



吉首大学

JISHOU UNIVERSITY

硕士学位论文

基于深度学习模型的秦简文字识别方法研究

作者 汪政阳

二〇二三年五月

分类号_____

密 级 公 开

UDC_____

单位代码 10531

基于深度学习模型的秦简文字识别方法研究

研究生姓名 汪政阳

指导教师 陈炳权

专业领域 人工智能

研究方向 图像处理

学位类别 电子信息硕士

论文提交日期 2023 年 6 月 2 日

论文答辩日期 2023 年 5 月 16 日

答辩委员会主席 _____ 论文评阅人 _____

Research on Qin Bamboo Slips Character Recognition Based on Deep Learning Model

Thesis Submitted to Jishou University for the Degree of
Master of Engineering

By

WANG Zhengyang

Supervisor: CHEN Bingquan

May 10, 2023

摘要

目前秦简文字的相关研究大多集中于文学、历史学等方面，在计算机视觉方面的研究仍然相对滞后，尚未形成一套系统的研究理论与方法，这使得开展秦简文字数字化工作变得困难重重。近年来，深度学习技术在文字识别领域已经取得了较为成熟的发展，但是针对秦简文字识别的研究仍有很大的提升空间，这也是本文研究的重点之一。利用深度学习技术可以推进秦简文字的数字化工作，促进秦汉文化遗产的保护和传承，同时也为相关领域的研究提供更加丰富和可靠的数据资源。

在这个研究背景下，本文着重从秦简文字样本的数据库构建、预处理与识别三方面工作入手，自建了秦简文字样本库，提出了一种秦简文字预处理流程和二值化算法，设计并实现了基于深度学习模型的秦简文字单个字符的识别算法，以解决目前秦简文字样本库缺失、图像质量欠均衡与文字识别准确率低下三个问题。

针对目前缺乏统一的秦简文字样本数据库问题，参照《里耶秦简（壹）》和《秦简牍合集》等书籍，整理出 66973 幅有效字符图像，并按照特定命名规则对图像文件进行命名。再通过腐蚀膨胀、仿射变换、旋转、平移、添加随机高斯噪声等形态学变换进行样本增量，实现了样本库的扩充，并设计了样本库的分类存储方式。

针对图像质量不平衡的秦简文字图像提出了一种基于图像信噪比自适应阈值模型的二值化算法。首先，通过尺寸归一化、灰度变换、降噪滤波等技术对图像进行处理；其次，根据图像信噪比（SNR）大小自适应设置阈值，分别采用 OTSU 算法和 Bernsen 算法进行二值化处理；最后，由峰值信噪比（PSNR）与结构相似性（SSIM）评价指标择优选取二值化图像，从而准确地提取秦简图像二值化后的文字轮廓。测试结果表明，该算法的二值化结果保留了更多的秦简文字细节特征和文字轮廓，其峰值信噪比和精确率也分别达到 25.61dB 和 76.67%，相较其它经典文本图像二值化算法，其性能指标均有明显提升。

以样本少且欠均衡的高信噪比秦简文字图像为研究对象，提出了一种基于改进的轻量级 AlexNet 网络模型的秦简文字识别算法。首先，调整 AlexNet 网络模型结构，构建四层卷积层与池化层的顺序连接；其次，在前两层卷积层分别融入 InceptionA 与 InceptionC 结构，以分解卷积的形式对秦简文字进行局部特征提取，选取 ReLU 函数作

为模型的激活函数，并在全连接层融入 Dropout 层进一步防止模型出现过拟合现象；最后，调用 Softmax 分类器完成秦简文字图像识别。仿真结果表明，该网络模型在轻量化与识别准确率的表现上具有明显的优势，平均识别耗时为 435ms，识别准确率达到 99.89%，识别效果表现良好，为秦简文字识别理论研究提供参考与借鉴。

关键词：秦简文字；图像二值化；深度学习；文字识别

ABSTRACT

At present, most of the relevant research on Qin Bamboo Slip characters is focused on literature, history, and other fields. Research on computer vision is relatively lagging behind, and a set of systematic research theories and methods have not yet been formed, making it difficult to carry out the digitization of Qin Bamboo Slip characters. In recent years, deep learning technology has achieved relatively mature development in the field of text recognition. However, there is still significant room for improvement in the research on text recognition of Qin bamboo slips, which is also one of the focuses of this paper. The use of deep learning technology can promote the digitization of Qin Bamboo Slips, facilitate the protection and inheritance of Qin and Han cultural heritage, and provide more abundant and reliable data resources for research in related fields.

In this research context, this paper focuses on the three aspects of database construction, preprocessing and recognition of Qin Bamboo Slip text samples, builds a Qin bamboo slip text sample database, and proposes a Qin Bamboo Slip text preprocessing process and binarization algorithm, designed and implemented a recognition algorithm for a single character of Qin bamboo slips based on a deep learning model, in order to solve the three problems of lack of sample database of Qin Bamboo Slips, unbalanced image quality and low text recognition accuracy.

Aiming at the current lack of a unified sample database of Qin Bamboo Slips, referring to books such as "Liye Qin Bamboo Slips(1)" and "Qin Bamboo Slips Collection", sorted out 66,973 valid character images, and named the image files according to specific naming rules. Then, morphological transformations such as erosion and expansion, affine transformation, rotation, translation, and adding random Gaussian noise were applied to the sample to increase the size of the sample library, and a classification storage method for the sample library was designed. Proposes a binarization algorithm based on the adaptive threshold model of the image signal-to-noise ratio for the unbalanced image quality of the Qin Bamboo Slips text images. The image is first processed by size normalization, grayscale transformation, noise reduction filtering and other technologies; then, the threshold is adaptively set according to the image signal-to-noise ratio(SNR), and the binarization process is carried out by using the OTSU algorithm and the Bernsen algorithm respectively. Finally, the binarized image is selected according to the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) evaluation index, so as to accurately extract the

text outline of the binarized Qin Bamboo Slips text image. The test results show that the binarization result of the algorithm retains more details and outlines of the Qin Bamboo Slips, and its peak signal-to-noise ratio and accuracy rate reach 25.61dB and 76.67%. Respectively, which is significantly better than other classic text image binarization algorithms.

Taking Qin Bamboo Slips text images with few samples and underbalanced high signal-to-noise ratio as the research object, this paper proposes a text recognition algorithm for Qin Bamboo Slips based on an improved lightweight AlexNet network model. The structure of the AlexNet network model is adjusted to build a sequential connection of four convolutional layers and pooling layers, and the InceptionA and InceptionC structures are integrated into the first two convolutional layers to perform partial localization of the Qin Bamboo Slips in the form of decomposed convolution. The ReLU function is selected as the activation function of the model, and the Dropout layer is integrated into the fully connected layer to further prevent the model from overfitting. The Softmax classifier is called to complete the text image recognition of the Qin Bamboo Slips. The simulation results show that the network model has obvious advantages in light weight and recognition accuracy, with an average recognition time of 435ms, a recognition accuracy of 99.89%.The recognition effect is good, which provides reference and reference for the theoretical research of Qin Bamboo Slips character recognition

Key Words: Qin bamboo slips, Image Binarization , Deep Learning, Text Recognition

目 录

第一章 绪论.....	1
1.1 课题背景及研究意义.....	1
1.2 国内外研究现状.....	2
1.2.1 图像预处理研究现状.....	2
1.2.2 手写体字符识别研究现状.....	4
1.2.3 秦简文字识别研究现状.....	6
1.3 本文研究内容和难点.....	7
1.3.1 研究内容.....	7
1.3.2 难点.....	8
1.4 本文组织结构.....	8
第二章 相关理论技术概述.....	10
2.1 图像预处理技术.....	10
2.1.1 伽马变换.....	10
2.1.2 非局部均值滤波.....	10
2.1.3 OTSU 二值化.....	11
2.1.4 Bernsen 二值化.....	11
2.2 深度学习及卷积神经网络.....	12
2.2.1 深度学习及神经网络.....	12
2.2.2 CNN 卷积层.....	13
2.2.3 池化层.....	15
2.2.4 全连接和分类.....	16
2.2.5 Inception V3 模块.....	17
2.3 本章小结.....	18
第三章 秦简文字样本库建立与样本预处理.....	19
3.1 样本库的组成.....	20
3.2 样本库建立过程.....	21
3.3 秦简文字图像预处理.....	22
3.3.1 图像增强.....	23
3.3.2 图像去噪.....	24
3.3.3 改进的二值化算法.....	25
3.4 样本增量与样本库存储.....	30
3.5 本章小结.....	31
第四章 基于轻量级 AlexNet 模型的秦简文字识别.....	32
4.1 传统 AlexNet 网络模型.....	32

4.2 改进的轻量级 AlexNet 网络模型	33
4.2.1 AlexNet 网络模型结构改进与参数调整	33
4.2.2 Inception V3 模块融合策略	33
4.3 秦简文字图像识别算法与模型	36
4.4 训练与测试	37
4.4.1 训练集与测试集的构建	37
4.4.2 实验环境设置	39
4.4.3 模型损失函数的建立	39
4.4.4 识别算法评估指标	40
4.5 结果与分析	41
4.5.1 不同 Inception 模块对网络性能的影响	41
4.5.2 本文算法与其它识别算法对比	43
4.6 本章小结	45
第五章 总结与展望	45
5.1 全文总结	45
5.2 本论文主要创新点	46
5.3 后续工作展望	46
参考文献	48

表 目 录

表 3.1 降噪性能对比.....	25
表 3.2 信噪比值与图像质量关系.....	26
表 3.3 各二值化算法在 QBS Text Dataset 的评估结果（平均值）.....	29
表 4.1 InceptionA 结构的参数及计算量变化.....	34
表 4.2 InceptionA 网络结构参数.....	34
表 4.3 InceptionC 结构的参数及计算量变化.....	35
表 4.4 InceptionC 网络结构参数.....	35
表 4.5 秦简文字图像识别网络模型参数.....	36
表 4.6 不同 Inception V3 模块在验证集上测试结果的均值.....	42
表 4.7 不同模型下文字识别网络评估指标结果对比.....	44

图 目 录

图 2.1 卷积核的卷积过程.....	14
图 2.2 全连接与局部连接.....	14
图 2.3 权值共享的过程.....	15
图 2.4 池化层池化过程.....	16
图 2.5 卷积神经网络特征信息逐层传递的过程.....	16
图 2.6 InceptionA 网络结构.....	17
图 2.7 InceptionC 网络结构.....	18
图 3.1 从左至右: 里耶秦简壹、里耶秦简贰、秦简牍合集壹(上).....	20
图 3.2 里耶秦简壹中截取的部分文字图像数据.....	21
图 3.3 从上至下: 高信噪比样本图、低信噪比样本图、模糊样本图.....	22
图 3.4 秦简图像-文释对照关系.....	22
图 3.5 秦简单文字图像预处理过程.....	23
图 3.6 高低信噪比图像二值化算法仿真对比.....	26
图 3.7 秦简文字图像信噪比自适应阈值模型.....	27
图 3.8 不同二值化算法仿真结果对比.....	28
图 3.9 增广后的秦简文字样本图.....	30
图 3.10 秦简文字样本库结构框图.....	31
图 4.1 传统 AlexNet 网络模型结构与参数.....	32
图 4.2 改进后的 AlexNet 网络结构.....	33
图 4.3 秦简文字图像识别网络结构.....	36
图 4.4 训练集与测试集文件夹的分类及命名格式.....	38
图 4.5 训练集与测试集样本的分类及命名格式.....	38
图 4.6 消融条件下的 Loss 值对比.....	41
图 4.7 消融条件下的识别准确率对比.....	42
图 4.8 R ₄ 模型下的秦简文字图像识别效果.....	43
图 4.9 不同模型的损失函数值结果对比.....	44

第一章 绪论

1.1 课题背景及研究意义

在中华民族上下五千年的历史当中，古籍文献记录了历代王朝的繁荣与落败，是了解当时历史文化的宝贵参考资料，也是传承优秀精神文明的重要载体。但随着时间来到现代，许多古籍资料经过自然侵蚀、人为破坏等因素的影响，保有率大大降低，部分流传下来的古籍文献也面临着残破不堪、释读困难的窘境，亟待进行人工修复和翻译。为了解决这个问题，古籍数字化技术应运而生。通过将古籍文献以电子数据形式保存下来，使得古籍资料能够永久留存后世，焕发出源源不断的活力，不仅能够保留中华民族的文化记忆，更能传承中华民族的文化智慧，是一项重要的文化事业和保护与传承中华民族的文化遗产的必要途径。

秦朝是中国历史上第一个统一的封建王朝，源自于战国时期的秦国。在秦统一中国后，秦始皇为了巩固其统治地位，采取“书同文”的方式统一了全国范围内的流通文字，也就是现在所说的小篆字体。这项重大改革不仅消除了不同书写体之间的混乱，也巩固了秦朝在政治、文化和教育方面的领导地位，尽可能消除了由于长期分裂割据造成的地区差异，以利统一。同时在秦朝还并未出现造纸术，秦人依旧使用丝帛、竹简和木牍作为书写材料，但丝帛的造价高昂，一般在记录重要资料时才会使用。因此竹简和木牍在秦朝的王室和民间中使用的更为广泛，从而形成了独到的秦简牍文字。秦简文字的发现对于研究秦的历史有着无法替代的重要作用，特别是对秦法律的研究尤其重要。

战国时期各国文字的基本结构虽然相同，但在细节上仍存在差异性。秦国丞相李斯以秦国原来使用的大篆籀文为基础，根据其他六国的字体繁简和偏旁位置进行了改革，制定出小篆字体并在本国内流通。最终制定出的秦简文字字形线条粗细均匀，平行排列均衡，整体圆润浑厚，结构以正、扁方形为主，长方形为辅，内紧外松，一字多形、异体字较少，构形严谨有序，讲求工稳与平衡。这项文字改革促进了文字的标准化和统一，消除了文字上的混乱和障碍，为大一统国家的建立打下了基础。同时秦国圣贤所创造出的知识，也能被记录在简牍中，传唱至今。但由于受到自然侵蚀、人为损坏等因素影响，许多秦简牍都遭到了不同程度的破损。因此整理和保护现存的古籍文献是传承和发扬传

统文化的重要手段。目前,虽然大量的秦简古籍已经被扫描制作成书并进行了人工释读,但是这些研究还只停留在初步阶段。随着信息时代的发展,对秦简牍文字进行数字化保护已经成为刻不容缓的任务。通过秦简文字识别技术,能够使为研究秦简文字的学者提供阅读秦文古籍的新方式,有利于促进秦简文字的信息化进程,对秦简文字的保护和传承具有重要的现实和历史意义。湖南省湘西龙山县里耶博物馆存放着大量的秦简文字简牍,这些简牍具有很高的历史和文化价值,是研究秦朝文化的重要依据。这些简牍的保存和展示,不仅可以帮助人们更好地了解秦朝的政治、文化和社会制度,而且还能为秦简文字的研究和传承提供有力的支持。因此,本文从湖南省文物考古研究所出版的《里耶秦简壹》、《里耶秦简贰》以及武汉大学出版社出版的《秦简牍合集》中选取了大量的秦文简牍扫描图像,首先构建秦简文字样本库,其次对这些图像进行预处理,最终完成对简牍中秦简文字单个字符的识别。

1.2 国内外研究现状

1.2.1 图像预处理研究现状

在秦简文字字符识别过程中,图像预处理是不可或缺的一步。针对图像中的文字,适当的预处理流程可以大幅提升识别算法的鲁棒性。而对于秦简文字图像来说有两个关键的预处理步骤,就是完成对于图像的增强操作和二值化处理。前者旨在通过抑制噪声来强化目标信息,从而为后续步骤提供优质的输入。而图片二值化则用来将前景与背景进行分割,将有用信息与无用信息有效地进行隔离;从而使得识别效率大大提升。

随着深度学习技术的迅速发展,神经网络在图像增强领域的应用也日趋广泛,其中图像分割、视频预测以及图像修复等都是典型应用场景,对于提高图像处理的效率和质量具有重要意义。特别是在秦简文字预处理过程中,神经网络能够在有效强化目标信息的同时抑制杂质,从而为后期识别步骤带来巨大便利。康晓凤等^[1]提出了一种基于边缘学习的二阶有条件 GAN 模型,该模型利用边缘生成器生成边缘图像,并将其作为条件输入到图像补全网络中,从而实现图像的自动补全。与传统的图像补全算法^[2]相比,该算法的生成结果更加准确、清晰,且具有更高的视觉质量。此外,基于深度学习的超分辨率重建算法具备更强的自适应能力和更高的计算效率,能够在保证图像质量的同时,有效地提升图像的分辨率和细节清晰度。何鹏浩等^[3]提出了一种轻量级的单图像超分辨

率网络模型，通过在动态多尺度金字塔网络结构中融合子空间注意力模块来实现。该模型有效降低了网络模型的参数量，在实验仿真结果中，该模型也表现出较好的性能。陈清江等^[4]提出了一种基于端到端双网络的低照度图像增强方法，它由两个模块组成：**Inception** 模块以及 **URes-Net** 模块。利用双模块进行特征提取与融合，以此提升图像背景亮度。仿真结果表明，该方法能够保留更多的图像细节信息，并且具有较强的泛化能力，能够达到更好的视觉效果。

