华南理工大学

South China University of Technology

# 硕士学位论文

## 基于深度学习的甲骨文检测与识别研究


作 者 姓 名　王浩彬

学 科 专 业　信号与信息处理

指 导 教 师　黄双萍 教授

所 在 学 院　电子与信息学院



论文提交日期　2019 年 4 月



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

# The Research of Oracle Bone Script Detection and Recognition Based on Deep Learning Methods

A Dissertation Submitted for the Degree of Master

**Candidate：Wang Haobin**

**Supervisor：Prof. Huang Shuangping**

South China University of Technology

Guangzhou, China

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

**分类号：TP391** **学校代号：10561**

**学　号：201620108880**

华南理工大学硕士学位论文

# 基于深度学习的甲骨文检测与识别研究

作者姓名：王浩彬　　指导教师姓名、职称：　黄双萍　教授

申请学位级别：工学硕士　　学科专业名称：信号与信息处理

研究方向：智能信息处理系统与模式识别

论文提交日期：　2019　年　4　月　10　日　　论文答辩日期：　2019　年　6　月　4　日

学位授予单位：华南理工大学　　学位授予日期：　　年　　月　　日

答辩委员会成员：

主席：　刘付强

委员：　金连文　　黄双萍　　薛洋

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

# 华南理工大学
# 学位论文原创性声明

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律后果由本人承担。

作者签名：王浩彬　　日期：2019年6月4日

# 学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属华南理工大学。学校有权保存并向国家有关部门或机构送交论文的复印件和电子版，允许学位论文被查阅（除在保密期内的保密论文外）；学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。本人电子文档的内容和纸质论文的内容相一致。

本学位论文属于：

□保密（校保密委员会审定为涉密学位论文时间：___年__月__日），于___年__月__日解密后适用本授权书。

☑不保密，同意在校园网上发布，供校内师生和与学校有共享协议的单位浏览；同意将本人学位论文编入有关数据库进行检索，传播学位论文的全部或部分内容。

（请在以上相应方框内打“√”）

作者签名：王浩彬　　日期：2019.6.4.

指导教师签名：黄2021年　　日期　2019.6.4.

作者联系电话：　　电子邮箱：

联系地址（含邮编）：

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

# 摘　　要

甲骨文是中国乃至东亚至今已知的最早成熟文字体系，甲骨文研究不仅对了解中国汉字起源很重要，对中国乃至全世界历史的探究及文化的传承也有重要意义。

甲骨文字符检测识别一直是甲骨文研究中重要的研究内容，任务是从甲骨拓片中找到甲骨字符并确认其字符类，这是字形破译的前提。目前，这些工作都需要甲骨学专家们人工完成，效率低下，且耗费昂贵的专家资源。因此研究甲骨文字符的自动检测识别很有价值。现在，甲骨文检测领域尚是空白，而甲骨文字符识别依赖专家的特征工程，设计的系统步骤繁复。

鉴于深度学习在计算机视觉领域的成功，在甲骨文检测领域，本文首先构建了首个甲骨文检测数据集 OBCD（Oracle Bone Character Detection），为深度学习应用于甲骨字符检测工作提供基准数据库。接着，论文结合基于区域的全卷积网络（Region-based Fully Convolutional Network）和特征金字塔网络（Feature Pyramid Network），设计并搭建了基础的甲骨字符识别算法框架。进一步，针对复合字符难检测且标注良好的训练数据规模太小的问题，提出了基于字模数据的动态增广算法，在增加整体数据规模的基础上针对性地增加难检字的训练数据，强化检测模型对难检字的关注。针对类似字符的甲骨划痕易误检的问题，本文提出甲骨文字符识别辅助检测算法，过滤错误预测结果。实验证明，本文提出的动态增广算法令检测模型能更鲁棒地预测难检字，而辅以识别模型的检测算法可以召回原本置信度低的正确预测，并过滤错误预测，最终 F-measure 指标达到 83.79%。

在甲骨文识别领域，论文首先构建了一个类别数达 306，示例数达 309551 的甲骨文单字符识别数据集 OBC306，这是至今数据规模最大的甲骨文字符识别数据集，为利用深度学习解决甲骨字形识别提供了丰富的数据。接着论文建立了基于 Inception-v4 深度卷积网络的字符识别算法框架，并针对字符类样本分布不均衡问题，提出基于循环式生成对抗网络的字符增广算法，用以合成更加多样化的甲骨字符图像，平衡类样本分布，解决长尾效应对甲骨字符识别性能的影响。实验证明，本文提出的方法在数据库 OBC306 上取得 86.54%的平均识别精度。

**关键词**：甲骨文；字符检测；字符识别；生成对抗网络；深度学习



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

# Abstract

The oracle bone script is the oldest mature writing system in China and even the East Asia. The study of oracle-bone inscriptions is of great importance not only for Chinese etymologies but also for learning about the culture and history of ancient China, and even the whole world.

The detection and recognition of the oracle bone script that are the preconditions of deciphering the oracle bone scripts is one of the most important parts in the study of oracle bone scripts. However, these works have to be executed manually by the experts, so they are inefficient as well as cost the valuable energies of experts. Therefore, it is of great value to study how to automatically detect and recognize the oracle bone scripts. Nowadays, there are few works in the field of oracle bone script detection, and the works on the oracle bone script recognition rely on feature engineering conducted by experts, and the designed systems are usually complex and consist of many components.

Considering the success of deep learning in the field of computer vision, in the field of oracle bone script detection, we first construct an oracle bone script detection dataset called OBCD (Oracle Bone Character Detection), and provide the baseline dataset for the works that apply deep learning methods to oracle bone script detection. Then, we combine Region-based Fully Convolutional Network and Feature Pyramid Network to construct the basic framework for oracle bone script recognition. Further, given the problem that the oracle bone characters with multiple components are difficult to detect and the dataset is lack of the training data of these characters, we propose a dynamic data augmentation algorithm based on the type matrices to increase the overall training data and create the difficult instances specifically to help the model focus on the difficult characters. Besides, to solve the problem that many meaningless nicks on oracle bones look like the oracle bone scripts and confuse the models, we also propose an algorithm to refine our detection result with the recognition model and help filter the false predictions. Experiments show that the proposed dynamic data augmentation algorithm helps the detection model locate the difficult instances more robustly, and the proposed algorithm of refining detection with recognition can help recall the correct predictions with low confidences and filter several false predictions, and we finally achieve a F-measure of 83.79%.

Besides, in the field of oracle bone script recognition, we construct an oracle bone script dataset in character-level with 306 classes and 309551 samples, and this dataset is the largest



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

set in existence and provides the data for the works adopting deep learning methods to recognize oracle bone scripts. Then, we construct the framework for the oracle bone script recognition based on Inception-v4 model. Given the problem that the distribution of the dataset classes is quite imbalanced, we propose a data augmentation algorithm based on Cycle Generative Adversarial Network to synthesize oracle bone character images in diversity, balancing the distribution of the dataset classes and eliminating ill effects that the long tail effect causes. Experiments show that our proposed method achieve an average accuracy of 86.54% on OBC306.

**Keywords: Oracle bone script; Character detection; Character recognition; Generative Adversarial Network; Deep Learning**



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

# 目　　录





(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

# 第一章 绪论

## 1.1 研究背景及意义

甲骨文是中国乃至东亚至今为止已知的最早形式的成熟文字体系，因刻在龟壳以及兽骨上而得名。甲骨文最早于 1899 年由晚清官员王懿荣在河南省安阳市出土的甲骨上发现。河南省安阳市西北殷都区小屯村，是中国商朝晚期都城遗址“殷墟”的所在地，这些甲骨上的字符主要是商朝王室在占卜凶吉以及记录叙事的时候镌刻的，记录了殷商时期这一段距今 3600 多年的历史。因此，甲骨文对于了解中国乃至整个世界的过去有着非常重要的意义。甲骨文的发现促进了各国学者对于上古历史以及古代汉字等领域的研究，同时也促成了新学科“甲骨学”的诞生。

自甲骨文第一次发现出土之后，对于甲骨的挖掘活动便逐步开展， 但是由于早先政府并未开始重视与保护，因此许多甲骨都被民间私下挖掘，并被转手卖至各地古董商人与收藏者。据统计，截止到 1928 年，流散于世界各地的甲骨文总数已经达到上万片。之后由于日军侵华，又有许多甲骨被运往日本。1928 年，中央研究院历史语言研究所建立，国内的科学考古挖掘逐步开展，更多的甲骨逐渐重见天日。目前为此，在河南安阳殷墟被发现的带有文字图案的甲骨已经达到了 15 万余片[1]，其中国内大陆藏有 10 万余片，台湾有 3 万余片，香港藏有 100 片左右，其他国家如美国、日本、英国、加拿大等 12 个国家总共藏有 27000 片左右[2]。

在甲骨文首次出土之后，短短数年之间，便有许多学者开始进行甲骨文研究。《铁云藏龟》是甲骨文的第一本实体著录，该著录于 1903 年面世。此后，随着更多学者投入甲骨文研究，许多重要著录一一出版[3-10]。由于相当数量的甲骨片流散到世界各地，因此除了国内，国外也有许多学者进行甲骨文研究，许多相关的著录也逐步出版[5,6,8-10] 。在世界各地学者的努力下，现今国内外出版的著录基本涵盖了所有出土的甲骨片。

甲骨文相关研究领域比较广阔，基础性工作包括甲骨片挖掘、拼接缀合、辨伪等，文字考释工作则包括分期断代、字形字义识别、词汇、语法和书法风格分析等。甲骨文研究领域发展了一百多年，各个细分领域都取得了一定的成果。例如分期断代领域最具代表性的著作是董作宾发表的《甲骨文断代研究例》[11]，在该著作中，他提出十项断代标准，将出土的所有甲骨按时间分为五个不同时期，确立了基本的甲骨文分期断代体系；陈梦家在《殷墟卜辞综述》[12]讨论了出土区域与甲骨时代的关系，并进一步讨论每个时期的断代论据；李学勤等作者在《殷墟甲骨分期研究》[13]提出在分期断代时参考地层学



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

信息。这些理论促进了甲骨文分期断代的发展，也为甲骨文的字形字义解析等工作提供了帮助。

尽管在甲骨文发现之后，甲骨学发展迅速，在多个研究方向上已取得一定的研究成果，但作为距今 3000 多年的文字体系，甲骨文还有许多内容有待破译，例如破解甲骨字符的含义方面，虽然目前已有超过 15 万片甲骨被挖掘出来，在这些甲骨片上已发现了约 4500 个不同的甲骨文字符，但仅有 2000 个左右被成功破译解读[14]。因此，当前甲骨文研究仍有广阔的前景。

## 1.2 研究问题

甲骨文的破译一直受到甲骨学领域乃至整个社会的关注，而作为破译的前提工作，甲骨文的检测与识别也同样是甲骨学领域的重要研究内容，它们的任务是从甲骨上找到甲骨字符位置以及确定其字符类别。尽管甲骨文的检测与识别工作已取得一定的成果[15-19]，但当前仍然面临着挑战，因此也影响了整体破译工作的进展。

目前限制甲骨文检测与识别任务的挑战主要有以下几个方面。首先，目前甲骨拓片上每一个字形字符的检测与识别，都需要甲骨学专家们手动查阅大量相关资料并结合自身经验才可以完成，这个过程消耗了大量的时间以及昂贵的专家资源，因此，研究效率低下且成本极高，这严重阻碍了甲骨文研究工作的顺利进展。

如果需要提高这方面的工作效率，实现甲骨文自动检测与识别技术是十分必要的。而要实现甲骨文字形研究的自动化，需要借助计算机技术，因此需要大数据量的标注数据集来架起计算机技术以及甲骨文字形研究的桥梁。尽管目前许多已经有许多甲骨文数字化项目，许多甲骨文资料都已通过数字化存储于线上，但是目前仍然缺少数据量大的、具有完备标注的甲骨文字符级别数据集，这一点影响了自动检测识别算法的研究。因此，总的来说，当前甲骨文的检测与识别工作需要自动化技术来提升自身的工作效率，同时也需要足够的甲骨文字符字形数据作为研究的数据基础。

## 1.3 本文主要工作及贡献

本文的总体目标是针对当前甲骨文检测与识别工作的问题需求，结合机器学习理论和计算机视觉理论，对甲骨文检测与识别的自动化技术进行深入研究。根据这个目标，本文的主要工作及贡献包括如下几点：

1. 在甲骨文检测领域，本文首先构建了甲骨文检测数据集 OBCD（Oracle Bone Chracter Detection），为深度学习技术应用于甲骨字符检测工作提供基准数据库。该数据库是



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

首个甲骨字符检测数据集，其素材全部来源于现有的甲骨著录[3]和数字化项目成果[20]。OBCD 数据库包含甲骨拓片图像 5838 张，甲骨字符示例 48821 个，用矩形框进行甲骨字符的标注。

