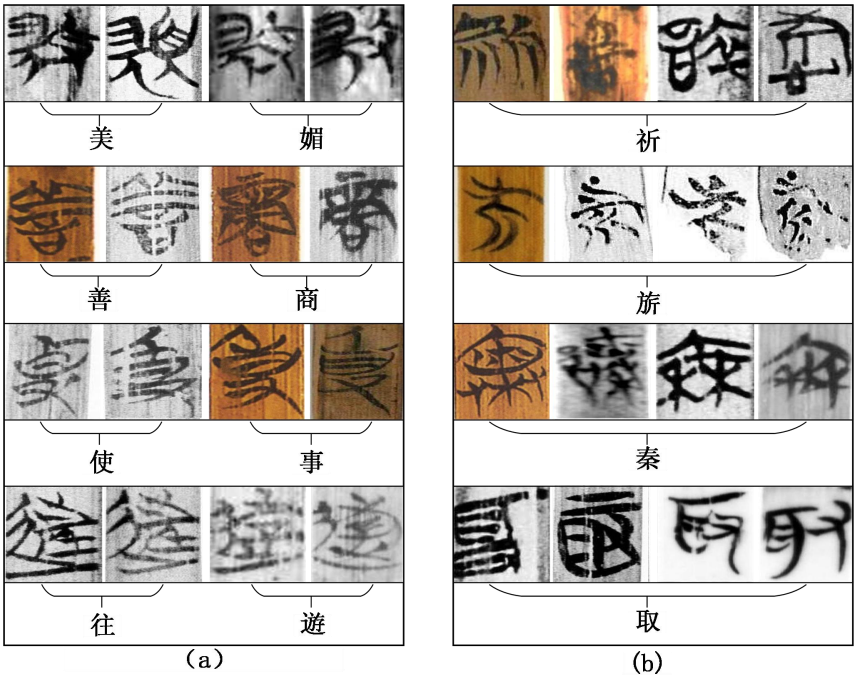

图 3-3（a）类间相似性（b）类内差异性

3.3 实验配置

3.3.1 实验环境及参数设置

本文实验所用的操作系统为 Windows，处理器型号为 Intel(R) Core(TM) i3-10105F CPU @ 3.70GHz；显卡型号为 NVIDIA GeForce RTX 3060；在 Pytorch 上进行训练与测试。选择使用学习速率为 0.2 的随机梯度下降法进行优化，动量值为 0.6。批尺寸大小设置为 64，100 个 epoch 后训练停止。实验环境配置如表 3-2 所示：

中国知网 https://www.cnki.net

表 3-2 实验平台和环境配置表

名称	配置
操作系统	Windows 10
CPU	Intel(R) Core(TM) i3-10105F CPU @ 3.70GHz
GPU	NVIDIA GeForce RTX 3060
开发环境	Python3.8+Pytorch+CUDA11.1+CUDNN8.0.4

3.3.2 评价指标

本研究采用文字识别准确率作为模型测试的评价标准，文字识别准确率反映了所有识别正确的样本数占总测试样本数的百分比，具体计算如下公式 3-1：

$$Accuracy = \frac{1}{M}\sum_{i=1}^{M}\frac{\left|Y_i \cap G_i\right|}{\left|Y_i \cup G_i\right|} \tag{3-1}$$

其中，M 为所有测试集样本数量，Y_i 代表测试样本 $n_i \in \{n_1, n_2, \cdots, n_k\}$ 的预测标签集合，G_i 代表 n_i 的真值标签集合。这个评价标准能够直观地反映出模型在文字识别任务上的表现，准确率越高，代表模型的性能越好。

本研究在准确率的基础上提供了模型的参数量和计算量，参数量用来衡量模型的大小，计算量用来衡量模型的复杂度。

参数量计算公式如下：

$$\text{卷积层：}\ Params = C_o \times \left(C_i \times K_h \times K_w + 1\right) \tag{3-2}$$

$$\text{全连接层：}\ Params = \left(I + 1\right) \times O \tag{3-3}$$

计算量计算公式如下：

$$\text{卷积层：}\ FLOPs = 2 \times C_i \times K_w \times K_h \times W \times H \times C_o \tag{3-4}$$

$$\text{池化层：}\ FLOPs = H \times W \times C_i \tag{3-5}$$

$$\text{全连接层：}\ FLOPs = [I + (I - 1) + 1] \times O \tag{3-6}$$

其中 H, W 为特征图的高和宽，C_i 为输入通道数，C_o 为输出通道数，K_h 代表卷积核宽，K_w 代表卷积核宽，当卷积核有偏置项时则＋1，否则不加。全连接层计算公式中，I 为输入向量的维度，O 为输出向量的

中国知网 https://www.cnki.net

维度。由于池化层为下采样操作，因此不考虑池化层的参数量。

3.4 基准实验结果与分析

3.4.1 基准实验

（1）ResNet 网络

在图像分类任务中，传统卷积神经网络由于网络层数的增加而导致梯度消失，使得模型的学习变得困难。而 ResNet[43]的残差块可以处理浅层特征，并进一步提取更多有用的特征信息，常被用于图像分类和识别任务中作为主要的特征提取结构。本章选用 ResNet50 来获取图像特征，这种网络结构解决了卷积神经网络过深导致的梯度消失和退化问题，并具有更快的收敛速度和更易学习的特点。

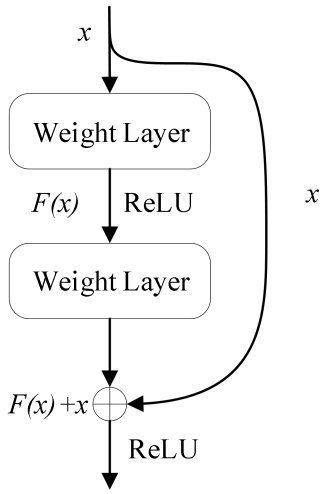

图 3-4 残差单元结构

设输入为x时所学特征为$H(x)$，$F(x)=H(x)-x$为网络学习残差，则残差单元可以表示为公式 3-7：

$$\begin{aligned} y_l &= x_l + F(x_l, \{W_l\}) \\ x_{l+1} &= f(y_l) \end{aligned} \tag{3-7}$$

其中，x_l, y_l示第l的输入和输出，W_l为权重矩阵，$f(\cdot)$表示 ReLU 激活函数。对于L层残差单元，从浅层到深层的学习特征为公式 3-8：

$$x_L = x_l + \sum_{i=l}^{L-1} F(x_i, \{W_i\}) \tag{3-8}$$

中国知网 https://www.cnki.net

通过学习残差，将部分原始输入的信息传递到恒等映射层，然后再直接输入下一层，解决了随着网络深度的增加而出现的梯度消失的问题。残差单元结构如图 3-4 所示。由于融合了上一层卷积网络，可以在输入特征的基础上进一步学习到新的特征，从而提高了模型的性能，使得模型在处理复杂数据时具有更强的表达能力和学习能力。

（2）VGG 网络

VGG[44]主要利用较小的卷积核来搭建网络，包括卷积层、池化层和全连接层，卷积层和池化层将原始图像数据映射到特征空间中，全连接层将特征映射进行分类，VGG16 网络结构如图 3-5 所示。该网络使用多个小卷积核代替单个大卷积核。具体地，使用两个 3×3 卷积核代替一个 5×5 卷积核，使用三个 3×3 卷积核代替一个 7×7 卷积核，这种方法能够减少网络参数量，增加网络的深度，并在一定程度上提高网络的特征提取能力，在最后使用了 3 个全连接层，使用 Softmax 函数进行分类。

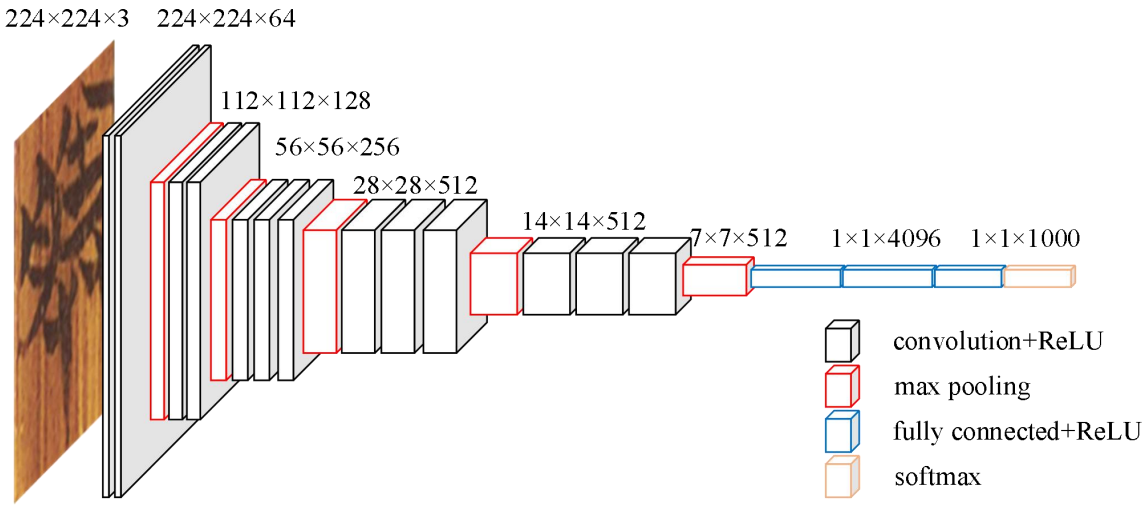

图 3-5 VGG16 结构图

（3）Swin Transformer

Transformer[45]是以注意力机制为核心的一种自然语言处理模型。将 Transformer 从自然语言处理领域转移到图像处理领域需要解决许多问题。例如，不同图像的大小差异可能很大、图像具有空间信息而非线性排列、以及图像高分辨率导致的计算量增大等。Swin Transformer[46]的出现克服了上述问题。Swin Transformer 以 Transformer 为主体思想，其通过使用

中国知网 https://www.cnki.net

滑动窗口的方式，在允许局部窗口之间信息流通的同时将自注意力的计算开销减小到与图像大小呈线性复杂度。

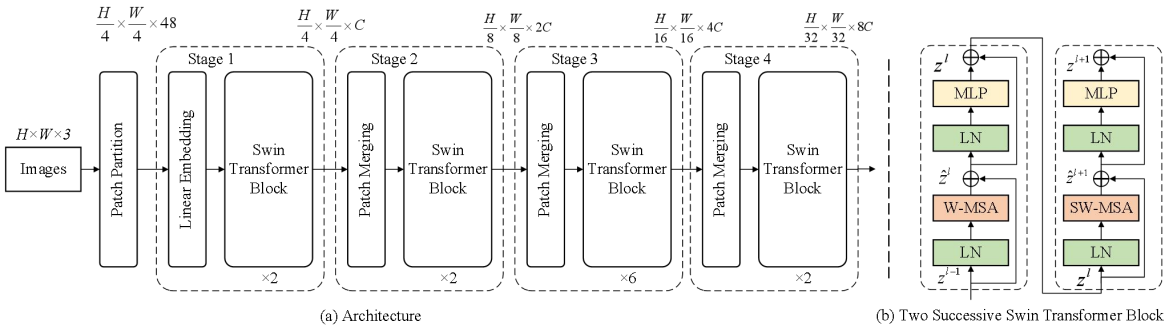

图 3-6 Swin Transformer 结构图

Swin Transformer 的结构如图 3-6 所示，其中虚线左侧是模型的整体结构，主要分为四个阶段，每个阶段中都包含 Swin Transformer 块，该模块的展开结构如虚线右侧所示。由于相邻两个模块在进行自注意力运算时有所不同，因此虚线右侧是连续两个该模块的整体结构，该模块总是成对出现。

表 3-3 对比实验结果

网络结构	参数量 /M	计算量 /G	准确率/%				
			Top-1	Top-2	Top-3	Top-4	Top-5
SqueezeNet[47]	1.10	2.95	67.49	75.72	79.57	81.76	83.55
ShuffleNet[48]	1.09	0.05	82.30	87.99	90.09	91.53	92.35
RegNet[49]	2.59	0.21	82.76	88.44	90.38	91.62	92.55
AlexNet[50]	16.08	0.31	83.25	89.16	91.33	92.42	93.12
GoogleNet[51]	12.55	1.58	86.09	90.91	92.55	93.44	94.04
Vgg16[52]	137.25	15.47	86.51	91.23	92.91	93.84	94.60
MobileNet[53]	3.16	0.33	86.51	91.33	92.94	93.93	94.60
DenseNet121[54]	7.70	2.91	87.15	91.32	92.94	93.91	94.41
Swin transformer[46]	87.49	15.47	87.18	91.28	92.94	93.69	94.24
Vision transformer[55]	86.36	17.58	83.32	89.04	90.93	91.99	92.47
ResNext50[56]	24.48	4.30	87.15	92.00	93.49	94.29	94.85
ResNet50[43]	25.00	4.14	87.42	91.92	93.64	94.48	94.90