在文本图像的文字识别过程中，二值化算法的优劣，不仅能验证预处理前期步骤中图像增强降噪的效果，同时也为后续的检测识别奠定基础，因此文本图像二值化算法研究已成为当前人工智能研究领域的经典问题之一。文字在保存中难免会遭到人为或非人为的损坏，其图像的质量不容乐观，存在诸如噪声复杂、字迹模糊、文本载体残缺等质量不均衡现象，且给文字图像 PSNR 值和二值化精确率的提升带来诸多挑战，直接影响文字检测识别的准确率与速度。文献调研可知^[5-7]，数字图像二值化算法通常分为全局与局部两大方法，其中经典的全局二值化算法代表为 **OTSU** 算法^[5]，局部二值化算法主要包括 **Bernsen** 算法^[6]和 **Niblack** 算法^[7]等，**Niblack** 算法的优势在于能够很好地处理光照不均或噪声点复杂的图像，使得分割出的字符轮廓和细节可以最大程度得以保留，但缺点在于容易产生伪噪声，且伪噪声过度依赖于滑动窗口的大小，过小的窗口便会产生伪噪声。**Sauvola** 算法^[8]是在 **Niblack** 算法基础上进行的改进，将其像素点邻域内点的均值作为计算标准，辅以局部标准差再对其进行调整，这样便可以处理掉图像背景中的噪声点，但在处理低信噪比图像时其效果不佳。**Su** 等^[9]提出了一种自适应图像对比度的二值化方法，以处理背景退化情况。该方法构造了自适应对比度图，通过对比度值的分布情况来识别背景变化区域，从而有效降低背景变化对图像处理的影响。**Moghaddam** 等^[10]提出了一种全局自适应形式的 **OTSU** 二值化算法，该算法可以自动调整阈值，这是传统 **OTSU** 算法无法实现的，作者通过将阈值设置为背景估计值的一个倍数，实现阈值的自适应估计，有效地提高背景估计的准确性和鲁棒性。**Ntirogiannis** 等^[11]提出了一种全局和局部相结合的连通分量自适应二值化方法。该方法首先对图像中的背景噪声进行消除并增强图像对比度，其次利用全局二值化技术得到二值化图像的初步分割结果，再采用自适应二值化方法对图像进行局部分割，然后将全局与局部二值化处理结果相连通，得到最终有效的二值化图像轮廓。该方法在处理复杂文本图像时表现出了较好的性能。**Mitianoudis** 等^[12]基于局部特征和高斯混合模型，采用三个阶段提高图像二值化处理的准

确性：首先对图像中的背景噪声进行抑制消除，其次采用双态高斯混合等方法，区分字符像素和背景像素，得到更准确的二值化结果。最后融合形态学算子降低误分类点对模型造成的影响，进一步提高图像二值化的准确性。近几年，部分学者也在不断研究基于深度学习的图像二值化处理算法，Liao 等^[13]提出一种可微分的二值化算法，可以更加灵活地调整二值化的参数阈值。

国内外有关秦简文字二值化方面的数字化研究还处于起步阶段。陶珩等^[14]提出了针对秦简文字图像的数字化检测技术，利用最稳定极值区域算法和非极大值抑制算法来检测出秦简文字，但在秦简文字二值化处理方法上并未做深入研究。参照其他古籍图像与少数民族文字图像的二值化方法，Lu 等^[15]提出了基于背景估计和笔划宽度估计的二值化算法，首先选用一维的迭代高斯滤波算法估计古籍文档图像背景，再利用估计出的背景对不同情况的退化古籍文档做相应的补偿处理，最后根据图像的像素点平均值与笔画结果进行二值化处理。熊炜等^[16]针对存在墨迹类噪声点的文档图像提出了一种融合背景估计与能量函数的二值化算法，能够有效的处理文档中的噪声点，性能指标均优于 OTSU 等经典二值化算法。冯炎等^[17]提出了基于局部对比度和相位保持降噪的古籍图像二值化算法，在对古籍图像的局部对比度进行构建时，采用对局部像素值的最大值与最小值做归一化处理；在降噪处理上选用相位信息降噪处理，得到更好的估计结果；在上述基础上对古籍图像做增强处理后得到二值化结果。Howe^[18]提出了一种基于拉普拉斯图像的全局能量函数最优化二值化算法，该方法使用预先训练好的图像去选定目标函数，使其图像二值化结果的精确率更高。但该算法的鲁棒性不理想，特别是对于噪声点严重的低信噪比图像，其二值化仿真结果不尽人意。

1.2.2 手写体字符识别研究现状

文字识别是计算机视觉研究领域的分支之一，归属于模式识别和人工智能，是计算机科学的重要组成部分。图像文字识别技术^[19]是计算机视觉领域的主要研究内容。手写体字符识别(Handwritten Character Recognition)算法^[20]是当前的热点研究方向。许多汉字手写体识别框架已经逐渐商业化，国内如北京大学等高校也都对汉字手写体识别展开了深入研究。此外，部分前沿科技企业也源源不断地研发出创新科技产品^[21]，涵盖了图像识别、文字识别、自然语言处理等多个领域。这些产品都经过了严格的测试和验证，具有高效、准确的识别能力，得到了社会广泛的应用和认可。随着深度学习的兴起，手写体

识别问题也得到了更多有效的解决方案。深度学习技术可以自动地从原始数据中学习特征，避免了传统方法中需要手动设计特征的繁琐过程。近年来，基于深度学习的手写体识别方法在准确率和鲁棒性方面都获得了突破性的进展，一跃成为该领域的研究热点。Shi 等^[22]提出的卷积循环神经网络(CRNN)框架是最具代表性的文本行识别算法。该算法首先使用卷积神经网络(CNN)提取图像特征信息，其次采用双向长短期记忆网络(BiLSTM)对特征图序列进行处理，提高识别的准确性和鲁棒性，最后采用 CTC 解码有效地避免标注数据中的对齐问题，从而实现更加准确的文本行识别。该算法在水平文本行上取得了优异的效果，具有更好的适应性和泛化能力。Yang 等^[23]提出了一种针对中文古籍的文本行识别方法。作者在 CRNN 框架的基础上，将主体框架简化为 CNN+CTC 两部分，去掉了 BiLSTM 框架，因为 CNN 在特征提取时也可以捕获文本行的前后文信息，与 BiLSTM 相比计算量更小且容易训练，从而能够在文本行识别任务上达到较高的识别准确度。Shi 等^[24]提出了一种针对弯曲文本的识别网络模型。首先该模型的文本定位网络会生成一些推荐的采样位置，其次通过采样器对图像进行重新采样以得到被正向的图片，最后采样器将新图像进行采样，使其变为正向，然后将正向处理后的图片送入 CRNN 框架进行识别，该方法在弯曲文本的识别任务上取得了非常优秀的量化指标结果。Wan 等^[25]提出一种基于字符分割的场景文字识别方法 TextScanner，通过将字符分割为单独的部分，可以更好地理解字符之间的关系，并且可以将字符的空间信息更加准确地编码到模型中。该模型最大的优势在于可以通过文本序列来监督识别的过程，从而形成一个互监督信号，使得该方法在多数数据集上获得最高的识别准度。

AlexNet 网络模型由于能解决过拟合问题，并且可以利用多 GPU 加速计算，因此在目标分类识别领域也得到了应用。李江等^[26]针对弹道椎体目标分类问题，提出了 AlexNet 网络融合长短期记忆网络的模型结构，实现了对于弹道椎体目标的微动时频图智能分类；郭敏钢等^[27]对 AlexNet 网络的归一化、优化器和激活函数三方面进行改进，提升了 AlexNet 模型的训练收敛速度和识别准确率，在 MNIST 数据集上的识别准确率达到 98.78%；Zhang 等^[28]针对于训练样本量不足的问题，提出一种基于 AlexNet 网络的自动调制分类方法，实现了对于星座样本数据的扩充，并将分类准确率有效地提高到 90.5%；钟桂凤等^[29]通过在 AlexNet-2 网络中融入注意力机制，并使用 Word2Vec 对文本词特征进行嵌入表示和词向量训练，提升了文本分类的性能和运行效率，其在 20NG 数据集上的微观 F1 测度与宏观 F1 测度分别达到了 85.1%和 83.2%。然而，传统的 AlexNet 网络

模型架构较为复杂以及参数量过多过大,导致无法在样本量较少的小型数据集上取得满意的性能评估结果。因此在提升文字识别准确率的同时,应探究 AlexNet 模型的轻量化改进方法,使其能够在小型数据集上也有较好的性能体现。

随着对中英文数据集中字符识别探索的深入,国内部分学者也开始针对少数民族文字和古文字的识别展开相关研究。党红恩等^[30]结合小波变换和笔画合成技术方法对藏文的字形结构进行检测,并根据模糊集的模糊性程度实现对于边缘阈值的选取,最终完成对藏文字形结构的特征提取;刘卫等^[31]提出了一种改进的维吾尔文字母识别方法,用多模板归一化的方法对维吾尔文字母进行识别,可以更准确地识别不同字形的维吾尔文字母,提高了识别的准确性和鲁棒性。任宏宇等^[32]提出了一种基于手写维吾尔文字符运动轨迹的特征提取方法,可以更好地表征维吾尔文字字符的形态和结构特征。苏向东等^[33]提出了一种新的蒙古文古籍单词识别方法。作者采用字形基元切分方法对蒙古文古籍单词进行处理,再融合知识策略,对切分后的部分进行识别,提高了蒙古文古籍单词识别的准确性。Zhang 等^[34]提出了一种基于古汉字偏旁部首的识别方法,作者归纳出了 10 种古汉字的常见偏旁部首结构,将字符整体拆分为部分的偏旁部首进行识别,降低了分类器中的分类数量,既能完成对于单个古汉字字符的识别,也能完成对于文本行的整体识别,同时为少样本的古汉字识别问题研究提供了新的思路。

1.2.3 秦简文字识别研究现状

目前采用深度学习算法进行手写体字符识别的研究已经成为图像文字识别技术科研领域的主流^[35],提出的一系列深度学习算法,如 VGG^[36]、DenseNet^[37]、Res2Net^[38]、Vision Transformer^[39]和 Swin Transformer^[40]等均被应用于图像文字识别领域并取得了很好的表现效果。然而上述深度学习算法在对文字图像进行特征提取时,存在着文字特征性信息提取不精确导致最终识别准确率较低的问题^[41]。秦简文字其本身具有古籍类文字共性存在的字体结构多样,字形变化多端等特点,且个人书法习惯不一,加以文字图像采集过程中出现字迹磨损等现象,导致特征信息提取不精确,这都给秦简文字图像识别带来了挑战。因此,要结合秦简文字字形特征与手写体差异性等因素,探究有效识别秦简文字图像的深度学习模型,提高手写秦简字符识别准确率,具有重要的理论与实际意义。

当前秦简文字识别研究较少,但也有部分研究人员针对秦文进行了先期探索。吴峥

[42]采用 K-最近邻(K-Nearest Neighbor, 简称 KNN)分类算法实现了秦简文字图像识别, 识别准确率达到 70.53%。参照秦简文字与其他少数民族和古文字的共性, 又对部分文献进行了调研。Zhang 等[43]提出了一种基于小样本量的满文文本识别技术, 利用深度卷积神经网络模型进行文本识别, 使用滑动窗口代替人工分割, 满文文本的识别准确率达到 98.84%。仁青东主等[44]利用 CRNN+CTC 算法实现了对自然场景下的藏文字符识别, 提升了藏文字符识别效果, 识别准确率达 93.24%。Rahmati 等[45]提出了一种基于长短时记忆神经网络的波斯语文字识别算法, 通过对模型参数优化调整, 在字母级的文字识别准确率平均值达 99.69%。为了提高秦简文字识别准确率, 本文将在深度学习网络模型的基础之上进行系列改进, 为手写体古文字符识别领域的研究和实践提供有力的支持和新思路, 具有重要的研究意义和实际应用价值。

1.3 本文研究内容和难点

1.3.1 研究内容

本文的主要研究内容如下:

- (1) 构建秦简牍文字样本库, 完成对秦简文字样本的采集及标注。
- (2) 研究针对背景光照欠均衡和噪声复杂的秦简文字图像预处理算法。本文对秦简文字图像的预处理操作包含增强与二值化两方面。首先将图像进行灰度转换、调整亮度和降噪等一系列二值化前的预先处理, 然后对处理后的图像采用了一种基于图像信噪比自适应阈值模型的二值化算法, 实验结果表明, 该算法的二值化结果保留了更多的秦简文字细节特征和文字轮廓, 相较其它文本图像二值化算法, 其性能指标均有提升。
- (3) 研究基于卷积神经网络的秦简单文字字符识别模型, 以样本少且欠均衡的高信噪比秦简文字图像为研究对象, 提出一种基于改进轻量级 AlexNet 网络模型的秦简文字识别算法。通过调整传统 AlexNet 网络模型结构, 并在网络卷积层融合 inception V3 模块, 用分解卷积的形式代替传统卷积, 以有效提升模型进行特征提取的能力。实验结果表明, 本文最终采用的识别模型对秦简文字数据集中的单个字符识别准确率达到 99.89%。

1.3.2 难点

当前关于秦简文字图像的建库、预处理和识别研究还处于起步阶段,可供参考的文献资料有限,因此本文在样本采集、数据库建立、预处理和最终实现识别的过程当中都存在较大的困难,具体表现在以下三个方面。

(1) 秦简文字样本库的整理和标注存在困难。当前国内已经出土的秦简牍数量庞大,但能够整理成册,以公开形式获得的样本数量偏少,样本的获取途径不够丰富。其次研究秦简文字相关领域的专家也十分稀缺。此外在对秦简文字字符进行采集、标注、分类的过程中,需要多位研究人员耗费财力与精力共同完成,并且还需要规定字符采集、标注、分类的标准,避免研究人员所整理的样本之间存在较大差异性。这些困难不仅影响了秦简文字的研究和应用,也制约了秦简文字样本库的建设和发展。

(2) 多数秦简牍由于长期被埋入地下的原因,受到了自然侵蚀因素的干扰,加上后续出土和保存过程中人为破坏因素的影响,其上的文字区域存在严重的噪声。这些噪声都影响到了进行文字识别时的速度和准确率,因此在进行识别之前,需要充分考虑噪声的类型和分布规律,选取适当的预处理算法,有效地降低图像噪声,提高图像的质量和可识别性,为后续的秦简文字识别工作打下坚实的基础。

(3) 秦文字符集样本不均衡。在本文整理出的 1080 个字符类别的秦简文字样本中,698 个字符类别中包含多幅相同字符,382 个字符类别中只有单幅文字图像样本。这种样本量的不均衡导致文字识别的训练模型表现效果不佳。因此在构建模型的过程中,需要采用更加均衡的样本集,以提高模型的泛化能力和识别准确率。同时也需要探索更加有效的特征提取和分类算法,以提高模型的鲁棒性和识别精度。

1.4 本文组织结构

本文正文部分共有 5 章主要内容,每章中的具体安排如下所示:

第 1 章 绪论。首先介绍秦简文字识别的课题背景及其研究意义,引出本文的研究目的,其次介绍文字识别领域的国内外研究现状,包括传统的图像处理方法和基于深度学习的方法,并总结其优缺点和适用范围,在此基础上,提出本文的研究思路和方法。最后说明各章节的主要贡献和意义。

第 2 章 相关理论技术概述。首先对本研究要采用的预处理技术进行阐述,包括图

像增强和二值化算法；其次介绍了文本识别的相关算法，包括传统的图像处理方法和深度学习的方法，重点阐述深度学习的主要特点；最后对 Inception V3 模块做介绍说明。为后续的实验和研究提供了重要的理论基础和指导意义。

第 3 章 秦简文字样本库介绍。对秦简文字样本库的构成结构进行阐述，并介绍了图像样本的来源信息以及建立过程，包括预处理过程，增量过程和样本库存储形式，并提出了一种改进的秦简文字图像二值化算法，提高图像的质量和后续识别准确率。

第 4 章 基于卷积神经网络模型的秦简文字单字符识别。分析传统 AlexNet 网络模型的不足之处，并在此基础上进行改进，并通过实验数据证明本文所改进后的识别模型能够对秦简文字字符进行快速、准确的识别。

第 5 章 总结与展望。对全文的研究内容进行总结，并梳理归纳了本研究当中的创新点，同时对后续亟待开展的工作进行了概括与展望。

第二章 相关理论技术概述

2.1 图像预处理技术

2.1.1 伽马变换

伽马变换是一种常用的非线性变换。在图像处理中,利用伽马变换可以对曝光过度或者曝光不足的灰度图像进行对比度调节,将灰度过高或者灰度过低的图像区域进行矫正,增强图像整体细节。