2. 在甲骨文检测领域，论文引入基于区域的全卷积网络 R-FCN（Region-based Fully Convolutional Network）算法，结合特征金字塔网络 FPN（Feature Pyramid Network）这种网络结构的多尺度设计，搭建了基础的甲骨字符检测算法框架。进一步，针对复合字符难检测且标注良好的训练数据规模不够大的问题，提出了基于字模数据的动态增广算法，在增加整体数据规模的基础上针对性地增加难检字的训练数据，强化检测模型对复合难检字的关注。针对类似字符的甲骨划痕易误检的问题，本文提出甲骨文字符识别辅助检测算法，在存在许多容易误检的甲骨划痕的情况下帮助检测模型减少错误的检测结果。实验证明，本文提出的动态增广算法令检测模型能更鲁棒地预测难检字，而辅以识别模型的检测算法可以召回原本置信度低的正确预测，并过滤错误预测，最终的综合指标 F-measure 达到 83.79%。
3. 在甲骨文字符识别领域，论文构建了一个数据量大、具有完备标注的甲骨文单字符数据集 OBC306（Oracle Bone Character 306），该数据库包括 309551 张甲骨字符图像，覆盖 306 个不同类的甲骨文字，这是目前为止数据量最大的甲骨文单字符数据集。OBC306 数据库样本均从原始甲骨拓片或者数字化图像中切割得到，体现了甲骨字符识别问题的重要挑战，包括多样性甲骨字符的外观呈现，素材获取和字符样例制作过程导致的复杂噪声类型，因甲骨字符字频导致的类分布不均衡的长尾效应等。该数据库为甲骨字符识别提供了数据基础，为采用深度学习解决甲骨识别困难提供了较为丰富的数据，为计算机视觉技术应用于甲骨文字形研究工作提供领域知识桥梁。
4. 在甲骨文字符识别领域，针对甲骨文识别任务中类别样本分布不平衡问题，论文提出一种基于循环式生成对抗模型（Cycle Generative Adversarial Net，CycleGAN）的甲骨字符数据扩充算法，增广样本量较少的字符类训练样本，解决长尾效应对甲骨字符识别性能的影响。实验证明本文提出的甲骨文字符数据增广算法可以提升整体识别效果，最终在数据库 OBC306 上取得 86.54%的平均识别精度。

## 1.4 论文主要内容与组织架构

本文的结构安排如下：



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

第一章阐述了论文研究背景、研究价值以及研究内容，指出目前甲骨文检测和识别研究中存在的问题。

第二章概述了甲骨文研究现状以及甲骨文字符检测识别技术，包括甲骨文素材收录情况，甲骨文字符检测识别工作，并联系具有相似挑战的自然场景文本检测领域工作进行阐述，并对深度学习相关技术作出概述。

第三章介绍甲骨文字符检测任务方面的工作，包括构建字符级别的甲骨文检测数据集 OBCD、基于 R-FCN 和 FPN 的检测算法框架、基于字模数据的甲骨文字符动态增广算法以及甲骨文字符识别辅助检测算法。

第四章介绍甲骨文字符识别任务方面的工作，包括构建的大数据量的字符级别甲骨文数据集 OBC306、基于循环式生成对抗网络的甲骨文字符数据增广算法，在此基础上构建了基于卷积神经网络 Inception-v4 的甲骨字符识别框架。



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

# 第二章 甲骨文研究现状与深度学习概述

## 2.1 甲骨文素材收录情况

甲骨在第一次被挖掘出土之后，经过多代甲骨学研究者的努力，大部分都已经收录发表[3-10]，这些著录是甲骨文字形研究的数据基础和领域知识的重要承载体。在计算机技术蓬勃发展之前，甲骨文都是通过出版实体著录的方法记录，著录的形式主要有三种：甲骨拓片，甲骨照片以及甲骨摹本。

甲骨拓片是通过使用染料将甲骨表面的纹路复制到纸张上。除了甲骨，其他中国传统器物包括铜器物同样采用这种形式进行内容记录。甲骨拓片这种内容记录形式的优点在于可以方便地将器物表面的纹路按照原有样子记录，缺点是不能表现出甲骨等器物的立体感与真实感，也难以记录原本器物上具有的微小纹路细节。国内第一本以拓片形式记录甲骨文的著录是《铁云藏龟》[21]，目前许多广受认可与广泛使用的甲骨文著录[3-10]中，甲骨文拓片仍是主要的展示形式。

甲骨照片是使用相机拍摄甲骨获得的甲骨图片，该内容记录方法的优点在于可以将甲骨本身的微小细节记录下来，而这些细节很难用上述墨拓方式完整记录。缺点在于拍摄的照片里字符不如拓片清晰，且甲骨本身的脏污干扰都被拍摄在甲骨照片中。到目前为止，以拍摄方式记录甲骨文的著录主要有《殷虚书契菁华》[22]、《双剑誃古器物图录》[23]等。

甲骨摹本是通过专家手动临摹甲骨上的纹路并编纂的甲骨著录形式。甲骨摹本的优点在于比照片更清晰，缺点在于摹本无法保留甲骨片的真实风貌，失去原本甲骨片的环境细节，同时对临摹者的专业水平要求非常高，与前两种甲骨文记录形式相比更消耗时间与昂贵的专家资源。第一本使用临摹形式记录甲骨文的是《殷虚卜辞》[24]，除此之外，1988 年出版的《苏德美日所见甲骨集》同样使用临摹形式记录了苏联等四个国家所藏甲骨片。

总之，在计算机技术未得到蓬勃发展之前，甲骨文主要以上述三种形式记录和呈现，编纂为实体书籍出版和供相关专家查阅利用。随着计算机技术与互联网的迅速发展，为应对数量日益增长的甲骨材料以及著录，国内外许多研究机构发起了甲骨文数字化项目，将甲骨文资料以数字形式储存于计算机或电子介质中。这种方法的优点在于储存更安全，同时让各个地方的研究者更方便地查阅资料，也有利于向社会大众普及甲骨文知识。目前已建立的甲骨文数据库有许多，例如汉达文库[20]是香港中文大学建立的集著录、检索



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

等多种功能的大型甲骨文数据库，其收录了 9 本甲骨文实体著录，所收录的卜辞数量达到 67683 片。其他重要的甲骨文数据库还包括小学堂甲骨文资料库[25]、大英图书馆的甲骨文数据库[26]等。

到目前为止，大部分出土的甲骨都已经以实体著录或者数字项目的形式收录，但是目前现有的甲骨文资料仅有图片，而且多以整片甲骨为单位，仍然缺少具有字符级别标注的数据集。因此构建一个大数据量的字符级甲骨文数据集是一项很有价值的工作，对甲骨文字符检测与识别工作具有较大的推动作用。

## 2.2 甲骨文字符检测技术

甲骨文字符检测是甲骨文字形研究的基础任务，其目的是从甲骨片图像上定位甲骨文字符的位置，为字形识别和语义破译提供位置信息。目前甲骨学研究基本上都是基于由甲骨拓片或照片构成的著录等资料，如果需要对甲骨图像上的文字进行进一步研究，则首先定位图像上的甲骨文字符，而这项工作如果靠甲骨学专家人工完成，将会是一个费时费力、价格昂贵的过程。因此，实现甲骨文字符的自动检测对字形的进一步研究具有重要意义。

目前甲骨文字符检测领域基本上是空白，既没有规范标注的检测数据集，也没有研究工作进行方法和技术的探索。甲骨文作为古汉字，虽然其体系还不够成熟，字符的空间分布离散，笔画不够规整，尺寸多变，但与现代文本一样已经具有相对稳定的拓扑结构。除此之外，由于长期掩埋以及早期私掘的缘故，甲骨表面变得不够平整，甲骨拓片图像包含许多噪声，检测甲骨文字符时需要应对复杂的背景环境。自然场景文本检测任务是计算机视觉中一项重要的基本任务，从随意拍摄的各种自然场景图像中检测文本内容，这个问题需要克服复杂背景环境带来的文本定位困难以及字体外观多样性等挑战，因此自然场景文本检测是与本论文研究的甲骨字符检测较为相近的视觉任务。现代文本检测技术特别是自然场景文本检测技术具对甲骨字符检测极具借鉴意义。近 20 年来，自然场景文本检测技术发展迅速，其方法主要可以分为三类：基于纹理的方法，基于连通域的方法以及混合方法，具体介绍如下。

基于纹理的方法[27-33]将文本看作一种特殊的图案纹理，利用一系列纹理特征区分出文本区域，例如小波系数[30, 31]、局部图像亮度[32, 33]等。Kim 等[27]直接使用图像像素值构造文本纹理特征，并结合支持向量机[34]以及自适应均值漂移算法[35]定位文本区域。Zhong 等[28]使用在离散余弦变换域中编码的局部图像方差直接对压缩后的视频图像作



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

标题文本的定位追踪。Zhong 等[29]在颜色复杂的图像上使用局部图像方差获取初步候选区域，再结合边缘检测，使用配对边缘定位文本区域。基于纹理的方法是自然场景文本检测任务的常用方法之一，但这种方法的缺点在于计算量较大，同时对字符尺度和字符方向敏感，而甲骨文的字形大小相比现代文字更加随意，字符尺度的变化范围大，因此基于纹理的方法对甲骨文字符检测的作用有限。

基于连通域的方法[36,37]首先通过图像灰度值等方式获取初步候选区域，然后再进行过滤得到最终的文本区域。这种方法的优点在于其计算量比基于纹理的方法少，同时对文字尺度、字符旋转角度以及字符结构的变动足够鲁棒，而这一点实际上与甲骨文检测任务契合。基于连通域的方法中最具代表性的有笔画宽度变换 SWT（Stroke Width Transform）[36]和最大稳定极值区域 MSER（Maximally Stable Extremal Regions）[37]。其中，SWT [36]认为同一文本具有相对稳定的笔画宽度，借助这一笔画特征获取初步的候选框，再利用一系列规则进行筛选以及合并，从而获得最终的文本区域，但实际上这一点对甲骨文字符并不适用，甲骨文是通过尖利器物刻在甲骨上的，力道难以控制，因此笔画宽度不如手写的现代文本一样稳定，甚至有可能出现笔画断裂，因此 SWT 方法并不适用于甲骨文字符检测。另一方面，MSER[37]认为图片在受不同阈值分割时，其文本区域相对保持稳定，基于这一点筛选出最大稳定极值区域，再利用其他规则进行筛选与合并，从而获得完整的文本区域。与 SWT 相比，MSER 不需要依赖字符的笔画宽度稳定性，然而由于甲骨拓片经过多年掩埋，其表面会出现许多划痕，而这些划痕与甲骨文字符的刻痕比较容易混淆，因此直接使用 MSER 方法进行甲骨文字符检测容易引入较多噪声。

混合方法[38,39]结合了基于纹理以及基于连通域两种方法的优点，每种方法各自负责流程的其中一个步骤，例如使用基于连通域的方法初步提取候选区域，再利用纹理特征进行筛选，或者相反。Liu 等[38]首先利用自己设计的边缘检测算法获取所有可能为文本的像素块，再利用连通域分析进行像素分组，最后利用小波域纹理对候选区域进行进一步筛选，从而从复杂背景中定位文本。Pan 等[39]则先使用小波变换结合连通域分析获取初步筛选框，再借助使用纹理特征训练好的支持向量机进行进一步筛选分类，从而获得最终的文本区域。

随着深度学习在计算机视觉领域的成功，许多自然场景文本检测工作也开始引入深度神经网络。引入深度学习的优点在于神经网络模型自身能学习更抽象的特征，不需要依赖手动提取的文本特征，使整个系统更趋于一体化，检测流程不再繁琐。基于深度神



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

经网络的自然场景文本检测工作大多受通用目标检测工作的启发，在其基础上根据文本自身的特性进行网络结构等机制上的设计。Liao 等[40]受通用目标检测领域的 SSD 网络[41]启发，实现了仅用一个深度神经网络检测多种尺寸的自然场景文本词语，摆脱了过去需要多个步骤、包含多种算法的繁杂系统，在保证具有较高准确度的同时也具有较快速度。Liu 等[42]针对自然场景中形状、倾斜角度多变的文本行，在文献[41]的基础上设计新的目标框格式与坐标回归机制，实现自然场景中多方向文本的检测。Zhou 等[43]借鉴 PVANet[44]的思路通过全卷积网络提取多尺度图像表征并进行融合，最后用于文本位置的预测与预测框的筛选。这些深度学习用于自然场景文本的成功，启发论文探究深度学习进行甲骨文字符检测的技术和方法。

