分类号_____ 密级_____

UDC _____ 编号 10736

西北师范大学

硕士学位论文

（专业学位）

基于深度学习的简牍文字识别与检测方法研究

研 究 生 姓 名：胡毓博

指导教师姓名、职称：张强 教授

实践指导教师姓名、职称：苏永恒 高级工程师

专 业 学 位 类 别：电子信息硕士

专 业 学 位 领 域：计算机技术

专 项 计 划：

二〇二四年五月

Research on deep learning character recognition and detection methods for Bamboo and Wooden Slips

A Thesis Submitted to

Northwest Normal University

in partial fulfillment of the requirement

for the degree of

Master of Electronic Information

by

Hu Yubo

Supervisor :（Associate）Zhang Qiang

Supervisor for Practice Guiding: Su Yongheng

May, 2024

摘 要

简牍是秦汉时期的珍贵历史档案，亦是知识宝库，随着文化数字化建设的兴起，现代科技被用于加强古代文献的保护、修复和综合利用。建立可靠的简牍文字识别与检测模型可以帮助研究人员更高效、准确地识别简牍文字。简牍文字与现代手写文字在字的尺度、形态、结构等书写风格上存在明显的差异性，且其因古籍文物属性带来的字迹退化等问题，为简牍文字的识别带来了一定的困难。此外，在整条简的文字检测中，由于其文字大小变化多样、排版复杂多变，使得检测难度亦显著提升。针对上述问题，本文以居延新简作为数据来源，对简牍文字的识别和检测方法展开研究，具体内容包括：

（1）面向字型多变简牍单文字的可变形卷积分类识别模型。针对简牍文字中文字多变的字型、文字尺度和形态多变以及长期的掩埋导致文字图像噪声严重的问题，采用双边滤波降低图像噪声，并引入可变形卷积，构建结合可变形卷积和 Swin Transformer 的简牍文字识别模型 DeConv Swin。利用可变形卷积的非规则采样特性，来解决文字字型和形态多变的问题。Swin Transformer 的层级化特征表达特性，来解决文字尺度变化带来的识别困难的问题。结果表明，简牍文字识别模型相较于单一神经网络对简牍文字的识别精度有所提高，准确率为 83.5%，在一定程度上解决简牍文字中字的字型、文字尺度和形态多变的问题。

（2）面向复杂版面的单简多文字 YOLO 检测模型。针对简牍版面复杂导致的图像中文字大小、位置多变问题和断裂、腐化导致的简牍大小不一的问题，引入可变形卷积，构建结合可变形卷积和 YOLOv8 的简牍文字检测模型 DeConv YOLO。利用可变形卷积的非规则采样和可学习 ROI 特性，得到更准确的简牍文字大小和位置的变化情况；YOLOv8 的多尺度融合结构，来解决单简尺度的不同变化；利用 SIoU 损失函数约束检测模型，得到更准确的文字位置。结果表明，本文提出的 DeConv YOLO 文字检测模型相较于单一文字检测模型或目标检测模型，在对简牍图像中文字位置的检测结果更准确，准确率为 87.90%，在一定程度上解决简牍图像中文字大小和位置多变的问题。

（3）简牍文字识别软件设计与开发。基于上述的识别与检测模型的理论，进行简牍文字识别软件平台的设计和开发工作。该软件平台由四个主要模块构成，各个模块的主要功能分别是数据导入、文字位置检测、检测结果的编辑、检测结果的保存、文字识别结果展示和编辑，文字识别结果的存储。这些模块协同工作，有效

实现简牍文字的识别与检测应用。

关键词：简牍；可变形卷积；图像分类；古文字识别；目标检测；古文字检测

Abstract

The bamboo and wooden slips are precious historical archives from the Qin and Han dynasties and a treasure house of knowledge. With the rise of cultural digitization, modern technology has been used to strengthen the protection, restoration, and comprehensive utilization of ancient documents. Establishing a reliable model for recognizing and detecting wooden slips can help researchers identify them more efficiently and accurately. There are obvious differences between bamboo slips and modern handwriting in scale, shape, structure, and other writing styles, and problems such as handwriting degradation caused by the nature of ancient books have brought some difficulties to the recognition of bamboo slips. In addition, in the text detection of the whole text, the difficulty of detection is also significantly increased due to the variety of text sizes and complex typesetting. Given the above problems, this paper takes Juyan's new script as the data source and studies bamboo slips' identification and detection methods. The specific contents include:

A deformable convolution classification and recognition model for monosyllabic letters with variable font type. In order to solve the problem of text image noise caused by changeable font, text scale and shape, and long-term burying, bilateral filtering is adopted to reduce image noise and variable convolution is introduced to construct the text recognition model DeConv Swin, which combines deformable convolution and Swin Transformer. The irregular sampling characteristic of deformable convolution is used to solve the problem of changeable font and shape. Swin Transformer features hierarchical feature expression to solve the problem of complex recognition caused by changes in text scale. The results show that compared with a single neural network, the recognition accuracy of wooden slips is improved to 83.5%, which can solve the problems of changeable font, scale, and shape in wooden slips to a certain extent.

Simple and multi-text YOLO detection model for complex layout. Aiming at the problem of variable text size and position in images caused by the complex layout of wooden plates and the problem of different sizes of wooden plates caused by breakage and corruption, deformable convolution was introduced, and DeConv YOLO was constructed, which combined deformable convolution and YOLOv8. Using the deformable convolutional irregular sampling and learnable ROI characteristics, we can

get more accurate changes in the size and position of wooden texts. YOLOv8 multi-scale fusion structure is used to solve the different changes of simple scale; a more accurate text position is obtained using SIoU loss function constraint to test the model. The results showed that the DeConv YOLO text detection model proposed in this paper was more accurate, with an accuracy of 87.90% than the single text detection model or the target detection model, which solved the problem of changeable text size and location in the bamboo Slips images to a certain extent.

Design and development of bamboo and wooden slip recognition software. Based on the above recognition and detection model theory, we design and develop a software platform for recognizing bamboo slips. The software platform is composed of four main modules. The main functions of each module are data import, text position detection, detection box result editing, and text recognition. These modules work together to realize the recognition and detection application of bamboo slips effectively.

Keywords: Bamboo & Wooden Slips; Deformable Convolution; Image Classification; Paleographic Recognition; Object Detection; Paleographic Detection;

目 录

第1章 绪论

1.1 研究背景和意义

1.1.1 研究背景

竹木简牍是秦汉时期留下的宝贵档案，它们全面记载那一时期的社会文明与文化发展。这些历史遗产不仅承载两千年前的文献资料，还反映古代社会的伦理观念、宗教信仰、交通运输、饮食传统、医学知识以及书法艺术等多个层面。竹木简牍被认为是中古时期中国的知识宝库，具有极高的研究价值。习近平总书记在考察中国人民大学时，强调应用现代科技强化古籍典藏保护、修复及其综合利用[1]。在哲学社会科学座谈会上，他提出中国古代文献蕴含治国智慧，为先人理解与改变世界提供基础。简牍等古文献也贮藏丰富的治国知识。2022 年 5 月，中共中央办公厅和国务院办公厅联合发布《关于推进实施国家文化数字化战略的意见》[2]，标志着我国文化数字化建设的重要发展机遇。文化计算作为新动力，对建设文化强国至关重要。在此背景下，以适应时代的需求，要求运用现代科技手段加强简牍古籍典藏的保护修复和综合利用。

在古代，简和牍是在纸张未发明和未普遍使用之前用于书写文字的主要材料。简牍作为我国古代重要的文字载体，承载着丰富的政治、经济、文化和社会等方面的信息，是我国古代文化的重要组成部分[3]。近年来，对出土的简牍进行整理与研究已经在文字学、书法学和史学领域取得重要的突破和进展。通过多学科的研究合作，推动简牍学的深入发展，内容涵盖文字的考释、书学价值的挖掘以及古史信息的开掘等多个方面。甘肃省以其大量出土汉代竹简而著称，占全国 80%以上，被誉为“汉简之都”[4]。这些竹简的学术价值极大，对于研究汉代的文化、国际关系、民族融合、民俗、邮政、丝绸之路贸易及古环境等多个领域提供珍贵的一手资料。作为古代文化遗物，出土的简牍不仅是宝贵的历史文献，也是考古发掘出土的珍贵文物。它们对于了解古代社会、文化和历史具有重要意义。

但是，简牍的研究一直以来都依赖于对古文字有专业了解的学者。每个文字的识别都需要进行繁琐的检索过程，这增加了人工识别的工作量，并给非专业人员的学习带来困难。由于古文字的复杂性和多样性，对于非专业人员来说，准确地辨认和解读简牍中的文字是一项艰巨的任务。近期，人工智能尤其是深度学习的快速进展，为古文字识别提供新的视角。深度学习通过分析庞大的数据集，能自动辨识文

字的形状、结构和语义，从而实现高效精确的文字识别。在汉字识别领域，已有学者利用深度学习技术在手写汉字识别方面取得显著的成果，这表明将人工智能应用于古文字形体识别是可行且具有巨大潜力的[5]。通过训练深度学习模型，可以使其具备对古文字形态的识别能力，辅助研究人员进行简牍的识别工作。

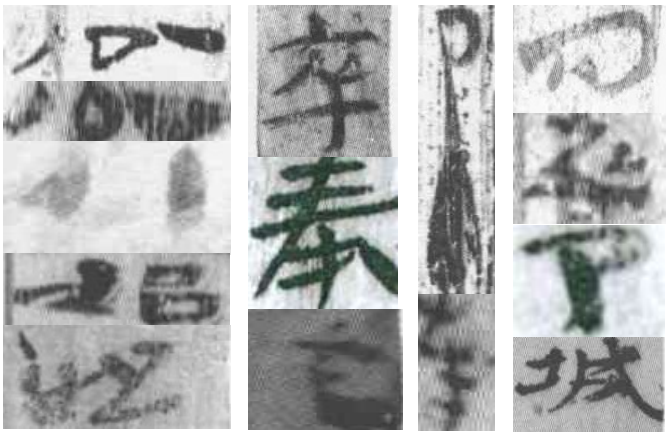


图 1-1 部分文字样本与简牍样本示例

这种技术的应用将大大提高减轻研究人员的工作负担，同时也为非专业人员提供更便捷的学习途径，让更多人能够积极参与到简牍研究中来。由于简牍文字与现代手写体文字存在着明显的区别，如图 1-1，如文字尺度、形态和结构等方面的差异，因此将深度学习应用于古文字形体识别时，需要充分考虑这些特点。本文针对简牍文字尺度、形态和结构的特点，将可变形卷积与 Swin Transformer 结合实现对简牍文字的识别。但要将深度学习应用于简牍文字识别，仍需专业人士的参与和指导。他们的专业知识和经验帮助训练和优化深度学习模型，确保识别结果的准确可靠。通过深度学习在古文字识别领域的应用，加快简牍综合利用的进展，促进古代文化的传承与发展。为研究人员提供更多宝贵的研究素材，推动古文字领域的深入探索，这也为研究者更深入地了解汉代社会、文化和历史等方面提供新的途径。

图 1-2 居延新简样本示例

本文针对简牍版面混乱、文字模糊不清、字体繁乱等特点，如图 1-2，结合可变形卷积与 YOLOv8 模型，实现简牍文字检测模型。该模型能够快速准确地检测出简牍图像中文字的位置，而且整个过程仅需要几毫秒的时间。通过将深度学习技术应用于简牍文字识别，大大减少研究者的参与，有效地提高古文字研究工作的效率和准确性。传统的文字识别方法需要依赖专业人员进行繁琐的检索和分析，而深度学习模型可以在短时间内自动完成文字的定位和识别，极大地减轻研究者的工作负担。此外，深度学习模型能够自动学习和适应不同的文字形态和特征，从而准确地识别出简牍图像中的文字内容。这为研究者提供更可靠和准确的文字释文，推动古文字研究的深入发展。

综上所述，通过将深度学习技术应用于简牍文字识别，可以快速准确地检测出文字的位置，并提供文字的释文识别结果。这种方法大大提高古文字研究的效率和准确性，为研究者们提供更多宝贵的研究素材和线索，推动古代文化的传承与发展。

1.1.2 研究意义

本文聚焦简牍文字释文解译过程中的文字识别与检测的现实问题，通过运用深度学习技术，利用图像识别和目标检测算法对研究对象进行建模分析，实现准确，高效的文字识别和检测工作。为简牍文字研究提供一种高效识别与检测方案。应用深度学习技术于简牍文本的识别和研究在实践和学术上至关重要。它对文化遗产保护、历史研究的深化和工作效率及准确性的提升极为关键。科学视角下，深度学习技术在人文科学的应用推动新技术的创新和学科交叉的深入合作，意义重大。

从现实意义的角度来看，简牍文本是人类文化遗产的核心组成部分。准确研究和识别这些文本对文化遗产的保护、传承、展示和传播至关重要，为这些领域提供新的方法和工具。这不仅有助于保存历史，还增进研究者对人类历史和文化演变的理解。研究简牍文字对于历史学、考古学和文化人类学等学科领域具有重要意义，通过深度学习技术的应用，可以提供更多准确的研究素材和线索，推动历史研究的深入发展。此外，利用深度学习技术的文字识别模型能够自动完成文字的定位和识别，大大提高研究的效率和准确性，减轻研究者的工作负担，也促进研究成果的传播和共享。

从科学意义的角度来看，通过将深度学习技术应用于简牍文字的识别和研究，可以推动深度学习在人文领域的应用，探索更多人文领域的研究问题和挑战。研究简牍文字的释文识别和应用深度学习技术，为其他领域的图像识别和文字识别研究提供借鉴和启示，促进相关技术的发展。此项研究的进行必须借助多学科领域的知识体系和研究方法，通过跨领域的合作与沟通，加速不同学科的整合与创新。

1.2 国内外研究现状分析

1.2.1 图像识别算法研究现状

在 20 世纪 80 年代，多层感知机（Multilayer Perceptron，MLP）的发展标志着计算机数字识别能力的显著提升。然而，受限于早期计算机的处理能力，特别是 CPU 和内存资源的限制，早期 MLP 只能应用于小规模数据集。这限制它们的表征能力，难以处理复杂的图像识别任务。为克服早期 MLP 的局限，Geoffrey Hinton 等人[6]引入逐层预训练。这种方法通过无监督学习初始化网络权重来增强网络特征提取能力，采用贪婪逐层训练策略：首先独立训练每一层网络以重建输入数据，再将这些层合并成深度网络。每层作为自编码器或受限玻尔兹曼机训练后，能将高维输入简化为低维表示，减少维度并提取有用特征，该方法使深层网络能在不被复杂数据淹没的情况下，学习更抽象的表示，为深度学习发展铺路，让训练深度神经网络变得可行，并在多个领域实现重大突破。21 世纪初深度学习崛起，卷积神经网络（Convolutional Neural Networks，CNN）在计算机视觉领域成为主流。大型科技公司在深度学习系统的开发上投入巨资，推动算法的创新和技术的商业化。他们的贡献不仅在于算法，还包括构建大规模数据集和提供计算资源。CNN 的训练依靠反向传播[7]算法，这是一种有效的优化方法，它通过计算损失函数对网络权重的梯度来更新权重，以此最小化损失。反向传播利用链式法则和梯度下降策略，能够从输出层开始，逐层反传误差信息，持续调整各层权重。

LeNet[8]卷积神经网络显示通过训练深层神经网络来自动从图像中提取特征的潜力，并且在 MNIST 手写数字识别数据集上取得很好的效果；Alex Krizhevsky 等人在 2012 年提出 AlexNet[9]，该模型采用 ReLU 激活函数，使用 Dropout 减少过拟合，同时引入数据增强和局部响应归一化技术，共有五个卷积层和三个全连接层，显著提升大规模视觉识别任务的性能；2013 年 Matthew Zeiler 和 Rob Fergus 提出 ZFNet[10]通过调整卷积层中的过滤器大小和步长，以及增加中间层的可视化，来优化网络结构，提高图像识别的准确率；2014 年牛津大学的视觉几何组提出 VGGNet[11]，该模型的关键特点是其使用多达 19 层的 3x3 卷积核和 2x2 的池化层，通过重复的堆叠简化网络结构，同时深化网络以提升性能；同年谷歌公司发布 Inception 模型[12–15]，并且之后经过不断通过改进网络结构来优化性能和效率提升模型的准确率和泛化能力；2015 年何凯明等人提出 ResNet[16]通过使用跳跃连接（Skip Connection）或短路（Shortcut）直接将输入添加到后续层，使得网络可以有几十甚至上百层，大幅提升图像识别和分类的准确性；2017 年 Gao Huang 等人提出 DenseNet[17]其特点是在每个层之间都建立直接连接，与传统的卷积网络不同，

每个层都接收到之前所有层的特征图作为输入，这种密集连接机制强化特征的传递，提高效率，减少参数数量，并且通过特征重用降低网络的冗余；同年 Andrew G. Howard 等人提出 MobileNet[18–20]采用深度可分离卷积来显著减少计算量和模型大小，同时仍保持相对较高的准确性；之后 Jifeng Dai 等人提出可变形卷积网络[21]（Deformable Convolutional Networks，DCN）通过在标准卷积的基础上引入额外的可学习偏移量，增强卷积神经网络对几何变换的建模能力，有效改善对象检测和语义分割等视觉任务的性能；2018 年 Xizhou Zhu 等人提出 DCNv2[22]，在 DCNv1 的基础上增加调制机制与特征模仿训练，在可变形卷积中加入调制机制，其每个采样点不仅会学习到的偏移，并通过学习到的特征幅度进行调制，使得网络模块能够变化样本的空间分布和相对影响力；为有效利用增强的建模能力，DCNv2 受到知识蒸馏的启发，使用教师网络来提供训练期间的指导。

研究者发现，应用于自然语言处理领域的 Transformer[23]模型经过适当调整后也能够应用于图像任务中，并且在部分任务中表现出优于 CNN 模型。2020 年 Alexey Dosovitskiy 等人提出视觉 Transformer[24]（Vision Transformer，ViT），通过将图像切分为 16x16 像素的块，直接使用 Transformer 架构处理，无需卷积网络，在大规模预训练后，在多种图像识别数据集中达到或超越其他算法，但自注意力存在很大的计算冗余，当序列长度增加时其计算量爆炸式增长；2021 年 Ze Liu 等人提出 Swin Transformer[25]模型，Swin Transformer 相较于 ViT 大大减少计算冗余，利用移位窗口，实现图像尺寸的线性计算，并且采用分层架构，具有较好的灵活性，能够在不同的尺度上建模，有效适用于图像分类、目标检测和语义分割等计算机视觉任务；2022 年 Zhuang Liu 等人提出 ConvNeXt[26]模型通过优化卷积层的设计，引入类似于 Transformer 的层归一化和残差连接等，增强训练稳定性并提高性能。