2.1.2 非局部均值滤波

非局部均值滤波(NL-Means,简称 NLM)是目前较为主流的降噪技术之一。采用非局部均值滤波降噪后,不仅能够提升图像的清晰度,且能够有效地保存文本轮廓和细节。在非局部均值滤波算法中,采用了均方误差(MSE)作为判定邻域块相似度大小的标准,MSE 计算公式如式 2.1 所示。

$$MSE(A, B) = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (A(i, j) - B(i, j))^2 \quad (2.1)$$

m, n 表示邻域块的行列数; $A(i, j)$ 为 A 邻域块中的点 (i, j) 的像素值; $B(i, j)$ 为 B 邻域块中的相同位置点的像素值。

NLM 的权重根据邻域块相似度大小计算而来,相似度越大则权重越大,具体计算公式如式 2.2 所示。

$$W_{\text{NLM}} = \frac{1}{\text{sum}} \exp\left(-\frac{MSE(A, B)}{\sigma^2}\right) \quad (2.2)$$

sum 是整个搜索框内所有领域窗口的 MSE 之和; σ 是高斯系数权值,其值越大则最终的降噪效果越好。得到了每个像素块的权重,就可以计算出滤波后新的像素值,其计算公式如式 2.3 所示, I_p 为原像素, I_q 为新像素。

$$I_q = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} W(i, j) \times I_p \quad (2.3)$$

2.1.3 OTSU 二值化

OTSU 二值化算法采用基于聚类的思想, 根据图像总的像素点数在全局范围内实现阈值的最优化计算, 使其计算出的二值化阈值可以达到类之间的方差最大化。对于一幅 $M \times N$ 大小的输入图像 $I(x,y)$, 选取一个阈值 T , 使用 T 将 $I(x,y)$ 阈值化分割为前景和背景两部分区域, 将属于前景的像素点数占整幅图像的比例记为 c_1 , 其平均灰度为 μ_1 ; 属于背景的像素点数占整幅图像的比例记为 c_2 , 其平均灰度为 μ_2 。图像的全局灰度均值记作 μ , 类间方差记为 g 。图像中像素的灰度值小于阈值 T 的像素个数记作 N_1 , 像素灰度值大于阈值 T 的像素个数记作 N_2 , 则有图像算法的原理过程如下:

$$c_1 = \frac{N_1}{M \times N} \quad (2.4)$$

$$c_2 = \frac{N_2}{M \times N} \quad (2.5)$$

$$N_1 + N_2 = M \times N \quad (2.6)$$

$$c_1 + c_2 = 1 \quad (2.7)$$

$$\mu = \mu_1 \times c_1 + \mu_2 \times c_2 \quad (2.8)$$

$$g = c_1 \times (\mu - \mu_1)^2 + c_2 \times (\mu - \mu_2)^2 \quad (2.9)$$

将式 2.8 代入式 2.9, 可以得到最终化简后的公式为:

$$g = c_1 \times c_2 \times (\mu_1 - \mu_2)^2 \quad (2.10)$$

通过对图像进行遍历, 分别计算不同阈值下的类间方差, 最终选择能使之达到最大差值的阈值 T , 作为最终图像的全局阈值, 使得图像中两类像素的区分度最大。OTSU 算法在单峰图像的二值化处理结果中表现出色, 但在多峰图像中, 由于图像存在多个波峰, 而 OTSU 算法只能选取一个全局阈值, 无法很好地适应图像中的多个峰值, 因此在多峰图像中的二值化处理结果表现较差。

2.1.4 Bernsen 二值化

Bernsen 算法是一种局部二值化算法, 在光照欠均衡与噪声点较多的图像上的表现效果优于 OTSU 算法。其算法步骤是首先设置 2 个默认参数值, 其次计算图中以某一像

素点值 P 为中心, 窗口大小设置为 $k \times k$ 的像素参数值, 并与预先设置的 2 个默认参数值进行比较, 以数值大小判定选取该像素点处的灰度值。这种方法可以较好的处理光照不均匀的多峰图像。

2.2 深度学习及卷积神经网络

2.2.1 深度学习及神经网络

深度学习(Deep Learning, DL)是一种机器学习算法, 其核心思想是通过多层神经网络来模拟人类大脑进行信息处理和学习。深度学习的目标是实现对数据的分类、识别、预测等任务, 通过学习数据中的特征和规律来达成这一目标。与传统的机器学习算法相比, 深度学习算法具有更强的自适应性和泛化能力, 能够处理大规模、高维度的数据。在许多领域中, 如图像识别、语音识别、自然语言处理等, 深度学习算法都取得了非常优秀的表现, 成为人工智能领域的一个重要分支。

深度学习的核心是神经网络(Neural Networks, NNs)。神经网络的结构和功能受到人类神经系统的启发。神经网络由多个节点(也称为神经元)组成, 这些节点通过连接形成多个层次。在神经网络中, 每个节点接收来自前一层节点的输入, 并通过加权求和与激活函数进行处理, 产生节点的输出。每个节点的输出又可以作为其他节点的输入, 形成多层次的信息传递和处理。通过反复迭代训练, 神经网络能够逐渐学习数据中的特征和规律, 从而实现对数据的分类、识别、预测等任务。神经网络具有更强的自适应性和泛化能力, 能够处理大规模、高维度的数据, 并能够从数据中学习复杂的非线性关系。

深度神经网络(Deep Neural Network, DNN)是一种神经网络模型, 具有多个层次(深度)结构, 可用于处理高维度和复杂的数据。每个层级由多个节点组成, 节点之间通过连接进行信息传递和处理。与传统的神经网络相比, DNN 具有更多的层级, 因此可以更好地学习和理解数据中的抽象特征和非线性关系, 从而实现更高级别的特征提取和表示。其在图像识别、自然语言处理、语音识别等领域中得到了广泛应用, 并在许多任务中取得了优异的结果。

卷积神经网络(Convolutional Neural Network, CNN)是一种针对图像处理和计算机视觉任务设计的神经网络模型。它的最显著的特点是卷积层和池化层的使用, 能够自动地从原始图像中提取特征, 具有位置不变性和参数共享的特点, 因此非常适合用于图像分

类、目标检测、图像分割等任务。此外 CNN 还可用于自然语言处理等领域，因其能够自动发现文本数据中的相关特征，进而提高文本分类和情感分析的准确性。卷积神经网络的训练通常使用反向传播算法，通过不断调整卷积核和池化层的参数，最小化预测误差。同时，为了避免过拟合，卷积神经网络还采用了一些正则化技术，如 dropout、L1 和 L2 正则化等。同时需要注意的是，卷积神经网络的训练需要大量的计算资源和数据，并且需要仔细设计网络结构和超参数，以避免出现梯度消失或梯度爆炸等问题。

2.2.2 CNN 卷积层

卷积层是卷积神经网络的核心组成部分。它由多个卷积核组成，每个卷积核可以看作是一种特征提取器。卷积运算可以从输入的图像中提取出相应的特征。当卷积核在图像上滑动时，会对每个位置的像素进行一系列的乘法和加法运算，从而得到一个新的特征图。卷积层可以多次堆叠使用，每一层的特征图都可以作为下一层的输入。这种层层堆叠的方式，可以让网络逐渐地提取更加抽象和复杂的特征，从而更好地区分不同类别的图像。

在卷积神经网络中，每个卷积层都会使用卷积核对输入的图像进行窗口滑动，以提取图像中的特征。这种操作可以通过调节移动的步长来适应不同的图像特征。在卷积神经网络中，每个卷积层都会执行这种操作，将输入的图像用设定好的卷积核（即滤波器）进行窗口滑动。移动的步长可以进行调节，以便更好地捕捉图像中的特征。如图 2.1 中，卷积核作为滤波器，对左侧的原始输入数据进行滑动，得到右侧的新的二维数据。这种操作可以在多个卷积层中进行堆叠，以逐渐提取更加抽象和复杂的特征，从而更好地区分不同类别的图像。通过不同的滤波器组合，卷积神经网络可以提取出不同的特征，从而实现了对图像的分类、识别等任务。

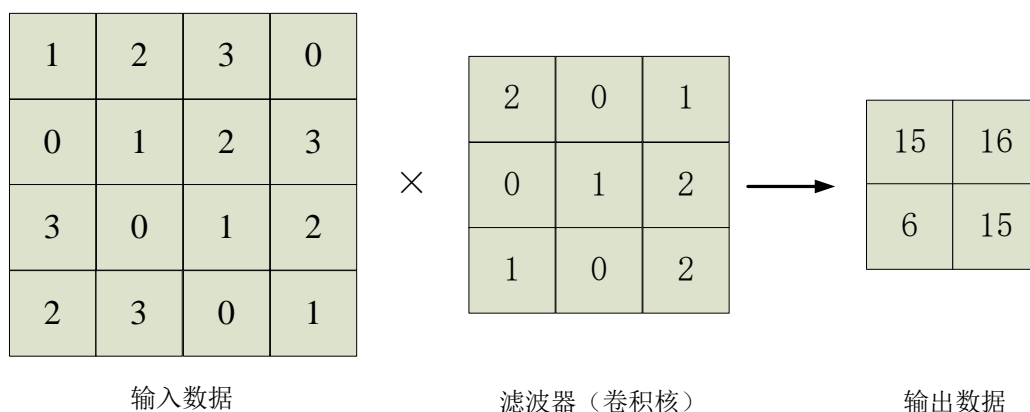


图 2.1 卷积层进行卷积操作的工作原理

在对卷积操作进行了解之后，可以进一步推出卷积神经网络的两大特点：局部感知和权值共享。神经网络中的局部感知思想与生物学视觉系统之间存在着紧密的联系。人类在对外界环境做出反应时，一般都是从离自身较近的部分向外界的整体进行感知反馈，而图像中也存在相似情况：相邻区域之间相关性更强，而远隔区域之间相关性则弱。卷积层在滑动卷积核时，只关注局部区域，并从中提取相应的特征，从而实现局部感知。如图 2.2 所示，图 2.2(a)为全连接过程，图 2.2(b)为局部连接过程。此外，卷积层中的权值共享机制也能够有效地减少模型的参数数量，提高模型的训练效率。这些特点是 CNN 相对于其他神经网络的独特之处，使得 CNN 在图像识别等领域具有很强的优势。

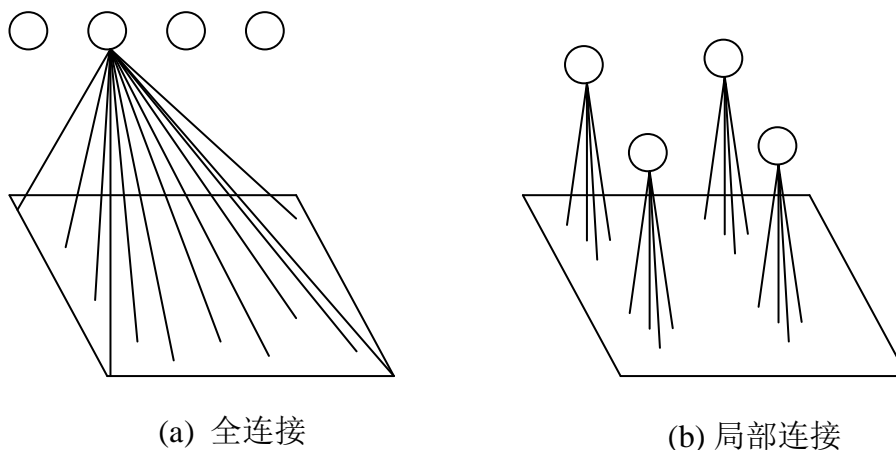


图 2.2 全连接与局部连接

由图 2.2(a)可以看出，全连接网络结构的每个节点都与前一层的所有节点相连。这种结构的优点在于它可以学习到输入样本中的所有信息，并通过整合前一层网络提取的特征将这些特征映射到样本标记空间上，从而产生相对准确的分类结果。由 2.2(b)可以看出，局部连接当中的神经元仅与前一层的一部分神经元相连接，即后一层神经元只能感知到前一层神经元的局部信息。这种结构的好处是可以减少神经网络需要学习的参数

数量，提高网络的计算效率和泛化能力。在图像处理中，局部连接常常被用来处理大规模的图像，因为图像通常具有很高的分辨率和维度。通过局部连接，神经网络可以更加高效地处理这些大规模的图像数据，从而提高网络的分类和识别性能。但即便采用局部连接的方式，在处理过程中的网络参数依旧较大，因此需要引入权值共享的概念。通过采用权值共享原理来减少对训练时设备所要消耗的大量参数、冗余信息以及准备时间。如图 2.3 所示为权值共享的示意图。

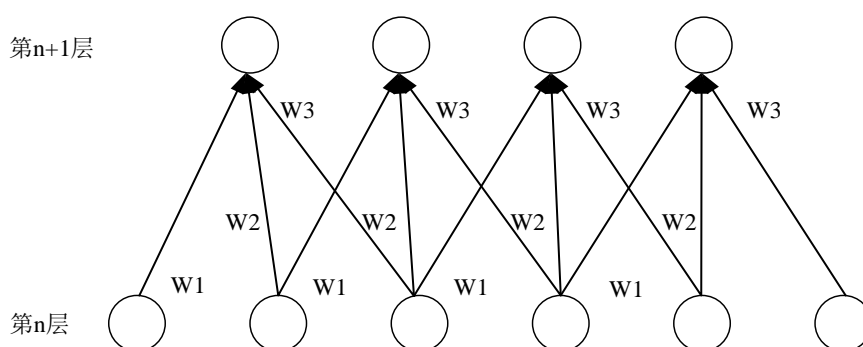


图 2.3 权值共享的过程

权值共享技术中，多个神经元共享相同的权值参数，这种共享的方式可以在一定程度上减少神经网络需要学习的参数数量，并提高网络的计算效率。通过权值共享，神经网络可以共享一组较小的参数，而不是为每个神经元单独学习一组大量的参数。这可以在一定程度上降低神经网络的计算复杂度，并提高网络的泛化能力。图 2.3 中，第 $n+1$ 层神经元只与第 n 层中的 3 个神经元进行连接，每一个神经元与上一层的神经元进行连接时，权值是相同的。 W_1 、 W_2 、 W_3 是不同神经元共享的权值。在卷积层中，权值共享是指同一深度切片的神经元共享权值。然而通常会采用多组卷积核来提取不同深度切片的特征，不同深度切片的神经元权值是不共享的。这表明图像的不同部分具有相同的统计特性，学习到的特征在不同位置上也同样适用。

2.2.3 池化层

在对图像进行处理的过程中，如果对图像中的每个像素点都进行计算，会大量增加网络模型的冗余参数和运行成本，造成不必要的浪费。因此需要选择性的提取图像中的具有代表性的特征信息。为了达到这个目的，在卷积神经网络中可以选择使用池化操作。在对图像进行卷积操作后，池化可以进一步提取图像中的关键特征。目前较为主流的池化方法有最大池化和平均池化两种方式。最大池化会在某一区域内选取最大值作为输出，

而平均池化则会计算某一区域内所有像素的平均值作为输出。这些方法可以有效地减少特征图的大小，并保留重要的特征信息。通过池化操作，神经网络可以在不丢失关键信息的情况下，大大减少网络的计算复杂度，并提高网络的泛化能力，不仅能够快速地从图片中抽取出重要特征，而且还能保留一定的信息，进一步减少了网络需要学习的参数。图 2.4 展示的是一个最大池化的池化过程。

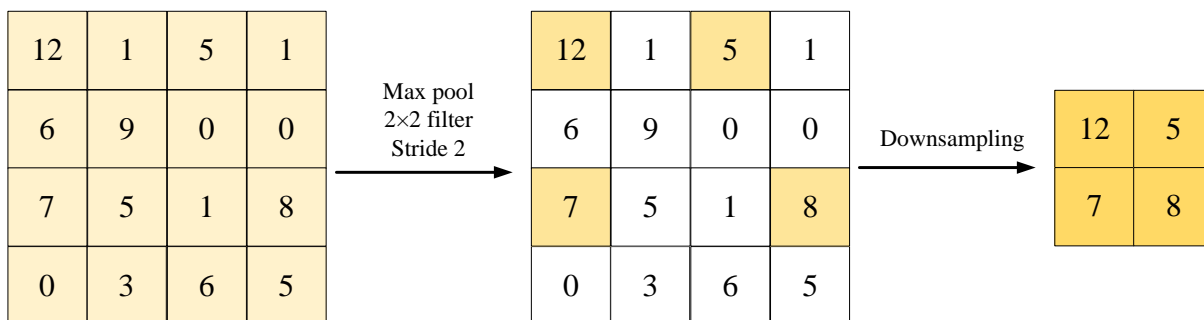


图 2.4 最大池化的池化过程

2.2.4 全连接和分类

全连接层的作用是将之前卷积层或者池化层提取出来的特征信息进行整合和压缩，并通过多项式逻辑回归预测输出值。在全连接层中，每个神经元都有一个权重向量和一个偏置项。这些权重和偏置项是通过反向传播算法进行训练的，目标是使得网络的输出结果与实际结果之间的误差最小化。为了进一步提升网络的性能，全连接层中每个神经元的激励函数一般采用 ReLU 函数，这样可以使得网络的非线性特征更加明显。全连接层可以大大减少网络的参数数量和计算复杂度，同时也可以提高网络的泛化能力和准确率，使得网络在处理图像识别等任务时表现更加优秀。图 2.5 以阿拉伯数字图像的识别为例，展示了卷积神经网络的整体工作流程。