## 2.3 甲骨文字符识别技术

甲骨文字形识别是甲骨学研究中的一个重要分支，目的是进行甲骨字符的类分析，在此基础上，借鉴已有的甲骨字符破译成果，破译更多的甲骨字符变形变种。甲骨字符字形识别将加速推进甲骨破译工作，将其从低效昂贵且困难的境地中解放出来。目前甲骨文字符还未和现代汉字一样被编入国家标准或国际标准，所以只能以图像形式展示。如果要实现甲骨文字形的快速检索，甲骨文的字形识别是必不可少的工作。除此之外，在借助大数据进行甲骨文字形研究的过程中，需要大数据量的、具有标注的甲骨文数据集，而通过甲骨学专家手动标注是一项人力消耗非常大的工作，实现甲骨文的自动识别可以有效减少人力损耗，提升工作效率，因此甲骨文自动识别也是一项当前较迫切的工作。

甲骨文字形的研究基础还相对薄弱，在 20 世纪 90 年代以前，包括甲骨文，中国古汉字研究都还是空白[17]。随着计算机技术的快速发展，借助计算机技术进行甲骨文字形研究开始成为甲骨学热门的研究方向，但相较于现代汉字研究领域，甲骨文字形研究工作还比较少。

到目前为止，大部分甲骨文字形识别工作的基本思路是围绕图论以及拓扑学的理论进行人工特征提取。提取特征之后，接下来比较普遍的做法是对特征进行编码，然后直接进行匹配和比较，从而实现识别。李锋等[17,18]认为甲骨文具有相对稳定的拓扑结构，具备明确的点和线，因此提出多层级的图论特征编码，然后通过编码序列的比较进行识别分类；栗青生等[15]将甲骨文看作无向图，构造多层级识别特征，在编码时使用拟邻接矩阵，以减少重码问题；顾绍通[16]认为甲骨文字形的拓扑结构一定程度上较稳定，通过



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

描述拓扑顶点之间的关系构造特征并进行编码，最后通过计算拓扑之间的距离进行识别；之后顾绍通[19]又针对原有特征编码效率低，重码较多的问题，提出从分形几何的角度总结甲骨文字符的特征，以识别甲骨文字形。上述工作使用图论与拓扑学特征结合简单编码进行甲骨文匹配识别，在数据量较少的情况下可以获得较好的结果，但数据量大的情况下，简单的图论特征与手动编码容易出现欠拟合。

由于机器学习算法在计算机视觉领域获得很大成功，因此除了手动编码进行比较和匹配，部分甲骨文字形识别工作也开始引入机器学习算法，以获取更抽象的特征表达。吕肖庆[45]提出一种基于曲率直方图的傅立叶描述子以提取字形特征，然后将特征输入到支持向量机模型[34]进行字形分类；Guo 等[46]将甲骨文字形识别问题看作草图识别问题，构建多层级特征表征，然后引入卷积神经网络算法作为分类器；高峰[47]针对甲骨上模糊字形的识别，提出了一种基于上下文语境的统计学分析和霍普菲尔得神经网络结合的识别方法，提升模糊字形的识别效果；刘永革[48]以分块直方图的方式提取特征，并引入支持向量机作为模型进行甲骨文字形分类，Meng 等[49]通过霍夫变换以及聚类提取特征，再通过计算距离与模板图像匹配，从而实现识别。上述工作引入机器学习算法后使模型获得了更强的特征表达能力，但仍然依赖专家手动的特征工程，同时需要设计一套步骤复杂的识别系统。

## 2.4 深度学习相关技术

深度学习属于机器学习的分支之一，通过构建具有多层级的人工神经网络以实现人工智能，这项技术目前已在学术界与工业界掀起了热潮，并在多种领域中都取得了成功。卷积神经网络是深度学习技术中的形式之一，具有强大的特征表达能力，目前已被广泛应用于计算机视觉、语音处理等多个领域。卷积神经网络的理论基于 1962 年 Hubel 和 Wiesel 提出的感受野概念[70]，此后 1980 年 FukuShima 提出的神经认知机[71]可以看作是最早的卷积神经网络，而近年来得益于多种网络结构[]的提出，卷积神经网络在多个领域取得较大成功。图 2-1 展示了典型的卷积神经网络 LeNet-5 网络[72]，卷积神经网络主要有三种网络层，分别是卷积层、池化层和全连接层。

卷积层的示意图如图 2-2 所示，其作用在于提取输入数据的特征，方式是在输入数据上进行窗口滑动，在每个窗口中使用卷积核进行计算。卷积层的特点主要有两点：局部连接与权值共享。传统的全连接层输出向量的每一个值都会与输入向量的所有值关联，导致全连接层需要训练的参数非常多，令神经网络非常难以训练，而卷积层在每一次滑



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

动窗口时，仅与当前窗口覆盖的局部像素值连接，因此输出值仅与局部窗口输入相关，可以大大减少训练参数量。除此之外，同一个卷积核在输入数据上进行窗口滑动时，对所有窗口均使用相同的卷积核参数，这样可以使卷积核的训练参数量进一步减少，也和输入数据的尺寸无关。凭借这两个特点，卷积层得以通过较少的参数实现较好的特征提取效果，同时可以实现特征的位移、尺度和形变不变性。这个特性对目标特征的提取非常重要，以甲骨文识别为例，甲骨文字在图像中的位置、大小以及形状变化都不会改变甲骨文字所属的字类，因此我们希望网络层输出的特征不会因为这些因素而发生改变。而通过堆叠多层卷积层，卷积神经网络可以获得更强大的特征提取能力，实现更加抽象的特征提取。

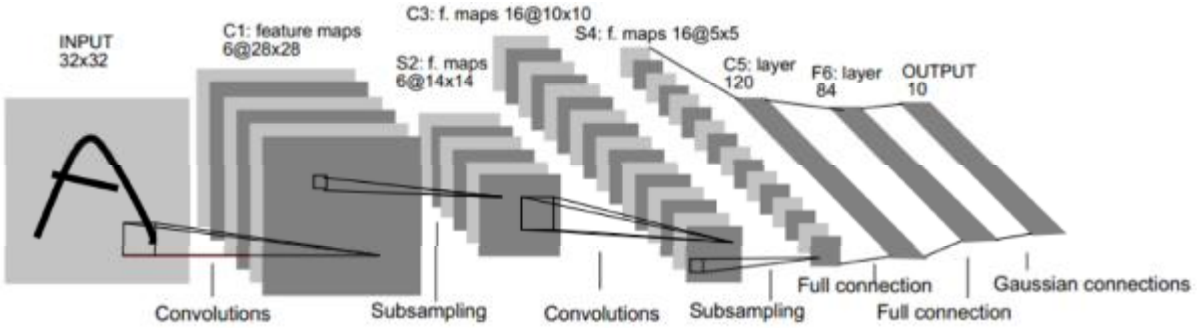



图 2-1 LeNet-5 网络框架

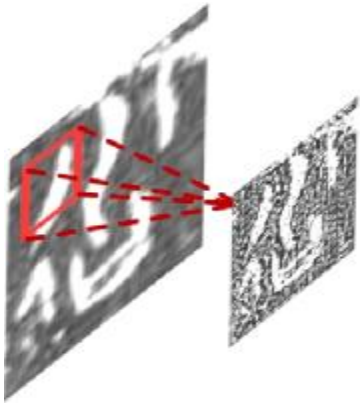


图 2-2 卷积层示意图

池化层的位置一般在卷积层后面，其任务在于对输入数据进行下采样，降低数据的维度。与卷积层类似，池化层同样是通过在输入数据上滑动窗口，并在每一个窗口中获取输出值，输出值可以是当前窗口区域的最值、均值以及求和等，但池化层自身不包含训练参数。池化层的作用主要有以下几点：首先它是通过对数据进行下采样，降低模型计算的复杂度，防止模型的过拟合；其次，它与卷积层相同，也引入了特征的不变性，图像目标一定程度内的变化不会改变池化层的输出。

全连接层在卷积神经网络的位置一般位于网络末端。传统人工神经网络通过堆叠全



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

连接层实现高层特征提取，而在卷积神经网络中，参数更少的卷积层替代全连接层进行主要的特征提取工作，全连接层一般负责在最后阶段将卷积层提取的局部抽象特征进行融合，将特征映射到标签空间。

## 2.5 本章小结

本章首先介绍了当前甲骨文资料著录情况及其收录形式,接着介绍了甲骨文字形检测领域工作，并结合具有类似挑战的自然场景文本检测工作进行阐述，然后介绍了甲骨文字形识别工作，最后介绍了深度学习的相关技术。



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

# 第三章 甲骨文字符检测

本章将详述论文在甲骨文字符检测方面的研究工作。本文首先构建了甲骨学研究领域首个甲骨文单字符检测数据集 OBCD（Oracle Bone Character Detection），分析了数据集的统计属性以及样本特征，总结了甲骨文字符检测任务的难点和挑战。接着，本文引入基于区域的全卷积网络 R-FCN （Region-based Fully Convolutional Network）深度检测模型，并采用特征金字塔网络 FPN（Feature Pyramid Network）进行多尺度网络结构设计，搭建甲骨图像字符检测算法框架。进一步，针对复合字符难检测问题，提出一种基于字模数据的甲骨文字符动态增广算法，强化检测模型对复合难检字的关注和优化偏好，提高复合字符检测准确率。最后，我们提出甲骨文字符识别辅助检测算法，在存在许多容易误检的甲骨划痕的情况下帮助检测模型减少错误预测结果。

## 3.1 构建甲骨文单字符检测数据集 OBCD

数据集 OBCD 的构建过程包括如下几个步骤：首先选取广受学术界认可的甲骨文集成性资料汇编著录《甲骨文合集》[3]作为数据集的素材来源，接着将著录的页面扫描为数字图像并以拓片为单位切割图像，然后在“甲骨文字典”软件[50]的辅助下，找到每一张拓片包含的甲骨文字符，使用 Labelimg 标注工具[51]进行标注。下面将详述构建过程。

### 3.1.1 构建过程

#### 3.1.1.1 数据集图像来源

在构建数据集 OBCD 的过程中，论文选取《甲骨文合集》[3]作为数据集图像来源。《甲骨文合集》[3]在 1978 年到 1982 年由中华书局出版，这是一本关于甲骨学的集成性资料汇编，它覆盖了自甲骨文首次出土以来，除了小屯南地以外的所有甲骨拓片，总共包含 41956 片甲骨拓片。本文对《甲骨文合集》的所有页面进行扫描，获取每份页面的数字图像，以便后续使用计算机工具进行数据标注。最终扫描获取的数字图像一共有 5242 页，包含了全部 41956 片甲骨拓片，由于甲骨的尺寸大小不一，有些是不完整的甲骨碎片，因此有些页面包含了多张较小的甲骨拓片。为了将数据集图像统一以甲骨拓片为单位，我们使用裁剪工具将包含多张甲骨拓片的扫描图片裁剪为多张以甲骨为单位的图片，裁剪过程尽量保持边界贴近甲骨拓片。最终经过扫描裁剪，得到的拓片图像数量为 41956 张。



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

#### 3.1.1.2 检测标注

本节的任务是对上述扫描裁剪后的甲骨拓片数字图像进行检测标注，对拓片中的甲骨字符用矩形框框定，为检测模型训练和验证提供监督和评价准则。标注过程使用两个重要的工具，分别是“甲骨文字典”软件[50]以及开源标注工具 Labelimg[51]。

“甲骨文字典”[50]是一个电子字典工具软件，由安阳师范大学甲骨文顾问研究团队制作，为甲骨文研究者提供便利的甲骨文资料检索工具。该工具基于香港中文大学中国文化研究所的甲骨文电子资料库[20]，能展示每张拓片上的甲骨文字符及其现代汉语翻译（如果该字符已经破译）。除此之外，“甲骨文字典”也提供了检索功能，每张甲骨拓片在进行收录时都会进行编号，编号表明其著录来源，页序等信息，根据每张甲骨拓片的编号就可以在“甲骨文字典”中获取每张拓片包括的甲骨文字符信息。借助这个工具，我们可以在对每张拓片图像作标注的时候获知当前拓片上包括哪些甲骨文字符，从而保证标注数据的准确性。

Labelimg[51]是一个开源的图像数据标注工具，其提供了人性化图形操作界面。借助这个工具，我们可以方便地使用矩形框对每一张图片中的甲骨文字符进行框选，这个工具将自动记录框选位置坐标，并且会将标注信息转化为目标检测领域应用广泛的 PASCAL VOC 数据集[52]标注格式进行保存，减少不同应用指向的检测工作之间的迁移成本，提高工作效率。


图 3-1 甲骨文检测数据集构建过程

数据集 OBCD 的标注流程如图 3-1 所示。对于每一张甲骨拓片，为了进行检索，需要获取其拓片编号，而拓片编号由两部分组成，即著录的字母缩写与拓片在著录的五位编号，这里以《甲骨文合集》中第 4 张甲骨拓片为例，《甲骨文合集》的字母缩写为 h，因此第 4 张甲骨拓片编号为 h00004。