中国知网 https://www.cnki.net

3.4.2 实验结果与分析

为了验证本研究所提出的楚系简帛古汉字数据集是否适用于深度学习方法，将一系列经典的卷积神经网络方法应用于该数据集，并将不同网络的识别结果进行比较，如表 3-3 所示。ResNet50 在对比的网络中实现了最高的识别准确率，而 SqueezeNet 由于深度特征提取有限而表现出最低的准确率。传统的 CNN 模型具有较低的计算复杂度，但识别精度不足。

3.5 本章小结

本章首先介绍了本研究所使用的数据集，数据集的背景、数据集的统计与分析，并介绍了本研究的实验环境配置和评价指标，为下文的结果阅读做出介绍，并在经典的卷积神经网络上对数据集进行了识别分类，实验结果证明本文的数据集适用于深度学习，为下文研究奠定基础。

中国知网 https://www.cnki.net

中国知网 https://www.cnki.net

第 4 章 基于离散小波变换的楚系简帛古汉字识别方法

4.1 研究背景

楚系简帛古汉字是湖北地区在春秋战国时期使用的文字，是甲骨文字符到金文的过渡性文字，其书写形式与一般的古汉字差别较大，因此识别过程面临着多样的挑战。其识别难点主要为以下两方面：（1）不同于自然场景文字，简帛古汉字所处的竹简或帛书经过数千年的沉积和腐蚀，大量文物表面出现损伤和变形，部分文字图像模糊甚至出现退化残缺，背景干扰、文字残缺等情况普遍存在。（2）不同于手写文字，简帛古汉字的繁简程度较为多变，字形相似且没有标准化规范，因此很容易出现混淆和识别错误。此外，简帛古汉字的年代跨度广，不仅包含了不同地域和文化背景下的多种不同风格汉字，还存在许多形态各异、难以辨认的文字，如图 4-1 所示。因此，如何高效、准确地识别这些古汉字，成为了古代文献研究的重要课题。

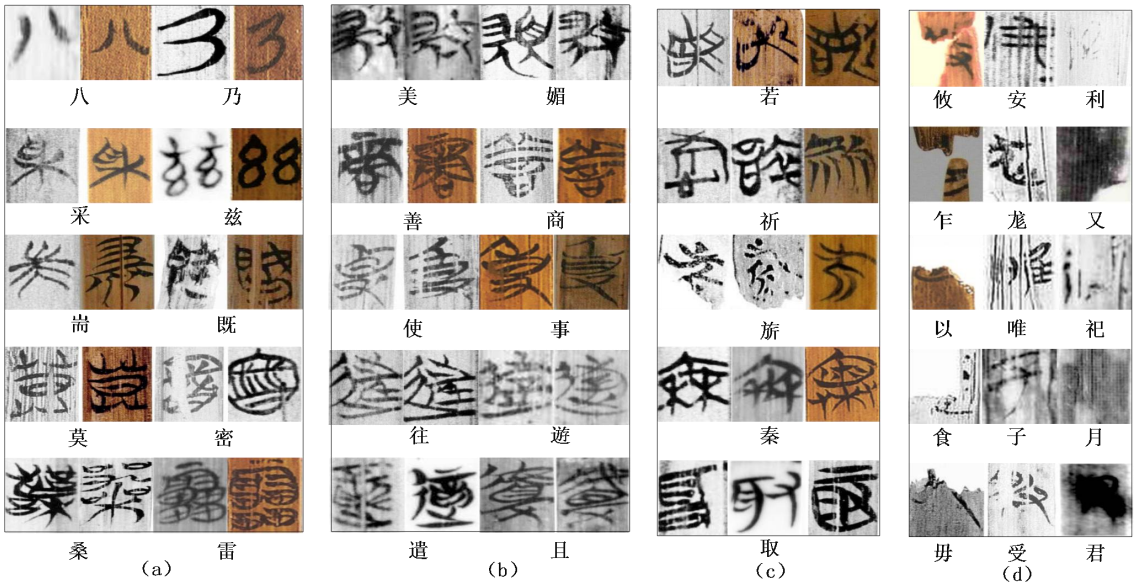

图 4-1（a）文字图像结构由易到难（b）文字类间相似性（c）文字类内差异性（d）文字残缺、退化、模糊

目前，对于甲骨文等古文字的智能识别研究已取得了初步的进展，但针对楚系简帛古汉字的智能识别研究目前较为缺乏，楚系简帛古汉字

中国知网 https://www.cnki.net

是一种具有极强笔画特性和明显纹理特征的语言文字，在识别过程中可以使用离散小波变换来对高频和低频信号进行分解，从而得到更加精细的纹理特征描述。同时，离散小波变换也具有压缩和降噪能力，可以降低图像在传输和保存过程中所带来的噪声。在使用离散小波变换进行特征提取后，还可以应用注意力机制，引入更多包含文字关键结构特征的特征图，从而提高识别准确率。针对以上分析，本章提出了使用传统的离散小波变换、卷积神经网络和注意力机制相结合的方法来提升识别准确率。主要工作如下：

（1）提出了融合离散小波的残差块 W-bneck。本章提出的 W-bneck 模块包含了离散小波卷积，SE 注意力和平均池化三部分，解决了特征信息丢失和结构特征忽略的问题。考虑到离散小波变换的低频子带包含大量的近似特征信息可以保留较多的全局信息，提出了使用离散小波变换的特征子带来替代卷积保留全局信息。考虑到原始残差连接下采样操作中会损失大量特征，使用平均池化保留特征图像的更多信息，并加入注意力关注重点特征。

（2）提出了融合离散小波变换的注意力机制。考虑到传统注意力机制使用平均池化和最大池化提取特征，但存在特征信息丢失和结构特征忽略等问题，提出了使用离散小波变换提取近似特征和结构特征的特性来替换注意力中的池化，实现了保留结构特征的最大化。使用离散小波注意力进行层间特征关注，重点关注通道增加时的关键特征。本文提出了小波注意力机制，使用离散小波注意力进行层间特征关注，重点关注通道增加时的关键结构特征。

4.2 相关工作

4.2.1 注意力机制

人类视觉系统可以快速捕捉关键区域，因此许多研究引入了注意力机制来使模型聚焦于关键信息并减少冗余。注意力机制是一种特征权重再分配策略，主要有空间注意力和通道注意力两种类型。Hu 等人[36]提出的 SENet 是经典的通道注意力机制，它通过在空间维度上进行全局自

中国知网 https://www.cnki.net

适应平均池化和全连接层来学习通道权值。Li 等人[57]在 SENet 基础上提出了 SKNet，使用不同的卷积核来动态计算通道权值。Wang 等人[37]提出了 ECANet，使用一维卷积层替代 SENet 中全连接层的通道缩减方法。同时，空间注意力也得到了广泛的关注。空间注意力是一种自适应的空间区域选择机制，Wang 等人[58]提出了 NLNet 利用自注意力机制模拟像素级的成对关系，构建密集的空间特征映射。A2-Net[59]和 CCNet[60]通过不同的方式对 NLNet 进行了改进。由于空间注意力关注空间特征，忽略通道信息，因此，混合注意力的出现解决了以上问题，即将通道注意力和空间注意力按照串行或并行的方式进行组合。Rark 等人[61]提出了 BAM 方法，按照并行的方式组合了通道注意力模块和空间注意力模块。而 Woo 等人[38]提出的 CBAM 则按照顺序的方式组合了通道注意力模块 CAM 和空间注意力模块 SAM。注意力机制在多种应用上获得了较好效果。这些注意力模型利用平均池化和最大池化或卷积网络将注意力集中在感兴趣的区域，但会损失特征信息，忽略结构特征。

4.2.2 小波变换

小波变换是一种有效的时空分析工具。离散小波变换具有计算高效和能够处理非平稳信号的优点，在图像处理和时间序列分析等领域得到广泛应用[62, 63]。一些研究者尝试将小波变换引入到深度网络中。例如，ScatNet[64]采用非线性模和平均池化的小波变换进行级联，提取具有平移不变性的特征，对图像分类任务具有很好的性能。Fujieda 等人[65]通过 DWT 提取特征，结合卷积神经网络特征进行图像分类。为解决最大池化方法可能导致的过拟合问题，研究者提出了小波池化方法[66, 67]，并证明了小波池化能够超越传统的平均池化和最大池化方法。

传统的深度学习和小波神经网络方法可以获得较高的分类精度。Choi 等人[68]提出了一个小波注意力嵌入网络，该模型通过对小波嵌入网络中的四个不同分量分别进行处理，提取增强的空间特征。Deeba 等人[69]提出了一种基于超分辨率 SR 的字典训练过程，以实现 CT 图像重建中的高分辨率重建。Li 等人[70]提出了各种小波的 DWT 和 IDWT 层，以

中国知网 https://www.cnki.net

增强网络的噪声鲁棒性。虽然近年来深度语义分割网络在目标检测中得到了广泛的应用，但不能充分利用图像的全局信息。为了解决这一问题，Huang 等人[71]提出了一种基于小波和注意机制的卷积神经网络新方法，称为 WA-CNN 算法。这些研究突出了深度学习与小波变换技术相结合的潜力，因此，卷积神经网络与小波相结合的研究是有效的。本章提出了使用离散小波变换、卷积神经网络和注意力机制相结合的方法在有限数据集下充分提取图像特征。

4.3 基于离散小波变换的楚系简帛古汉字识别方法

本章提出基于 ResNet 和 DWT 的楚系简帛文本识别方法 WA-Net，针对文字退化模糊、噪声干扰、残缺等情况导致传统卷积神经网络的文字识别准确率不高，提出了使用离散小波变换和卷积神经网络结合的方法，在图像处理中，离散小波变换能够提供很好的图像压缩和图像降噪能力，将图像转换为一组系数，能在不同的空间和频率上表示图像特征，可以有效减少噪声的干扰，以生成能充分揭示关键特征信息的特征图。

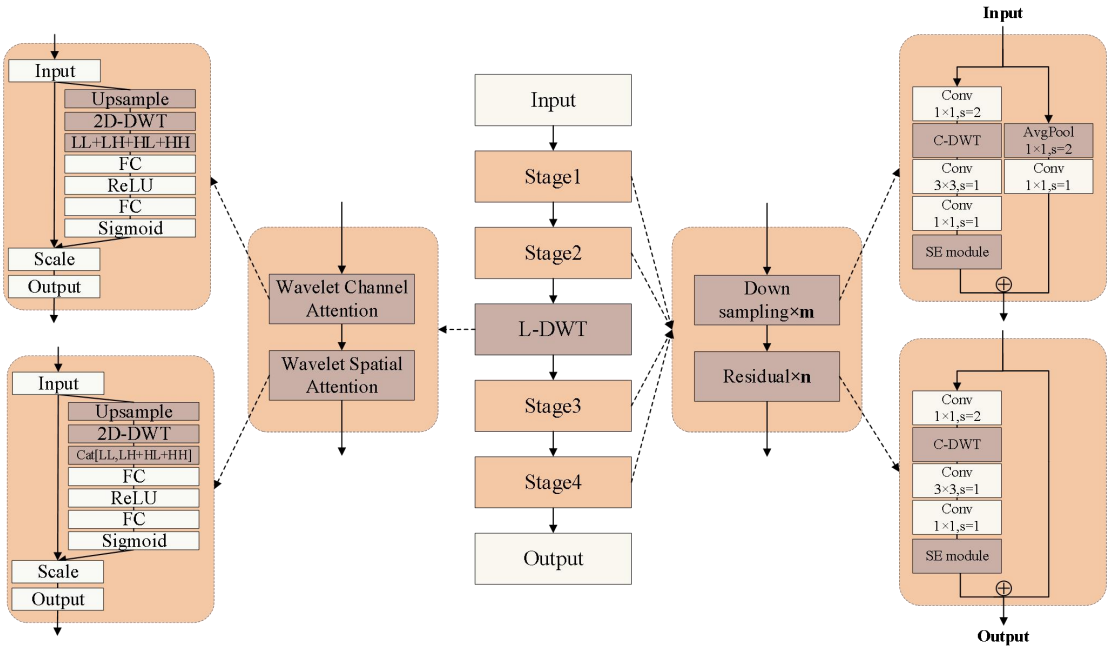

图 4-2 WA-Net 识别流程图

如图 4-2 所示，本章的 WA-Net 网络结构包含 7 个阶段，这里分别

中国知网 https://www.cnki.net

记为 Input、Stage1、Stage2、L-DWT、Stage3、Stage4 和 Output。识别流程如下：输入的文字图像经过数据增强，任取图片进行随机裁剪和缩放，进行卷积和最大池化操作，Stage1、2、3、4 使用改进后的残差块进行堆叠，L-DWT 层为离散小波通道注意和离散小波空间注意模块，经过多层的卷积操作，进行平均池化和分类输出识别结果。