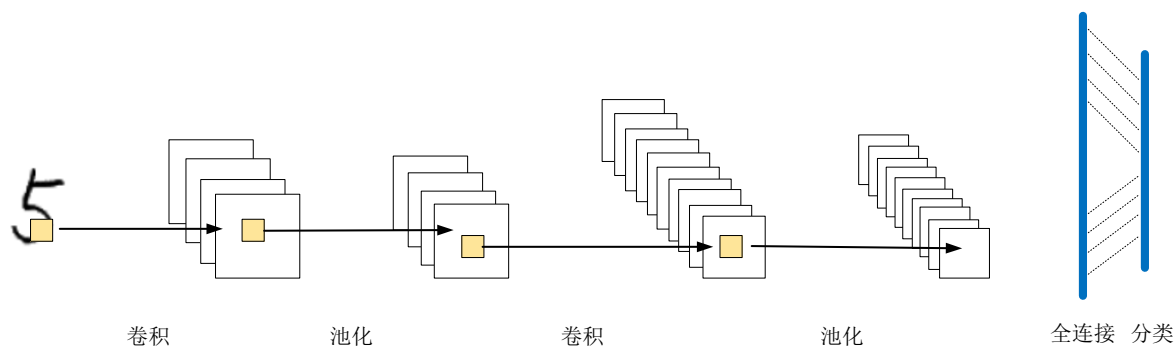


图 2.5 卷积神经网络的整体工作流程

2.2.5 Inception V3 模块

Inception 模块^[46]的核心思想是将输入的图像特征分配在多个不同卷积层进行卷积，并将多个卷积结果合并在一起。通过多个卷积层并行卷积操作，然后再将得到的特征提取结果连接起来，形成一个深层次的矩阵。在保持原模型结构稀疏性的同时，又利用密集矩阵的高计算性能，提升了模型识别准确率的同时避免了过拟合现象。

Inception V3 模块创新性的引入了分解卷积的概念，替换掉了在卷积层中一层接一层的传统特征提取操作，把大卷积因式分解成小卷积和非对称卷积，在保持感受野不变的前提下，减少整体模型中参数的计算量。此外在最大池化下采样的过程中计算输入值的卷积结果和池化结果，并将二者用 `concat()` 方式做数组合并，在减少计算量的同时降低了信息特征提取的损失量。

Inception V3 模块共有 6 种网络结构分支，其中 InceptionA 与 InceptionC 两种轻量级模块的基础网络结构如图 2.6 和图 2.7 所示。

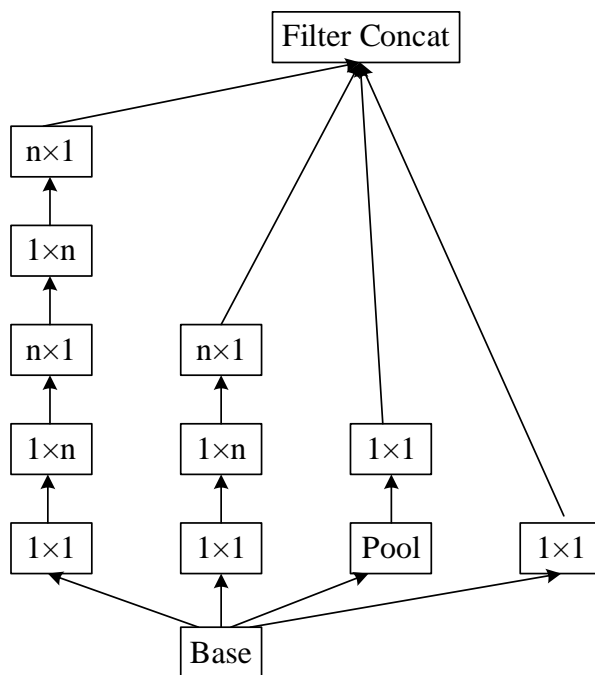


图 2.6 InceptionA 网络结构

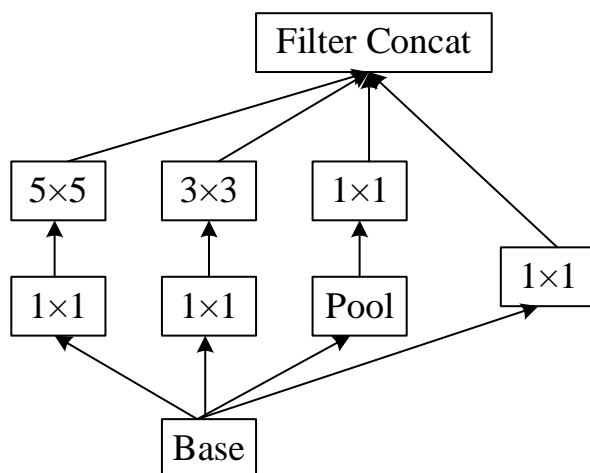


图 2.7 InceptionC 网络结构

2.3 本章小结

本章主要介绍了图像预处理技术和深度学习卷积神经网络的发展历程和工作原理。其中图像预处理技术从图像增强和二值化两个方面进行了详细介绍。这些预处理技术可以在一定程度上优化图像质量和特征提取效果，为后续算法提供更好的数据基础。其次对深度学习与卷积神经网络进行详细阐述，并在本研究的实验当中对上述技术进行使用，提供更加清晰的实验思路和操作指导。

第三章 秦简文字样本库建立与样本预处理

在文字识别领域,建立样本数据集是进行后续研究的基础,其过程至关重要。首先,识别算法需要通过统计分析得出具有鉴别力的特征向量,因此需要提供大量样本集进行学习训练,以此来识别文字。其次,随着深度学习的发展,CNN的识别模型训练阶段变得更加重要。而数据集的规模、质量和多样性都是影响训练效果的重要因素。为了训练出具有高准确率的模型,需要大量的多样性数据作为训练样本。最后,使用统一的数据集进行测试,可以为各类识别算法的比较和评估提供一个标准基础。只有采用相同的数据集进行比较,才能够恰当地表明各类识别算法的优劣势,从而更好地指导和推动文字识别技术的发展。

近年来,随着手写输入笔划先后次序动态识别技术的发展,联机手写文字数据集在手写文字识别领域中也成为研究的热点。这些数据集涵盖了手写输入的各种情况,如不同字体、不同书写方式等,可以为手写文字识别算法的研究提供更多的挑战和数据支持。联机手写文字数据集的发展也为手写文字识别领域的研究提供了更加丰富的数据基础,也为算法的研究提供了更多的可能性。这种趋势表明,手写文字识别技术将会更加注重动态识别的能力,而相关数据集的建立将会成为未来的发展方向。

建立成熟的秦简文字样本数据集,可以体现其在多个方面的价值。首先,这样的数据集可以为秦简文字研究领域提供一个统一的实验标准,从而提高该领域的研究水平和影响力。其次,建立该数据集可以为算法的研究提供更多的挑战和数据支持,从而推动该领域的技术创新和进步。然而目前秦简文字识别领域没有一个标准的数据集,无法对不同的识别算法进行公正的比较和评估,也无法为该领域的研究提供一个共同的基础。此外,不同的研究人员使用的数据集可能存在差异,这会影响研究结果的可复现性和可靠性。这表明建立秦简文字样本数据集的需求迫在眉睫,这将有助于推进秦简文字识别领域的研究,并有望在人工智能领域引起更广泛的关注。

本文参考孟一飞^[47]对于西夏古文字图像样本库的建立步骤,初步完成了对秦简文字图像中单个字符样本数据库的建立。这些工作的完成有助于解决秦简文字识别领域目前缺乏统一标准的问题,为该领域的研究提供了重要的支持。所构建的单字样本数据集和文本样本数据集是针对秦简文字识别任务而设计的,可以为研究者提供一个统一的实验平台。值得注意的是,这只是秦简文字样本数据库建立的一个初步尝试,还需要进一步

扩充和完善。我们将继续进行这方面的工作，以期为秦简文字识别领域提供更加全面和可靠的支持。

3.1 样本库的组成

本文所构建的秦简文字样本数据库主要包括 2 个部分：分别是通过秦简古籍文献资料扫描获取和从秦简古籍中提取后通过数据增广样本获取。本章针对这两种方法进行介绍。

目前秦简文字数据集缺乏统一的数字化平台，使得秦简文字的数据采集、整理和归档变得非常困难，也限制了秦简文字研究的深入展开。建立起一个全面、准确的数据集，才能为秦简文字的研究和保护提供更可靠的基础和更多的可能性。当前部分研究工作者对出土的秦简牍进行总结归纳，并对简牍上的文字做出释读，最后以书籍的形式向社会开放展示，如图 3.1 所示。



图 3.1 从左至右：里耶秦简壹、里耶秦简贰、秦简牍合集壹(上)

2012 年，《里耶秦简（壹）》出版，该书搜集整理了在湖南省湘西州龙山县里耶镇里耶古城 1 号井第 5 层、第 6 层和第 8 层出土的秦简牍图像，总计收录 17000 余个秦简文字符。2018 年，《里耶秦简（贰）》出版，该书主要搜集整理了里耶古城 1 号井第 9 层出土的秦简牍扫描图像，总计收录约 18000 个秦简文字符，两本书中字符基本包含了大多数秦简古籍中包括的字符。2014 年，武汉大学简帛研究中心等编写了《秦简牍合集》，该系列共有 6 本书籍，包括对睡虎地、周家台、龙岗等地出土的秦简牍进行整理，总计

收录约 32000 个秦简文字符，堪称秦简牍整理工作的新里程碑。本文通过扫描秦简古籍文献资料所收集的秦简文字图像数据，便主要来源于以上 8 本书籍。

3.2 样本库建立过程

秦简文字样本库的建立过程包括截取图片、类别划分与建立图像-文释对照关系 3 方面内容，具体制作方法如下：

(1) 截取图片。首先将书籍中秦简图片进行扫描制作成电子文档，其次利用 HyperSnap 和 Photoshop 软件当中的截图抓取工具，对于扫描图像中的文字区域进行单个截取，如图 3.2 所示：



图 3.2 里耶秦简壹中截取的部分文字图像数据

(2) 按测试要求分类。根据清晰度与测试需求分为模糊图像样本库、低信噪比图像样本库与高信噪比图像样本库。图片分类标准参照信噪比(SNR)参数值设置， $SNR \geq 50dB$ 归为高信噪比样本， $30dB \leq SNR < 50dB$ 归为低信噪比样本， $SNR < 30dB$ 则归为模糊样本。最终取得高信噪比样本图 13134 幅，低信噪比样本图 35754 幅，模糊样本图 18085 幅。如图 3.3 所示：



图 3.3 从上至下：高信噪比样本图、低信噪比样本图、模糊样本图

(3) 建立图像-文释对照关系。该对照关系可以准确的检索到图像所对应的书籍出处位置，在图像文件的命名规则上采取了归一化处理，格式统一为：“UnearthedlandN_page_Num_num.jpg”，其中 Unearthed landN 代表该秦简书籍上所标注的出土地来源，该书籍的页编号用 page_Num 表示，在该页上所截取的字符编号用 num 表示。统一了图像与文释对照关系后，就可以建立每幅文字图像与书籍出处的联系，再按照归一化后的文件命名便可以迅速准确的找到图像来源。如图 3.4 所示。



图 3.4 秦简图像-文释对照关系

3.3 秦简文字图像预处理

针对整理出的秦简文字图像的特点，本文需要进行图像预处理，以降低自然侵蚀、

光照欠均衡、人为破坏等因素的影响，从而提高图像质量。为了提升秦简文字识别准确率，本文首先针对秦简图像光照背景欠均衡和噪声点复杂等问题，提出一种预处理流程，通过平衡背景亮度和降噪，达到提升图像质量的目的，为提升秦简文字识别效果奠定前期基础。本文选用的预处理操作主要包括：图像尺寸归一化、灰度变换、降噪、二值化等。如图 3.5 所示为秦简文字单个字符图像预处理过程。

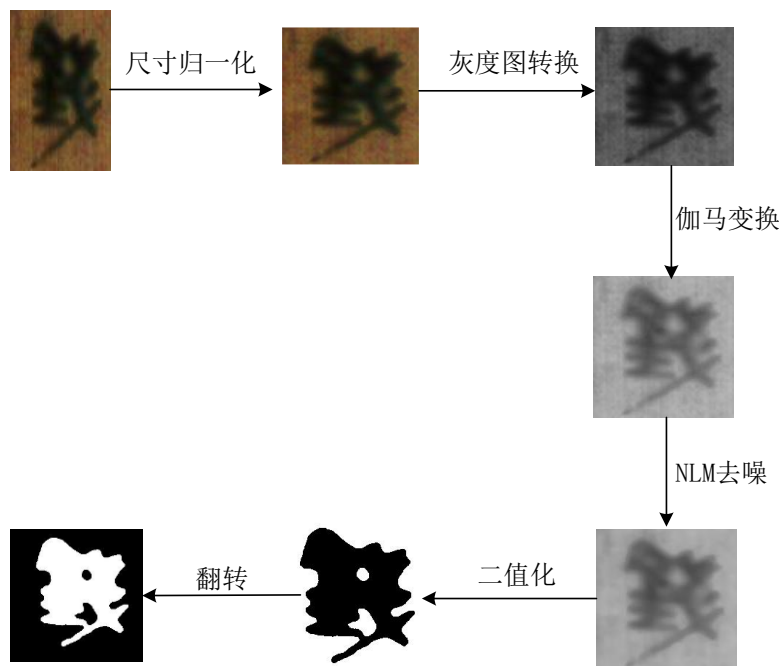


图 3.5 秦简单文字图像预处理过程

其中，针对图像质量不平衡的秦简文字图像提出了一种基于图像信噪比自适应阈值模型的二值化算法。为了能够突出显示图像当中的有效文字区域，方便后续处理，需要对图像进行归一化操作，确保在图像当中能够捕捉到文字信息。经裁剪过后的所有秦简文字单字符图像大小均为 224 像素×244 像素。经过图像归一化处理，可以得到一批清晰、标准化的单字图像数据集。对这些单字图像进行处理和分析，可以更加深入地了解文字的特征和规律，从而为相关领域的学术研究和应用提供更为可靠的数据支撑。

3.3.1 图像增强

在复杂环境下，由于受到光照强度、自然侵蚀、拍摄角度或人为损坏等因素影响，所以导致原始秦简文字图像会产生各种复杂背景与噪声点，使得后续秦简文字图像二值化的精确率严重下降。因此需对秦简文字图像进行增强处理，提高图像的清晰度，凸显秦简文字轮廓和细节。

图像灰度转换。为提升文本图像质量，通常将多通道的图像变为每个像素点仅由一

个灰度值表示的单通道图像。图像灰度转换能够将多通道的 RGB 图像转换为单通道灰度图像，由于大多数二值化算法都是基于灰度图像进行操作，所以要将输入的彩色图像进行灰度化处理。本文采用经典灰度转化算法，即 OpenCV 中封装的 `cvtColor()` 函数，实现彩色图像的灰度化转换处理。其转换公式如式 3.1 所示：

$$Gray = R \times 0.299 + G \times 0.587 + B \times 0.114 \quad (3.1)$$

式中： $Gray$ 表示转换后的图像像素的灰度值， R 、 G 、 B 分别表示原图像像素中的红、绿、蓝三基色分量。

利用伽马变换调整秦简文本图像亮度背景。具体过程：对输入图像每一像素点的灰度值做指数变换，使得输出图像中像素点灰度值与输入时的灰度值呈非线性关系，具体变换公式：

$$S = cr^\gamma \quad (3.2)$$

式中： r 为灰度图像的输入值，通常取值为 $[0,1]$ ； S 为经过伽马变换后的灰度输出值； c 为灰度调整系数，通常取 1； γ 为伽马因子，控制整个变换的调整程度。

本文采用自适应方法调整其 γ 值大小，调整为：

$$\begin{cases} \gamma = 0.5, & 0 \leq r < 0.5 \\ \gamma = 1, & r = 0.5 \\ \gamma = 2, & 0.5 < r \leq 1 \end{cases} \quad (3.3)$$

在输入图像的像素点灰度值 r 位于 $[0,0.5]$ 时，图像的亮度整体偏暗，此时将 γ 设定为 0.5 以放大灰度值；当灰度值 $r=0.5$ 时，图像此像素点处亮度较为均衡，因此将 γ 值设定为 1；当灰度值 r 位于 $(0.5,1]$ 时，图像的亮度整体曝光度过高，此时将 γ 设定为 2 以缩小灰度值，使最终图像的亮度背景处于较为均匀的状态，更适合人眼的观察范围和后续降噪处理。

3.3.2 图像去噪

本文在对 NLM 降噪参数设置过程中，将目标像素块的相似半径设置为 2×2 ，像素块的搜索区域半径设置为 5×5 ，高斯函数平滑参数设置为 10。经实验仿真结果分析，运用 NLM 算法对秦简文字进行去噪，可以在降噪声的同时较好地保留秦简文字区域的笔划细节。