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

接下来我们根据拓片编号使用 “甲骨文字典”工具进行检索，获得编号 h00004 的拓片包含的甲骨字原文。在“甲骨文字典”展示的原文内容中，开头与结尾的省略号表示在这张甲骨拓片以外还存在其他甲骨文字符；括号中的甲骨文字符是当前甲骨拓片包含的残缺严重的字符，这部分字符在拓片上几乎无法看清，因此在标注过程中忽略；而未在括号中的字符即为当前拓片上可以看到的甲骨文字符，也是标注过程中需要标注的字符。

获取拓片上的字符之后，这里使用 Labelimg[51]标注工具对图片上对应的甲骨文字符进行框选，框选过程中尽量贴紧甲骨文字符，工具将自动记录框选坐标。因此在数据集 OBCD 中，每个字符的标注信息即为其框选位置的坐标$(x_1, y_1, x_2, y_2)$，其中（$x_1, y_1$）为矩形框左上角顶点的坐标，（$x_2, y_2$）为矩形框右下角的坐标。

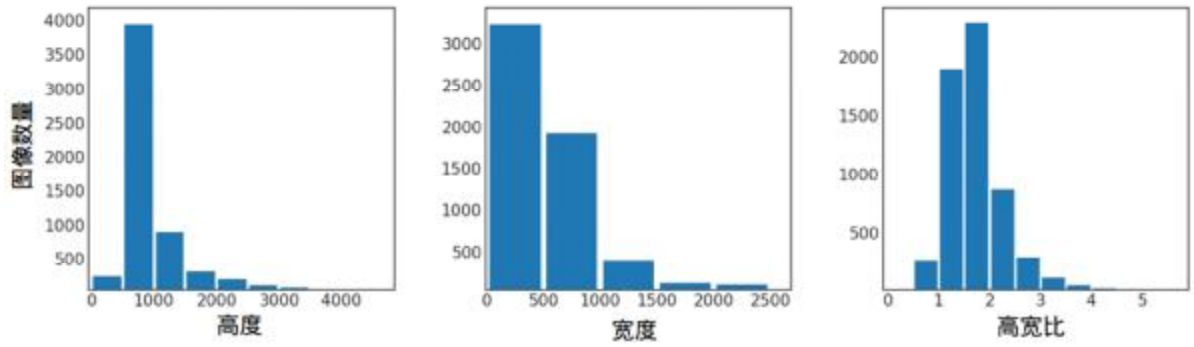



图 3-2 甲骨文检测数据集拓片图像尺寸分布

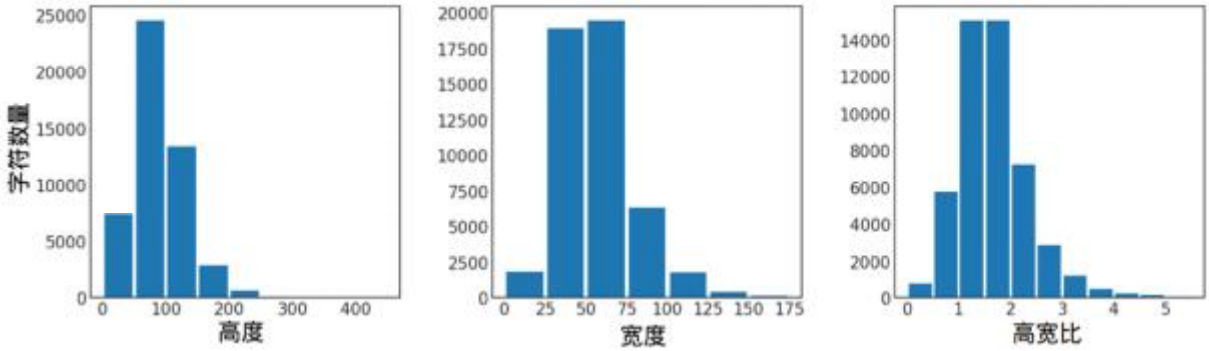



图 3-3 甲骨文检测数据集字符实例尺寸分布

## 3.1.2 数据属性与挑战

### 3.1.2.1 样本统计属性

鉴于检测标注需要大量的人力和费用投入，到目前为止，我们只对 41956 甲骨拓片图像中的 5838 张进行了字符检测标注，即 OBCD 数据集总共包括 5838 张拓片图像，在这些拓片图像中用矩形框标注了 48821 个甲骨文字符，平均每张拓片图像包括 7 个左右的字符示例。OBCD 中拓片图像尺寸分布如图 3-2 所示，图像高度最大部分集中在 0 到



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

1000 像素之间，其次集中在 1000 到 2000 像素之间，而最大高度为 5983 像素。宽度集中在 0 到 500 像素之间，其次集中在 500 到 1000 像素之间，其中最大宽度为 3479 像素。总体说来，图像尺寸跨度较大。拓片图像长宽比集中在 1 到 2 之间，表明大部分图像是形状瘦高的图像。标注框尺寸分布如图 3-3 所示，字符高度集中在 50~150 像素之间，宽度集中在 25~75 像素之间，高宽比集中在 1 到 2 之间，可见甲骨文字符多为高瘦型目标。从标注字符的角度看，这些甲骨字符对象基本是小目标，且呈现显著尺度多样性，因此，甲骨字符检测归属于小视觉目标检测且需考虑多尺度字符检测，这是检测方法设计过程中应予以重点考虑的因素。

#### 3.1.2.2 甲骨文字符检测难点

图 3-4 给出了 OBCD 数据集的一些甲骨拓片图像及字符标注示例，其中绿色矩形框即为标注框。由图 3-4 可知，由于多种影响因素（例如，挖掘过程，掩埋地理条件，内容记录的人为因素等），一些拓片图像上的甲骨文字符较难看清，因此 OBCD 数据集是一个具有挑战性的检测数据集。接下来本文对数据集的几个难点进行详述。

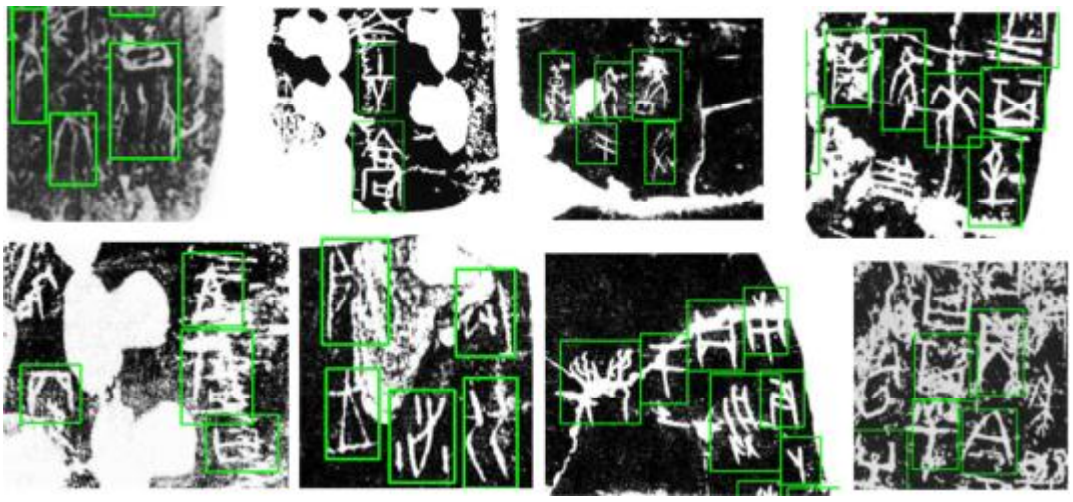


图 3-4 OBCD 数据集拓片图像及标注示例

首先是小目标。当前甲骨文检测数据集包括大量几十个像素长宽的小目标，会带来检测困难。数据集中大尺寸的图片较多，这些图片包含了一整片较大的龟甲或兽骨，因此甲骨字符区域相对较小。图 3-5 为不同尺寸的样本图像包含的目标数量以及小目标数量，横轴为图像短边，纵轴为字符目标数量，此处小目标定义为短边小于 36 像素的目标。由图 3-5 可知，字符目标主要集中在短边为 500 像素左右的图像上，同时图像尺寸越大，小目标占比越大。

其次是复合字。甲骨文属于象形字，与现代汉字类似，具有类似偏旁部首的部件，有些字形由多个部件组成，以表达较复杂的含义，这里称之为复合字。但甲骨文也与现



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

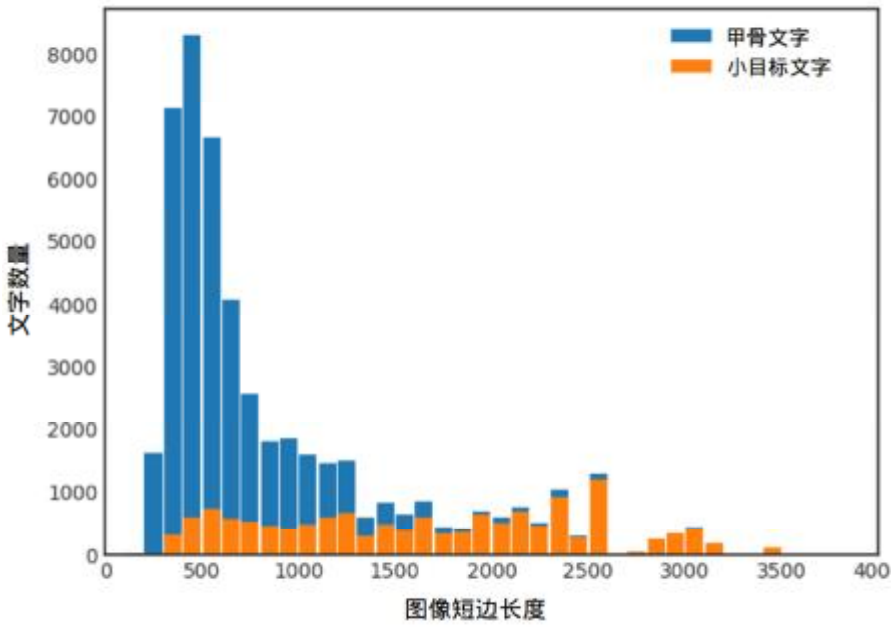
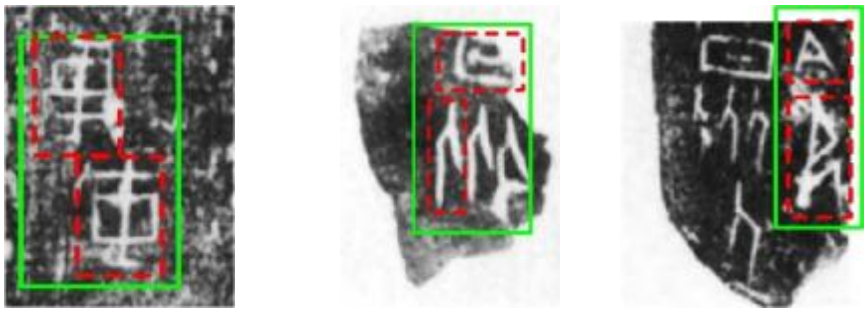



图 3-5 甲骨文检测数据集中小目标的分布

代汉字不同，部件的字形并没有因成为“偏旁”而变形，其字形仍与作为单字的字形一样，因此含有多个部件的复合字容易被误检测为多个甲骨文单字。而且甲骨文的文字体系仍未成熟，字形较为多变，部件的相对位置比较自由，没有排列在同一水平面或垂直面上，同时相对距离有时离得较远，这也提升了复合字的检测难度。数据集中的复合字示例如图 3-6 所示，每张子图中绿色矩形框为正确的甲骨文实例，而红色虚线框为甲骨文实例的部件，每个虚线框中的部件单独存在时也是一个具有意义的甲骨文字符。

除此之外还有字形破损的问题。甲骨表面噪声严重，部分字形模糊残缺。由于掩埋环境与私掘等原因，有许多出土的甲骨片已经破碎，即使经过甲骨学研究者的努力缀合，但裂痕仍然存在，而部分裂痕经过甲骨文字符，使字形变得残缺，具体如图 3-7 的 a)、b)。除此之外，许多甲骨片表面出现磨损，字形变得较难辨认，例子如图 3-7 的 c)所示。

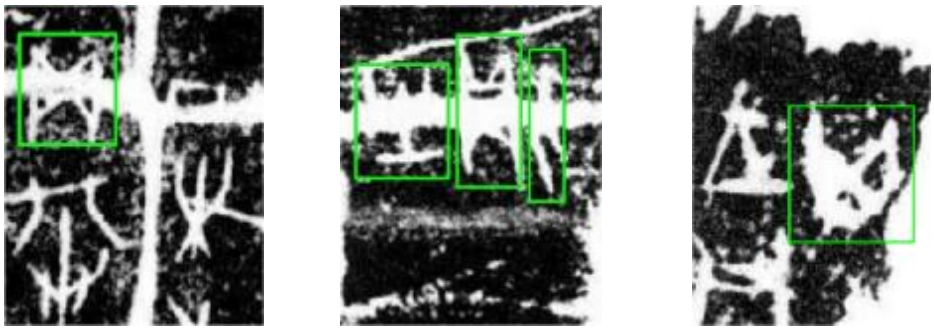


a） b） c）

图 3-6 甲骨文检测数据集中复合字示例



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net



a) b） c）

图 3-7 甲骨文检测数据集中噪声严重的样本示例

另外，由于年代久远等原因，甲骨片表面出现许多类似甲骨文字的痕迹，这种划痕容易与甲骨文字混淆，导致错误检测，具体样例如图 3-8 所示。

a) b） c）

图 3-8 甲骨文检测数据集中类似甲骨文字的痕迹示例

## 3.2 甲骨文字符动态增广算法

复合字是当前数据集的检测难点之一，针对这个难点，本文以甲骨文字模数据集为数据来源，提出一种甲骨文字符动态增广算法，在提升整体数据规模的基础上针对性增加每张训练图像的复合字实例数量。下面详细介绍该算法。