表 4-1 WA-Net 网络结构

层名称	输入特征	操作	SE 注意力	输出通道数
Input	224×224×3	Conv2d	-	64
	112×112×64	MaxPool	-	64
Stage1	56×56×64	W-bneck	1	256
	56×56×256	W-bneck	1	256
	56×56×256	W-bneck	1	256
Stage2	56×56×256	W-bneck	1	512
	28×28×512	W-bneck	1	512
	28×28×512	W-bneck	1	512
	28×28×512	W-bneck	1	512
L-DWT	28×28×512	WC-Attention	-	512
	28×28×512	WS-Attention	-	512
Stage3	28×28×512	W-bneck	1	1024
	14×14×1024	W-bneck	1	1024
	14×14×1024	W-bneck	1	1024
	14×14×1024	W-bneck	1	1024
	14×14×1024	W-bneck	1	1024
	14×14×1024	W-bneck	1	1024
Stage4	14×14×1024	W-bneck	1	2048
	7×7×2048	W-bneck	1	2048
	7×7×2048	W-bneck	1	2048
Output	-	AvgPool	-	-
	2048	FC	-	730

中国知网 https://www.cnki.net

4.3.1 网络结构

在图像分类任务中，传统卷积神经网络由于网络层数的增加而可能导致梯度消失，使得模型的学习变得困难。而 ResNet[72]的残差块可以处理浅层特征，并进一步提取更多有用的特征信息，常被用于图像分类和识别任务中作为主要的特征提取结构。本章选用 ResNet50 来获取图像特征，网络结构如表 4-1 所示，与其他注意力模块相比，如 CBAM[38]反映了空间域中不同部分的权重，本章提出的包含 C-DWT 的 W-bneck 模块和两个小波注意力模块：WCA（Wavelet Channel Attention）和 WSA（Wavelet Spatial Attention）可以更好地表达不同分辨率下的频域权重[68, 73]。此外，在 W-bneck 块中引入了一些功能模块，如小波卷积[65, 74]，平均池化[75]和 SE 注意力[36]，从而能够重点关注文本图像的结构特征。WCA 将低频和高频统计特征作为通道标量，更好地表征不同通道的重要性，WSA 则能有效地聚焦关键的空间结构特征。

4.3.2 离散小波变换

二维离散小波变换是一种应用于二维图像数据的信号处理方法，可以将图像分解成不同频率子带，进而抽取不同分辨率和频率的子带下的信号特征，使得这些特征更加明显。在图像处理中，给定一幅图像 I，使用离散小波变换将图像 I 解为具有不同频率特征的低频子带 (LL)、垂直高频子带 (HL)、水平高频子带 (LH) 和对角高频子带 (HH)。低频子带包含了图像 I 的近似图像信息，该图像包含了原始图像 I 的大部分平滑信息和图像结构，也可以理解为对原始图像进行压缩得到的有损失的图像。垂直高频子带包含图像 I 的纵向的边缘特征、水平高频子带包含图像 I 的横向的边缘特征、对角高频子带包含图像 I 的对角特征。这些高频子带包含图像中的细节、边缘和纹理等关键信息，可以用于检测和提取图像的重要特征。由于高频子带的方向和频率不同，它们有助于有效地反映图像的多维特性。如图 4-3 所示：

中国知网 https://www.cnki.net

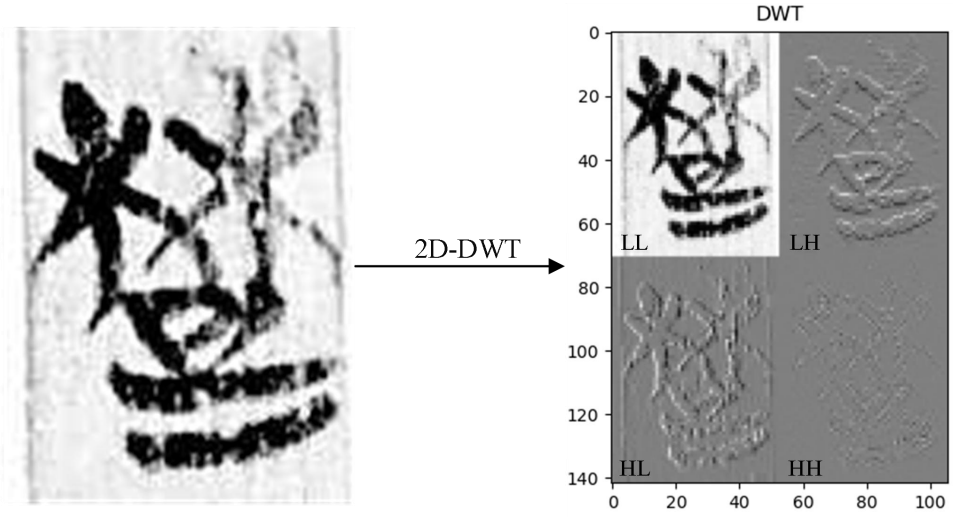

图 4-3 离散小波变换，LL 近似特征，LH，HL 结构特征，HH 纹理特征

对于图像 $f(x,y)(M\times N)$，包含尺度函数和平移基函数的离散小波变换可以定义为:

$$W_{\varphi}(j_0,m,n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\cdot\varphi_{j_0,m,n}(x,y) \tag{4-1}$$

$$W_{\psi}^{i}(j,m,n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\cdot\psi_{j,m,n}^{i}(x,y) \tag{4-2}$$

其中，$i=\{H,V,D\}$，$0\le x\le M-1,0\le y\le N-1$，$j_0$是一个任意的开始尺度，$W_{\varphi}(j_0,m,n)$ 系数定义 $f(x,y)$ 在尺度 j_0 处的近似。$W_{\psi}^{i}(j,m,n)$ 系数对尺度 $j\ge j_0$ 附加了水平、垂直和对角方向的细节。$\varphi_{i_0,m,n}(x,y)$ 是尺度函数，ψ^{*} 表示平移基函数，ψ^{H} 表示小波度量函数沿着列方向变化，ψ^{V} 表示小波度量函数沿着行方向变化，ψ^{D} 表示小波度量函数沿着对角线方向变化。图像 $f(x,y)(M\times N)$ 可通过离散小波逆变换得到公式 4-3：

$$\begin{aligned}f(x,y)=&\frac{1}{\sqrt{MN}}\sum_{m}\sum_{n}W_{\varphi}(j_0,m,n)\varphi_{j_0,m,n}(x,y)\\&+\frac{1}{\sqrt{MN}}\sum_{i=H,V,D}\sum_{j=j_0}^{\infty}\sum_{m}\sum_{n}W_{\psi}^{i}(j,m,n)\psi_{j,m,n}^{i}(x,y)\end{aligned} \tag{4-3}$$

正交小波基的尺度函数为：$\varphi(x,y)=\varphi(x)\varphi(y)$，当且仅当 $\varphi(x)\varphi(y)=1$时，由公式 4-1 得:

中国知网 https://www.cnki.net

$$\begin{aligned} W_{\varphi}(j_0,m,n) &= \frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\cdot\varphi_{j_0,m}(x)\varphi_{j_0,n}(y) \\ &= \frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y) \\ &= \frac{1}{\sqrt{MN}} SUM(f(x,y)) \\ &= \sqrt{MN}\,GAP(f(x,y)) \end{aligned} \tag{4-4}$$

其中$W_{\varphi}(j_0,m,n)$表示采用小波基的二维离散小波的低频子带。由此可见，低频子带特征系数之和与GAP成正比。

4.3.3 W-bneck 模块

卷积神经网络中残差模块的主要目的是提高模型的计算效率，通过减少模型参数来防止过度拟合，同时保证模型的准确性。该模块采用1×1卷积、3×3卷积和1×1卷积的顺序连接，以减少特征图的通道数，并在3×3卷积中使用了补零和批归一化来提高模型的稳定性。考虑到离散小波变换的高频子带可以捕获详细信息，低频子带保留基本结构，通过子带叠加可减少特征损失。本章在卷积过程中添加$C-DWT$模块提取图像详细特征信息，使用通道注意力模块提高有效特征信息占比的同时减少无效特征的数据信息，并在需要下采样的残差块中进行平均池化下采样替换卷积获取更多原始图像信息。结构如图 4-4 所示。

由于二维离散小波变换生成的特征子带会降维，为了保证大小为$H\times W$的特征图尺度大小不变，在$C-DWT$模块中添加上采样操作，通过二维离散小波变换生成和输入特征图相同大小的子带，通过子带叠加保证特征F'包含信息的完整性。如式 4-5 所示：

$$F' = \sum_{i=0}^{H}\sum_{j=0}^{W}(DWT_2D(UpSample(F))) \tag{4-5}$$

特征经过多层卷积后进行通道注意力获取关键信息，通过压缩操作，将空间维度为$H\times W$的特征图聚合成一个特征描述符，该描述符包含了通道维度上特征响应的全局分布情况。接下来，通过使用激活函数和基

中国知网 https://www.cnki.net

于通道相关性的门限机制来学习并控制每个通道的激活状态，并将重新

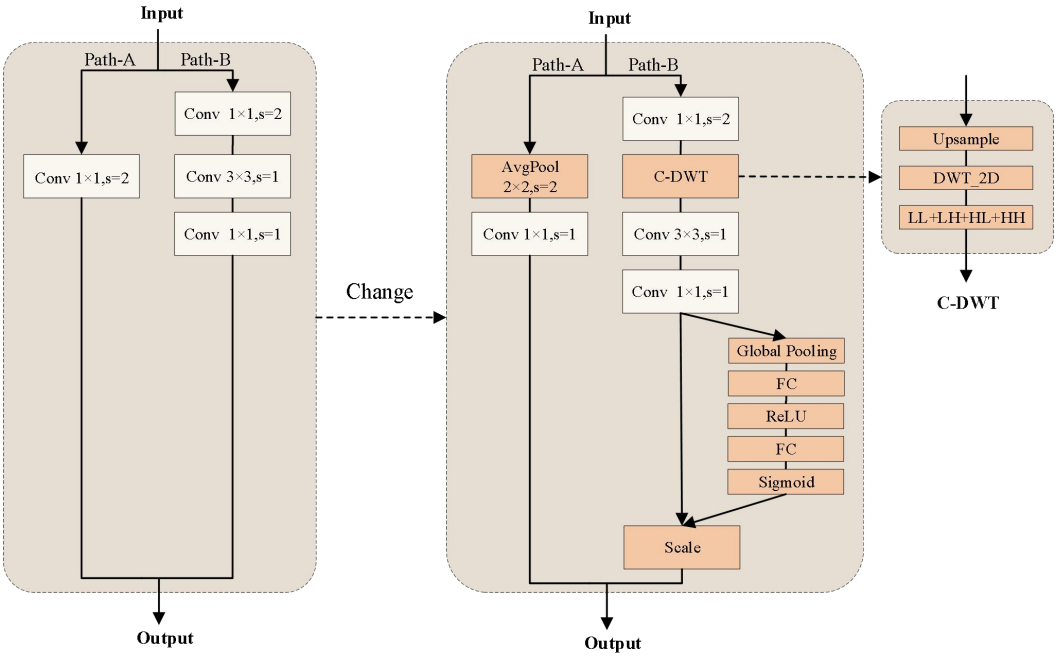

图 4-4（a）原始的残差块结构图（b）改进后的 W-bneck 模块

赋权重的特征图输出到下一层，在输出之前，这个过程会被分成压缩和激励两部分。通过使用通道注意力机制，可以更好地调整每个通道之间的激活状态，提高了网络对具有显著细节结构的重要特征的关注度，压缩无效特征信息，从而提高网络的性能。

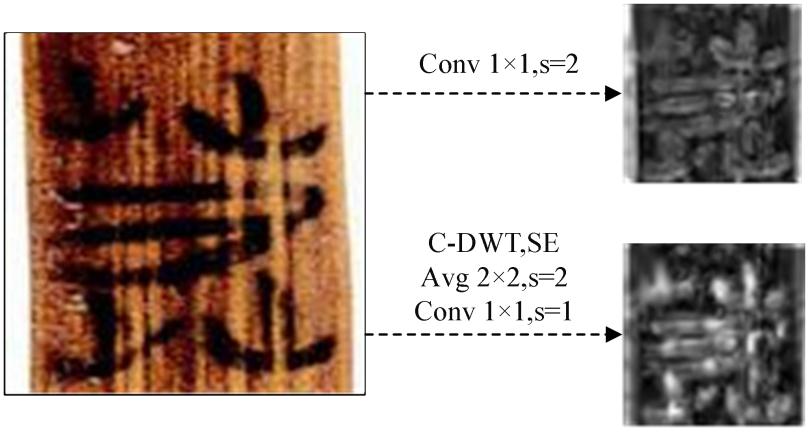

图 4-5（a）大小为 1×1，步长为 2 的卷积操作（b）大小为 2×2，步长为 2 的平均池化操作

中国知网 https://www.cnki.net

特征经过多层卷积后进行通道注意力获取关键信息，通过使用通道注意力机制，可以更好地调整每个通道之间的激活状态，提高了网络对具有显著细节结构的重要特征的关注度，压缩无效特征信息，从而提高网络的性能。在下采样操作中，原来使用的卷积核大小为1×1，步长为 2 的卷积忽略了 3/4 的输入特征图，如图 4-5（a）所示，为了减小信息损失，改进后在1×1的卷积之前添加了一个卷积核大小为2×2，步长为 2 的平均池化层，如图 4-5（b）所示，以保证获取更多的特征图信息。这一改进的细节，在图 4-6 中有所示。