为了体现出本文所选用的降噪算法的优势，文中选取了目前经典降噪算法（如中值滤波^[48]、高斯滤波^[49]、最小值滤波^[50]等）与 NLM 算法作对比分析。在选取横向对比降噪能力的评估指标上，本文采用基准指标即峰值信噪比(PSNR)。

从构建的秦简文字数据集中，选取了 100 幅图像对其采用各个降噪算法进行降噪处理，单独计算每幅图像的 PSNR 值后再对其整体求平均值，根据计算结果横向比较它们各自降噪能力的高低。降噪性能结果如表 3.1 所示。

表 3.1 降噪性能对比

降噪算法	PSNR/dB
中值滤波	29.161
最大值滤波	18.875
最小值滤波	20.639
高斯滤波	27.453
非局部均值滤波	30.216

通过对比分析发现，相较于其他传统降噪算法，采用非局部均值滤波降噪后的秦简文本图像取得了最高的 PSNR 均值，证明其能够最大限度降低秦简文本图像噪声，为后续的二值化处理提供最有利的条件。

3.3.3 改进的二值化算法

OTSU 算法的算法特性导致其在噪声点过多的低信噪比图像中，会将大量噪声点判定为需要提取的前景文本区域，且不能够针对于多个目标进行分割处理，在目标与背景大小呈现不均衡比例的情况下，其类间方差的计算结果会表现为多峰直方图形态，导致算法二值化效果不好。因此对于存在过多噪声点的秦简文字图像需使用其它二值化算法来进行二值化处理。Bernsen 算法在光照欠均衡与噪声点较多的图像上的表现效果优于 OTSU 算法。但 Bernsen 算法也存在一些不足，在处理高信噪比文字图像时，不能很好的处理噪声区域，导致文本区域和噪声区域产生大量黏连，最终显示出的二值化效果较差。

2 种算法对于秦简文字图像二值化处理的仿真对比如图 3.6 所示。其中图 3.6(a)为经过预处理后的较清晰图像的 OTSU 算法和 Bernsen 算法的仿真结果，图 3.6(b)为经过预处理过后仍存在复杂背景与较多噪声点图像的算法仿真结果。由仿真图可以看出，OTSU

算法在处理高质量图像时的表现效果普遍优于 Bernsen 算法，而在处理低质量图像时 Bernsen 算法的表现效果则普遍更好。

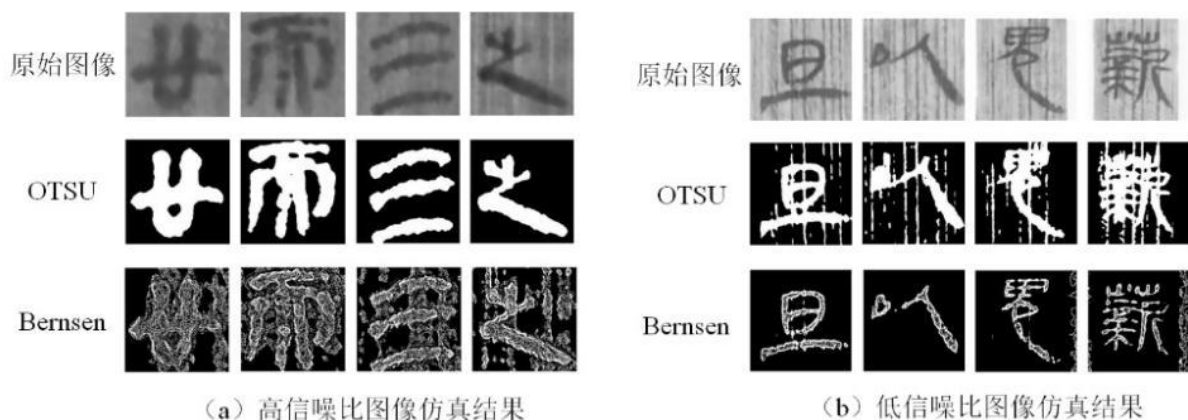


图 3.6 高低信噪比图像二值化算法仿真对比

因此，本文将 OTSU 算法与 Bernsen 算法进行结合，依据图像信噪比大小设置自适应阈值进行二值化处理，最后按照评价指标表现性能大小选取最佳二值化图像。

在进行秦简文字图像信噪比阈值确定的过程中，判定标准有三项：首先能够将所需文本区域从背景当中提取出来；其次能够保留更多的文字细节特征；最后能够定量分析强度。本文所使用的秦简文字图像数据集中的图像数据以数码相机拍摄的方式进行收集，因此采用通用相机的信噪比衡量标准进行阈值选取。信噪比值与图像质量关系如表 3.2 所示。

表 3.2 信噪比值与图像质量关系

SNR	图像质量
20dB	有严重噪声的图像
30dB	有大量噪声的图像
40dB	有精细颗粒的图像
50dB	有少量噪声的图像
60dB	无噪声的图像

在数据集中，将 $30\text{dB} \leq \text{SNR} < 50\text{dB}$ 的图像归为低信噪比图像，将 $\text{SNR} \geq 50\text{dB}$ 的图像归为高信噪比图像。因此本文根据实验要求，再结合数据集中数据与表 3.2 对应关系，将算法自适应阈值设定为 50dB 来进行仿真测试。

本文算法将 OTSU 算法与 Bernsen 算法结合，通过秦简文字图像信噪比设定自适应阈值进行二值化处理。算法流程步骤如下：

(1) 将输入的秦简文字图像依据信噪比值为 50dB 的自适应阈值进行高、低信噪比图像判定。将 $30\text{dB} \leq \text{SNR} < 50\text{dB}$ 的图像归为低信噪比图像，将 $\text{SNR} \geq 50\text{dB}$ 的图像归为高信噪比图像；

(2) 针对阈值高于 50dB 的高信噪比图像采用 OTSU 算法进行二值化处理；

(3) 针对阈值位于 $30\text{dB} \leq \text{SNR} < 50\text{dB}$ 的低信噪比图像首先采用 Bernsen 算法进行二值化处理，并计算图像的 PSNR 值与结构相似性(SSIM)值,再与 OTSU 算法得到的二值化图像的 PSNR 值与 SSIM 值进行对比，选择评价指标更优的二值化图像。

其模型流程思想如图 3.7 所示。

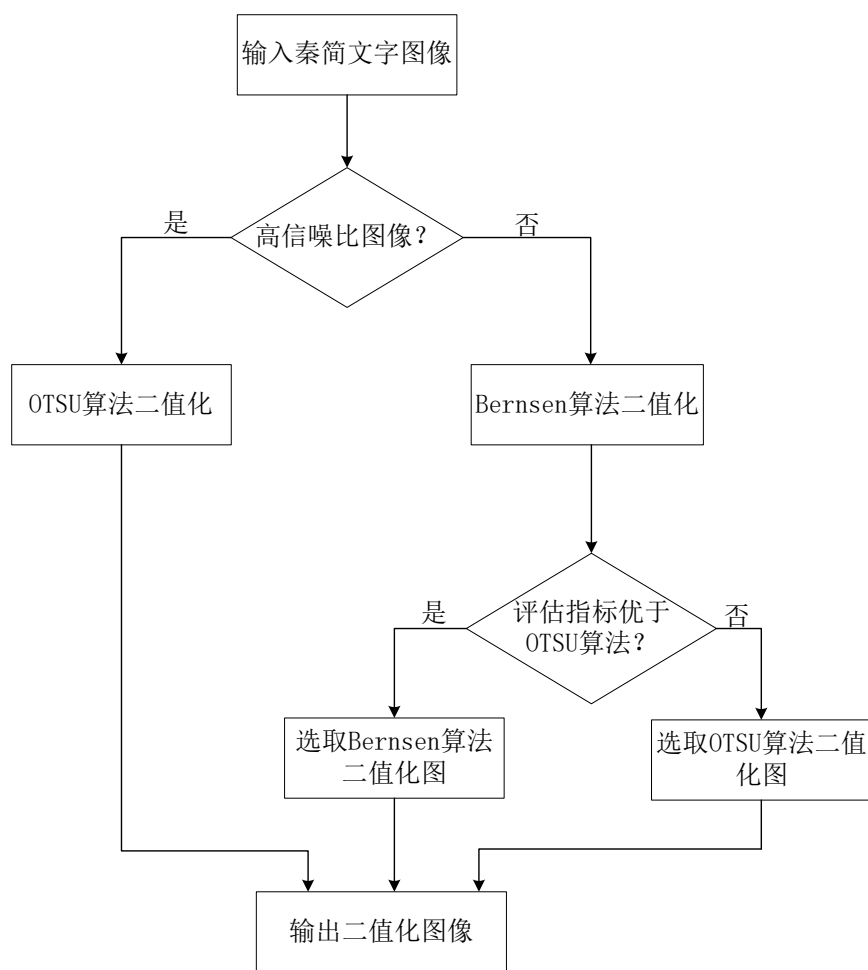


图 3.7 秦简文字图像信噪比自适应阈值模型

为了体现本文改进后算法的优势，并寻找与其他二值化算法之间的差别，从验证集中的不同信噪比样本图像库中选取了 10 张有代表性的实验图像，图 3.8 是本文改进后算法与其他算法的仿真结果展示。



图 3.8 不同二值化算法仿真结果对比

图 3.8(a)是较为清晰的高信噪比样本图与背景复杂或噪声点较多的低信噪比样本图。OTSU 和三角阈值分割算法对于高信噪比的清晰图像二值化处理效果较好,但对于带有复杂背景和较多噪声点的低信噪比图像处理没有达到预期效果;Bernsen 和 Integral 算法在处理高信噪比图像时会大量将文本区域错判为污渍区域而覆盖掉;Sauvola 和直方图算法尽管在一定程度上降低了噪声,但对于文本区域的黏连效果处理不太理想;Niblack、自适应阈值分割和最大熵阈值分割算法由于算法特性导致对噪声点敏感,会将大量的噪声点判定为需要保留的前景文本区域。本文提出的二值化方法在处理有亮度背景不均衡和噪声点干扰的秦简文字图像问题时,相比其他算法有了较好的改善,在一定程度上取得了相对满意的结果。

在构建的秦简文字数据集中,选取 1000 个验证数据与其它二值化算法结果对比,各项评估指标的平均值结果如表 3.3 所示。其中精确率指标的计算公式如下:

$$Precision = \frac{TP}{TP + FP} \quad (3.4)$$

表 3.3 各二值化算法在 QBS Text Dataset 的评估结果 (平均值)

算法	F-measure	PSNR	Precision	SSIM
OTSU	80.02	24.05	66.67	0.82
Bernsen	16.51	11.86	9.09	0.31
Niblack	37.53	6.17	23.08	0.68
Sauvola	80.02	18.53	66.67	0.97
Integral ^[51]	28.62	8.25	16.67	0.76
直方图 ^[52]	82.51	25.07	71.78	0.77
三角阈值分割 ^[53]	72.18	12.44	62.59	0.32
自适应阈值分割 ^[54]	25.96	8.25	21.17	0.32
最大熵阈值分割 ^[55]	27.06	10.58	19.54	0.35
本文算法	86.81	25.61	76.67	0.82

由表知,本文算法所处理的秦简文字图像的精确率达到 76.67%,F 值达到 86.81%,PSNR 为 25.61dB,SSIM 为 0.82。在 PSNR、F-measure 和 Precision 的评估指标结果中,本文方法都达到了相对最优值。在比较于各项评估指标次优的 OTSU 算法时,本文方法的 F-measure、PSNR 和 Precision 分别提高了 8.49%、6.49%和 15%,但在 SSIM 值的对比算法数据中还未达到最优,与最优的 Sauvola 算法相差 0.15,与次优的 OTSU 算法持平。

3.4 样本增量与样本库存储

为了提高模型的泛化性能和准确率,本文采用了传统形态学变换对收集到的样本进行了数据增广。这些形态学变换包括随机翻转、添加随机高斯噪声等,可以使得数据集更加真实、多样化,并且可以有效地缓解过拟合的问题。通过数据增广,可以在有限的数据集上获得更多的样本,从而提高模型的鲁棒性和准确率。这样的措施在深度学习中非常常见,也是提高模型效果的重要手段之一。在对样本进行增量操作后,同样归一化为 224 像素×224 像素大小。增量图像的文件命名规则在原图像的规则上新增一项增广编号,格式为:“UnearthedlandN_page_Num_num_Aug.jpg”, Aug 即表示增广编号,从 1 到 8 进行编号,分别代表不同的形态学变换方式。图 3.9 为秦简文字图像样本的不同形态学变换增广方式。

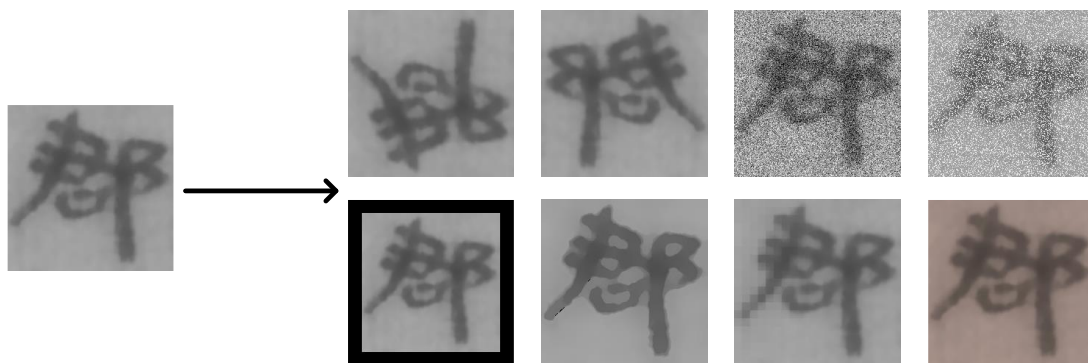


图 3.9 增广后的秦简文字样本图

样本增量完成后,要对所有样本进行数据存储。通过秦简古籍文献资料扫描获取和通过数据增广两种方式获取样本后,最终取得库中图像共计 512480 幅。在存储方式上首先分为秦简文字扫描库与秦简单文字字符库,在扫描库中存储的为秦简古籍文献资料的扫描图像,其次,在秦简单文字字符库中,根据图像文字清晰度的高低分为 3 个子库,其中高信噪比样本库共有 91472 幅图像,低信噪比样本库共有 276552 幅图像,模糊样本库中共有 144456 幅图像。最后,在高信噪比样本库中,又根据图像中文字的可识别性,将 91472 幅图像分为可识别文字与未识别文字两个子库,共计 1095 个字符类别。其中可识别文字库中共有 1080 个不同的字符类别,未识别文字库中共有 15 个不同的字符类别。本文最终所建的秦简文字样本库结构如图 3.10 所示。

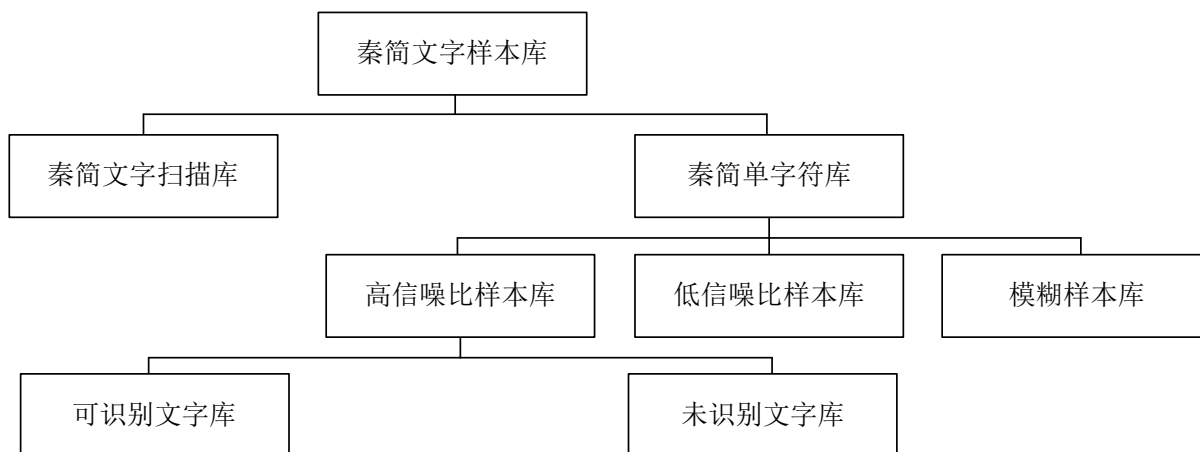


图 3.10 秦简文字样本库结构框图

3.5 本章小结

本章首先介绍了秦简文字样本库的构成，然后介绍了文字图像的收集整理、预处理过程，并提出一种基于图像信噪比自适应阈值模型的二值化算法，最后介绍了进行样本增量的传统形态学变换方法与样本库存储方式，使得样本库的规模得以快速扩大，同时也保证了数据的可靠性和高效性。最终能够满足后续秦简文字识别的需求。

第四章 基于轻量级 AlexNet 模型的秦简文字识别

4.1 传统 AlexNet 网络模型

AlexNet 是基于卷积神经网络(Convolutional Neural Networks, CNN)的深度学习模型。其最大优势体现在首次将模型部署在 GPU 上进行训练,对比用 CPU 进行训练时提升了模型的加速训练能力,由于采用计算复杂度更低的 ReLU 函数作为模型的激活函数,并在全连接层添加 Dropout 层随机移除神经网络训练元,消除了因模型参数过大而训练样本较少时出现的过拟合现象。另外,该网络由于构建了批量标准化层,减少了奇异训练样本造成的模型梯度弥散问题,从而增强了模型的泛化能力,实现了对输入图像数据的分类识别功能。