### 3.2.1 增广数据来源

合成甲骨文字符图像的过程是以甲骨文字模数据集作为数据来源，这是一个字符级别的字模数据集，由安阳师范学院的甲骨学专家参考 13 本甲骨文著录上的拓片样本，根据甲骨上的字符手描获得。甲骨文字模数据集一共有 41980 张字模图像，共覆盖了 5490 类字形。该数据集中所有字模图像的尺寸均为 500×500，都是白底黑字的二值化图像。图 3-9 给出了字模数据集的样例。对比图 3-6、3-7，字模数据与真实拓片上的字符图像差别较大，字模图像中的字符非常完整清晰，没有包含噪声，为生成新的甲骨文字符图像提供质量较好的数据基础。



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

图 3-9 甲骨文字模数据集图像样例

字模数据的类样本数量分布如图 3-10 所示。由图 3-10 可知，有 2905 个类仅有 1 张字模数据，即数据集中有 52.91%的类别仅有 1 张样本，且这部分的样本量仅占总样本量的 6.91%，由此可知，数据集中有大量低频的字符类别；有 1731 个类的数据量在 2 到 9 之间，这部分类别数占总类别数的 31.53%，数据量一共有 6592 张，约占总数据量的 15.7%；除此之外，一共有 784 个字类的样本量在 10 到 99 张之间，这部分类别数约占总类别数的 14.28%，这部分样本量一共有 21676 张，约占总样本量的 51.63%；最后，仅有 70 类字拥有超过 100 张样本，类别数约占总类别数的 1.28%，但这部分样本量约占总数据量的 25.74%。由此可知，当前数据集中在拥有大量低频类别的同时，也存在一部分相对高频的类别，总的来说数据集类别之间的样本量差距较大，大部分类别的样本偏少。

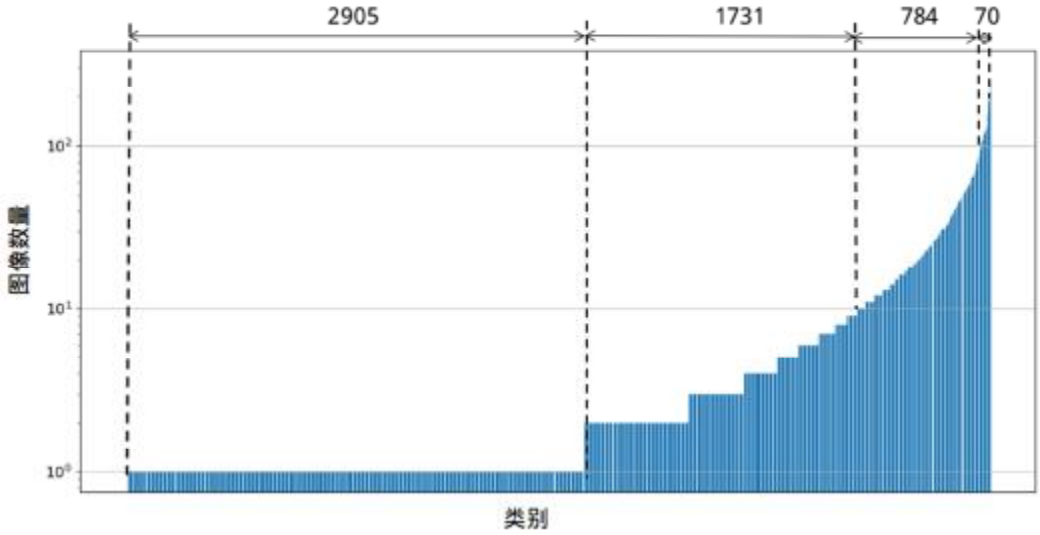



图 3-10 甲骨文字模数据集样本数量分布



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

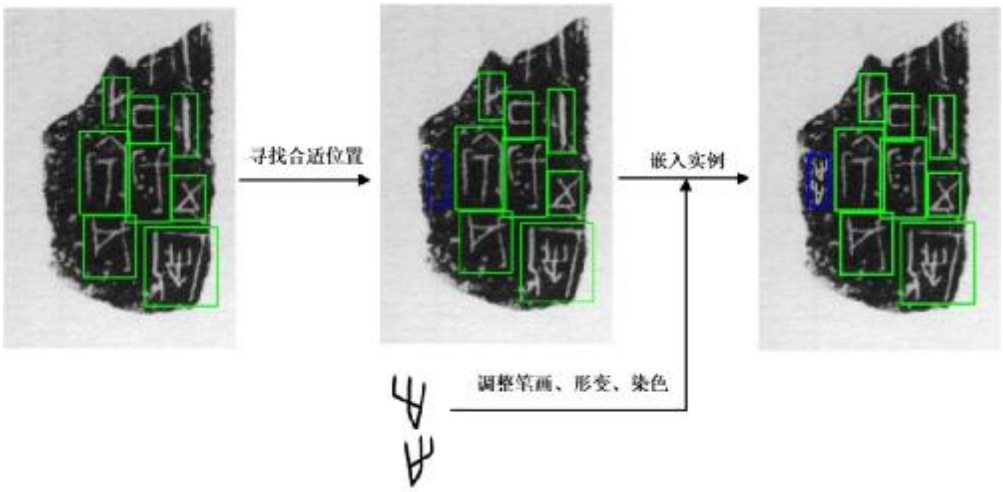



图 3-11 基于字模的甲骨文动态增广算法流程

### 3.2.2 数据增广算法

本文提出的动态增广算法流程如图 3-11 所示，基本思路是以字模作为合成数据来源，在训练拓片图像上嵌入新的字符示例，以增加训练示例数量，从而提升检测效果。在嵌入过程中，本文考虑了合成文字与拓片图像的融合程度，尽量令合成文字在拓片图像中不显得突兀。针对检测数据中复合字相对较难检测的问题，我们选取了字模数据集中的复合字字模作为主要数据来源。具体嵌入步骤如下：

1. 随机选择拓片图像中的一个原标注框，获取框内字符的灰度值，用于后续合成新字符。示意图如图 3-12 所示，考虑到图像中字符与背景的对比度相对较大，而且标注框中字符的灰度值相对稳定，同时由于标注过程中标注框尽量贴近文字，使甲骨文字符占据标注框内的大部分区域，所以标注框区域的灰度直方图中，靠右的波峰对应的灰度值即为该字符的灰度值。在这一步骤中我们通过绘制标注框的灰度直方图，然后利用信号处理中的历史峰值查找算法[53]获取文字部分的灰度值。

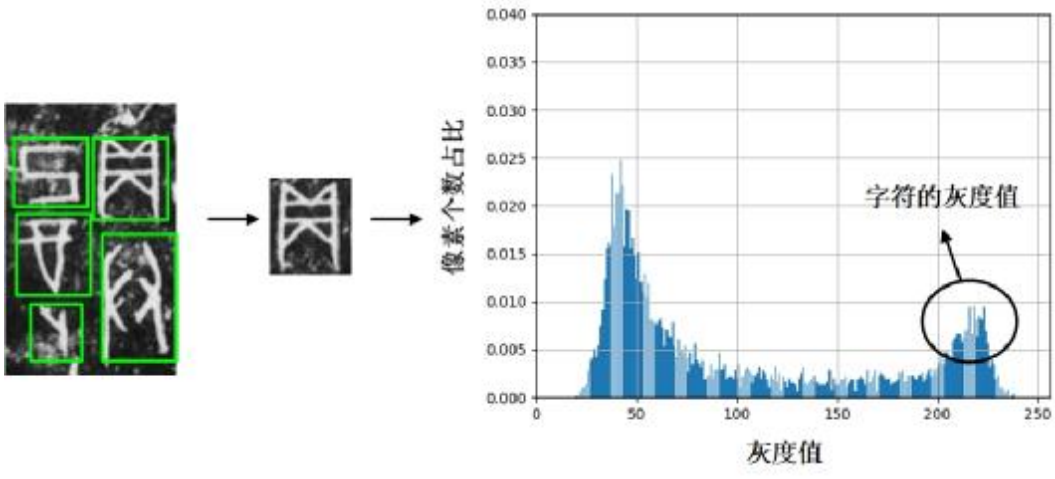



图 3-12 对原标注框绘制灰度直方图获取字符灰度值的示意图



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

2. 合成新的字符实例。本文随机选取一个甲骨文字模，使用上一步骤获得的灰度值对字模的文字部分进行染色，这样可以使新的字符实例在风格上与真实文字实例相似，便于后续融入训练图像。为了增加文字实例的多样性，在每一次合成的时候，我们会随机调整文字示例的笔画粗细，并对新文字实例进行仿射变换，包括放缩、旋转、倾斜和翻转，同时加入随机噪声。
3. 在训练图像中寻找可以嵌入新字符实例的位置。可以嵌入新字符实例的位置需要符合几个条件。首先新字符实例的位置应当在拓片图像的甲骨片部位中。由图 3-11 可知，拓片图像的甲骨部位周围存在白色空白区域，不可能出现甲骨文字符，因此选择位置时需要避开这些区域，选择在甲骨片区域中的位置。但同时这里允许新文字实例的位置被空白区域部分覆盖，但不会被遮盖过多，防止字形过分残缺。由 3.1.2 节可知，由于甲骨片出现破碎，一些甲骨文字符分布在甲骨片边缘，被空白区域遮盖了部分字形，本文希望在合成数据时模拟这种情况。其次，新字符实例的位置不能与原有的字符重叠。字符实例重叠会相互影响，导致无法看清字形，影响训练效果。综合上述条件，本文的具体方法是利用整张图像的灰度均值作为阈值过滤甲骨片周围的空白区域，接着随机生成若干个候选框，然后根据候选框内部的灰度均值与方差，过滤被空白区域覆盖过多的候选框，最终得到适合嵌入字符实例的候选框。这里为了让字符实例的尺寸与真实文字实例相近，所有候选框的尺寸都由当前图像的真实标注框的尺寸均值结合随机数决定。
4. 将合成的文字实例嵌入符合条件的候选框。在嵌入过程中我们保持文字实例的长宽比，防止文字的变形扭曲。除此之外，考虑到甲骨片许多位置存在较大的白色块状噪声，为了保留环境中的白色块状噪声，使这些文字实例足够真实，我们嵌入文字实例时仅对候选框中文字实例的笔画部位进行灰度值的赋值，其他部位不做改动。

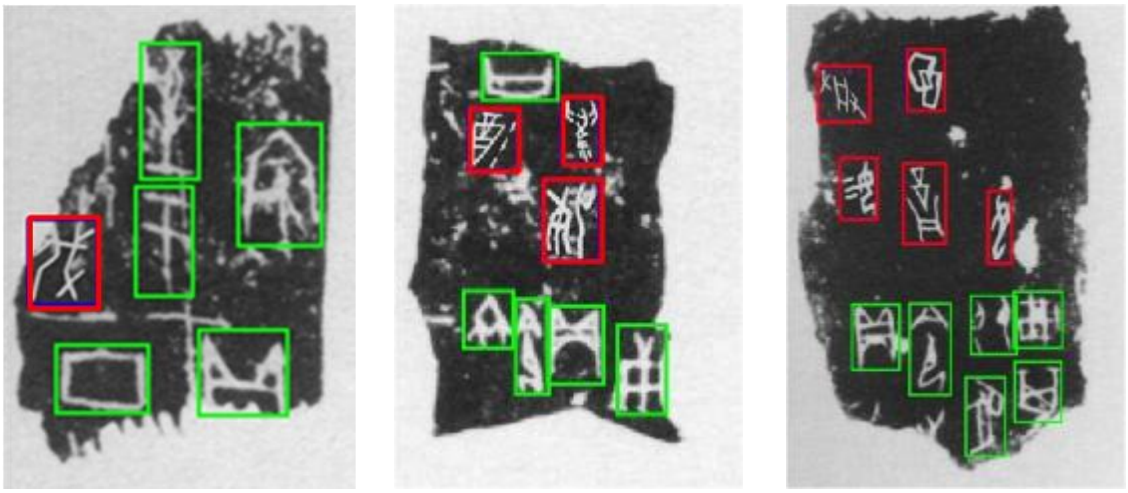
为提升训练样本的多样性，我们在模型训练过程中动态进行上述合成嵌入步骤。在训练的每次迭代中，算法合成新的文字实例，在每张训练图像上动态选择候选区域并嵌入实例，每张图嵌入 $N_{syn}$ 个新文字实例，并且这些合成的文字实例仅在当前迭代里生效，在当前迭代的训练结束后就会消除，防止图像随着训练的进行逐渐被合成实例占满，为后续的合成实例保留位置。除此之外，由于部分图像上的甲骨拓片面积较小，没有足够空间嵌入合成实例，因此算法在寻找合适的嵌入位置时设置最大尝试次数 $N_a$，防止训练过程效率过低或陷入死循环。