1.2.2 目标检测算法研究现状

目前，一些目标检测算法有着很高的精度。R-CNN[27]是深度学习在目标检测领域的开山之作，R-CNN 通过结合区域提议方法和 CNN 网络，显著提高目标检测的精度，引入从图像中识别出潜在的兴趣区域的机制，独立地利用卷积神经网络提取特征和分类这些区域，从而检测出图像中的对象；随后 SPP-Net[28]提出，空间金字塔池化（Spatial Pyramid Pooling，SPP）是该算法的一个关键创新，它允许网络处理任意尺寸的输入图像，而无需修改网络架构，这解决 CNN 的一个重要限制，CNN 需要输入图像具有固定的尺寸，主要原因是全连接层在接受前一层的输出时需要一个固定长度的特征向量；Fast R-CNN[29]实现端到端的训练，通过梯度回传机制训练两阶段检测模型，取消对中间特征存储的需求，它利用边界框回归精确定

位目标，并替换支持向量机（Support Vector Machine，SVM）为 Softmax 分类器，进一步简化训练过程并提高目标检测的效率和精度；之后，Faster R-CNN[30]抛弃选择性搜索，改为区域建议网络（Region Proposal Network，RPN）实现建议框的快速、高效生成，解决性能上限，RPN 与 Fast R-CNN 共享卷积特征，提高处理速度，同时通过锚点策略生成多尺度、多宽高比的建议框；此外，2017 年 Lin T.-Y.等[31]提出特征金字塔网络（Feature Pyramid Network，FPN），FPN 通过自上而下的架构和侧向连接，巧妙整合高阶语义和低阶细节，显著提升跨尺度的特征表征能力，这使得 FPN 在处理不同尺寸的目标时表现突出，虽然 FPN 提高小目标检测的准确性，但其模型复杂度和计算消耗也随之增加，可能影响模型的效率；为充分利用特征提取网络中的特征图信息，在 2018 年 Shu liu 等[32]，提出路径聚合网络（Path Aggregation Network，PANet），PANet 增加自底向上的信息路径，提升低级特征向高层的信息流动，增强特征的利用效率和多尺度表示能力，引入自适应特征池化，促进不同尺度特征的整合，进一步改进小目标检测性能。

2016 年 Redmon J 等，提出 YOLOv1[33]引领单阶段目标检测算法的潮流，有效地折衷检测速度和精度，相比于 Fast R-CNN[29]依赖区域建议算法，Faster R-CNN[30]采用 RPN 来改进这一过程，与 Faster R-CNN 不同，YOLOv1 将目标分类和边界框回归这两个阶段融合为一步，提升处理速度。在处理每张待检测图像时，YOLOv1 将图像划分为$k\times k$ 个网格，每个网格负责预测多个边界框，每个边界框由五个参数定义，四个用于确定边界位置，一个反映物体的置信度；在 2017 年 Redmon J 等，提出的 YOLOv2[34]，YOLOv2 更换特征提取网络，使用 DarkNet-19 作为其特征提取网络，引入批量归一化来提高收敛速度并减少对其他形式正则化的依赖，并借鉴 Faster R-CNN[30]，使用锚点框来预测边界框，提高对不同尺寸和比例物体的检测精度，还采用更好的分辨率输入和细粒度特征，以及新颖的多尺度训练策略；YOLOv3[35]采用多尺度预测和 Darknet-53，一个更深、更强的卷积网络作为特征提取器，以提高对小尺寸目标的检测能力重新设计出特征提取能力更强的网络。该网络使用 53 个卷积层，故命名为 DarkNet-53，YOLOv3 还引入三个不同尺度的特征图来检测大、中、小尺寸的物体，增强模型的尺寸适应性。同时，它通过使用逻辑回归来预测物体类别，而非 Softmax 方法，改善模型对多标签分类的处理；2020 年 Bochkovskiy Alexey[36]提出 YOLOv4，采用 CSPDarknet53 作为更强大的骨干网络，使用 Mish 激活函数，引入新的正则化方法 Cross mini-Batch Normalization（CmBN），Self-Adversarial Training（SAT），以及 Mosaic 数据增强来提高模型的泛化能力，YOLOv4 还融入 SPPNet[28]和 PANet[32]结构来改善特征提取，以及 CIOU Loss[37]来提高边界框回归的准确度；同年 Ultralytics 团队发布 YOLOv5，Ultralytics 对模型

结构和训练流程进行优化，使其在目标检测任务上实现快速且准确的性能；YOLO 系列的目标检测算法在后续又相继发展出 YOLOX[38]，YOLOv6[39]，YOLOv7[40]和 YOLOv8，其中 YOLOv8 在之前成功的 YOLO 版本的基础上，引入新的骨干网络和新的 Ancher-Free 检测头，但是其对于小目标的检测效果不佳。

在 ViT[23]提出之后相关研究者将 ViT 应用于目标检测任务，2020 年 Carion N 等[41]提出基于 Transformer 的端到端目标检测模型 DETR，取消目标检测的后处理和先验知识的束缚，简化目标检测过程；2021 年 Fang Y 等[42]针对视觉变换器在视觉任务中是否能进行最小解 2D 空间结构的对象和区域级别识别的问题，提出 YOLOS 算法，该算法基于原始的 Vision Transformer 架构，通过最少的修改实现在 COCO 对象检测数据集上最优性能；之后 2022 年 Li Y 等[43]针对如何将原始的 ViT 用作目标检测的骨干网络这一问题，提出 ViTDet 算法，通过构建简单的特征金字塔以及使用窗口注意力机制，ViTDet 在 COCO 数据集上达到较好的性能。

1.2.3 古文字研究现状

传统的古文字识别方法主要采用提取文字的拓扑特征和图像特征的方式进行识别[44–48]。这些古代文字识别方法主要集中于提取文字的关键特征，并尝试使用这些局部特征来表示文字的整体特征。通常情况下，这些研究所依赖的数据集主要由手工摹写的古代文字构成，而不是直接从古代文字图像中提取特征进行识别。这种方法的一个显著限制是其依赖于手工特征的选择和提取，这可能无法全面捕捉文字的所有细节和变化。因此，传统的古代文字识别技术往往在识别性能上无法达到理想的水平。

基于深度学习的计算机视觉研究为古文字识别带来新的解决思路，2020 年田园[49]为解决战国时期竹简文字识别难题，研究者建立名为 Bambooslips 的数据集，并在此基础上提出一种识别算法，该算法适用于小样本学习，通过孪生网络进行深度度量学习，并结合数据增强手段扩充样本量，提升对高度相似战国简文字的识别准确率；同年刘梦婷[50]提出基于改进 AlexNet 的甲骨文字识别方法，专门针对拓片上手写体甲骨文的特点设计，采用条形卷积核替代部分方形卷积核，并通过堆叠特征图来加深网络深度并减少参数，从而更有效地提取甲骨文字特征，改进后的网络模型识别效果明显优于传统方法；2021 年张颐康等[51]为提升甲骨文拓片的识别准确性，引入基于跨模态深度度量学习，利用高品质临摹甲骨文样本进行训练，建立一个能表征临摹与拓片文本的共享特征空间，实现拓片甲骨文的跨模态识别，通过最近邻方法在共享空间中分类，最终在甲骨文识别上优于传统单模态技术，并有效识别新类别甲骨文；2021 年 Wulingjing 等[52]针对先秦文字的自动识别问题，提出

一种结合注意力机制的识别算法，该算法替换传统的卷积操作，并为古汉字图像设计专门的数据增强方法，尽可能保持汉字书写形式的同时增加数据多样性；同年林小谕等[53]针对甲骨文中的字形结构多样性和异体字识别难题，提出基于深度学习的 BN-LeNet 和 OraNet 模型，这些模型通过最大极值稳定区域选取与迁移学习相结合的方法，显著提高甲骨文单偏旁及合体字的识别准确率，有效解决甲骨文识别的精确性问题，由于通过拼接甲骨文偏旁生成大量的甲骨文合体字用于模型的训练，导致模型的泛化能力较差；2022 年朱旭[54]针对拓片甲骨文样本少、类间不平衡和标识获取难的问题，提出基于改进的 PUGAN 和 CNN 模型的识别算法，该算法通过实现手写甲骨文到拓片甲骨文跨域自适应，实现对拓片甲骨文的有效识别；同年吴炫奇[55]研究在 VGG、ResNet 等算法中，通过在模型中引入注意力机制，即在网络的数据输入层集成一个注意力空间域的空间变换网络，可以进一步增强模型的性能；2022 年石佳钰[56]，针对手写蒙古文字元识别的问题，提出基于生成对抗网络的方法，解决数据集规模小和多样性差，以及手写文字风格多变对识别准确性影响的问题；同年刘绪兴[57,58]针对小样本条件下古文字的高效识别问题，提出基于多尺度特征融合的异体字识别网络和基于深度度量学习的小样本识别方法，提高对甲骨文、金文在内的古文字的识别准确率，解决传统方法在样本数量稀少和异体字字形变化大的情况下识别效果不佳的问题；之后 Assael Yannis 等人[59]提出 Ithaca 的深度神经网络，用于恢复古代希腊铭文，以及确定这些铭文的地理和时间属性；2023 年李沿增[60]针对古文字识别中的长尾效应、样本不足和字符复杂性问题，提出融合目标检测和知识图谱的模型，该模型通过细粒度部件识别和知识推理，提高甲骨文等古文字的识别准确率，有效解决样本不平衡问题；同年郝超华[61]针对西夏文字的识别问题，提出基于无监督双视图对比学习和无监督 Transformer 的算法，解决西夏文字数据集难以标注、相似文字识别准确率低以及残缺文字识别准确率不高的问题；毛亚菲等[62]针对甲骨文图像识别，特别是拓片中甲骨文字的微小局部特征提取难题及高相似度文字识别准确率不高的挑战，提出一种基于优化的 ResNeSt 网络算法，该算法通过融入跳连结构和坐标注意力机制，和优化分类器，有效增强模型对细节特征的感知能力和对相似文字的辨识精确度。

在传统方法中古文字检测通常使用人为设计的特征提取方式进行文字特征提取，史小松[63]使用基于阈值的连通区域方法检测甲骨文，这方法在减少手工标注误差方面有所帮助，但在背景复杂或噪声较大的拓片中效果受限；潘振赣[64]运用模糊 C 均值（Fuzzy C Means，FCM）算法进行拓片图像的分割；王书敏[65]综合连通区域分析、边缘检测和纹理分析等技术定位古文字，但甲骨文的独特形态和复杂背景限制该方法的适用性。随着深度学习技术的快速进步，在场景文字检测领域得

到广泛的应用[66–70]。一些研究者已经开始探索如何使用深度学习技术来检测文物图像中的字符[71,72]。这些研究尝试将深度学习模型应用于古文字的自动识别和定位，以期望能够处理复杂的背景和文字形态，提高检测的准确性和效率。古代文字检测面临着与现代手写文字检测截然不同的挑战。现代文字通常遵循一定的书写规范，而且相关的数据集数量庞大。相反，出土文献如简牍中的古文字，其尺寸、形状和位置各异，且在文档中的分布极其不规则。此外，文字种类繁多，但每种文字出现的频率很可能极不平衡，导致类别间不均衡，针对这些问题，传统的文字检测方法可能无法有效应对古文字的多样性和复杂性，更倾向于采用目标检测和文字检测算法。2019 年 Lin Meng[71]为从拓印图像中分割和识别甲骨文，通过改进单次多框检测器算法以增强其对甲骨文字符的检测能力；2020 年王浩彬[72]设计基于区域的全卷积网络和特征金字塔网络的甲骨文字符检测框架，为解决复合字符难检测问题，发展动态数据增广算法，提出辅助的字符识别算法，提升模型性能，但是并不能正确检测到一些小且易混淆的甲骨文；2020 年陈善雄[73]等针对彝文古籍文献中字符检测的复杂性问题，提出一种基于最大极值稳定区域和卷积神经网络的字符检测方法，该方法通过预处理、二值化、文本区域提取和字符检测的流程，有效地分离文本与非文本区域，在单字符检测中获得较高的准确率和召回率，解决彝文古籍字符识别的检测问题；2020 年邢济慈[74]针对从甲骨拓片中定位每一个字符的检测问题，提出 SPPG-YOLO 和 ASPP-YOLO，解决甲骨文字自动检测的精确性与效率低下的问题；2021 年刘芳[75]等针对甲骨文拓片的自动检测与识别问题，该算法通过三元组损失函数和旋转角度回归技术的优化，提出基于 Mask R-CNN 的改进算法；同年殷航等[76]针对自然场景中中文文本的倾斜、模糊和光照等检测难题，提出基于 YOLOv3 和 MSER 的 YOLOv3-M 算法，该算法改善 YOLOv3 无法检测倾斜目标的不足，并优化 MSER 容易受复杂场景干扰的问题，实现更准确和快速的中文文本检测；2023 年李健昱[77]等针对复杂纹理背景下骨签文字特征提取困难以及文字密集粘连导致的检测框冗余问题，提出一种融合自注意力卷积和改进损失函数的检测算法。该算法通过在 YOLOv5 特征提取时引入自注意力卷积模块增强对文字特征的识别，并使用 Focal-EIOU 损失函数优化，有效解决密集粘连文字的检测框冗余问题。

从以上相关研究来看，虽然目前对于古文字识别与检测的智能应用和理论技术处于快速发展阶段，深度学习、图像识别、目标检测等技术得到很好的应用，但是对于汉代简牍的相关研究较少，并且存在以下问题需要考虑和解决：

（1）简牍文字形态、尺度和结构多变。简牍文字与现代手写体文字识别和古代摹本文字识别有着显著的差异。简牍文字的特殊性源于其所处的年代和载体，导

致同一文字出现多种书写方式。此外，书写载体的大小和单个载体的篇幅限制也导致文字形态、尺度和结构的多样性。长期的掩埋使得文字墨迹经历退化甚至消失，同时也引起载体性质和形态的变化。最终采集到的影像存在许多噪声，并且文字也发生变形。需要一种针对上述问题的简牍文字识别模型实现对文字的自动识别。

（2）简牍文字版面混乱，导致文字的大小不一且位置多变。长期的掩埋简牍断裂，腐化导致简牍大小不一，这使得直接应用目标检测模型或文字检测模型容易导致文字漏检和错检的问题。因此需要研究面向上述问题的简牍单简文字检测模型，简牍图像中文字密集排列导致模型在检测文字时，不仅需要考虑检测文字的准确性，还要考虑检测区域内模型对文字捕获的完整性。

为解决上述问题，本文将图像识别、目标检测应用到汉代简牍文字识别与检测研究中，构建基于深度学习的居延新简文字识别与检测方法，为简牍文字的数字化保护和利用提供新的途径。

1.3 主要研究内容

适应时代的需求，要求研究者运用现代科技手段加强简牍古籍典藏的保护修复和综合利用。本研究针对简牍文字尺度、形态和结构等方面变化大和简牍复杂版面，文字的大小不一且位置多变等问题，提出基于深度学习的居延新简文字识别与检测方法研究。

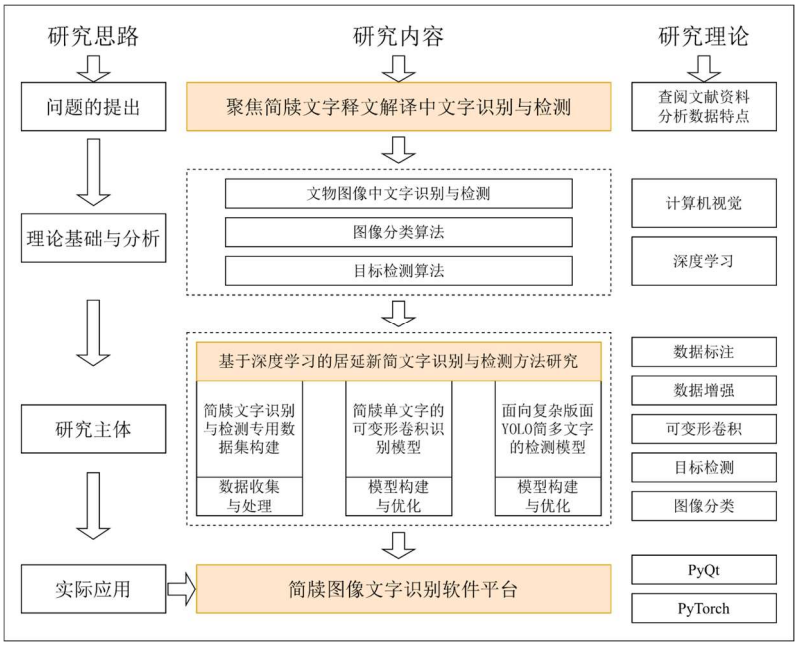


图 1-3 研究技术路线图

按照提出问题-理论研究-实际应用的研究思路，将深度学习，图像分类，目标

检测模型算法引入简牍文字识别与检测，通过研究方案、简牍文字识别与检测数据集构建、简牍文字识别与检测算法的研究和简牍文字识别与检测软件平台的开发等工作，可以推动深度学习在人文领域的应用，探索更多人文领域的研究问题和挑战，推动古文字研究的深入发展。本文的研究路线图如图 1-3 所示。

具体的研究内容包括以下三个部分：

（1）面向字型多变简牍单文字的可变形卷积分类识别模型

简牍文字的特殊性导致同一文字有不同写法，尺度、形态和结构多变。现有古文字识别方法大多采用基于卷积神经网络的深度学习模型，但普通卷积神经网络的平移不变性无法满足简牍文字的多样性和载体形变的问题。考虑到标准卷积神经网络受限于其固有的稀疏连接和固定的局部感受野，难以根据图像内容动态调整感受野的大小，本文提出可变形卷积和 Swin Transformer 简牍单字识别模型 DeConv Swin。该模型特别设计以适应文字的多样形态、结构、变形和尺寸。具体来说，该网络具备以下特征：采用多尺度或自适应感受野策略，有效处理包括旋转和仿射变换等各类几何变形。此外，可变形卷积技术已在目标检测和图像分类任务中显示出显著效用，它能够灵活适应文字的形态变化、结构特征和变形情况。此外，Swin Transformer 的层级特征表达结构也能很好地适应简牍文字的尺度变化。这种模型能够有效识别简牍文字并适应其特殊的形态和结构变化。

（2）面向复杂版面的单简多文字 YOLO 检测模型

在深度学习技术成熟之前，古文字与计算机的研究通常采用手工特征或机器学习方法。然而，这些方法在处理大规模样本训练时效果有限。随着目标检测算法的发展，一些学者开始尝试使用深度学习技术来检测古文字。但是，古文字的检测面临特殊挑战，与现代文本检测有所不同。简牍文本版面复杂，文字大小和位置多变，直接应用目标检测模型会导致漏检和错检。为解决这个问题，本研究提出基于可变形卷积的 YOLOv8 单简多字检测模型 DeConv YOLO，适应文字的多尺度变化和位置多变。可变形卷积层能够自适应地调整卷积核形状，以应对古文字的多样性和不规则性。本文引入可变形卷积层并结合 YOLOv8，并调整损失函数，以更好地处理简牍文字的大小和位置变化。这种模型能够有效地检测古文字，并适应其特殊的形态和结构变化。