传统的 AlexNet 网络模型包括 1 个输入层、5 个卷积层、3 个池化层、2 个全连接层与 1 个输出层。模型结构如图 4.1 所示。AlexNet 模型在对文字进行识别通过多个卷积层的“卷积-激活-池化-归一化”操作,提取到文本区域的特征图,进而送入全连接层并调用分类器,计算输入图像属于某字符类别的概率,最终实现文字图像识别。

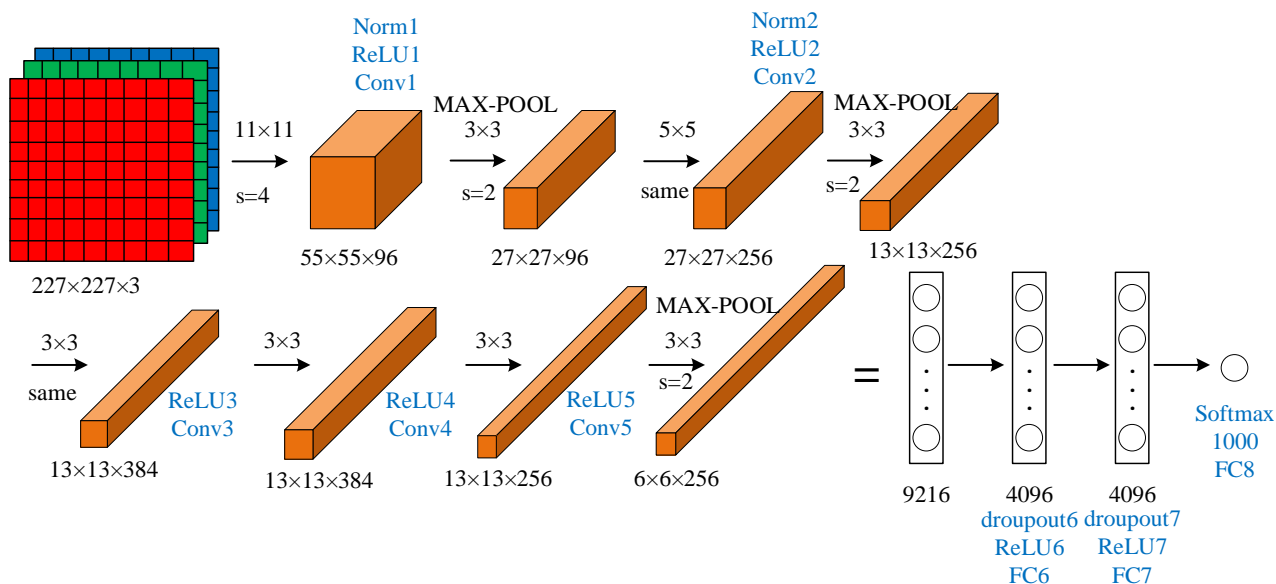


图 4.1 传统 AlexNet 网络模型结构与参数

4.2 改进的轻量级 AlexNet 网络模型

4.2.1 AlexNet 网络模型结构改进与参数调整

秦简单文字图像相较于文本区域内多文字的图像，其文本特征较易提取，但是，传统的 AlexNet 网络模型复杂度较高，难免提高了非必要的训练成本。因此本文根据秦简文字图像特征，将原始 AlexNet 模型的 5 层卷积层简化为 4 层卷积层，通过减少深度以降低模型复杂度，并添加 1 层池化层，设计成 4 层卷积层与 4 个池化层的顺序连接。在参数调整上，将输入图像的像素大小设置为 $224 \times 224 \times 1$ ，第 1 个卷积层设计了 11×11 大小的卷积核，在第 2 卷积层设计了 5×5 大小的卷积核，第 3、4 个卷积层设计了 3×3 大小的卷积核来提取局部图像特征，将池化层的池化大小设置为 2×2 ，并根据获得的神经元数量添加了一层全连接层，最终实现轻量级网络的搭建。图 4.2 所示为本文改进后的 AlexNet 网络结构。

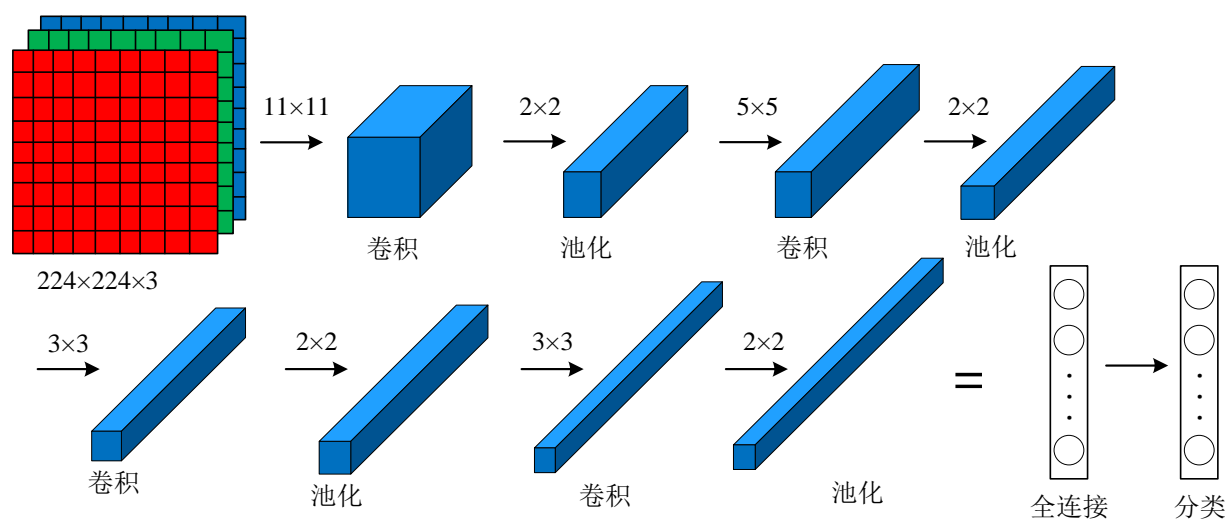


图 4.2 改进后的 AlexNet 网络结构

4.2.2 Inception V3 模块融合策略

Inception V3 共有 6 种网络结构分支，本文的秦简文字识别模型在卷积层融入其中的 2 种分支结构，在前 2 层卷积层分别融入了 InceptionA 和 InceptionC 结构，以分解卷积的形式代替传统卷积进行特征提取，减少网络计算成本，提升模型识别准确率。

结合秦简图像文本区域内文字为单个文字，进而笔划特征较易提取的特点，在

InceptionA 结构中, 本文分别选用 5×5 、 3×3 、 1×1 , 3 种不同尺度的卷积核代替初始的 11×11 大小的卷积核进行多通道的特征提取, 并计算输入值的池化结果, 最后将各通道进行融合。其模块的参数量及计算量变化如表 4.1 所示。

表 4.1 InceptionA 结构的参数及计算量变化

卷积核大小	参数量	计算量
11×11	$121 \times C^2$	$121 \times WHC^2$
5×5	$25 \times C^2$	$25 \times WHC^2$
3×3	$9 \times C^2$	$9 \times WHC^2$
1×1	C^2	WHC^2

其中 C 表示特征图的输入通道数, W 表示特征图的宽度, H 表示特征图的高度。使用 1 个 5×5 、2 个 3×3 和 3 个 1×1 的卷积核代替 11×11 的卷积核做分解卷积, 可以使参数量和计算量从 121 个单元下降至 46 个单元。其具体网络结构参数图如表 4.2 所示。输出通道数 c_{out} 的计算公式如式 4.1 所示:

$$c_{out} = \frac{c_{in} + 2 \times p - f \times f}{s} + 1 \quad (4.1)$$

式中: c_{in} 为输入的特征图大小, p 为特征图的填充圈数, $f \times f$ 为每层所使用的卷积核大小, s 为卷积的步长, 所采用的卷积方式统一为 same 卷积。

表 4.2 InceptionA 网络结构参数

网络层	类型	输入通道数(c_{in})	卷积核大小($f \times f$)	填充(p)	输出通道数 (c_{out})
输入层	输入特征图	30	-	-	30
分解卷积层 1	$1 \times 1_1$	30	1×1	-	48
分解卷积层 2	$5 \times 5_1$	30	1×1	-	48
	$5 \times 5_1$	48	5×5	2	64
分解卷积层 3	$3 \times 3_1$	30	1×1	-	48
	$3 \times 3_2$	48	3×3	1	64
	$3 \times 3_3$	64	3×3	1	80
池化层	池化通道	30	1×1	-	16
输出层	融合通道数	$48+64+80+16$	-	-	208

在 InceptionC 结构中,保持底层 3 个 1×1 卷积核不变的基础上,使用了 3 组 1×7 和 7×1 的卷积核进行分解卷积,并且根据新输入层的特征图数量,将结构内的通道数增加至 96。其模块的参数量及计算量变化如表 4.3 所示。

表 4.3 InceptionC 结构的参数及计算量变化

卷积核大小	参数量	计算量
5×5	$25 \times C^2$	$25 \times WHC^2$
1×7	$7 \times C^2$	$7 \times WHC^2$
7×1	$7 \times C^2$	$7 \times WHC^2$
1×1	C^2	WHC^2

由表 4.3 可知,使用了 3 组 1×7 和 7×1 的卷积核代替 5×5 大小卷积核进行分解卷积时,可以使参数量和计算量从 25 个单元下降至 21 个单元。其具体网络结构参数图如表 4.4 所示。结果表明采用融合 Inception V3 模块策略可以有效降低模型复杂度。

表 4.4 InceptionC 网络结构参数

网络层	类型	输入通道数 (c)	卷积核大小 ($f \times f$)	步长(s)	填充(p)	输出通道数 (c)
输入层	输入特征图	224	-	-	-	224
分解卷积层 1	$1 \times 1_1$	224	1×1	-	-	96
分解卷积层 2	$7 \times 7_1$	224	1×1	-	-	64
	$7 \times 7_2$	64	1×7	-	(0,3)	64
	$7 \times 7_3$	64	7×1	-	(3,0)	96
分解卷积层 3	$7 \times 7_4$	224	1×1	-	-	64
	$7 \times 7_5$	64	7×1	-	(3,0)	64
	$7 \times 7_6$	64	1×7	-	(0,3)	64
	$7 \times 7_7$	64	7×1	-	(3,0)	64
分解卷积层 4	$7 \times 7_8$	64	1×7	-	(0,3)	96
	$7 \times 7_9$	224	1×1	-	-	96
池化层	平均池化	384	1×1	1	1	384
输出层	融合通道数	$96+96+96+96$	-	-	-	384

4.3 秦简文字图像识别算法与模型

秦简文字图像识别算法模型结构与参数调整如图 4.3 与表 4.5 所示，以 AlexNet 模型为基础网络架构，通过在卷积层融合 Inception V3 模块实现秦简文字图像识别。具体步骤如下：

(1) 对使用的数据集中的秦简文字图像进行预处理操作，包括尺寸归一化、增强去噪、二值化和编码转换等，得到一幅 224×224 的单通道二值化图像，以此作为原始图像的输入；

(2) 对 AlexNet 网络模型进行改进，构造 4 层卷积层与池化层的顺序连接，并根据获取的神经元数量添加 1 层全连接层，构建 3 层全连接层将特征空间更好映射到样本标记空间；

(3) 在前 2 层卷积层进行特征提取时，分别融入 Inception V3 模块中的 InceptionA 与 InceptionC 结构，以分解卷积的形式代替传统卷积形式，降低模型权重和提高模型识别精度，以降低识别误差率，并通过池化层降低维度来提升模型容错率；

(4) 将经过特征提取处理过的特征图像送入全连接层，调用 Softmax 分类器，计算输入图像属于某字符类别的概率，最终实现秦简文字图像识别。

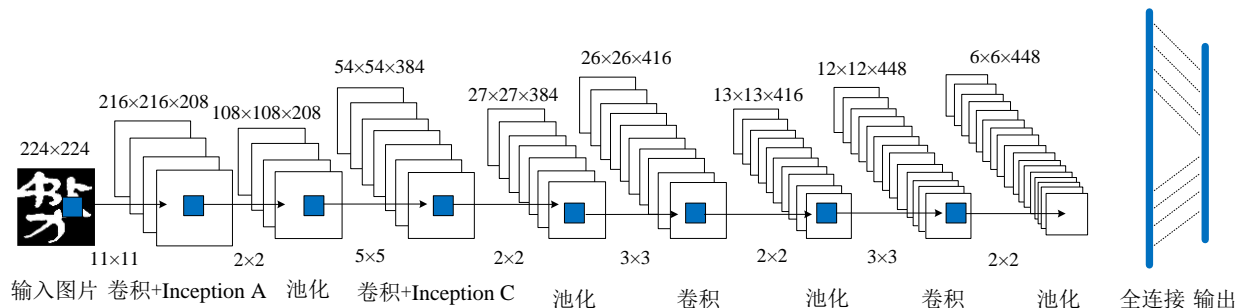


图 4.3 秦简文字图像识别网络结构

表 4.5 秦简文字图像识别网络模型参数

网络层	类型	输入大小(宽 W , 高 H)	输出通道数 (c)	卷积核大小 ($f \times f$)	步长 (s)	填充大小 (P)	输出大小(宽 W , 高 H , 通道数 c)
输入层	输入图像	224×224	1	-	-	-	$224 \times 224 \times 1$
层 1	卷积层	224×224	30	11×11	3	2	$216 \times 216 \times 30$
	InceptionA	216×216	208	$\begin{pmatrix} (1 \times 1) \\ (5 \times 5) \\ (3 \times 3) \end{pmatrix}$	-	2	$216 \times 216 \times 208$

续表 4.5

	池化层	216×216	208	2×2	-	-	$108 \times 108 \times 208$
层 2	卷积层	108×108	224	5×5	1	2	$108 \times 108 \times 224$
	InceptionC	108×108	384	$\begin{pmatrix} (1 \times 1) \\ (1 \times 7) \\ (7 \times 1) \end{pmatrix} \times 3$	-	$\begin{pmatrix} (0,3) \\ \times 3 \\ (3,0) \end{pmatrix}$	$54 \times 54 \times 384$
	池化层	54×54	384	2×2	-	-	$27 \times 27 \times 384$
层 3	卷积层	27×27	416	3×3	1	1	$26 \times 26 \times 416$
	池化层	26×26	416	2×2	-	-	$13 \times 13 \times 416$
层 4	卷积层	13×13	448	3×3	1	1	$12 \times 12 \times 448$
	池化层	12×12	448	2×2	-	-	$6 \times 6 \times 448$

4.4 训练与测试

4.4.1 训练集与测试集的构建

本实验在可识别文字样本库中选取样本数量相对较多的 50 个类别的字符进行训练测试。对选取的字符样本数量做归一化处理，使每一类别下的数量达到较为均衡的水平，最终的样本量总数达到 149713。按照 4:1 的比例划分出训练集和测试集，训练集有 119770 个样本，测试集有 29943 个样本。所有图像经过标准化处理，缩放为[-1,1]范围，实现字符样本分类识别。在文件及文件夹的命名上，文件采取数字顺序的排列命名方式，文件夹采取“num_text”的命名方式，其中 num 表示文件夹序号，text 表示该文件夹下字符的中文拼音。最终构建的测试集与训练集文件夹的分类及命名格式如图 4.4 所示，样本的分类及命名格式如图 4.5 所示。