本章实现的动态增广算法生成的图像样本如图 3-13 所示，图中绿色框为原标注框，



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

红色框为嵌入的新标注框。从肉眼上看，合成的文字实例与真实的文字实例在风格上非常相近，与整体图像的融合效果较好。



a) b) c)

图 3-13 嵌入新文字实例后的图像样本

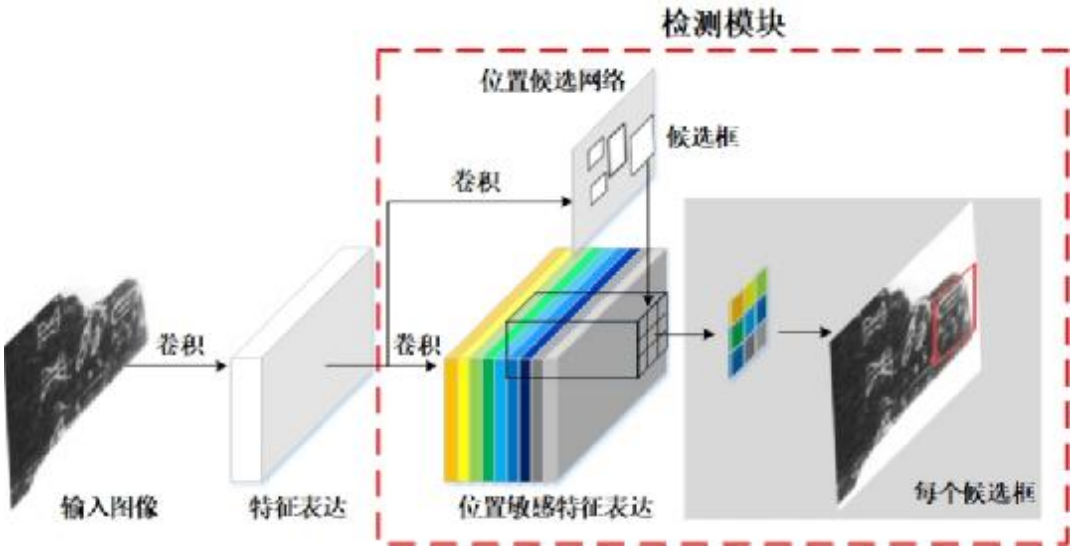



图 3-14 基于 R-FCN 的甲骨文字符检测网络结构

## 3.3 一种识别辅助的检测算法

### 3.3.1 多尺度 R-FCN

本文的甲骨文单字符检测算法框架以 R-FCN 深度神经网络[54]为基础。R-FCN 网络[54]由 Dai 等人在 2016 年提出，该网络针对以往工作对目标位置不够敏感的问题，增加了位置敏感卷积网络模块 Position-Sensitive Convolution Network，从而实现更加精确的通用目标检测效果。在当前甲骨文检测任务中，我们将甲骨字符实例看作一种特殊的目标，并且考虑到 R-FCN 网络在通用目标领域取得的良好效果，本文将 R-FCN 网络引入到甲骨字符检测算法框架搭建过程中。在甲骨文字符检测任务中，基于 R-FCN 的检测网络



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

结构如图 3-14，网络以整张拓片图像作为输入，经过多层卷积之后获得整张拓片图像的特征表达，作为后续检测模块输入。检测模块共有两个分支，区域候选网络 RPN（Region Proposal Network）分支和位置敏感卷积网络 PSCN（Position Sensitive Convolution Network）分支，RPN 分支根据输入的全图特征表达，输出初步的甲骨文字候选区域。而 PSCN 分支则根据输入的全图特征表达，生成全图的目标位置敏感特征表达图，并以此修正 RPN 分支预测的甲骨文字初步候选区域，使候选区域更精准地定位甲骨文字。

R-FCN 模型在位置精修方面具有较强能力，但如果仅仅使用 R-FCN 模型，由于输入检测模块的全图特征表达图经过多层卷积层处理后，分辨率有所下降，较小目标的特征分辨率会变得过低，不利于小目标检测。由 3.1.2.节可知，数据集 OBCD 中存在许多小的甲骨文字符，特征表达分辨率有限的情况下，这些字符将很容易被漏检。为了涵盖多种小尺度的甲骨文字符实例，本章在 R-FCN 模型的基础上引入特征金字塔网络 FPN。特征金字塔网络[55]是由 Lin 等作者在 2017 年提出的通用目标检测工作，其基本思想是通过设计多层级的 U 型网络结构，更高效地利用深度卷积网络产生的不同尺度特征表达，从而提升多尺度目标的检测效果。在当前任务中，结合 FPN 的甲骨文字符检测网络结构图如图 3-15，输入图像经过多层卷积层处理后，获得多级尺寸不同的特征表达，再经过多层反卷积层逐渐恢复特征表达的尺度，每一层反卷积层的特征表达与对应卷积层的特征表达相互融合，并输入到负责对应尺度目标的检测模块分支，最终多个分支的结果组合并通过非最大值抑制算法（Non-Maximum Suppression）[56]过滤冗余候选框，从而得到最后的输出结果。FPN 网络中的特征表达图分辨率从上往下逐渐增加，较上层的特征表达用于检测尺寸较大的甲骨文字，较下层的特征表达用于检测尺寸较小的甲骨文字，本文将这种方式引入到甲骨文字符检测任务，提升了特征表达的利用率，同时避免因特征表达图的分辨率不足导致的目标检测失败。

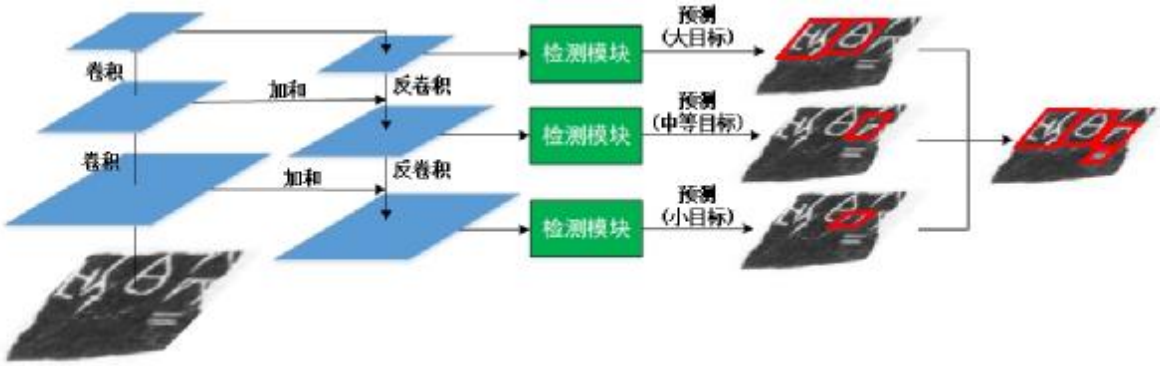



图 3-15 结合 FPN 的甲骨文字符检测网络结构



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

### 3.3.2 识别辅助检测算法

由 3.1.1 节知，当前甲骨文检测数据集的数据标注信息仅有文本框的位置，没有标明每一个文本框是什么文字，因此当前甲骨文检测任务中训练的字符检测模型在分类分支上只能进行二分类任务，缺乏在字形方面更细致的分类能力，对甲骨拓片上类似甲骨文字的划痕容易错误判断。为了减少检测任务中划痕的误检情况，我们提出一种识别辅助检测算法，使用具有 306 类字符的甲骨文识别数据集 OBC306 训练出字符识别模型，利用这个具备多类字符先验知识的识别模型对当前检测任务的预测结果进行辅助修正。数据集 OBC306 及识别模型的详细介绍详见后续 4.1 节。辅助检测的算法流程主要包含四个步骤，下面一一介绍。

首先，本文使用 OBC306 数据集训练出一个字符识别模型，训练直至模型收敛。于 OBC306 数据集包括 306 类字符，因此字符识别模型具有字符多分类能力。由于在后续第四章使用 OBC306 数据集的实验中，Inception-v4 网络相比其他深度神经网络的表现更好，因此这里的模型结构使用 Inception-v4[57]。

其次，考虑到 OBCD 数据集与 OBC306 数据集的数据分布不同，因此本文在经 OBC306 训练的识别模型基础上，再使用 OBCD 的文字实例进行进一步精调。进一步精调训练的任务是二分类，本文将 OBCD 训练集的文字实例作为正样本，并从 OBCD 数据集上生成负样本，具体方法是从每张检测图像随机切取图像区域，并取图像区域与检测图像上每个字符实例比较，如果该图像区域与每个字符实例的重叠面积都小于该字符实例面积的 0.1 倍，且重叠面积之和小于该图像区域的 0.1 倍，则该图像区域即为训练任务中的负样本。每一次切取图像区域时，区域尺寸与检测图像上随机一个标注框相同。由于部分检测图像没有足够空间获取符合条件的负样本，本文从每张检测图像上随机取最多 5 个图像区域作为负样本，最终这里获得的负样本共 27009 个，正样本一共 36377 个，本文将正负样本各取 75%作为训练集，剩下 25%作为测试集，将经过 OBC306 数据集训练的识别模型进一步调优训练。为了不破坏预训练模型学习到的多类字符先验知识，本次训练过程使用的初始学习率设为识别任务的 0.5 倍，训练过程直至模型收敛。

然后，我们取检测模型对测试集进行预测，并获得初步预测框及其置信度分数 $S_1$，其中每张测试图像输出 100 个初步预测框。对于每一个初步预测框，我们使用上一步骤中训练好的识别模型进行识别分类，每个初步预测框都会获得识别模型的置信度分数 $S_2$。

最后，对于每一个初始预测框，我们使用置信度分数 $S_1$ 和 $S_2$ 计算综合分数 $S$，计算公式如公式（3-1）所示，即综合分数 $S$ 为 $S_1$ 和 $S_2$ 的加权调和平均数。在当前任务中，



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

我们将 $m_1$ 和 $m_2$ 分别设为 0.9 和 0.1。当初步预测框的综合分数 $S$ 超过置信度阈值，则这个初步预测框作为检测模型的正式预测框之一。

$$S=\frac{\sum_{i=1}^{2}m_i}{\sum_{i=1}^{2}\frac{m_i}{S_i}} \tag{3-1}$$

## 3.4 实验结果与分析

本文接下来通过实验测试上文介绍的基于字模数据的甲骨文字符动态增广算法以及利用识别模型辅助的检测算法的性能。

本节实验使用 3.1 节中构建的甲骨文单字符检测数据集 OBCD 作为实验数据集。这里随机选择 75%的检测数据作为训练集，25%的数据作为测试集，训练和测试的样本数量和实例数量如表 3-1 所示。

表 3-1 甲骨文单字符检测实验训练和测试数据统计

| | 样本数量 | 字符实例数量 |
|---|---|---|
| 训练数据 | 4380 | 36377 |
| 测试数据 | 1458 | 12444 |
| 总量 | 5838 | 48821 |

在评价指标方面，本节实验使用目标检测领域通用的准确率（$P$）、召回率（$R$）和 F-measure 作为评价检测效果的指标，F-measure 是一个综合考虑准确率与召回率的指标，这三个指标的计算方法如公式（3-3）~（3-5）所示，

$$P=\frac{TP}{TP+FP} \tag{3-3}$$

$$R=\frac{TP}{TP+FN} \tag{3-4}$$

$$F=\frac{2*P*R}{P+R} \tag{3-5}$$

其中 $TP$ 为正确预测数，$FP$ 为错误预测数，$FN$ 为未被检测到的字符实例数。因此准确率为正确预测数（$TP$）与总预测数（$TP+FP$）的比值，召回率为正确预测数（$TP$）与文字实例总量（$TP+FN$）的比值，而 F-measure 为准确率和召回率的调和平均数。



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

### 3.4.1 检测算法框架与动态增广实验

本文首先探究 3.2 节介绍的动态增广算法与 3.3 节介绍的 R-FCN 结合 FPN 的检测算法框架在甲骨文检测实验中的性能。本节实验在模型结构方面使用 R-FCN 和 FPN，在动态增广算法方面，设置每张训练图像随机嵌入实例个数 $N_{syn}$ 为 10。