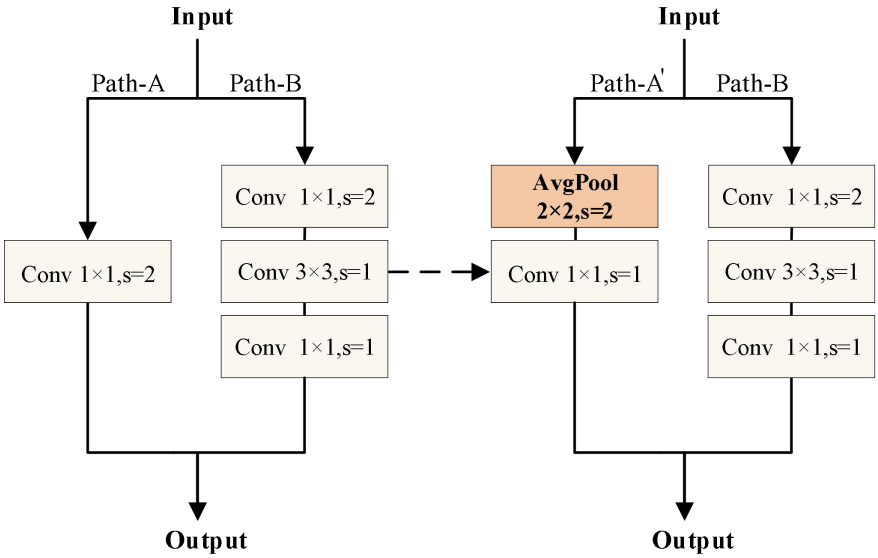

图 4-6 添加平均池化的残差块

4.3.4 小波注意力

二维离散小波将图像特征分解为低频子带和高频子带，分别存储图像的结构、细节信息和噪声。现有的注意力模型主要利用平均池化或最大池化来关注感兴趣区域，但存在特征信息丢失和结构特征忽略的问题。利用小波变换的压缩特性和结构特征提取能力，本章提出小波注意力，使用小波通道注意力可以自适应地调整高频信息的标量，以分配不同通道的较优权值，有效地利用低频和高频特征构建注意模型。使用小波空间注意力可以有效地关注关键的空间结构特征，有效地利用小波变换捕捉到的结构的近似和细节特征，提高分类的准确性。与传统的空间注意

中国知网 https://www.cnki.net

机制相比，小波空间注意机制在保持结构特征方面表现出上级性能。小波注意力结构如图 4-7 所示。

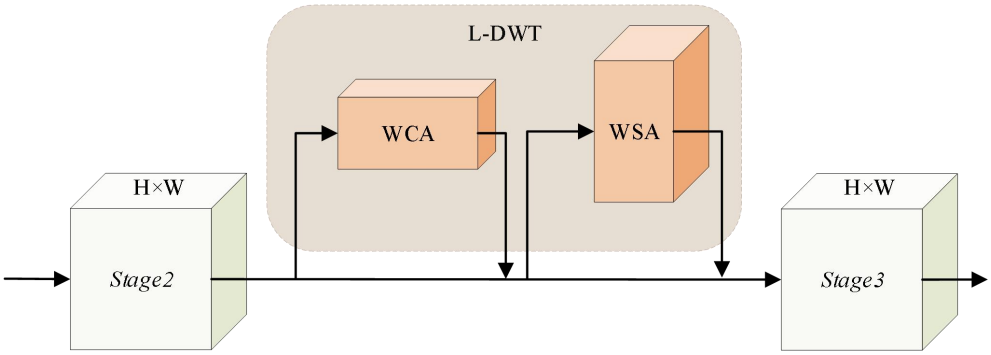

图 4-7 小波注意力，WC 为小波通道注意力，WS 为小波空间注意力

小波通道注意：小波通道注意模块主要使用 DWT 将特征分解为高频和低频特征子带。它利用高频和低频特征的统计特征作为通道标量，代替 SENet 中的平均池化操作，如图 4-8 所示。由于小波操作会使特征图降维，为了保证特征图处理后大小不变，采用双线性插值法进行上采样操作，该方法具有低通滤波性质，可以减弱图像高频分量。

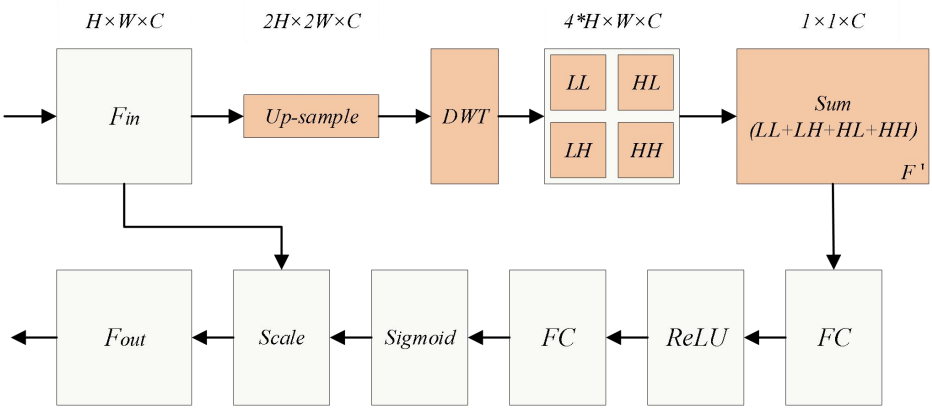

图 4-8 小波通道注意

大小为 $H\times W$ 特征 F_{in} 传入小波通道注意力，通过上采样操作得到 $2H\times 2W$ 的特征图，经过二维离散小波变换分解出四个 $H\times W$ 的特征子带，对四个子带求和来替换平均池化操作。通过两个全连接层 W_0,W_1 对向量 F'' 进行处理，得到通道权重值 S，经过两层全连接层后，S 中不同的数值表示不同通道的权重信息，从而赋予通道不同的权重。最后通过标量 S 与特征 F' 相乘。得到输出特征 F_{out}。其中 + 表示元素的求和。

中国知网 https://www.cnki.net

$$F_{out} = F_{in} sigmoid(W_1 \operatorname{Re} LU(\frac{W_0}{M \times N} \sum_{i=0}^{H} \sum_{j=0}^{W} (LL + LH + HL + HH) F^{'})) \quad (4\text{-}6)$$

小波空间注意：小波空间注意模块主要利用小波分解后的高频和低频特征子带具有提取图像空间特征的特性。通过小波分解得到图像的关键点和结构特征，使用低频子带和其他三个子带的叠加代替空间注意力中的平均池化和最大池化运算。与原空间注意方法相比，聚合低频和高频特征可以提供更多关键特征的特征图，如图 4-9 所示。

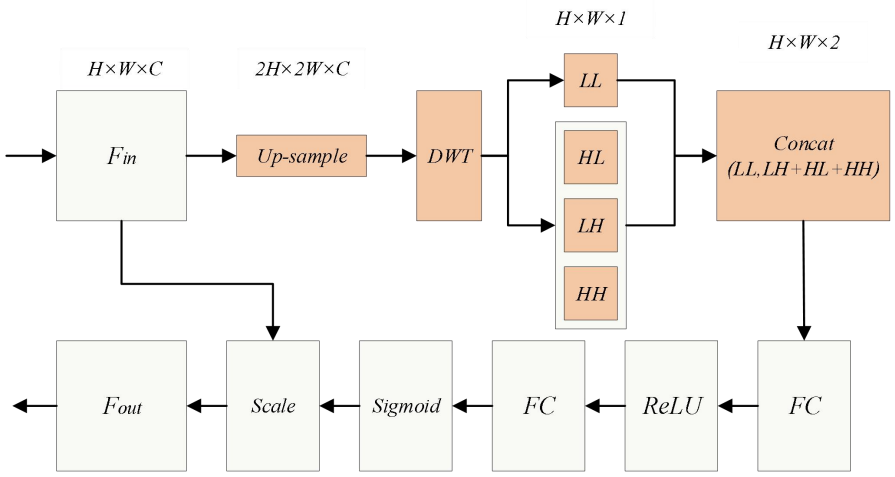

图 4-9 小波空间注意

将大小为$H \times W$的特征F_{in}传入小波空间注意力，通过上采样操作得到$2H \times 2W$的特征图。随后，经过二维离散小波变换分解出四个大小为$H \times W$的特征子带，将三个高频子带进行叠加求，并与低频子带特征拼接生成一个特征描述符。然后，利用通道数为 1 的卷积进行连接和卷积操作，以调整通道数。最后，使用 Sigmoid 激活函数为输入特征图的每个特征点分配权重值，将权重通过乘法逐通道加权到输入特征层，得到输出特征F_{out}。

$$F_{out} = F_{in} sigmoid(W_1 \operatorname{Re} LU(W_0 Cat(LL, \sum_{i=0}^{H} \sum_{j=0}^{W} (LH + HL + HH)))) \quad (4\text{-}7)$$

其中W_0, W_1表示两个全连接层。小波空间注意模块可以生成不同通道的注意图，可以更好地曲线化每个通道的空间注意，它可以在一定程度上保持比平均池操作和最大池操作更丰富的空间信息。小波通道注意、小波空间注意均可嵌入卷积神经网络的不同特征层。

中国知网 https://www.cnki.net

4.4 实验结果与分析

4.4.1 对比实验

为了评估不同网络的识别性能，将所提出的网络与其他经典卷积神经网络识别结果进行比较，如表 4-2 所示。ResNet50 在对比的网络中实现了最高的识别准确率，而 SqueezeNet 由于深度特征提取有限而表现出最低的准确率。传统的 CNN 模型具有较低的计算复杂度，但识别精度不足。与最先进的对象识别模型（如 Vision Transformer）相比，本章的方法将 Top-1 准确率提高了 4.92%。对网络的参数和计算要求进行了比较和分析。由于小波变换涉及的运算量大，该模型在参数和计算效率方面需要大量的资源。然而，实验结果表明，该模型的识别精度从 Top-1 到 Top-5 有显着的改善。

表 4-2 对比实验结果

网络结构	参数量/M	计算量/G	准确率/%				
			Top-1	Top-2	Top-3	Top-4	Top-5
SqueezeNet	1.10	2.95	67.49	75.72	79.57	81.76	83.55
ShuffleNet	1.09	0.05	82.30	87.99	90.09	91.53	92.35
RegNet	2.59	0.21	82.76	88.44	90.38	91.62	92.55
AlexNet	16.08	0.31	83.25	89.16	91.33	92.42	93.12
GoogleNet	12.55	1.58	86.09	90.91	92.55	93.44	94.04
Vgg16	137.25	15.47	86.51	91.23	92.91	93.84	94.60
MobileNet	3.16	0.33	86.51	91.33	92.94	93.93	94.60
DenseNet121	7.70	2.91	87.15	91.32	92.94	93.91	94.41
Swin transformer	87.49	15.47	87.18	91.28	92.94	93.69	94.24
Vision transformer	86.36	17.58	83.32	89.04	90.93	91.99	92.47
ResNext50	24.48	4.30	87.15	92.00	93.49	94.29	94.85
ResNet50	25.00	4.14	87.42	91.92	93.64	94.48	94.90
WA-Net	128.31	12.29	89.24	93.22	94.61	95.11	95.70

通过与经典 SE 注意力和 CBAM 注意力的对比实验，验证了小波注

中国知网 https://www.cnki.net

意力在网络中的有效性。实验结果表明，本章提出的基于小波的注意力方法比经典的注意力机制获得了略高的识别准确率，如表 4-3 所示。具体而言，当单独使用小波通道注意力或小波空间注意力时，识别精度仅略有提高。然而，当两种注意力同时应用时，Top-1 增加了 0.65%。此外，当与本章提出的其他方法结合使用时，Top-1 增加了 1.24%。

表 4-3 注意力对比实验结果

网络结构	准确率/%				
	Top-1	Top-2	Top-3	Top-4	Top-5
ResNet50	87.42	91.92	93.64	94.48	94.90
ResNet50+SE	88.02	92.29	93.74	94.68	95.13
ResNet50+WC	88.04	92.30	93.68	94.48	94.90
ResNet50+SAM	87.63	91.75	93.19	93.98	94.65
ResNet50+WS	87.67	91.99	93.52	94.56	95.05
ResNet50+SE+SAM	87.35	92.04	93.49	94.08	94.66
ResNet50+WC+WS	88.00	92.35	93.71	94.65	95.16
Our method	89.24	93.22	94.61	95.11	95.70