（3）简牍文字识别软件设计与开发

本文以减少对简牍文字识别过程中人工识别的工作量的实际需求为导向，基于上述研究内容中提出的基于深度学习的简牍文字识别与检测模型，设计开发一款简牍文字识别软件平台进行实际应用。该软件平台包括数据导入、文字检测、检测框修改和文字识别等四个主要模块，通过这些模块的配合，实现对简牍文字的识

别和检测的实际应用。

1.4 论文结构安排

简牍文字的识别与检测相较于现代汉字的相应研究，面临更复杂的挑战。因此，设计不需要专业知识依赖的文字识别与检测模型具有重要意义。本文简牍文字识别与检测中主要存在以下难点：简牍文字尺度、形态和结构等方面变化大和简牍上的文字规整度较差，文字的大小不一且位置多变等问题。本文针对以上难点提出相应的解决办法，为后续相关研究提高有效的宝贵经验。本文的结构安排如下：

第 1 章 绪论。本章首先全面阐述简牍文字识别与检测的研究背景及其重要性。并概要介绍图像分类、目标检测算法，以及国内外古文字识别与检测的研究进展。以及总结本文的研究内容及其组织架构。

第 2 章 面向字型多变简牍单文字的可变形卷积分类识别模型。本章针对简牍文字尺度、形态和结构多变的特点，提出一种可变形卷积分类识别模型。首先介绍本文构建的简牍文字识别数据集的过程，并详细说明数据的预处理和增强方法。其次，对居延新简文字识别模型的关键模块进行详细介绍，包括特征提取、可变形卷积和分类器等。最终，通过大量的实验验证本文提出的模型的有效性和识别性能。

第 3 章 面向复杂版面的单简多文字 YOLO 检测模型。本章主要解决简牍文字规整度差、大小和位置多变的问题，提出基于 YOLOv8 的单简多字检测与识别模型。首先介绍构建简牍文字检测与识别数据集的过程，并详细说明数据的预处理过程。然后，对模型的各个模块进行详细的介绍，包括特征提取、检测和识别等。最后，通过实验验证本文提出的模型的有效性和检测识别能力。

第 4 章 简牍文字识别软件设计与开发。本章基于前面章节中开发的识别模型和检测模型，设计和开发一款简牍文字识别软件平台，以实现对简牍文字的实际应用。该软件平台包括数据导入、文字检测、检测框修改和文字识别等四个主要模块，通过这些模块的配合，实现对简牍文字的识别和检测的实际应用。

总结与展望。对本文完成的工作和研究内容进行总结，并且在未来展望中对内容的不足之处和后续的研究方向进行说明。

第2章 面向字型多变简牍单文字的可变形卷积识别模型

不同于现代手写体文字识别和古代摹本文字识别，简牍文字所处年代和载体的不同，同一文字有不同写法，导致文字形态和结构多变。由于书写载体的大小不一和单个载体篇幅限制导致文字尺度多变。目前常见的古文字识别方法大都采用的是基于标准卷积神经网络的深度学习模型，通过分析标准卷积神经网络，不难发现标准卷积的规则采样方式无法满足简牍文字形态、结构、尺度多变等问题。并且 CNN 网络稀疏特性和局部连接特性，往往其感受野是被限制死的，并不根据图像内容自适应的调节。基于上述问题，本章构建 DeConv Swin 模型可以应对文字形态、结构和尺度多变等问题的简牍单文字识别模型。要实现这一目的，所设计的网络应该满足一下特点：感受野满足多尺度或尺度自适应特性，可处理旋转、仿射等几何形变引起的变化。可变形卷积在目标检测和图像分类中的成功应用，非规则的采样方式解决文字形态和结构多变的问题，而 Swin Transformer 的具有层级的特征表达方式结构对于简牍文字的尺度变化也有很好的适应性。

本章的主要工作和安排如下：在 2.1 小节，介绍如何构建简牍单字识别数据集以及数据集的图像增强，2.2 小节对构建的可变形卷积 Swin Transformer 识别模型细节进行介绍，2.3 小节中则给出设置的模型评估指标、实验环境和相关参数，以及简牍单字识别模型在数据集上的表现结果，以及外部验证结果四部分内容，最后的 2.4 小节进行章节小结。

2.1 简牍文字识别数据集

2.1.1 数据集构建

本文的研究对象主要为居延新简。在数据集的编制过程分以下几个阶段进行，每个阶段如图 2-2 所示，分别为数据采集、数据预处理、数据注释、数据验证和输出文字图像。第一阶段是通过高光谱相机扫描简牍实物来获取数据。通过与博物馆合作，获取其简牍实物。再通过使用高光谱相机拍摄简牍实物，最后获取高清晰度字迹清晰简牍红外数字图像。图 2-1 显示每枚简牍实物的扫描过程。这个过程产生 8049 张简牍图像。

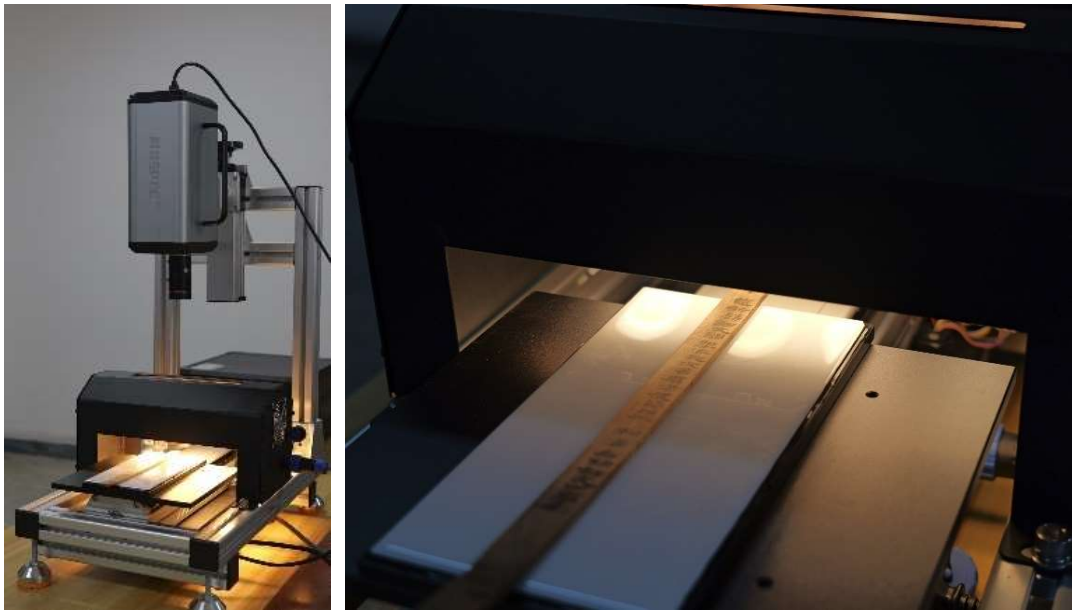


图 2-1 高光谱相机扫描简牍文物以获取其红外数字图像

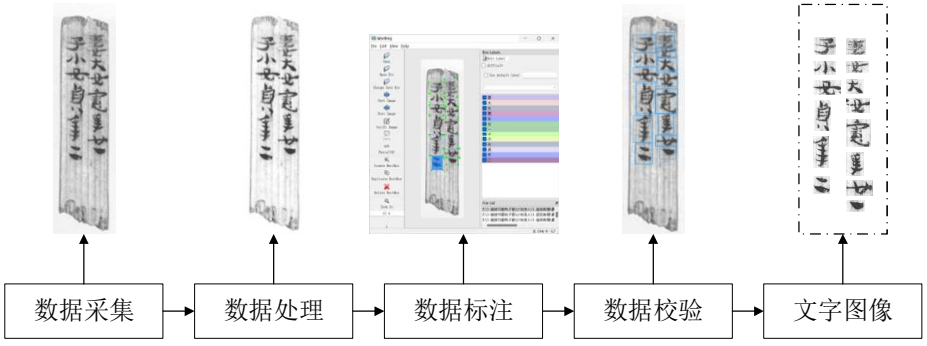


图 2-2 生成数据集的处理步骤概述

在进行简牍文字图像的标注时，每个文字区域被精确地界定，并用四个坐标来标识：分别为左上角的横坐标与纵坐标（左上 x，左上 y），以及右下角的横坐标与纵坐标（右下 x，右下 y），同时记录所对应的现代文字。此过程由简牍相关的专业人员采用 LabelImg 图像标注工具完成标注工作，LabelImg 界面如图 2-3。标注工作完成后，进行一次非技术性的校对工序，其目的在于确认标注数据的完整性与准确性，包括核对标注信息与图像中文字的一致性，以及排查可能存在的遗漏或错误情况。最终标注简牍文字图像 221 类，总共 60000 张文字图像。

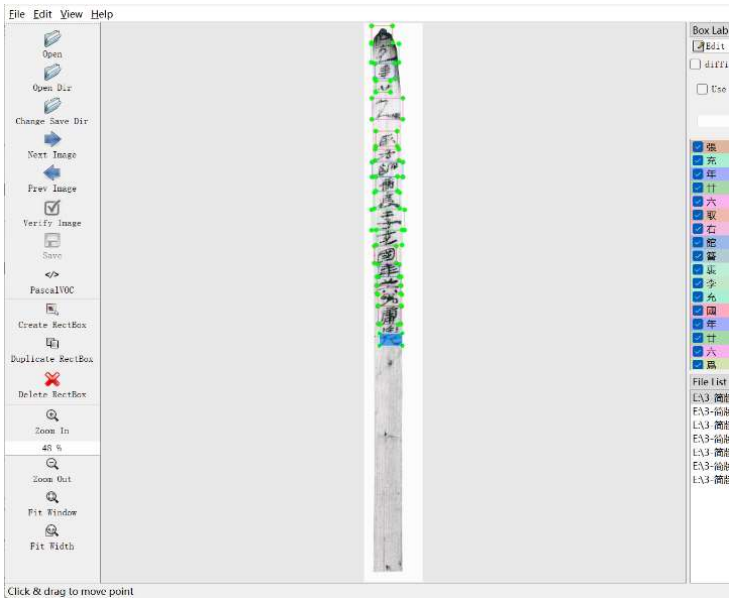


图 2-3 LabelImg 对简牍图像进行标注

标注过程完成后，依据所得的标注结果，将原始图像分割成若干单字图像，并将相同字符的图像汇总至特定的文件夹中。在此基础上，对字体模糊或字体残缺的图像进行筛查，并予以剔除，以确保所用图像的质量，部分文字图像数据如图 2-4 和图 2-5 所示，统计并可视化 100 次以上的文字如图 2-6。

图 2-4 居延新简部分“前”字样本

图 2-5 居延新简部分文字样本

图 2-6 居延新简数量 100 以上的文字词频图

2.1.2 数据集处理

数据集的多样性和质量对深度神经网络的实际训练和性能表现具有直接的影响。充足的数据量有助于提升模型的泛化能力，而高质量的图像能够帮助模型更好地学习图像特征。因此，本文的重点将放在图像增强和图像降噪这两个方面，以改进文字图像的数据处理。

（1）图像增强：数据扩展在模型迭代过程中实时执行。

图像增强对原始图像进行一系列的变换和处理，改善图像的质量和可视化效果。这包括调整亮度、对比度、大小等，以增强图像的细节和清晰度。通过图像增强，可以改善文字图像的可读性，使得深度神经网络能够更准确地识别和理解图像中的文字内容。主要使用原始图像的大小缩放、平移和围绕中心旋转一定角度来模仿文字书写过程中的形变。之后在一定范围内随机对图像亮度和对比度进行调整。图像增强过程在每个轮次中随机抽取不同增强方法，虽然数据总量没有改变但输入模型的训练图像每个轮次都不相同，间接增加样本数量。数据增强效果如图 2-7

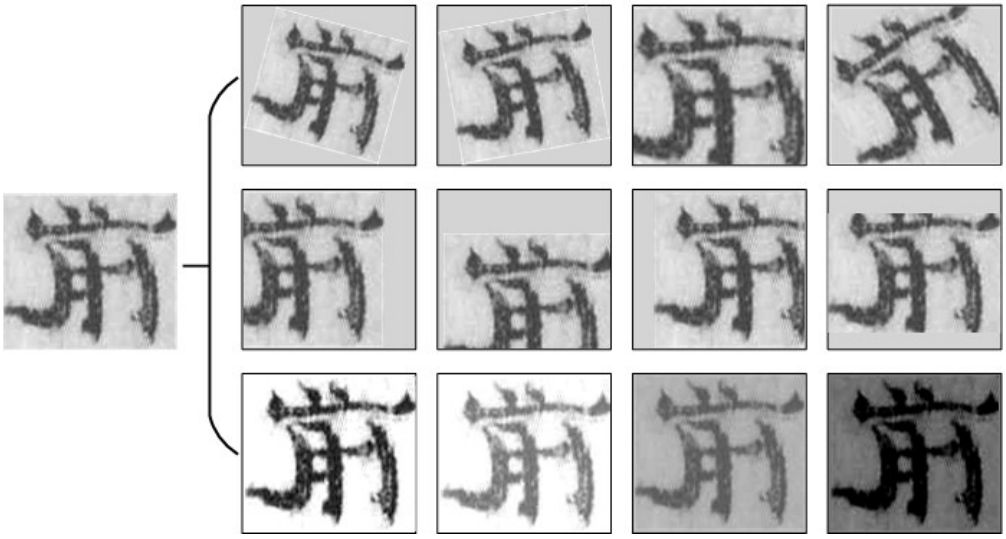


图 2-7 数据增强

（2）图像降噪：

图像降噪技术可以有效减少图像中的噪声和干扰，提高图像的清晰度和准确性。对于文字图像而言，降低噪声有助于消除文字边缘的模糊和细节丢失现象，从而提高深度神经网络对文字的识别和分析能力。本章对于数据集的降噪算法是双边滤波，选择双边滤波的主要原因是经过统计分析以后简牍单个文字图像的平均峰度为 5.3，平均偏度为-1.4 其噪声类型符合均值偏移高斯白噪声，对于这种类型的图像噪声可以采用高斯滤波和双边滤波，但是高斯滤波会对图像中的边缘信息造成一定程度的模糊，而双边滤波是一种基于局部相似性的非线性滤波器，可以有效地抑制噪声，而对图像中的边缘信息影响较小，因此选用双边滤波算法，其计算方式如式（2-1）所示

$$I'(x)=\frac{1}{W_p}\sum_{x_i\in S} I(x_i) f_r(\| I(x_i)-I(x)\|) g_s(\| x_i-x\|) \tag{2-1}$$

其中，$I(x)$是输出图像在位置 x 的像素值；$I(x_i)$输入图像在位置 x_i 的像素值；S 是

以 x 为中心的窗口内所有像素的集合；f_r 是关于像素值相似度的高斯函数，用于测量中心像素 $I(x)$ 与邻域像素 $I(x_i)$ 之间的强度差异；g_s 是关于空间距离的高斯函数，用于测量中心像素 x 与邻域像素 x_i 之间的空间距离；W_p 是为确保权重的和为 1 而引入的归一化因子，保证滤波后的像素强度值不会因为权重的变化而产生较大的偏移，其计算方式如式（2-2）所示

$$W_p = \sum_{x_i \in S} f_r(\| I(x_i) - I(x) \|) g_s(\| x_i - x \|) \tag{2-2}$$

其中，$f_r(\| I(x_i) - I(x) \|)$ 是一个基于像素值差异的权重函数，也称为强度权重，$\| I(x_i) - I(x) \|$ 表示强度的差异；$g_s(\| x_i - x \|)$ 是一个基于空间距离的权重函数，也称为空间权重，$\| x_i - x \|$ 表示中心像素和邻域像素之间的空间距离。

最终经过双边滤波降噪之后的部分图像如图 2-8 所示，可以看到在过滤背景噪声之后同样保留文字的边缘信息。

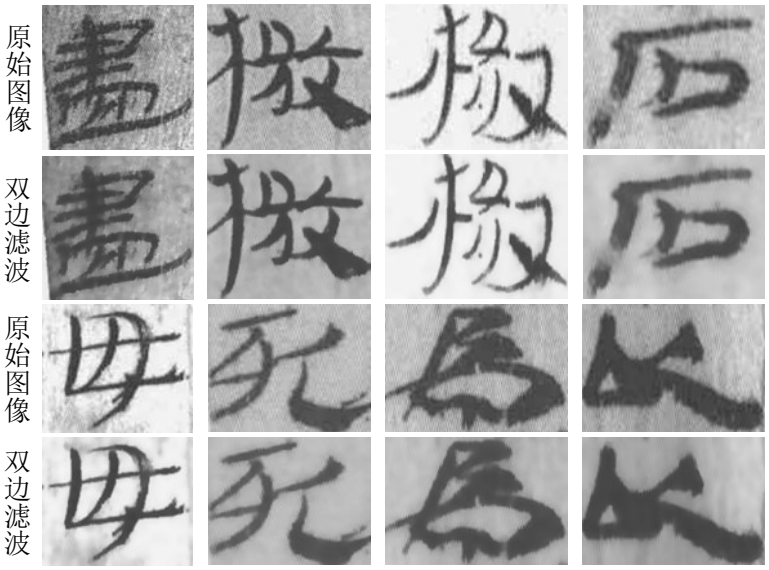


图 2-8 居延新简部分文字降噪结果图

2.2 简牍文字识别模型构建

本章提出一种结合可变形卷积和 Transformer 的融合模型 DeConv Swin，用于简牍文字识别。DeConv Swin 整体网络是基于 Swin Transformer 建立，DeConv Swin 主要有以下三部分组成：用于提取简牍文字的形态、结构、文字变形的可变形卷积层，用来提取不同尺度简牍文字图像信息的 Swin Transformer 骨干网络和用于输出分类结果的分类输出头。

对于简牍文字形态多变、结构扭曲变形和文字大小不一的特点来说，需要更好地学习文字结构层面的语义信息和更深层次的抽象信息，以缓解简牍文字多变导

致文字结构相似导致模型识别为同一文字的问题。因此，必须更加细致的收集图像中的信息，并且利用层级结构提取更为深层的语义信息保证不同尺度文字信息不被忽略。然而，标准卷积操作具有局限性，卷积固有的规则采样限制其在对旋转和透视等几何特性方面的适应能力，相较于标准卷积，可变形卷积在标准卷积的基础上引入可学习的偏移量，使得可变形卷积对于旋转和透视等几何特性方面的适应能力增加，可以更好地适应简牍文字变化多的特点。另一方面，由于简牍文字书写在不同形制的简牍之上，导致不同简牍上的文字尺度差异巨大，本章利用 Transformer 可以真正的关注全局信息，并且 Swin Transformer 的层级结构保留不同尺度的文字深层意义信息。本章尝试从更好适应简牍文字变化和提高不同尺度文字的语义信息提取方面入手，提出 DeConv Swin 模型来更有效对简牍文字进行识别。