测试分为单尺度测试以及多尺度测试，单尺度测试时图像输入尺寸在 800×1280，多尺度测试时图像输入尺寸共有 6 组，分别为 480×800、576×900、688×1100、800×1280、1200×1600 和 1400×2000。最终检测实验结果如下表 3-2 所示。从表 3-2 可得，检测算法框架引入 FCN 后的 F-measure 指标比仅使用 R-FCN 的时候更高，特别是在多尺度测试的条件下，FPN 根据自身结构特点，能更好地覆盖字符的尺度范围，从而较大地提升了甲骨文字的召回率，说明 FPN 对检测多种尺度的甲骨文字符具有明显的提升作用。而动态增广算法方面，在单尺度测试的条件下，使用动态增广算法使整体的 F-measure 有所下降；但在多尺度测试的条件下，使用动态增广算法后令 F-measure 有所提升。这是因为单尺度测试时，检测模型仍然不具有检测过大与过小的甲骨文字符的能力，能检测的字符尺度范围内的难检字数量有限，即使检测模型能更鲁棒地预测难检字，对 F-measure 的贡献也比较少，而且使用动态增广算法影响了整体数据分布，因此最终整体的 F-measure 下降；而在多尺度测试的条件下，能检测的字符尺度范围增大，难检字数量增多，数量比重增加，对 F-measure 的贡献提升，而使用动态增广算法使检测模型更鲁棒地预测难检字，因此最终使用动态增广算法令整体 F-measure 提升。

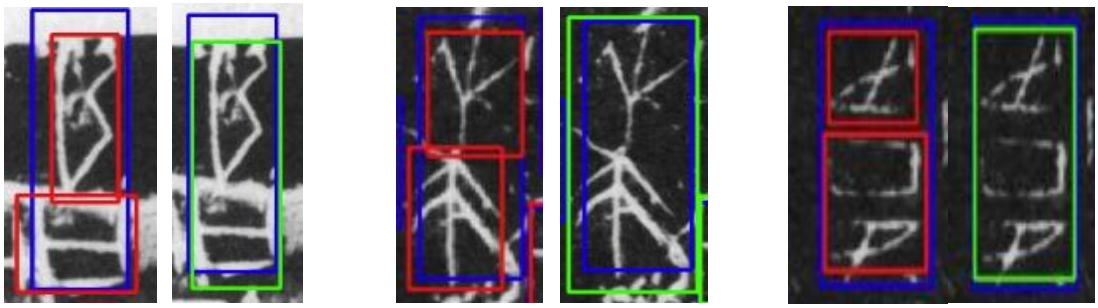
表 3-2 甲骨文单字符检测实验结果

| | 单尺度测试 | | | 多尺度测试 | | |
|---|---|---|---|---|---|---|
| | *P* | *R* | *F* | *P* | *R* | *F* |
| R-FCN | 87.96% | 69.33% | 77.54% | 85.45% | 74.64% | 79.68% |
| R-FCN + FPN | 89.64% | 70.81% | 79.12% | 82.15% | 85.10% | 83.13% |
| R-FCN + FPN + 增广 | 90.83% | 69.19% | 78.55% | 83.39% | 83.58% | 83.49% |

图 3-16 是使用动态增广算法前后，R-FCN+FPN 模型对复合字符的检测样例图，其中每个子图的左图为表 3-2 中“R-FCN + FPN”方法的预测结果，蓝色框为标注框，红色框为预测结果，右图为表 3-2 中“R-FCN + FPN + 增广”方法的预测结果，蓝色框为标注框，绿色框为预测结果。从图 3-16 可以看到，使用动态增广算法之后，原本被错检或漏检的复合字都能被正确检测。因此可以看出，当前基于字模数据的甲骨文动态增广算法成功地令检测模型更鲁棒地预测难检的复合字符。



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net



a) b) c)

图 3-16 使用动态增广算法前后的检测结果对比

表 3-3 不同的动态嵌入个数的检测实验结果

| $N_{syn}$ | $N'_{syn}$ | 单尺度测试 | | | 多尺度测试 | | |
|---|---|---|---|---|---|---|---|
| | | *P* | *R* | *F* | *P* | *R* | *F* |
| 0 | 0 | 89.64% | 70.81% | 79.12% | 82.15% | 85.10% | 83.13% |
| 5 | 3.6131 | 90.84% | 69.93% | 79.02% | 82.99% | 83.96% | 83.47% |
| 10 | 6.1521 | 90.83% | 69.19% | 78.55% | 83.39% | 83.58% | 83.49% |
| 15 | 7.9141 | 90.73% | 68.68% | 78.12% | 82.73% | 83.81% | 83.27% |
| 20 | 9.0804 | 91.10% | 68.71% | 78.34% | 82.37% | 83.94% | 83.14% |

接下来本文进一步探究每张图嵌入的合成文字实例个数 $N_{syn}$ 对实验结果的影响，这里令其他实验条件保持不变，然后设置 $N_{syn}$ 分别为 0、5、10、15、20 进行实验。实验结果如表 3-3 所示。从实验结果可得，首先由于部分图像没有足够空间，平均每张训练图像的实际嵌入个数 $N'_{syn}$ 未能达到 $N_{syn}$。然后在单尺度测试的情况下，$N_{syn}$ 为 5 的效果较好，但由于能检测的字符尺度范围内的难检字数量有限，即使检测模型能更鲁棒地预测难检字，对 F-measure 的贡献也比较少，而且使用动态增广算法影响了整体数据分布，因此使整体的 F-measure 比未进行动态增广时的略低；但是多尺度测试的情况下，$N_{syn}$ 为 10 的效果在所有实验中达到最好，而且所有使用动态增广算法的实验结果都高于使用之前的结果。除此之外，$N_{syn}$ 高于 10 以后，随着 $N_{syn}$ 和 $N'_{syn}$ 的提升，由于合成数据过度影响了训练数据分布，算法的检测结果逐渐下降。

### 3.4.2 识别辅助检测实验

本节对 3.3 节提出的识别辅助检测算法进行性能上的探究。本节实验仍然使用 R-FCN 结合 FPN 作为检测算法框架，并在加入 3.2 节提出的甲骨文动态增广算法的基础上进行探究，每张图嵌入的新文字实例个数 $N_{syn}$ 设为 10。实验结果如表 3-4 所示。由表 3-4 的结果可得，在进行单尺度测试情况时，使用识别任务辅助可以召回一些因分辨

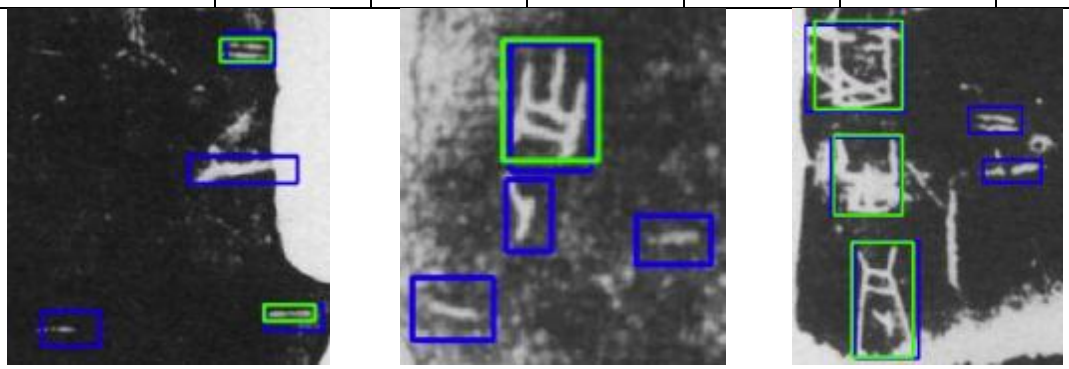


(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

率不足而被检测模型舍弃的正确预测框，较好地提升检测的召回率，从而提升整体检测性能；在进行多尺度测试，整体预测候选框较多的情况下，使用识别任务辅助可以有效提升检测准确率，从而提升整体的检测性能。

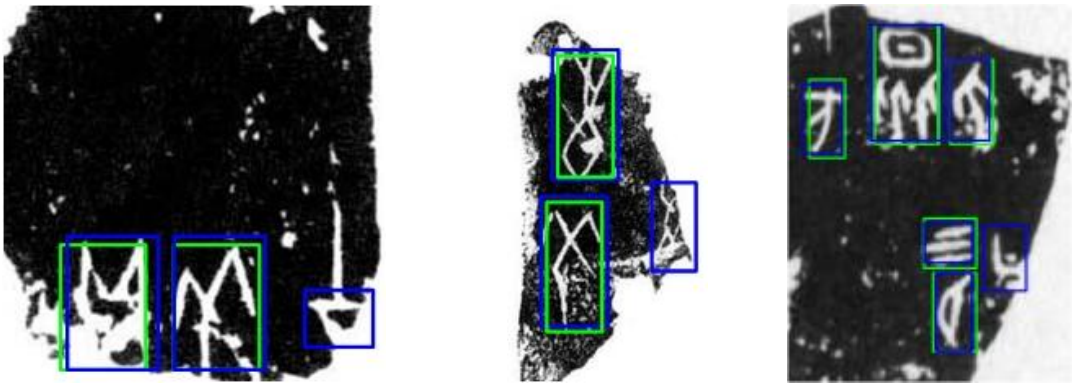
表 3-4 使用识别辅助检测前后的实验结果

| | 单尺度测试 | | | 多尺度测试 | | |
|---|---|---|---|---|---|---|
| | *P* | *R* | *F* | *P* | *R* | *F* |
| 不使用识别辅助 | 90.83% | 69.19% | 78.55% | 83.39% | 83.58% | 83.49% |
| 使用识别辅助 | 90.57% | 70.64% | 79.37% | 84.69% | 82.90% | 83.79% |



a） b） c）

图 3-17 类似“一二三”的甲骨文字的检测失败示例



a） b） c）

图 3-18 字形残缺的甲骨文字符的检测失败示例

### 3.4.3 问题总结与分析

从上述实验的结果可得，本文提出的甲骨文动态增广算法成功地让检测模型能够更好地检测难检字，而识别辅助检测算法也成功地帮助检测模型进一步提升了甲骨文检测任务的 F-measure 指标。但当前检测任务中还有一些问题未得到解决，主要有以下几点：

首先，数据集中仍有许多较小且易混淆的甲骨文字符不能被正确检测。甲骨拓片上



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

存在类似汉字“一二三”的甲骨文字符，这些字符在甲骨片上用于表示该卜辞被用于占卜的次数，与卜辞内容无关，所以它们并没有像其他文字一样排列或聚集，在空间分布上比较随机，又由于甲骨片上还分布着许多无意义的裂痕和刮痕，因此这些类似“一二三”的甲骨文字符容易和这些痕迹混淆，不容易正确检测。除此之外，这些字符在较大的甲骨拓片上分布较多，和整张图片相比，其标注框的相对尺寸较小，即使本章引入在 3.3 节介绍的特征金字塔网络 FPN，充分利用了多种分辨率的特征表达图，也无法完全覆盖这些较小的甲骨文字，这些字符的检测难度仍然存在。图 3-17 即为类似“一二三”的小文字的字符样例，其中蓝色框为标注框，绿色框为算法的预测框，由图 3-17 可得，目前这些类似汉字“一二三”的甲骨文字符的检测召回率较低。

除此之外，如 3.1.2 节所述，数据集中有较多残缺字符，这些字符处于拓片边缘，或者被较大的裂痕穿遮盖，其字形很难看清楚，当前算法仍然容易忽略或者错误检测这些文字，图 3-18 为残缺字符的样例，其中蓝色的为标注框，绿色为当前算法的预测框。

## 3.5 本章小结

本章介绍了甲骨文字符检测方面的探究工作，首先介绍了构建的甲骨学领域首个甲骨文单字符检测数据集 OBCD，并针对其难点挑战，提出了一种基于字模的甲骨文字符动态增广算法，令检测模型更鲁棒地预测出难检的复合字符，同时还提出了一种识别辅助检测算法，在易与甲骨文字符混淆的甲骨划痕较多的情况下过滤错误预测框，进一步提升检测算法的整体性能。



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

# 第四章 甲骨文字符识别

本章将介绍论文在甲骨文字符识别方面的探究工作。本章首先构建了至今为止数据规模最大的甲骨文字符识别数据集，对样本进行统计属性分析以及难点分析，然后针对当前任务中字符类样本分布不均衡的问题，提出一种基于循环式生成对抗网络（CycleGAN）的甲骨文数据增广算法，通过合成甲骨文字符图像平衡类样本分布，解决长尾效应对识别任务的影响。