表 4-4 为本章提出模块的参数量和计算量。WCA 和 WSA 模块的参数数量相同；然而，由于 WSA 使用低频子带作为 GAP 和高频子带作为最大池化而引起更高的计算量。W-bneck 模块采用多次卷积运算，利用小波卷积特性有效抑制图像噪声。它结合平均池化和 SE 注意力进行全面精细的特征提取，导致模块参数和计算要求相对较高。

表 4-4 不同模块的参数量和计算量

模块	Param.(M)	Flops(G)
W-bneck	0.24552	9.47324
WCA	0.00001	0.00060
WSA	0.00001	0.00150

图 4-10 展示了 ResNet 识别错误而 WA-Net 识别正确的部分文字样本及其置信度，图中文字部分分别为“样本真实标签_ResNet 识别结果（置信度）_WA-Net 识别结果（置信度）”。可以看出，具有少量噪声

中国知网 https://www.cnki.net

的样本依然可以正确识别，具有较高的置信度，如“雀_隹（0.58）_雀（0.84）”，“文_癸（0.74）_文（0.99）”，“攸_逆（0.43）_攸（0.83）”。可以看出，类间差异性较大的文字图像也可以在一定程度上正确识别，如“八_行（0.58）_八（0.55）”和“八_月（0.53）_八（0.54）”。部分模糊或残缺的样本依然可以正确识别，如“又_不（0.19）_又（0.2）”，“左_右（0.56）_左（0.43）”，“子_之（0.4）_子（0.72）”，“曰_日（0.62）_曰（0.8）”。



图 4-10 部分识别正确图像

部分识别错误的文字样本及其置信度如图 4-11 所示，识别错误的样本具有以下特征：（1）严重的退化模糊、载体残缺的样本。由于这类样本图像质量较差，网络在样本中难以提取有效的特征信息，因此识别正确率较低，如“高_未（0.28）”，“南_高（0.32）”，“食_敢（0.29）”。（2）文字结构相似或差异较大的样本。由于古汉字缺乏统一规范，导致不同文字结构相似或同一文字差异较大的，这种差异主要源于不同书写者的字体风格不同，如“母_毋（0.92）”，“有_又（0.99）”等，容易造成分类错误。

中国知网 https://www.cnki.net

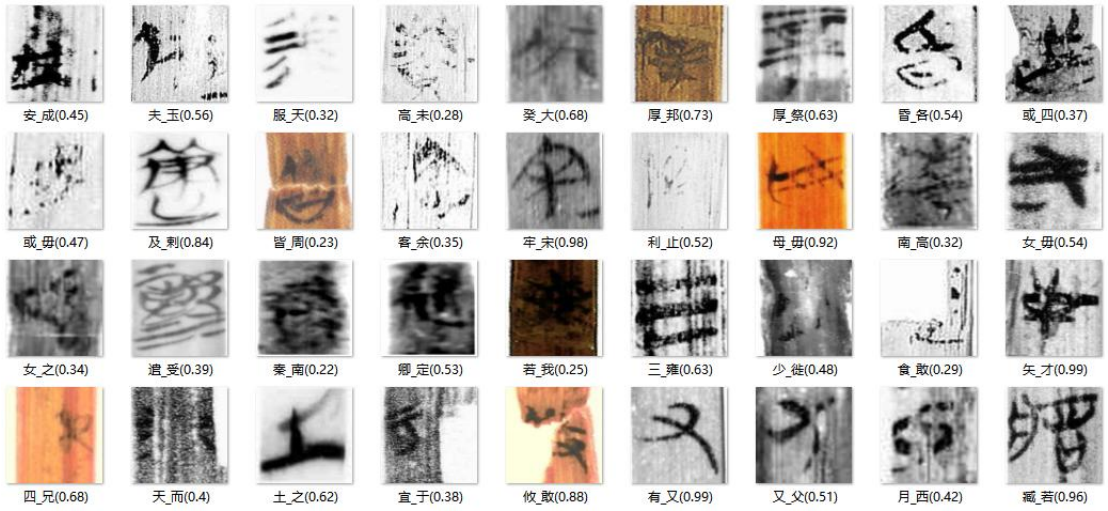

图 4-11 识别错误图像

本章提出的方法包含了小波包对特征图像的处理，在识别过程中使用离散小波变换得到更加精细的纹理特征描述。离散小波变换具有压缩和降噪能力，可以一定程度上降低图像在传输和保存过程中所带来的噪声。以上结果表明，本章提出的 WA-Net 实现了楚系简帛古汉字的识别。即使是具有高噪声和模糊的样本图像也能够被正确识别，但对于残缺和严重退化的图像则无法准确识别，也说明本章在前处理和特征提取方面还具有一定的局限性。

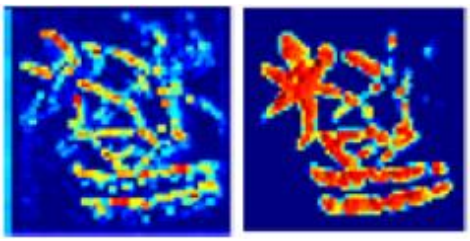

图 4-12 ResNet50 和 WA-Net 在 stage3 的特征可视化

为了证明本章提出的小波注意力网络有效性，选取阶段 3 的特征图进行可视化，并对该阶段的两个网络可视化特征图进行对比，如图 4-12 所示，ResNet50 可以提取文字主要特征，但是主要特征周围有很多区域也被关注，但是 WA-Net 降噪效果明显，更加集中于文字本身，忽略了文字之外的区域。此外，在图 4-13 中，对比了 ResNet50 和 WA-Net 的深层特征图，结果显示，本章提出的 WA-Net 模型对文本区域有更明显的突出显示。

中国知网 https://www.cnki.net

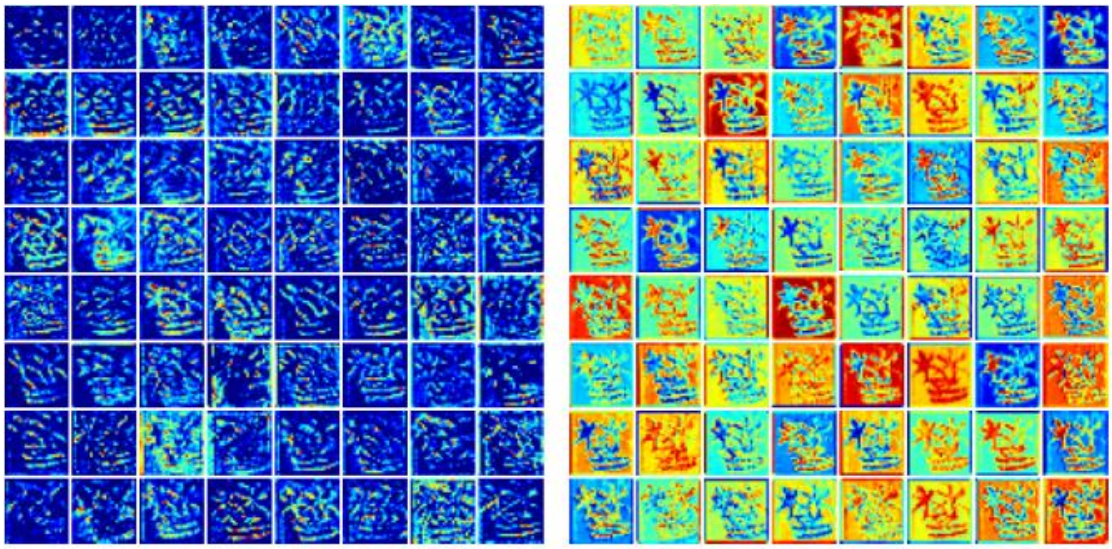

图 4-13 ResNet50 和 WA-Net 的特征可视化

4.4.2 消融实验

表 4-5 消融实验结果

模块				准确率/%				
SE	AvgPool	L-DWT	C-DWT	Top-1	Top-2	Top-3	Top-4	Top-5
-	-	-	-	87.42	91.92	93.64	94.48	94.90
✓	-	-	-	87.18	91.62	93.47	94.43	94.93
-	✓	-	-	87.47	92.00	93.57	94.50	95.21
-	-	✓	-	87.78	92.24	93.98	94.68	95.20
-	-	-	✓	88.86	92.74	94.08	94.90	95.40
-	✓	✓	-	87.13	92.19	93.74	94.5	95.03
✓	✓	-	-	87.27	92.25	93.79	94.43	94.86
-	-	✓	✓	87.33	91.85	93.68	94.75	95.13
-	✓	-	✓	87.63	91.99	93.73	94.63	95.11
✓	-	-	✓	87.82	91.90	93.57	94.19	94.68
✓	✓	-	-	88.67	93.01	94.06	95.03	95.52
✓	-	✓	✓	87.27	92.05	93.59	94.50	95.18
-	✓	✓	✓	87.97	92.34	93.73	94.33	94.98
✓	✓	✓	-	88.89	92.67	94.01	94.88	95.55
✓	✓	-	✓	89.27	93.12	94.23	94.96	95.40
✓	✓	✓	✓	89.24	93.22	94.61	95.11	95.70

为了证明离散小波对特征提取的有效性，分析各个模块对结果的影响，本章在 ResNet50 网络上进行了消融实验，结果如表 4-5 所示。在瓶颈模块增加通道注意力模块时，Top-1 到 Top-5 有轻微下降，其他三个

中国知网 https://www.cnki.net

模块均对准确率提升有所帮助，尤其是单独使用 C-DWT 模块时，可提高 1.44%。通过实验发现，当 L-DWT 和 SE 注意力同时使用，可在 L-DWT 提升基础上再提升 0.9%，故 L-DWT 与 SE 注意力配合使用效果较好。当平均池化和 C-DWT 模块一起使用时，准确率有提升，当 SE 注意力和平均池化以及 C-DWT 模块一起使用时，准确率提升了 1.64%，故结合了 SE 注意力和平均池化和 C-DWT 模块的模型识别效果显著，当四个模块同时使用时，Top-1 相比于不使用 L-DWT 模块的方法有 0.03%的降低，但是 Top-2 到 Top-5 均有提升。

4.5 本章小结

本章提出了一种针对楚系简帛古汉字智能识别的方法—小波注意力网络（WA-Net）。利用小波变换的多尺度特征提取能力，结合注意机制捕获更多有区别的字符结构和纹理特征，解决了在低质量和结构相似的图像中识别精度低的挑战。本章提出的 L-DWT 和 C-DWT 模块将神经网络的学习空间扩展到频域，并通过结合低频和高频子带来减轻文本纹理和结构的信息丢失，而不是采用传统注意机制中的平均池化或最大池化。实验结果显示，小波注意力网络在楚系简帛古汉字数据集上表现出较高的识别准确率，尤其在处理退化和模糊的文本图像时表现突出。

中国知网 https://www.cnki.net

第 5 章 基于长尾分布的楚系简帛古汉字的特征平衡损失

5.1 研究背景

作为春秋战国时期楚国所使用的文字，楚系简帛古汉字的样本非常珍贵和稀少，由于文本图像样本难以获得，收集的楚系简帛古汉字数据集 Chu730 存在严重的类别不均衡问题。本章在图 5-1 中展示了本文数据集的样本频率分布图。Chu730 最大类别的文字图像有 1758 个样本，与此形成鲜明对比的是，其他较多的尾部类别只有一两个样本。因此，在这类数据集上训练的卷积神经网络往往表现出较差的性能，特别是在识别尾部样本方面。

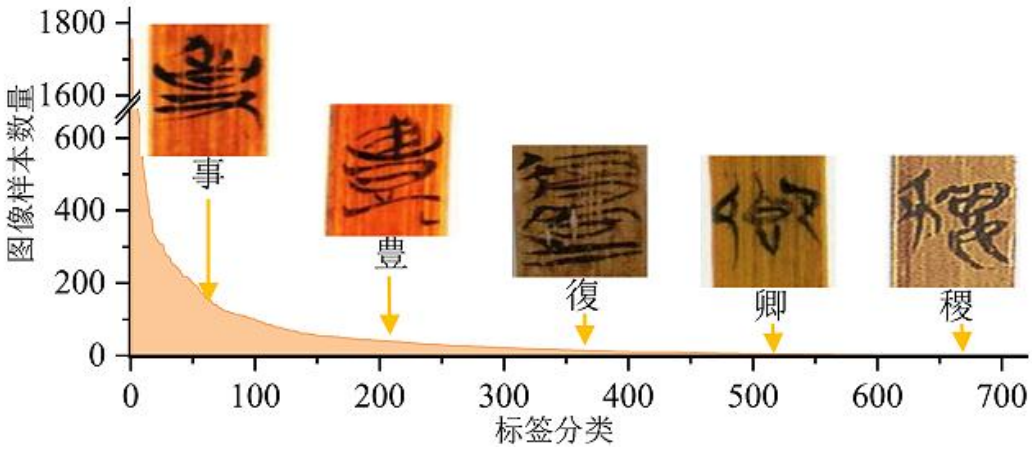

图 5-1 Chu730 数据集类别排序

为了解决数据分布不均衡所带来的问题，在深度学习中有各种各样的建议[76-79]。其中，基于混合的方法由于其简单而有效的性质，近年来得到了应用[77, 80, 81]。Zhang 等人[81]首先提出了混合方法，通过对原始输入样本及其对应的标签进行线性插值，直接生成样本。随后，有研究者开发了 Manifold Mix-up[80]，在特征空间中对随机选取的层进行插值。最近，Remix[78]被设计用于分离特征和标签之间的混合权重，进一步提高了非平衡数据集的分类性能。