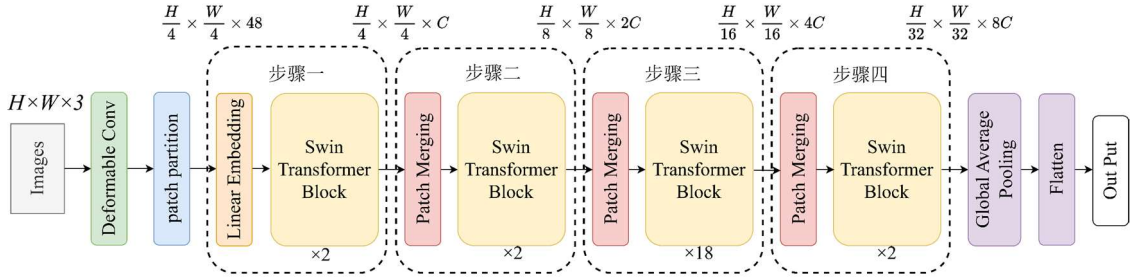


图 2-9 DeConv Swin 整体结构

DeConv Swin 的整体结构如图 2-9 所示，模型主要包含三个部分：可变形卷积特征提取层，Swin Transformer 多尺度特征提取层和简牍文字识别输出模块。可变形卷积层的引入增加模型对简牍文字变化多特点的适应能力，Swin Transformer 模块为识别提供全局信息和不同尺度简牍文字的深层语义信息。最终输出层采用全局平均池化（Global Average Pooling，GAP）代替全连接层，使得模型可以接受任意尺度的简牍文字图像。

2.2.1 文字字型特征提取模块

简牍是由不同的人进行书写的，同一个文字之间存在着不同的写法，使得文字产生多种形态。此外，由于简牍的书写载体大小不一以及单个载体的篇幅限制，导致简牍文字的形态、尺度和结构变化多样。这种由书写带来的变形以及文物长时间掩埋在地下导致的墨迹退化，使得文字之间更加难以辨认。

传统 CNN 利用局部连接和参数权重共享的机制，对图像进行位置上的规则网格采样，并对采样像素进行卷积运算，以生成对应位置的输出。因此传统 CNN 模型具有平移不变性。对于简牍文字而言，由于简牍文字形态和结构多变，其采样点

在空间上并非是规则形状，如果直接使用 CNN 模型利用卷积实现特征提取，导致其无法适应简牍文字多变的特点。正是 CNN 固有的规则格点采样方法，限制其在保持尺度、旋转和透视等几何特性方面的能力，虽然可以通过金字塔型网络设计或对数据集实施几何变换来增强数据，从而使 CNN 模型识别和学习特定的几何特性，但这些学习到的特性仅限于预设的某些特定尺度和旋转角度的几何变换条件，这意味着 CNN 模型仍然不能对任意几何变换保持不变性。

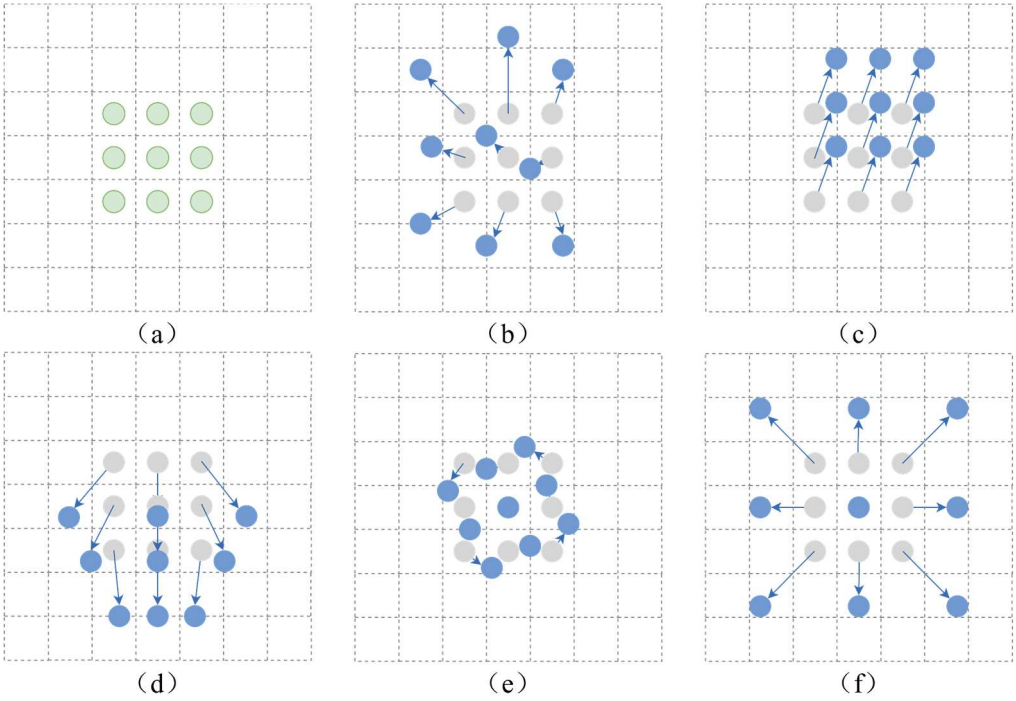


图 2-10 标准卷积和可变形卷积采样点（a）标准卷积规则采样点（绿色点表示）；（b）可变形卷积中带有偏移量（蓝色箭头表示）变形采样点（蓝色点表示）；（c）-（f）：是（b）的特殊情况，表明可变形卷积可以用于处理各种几何变换，如平移、透视、旋转和缩放

可变形卷积是一种新的方法，其卷积核具有自适应调整形状的能力，可变形卷积中，每个卷积核采样点都引入偏置量，使其能够根据不同的物体调整自身的形状，这种自适应调整形状的能力使得可变形卷积在处理具有不规则形状的目标时表现出更大的灵活性，图 2-10 所示的示例中，可变形卷积的卷积核能够根据目标的形状自动调整自身的形状，以适应不同的目标形状。

可变形卷积具有自适应调整形状的特点，使得它在提取简牍文字特征的过程中能够更加精细地捕捉文字的形态变化。通过自动调整卷积核的形状，可变形卷积可以更好地适应不同文字之间的写法差异、形态变化以及尺度和结构的多样性，从而提高对简牍文字的识别准确性。通过对简牍文字进行形变建模，可变形卷积能够更好地捕捉并区分文字之间的细微差异，从而提高识别的准确度。它能够灵活地适应文字的多样性，使得在复杂的简牍文字识别任务中，可变形卷积能够更好地处理

形态、尺度和结构的变化，提升对文字的准确识别能力。可变形卷积的计算方式如式（2-3）：

$$\hat{y}(p_0)=\sum_{p_n\in\mathcal{R}} w(p_n)\cdot x(p_0+p_n+\Delta p_n) \tag{2-3}$$

其中，$\hat{y}(p_0)$ 表示输出特征图在位置 p_0 的像素值；$w(p_n)$ 是卷积核在位置 p_n 的权重；$x(\cdot)$ 表示输入特征图；p_0 是输出特征图中的中心位置点；p_n 是标准卷积核内的位置偏移，属于预定义的邻域 R；Δp_n 是在位置 p_n 的可学习偏移量，它由网络通过学习得到，用于适应输入特征图的几何变形。

2.2.2 文字尺度特征提取模块

传统的 CNN 模型中，卷积层输出的特征图（Feature Map）在流经最后一个卷积层后，通常会被连接到多个全连接层，这一过程负责将特征图转换为一个固定长度的特征向量，适配于图像的分类或回归任务。例如，对于图像分类，如 AlexNet 模型会输出一个 1000 维向量，每一维代表输入图像属于某类的概率。然而，对于简牍文本的处理，由于简牍的物理尺寸和书写内容的多样性，文字的尺度差异很大。如果将所有简牍图像直接缩放到相同大小，可能会导致细节信息的大量丢失，如图 2-11 所示，这将阻碍模型学习到有效特征。

由于古文字图像的尺度变化较大，传统 CNN 模型会受到固定感受野的限制，导致对不同尺度的古文字图像处理效果不佳。而 Swin Transformer 通过引入金字塔的结构，可以处理不同尺度的特征，可以适应不同尺度的古文字图像，从而有效解决尺度多变问题。此外，传统 CNN 模型在处理大尺度图像时也存在劣势。由于传统 CNN 模型的卷积操作是基于局部感受野的，对于大尺度图像，需要使用较大的感受野来捕捉全局上下文信息，但这会导致计算和存储开销的增加。

图 2-11 文字尺度变化

Swin Transformer 引入分层的注意力机制，如图 2-12，使得模型能够在不增加计算和存储开销的情况下有效地建模大尺度图像的全局上下文信息。另一个传统 CNN 模型的劣势是固定的局部感受野。传统 CNN 模型通常使用固定大小的卷积核进行卷积操作，这限制模型对不同尺度古文字图像进行建模。

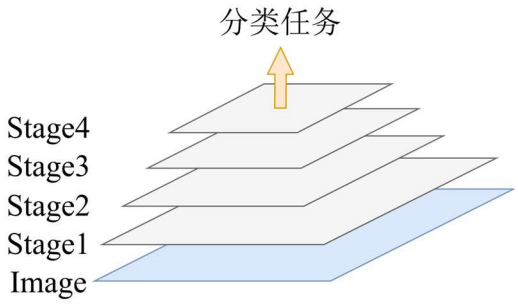


图 2-12 Swin Transfomer 分层注意力机制

而 Swin Transformer 引入基于窗口的自注意力机制，通过在输入特征图上划分不同的窗口，并在窗口内进行自注意力计算，使得模型能够对不同尺度的古文字图像进行建模。Swin Transformer 还具有更好的并行计算能力。传统 CNN 模型在处理大尺度图像时，需要将图像切分成小块进行处理，然后将结果拼接在一起。这种切分和拼接的操作会导致计算和通信的开销。而 Swin Transformer 通过自注意力机制的引入，使得模型可以直接在整个图像上进行并行计算，避免切分和拼接的开销。基于窗口的自注意力机制其主要分为两部分，一部分为窗口内部的自注意力（Window-based Multi-head Self-Attention，W-MSA）和位移窗口自注意力（Shifted Window-based Multi-head Self-Attention，SW-MSA），二者的计算方式分别如式（2-4）和式（2-5）

$$Attention(Q,K,V)=Softmax\left(\frac{QK^{T}}{\sqrt{d}}\right)V \tag{2-4}$$

$$Attention(Q,K,V)=Softmax\left(\frac{QK^{T}}{\sqrt{d}}+M\right)V \tag{2-5}$$

其中，Q、K、V 分别是查询（Query）、键（Key）、值（Value）矩阵，它们输入序列经过线性变换计算得来；d 是 Q 和 K 矩阵的维度分量，用于缩放点积的大小，通常是模型隐藏层维度；Softmax 用于归一化权重；T 表示矩阵转置；M 是一个可选掩码，用于 SW-MSA 中实现循环移位窗口机制，它允许跨窗口进行自注意力计算，同时保持窗口边界的连续性。

2.2.3 损失函数

本章在训练模型过程中采用交叉熵损失函数（Cross Entropy Loss）其计算方式如式（2-6）所示：

$$L=-\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{M}y_{i,c}\log(p_{i,c}) \tag{2-6}$$

其中，L 是计算得到的损失值；N 是数据集中样本的总数；M 是类别总数；$y_{i,c}$ 是一个指示器，当样本 i 属于类别 c 时为 1，否则为 0；$p_{i,c}$ 是模型预测样本 i 属于类别 c 的概率。采用交叉熵损失函数在模型输出每个类别的概率，这有助于在训练过程中直接优化分类概率，使得训练更加直接和高效，对于不正确的分类预测给予较大的惩罚，尤其是当模型对于实际类别的预测概率很低时，在使用 Softmax 函数的情况下，模型输出的概率会自动归一化，确保了输出可以解释为概率分布。Softmax 函数的计算方式如式（2-7）所示：

$$p_{i,c}=\frac{e^{z_{i,c}}}{\sum_{j=1}^{M}e^{z_{i,j}}} \tag{2-7}$$

其中，$p_{i,c}$ 是样本 i 在类别 c 上的预测概率；$z_{i,c}$ 是模型输出的原始值；M 类别总数。

2.3 实验结果与讨论

2.3.1 模型评估指标

为保证评估模型对多变字型的简牍文字图像的识别精度，本章通过分类准确率，召回率，精确率、F1-Score 并结合混淆矩阵来判断多变字型的简牍文字识别网络性能。准确率可以直观体现模型性能，它表示在所有分类决策中，正确的分类比例；召回率度量实际正类中有多少被模型正确识别的比例；精确率计算的是被模型预测为正类的样本中，实际为正类的比例；F1-Score 是精确率和召回率的调和平均值，平衡二者的影响；混淆矩阵提供一个详细的分类结果展示，它显示每个类别的实际与预测值的对应情况。混淆矩阵如表 2-1 所示：

表 2-1 混淆矩阵

	正样本	负样本
预测正例	*TP*	*FP*
预测反例	*FN*	*TN*

其中，TP：真正例，分类器正确地将文字预测为正确文字的数量；TN：真负例，分类器正确地将负类预测为负类的数量；FP：假正例，分类器错误地将其他文字预测为正确的数量；FN：假负例，分类器将正类文字预测为错误文字的数量。

分类准确率（Accuracy）如式（2-8）：

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2-8}$$

精确率（Precision）计算如式（2-9）：

$$Precision = \frac{TP}{TP + FP} \tag{2-9}$$

召回率（Recall）计算如式（2-10）：

$$Recall = \frac{TP}{TP + FN} \tag{2-10}$$

F1 分数（F1 Score）计算如式（2-11）：

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{2-11}$$

2.3.2 实验环境及参数设置

本章实验使用如表 2-2 所示的实验环境来实现和训练网络。本算法使用 SGD 来优化网络权重，momentum=0.9，weight decay=0.0005，学习率设置为 0.0001。利用 GPU 加速训练过程，将 batch size 设置为 8，并将训练迭代轮次为 40 次。在训练开始前对数据进行筛选保留单类数量超过 100 以上的图像类别，最终有 221 类文字图像满足上述条件，并且采用 8：1：1 的比例划分训练集、测试集和验证集。

表 2-2 实验环境与具体配置

实验环境	具体配置
操作系统	Ubuntu 20.04
CPU	Intel(R) Xeon(R) Silver 4210R CPU @ 2.40GHz
GPU	NVIDIA RTX 4090
显存	24G
内存	64G
深度学习框架	PyTorch 1.8.0

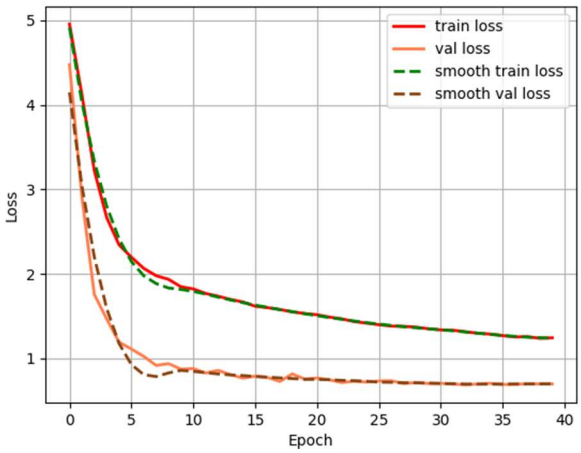


图 2-13 模型在训练集和验证集上的损失变化

在 DeConv Swin 训练过程中记录每一轮的训练集和验证集损失，来判断模型训练是否收敛，如图 2-13 所示模型在 15 次左右迭代时基本达到收敛，最终 40 次迭代完全收敛。

2.3.3 简牍文字识别模型评估

本实验研究基于可变形卷积的 Swin Transformer 的网络结构，验证不同结构的 Swin Transformer 对于简牍文字识别结果的影响，可变形卷积在模型中不同位置以及数量对于识别结果的影响。通过比较基于可变形卷积的 Swin Transformer 的不同结构完成模型性能评估。实验验证过程中由于可变形卷积计算消耗大，三层可变形卷积的加入导致模型无法完成训练，因此只对比一层和两层可变形卷积的加入的影响。

DeConv Swin v1：采用一层可变形卷积与 Swin-small 的组合，并且可变形卷积在 Patch Partition 之前。

DeConv Swin v2：采用一层可变形卷积与 Swin-small 的组合，并且可变形卷积在 Patch Partition 之后，在 Linear Embedding 之前。

Swin-small：Swin-small 的原始结构

2DeConv Swin v1：采用两层可变形卷积与 Swin-small 的组合，且两层都在 Patch Partition 之前。

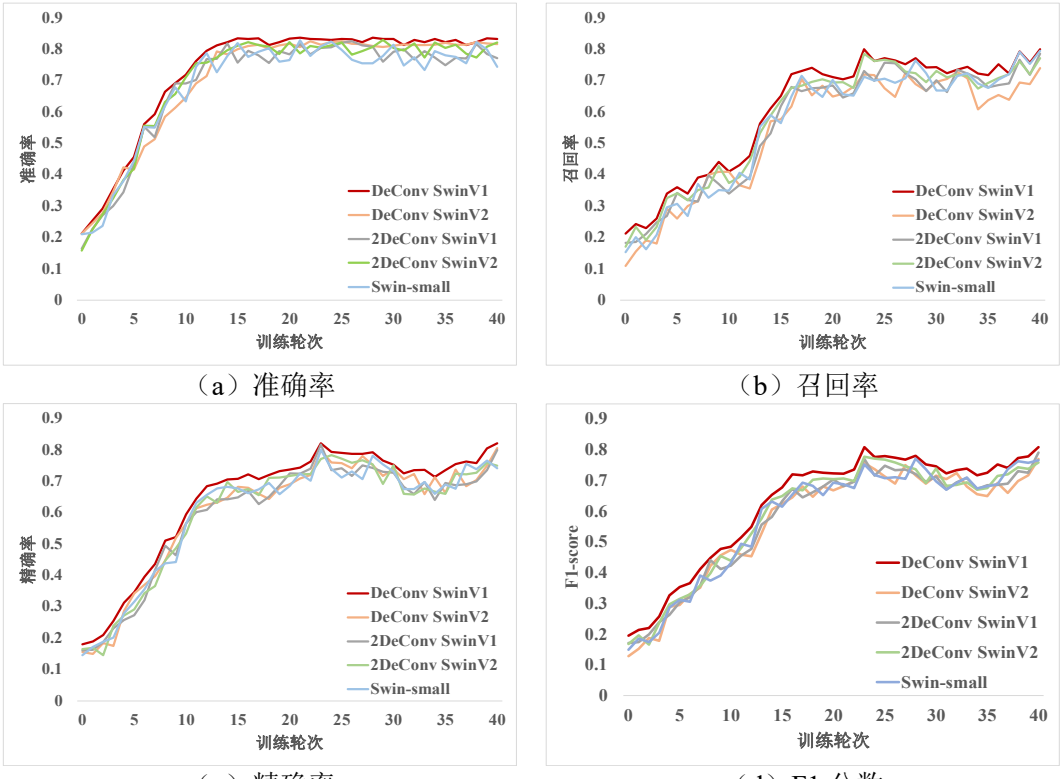
2DeConv Swin v2：采用两层可变形卷积与 Swin-small 的组合，将 Patch Partition 放在两层可变形卷积之间。