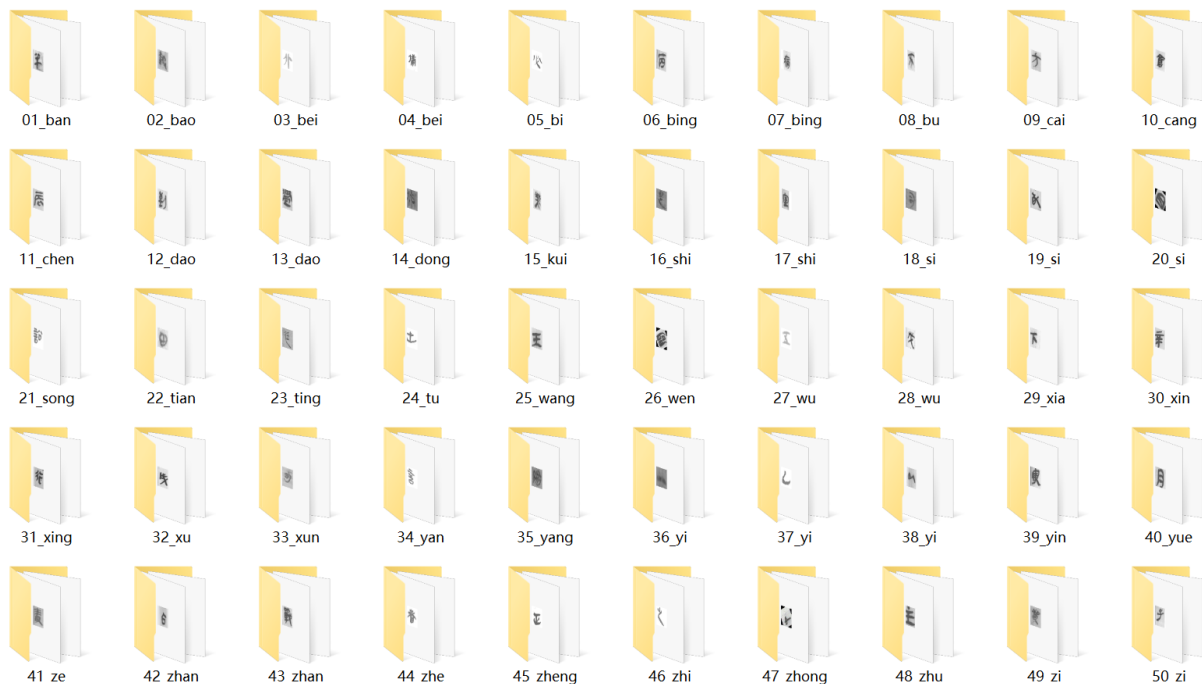


图 4.4 训练集与测试集文件夹的分类及命名格式

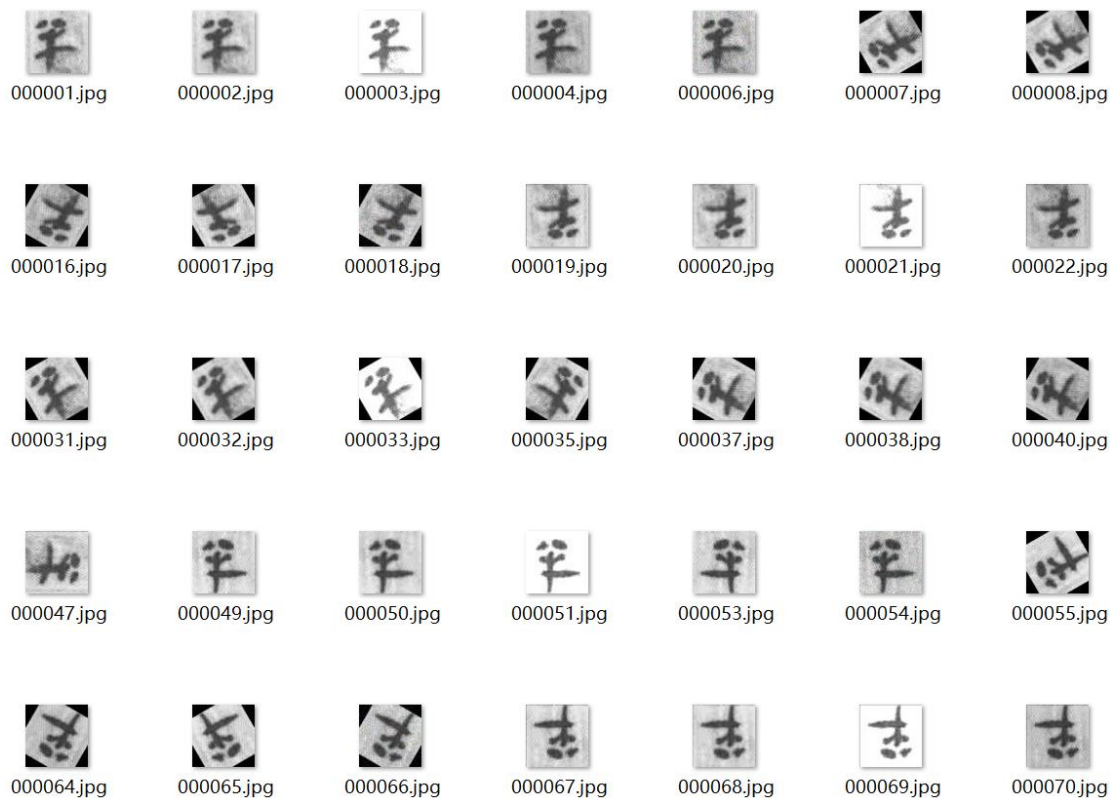


图 4.5 训练集与测试集样本的分类及命名格式

4.4.2 实验环境设置

本实验使用的所有图像样本像素大小均设置为 $224 \times 224 \times 3$ ，batch size 为 100。本次实验中使用的编译语言为 Python3.7，网络模型基于 Pytorch 深度学习框架建立，并将网络模型加载至 GPU 上进行，服务器显卡型号为 RTX3070。

在本文构建的秦简文字图像识别模型中，在 4 层卷积层中分别选取 11×11 、 5×5 、 3×3 和 3×3 大小的卷积核，最大池化层的池化尺寸设置为 2×2 ，激活函数采用 ReLU 函数。初始学习率设置为 0.01，每隔 10 轮训练变为原先的 0.5 倍，batch size 设置为 100，epoch 设置为 50，并在全连接层使用 Dropout 层防止出现过拟合现象，对应参数设置为 0.5。

实验主要包括两个部分：其一测试在 AlexNet 模型当中使用 Inception V3 模块与不使用时的仿真结果对比；其二测试改进后的 AlexNet 模型与其它图像识别算法对于秦简文字图像的识别效果。

4.4.3 模型损失函数的建立

模型采用 softmax 层作为训练的分类器，因此选用 softmax 函数的多分类交叉熵作为模型的计算识别损失函数可以保证模型的内聚性，进而判定模型预测值与真实值的偏差程度。原始的交叉熵损失函数公式如下：

$$L = \frac{1}{k} \sum_{i=1}^k (-\sum_{j=1}^n y_{i,j} \log p_{i,j}) \quad (4.2)$$

式中： L 为训练损失； k 为批量训练样本大小； N 为样本类别数； $y_{(i,j)}$ 为第 i 个样本在第 j 类上的真实值； $p_{(i,j)}$ 为第 i 个样本对第 j 类的预测值。本文构建的秦简识别模型为多分类模型，因此交叉熵可以简化为它们的求和平均，公式如下：

$$L = \frac{1}{k} \sum_{i=1}^k (-\log p_{i,Y(i)}) \quad (4.3)$$

式中： $Y_{(i)}$ 表示第 i 个样本所属的类别， $p_{i,Y(i)}$ 表示第 i 个样本在所属类别上的预测概率。使用 softmax 函数将识别模型全连接层的输出结果控制在 $[0,1]$ 的概率范围内，softmax 函数如下：

$$p_{i,j} = \frac{e^{l_{i,j}}}{\sum_{m=1}^N e^{l_{i,m}}} \quad (4.4)$$

式中： $p_{(i,j)}$ 为第 i 个样本对第 j 类的预测概率； $l_{(i,j)}$ 为全连接层对第 i 个样本在第 j 类的输出结果； N 为秦简文本的类别数。因此最终本文的秦简文字识别模型所使用的 softmax loss 整理后公式如下：

$$L = \frac{1}{k} \sum_{i=1}^k (-\log \frac{e^{l_{i,Y(i)}}}{\sum_{m=1}^N e^{l_{i,m}}}) \quad (4.5)$$

式中： L 为训练损失； k 为批量训练样本大小； n 为训练集上秦简文本的类别数； $l_{(i,Y(i))}$ 为全连接层对第 i 个样本在所属类别上的预测结果。

4.4.4 识别算法评估指标

本文选取 4 种评估图像识别算法模型优劣的性能指标，分别为准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 F 值(F-measure)，计算公式如下：

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.6)$$

$$Precision = \frac{TP}{TP + FP} \quad (4.7)$$

$$Recall = \frac{TP}{TP + FN} \quad (4.8)$$

$$F\text{-measure} = \frac{2 \times precision \times recall}{precision + recall} \quad (4.9)$$

式中： TP 、 FP 、 FN 和 TN 分别表示真实的正值、错误的正值、错误的负值和真实的负值。准确率是指正确样本数在样本总数中所占的比重，精确率是指在预测为正样本数量中真正的正样本数量所占的比重，召回率是指在所有真正的正样本数量中，所找回的正确样本所占的比值，F 值为精确率和召回率的加权调和平均值。同时为测试模型的轻量化效果，对模型的平均识别耗时、参数量以及模型大小同时进行测试。

4.5 结果与分析

4.5.1 不同 Inception 模块对网络性能的影响

本文首先对在 AlexNet 模型的前 2 层卷积层中融合的 Inception V3 模块做消融实验测试, 前 2 层卷积层中感受野的大小分别设置为 9×9 与 5×5 , 在此条件下分别测试了 R_1 : 不含 Inception V3 模块; R_2 : 只融合 InceptionA 结构; R_3 : 只融合 InceptionC 结构; R_4 : 同时融合 InceptionA 与 InceptionC 结构情况下的识别准确率与损失值。图 4.6 与图 4.7 分别给出了验证集上 4 个消融实验的仿真对比。实验数据结果如表 4.6 所示, R_4 实验下秦简文字图像的识别效果如图 4.8 所示。

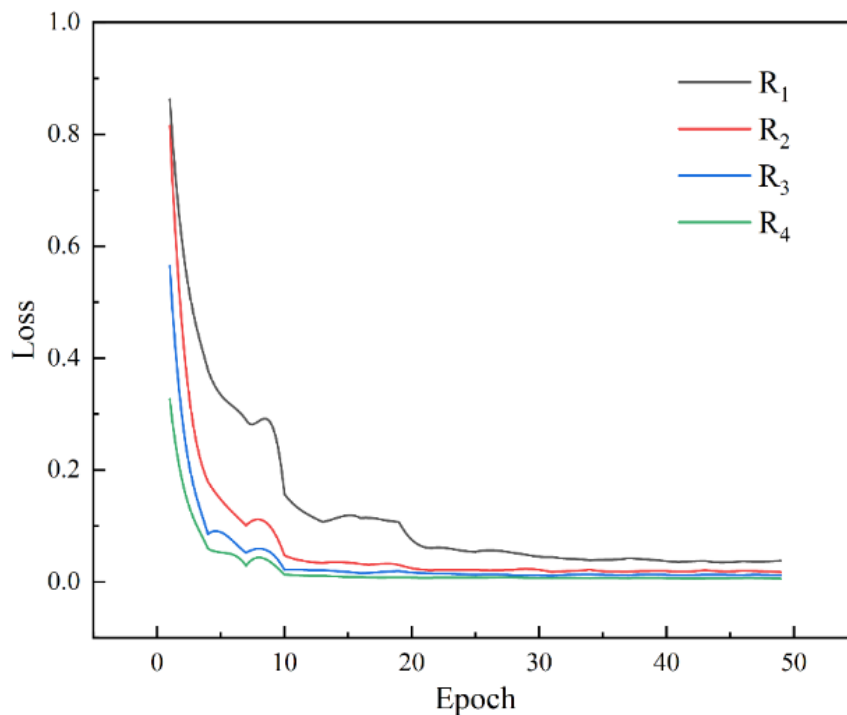


图 4.6 消融条件下的 Loss 值对比

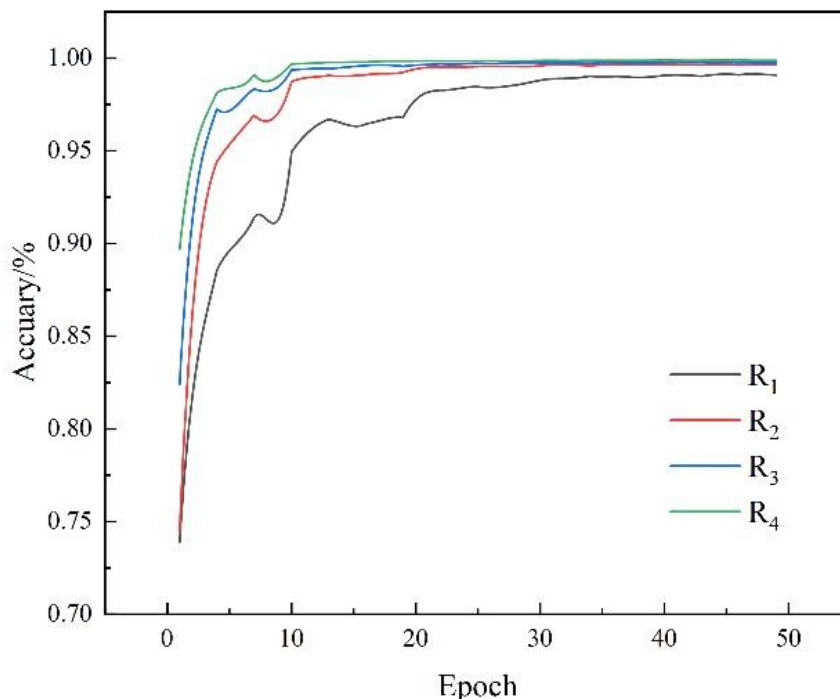


图 4.7 消融条件下的识别准确率对比

表 4.6 不同 Inception V3 模块在验证集上测试结果的均值

网络	InceptionA	InceptionC	Loss	Accuracy/%	耗时/ms	参数量/M	模型大小/MB
R ₁	—	—	0.0342	99.12	659	57	217
R ₂	√	—	0.0187	99.68	426	17	64.8
R ₃	—	√	0.0122	99.80	428	20	76.3
R ₄	√	√	0.0076	99.89	435	23.2	88.5

由仿真对比结果可知，在没有融入 Inception V3 模块的 R1 模型中，在 50 次迭代训练中训练集和测试集的 loss 值收敛性效果不佳，并且在 30 次迭代后准确率才达到一个较好的效果，测试集中 acc 值最高时为 99.12%；在 R2 模型中，第 20 次迭代训练时模型的 loss 值和 acc 值已取得一个较好的收敛性效果，acc 值最高时为 99.68%，但训练集上最初的 loss 值略高于 R1 模型；R3 模型与 R2 模型的仿真结果相似，均在第 20 次迭代训练时便达到较好的收敛效果，但相较于 R2 模型，其初始的 loss 值明显降低，acc 值最高时为 99.80%；在 R4 模型中，第 10 次迭代训练时便取得了非常理想的收敛效果，且最高时的 acc 值达到了 99.89%，相较于前三种对比模型，分别提高了 0.77%、0.21% 与 0.08%，最优时的 loss 值分别降低了 0.0266、0.0111 与 0.0046。

图 4.8 R₄ 模型下的秦简文字图像识别效果

在 R₄ 模型下进行测试的 50 幅秦简文字图像中, 有 49 幅图像的预测字符类别值与真实值相符。在验证模型的轻量化效果上, R₄ 模型的平均识别耗时为 435ms, 较于最长耗时的 R₁ 模型提升了 224ms, 但受模块数影响, 耗时仍比 R₂ 与 R₃ 模型分别高 9ms 和 7ms, 不过识别准确率仍为四者中最优; 参数量与模型大小方面, R₄ 模型相较 R₁ 模型均有大幅度的下降, 但为保证达到最优识别准确率, 并未做进一步轻量化处理, 因此相较 R₂ 与 R₃ 模型, R₄ 模型的权值和复杂度略高。

4.5.2 本文算法与其它识别算法对比

为了进一步评估本文改进后 AlexNet 模型的性能, 因此选取了 10 种经典识别网络模型算法与本文算法作对比。为了有效降低实验误差所产生的影响, 进行了 50 次图像识别实验, 并将 50 次实验结果的平均值作为最终的实验结果。图 4.9 是各个模型之间损失函数值的对比, 表 4.7 是各个算法之间评估指标的对比结果。

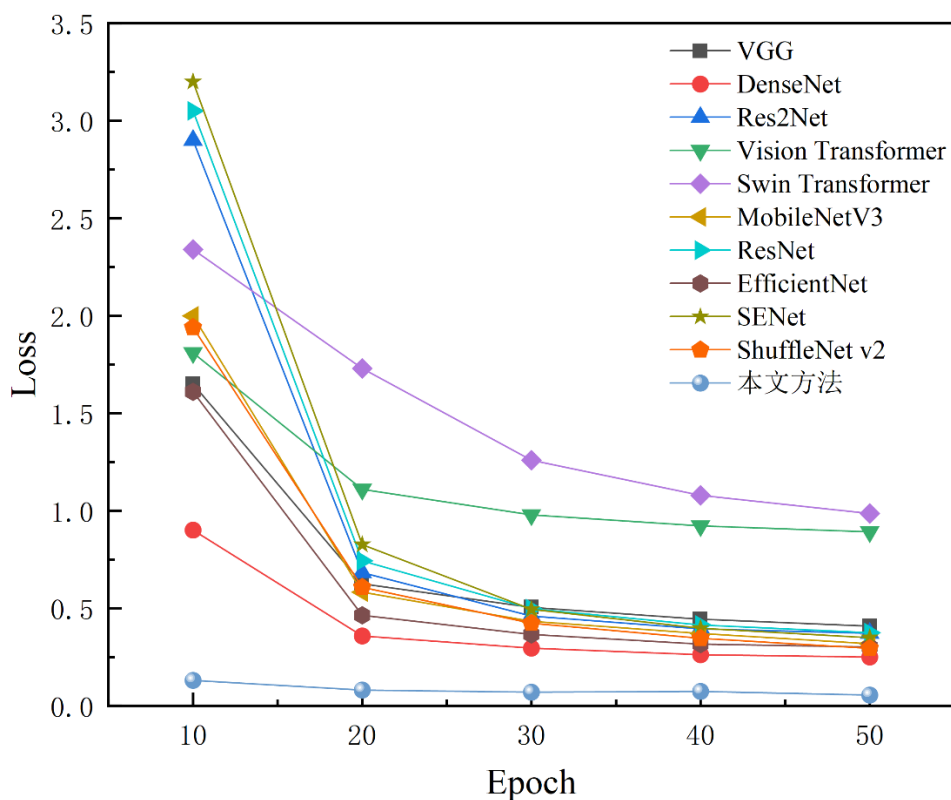


图 4.9 不同模型的损失函数值结果对比

表 4.7 不同模型下文字识别网络评估指标结果对比

模型	Accuracy/%	Precision/%	Recall/%	F-measure/%
VGG ^[36]	94.18	93.42	91.87	92.18
DenseNet ^[37]	98.67	98.33	98.57	98.42
Res2Net ^[38]	99.42	99.24	99.42	99.32
Vision Transformer ^[39]	97.70	97.92	97.42	97.53
Swin Transformer ^[40]	97.25	97.16	96.38	96.70
MobileNetV3 ^[56]	95.67	95.07	94.51	94.55
ResNet ^[57]	99.86	99.79	99.89	99.84
EfficientNet ^[58]	94.67	93.48	92.65	92.58
SENet ^[59]	99.77	99.70	99.74	99.71