基于上述研究，本章提出了一种新的损失函数，同时在图像层面和标签两个层面进行改进。首先，在图像层面，图像与图像之间存在类间

中国知网 https://www.cnki.net

相似性和类内差异性，不同类别文本图像的分类难易程度不同，为此，本章提出了一种困难样本重加权损失，来提升困难样本的分类准确性，其次，在类别层面上，由于收集的数据集存在严重的长尾分布问题，本章采用重加权的方式，对尾部标签施加更高的权重，来缓解这一问题。在 Chu730 数据集上的实验结果表明，本章提出的方法在尾部样本识别准确率的提升上是有效的。

5.2 相关工作

5.2.1 长尾分布

由于在现实世界应用中不平衡数据的普遍存在，长尾识别已经吸引了越来越多的关注[82-86]。解决长尾分布问题最直观的方法是通过类平衡采样[76, 87]或损失函数重加权[88-92]来重新平衡模型。研究人员提出的重采样方法可以实现更均衡的数据分布，这些方法包括对少数群体过采样[93, 94]，对大多数类进行欠采样[95]，以及根据每个类别的样本数量类平衡抽样[84, 96]。然而，这些方法会导致对尾部类的过度拟合，这必然会抑制模型的识别性能。最近，Cui 等人[92]提出了根据每一类的“有效数”对损失函数进行重新加权或对数据进行重新抽样的方法，这一方法已被证明对长尾分布数据集是有效的。另一方面，这种“有效数”策略并没有真正解决长尾数据特征分布[76, 82, 97, 98]不均匀的问题。随后，Cao 等人[89]利用标签分布感知边缘（label-distribution-aware margin，LDAM）对损失进行重新加权，可以提高尾类的泛化性能。

5.2.2 损失函数

损失函数会影响神经网络模型的泛化能力以及其他特性，损失函数的目的是通过调整损失值来重新平衡不同类别的权重值。它从两个角度解决了阶级不平衡的问题：样本和阶级。基于样本的方法[8,9]通过损失中的细粒度参数为困难样本分配较大的相对权重。例如，Focal loss[90]利用样品预测硬度作为损失函数的重加权系数。然而，一个样本的分类难度

中国知网 https://www.cnki.net

可能与其对应的类别大小没有直接关系。因此，抽样方法无法处理大规模且严重不平衡的数据。逐类方法[4,5,10]用与标签频率负相关的特定类的参数来分配损失函数。

5.3 基于长尾分布的楚系简帛古汉字识别方法

针对楚系简帛古汉字图像分布不均衡和部分图像类内差距较大的特点，本章提出了一个混合损失函数来提高传统的卷积神经网络对文字识别的准确率。整体结构如图 5-2 所示。主要包含四个部分，第一部分是骨干网络用于提取特征，第二部分为困难样本重加权 F_1，第三部分为权重重平衡 F_2，第四部分为网络输出。

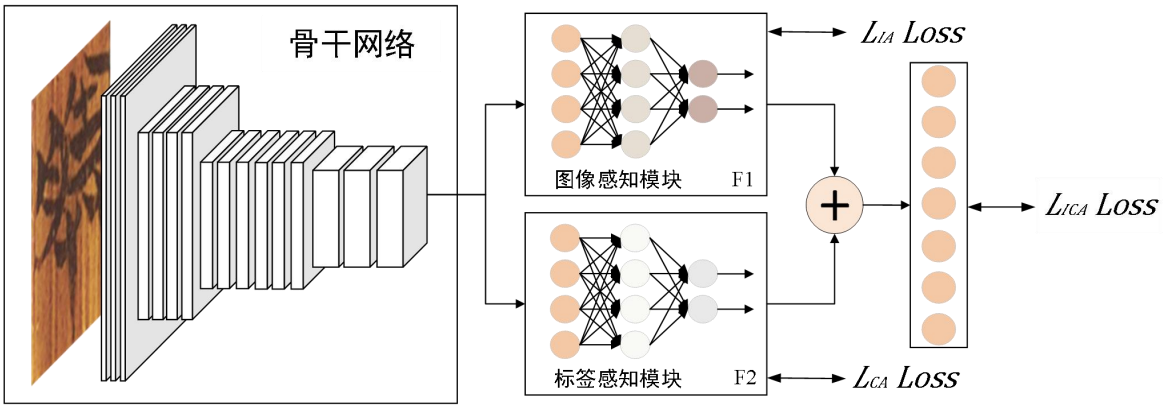

图 5-2 识别流程图

5.3.1 交叉熵损失函数

对于一个输入样本 $x_i \in R^{B\times C\times H\times W}$，其中 B 是输出的批次大小，C 是输出的通道数大小，H 和 W 分别是输出的宽和高，经过骨干网路和 Softmax 后，本章可以得到预测输出 p_i，使用交叉熵损失（L_{CE}）来度量模型预测值与真实值之间的差异。交叉熵损失函数的定义如公式 5-1：

$$L_{CE}(y_i, p_i) = -(y_i \log p_i + (1-y_i)\log(1-p_i)) \tag{5-1}$$

其中 p_i 表示模型对输入样本 x_i 的预测概率，y_i 表示输入样本 x_i 的实际标签。然而在古文本图像分类任务中，图像与图像之间，类别与类别之间的识别难易程度存在差异性。具体的，数据集中包含训练样本较少的尾

中国知网 https://www.cnki.net

部标签，以及难以准确识别的样本阻碍了高精度的古文本图像识别的发展。为了解决这一问题，本章提出困难样本加权损失和尾部样本重加权损失来解决长尾分布的难题以及提升困难样本的识别精度。

5.3.2 困难样本重加权损失

考虑到同一标签内，不同图像之间存在类内差异性，即同一文本图像文字差别较大，识别难度也不同，为了解决这一问题，本章提出困难样本加权损失，通过给难以识别样本更大的权重，提升难以识别样本的分类精度。

从图 5-2 中可以看出，图像经过骨干网络以及图像感知模块F_1后，得到输出$p_{hard_i} \in R^{B\times C'\times 1}$，$C'$表示数据集中标签的数量，则本章采用公式 5-2：

$$W_{hard_i} = \frac{1 - p_{hard_i}}{p_{hard_i} + \alpha} \tag{5-2}$$

判断图像x_i是否为困难分类样本。公式 5-2 中p_{hard_i}表示标签预测正确的概率，α为防止分母为 0 设置的超参数，即当p_{hard_i}为 0 时表示这一样本难以分类准确，p_{hard_i}为 1 时表示样本较易分类准确。本章将这一比值作为权重，与p_{hard_i}相乘求取困难样本重加权损失L_{IA}，如公式 5-3：

$$\begin{aligned} p'_{hard_i} &= W_{hard_i} \times p_{hard_i} \\ L_{IA} &= \sum_i L_{CE}(y_i, p'_{hard_i}) \end{aligned} \tag{5-3}$$

5.3.3 权重重平衡损失

为了平衡尾部样本数量较少，本章提高尾部样本的权重来提升样本分类的准确性，本章根据数据集分布特点取 $class_{hm}$，$class_{mt}$，将数据集分为头部 head，中部 medium，尾部 tail 三部分，并根据不同分类进行不同采样操作。重新平衡的权重的计算方式如式 5-4 所示：

$$W_{rebalce_i} = \frac{\max(num_i) - num_i}{\max(num_i)} + 1 \tag{5-4}$$

其中，$W_{rebalce_i}$表示由图像数量计算调整后的权重，num_i表示标签i在训

中国知网 https://www.cnki.net

练集中的图像数量，$\max(num_i)$ 表示训练集中拥有最多的图像的数量。图像经过骨干网络以及标签感知模块 F_2 后，得到输出 $p_{rebalce_i} \in R^{B\times C'\times 1}$，则本章应当长尾分布所采用的权重重平衡损失 L_{CA} 定义为公式 5-5：

$$L_{CA} = \sum_i (W_{rebalce_i} \times L_{CE}(y_i, p_{rebalce_i})) \tag{5-5}$$

5.3.4 图像标签损失

通过上述两个分支得到 p_{hard_i} 和 $p_{rebalce_i}$，本章的最终网络的最终输出 p_i 可以表示为公式 5-6：

$$p_i = \lambda_1 \times p_{hard_i} + \lambda_2 \times p_{rebalce_i} \tag{5-6}$$

其中 λ_1, λ_2 为网络训练过程生成的超参数。本章采用交叉熵损失来计算最终输出与真实值之间的误差，损失如公式 5-7：

$$L_{CE} = \sum_i L_{CE}(y_i, p_i) \tag{5-7}$$

本章提出的总损失 L_{ICA} 定义为公式 5-8：

$$L_{ICA} = L_{IA} + L_{CA} + L_{CE} \tag{5-8}$$

5.4 实验结果

5.4.1 对比实验

为了评估本章所提方法在不同网络的识别性能，在相同训练和测试条件下，将本章提出的新损失函数应用在各个卷积神经网络网络上与只使用交叉熵损失函数的卷积神经网络识别结果进行对比，结果如表 5-1 所示。可以看出，相较于交叉熵损失函数，本章所提出的损失函数在多个网络上的识别结果均有提升，其中在 AlexNet 上使用的时候，Top-1 提升了 0.67%，在 FrostNet 上使用的时候，Top-1 提升了 0.66%，在 ResNet50 上使用的时候，Top-1 提升了 0.82%，在 ResNet34 上效果最好，Top-1 提升了 1.66%。

表 5-2 为交叉熵损失函数和焦点损失函数以及本章改进的损失函数在 ResNet 网络上的对比实验，在文字识别方面，使用 Focal Loss 的网络识别结果比使用交叉熵损失函数的网络识别结果略低，本章提出的损失

中国知网 https://www.cnki.net

函数在 ResNet34 网络的 Top-1 上比 Focal Loss 提升 2.02%，比交叉熵损失函数提升了 1.66%。在 ResNet50 网络的 Top-1 上比焦点损失函数提升 1.64%，比交叉熵损失函数提升了 0.82%，且在 Top-2 到 Top-5 也均有提升。

表 5-1 对比实验结果

网络结构	准确率/%				
	Top-1	Top-2	Top-3	Top-4	Top-5
AlexNet	83.40	88.69	90.71	91.80	92.65
AlexNet+ours	84.07	89.39	91.23	92.52	93.17
FrostNet	86.56	91.27	93.02	94.08	94.70
FrostNet+ours	87.22	91.53	93.07	93.71	94.43
ResNet34	87.78	91.77	93.29	94.16	94.75
ResNet34+ours	89.44	93.22	94.60	95.33	95.77
ResNet50	87.63	91.95	93.61	94.43	94.90
ResNet50+ours	88.45	92.59	93.93	94.55	95.06

表 5-2 注意力对比

网络结构	准确率/%				
	Top-1	Top-2	Top-3	Top-4	Top-5
ResNet34+Focal Loss	87.42	92.35	93.83	94.63	95.10
ResNet34+Cross Entropy Loss	87.78	91.77	93.29	94.16	94.75
ResNet34+Ours Loss	89.44	93.22	94.60	95.33	95.77
ResNet50+Focal Loss	86.81	91.87	93.59	94.46	94.98
ResNet50+Cross Entropy Loss	87.63	91.95	93.61	94.43	94.90
ResNet50+Ours Loss	88.45	92.59	93.93	94.55	95.06

为了评价本章提出的损失函数在头部、中部和尾部三个由 Chu730 数据集拆分的三个子数据集上的识别效果，在这三个子数据集上对本章提出的方法和其他卷积神经网络方法进行对比实验。从表 5-3 可以看出，本章提出的方法在 ResNet34 上使用时，尾部样本的识别正确率提升最高，Top-1 提升了 2.62%，Top-3 提升了 2.05%，在 FrostNet 上时，尾部

中国知网 https://www.cnki.net

样本没有明显提升，但是中部样本 Top-1 提升最大，为 2.93%，在其他网络上的三个部分均有提升，尾部子数据集在 ResNet50 上，Top-1 提升了 1.09%，且头部子数据集和中部子数据也均有提升。