由表 2-3 中和图 2-14 的结果可以发现可变形卷积数量并非越多越好，从结果来看反而是 DeConv Swin v1 的测试结果最好准确率达到 83.5%；DeConv Swin v1 和 DeConv Swin v2 之间主要是验证 Patch Partition 层是否对于可变形卷积提取的

特征有影响，最终实验结果表明可变形卷积层在 Patch Partition 层之后反而精度下降，说明可变形卷积直接提取原始图像的特征要优于通过 Patch Partition 层提取之后再次提取的特征。DeConv Swin v1 和 2DeConv Swin v1 相比多一层的可变形卷积并没有带来积极的影响，反而使得最终结果有所下降。两层可变形卷积会增加模型的复杂度，从而导致过拟合。模型可能会学习到图像中的噪声和干扰，从而导致泛化性能下降。2DeConv Swin v2 与 2DeConv Swin v1 相比只是改变其中一组可变形卷积的位置之后对模型结果并未有所影响，说明两层可变形卷积增加模型的复杂度，从而导致过拟合的影响大于其位置改变的影响。因此一层可变形卷积并且在 Patch Partition 之前的效果最好，对模型后续的特征提取帮助最大。

表 2-3 简牍文字识别模型评估

模型方法	准确率	召回率	精确率	F1 分数
Swin-small	82.6%	79.3%	80.1%	79.7%
DeConv Swin v1	**83.5%**	**79.9%**	**81.9%**	**80.9%**
DeConv Swin v2	82.4%	73.8%	80.5%	77.0%
2DeConv Swin v1	82.6%	78.4%	79.6%	79.0%
2DeConv Swin v2	82.7%	78.5%	80.9%	79.7%



（a）准确率　（b）召回率

（c）精确率　（d）F1 分数

图 2-14 消融实验结果可视化图

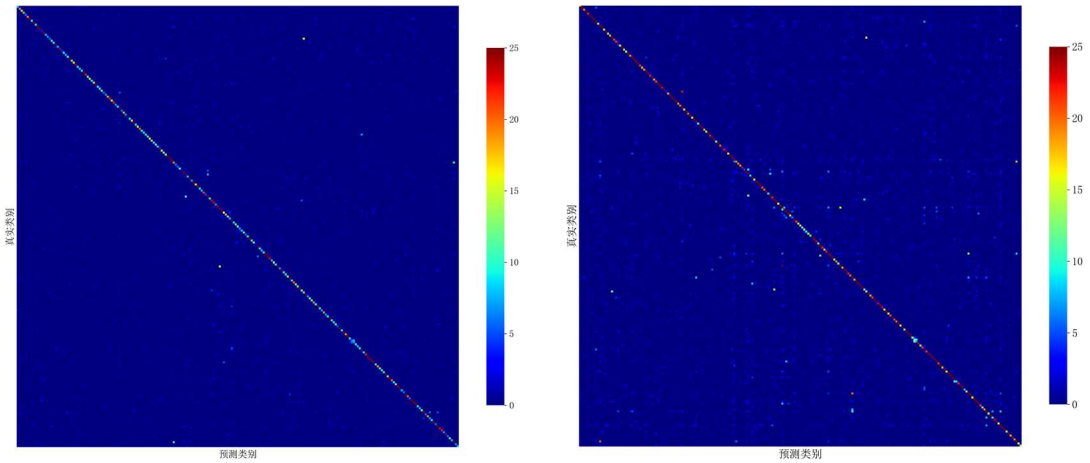
2.3.4 与经典图像分类算法对比分析

为探究提出模型与其他主流分类识别神经网络模型的优越性，本章在所构建的简牍文字识别数据集上评估主流分类神经网络模型，分别有 VGG 系列，ResNet 系列，InceptionV3、ViT、Swin Transformer 系列以及本文构建的 DeConv Swin。如表 2-4 所示为上述深度学习模型在简牍文字识别数据集训练之后，经过测试集测试之后的实验结果。

表 2-4 不同主流分类识别神经网络的识别结果

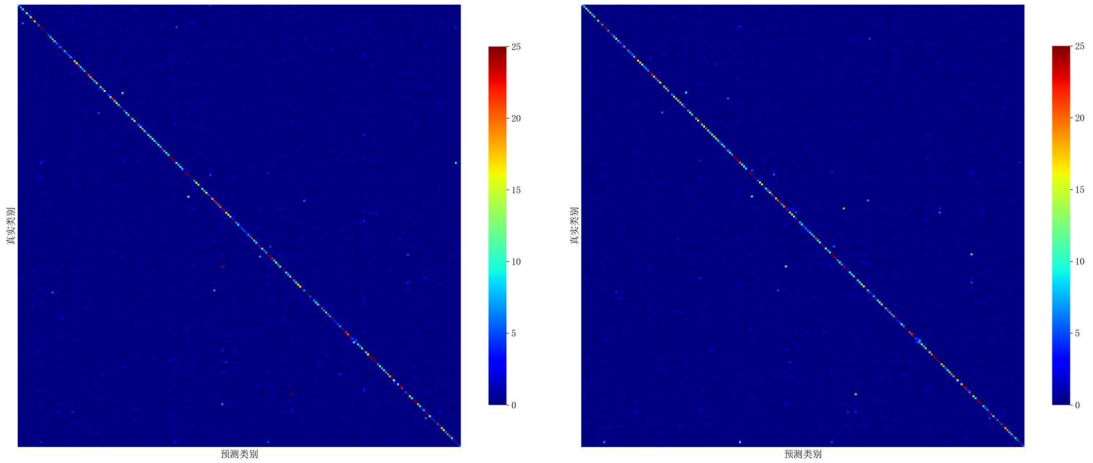
模型方法	准确率	召回率	精确率	F1
VGG16	71.7%	69.5%	70.1%	69.7%
VGG19	68.9%	55.6%	68.5%	61.3%
ResNet34	72.9%	70.2%	72.1%	71.1%
ResNet101	72.5%	68.5%	71.5%	69.9%
ViT	73.4%	69.3%	71.9%	70.5%
InceptionV3	70%	65.5%	68.8%	67.1%
Swin-small	82.6%	79.3%	80.1%	79.7%
DeConv Swin	**83.5%**	**79.9%**	**81.9%**	**80.9%**

由表 2-4 可知，通过与之前的主流的分类深度神经网络比较，本文构建的 DeConv Swin 模型相较于其他模型获得最好的识别性能。对于简牍单文字识别任务来说，可变形卷积的加入使得模型对文字的形态变化捕捉的更加精细；Swin Transformer 的基于窗口的自注意力机制，通过在输入特征图上划分不同的窗口，并在窗口内进行自注意力计算，使得模型能够自适应地对不同尺度的古文字图像进行建模，这也是本章提出的针对于面向字型多变简牍单文字的可变形卷积分类识别模型获得最佳性能的关键原因。

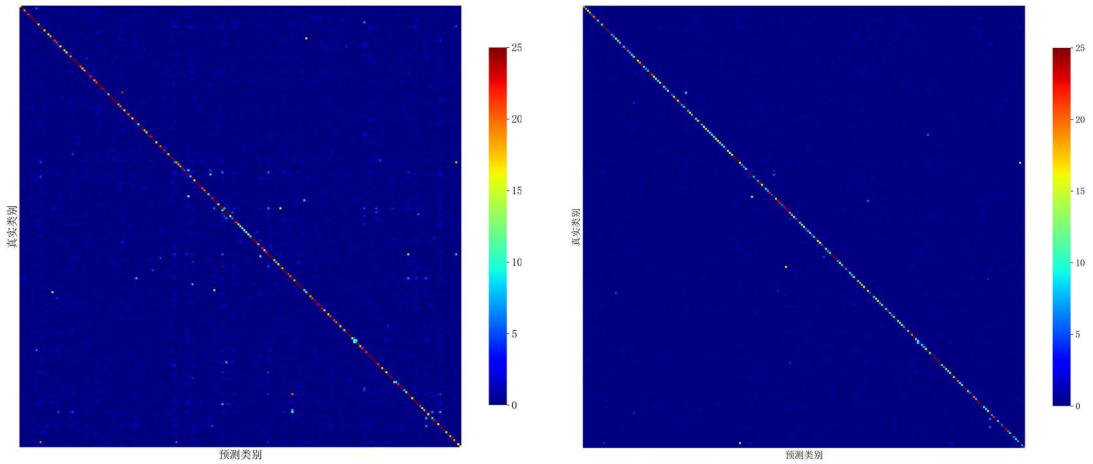


（a）Swin-s　　（b）ViT

图 2-15 混淆矩阵可视化结果



（c）VGG16　　（d）ResNet34



（e）InceptionV3　　（f）DeConv Swin

图 2-15（续）混淆矩阵可视化结果

并且通过不同模型在测试集上的混淆矩阵可以看出，图 2-15（b）中显示 ViT 对于不同简牍文字图像的特征学习相对一致，但对于不同图像之间以及不同类之间的差异学习较差，导致其虽然准确率是除 DeConv Swin 和 Swin-s 之外最高的达到 73.4%，但是 ViT 没有学习到文字与文字之间的差异性导致其从混淆矩阵可视化图来看，分类错误的数量较多；图 2-15（c）（d）是 VGG16 和 ResNet34 的混淆矩阵可视化结果，可以看出相比较于 ViT 虽然对于文字图像之间共通的特征学习较差，但是它们都能学习到文字图像之间的差异性，从 F1 分数的结果上来看它们与 ViT 的结果相差不多；图 2-15（f）是本章所提出的模型 DeConv Swin 通过可视化结果可以直观看出，DeConv Swin 的 FN 和 FP 的数量明显减少，从图 2-15（a）与（f）的对比中可以发现，本章提出的模型对于基准模型 Swin-s 来说其中几类的分类效果有提升，说明可变形卷积的加入是有效的。通过图 2-16 也可看出通过引入可变形卷积使得 Deconv Swin 提升了模型对于形近字区别能力，使得其可以正确识别出对应的文字。

Deconv Swin	毋	死	以	傲	椒
Swin Transformer	毋	死	以	椒	椒
真值	毋	死	以	傲	椒

图 2-16 部分文字识别结果

2.4 本章小结

本章提出一种基于可变形卷积神经网络的简牍识别模型。该模型将单个文字图像作为网络的输入，并通过网络提取图像的特征，然后通过分类模块输出图像中的文字内容。本章详细介绍数据集的构建过程和网络模型的构建过程，并对所提出的模型进行识别性能的比较分析。在网络模型的构建过程中，采用可变形卷积神经网络，通过引入可变形卷积操作，网络能够自适应地调整卷积核的形状和位置，从而更好地适应文字的变化。此外，模型的特征提取主干网络采用 Swin Transformer，利用 Swin Transformer 多尺度信息以适应简牍文字的尺度变化。实验结果表明，该模型能够准确地识别简牍文字图像中的文字内容。

第3章 面向复杂版面的单简多文字 YOLO 检测模型

随着目标检测算法的发展，场景文字检测成为研究的热点，一些学者开始尝试使用深度学习技术来检测古文字。但对于古文字的检测存在一些特殊挑战，长期掩埋导致文字载体形态发生变化，导致采集的影像存在大量噪声和文字形变。并且简牍上文字的大小不一且位置多变，导致直接应用目标检测模型或者文字检测模型导致文字漏检和错检的问题。针对以上问题，本章对居延新简数据采用双边滤波算法降低图像背景噪声，设计了一种 DeConv YOLO 模型应对简牍图像的尺度多变和文字位置多变的问题。通过引入可变形卷积，模型能够更好地适应简牍文字的位置多变，类 PANet 网络结构的多尺度特征融合模块来应对简牍图像的尺度多变，并对损失函数进行调整，以准确的检测居延新简中的文字位置。

本章的主要工作和安排如下：3.1 小节，介绍如何构建和处理简牍文字检测数据集，3.2 小节详细说明简牍文字检测模型的构建，3.2 小节中则给出设置的模型评估指标、相关参数，以及简牍文字检测模型在数据集上的表现结果，以及外部验证结果三部分内容，最后的 3.3 小节进行章节小结。

3.1 数据预处理

通过使用高光谱相机拍摄简牍实物，最后获取高清晰度字迹清晰简牍红外数字图像。这个过程产生 8049 张简牍图像。采用算法和人工相结合的标注方式以减少标注成本，所标注出的文本框共 103282 个，每张图像文字数量统计图如图 3-1，平均 13.03 个，但最多的图像中含有文本框 143 个，最少的包含文本框 1 个。部分数据示例如图 3-2 所示。

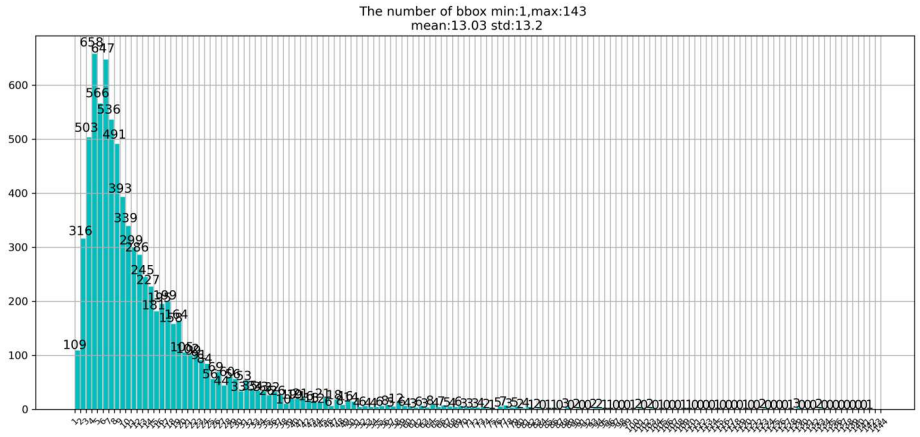


图 3-1 居延新简文字标注框统计图

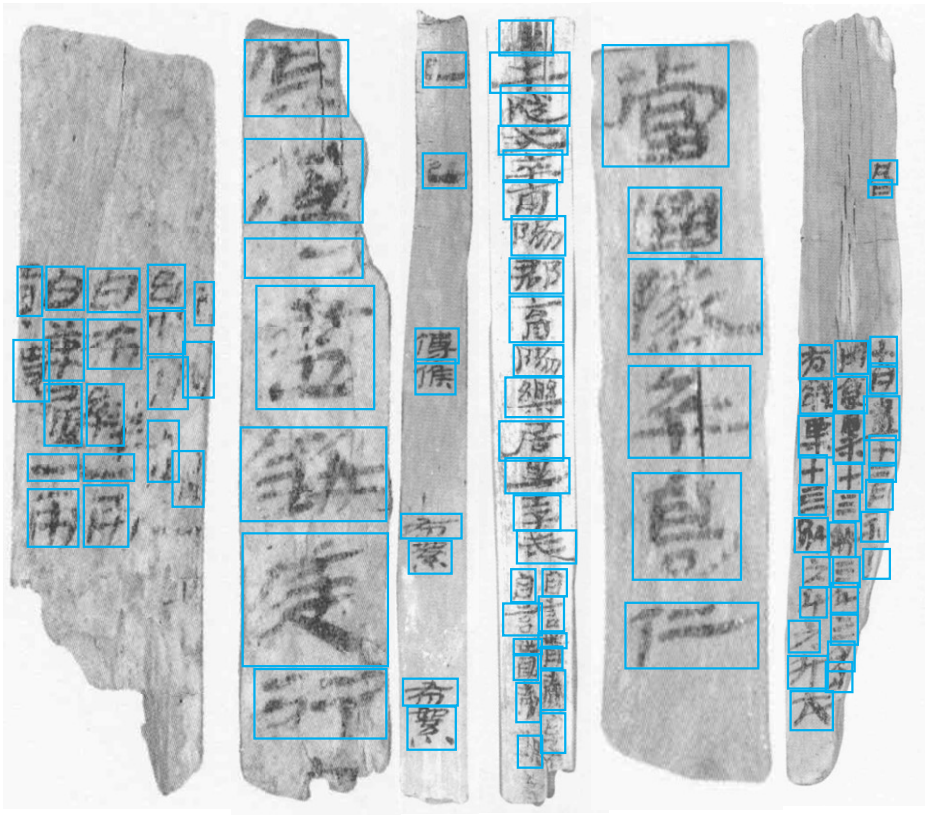


图 3-2 居延新简检测数据集各类样本标注示例

通过 2.1.2 小节对居延新简文字图像分析可知，简牍图像存在的噪声特性符合偏移高斯白噪声特性，通常这类噪声的去噪方法有两种分别是高斯滤波和双边滤波，高斯滤波的计算方式如式（3-1）：

$$G(x,y)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3-1}$$

其中，$G(x,y)$是二维高斯函数的值，表示对位置(x,y)的权重；x和y是像素位置相对于中心的坐标，其中中心像素的坐标为$(0,0)$；σ是标准差，它决定高斯函数的宽度，标准差越大，高斯函数的钟形曲线越宽，滤波器对图像的平滑效果越强。

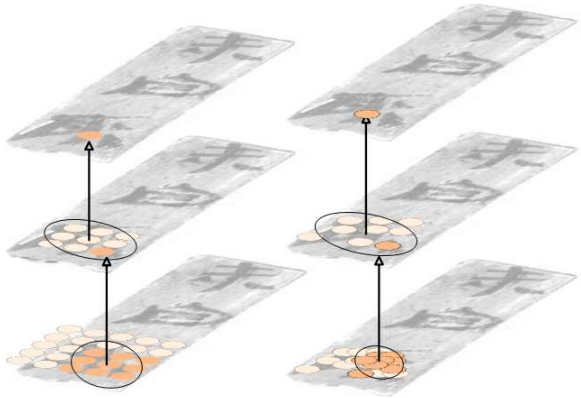
通过式（3-1）不难看出，高斯滤波降噪是一种线性平滑滤波器，其对边缘信息的保留能力差。居延新简图像中需要过滤背景噪声保留文字信息，但高斯滤波会在过滤背景信息的同时模糊文字边缘和背景之间的边界。虽然双边滤波计算成本比较高，但是双边滤波在平滑图像的同时能够保持边缘信息。最终采用双边滤波进行图像降噪如图 3-3 所示。

图 3-3 居延新简降噪效果图

3.2 简牍文字检测模型

本章提出一种基于 YOLOv8 和可变形卷积的融合模型 DeConv YOLO，用于简牍文字检测任务。DeConv YOLO 整体网络是基于 YOLOv8 构建的，YOLOv8 是由 Ultralytics 开发，综合目前发表的 YOLO 系列算法。DeConv YOLO 主要由以下三个部分组成：用于定位匹配文字位置特征的可变形卷积层，提取特征的骨干网络，加强特征提取网络和用于回归检测文字的检测头。