ShuffleNet v2 ^[60]	95.62	95.25	93.79	94.06
本文方法	99.89	99.87	99.86	99.86

由图 4.9 结果可知, 本文方法在损失函数值上达到了 10 种对比模型的最优值, 相较于次优的文献[5]方法中的 Loss 值降低了 0.2455。由表 4.7 可知, 在 4 项识别评估指标对比中, 本文方法的 3 项指标达到了最优。在准确率、精确率和 F-measure 值上相较于次优的文献[23]方法分别提升了 0.03%、0.08%和 0.02%, 在召回率上相较于最优的文献[22]方法相差 0.03%。相比于文献[4, 21, 23, 25], 本文方法的各项识别指标均有较大提升。综合考虑, 本文方法在对高信噪比的秦简文字图像进行识别时相较于其他对比模型可以取得较好的仿真效果。

4.6 本章小结

本章证明了所提出的秦简文字识别算法在提高识别率方面具有显著的有效性。通过对比实验结果, 发现采用改进后的轻量级 AlexNet 网络模型对秦简文字进行识别能够显著提高识别率与识别速度。这表明本算法能够有效地提取样本的特征和信息, 减少误差和歧义, 从而提高了识别的准确性和可靠性。同时, 本算法的实现也相对简单和高效, 具有一定的实用性和推广性。因此, 相信本算法能够在秦简文字识别领域中发挥重要的作用, 并为相关领域的研究提供有益的参考。

第五章 总结与展望

5.1 全文总结

本文通过对秦简文字识别问题进行分析和探讨后, 提出了一系列有效的解决方案和算法。研究内容包括: 初步完成了秦简文字样本单字数据集和文本数据集的建立; 对秦简文字图像进行预处理, 并针对图像质量欠均衡的秦简文字图像提出了一种基于图像信噪比自适应阈值模型的二值化算法。经仿真结果与量化指标分析可知, 该算法可以去除秦简文字图像中的大部分噪声和干扰, 使得后续识别模型更加关注核心信息, 从而提高了模型的准确性和鲁棒性; 完成基于深度学习模型的秦简文字识别工作, 包括模型选取、

模型改进和仿真测试。并通过消融实验与对比其他识别模型的仿真结果与量化指标等，验证了本文改进后的轻量级 AlexNet 网络应用在秦简文字图像当中的有效性。

5.2 本论文主要创新点

(1) 自建秦简文字图像样本数据库，填补了当前国内秦简文字数据库缺失的空白。

(2) 针对图像质量不平衡的秦简文字图像提出了一种基于图像信噪比自适应阈值模型的二值化算法。在秦简文字图像样本中的测试结果表明，该算法的二值化结果保留了更多的秦简文字细节特征和文字轮廓，相较其它经典文本图像二值化算法，其性能指标均有明显提升。

(3) 以高信噪比秦简文字图像为研究对象，对 AlexNet 网络进行改进，提出了一种基于轻量级 AlexNet 网络的秦简文字识别算法。模型在具备轻量化特点的同时达到对高信噪比文字图像的高识别准确率，验证了本文方法的有效性。

5.3 后续工作展望

目前秦简文字的数字信息化工作仍处于起步阶段。由于各方面条件限制，论文中的分析与研究还有待进一步提升和完善。具体而言，未来的研究工作可以从以下几个方面入手：

(1) 秦简文字样本数据集的扩充和完善。本文所讨论并建立的秦简文字样本数据集尚处在非常初级的阶段，样本类别不能包含所有秦简文字，当前所有样本都提取自《里耶秦简壹》、《里耶秦简贰》以及《秦简牍合集》在内的 8 本书籍，因此秦简文字的字形风格较为固定，无法体现出多样性。另外当前测试集与验证集中所采用的文件命名格式尚不够成熟，是基于可识别秦简文字的现代汉语拼音字母进行排序编号，随着后续样本库中样本量的增加，要不断完善和优化样本的命名格式，避免在出现同拼音字母的秦简文字时命名格式会冲突的情况，影响后续的认可准确率。同时还将探索更加有效的标注方法和数据处理技术，以进一步提高数据集的质量和可用性，为秦简文字识别技术的研究和应用打下更加坚实的基础。这些工作的完成将有助于扩充和完善秦简文字样本数据集，使其更加全面和准确。

(2) 识别方法的改进。本文模型对于低信噪比及模糊的秦简文字样本图像识别效果

不佳，其泛化能力一般，具体表现在模型在对模糊秦简文字样图像进行特征提取时，获取到的特征图质量较差，所以最终预测值与真实值的偏差较大。因此，需要针对秦简文字的特点，开展更加深入、系统的研究，开发适合不同类别秦简文字的识别算法和模型。同时，也需要建立更全面、更准确的秦简文字样本数据集，以提高秦简文字识别的准确性和鲁棒性。随着研究的深入和技术的不断成熟，相信秦简文字识别技术将会取得更加显著的进展，为秦简学研究提供更加有效和便捷的工具和手段。

(3) 秦简文字注译系统平台的搭建。本文尚未讨论秦简文字注译系统平台的搭建工作，而搭建秦简文字注译系统平台，可以辅助秦简学研究者辨识秦简文字。因此在后续工作中，本文将重点关注搭建秦简文字注译平台的工作。通过该平台，使用者可以借助本系统识别和检索相关的文字，同时也可以将对于这些文本的注释和解读作为样本添加到数据集中。这样使得本文中所讨论的秦简文字样本数据集成为一个动态增长的数据集，其规模和数据量也可以随着应用的广泛而不断增长和丰富。

以上亟待开展的工作表明，秦简文字识别研究仍面临许多挑战和问题，需要进一步开展深入的研究。这些问题涉及到秦简文字的形态变化、噪声干扰、字体多样性等方面，需要通过更加精细的算法和更加完善的数据集来解决。因此，我们需要投入更多的精力和资源，开展更加系统和细致的研究工作，以进一步完善秦简文字识别技术，并推动相关领域的发展。相信在更多研究学者共同不懈的努力下，秦简文字识别技术一定会取得更大的突破和进展，为相关领域的发展做出更大的贡献。

参考文献

- [1] 康晓凤,厉丹,朱安琪.基于边缘GAN的图像修复算法[J].电子器件,2022,45(05):1116-1122.
- [2] Tang X, Luo W, Wang X. Content-based photo quality assessment[J].IEEE Transactions on Multimedia, 2013, 15(8):1930-1943.
- [3] 何鹏浩,余映,徐超越.基于动态金字塔和子空间注意力的图像超分辨率重建网络[J].计算机科学,2022,49(S2):423-430.
- [4] 陈清江,李金阳,屈梅等.基于端到端双网络的低照度图像增强方法[J].计算机工程与科学,2022,44(11):2019-2026.
- [5] OTSU N. A threshold selection method from gray level histograms[J]. IEEE Transactions on Systems, Man and Cybernetics,1979,9(1):62-66.
- [6] LIU N, HAN J.Dhsnet:Deep hierarchical saliency network for salient object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition.2016:678-686.
- [7] RASYIDI H,KHAN S.Historical document image binarization via style augmentation and atrous convolutions[J].Neural Computing and Applications,2020,33(13):7339-7352.
- [8] SAUVOLA J,PIETIKAINEN M.Adaptive document image binarization[J].Pattern Recognition,2000,33(2):225-236.
- [9] Su B, Lu S, Tan C L. Robust document image binarization technique for degraded document images[J]. IEEE transactions on image processing,2012,22(4):1408-1417.
- [10] Moghaddam R F, Cheriet M. AdOtsu: An adaptive and parameterless generalization of Otsu's method for document image binarization[J]. Pattern Recognition,2012,45(6):2419-2431.
- [11] Ntirogiannis K, Gatos B, Pratikakis I. A combined approach for the binarization of handwritten document images[J]. Pattern recognition letters,2014,35:3-15.
- [12] Mitianoudis N, Papamarkos N. Document image binarization using local features and Gaussian mixture modeling[J]. Image and Vision Computing,2015,38:33-51.
- [13] Liao M, Wan Z, Yao C, et al. Real-time scene text detection with differentiable binarization[C]//Proceedings of the AAAI conference on artificial intelligence.2020,34(7): 11474-11481.
- [14] 陶珩,张钦科,刘浩男,等.湘西里耶秦简文字数字化检测技术研究及实现[J].电子制作,2021(18):56-57.
- [15] LU S,SU B,TAN C L.Document image binarization using background estimation and stroke edges[J].International Journal on Document Analysis and Recognition,2010,13(4):303-314.
- [16] 熊炜,王鑫睿,王娟,等.结合背景估计与能量函数的图像二值化算法[J].计算机工程与设

- 计,2019,40(7):1984-1989.
- [17] 冯 炎.基于局部对比度和相位保持降噪的古籍图像二值化算法[J].计算机应用与软件,2022,39(2):162-166.
- [18] HOWE N.Document binarization with automatic parameter tuning[J].International Journal on Document Analysis and Recognition(IJDAR),2012,16:247-258.DOI 10.1007/s10032-012-0192-x.
- [19] RANA A, CHAUHAN K. Computer vision and machine learning for image recognition: A review of the convolutional neural network (CNN) model[J]. Asian Journal of Multidimensional Research, 2021, 10(10):1023-1029.
- [20] SUBEDI B, YUNUSOV J, GAYBULAYEV A, et al. Development of a low-cost industrial OCR system with an end-to-end deep learning technology[J]. IEMEK Journal of Embedded Systems and Applications, 2020, 15(2): 51-60.
- [21] 赵继印,郑蕊蕊,吴宝春,李敏.脱机手写体汉字识别综述[J].电子学报,2010,38(02):405-415.
- [22] 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.
- [23] Yang H, Jin L, Sun J.Recognition of Chinese text in historical documents with page-level annotations[C]//2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR). IEEE,2018:199-204.
- [24] Shi B,Wang X,Lyu P,et al.Robust scene text recognition with automatic rectification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition.2016:4168-4176.
- [25] Wan Z,He M,Chen H,et al.Textscanner:Reading characters in order for robust scene text recognition[C]//Proceedings of the AAAI conference on artificial intelligence.2020,34(7): 12120-12127.
- [26] 李江,冯存前,王义哲,等.基于 AlexNet-BiLSTM 网络的锥体目标微动分类[J].信号处理, 2019, 35(11):1835-1843.DOI: 10.16798/j.issn.1003-0530.2019.11.008.
- [27] 郭敏钢, 宫鹤. AlexNet改进及优化方法的研究[J]. 计算机工程与应用,2020,56(20):124-131.
- [28] ZHANG Chengchang, XU Yu, YANG Jianpeng, et al. Automatic modulation classification based on AlexNet with data augmentation[J/OL]. The Journal of China Universities of Posts and Telecommunications.<https://doi.org/10.19682/j.cnki.1005-8885.2022.0009>.
- [29] 钟桂凤, 庞雄文, 隋栋. 基于Word2Vec和改进注意力机制AlexNet-2的文本分类方法[J]. 计算机科学, 2022, 49(4):288-293.
- [30] 党红恩.藏文字形轮廓特征优化识别提取仿真研究[J].计算机仿真,2016,33(11):341-344.
- [31] 刘卫,李和成.基于多模板归一化的维吾尔文字母识别算法[J].中文信息学报,2016,30(01):156-161.
- [32] 任宏宇,袁保社,田园.基于BP神经网络的维吾尔联机手写字符识别[J].微电子学与计算

- 机,2010,27(08):238-241.DOI:10.19304/j.cnki.issn1000-7180.2010.08.059.
- [33] Su Xiangdong, Gao Guanglai. A knowledge-based recognition system for historical Mongolian documents[J]. International Journal on Document Analysis and Recognition,2016,19(3):221-235.
- [34] Zhang J, Du J, Dai L. Radical analysis network for learning hierarchies of Chinese characters[J]. Pattern Recognition, 2020, 103:1-13.
- [35] MUSHTAQ F, MISGAR M M, KUMAR M, et al. UrduDeepNet: offline handwritten Urdu character recognition using deep neural network[J]. Neural Computing and Applications, 2021, 33(22): 15229-15252.
- [36] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [37] 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.
- [38] GAO Shanghua, CHENG Mingming, Zhao Kai, et al. Res2net: A new multi-scale backbone architecture[J]. IEEE transactions on pattern analysis and machine intelligence, 2019, 43(2): 652-662.
- [39] 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.
- [40] LIU Ze, LIN Yutong, CAO Yue, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 10012-10022
- [41] 郑远攀, 李广阳, 李晔. 深度学习在图像识别中的应用研究综述[J]. 计算机工程与应用,2019, 55(12):20-36.
- [42] 吴峥.《岳麓书院藏秦简》语料库建设暨若干汉字智能识别相关研究[D].湖南大学,2019.DOI: 10.27135/d.cnki.ghudu. 2019.001467.
- [43] ZHANG Diandian, LIU Yan, WANG Zhuowei, et al. OCR with the deep cnn model for ligature script-based languages like Manchu[J]. Scientific Programming, 2021, 2021:1-9.
- [44] 仁青东主, 尼玛扎西. 基于深度学习的自然场景藏文识别研究[J].高原科学研究, 2019, 3(4): 96-103.
- [45] RAHMATI M, FATEH M, REZVANI M, et al. Printed Persian OCR system using deep learning[J]. IET Image Processing, 2020, 14(15):3920-3931.
- [46] SZEGEDY C, LIU Wei, JIA Yangqing, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015:1-9.
- [47] 孟一飞.西夏文字数字信息化若干问题研究[D].北京:北京交通大学,2019.

- [48] ZHU T, GU H, CHEN Z. A median filtering forensics CNN approach based on local binary pattern[C]//Proceedings of the 11th International Conference on Computer Engineering and Networks. Springer, Singapore, 2022: 258-266.
- [49] DING F, SHI Y, ZHU G, et al. Real-time estimation for the parameters of Gaussian filtering via deep learning[J]. Journal of Real-Time Image Processing, 2020, 17(1): 17-27.
- [50] 许 韬. 非线性扩散图像混合滤波去噪方法研究[J]. 计算机仿真, 2020, 37(12): 460-464.
- [51] MUSTAFA W A, YAZID H, ALKHAYYAT A, et al. Effect of direct statistical contrast enhancement technique on document image binarization[J]. CMC-COMPUTERS MATERIALS & CONTINUA, 2022, 70(2): 3549-3564.
- [52] HAN Y, LUO L, XIE B, et al. Nonparametric histogram segmentation-based automatic detection of yarns[J]. Textile Research Journal, 2020, 90(11-12): 1326-1341.
- [53] ASAMOAH D, MARFO R, OPPONG S. Measuring the severity of fungi caused disease on leaves using triangular thresholding method[J]. International Journal of Applied Information Systems, 2017, 12(1): 24-32.
- [54] PARSAI S, AHMADI M. New Local Binary Pattern Feature Extractor with Adaptive Threshold for Face Recognition Applications[J]. International Journal of Artificial Intelligence & Application, 2022, 13(4): 79-87.
- [55] YAZID H, BASAH S N, RAHIM S A, et al. Performance analysis of entropy thresholding for successful image segmentation[J]. Multimedia Tools and Applications, 2022, 81(5): 6433-6450.
- [56] HOWARD A, SANDLER M, GRACE C, et al. Searching for mobilenetv3[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 1314-1324.
- [57] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [58] TAN Mingxing, QUOC L. Efficientnet: Rethinking model scaling for convolutional neural networks[C]//International conference on machine learning. PMLR, 2019: 6105-6114.
- [59] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
- [60] MA Ningning, ZHANG Xiangyu, ZHENG Haitao, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 116-131.