表 5-3 对比实验结果

网络结构	准确率/%								
	Head			Medium			Tail		
	Top-1	Top-3	Top-5	Top-1	Top-3	Top-5	Top-1	Top-3	Top-5
AlexNet	90.58	95.50	96.65	82.69	90.57	92.31	71.02	82.26	85.74
AlexNet+ours	91.22	96.01	97.09	83.52	91.21	92.86	71.65	82.72	86.37
FrostNet	92.59	96.97	97.70	85.71	93.77	95.42	76.33	85.51	88.88
FrostNet+ours	92.81	97.03	97.61	88.65	94.41	95.42	76.33	85.17	88.14
ResNet34	92.94	96.49	97.35	88.55	94.69	95.60	78.10	86.71	89.56
ResNet34+ours	93.90	97.35	97.73	90.66	96.15	96.80	80.72	88.71	91.61
ResNet50	92.43	96.65	97.38	88.74	94.32	95.70	77.75	86.82	89.33
ResNet50+ours	93.49	97.32	97.70	89.47	95.15	96.43	78.84	87.11	89.50

此外，本章提供了网络各层的热力图，用以展示网络在汉字特征提取方面的表现，如图 5-3 所示。与基准网络相比，本文方法在特征提取方面表现更佳，尤其在提取浅层特征和深层特征时更加明显。这说明了本方法对于不同复杂度的古汉字具有更强的特征提取和区分能力，从而有效地提高了古汉字识别的准确率。

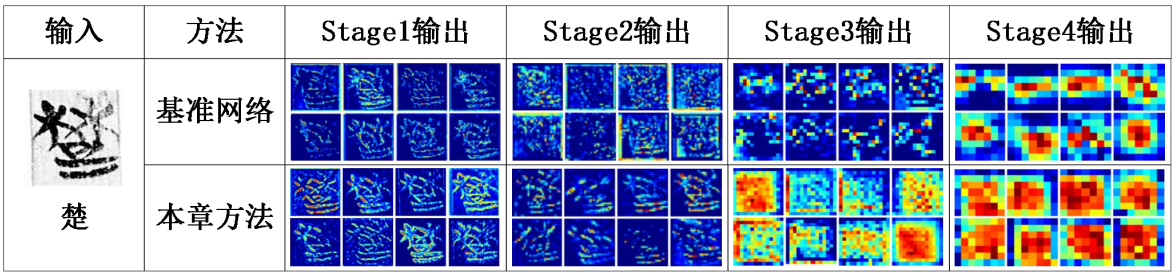

图 5-3 各层热力图对比

5.4.2 消融实验

为了验证所提出的损失函数用于特征提取的有效性，在识别效果最

中国知网 https://www.cnki.net

好的 ResNet34 上进行了消融实验，实验结果如表 5-4 所示。当 ResNet34 与损失函数 L_{IA} 结合使用的时候，Top-1 精度提高了 0.86%。当 ResNet34 与损失函数 L_{CA} 结合使用的时候，Top-1 精度提高了 1.15%。最后，当两者同时结合时，Top-1 到 Top-5 的识别准确率均有明显的提升，其中 Top-1 的精度提高了 1.66%，实验结果表明了本章所提出模块的有效性。

表 5-4 消融实验结果

模块	准确率/%				
	Top-1	Top-2	Top-3	Top-4	Top-5
ResNet34	87.78	91.77	93.29	94.16	94.75
ResNet34+L_{IA}	88.64	92.59	94.01	94.75	95.30
ResNet34+L_{CA}	88.93	92.76	94.21	95.05	95.58
ResNet34+L_{ICA}	89.44	93.22	94.60	95.33	95.77

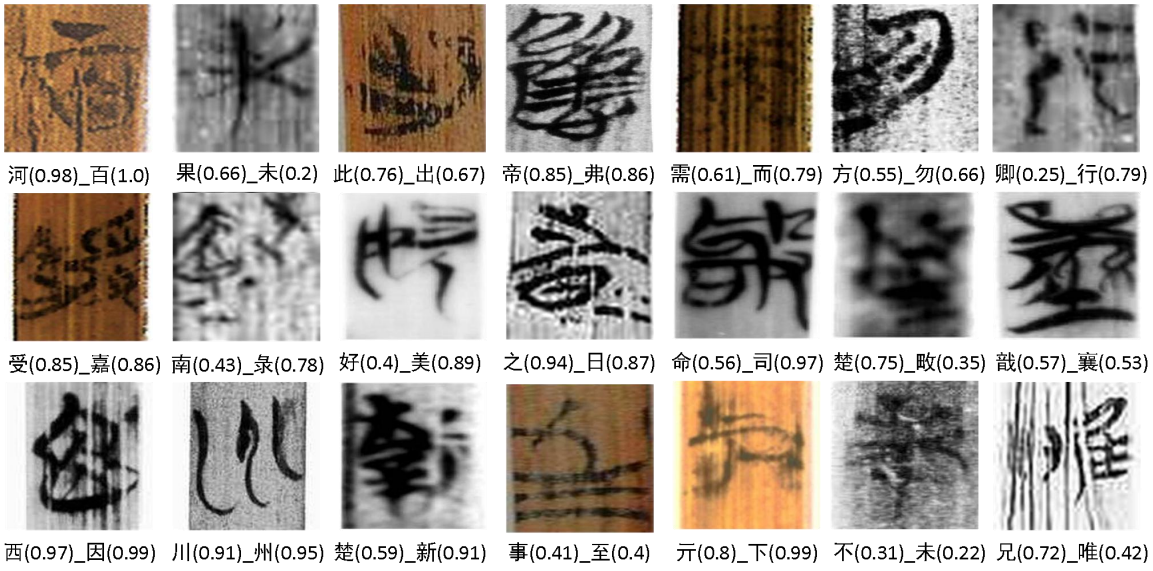

图 5-4 识别正确图像

图 5-4 显示了部分骨干网络识别错误，本章方法识别正确的图像，图像标注为识别错误文字（置信度）_正确文字（置信度），改进后的损失函数在 Chu730 上表现出显著的增强。

5.5 本章小结

本章为了应对真实场景下得楚系简帛古汉字分类中存在的长尾分布

中国知网 https://www.cnki.net

问题，提出了一种混合损失函数，该函数从图像和标签两个层面考虑。（1）困难样本重加权，通过在图像层面考虑分类问题，对难以识别准确的样本施加更高的权重。（2）重新平衡权重，通过在标签层面考虑数据集的长尾分布问题，对尾部样本重新调整权重。使用本章的新损失函数训练的模型比现有方法取得了显着的性能提升。

中国知网 https://www.cnki.net

中国知网 https://www.cnki.net

第 6 章 总结与展望

6.1 研究总结

汉字是世界上使用最广泛的语言之一，古汉字的智能识别对于文化遗产数字化保护应用具有极高的价值。相较于英文文本，汉字具有较大的字符数量，并且由于汉字笔划的组合形式，文本具有较强的类内差异性和类间相似性等特点。在实际研究中，手写汉字的识别难度相对于打印文本的识别更高。这主要归因于不同人的书写风格差异、古汉字文本的异体字和变形等因素，都增加了识别的复杂度。本文基于真实场景下的楚系简帛古汉字数据集，分别从数据层面与算法层面对古汉字文本识别任务进行了深入研究。在数据层面，采取了数据增强和重加权得方法来解决数据量不足和样本类别分布不均匀等问题。在算法层面，采用传统识别方法常用的小波变换和深度学习方法相结合的方法解决了强噪声高残缺的楚系简帛古汉字的识别问题。本文以卷积神经网络为基础工具，针对较少研究的楚系简帛古汉字识别问题展开了研究。主要包括数据集创建、基准方法选取、改进方法创新等方面。研究内容及成果如下：

（1）楚系简帛古汉字因其使用区域集中在当今湖北一带而保存较少，是出土量有限的古汉字之一。由于单一文字数据量较小，且文字的使用频率不同，导致文字样本的类别间存在较大差异。目前暂无公开的用于计算机识别的楚系简帛古汉字数据集。为此，通过导师的指导帮助，整理出了一个用于深度学习得楚系简帛古汉字数据集—Chu730，此数据集的详细分析已经在第三章中展开介绍。

（2）基于前文建立的数据集，设计并实现了基于离散小波变换的楚系简帛古汉字识别方法，使其适用于楚系简帛古汉字图像识别任务。同时为进一步提高网络的识别的精度，将经典卷积神经网络与基于小波变换的卷积神经网络进行识别效果对比，证明本文提出网络的识别正确率。随后针对本文提出的注意力模块在楚系简帛古汉字数据集上进行了消融实验，证明了本文提出的模块的对于楚系简帛古汉字图像识别的有效性。

中国知网 https://www.cnki.net

（3）基于前文建立的数据集，通过分析数据集，发现楚系简帛古汉字数据呈现出长尾分布的特点，本文设计并提出了一种针对长尾分布的解决方法。具体来说，从图像以及标签两个层面考虑。首先在图像层面考虑分类问题，对难以识别准确的样本施加更高的权重；其次重新平衡权重，通过在标签层面考虑数据集的长尾分布问题，对尾部样本重新调整权重。使用本文的新损失函数训练的模型比现有方法取得了显着的性能提升。

6.2 研究展望

本文针对楚系简帛古汉字的智能识别问题展开研究，通过分析任务的特点，整理了用于智能识别的楚系简帛古汉字数据集 Chu730，为后续研究奠定了基础。在此基础上，提出了两个识别方法进行汉字的智能识别，分别是基于小波变换的多模块结合方法与基于长尾分布的感知损失函数方法。与其他目标识别任务相比，对楚系简帛古汉字的直接研究相对较少，还有极大的研究空间和很多需要解决的问题。主要包括以下方面：

（1） 本文提出一个新的损失函数和一个基于小波变换的识别方法，其中最高的 Top-1 识别准确率达到 89.44%，但仍存在提升识别准确率的空间，且不同类别字符的识别准确率存在较大差异，这表明了当前方法在处理不同类别字符上存在一定的局限性，识别方法和模型有进一步研究的空间。

（2） 数据集的规模和样本的质量对最终结果和实际应用价值有着直接的影响。同时由于简帛年代久远，字体形态的改变将导致简帛古汉字中还存在着一些古文字字形与现代汉字不匹配的问题，给数字集的标注工作带来了相当的难度。一个高质量的数据集不仅能够提高研究成果的可信度，还能够为古汉字的数字化研究和实际应用提供有力支持，建立一个包含更多文字数量和类别的数据集，对深入学习解决楚系简帛中古汉字的研究问题具有意义。

（3） 在实际应用中，文字类别种类繁多，可能会出现部分文字未

中国知网 https://www.cnki.net

经过充分训练而导致识别错误的情况。在未来工作共，考虑通过在训练过程中引入聚类操作，并设定相关的阈值。当某些文字超过其类别中心的范围时，可以判定其为其他类别。另外，也可以设置一个额外的“待定”类别，用于归类所有未经过训练的文字，或者通过无监督学习的方法，由网络自动对其进行分类。

（4） 文本识别应分为文本检测和文本识别两个方面，本文的研究仅是单字符级别的古汉字文本识别，文字检测由手工标注完成。未来的研究可以探索实现简帛图像中文字的自动检测定位，以减少人工工作量。随着深度学习技术的发展，结合检测和识别的端到端方法有望实现对大篇幅古文档、书法作品等的自动检测与识别，推动史文资料的数字化应用。

中国知网 https://www.cnki.net

中国知网 https://www.cnki.net

参考文献

[1] Alashhab S, Gallego A J, Lozano M Á. Efficient gesture recognition for the assistance of visually impaired people using multi-head neural networks[J]. Engineering Applications of Artificial Intelligence, 2022, 114: 105188.

[2] Yang K, Yi J, Chen A, et al. ConvPatchTrans: A script identification network with global and local semantics deeply integrated[J]. Engineering Applications of Artificial Intelligence, 2022, 113: 104916.

[3] Coquenet D, Chatelain C, Paquet T. End-to-end handwritten paragraph text recognition using a vertical attention network[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(1): 508-524.

[4] Okarma K, Lech P. A method supporting fault-tolerant optical text recognition from video sequences recorded with handheld cameras[J]. Engineering Applications of Artificial Intelligence, 2023, 123: 106330.

[5] 王凯丽. 多场景古汉字识别方法研究[D]. 武汉大学, 2020.

[6] Gupta N, Jalal A S. Traditional to transfer learning progression on scene text detection and recognition: a survey[J]. Artificial Intelligence Review, 2022, 55(4): 3457-3502.

[7] Peng L, Liu C, Ding X, et al. Multi-font printed Mongolian document recognition system[J]. International Journal on Document Analysis and Recognition, 2010, 13: 93-106.