对简牍图像中文字尺度多变、位置多变这一特点来说，进行文字检测容易导致文字位置的漏检和错检，因此需要更好地匹配文字位置和图像尺度变化。由于卷积规则采样的影响，导致普通卷积无法完全匹配文字的变化，所以采用可变形卷积这一卷积的变体。可变形卷积在标准卷积的基础上引入可学习的偏移量，使得可变形卷积的采样点由标准卷积的规则采样变为不规则采样如图 3-4，使其更好地匹配文字多变的位置。另一点，由于简牍的不同形制和长期掩埋导致的腐烂和缺失，简牍图像之间尺度差异大，导致图像中简牍文字大小不一，因此骨干网络采用 Darknet-53 和路径聚合网络（Path Aggregation Network，PANet）结合的方式实现更好的多尺度特征提取。最后针对文字位置检测输出采用 SIoU 与特征点损失（Distribution Focal Loss，DFL）结合的方式进行文字位置的输出。



（a）常规卷积采样　　（b）可变形卷积采样

图 3-4 不同卷积采样点示意图

DeConv YOLO 整体结构如图 3-5 所示，模型主要分为可变形卷积提取层，Darknet-53 和 PANet 结合的特征提取层和 SIoU 与 DFL 结合的文字检测输出头组成。第一部分由可变形卷积与 Darknet-53 组成，使其更加适应简牍图像中文字变化特点；第二部分由 PANet 构成，其主要是对骨干网络提取的特征进行强化并融合；最终检测头接受三种不同尺度的融合特征进行文字位置信息的输出。

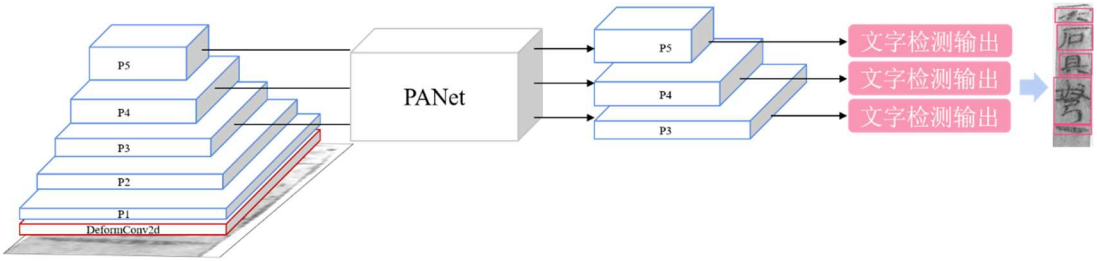


图 3-5 简牍文字检测模型框架

3.2.1-3.2.3 小节将会详细介绍模型框架的各个部分。

3.2.1 简牍文字检测骨干网络

当模型接受到简牍图像时，首先使用可变形卷积进行特征提取，对输入的图像进行卷积计算，生成具有自适应形状采样能力的输出特征图，并且通过可变形卷积的可变形区域兴趣池化（Deformable RoI Pooling）传入 YOLOv8 的原始骨干网络。可变形区域兴趣池化的计算公式如式（3-2）：

$$y(p_0) = max_{p_n \in \mathcal{B}(p_0)} x(p_n + \Delta p_n) \qquad (3\text{-}2)$$

其中，$y(p_0)$ 是池化后特征图在位置 p_0 的值；$\mathcal{B}(p_0)$ 表示围绕 p_0 的 bin 的区域；$x(p)$ 是输入特征图在位置 p 的值；Δp_n 是对于 p_n 位置学习到的偏移量。

可变形区域兴趣池化[21]完整过程如图 3-6 所示

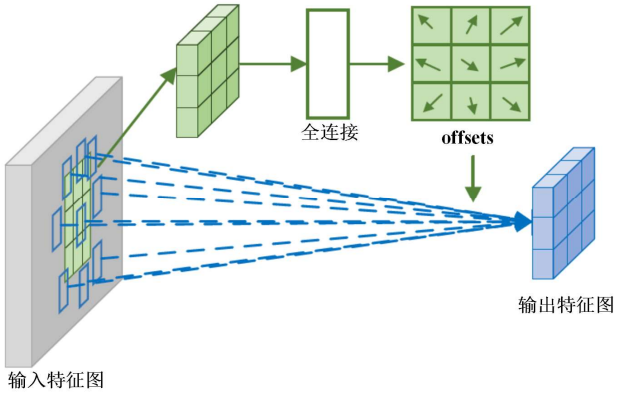


图 3-6 可变形区域兴趣池化

YOLOv8 的骨干网络是在 Darknet-53 网络结构的基础上删除下采样过程中维度为 1024 的卷积核，该层移除的作用主要是防止过拟合，增强泛化能力并且减少模型参数使之更容易训练，最终是由多个卷积层和 CSPLayer_2Conv 组成，整体流程如图 3-7 所示。

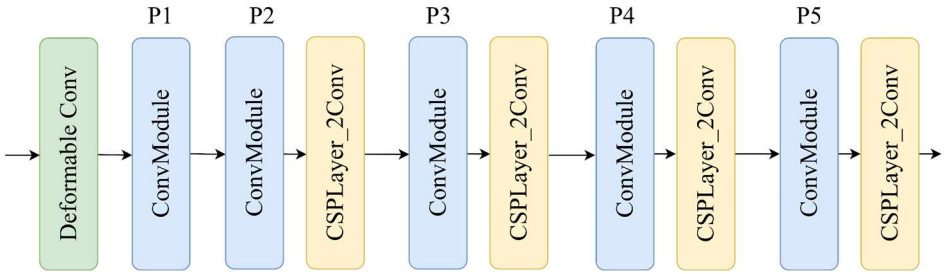


图 3-7 简牍文字检测模型骨干网络

YOLOv8 中 CSPLayer_2Conv 模块计算方式基本与 YOLOv5 中 C3 模块相似，其主要改动为将第一个卷积层的卷积核大小由 6x6 改为 3x3 虽然减少感受野范围，但是对于计算效率的提升明显。并且于 C3 模块相比 CSPLayer_2Conv 中增加更多的跳层连接和额外的分离操作，如图 3-8。

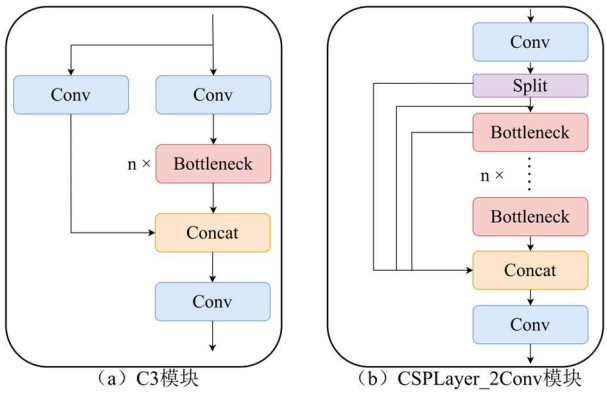


图 3-8 C3 和 CSPLayer_2Conv 结构图

3.2.2 多尺度特征融合网络

在简牍文字检测模型的特征融合阶段，采用路径聚合网络（PANet）结构。该方法在特征金字塔（FPN）网络自顶向下的结构中增加一个自底向上的金字塔，相比较于 FPN 结构，PANet 是对 FPN 的一个补充，使得低层的强定位特征可以向上传递，最终结果是浅层特征图与深层特征图聚合，使得模型对于多尺度特征的表达能力进一步增强。本章 DeConv YOLO 检测模型所使用的就是类 PANet 的特征融合网络如图 3-9 所示。由于检测对象为不同大小的文字，则需要不同尺度的特征图来提供相应的特征信息，提高对小尺度文字的检测准确性。

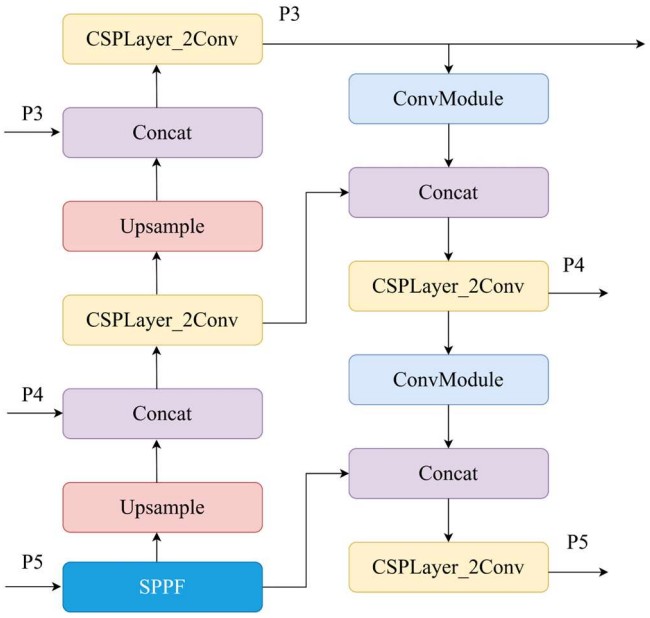


图 3-9 多尺度特征融合结构

如图 3-10 为多尺度特征融合结构的空间金字塔池化层（Syntax Parse Pruning Fast，SPPF）的结构，该结构的引入用于主要用于解决卷积神经网络输入图片尺寸固定的限制，同时解决图像区域剪裁、缩放操作导致的图像物体剪裁不全等问题。SPPF 层的计算方式如式（3-3）：

$$S = \bigoplus_{l=1}^{L} \bigoplus_{i=1}^{n_l} \max(F_{w_l,h_l}^{(i)}) \tag{3-3}$$

其中，S 为 SPPF 层的输出，是一个固定长度的向量，由所有池化结果连接而成；$\bigoplus$ 为连接操作；l 当前池化层级别索引；L 为空间金字塔池化层的总层数；n_l 为第 l 层中的网格数；i 为第 l 级金字塔中的第 i 个网格的索引；$F_{w_l,h_l}^{(i)}$ 为在特征图 F 中，相对应第 l 层的第 i 个网格的区域；w_l 和 h_l 为第 l 层每个网格的宽度和高度。

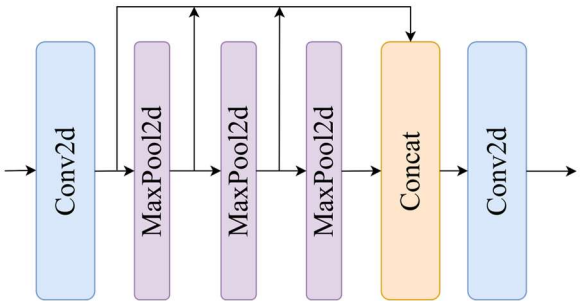


图 3-10 SPPF 模块

而在多尺度特征融合模块和骨干网络中 ConvModule 的主要组成是由一个二维卷积层、一个批标准化（Batch normalization，BN）层和一个 SiLU 激活函数，SiLU 是一个非单调的激活函数，有助于改善神经网络的学习能力，并且 SiLU 函数能够提供平滑的梯度，有利于梯度下降优化算法。

SiLU 激活函数的计算方式如式（3-4）：

$$f(x) = x \cdot \frac{1}{1+e^{-x}} \tag{3-4}$$

其中，x 是网络层的原始输入。

3.2.3 损失函数

YOLOv8 所使用的训练损失函数为 CIOU（Complete Intersection over Union，CIOU）与 DFL 结合进行优化边界框（Bounding Box）的预测，在 YOLOv8 中，Bounding Box 的预测是通过回归来实现的，而 DFL 则用于监督这个回归过程。DFL 计算方式如式（3-5）：

$$DFL(p,q) = -\sum_{c=1}^{C} q_c \cdot (1-p_c)^{\gamma} \cdot log(p_c) \tag{3-5}$$

其中，C 是类别的总数，在本章中的文字检测只有一类，因此 C 为 1；q_c 是目标类别 c 的真实分布通常是一个 one-hot 变量；p_c 是模型预测样本属于类别 c 的概率；γ 是一个调整参数，它减少易分样本的权重，增加难分样本的权重。

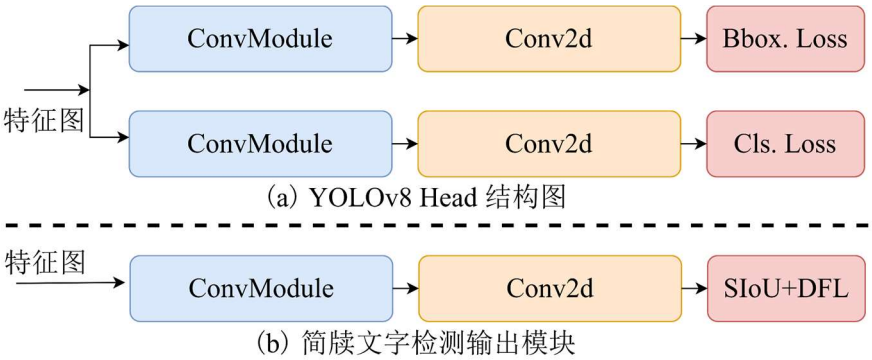
它的作用是使得预测的 Bounding Box 更加准确，从而提高目标检测的性能。CIoU 损失函数在计算边界框之间的相似性时，考虑完整的交并比（Intersection over Union，IoU）以及边界框之间的距离和角度差异。然而，对于文字检测任务，由于文字的尺度变化较大，CIoU 损失函数可能对小尺度文字的检测效果不够敏感，CIoU 的计算方式如式（3-6）。

$$LOSS_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (3\text{-}6)$$

其中 $\rho^2(b, b^{gt})$ 分别代表预测框和真实框的中心的欧式距离。c 代表的是能够同时包含预测框和真实框的最小闭包区域的对角线距离。

为解决这个问题，引入 SIoU（Scale-Invariant Overlap，SIoU）损失函数，SIoU 损失函数通过引入尺度不变性来解决这个问题，使得损失函数对目标的大小变化具有鲁棒性。SIoU 损失函数的计算基于目标框之间的重叠度量，即两个目标框的交并比。具体而言，SIoU 损失函数定义一个尺度不变的重叠度量，用于衡量预测框（预测的目标框）与真实框（标注的目标框）之间的相似程度，SIoU 损失函数的计算公式如式（3-7），这使得 SIoU 损失函数能够更好地适应文字检测任务中的尺度变化。通过使用 SIoU 损失函数，能够更准确地衡量检测边界框之间的相似性，并更好地适应不同尺度文字的检测需求。这样可以提高模型对小尺度文字的检测准确性，并增强模型对尺度变化的鲁棒性。因此，采用 SIoU 损失函数作为文字检测的损失函数，有助于改进模型的性能和稳定性。新的文本检测输出模块如图 3-11(b)所示。

$$LOSS_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (3\text{-}7)$$



(a) YOLOv8 Head 结构图

(b) 简牍文字检测输出模块

图 3-11 简牍文字检测输出模块

3.3 实验结果与讨论

3.3.1 模型评估指标

确保模型对简牍图像中文字定位的准确性，在本章构建的模型完成训练之后，本章节采用几个不同的性能指标来衡量模型的检测性能，包括准确率 Precision、召回率 Recall 和 Hmean 即 F1 值。几个性能评价指标的计算方式如下：

（1）准确率（Precision）的定义为准确检索到的文本框数量与所检测到文本

框数量的比值，假定准确检索到的文本框的数量为M，要被准确检测到文本框的数量为M_a其计算方式如式（3-8）：

$$Precision = \frac{M}{M_a} \tag{3-8}$$

（2）召回率（Recall）的定义为准确检索到的文本框数量语序要被准确检索的文本框数量的比值，假定准确检索到的文本框的数量为M，要被准确检测到文本框的数量为M_b其计算方式如式（3-9）：

$$Recall = \frac{M}{M_b} \tag{3-9}$$

（3）Hmean 是文本检测任务中应用最广泛的评测指标之一，因其计算检测准确率（Precision）与召回率（Recall）之间的调和平均数（Harmonic mean, H-mean），故得名 Hmean。记精度为P，召回率为R，则 Hmean 的计算方式如式（3-10）：

$$Hmean = 2 \times \frac{P \times R}{P + R} \tag{3-10}$$

3.3.2 实验参数设置

本章实验使用表 2-1 所示的实验环境来实现和训练网络。本算法使用 ADAM 来优化网络权重，$\beta 1$=0.9，$\beta 2$ =0.999，学习率设置为 0.0001。利用 GPU 加速训练过程，将 batch size 设置为 2。

3.3.3 YOLOv8 简牍文字检测模型评估

本实验研究基于 YOLOv8 的单简多字检测模型的网络结构，验证不同模块在改进前后，对于整个模型的重要性。由第二章实验可知可变形卷积计算消耗大，因此本实验只验证一层可变形卷积对于模型结构的影响。

DeConv YOLOv1：基于 YOLOv8 的单简多字检测模型，可变形卷积置于图 3-7 所示位置，并且采用图 3-11（b）所示的文字检测输出模块。

DeConv YOLOv2：不改变 DeConv YOLOv1 模型的文字检测输出模块的基础上，将可变形卷积层置于骨干网络 P1，P2 之间。

DeConv YOLOv3：不改变 DeConv YOLOv1 模型的骨干网络结构，采用原始 YOLOv8 的检测输出模块。

DeConv YOLOv4：不改变 DeConv YOLOv2 模型的骨干网络结构，采用原始

YOLOv8 的检测输出模块。

YOLOv8-D：原始 YOLOv8 模型，去除其预测输出模块中的分类分支，只保留位置检测分支。

由表 3-1 中实验结果可发现 DeConv YOLOv1 与 DeConv YOLOv2 中可变形卷积相较于提取普通卷积的特征图来说，直接对原始图像的特征提取，其特征提取效果更好，这一点从实验结果中也得到证实，v1 比 v2 的结果在准确率、召回率和 Hmean 中均有所提升。并且可变形卷积提取特征图的特征性能，可能不如普通卷积的提取性能，这一结论与表 3-1 中 DeConv YOLOv2 与 YOLOv8-D 的实验结果也能证明。DeConv YOLOv1 与 DeConv YOLOv3、DeConv YOLOv2 与 DeConv YOLOv4 这两组实验结果证明，本文中采用 SIoU 替换原本的 CIoU 进行模型训练对于模型最终检测结果有比较好的提升，该策略帮助模型更好的学习文字位置信息。

表 3-1 简牍文字检测模型评估结果

模型方法	准确率（Precision）	召回率（Recall）	Hmean
DeConv YOLOv1	**87.9%**	**82.1%**	**84.9%**
DeConv YOLOv2	86.3%	80.2%	83.1%
DeConv YOLOv3	86.4%	81.7%	84.0%
DeConv YOLOv4	85.3%	79.4%	82.2%
YOLOv8-D	86.5%	81.6%	84.0%

3.3.4 对比实验

在现实世界的应用中，目标检测任务的成功既依赖于准确性，也取决于检测效率，后者是决定技术能否实际部署的关键。高效的检测系统能即时处理输入并快速响应，确保机器在实时环境下做出正确决策，满足应用需求。因此，模型的推理时间（Processing time）成为衡量其性能的重要指标，即测试集中每个图像的平均处理时间，来对基于 YOLOv8 的单简多字检测模型进行综合评价。

表 3-2 简牍文字检测模型对比结果

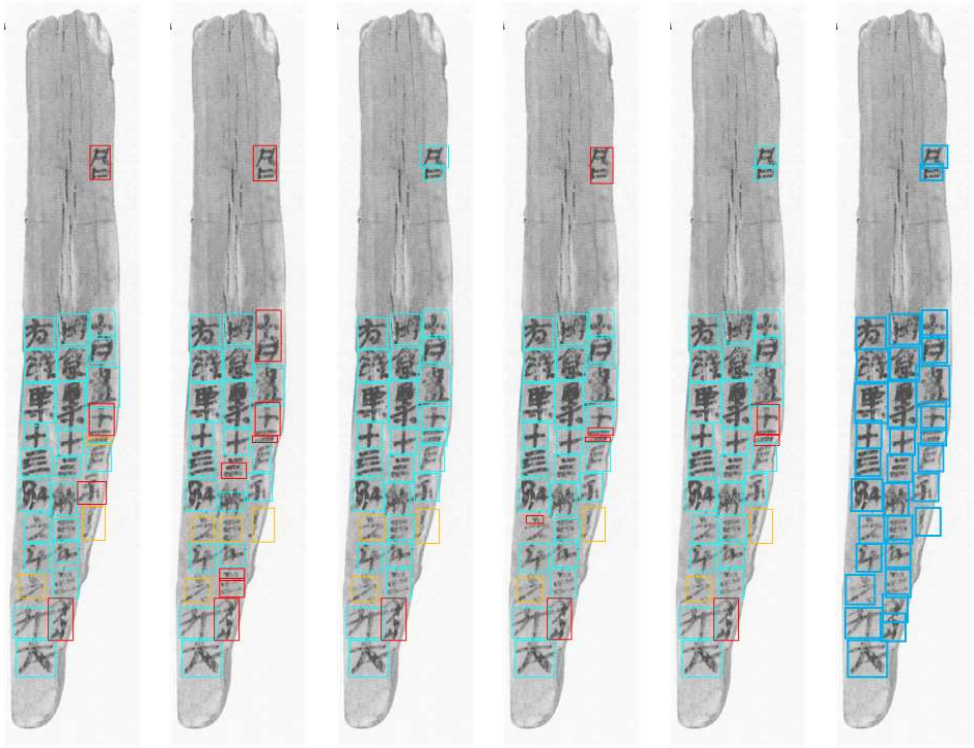
模型方法	准确率（Precision）	召回率（Recall）	Hmean	推理时间（ms）
DETR	86.3%	81.2%	83.7%	**47.3**
Faster R-CNN	82.1%	75.3%	78.6%	56.3
YOLOv8	86.5%	81.6%	84.0%	48.8

续表 3-2 简牍文字检测模型对比结果

模型方法	准确率（Precision）	召回率（Recall）	Hmean	推理时间（ms）
DBNet	86.1%	82.0%	84.0%	47.5
DeConv YOLO	**87.9%**	**82.1%**	**84.9%**	49.8

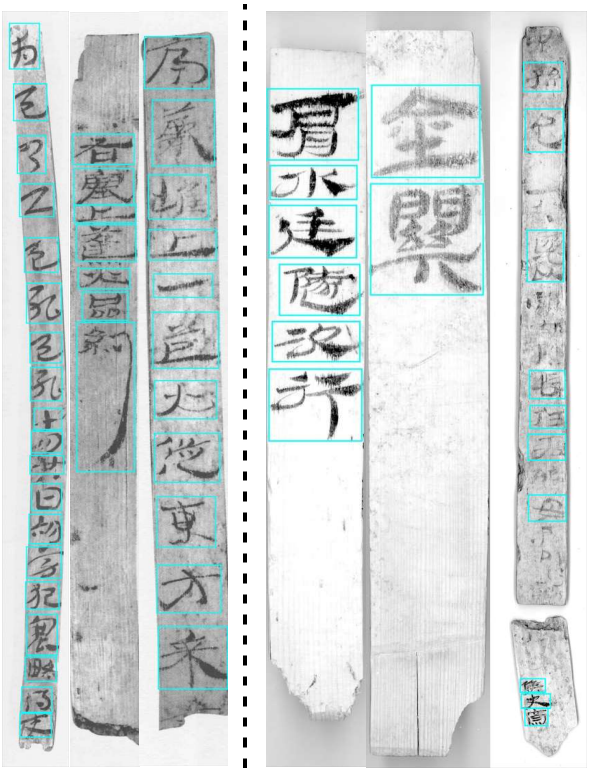
为探究提出模型与其他主流检测模型的优越性，本章在所构建的简牍文字检测数据集上进行评估，分别有目标检测算法 DETR、Faster R-CNN、YOLOv8 和文字检测算法 DBNet。图 3-12 展示不同模型在测试图像中的检测结果，可以看出 DeConv YOLO 模型基本可以检测出图像中的文字，虽然依旧有部分文字未检测出，相较于其他几种模型来说结果可以接受，红色显示为各个模型检测错误的文字，黄色显示各个模型漏检的文字。表 3-2 显示各种检测模型在简牍文字检测数据集上的评估结果，首先对比双阶段的检测方法 Faster R-CNN，由于其双阶段的模型结构导致其在计算速度上并未有任何优势，并且其检测结果相对也是最差的。其次，对比本章模型的基础模型 YOLOv8，虽然 YOLOv8 中引入可变形卷积导致其计算量增加从而导致推理时间的增加，DeConv YOLO 中可变形卷积引入对于适应文字多尺度变化和位置多变情况的考虑，因此在准确率、召回率和 Hmean 均有提升。之后便是与 CNN 结合 Transformer 的模型 DETR，该方法由于没有采用多尺度特征来检测导致对小目标检测能力较差，因此在简牍文字检测数据集中的表现略逊于本文提出的模型，且 DETR 模型的训练收敛时间远长于其他模型。最后与文字检测中经典算法 DBNet，虽然 DBNet 通过使用可微分的二值化模块，简化基于分割方法的文字检测方法，这一点从实验结果中也能证明其推理速度十分快速，但由于其结构特点却无法适应文字规整度差的简牍图像，从而导致准确率不如本文提出的模型。

其他部分居延新简检测结果，如图 3-13（a）所示，可以看出本章模型可以准确检测出简牍图像中文字的位置。本文收集的数据集主要基于居延新简，为验证本文模型的鲁棒性和泛化能力，也在其他地区出土的汉代简牍图像中进行检测，结果如图 3-13（b）所示，本章的模型基本可以检测出地湾汉简中的文字，证明本模型的鲁棒性和泛化性能较好，但是如果需要更好的检测效果，还是需要基于地湾汉简进行模型的微调训练。



DETR　Faster R-CNN　YOLOv8　DBNet　DeConv YOLO　真实标签

图 3-12 各算法在测试集上结果可视化



(a) 居延新简　(b) 地湾汉简

图 3-13 简牍文字检测模型结果展示

3.4 本章小结

本章的主要研究内容是针对简牍上文字规整度较差、文字大小不一且位置多变的问题进行探究。传统的检测算法在面对这些问题时容易出现漏检和错检的情况。为解决这一问题，首先引入可变形卷积，这是一种能够自适应调整卷积核形状的卷积操作。通过使用可变形卷积，DeConv YOLO 能够更好地适应简牍上文字的多样性和规整度较差的特点。这种卷积操作能够捕捉到文字在不同尺度和位置上的特征，从而提高文字的检测效果。此外 YOLOv8 作为基础架构，为适应简牍上文字的大小不一和位置多变的情况，通过调整损失函数，特别是引入 SIoU 损失函数，来提高模型对不同尺度目标的检测精度。SIoU 损失函数考虑目标的尺度变化，能够更好地处理简牍上文字的大小不一的情况。通过在简牍文字检测数据集测试，验证上述方法的有效性。实验结果表明，DeConv YOLO 模型在面对简牍图像中文字规整度较差、文字大小不一且位置多变的问题时具有较好的适应性。它能够提升检测结果的准确性，有效地解决传统算法容易出现的漏检和错检问题。

第4章 简牍文字识别软件设计与开发

为实际应用本文提出的算法，并验证其有效性，本章将详细讲述基于前文提出的文字识别和文字检测模型，开发的简牍文字识别和检测软件。本系统具备以下主要功能：简牍图像的导入、文字位置的自动检测、文字的准确识别，以及识别结果的图像保存。通过这个软件，用户可以上传简牍图像，系统将自动检测并识别图像中的文字内容，并将检测到的汉字以图像形式保存和展示。本文的研究基于 PyTorch 1.8.1 框架，使用 Python 3.8 编程实现所有算法。为确保算法顺利集成至软件平台，本章使用 PyCharm 和 PyQt5 进行系统界面设计与功能模块开发。PyCharm 作为 IDE，提供代码编辑、调试和版本控制的高级功能，提升开发效率。PyQt5 将 Qt 库与 Python 结合，便于跨平台 GUI 创建，其丰富的控件和直观设计工具，打造既美观又用户友好的界面。

本章首先对该系统的开发环境和平台总体框架进行简要介绍，随后详细介绍软件平台的各个功能模块，并展示相应的系统界面。最后分析总结整个系统平台的实际应用。

4.1 简牍文字识别软件设计

4.1.1 文字识别软件平台需求分析

在研究中观察到，当前进行文字文本标注的工具功能较为基础，操作者需手动进行文本框的拖动、尺寸调整以及角度校正，这些操作不仅重复且耗时。为解决这一问题，本文提出的居延新简文字识别软件旨在提供一个效率更高的文本标注解决方案。此软件能够自动展示检测到的文本框，并允许用户轻松地调整框的大小和位置，同时获取标准化的标注数据。通过这个平台，标注人员无需从头开始为每张图片画框。相反，他们可以直接在模型提供的预检测结果基础上，对不精确的文本框进行微调。这种方法不仅显著提高标注的效率，而且还有效降低人力成本。基于上述用户需求分析，文字识别软件平台的设计应当具备以下几点特性：

准确性：软件在检测到简牍图像后，可以准确检测图像中的文字位置，并可以在编辑器界面进行预测框的修改等操作，最终输出对图像中文字的识别结果；并且对于错误格式的文件可以返回对应的错误信息

便捷性：用户界面直观易用，即使是初次接触的用户也能快速上手，拥有自动

化的识别流程，可以高效处理大量的文本图像，并提供准确的识别结果。

4.1.2 文字识别软件平台整体框架

该平台是一个易于操作的简牍文字识别软件。其主要功能就是对输入的简牍图像上的文字进行检测与识别。首先需要加载硬盘中的简牍图像进行数据读取，并将获取的数据利用检测模型进行文字位置的检测；其次文字检测模型将检测到的文字位置信息在原始图像中标识；最终通过文字识别模型进行对单个文字图像进行识别并输出识别结果。该简牍文字识别系统的整体流程如图 4-1 所示。

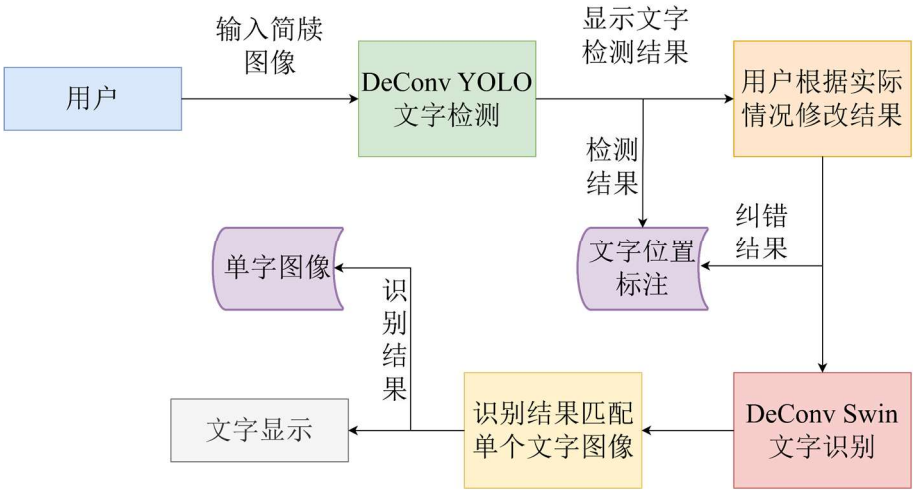


图 4-1 简牍文字识别软件技术路线图

（1）导入待识别的简牍图像，若导入为空或者不支持的格式，会有相应的提示；

（2）导入成功之后，利用 DeConv YOLO 模型进行检测，并显示图像中文字位置，若有检测错误的情况可由用户进行修改，并将最终检测结果保存为 xml 文件。

（3）用户修改检测结果，完成后进行单张文字图像的获取，利用 Deconv Swin 模型识别文字，最终输出识别结果，并将识别结果与文字图像匹配并以文件夹的形式保存。

4.2 文字识别软件平台实现

前三节对文字识别系统的开发环境、需求分析和平台总体架构做详细介绍，本节将从系统的图片导入、单文字图片获取、文字识别三个模块进行详细介绍，并展示其对应界面。

4.2.1 图像导入模块

该平台的简牍图像输入方式非常灵活，用户可以直接从本机导入一张待识别的汉代简牍图像。系统支持多种常用图像格式，例如 PNG、JPEG、BMP 等。这样的设计考虑到用户的使用习惯和图像存储的普遍格式，使得用户能够方便地将自己拥有的汉代简牍图像导入到软件平台中进行处理和识别。在图像导入过程中，系统会对用户的输入进行验证和判断。如果输入为空或者选择不支持的图像格式，系统会自动弹出对应的提示信息，提醒用户重新选择合适的图像文件。这样的设计能够避免用户因为错误的输入而导致操作失败或产生不必要的困惑。为更好地展示系统的图像导入功能和操作界面，图 4-2 展示在软件平台中进行图像导入的界面。界面简洁直观，用户可以清晰地看到导入按钮和相关的提示信息。这样的设计使得用户能够快速而准确地完成图像导入的操作。

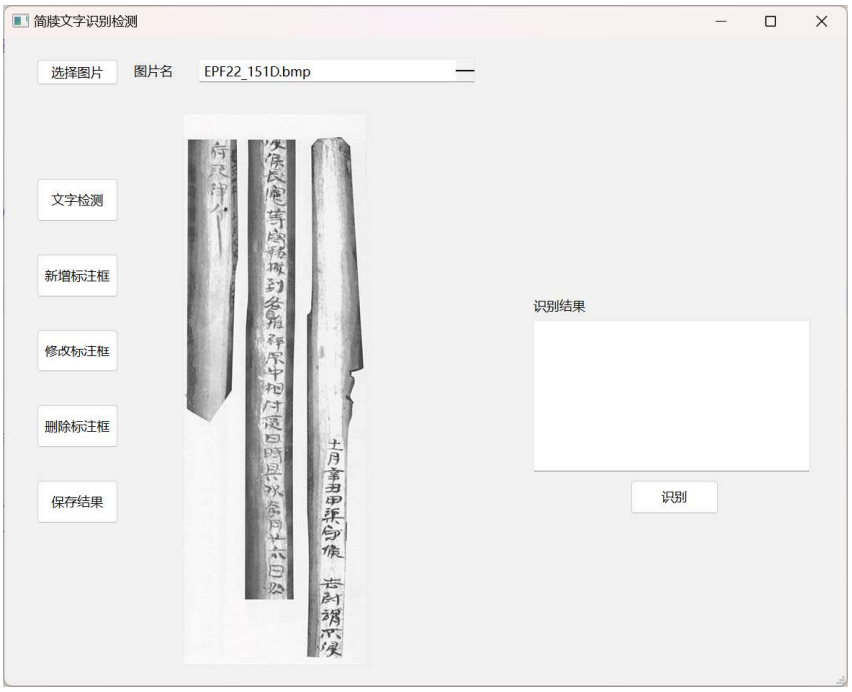


图 4-2 简牍图像导入界面

4.2.2 文字检测模块

文字检测模块在该平台中起着关键的作用，它能够检测简牍图像中每个文字在图像中的位置，为下一步的文字识别提供准确的边界框信息。然而，由于简牍图像存在噪声、纹路众多、墨迹退化等因素，文字的检测过程也受到一定的干扰。为解决这个问题，该模块引入手动干预的功能，使得用户可以对检测结果进行修正，包括修正错检和漏检的文字。同时，该模块还将修正后的文字框以蓝色显示在简牍图像中，并允许用户保存每个文字的图像。最后，每张简牍图像对应的文字图像会

被保存在一个文件夹中。

手动干预是为弥补自动文字检测算法的不足之处，并提供更准确的结果。用户可以通过界面中的交互操作，对检测结果进行修正。对于错检的文字，用户可以选择将其删除或者调整边界框的位置，以排除误检的情况。对于漏检的文字，用户手动添加相应的边界框，确保没有遗漏任何文字。这样的手动干预机制可以根据具体情况进行灵活调整，提高检测结果的准确性。

为便于用户查看和分析修正后的结果，修正后的文字框以蓝色显示在简牍图像中。这样，用户可以直观地观察每个文字的位置，并对其进行进一步的修改或确认。此外，用户还可以选择保存每个文字的图像，以便后续的研究和分析。这样的设计使得用户能够方便地获取并管理每个文字的图像数据。图 4-3 展示在软件平台中进行文字检测的界面。将经过调整后的结果保存到 xml 文件中。

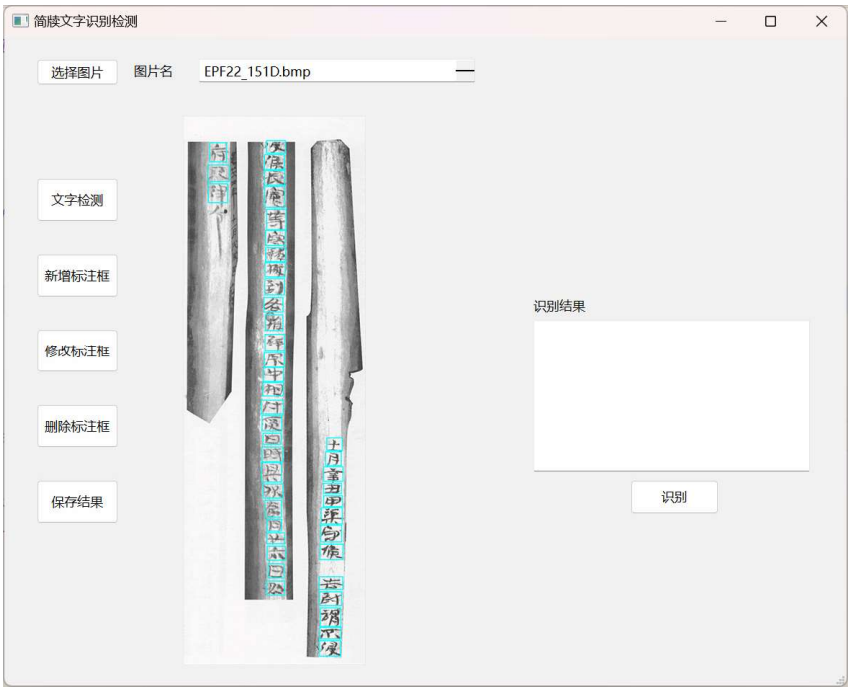


图 4-3 简牍图像检测界面

4.2.3 文字识别模块

文字识别模块，通过文字检测模块获取的文字框获取对应的文字图像，之后将文字图像输入文字识别模型中进行文字识别，并将每一个识别的文字在文本框中输出如图 4-4 所示。并且可以由用户进行结果的编辑，并将编辑之后的结果进行存储，同一类文字图像保存到对应命名的文件夹中。同时修改标注文件，为每个文字框添加具体的文字信息。