[8] Zhou X D, Wang D H, Tian F, et al. Handwritten Chinese/Japanese text recognition using semi-Markov conditional random fields[J]. IEEE transactions on pattern analysis and machine intelligence, 2013, 35(10): 2413-2426.

[9] Chu C, Nakazawa T, Kawahara D, et al. Chinese-Japanese machine translation exploiting Chinese characters[J]. ACM Transactions on Asian Language Information Processing, 2013, 12(4): 1-25.

中国知网 https://www.cnki.net

[10] Chen L. Research and application of chinese calligraphy character recognition algorithm based on image analysis[C]//2021 IEEE International Conference on Advances in Electrical Engineering and Computer Applications. IEEE, 2021: 405-410.

[11] Yalin M, Li L, Yichun J, et al. Research on denoising method of chinese ancient character image based on chinese character writing standard model [J]. Scientific Reports, 2022, 12(1): 19795.

[12] Li Z, Wu Q, Xiao Y, et al. Deep matching network for handwritten Chinese character recognition[J]. Pattern Recognition, 2020, 107: 107471.

[13] Huang M, Liu Y, Peng Z, et al. Swintextspotter: Scene text spotting via better synergy between text detection and text recognition[C]//proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 4593-4603.

[14] Bhunia A K, Sain A, Kumar A, et al. Joint visual semantic reasoning: Multi-stage decoder for text recognition[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 14940-14949.

[15] Wang P, Li H, Shen C. Towards end-to-end text spotting in natural scenes [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(10): 7266-7281.

[16] Wang K, Yi Y, Tang Z, et al. Multi-scene ancient Chinese text recognition with deep coupled alignments[J]. Applied Soft Computing, 2021, 108: 107475.

[17] Nikolaidou K, Seuret M, Mokayed H, et al. A survey of historical document image datasets[J]. International Journal on Document Analysis and Recognition (IJDAR), 2022, 25(4): 305-338.

[18] Jalali A, Kavuri S, Lee M. Low-shot transfer with attention for highly imbalanced cursive character recognition[J]. Neural Networks, 2021, 143: 489-499.

[19] 甘吉. 手写文字识别及相关问题算法研究[D]. 北京: 中国科学院大学,

中国知网 https://www.cnki.net

2021.

[20] 付飞飞. 基于深度学习的场景汉字识别问题研究[D]. 河南大学, 2020.

[21] 王德青,吾守尔·斯拉木,许苗苗. 场景文字识别技术研究综述 [J]. 计算机工程与应用, 2020, 56 (18): 1-15.

[22] 涂超虎,易尧华,王凯丽,等. 自适应多级特征融合的场景古汉字识别 [J/OL]. 武汉大学学报(信息科学版), 1-10.

[23] Hu J, Lim S G, Brown M K. Writer independent on-line handwriting recognition using an HMM approach[J]. Pattern Recognition, 2000, 33(1): 133-147.

[24] 任义. 基于贝叶斯网络分类器的随机优化研究[D]. 吉林大学, 2023.

[25] 樊继伟, 李朝锋, 吴小俊. 一种改进的主动支持向量机算法及其应用[J]. 计算机工程, 2009, 35(19): 178-180.

[26] LeCun Y, Boser B, Denker J, et al. Handwritten digit recognition with a back-propagation network[J]. Advances in neural information processing systems, 1989, 2.

[27] Shi B, Bai X, Yao C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(11): 2298-2304.

[28] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.

[29] Cheng Z, Bai F, Xu Y, et al. Focusing attention: Towards accurate text recognition in natural images[C]//Proceedings of the IEEE international conference on computer vision. 2017: 5076-5084.

[30] Lin Y, Wu J, Gao P, et al. Lsh-based large scale chinese calligraphic character recognition[C]//Proceedings of the 13th ACM/IEEE-CS joint confere Xu Y, Yin F, Wang D H, et al. CASIA-AHCDB: A large-scale Chinese ancient handwritten characters database[C]//2019 International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2019: 793-798.

中国知网 https://www.cnki.net

[31] Tian X, Sun T, Qi Y. Ancient Chinese character image segmentation based on interval-valued hesitant fuzzy set[J]. IEEE Access, 2020, 8: 146577-146587.

[32] Zhang C, Liu X. Feature extraction of ancient Chinese characters based on deep convolution neural network and big data analysis[J]. Computational Intelligence and Neuroscience, 2021, 2021.

[33] Jia B, Wu Z, Wu B, et al. Enhanced character embedding for Chinese named entity recognition[J]. Measurement and Control, 2020, 53(9-10): 1669-1681.

[34] 谢杰成, 张大力, 徐文立. 小波图象去噪综述[J]. 中国图象图形学报, 2002, 7(3): 209-217.

[35] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[36] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11534-11542.

[37] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[38] 邓杰荣,梁森,曹昕妍,等. 基于深度学习的汉字识别方法研究综述 [J]. 微纳电子与智能制造, 2020, 2 (03): 73-81.

[39] 孙华, 张航. 汉字识别方法综述[J]. 计算机工程, 2010, 36(20): 194-197.

[40] 李阳娟. 汉字识别技术的研究进展分析[J]. 科技和产业, 2022. 22(04): 180-187.

[41] 孙华. 基于多特征融合 SVM 的古汉字图像识别研究[D]. 长沙: 中南大学, 2010.

[42] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE conference on computer vision and pattern r

中国知网 https://www.cnki.net

ecognition. 2016: 770-778.

[43] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[44] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[45] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.

[46] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size[J]. arXiv preprint arXiv:1602.07360, 2016.

[47] 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.

[48] Schneider N, Piewak F, Stiller C, et al. RegNet: Multimodal sensor registration using deep neural networks[C]//2017 IEEE intelligent vehicles symposium (IV). IEEE, 2017: 1803-1810.

[49] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.

[50] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.

[51] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[52] 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.

[53] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolution

中国知网 https://www.cnki.net

al networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.

[54] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.

[55] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1492-1500.

[56] Li X, Wang W, Hu X, et al. Selective kernel networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 510-519.

[57] Liu M, Wang Z, Ji S. Non-local graph neural networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2021, 44(12): 10270-10276.

[58] Chen Y, Kalantidis Y, Li J, et al. A^ 2-nets: Double attention networks[J]. Advances in neural information processing systems, 2018, 31.

[59] Huang Z, Wang X, Huang L, et al. Ccnet: Criss-cross attention for semantic segmentation[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 603-612.

[60] Park J, Woo S, Lee J Y, et al. Bam: Bottleneck attention module[J]. arXiv preprint arXiv:1807.06514, 2018.

[61] Gholamalinejad H, Khosravi H. Vehicle classification using a real-time convolutional structure based on DWT pooling layer and SE blocks[J]. Expert Systems with Applications, 2021, 183: 115420.

[62] Jiao L, Yang Y, Liu F, et al. The new generation brain-inspired sparse learning: A comprehensive survey[J]. IEEE Transactions on Artificial Intelligence, 2022, 3(6): 887-907.

[63] Bruna J, Mallat S. Invariant scattering convolution networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2013, 35(8): 1872-1

中国知网 https://www.cnki.net

886.

[64] Fujieda S, Takayama K, Hachisuka T. Wavelet convolutional neural networks[J]. arXiv preprint arXiv:1805.08620, 2018.

[65] Williams T, Li R. Wavelet pooling for convolutional neural networks[C]//International conference on learning representations. 2018.

[66] Trevino-Sanchez D, Alarcon-Aquino V. Hybrid pooling with wavelets for convolutional neural networks[J]. Journal of Intelligent & Fuzzy Systems, 2022, 42(5): 4327-4336.

[67] Choi Y J, Lee Y W, Kim B G. Wavelet attention embedding networks for video super-resolution[C]//2020 25th International conference on pattern recognition (ICPR). IEEE, 2021: 7314-7320.

[68] Deeba F, Kun S, Ali Dharejo F, et al. Sparse representation based computed tomography images reconstruction by coupled dictionary learning algorithm[J]. IET image Processing, 2020, 14(11): 2365-2375.

[69] Li Q, Shen L, Guo S, et al. Wavelet integrated CNNs for noise-robust image classification[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 7245-7254.

[70] Huang S, Pu X, Zhan X, et al. SAR ship target detection method based on CNN structure with wavelet and attention mechanism[J]. Plos one, 2022, 17(6): e0265599.

[71] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[72] Huang H, Yu A, Chai Z, et al. Selective wavelet attention learning for single image deraining[J]. International Journal of Computer Vision, 2021, 129(4): 1282-1300.

[73] Fujieda S, Takayama K, Hachisuka T. Wavelet convolutional neural networks for texture classification[J]. arXiv preprint arXiv:1707.07394, 2017.

[74] Hu J, Zeng M, Wu E. Bag of Tricks with Quantized Convolutional Neural

中国知网 https://www.cnki.net

Networks for Image Classification[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.

[75] Kang B, Xie S, Rohrbach M, et al. Decoupling representation and classifier for long-tailed recognition[J]. arXiv preprint arXiv:1910.09217, 2019.

[76] Zhang Y, Wei X S, Zhou B, et al. Bag of tricks for long-tailed visual recognition with deep convolutional neural networks[C]//Proceedings of the AAAI conference on artificial intelligence. 2021, 35(4): 3447-3455.

[77] Chou H P, Chang S C, Pan J Y, et al. Remix: rebalanced mixup[C]//Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16. Springer International Publishing, 2020: 95-110.

[78] Zubair M, Woo S, Lim S, et al. Deep Representation Learning with Sample Generation and Augmented Attention Module for Imbalanced ECG Classification[J]. IEEE Journal of Biomedical and Health Informatics, 2023.

[79] Verma V, Lamb A, Beckham C, et al. Manifold mixup: Better representations by interpolating hidden states[C]//International conference on machine learning. PMLR, 2019: 6438-6447.

[80] Zhang H, Cisse M, Dauphin Y N, et al. mixup: Beyond empirical risk minimization[J]. arXiv preprint arXiv:1710.09412, 2017.

[81] Wang Y X, Ramanan D, Hebert M. Learning to model the tail[J]. Advances in neural information processing systems, 2017, 30.

[82] Zhou B, Lapedriza A, Khosla A, et al. Places: A 10 million image database for scene recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(6): 1452-1464.

[83] Mahajan D, Girshick R, Ramanathan V, et al. Exploring the limits of weakly supervised pretraining[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 181-196.

[84] Zhong Y, Deng W, Wang M, et al. Unequal-training for deep face recognition with long-tailed noisy data[C]//Proceedings of the IEEE/CVF confere

中国知网 https://www.cnki.net

nce on computer vision and pattern recognition. 2019: 7812-7821.

[85] Gupta A, Dollar P, Girshick R. Lvis: A dataset for large vocabulary instanc e segmentation[C]//Proceedings of the IEEE/CVF conference on compute r vision and pattern recognition. 2019: 5356-5364.

[86] He H, Bai Y, Garcia E A, et al. ADASYN: Adaptive synthetic sampling ap proach for imbalanced learning[C]//2008 IEEE international joint confere nce on neural networks (IEEE world congress on computational intelligen ce). Ieee, 2008: 1322-1328.

[87] Khan S H, Hayat M, Bennamoun M, et al. Cost-sensitive learning of deep feature representations from imbalanced data[J]. IEEE transactions on neu ral networks and learning systems, 2017, 29(8): 3573-3587.

[88] Cao K, Wei C, Gaidon A, et al. Learning imbalanced datasets with label-di stribution-aware margin loss[J]. Advances in neural information processin g systems, 2019, 32.

[89] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C] //Proceedings of the IEEE international conference on computer vision. 20 17: 2980-2988.

[90] Ren M, Zeng W, Yang B, et al. Learning to reweight examples for robust d eep learning[C]//International conference on machine learning. PMLR, 20 18: 4334-4343.

[91] Cui Y, Jia M, Lin T Y, et al. Class-balanced loss based on effective number of samples[C]//Proceedings of the IEEE/CVF conference on computer vis ion and pattern recognition. 2019: 9268-9277.

[92] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: synthetic minority ov er-sampling technique[J]. Journal of artificial intelligence research, 2002, 16: 321-357.

[93] Han H, Wang W Y, Mao B H. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning[C]//International conference on i ntelligent computing. Berlin, Heidelberg: Springer Berlin Heidelberg, 200

中国知网 https://www.cnki.net

5: 878-887.

[94] Drummond C, Holte R C. C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling[C]//Workshop on learning from imbalanced datasets II. 2003, 11(1–8).

[95] Shen L, Lin Z, Huang Q. Relay backpropagation for effective learning of deep convolutional neural networks[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14. Springer International Publishing, 2016: 467-482.

[96] Liu Z, Miao Z, Zhan X, et al. Large-scale long-tailed recognition in an open world[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 2537-2546.

[97] Tan J, Wang C, Li B, et al. Equalization loss for long-tailed object recognition[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11662-11671.

中国知网 https://www.cnki.net