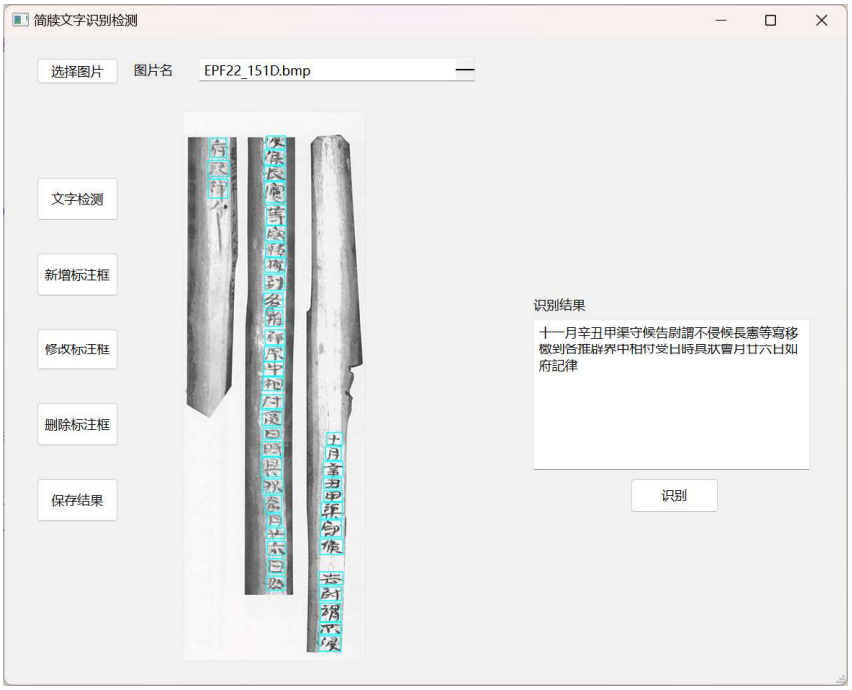


图 4-4 简牍图像识别界面

4.3 文字识别软件平台应用

该软件平台能够快速检测和识别简牍图像中的文字，并输出对应的汉字。这项工作实现对汉代简牍文字的研究与计算机文字识别和检测的结合，为非专业人士进行汉代简牍研究提供便利。目前，大多数汉代简牍文字的识别工作都依赖于专业人士的辨认，这个过程耗时且需要大量人力，大大增加研究成本。然而，通过使用该软件，非专业人士也能够快速准确地识别汉代简牍文字。

该软件平台的功能包括图像文字位置的检测和文字识别。它能够准确地检测出简牍图像中的文字位置，确定文字所在的区域。通过先进的计算机文字识别技术，软件能够将这些区域中的文字转换为对应的汉字。该过程快速高效，大大节省非专业人士进行汉代简牍研究的时间和精力。

使用这一软件平台，非专业人士可以通过导入待识别的简牍图像，快速获得对应的汉字信息，而无需依赖专业人士的辨认。这样的便利性对于推动汉代简牍研究的广泛参与具有重要意义。它降低研究的门槛，使更多的人能够参与到对汉代简牍文字的研究中来，推动学术界对汉代文化和历史的更深入理解。

总之，该软件的使用为非专业人士提供快速识别汉代简牍文字的能力，极大地促进汉代简牍研究的发展。它的便利性和高效性使得研究过程更加普及和高效，为汉代文化的研究和传承作出重要贡献。

4.4 本章小结

本章基于 PyCharm、PyQt5 和 Python 设计实现汉代简牍文字识别系统，可以快速、准确地满足简牍文字识别的需求，且也证明本文简牍文字基于 YOLOv8 的单简多字检测算法和面向字型多变简牍单文字的可变形卷积分类识别模型的可行性与实用性。

总结与展望

主要结果与创新点

本论文以可变形卷积为基础结合不同的神经网络模型，研究单个简牍文字图像的识别问题和完整简牍图像的文字检测问题。针对单个简牍文字图像的识别问题着重研究简牍文字形态、结构和尺度多变等问题，并提出对应的解决方案。对于完整简牍图像的文字检测，更加侧重于研究如何解决简牍文本的规整度较差导致的文字的大小和位置变化，导致漏检和误检的问题。基于以上研究思路，并按照问题分析，数据收集、模型研究、方法应用的整体思路，将深度学习、图像识别、目标检测等算法引入简牍文字识别与检测中，主要结果以及创新点包括以下部分：

（1）面向字型多变简牍单文字的可变形卷积分类识别模型。简牍文字所处年代和载体的不同，同一文字有不同写法，导致文字形态和结构多变。书写载体的大小不一和单个载体篇幅限制导致文字尺度多变。针对以上问题本文提出 DeConv Swin 简牍文字识别模型，模型包含两个显著特点，即可变形卷积的不规则采样能力和 Swin Transformer 的层级注意力机制的多尺度特征提取能力，可以有效解决因文字形态、结构和尺度多变导致的简牍文字识别难的问题。通过实验对比分析显示，DeConv Swin 能很好地进行简牍文字识别，模型的准确率、精确率和召回率分别为 83.5%、81.9%和 79.9%，并通过消融实验证明各个模块的有效性，证明模型设计合理。

（2）基于 YOLOv8 的单简多字检测与识别模型。长时间的掩埋使得简牍的物理形态发生改变，这导致采集的图像包含大量噪声和文字的扭曲。此外，简牍上文字的尺寸不统一，位置随意，这些因素使得传统的目标检测模型或文字检测模型在应用时容易出现漏检和误检的问题。针对以上问题，本文采用了双边滤波算法来降低图像的背景噪声，并提出一种 DeConv YOLO 模型来适应简牍图像中文字尺寸的多样性和位置的不确定性。模型通过集成可变形卷积，增强了对文字位置变化的适应能力；同时，类 PANet 的网络结构改进，实现了多尺度特征的有效融合，以适应不同大小文字。此外，对模型的损失函数进行了优化，以提高对居延新简文字位置的检测准确性。针对模型训练的需求，本文构建居延新简文字检测数据集。通过实验对比分析显示，模型能够很好的对简牍图像中的文字位置进行检测，模型的准确率和召回率分别为 87.9%和 82.1%。并通过消融实验证明各个模块的有效性，证明模型设计合理。

（3）居延新简文字识别软件平台。本文在对简牍文字识别与检测模型进行理论研究的同时，兼顾研究成果的现实应用，将理论研究与现实问题相结合，基于深度学习的居延新简文字识别与检测方法设计居延新简文字识别软件平台。该平台包括数据导入、文字检测、检测框修改和文字识别等四个主要模块，通过这些模块的配合，实现对简牍文字的识别和检测的实际应用。

总体来说，本文对于解决简牍文字识别存在的文字不同的写法，文字尺度、形态和结构多变问题，和简牍文字检测中文字的大小和位置不确定的问题提出相应的解决办法，并且在相关数据集上进行验证方法的有效性。这为后序相关研究提供有效的宝贵经验，为简牍文字加快数字化保护工作提供有效的解决途径。

研究展望

竹木简牍是秦汉时期留下的宝贵档案，是中古时期中国的知识宝库，国家越来越重视现代科技强化古籍典藏保护、修复及其综合利用。基于深度学习的居延新简文字识别与检测方法的研究为简牍文字的数字化保护和利用提供新的途径。然而，当前的研究还存在一些挑战和局限性，未来的研究可以从以下几个方面展望和改进。

（1）目前构建的居延新简文字识别和检测数据集可能还不够大规模和多样化，数据集的扩充和完善是未来研究的一个重要方向，简牍文字具有丰富的地区、时期和书法风格的变化，因此，进一步扩充数据集，包括不同地区、不同时期、不同书写风格的简牍文字样本，以提高模型的泛化能力和鲁棒性。

（2）除形态和结构外，简牍文字还包含丰富的笔画和书法特征，未来的研究可以探索将图像信息与文本信息相结合，多模态信息融合是提高简牍文字识别和检测准确性的关键，利用多模态数据，如书法笔迹数据、纸张材质数据等，进一步提高识别和检测的准确率和稳定性。

参考文献

[1] 习近平. 习近平：在哲学社会科学工作座谈会上的讲话(全文)[EB/OL]. (2016-05-17).http://www.scio.gov.cn/31773/31774/31783/Document/1478145/1478145.htm

[2] 中共中央办公厅 国务院办公厅印发《关于推进实施国家文化数字化战略的意见》_最新政策 _ 中国政府网 [EB/OL]. (2022-05-22).https://www.gov.cn/zhengce/2022-05/22/content_5691759.htm

[3] 王祖龙, 万子昂. 近年来出土简牍书迹的整理与研究综述[J]. 书法报, 2023: 012.

[4] 甄涛. 简牍中国丨“冷门”不冷 “绝学”有继 千年简牍焕发时代光彩[EB/OL].(2023-08-05) https://news.cctv.com/2023/08/05/ARTIWeofD2OzXns54f18YKHj230805.shtml.

[5] 莫伯峰, 张重生. 人工智能在古文字研究中的应用及展望[J]. 中国文化研究, 2023, (2): 47–56.

[6] Hinton G.E., Osindero S., Teh Y.-W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527–1554.

[7] Rumelhart D.E., Hinton G.E., Williams R.J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533–536.

[8] Lecun Y., Bottou L., Bengio Y., et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278–2324.

[9] Krizhevsky A., Sutskever I., Hinton G.E. ImageNet Classification with Deep Convolutional Neural Networks[C]//Advances in Neural Information Processing Systems. 2012, 25. 84–90.

[10] Zeiler M.D., Fergus R. Visualizing and Understanding Convolutional Networks[C]//Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13. Springer International Publishing, 2014: 818-833.

[11] Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J].Computer Science, 2014.DOI:10.48550/arXiv.1409.1556.

[12] 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.

[13] Ioffe S., Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[C]//Proceedings of the 32nd International Conference on Machine Learning. 2015: 448–456.

[14] Szegedy C., Vanhoucke V., Ioffe S., et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.

[15] Szegedy C., Ioffe S., Vanhoucke V., et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2017, 31(1). 4278–4284.

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

[17] Huang G., Liu Z., Van Der Maaten L., et al. Densely Connected Convolutional Networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2261–2269.

[18] 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.

[19] Howard A, Sandler M, Chu G, et al. Searching for mobilenetv3[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1314-1324.

[20] Sandler M., Howard A., Zhu M., et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 4510–4520.

[21] Dai J., Qi H., Xiong Y., et al. Deformable convolutional networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 764-773.

[22] Zhu X., Hu H., Lin S., et al. Deformable ConvNets v2: More Deformable, Better Results[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 9308-9316.

[23] Vaswani A., Shazeer N., Parmar N., et al. Attention is All you Need[C]//Advances in Neural Information Processing Systems. 2017, 30. 6000–6010.

[24] 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.

[25] Liu Z., Lin Y., Cao Y., et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows[C]//2021 IEEE/CVF International Conference on Computer Vision. 2021: 9992–10002.

[26] Liu Z., Mao H., Wu C.-Y., et al. A ConvNet for the 2020s[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 11976-11986.

[27] Girshick R., Donahue J., Darrell T., et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[28] He K., Zhang X., Ren S., et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.

[29] Girshick R. Fast R-CNN[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[30] Ren S., He K., Girshick R., et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[C]//Advances in Neural Information Processing Systems. 2015, 28. 1137-1149.

[31] Lin T.-Y., Dollar P., Girshick R., et al. Feature Pyramid Networks for Object Detection[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[32] Liu S., Qi L., Qin H., et al. Path Aggregation Network for Instance Segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.

[33] Redmon J., Divvala S., Girshick R., et al. You Only Look Once: Unified, Real-Time Object Detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[34] Redmon J., Farhadi A. YOLO9000: Better, Faster, Stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[35] Redmon J., Farhadi A. YOLOv3: An Incremental Improvement[J].arXiv preprint arXiv:1804.02767, 2018.

[36] Bochkovskiy A., Wang C.-Y., Liao H. YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[37] Zheng Z., Wang P., Liu W., et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 12993-13000.

[38] Ge Z., Liu S., Wang F., et al. YOLOX: Exceeding YOLO Series in 2021[J]. arXiv preprint arXiv:2107.08430, 2021.

[39] Li C., Li L., Jiang H., et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications[J]. arXiv preprint arXiv:2209.02976, 2022.

[40] Wang C.-Y., Bochkovskiy A., Liao H.-Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 7464–7475.

[41] Carion N., Massa F., Synnaeve G., et al. End-to-End Object Detection with Transformers[C]// European conference on computer vision. Cham: Springer International Publishing, 2020: 213-229.

[42] Fang Y., Liao B., Wang X., et al. You Only Look at One Sequence: Rethinking Transformer in Vision through Object Detection[J]. Advances in Neural Information Processing Systems, 2021, 34: 26183-26197.

[43] Li Y., Mao H., Girshick R., et al. Exploring Plain Vision Transformer Backbones for Object Detection[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 280-296.

[44] 周新伦, 李锋, 华星城, 等. 甲骨文计算机识别方法研究[J]. 复旦学报(自然科学版), 1996, (5): 481–486.

[45] 栗青生, 杨玉星, 王爱民. 甲骨文识别的图同构方法[J]. 计算机工程与应用, 2011, 47(8): 112–114.

[46] 刘志基. 简析古文字识别研究的几个认识误区[J]. 语言研究, 2019, 39(4): 89–95.

[47] 顾绍通. 基于拓扑配准的甲骨文字形识别方法[J]. 计算机与数字工程, 2016, 44(10): 2001–2006.

[48] 顾绍通. 基于分形几何的甲骨文字形识别方法[J]. 中文信息学报, 2018, 32(10): 138–142.

[49] 田园. 基于深度度量学习的战国简文字识别技术[D]. 河南大学,2020.

[50] 刘梦婷. 基于深度卷积神经网络的甲骨文字识别研究[D].郑州大学, 2020.

[51] 张颐康, 张恒, 刘永革, 等. 基于跨模态深度度量学习的甲骨文字识别[J]. 自动化学报, 2021, 47(4): 791–800.

[52] Wu L., Zhang C., Xu M., et al. Ancient Chinese Recognition Method Based on Attention Mechanism[C]//2021 7th IEEE International Conference on Network Intelligence and Digital Content. 2021: 309–313.

[53] 林小渝, 陈善雄, 高未泽, 等. 基于深度学习的甲骨文偏旁与合体字的识别研究[J]. 南京师大学报(自然科学版), 2021, 44(2): 104–116.

[54] 朱旭. 基于改进 PUGAN 和 CNN 模型的甲骨文分类算法研究[D]. 大连海事大学, 2022.

[55] 吴炫奇. 基于深度学习的商周金文文字识别研究[D]. 华东师范大学, 2022.

[56] 石佳钰. 基于生成对抗网络的手写蒙古文字元识别研究[D]. 内蒙古师范大学, 2022.

[57] 刘绪兴. 基于深度学习的古文字识别研究[D]. 西南大学, 2022.

[58] Liu X., Gao W., Li R., et al. One shot ancient character recognition with siamese similarity network[J]. Scientific Reports, 2022, 12(1): 14820.

[59] Assael Y., Sommerschield T., Shillingford B., et al. Restoring and attributing ancient texts using deep neural networks[J]. Nature, 2022, 603(7900): 280–283.

[60] 李沿增. 基于目标检测和知识图谱的古文字识别研究[D]. 吉林大学, 2023.

[61] 郝超华. 西夏文字的无监督识别算法研究与应用[D]. 北方民族大学, 2023.

[62] 毛亚菲, 毕晓君. 改进 ResNeSt 网络的拓片甲骨文字识别[J]. 智能系统学报, 2023, 18(3): 450–458.

[63] 史小松, 黄勇杰, 刘永革. 基于阈值分割和形态学的甲骨拓片文字定位方法[J]. 北京信息科技大学学报(自然科学版), 2014, 29(6): 7-10+24.

[64] 潘振赣. 基于模糊聚类的碑文拓片图像分割算法研究[D]. 苏州大学, 2011.

[65] 王书敏. 基于纹理特征方法的甲骨拓片文字定位研究[J]. 信息系统工程, 2020, (12): 141–142.

[66] Yu D., Li X., Zhang C., et al. Towards Accurate Scene Text Recognition With Semantic Reasoning Networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition . 2020: 12110–12119.

[67] He M., Liao M., Yang Z., et al. MOST: A Multi-Oriented Scene Text Detector with Localization Refinement[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 8809–8818.

[68] Yan R., Peng L., Xiao S., et al. Primitive Representation Learning for Scene Text Recognition[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition . 2021: 284–293.

[69] Fang S., Xie H., Wang Y., et al. Read Like Humans: Autonomous, Bidirectional and Iterative Language Modeling for Scene Text Recognition[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 7094–7103.

[70] Baek J., Matsui Y., Aizawa K. What If We Only Use Real Datasets for Scene Text Recognition? Toward Scene Text Recognition With Fewer Labels[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition . 2021: 3112–3121.

[71] Meng L., Lyu B., Zhang Z., et al. Oracle Bone Inscription Detector Based on SSD[C]//New Trends in Image Analysis and Processing–ICIAP 2019: ICIAP International Workshops, BioFor, PatReCH, e-BADLE, DeepRetail, and Industrial Session, Trento, Italy, September 9–10, 2019, Revised Selected Papers 20. Springer International Publishing, 2019: 126-136.

[72] 王浩彬. 基于深度学习的甲骨文检测与识别研究[D]. 华南理工大学, 2020.

[73] 陈善雄, 韩旭, 林小渝, 等. 基于 MSER 和 CNN 的彝文古籍文献的字符检测方法[J]. 华南理工大学学报(自然科学版), 2020, 48(6): 123–133.

[74] 邢济慈. 基于深度卷积神经网络的甲骨文字检测技术研究[D]. 郑州大学, 2020

[75] 刘芳, 李华飙, 马晋, 等. 基于 Mask R-CNN 的甲骨文拓片的自动检测与识别研究[J]. 数据分析与知识发现, 2021, 5(12): 88–97.

[76] 殷航, 张智, 王耀林. 基于 YOLOv3 与 MSER 的自然场景中文文本检测研究与实现[J]. 计算机应用与软件, 2021, 38(10): 168-172+195.

[77] 李健昱, 王慧琴, 刘瑞, 等. 复杂纹理背景下的密集骨签文字检测算法[J]. 液晶与显示, 2023, 38(9): 1293–1303.