## 4.1 构建甲骨文单字符识别数据集 OBC306

为构建数据集 OBC306，首先收集了广泛受学术界认可的甲骨文出版著录[3-10]作为数据集的素材来源，接着将著录的所有页面扫描为数字图像，然后借助“甲骨文字典”工具，从沈建华整理的《甲骨文字形表》[58]出发，检索出表中甲骨文字符在拓片上出现的所有位置，并手动裁剪出来作为每个字符的样本图像，汇集构建数据集 OBC306。下面将具体阐述构建过程。

### 4.1.1 构建过程

#### 4.1.1.1 数据集图像来源

为构建数据集 OBC306，选择了包括 3.1.1 节中使用的《甲骨文合集》[3]在内 8 本具有权威性的甲骨文出版著录，如表 4-1 所列，这些甲骨文著录基本涵盖了目前所有出土甲骨拓片。

在这些著录中，除了《甲骨文合集》[3]，《小屯南地甲骨》[4]收录了 1973 年中国社会科学院考古研究所安阳工作队在河南省安阳市小屯村南地的科学挖掘成果，包含了 4589 片甲骨，这本著录记载了许多以往甲骨文资料中未出现过的内容，而且是经过科学的田野挖掘获得，因此具有很大的学术价值；《甲骨文合集补编》[7]作为《甲骨文合集》[3]的补充，在 1999 年出版，收录了 13450 张甲骨拓片，包括《小屯南地甲骨》[4]《英国所藏甲骨集》[5]等新材料，和其他未被《甲骨文合集》[3]收录的素材。

除了收录国内所藏甲骨的著录，由于许多甲骨片流散于世界各地，还有一些著录收录了海外国家所藏甲骨。本文选取了 5 本国外甲骨的相关著录[5,6,8-10]，这些著录基本上覆盖了所有海外国家所藏甲骨。在这 5 本著录中，《英国所藏甲骨集》[5]一共收录了 2735 片甲骨片，基本覆盖了英国所藏甲骨，这些甲骨的文字信息丰富，风格上具有特点，具有较高学术价值；《苏德美日所见甲骨集》[6]收录了 692 片甲骨，来自苏联、德国、美国和日本四个国家，都是由甲骨学专家临摹记录；《怀特氏等收藏甲骨文集》[8]收录 1915



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

片由加拿大甲骨学者收藏的甲骨片，这些甲骨片记录了殷商时期的军制信息，因此具有较高学术价值；《东京大学东洋文化研究所藏甲骨文字》[9]收录了 1315 片甲骨，这些甲骨片来自三位日本收藏者，著录上的每一片都有自己的拓片和照片；《天理大学附属天理参考馆甲骨文字》[10]收录的甲骨同样来自日本，由天理大学附属天理参考馆收藏，共 747 片拓片。

选取了上述著录之后，我们使用相同分辨率对所有页面进行扫描。需要注意的是，部分著录内容存在重复，例如《甲骨文合集补编》[7]与《小屯南地甲骨》[4]《英国所藏甲骨集》[5] 部分内容重复，因此在扫描生成页面图像的过程中，我们进行去重操作，去除重复甲骨拓片。

表 4-1 构建 OBC306 所选甲骨文著录及其简称

| 著录名称 | 简称 |
|---|---|
| 《甲骨文合集》[3] | h |
| 《小屯南地甲骨》[4] | t |
| 《英国所藏甲骨集》[5] | y |
| 《苏德美日所见甲骨集》[6] | s |
| 《甲骨文合集补编》[7] | b |
| 《怀特氏等收藏甲骨文集》[8] | w |
| 《东京大学东洋文化研究所藏甲骨文字》[9] | d |
| 《天理大学附属天理参考馆甲骨文字》[10] | l |

#### 4.1.1.2 处理步骤

本文接下来阐述从扫描图像开始生成单字符图像的过程。在整个过程中，本文应用了两个重要的工具，分别是《甲骨文字形表》[58]以及“甲骨文字典”软件工具[50]。

《甲骨文字形表》[58]由沈建华、曹锦炎编著，这两位作者根据已有甲骨文资料，整理了 4024 个不同的甲骨文字符及其异体字形，目前已广泛受甲骨学界认可和使用。这本著录在编纂过程中运用了香港中文大学中国文化研究所甲骨文电子资料库[20]作为数据基础。

“甲骨文字典”软件工具[50]即为 3.1.1 节中所使用的工具，这个工具除了可以根据甲骨片编号检索甲骨上包含的甲骨文字符，还可以直接检索出每个甲骨文字符出现过的甲骨片，借助这个电子字典工具，我们不需要直接操作甲骨拓片原件，减少检索甲骨文字符



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

出现位置的时间成本，从而能够更有效率地构建高质量数据集。


图 4-1 数据集 OBC306 的构建步骤

构建数据集的具体处理步骤如图 4-1。对于甲骨文字形表的每个字符，首先使用“甲骨文字典”检索出字符在所有拓片上出现的位置，然后对于每个位置，手动使用工具裁剪出该字符实例，并作为该字符的单字符图像，每个字符即为一个单独的字类。裁剪过程中使用矩形框进行框选，同时尽可能让框紧贴该字符实例，使裁剪出来的图像不包含过多冗余信息。在整个处理过程中，所有步骤都在甲骨学专家的指导下进行，因此数据集具有较高标注质量。图 4-2 展示了数据集 OBC306 的图像样例。虽然受时间和人力成本所限，数据集 OBC306 并没有完全涵盖字形表中的所有甲骨文字符，但当前样本数量已达到 30 万级别，已是目前数据量最大的甲骨文单字符数据集。

图 4-2 数据集 OBC306 中的图像样本展示

## 4.1.2 数据统计属性与识别挑战

### 4.1.2.1 数据统计属性

数据集 OBC306 的样本数量分布如图 4-3 所示。OBC306 一共含有 309551 张字符样本图像，共涵盖 306 个类，每个类对应一个不同的甲骨文字符，虽然类平均样本量只略



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

高于 1000，但每一类的样本数量却相差显著，数据集样本分布不均。在该数据集中，样本数量超过 1000 张的类别共有 70 类，而这些类别的总样本量为 259460 张，占整个数据集的 83.62%。在这 70 个类别中，有 5 类的样本数超过 10000 张，这 5 类的总样本量达到 78963，其中，编号为 094023 的类别是数据集中样本量最大的类，其样本量达到 25898。从上述数据可以得知，数据集中具有一部分高频使用的甲骨文字符，而这些字符的总样本量在整个数据集中占据很大比重。其次，共有 126 个不同的甲骨文字符的样本量在 100 到 1000 之间，这部分甲骨文字符的总样本量为 47081，占整个数据集的 15.21%。样本量在 10 到 100 之间的类别共有 58 个，共有 2904 张样本，占总样本量的 0.94%。除此之外，还有 52 个甲骨文字符的样本量不超过 10 张，有 29 个字符的样本仅有 1 张。由此可以看出，在包含部分高频甲骨文字符的同时，数据集中还有大量低频的字符，整个数据集的样本在分布上严重不均，类别之间样本量差别很大，样本数量很少的类别很多，我们将分布按照类样本数升序排列画成图 4-3，这种图形的分布被称为“长尾效应”。接下来本文描述数据集中样本的尺寸大小。图 4-4 为数据集所有样本的高度、宽度和高宽比的分布图。由图可得，样本高度的中位数在 50 到 100 之间，而宽度则相对集中在 0 到 100 之间，高宽比的峰值在 1.0 到 2.0 之间。总体而言，数据集中的样本主要为尺寸较小的长条形图像。

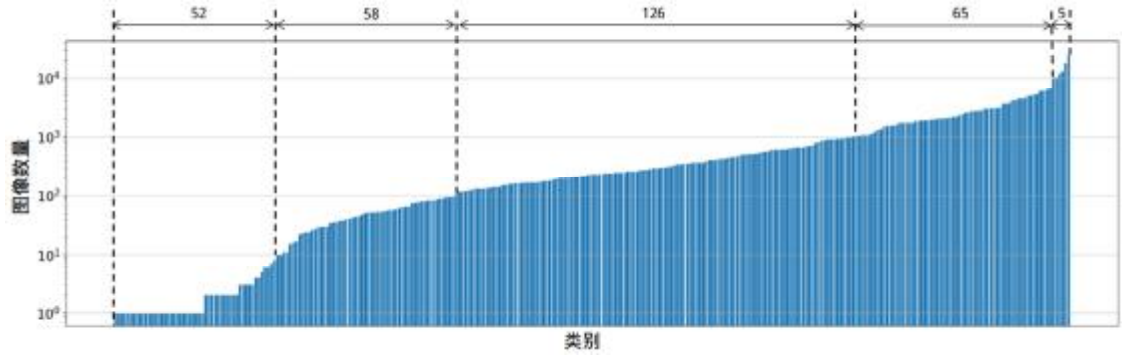



图 4-3 数据集 OBC306 中每个类的样本数量分布（升序排列）

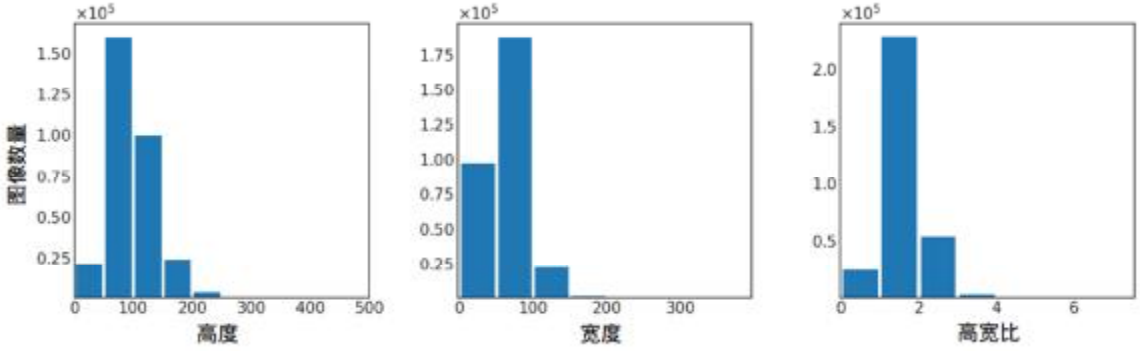
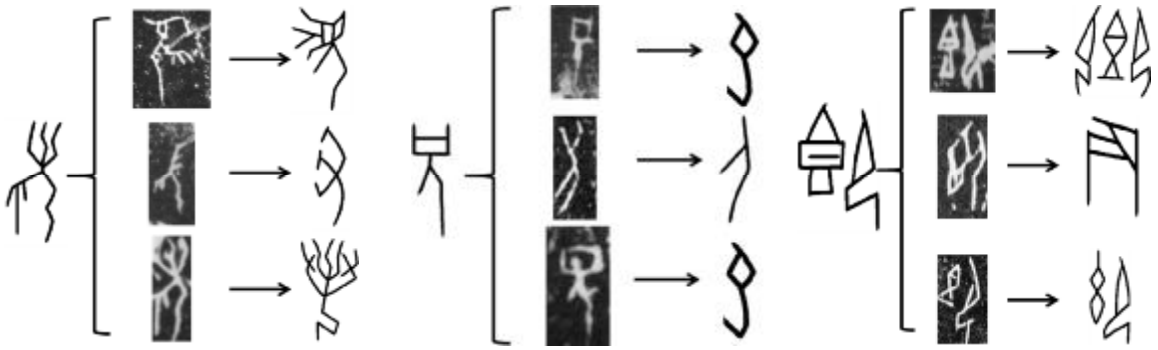



图 4-4 数据集 OBC306 中的样本尺寸分布



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net



a) b) c)

图 4-5 数据集 OBC306 中的异体字图像样例，这些样例同时也是 4.3 节实验的错误样例

#### 4.1.2.2 识别挑战

从上述统计数据可知，数据集 OBC306 数据量大，在大小和形状方面具有一定的多样性，同时也可知样本数量在每个类中严重分布不均，呈现出长尾效应，这些因素都能够体现出该数据集是一个具有挑战性的数据集。除此之外，从字形识别的角度上，该数据集中还有另外两个难点，接下来进行详细说明。

甲骨文字符识别的第一个难点在于甲骨文存在异体字。甲骨文属于象形文字，但同时还不够规整，相对自由，其字形更注重于描述事物的具体特征，因此字符中部位的相对位置并不固定，笔画的数量也会上下浮动，最终字形呈现就会出现变化。数据集 OBC306 中的样本也存在异体的情况，具体的字符样例如图 4-5 所示，括号左边的为样本所属字符，黑底白字的单字图像为异体字图像，箭头右端为后续实验的错误识别结果。

除此之外，噪声问题是另外一个难点。由于长年累月的掩埋等原因，挖掘出来的甲骨普遍存在受损情况，因此数据集 OBC306 的样本图像也受多种噪声影响，样本噪声实例如图 4-6，主要可以分成以下四种情况：

第一，许多拓片上的甲骨文笔迹部位出现破损，导致多个笔画被缺口粘连在一起，从而对应样本图像上的字符局部被白色噪声覆盖，导致字形不够清晰，具体如图 4-6 a)；

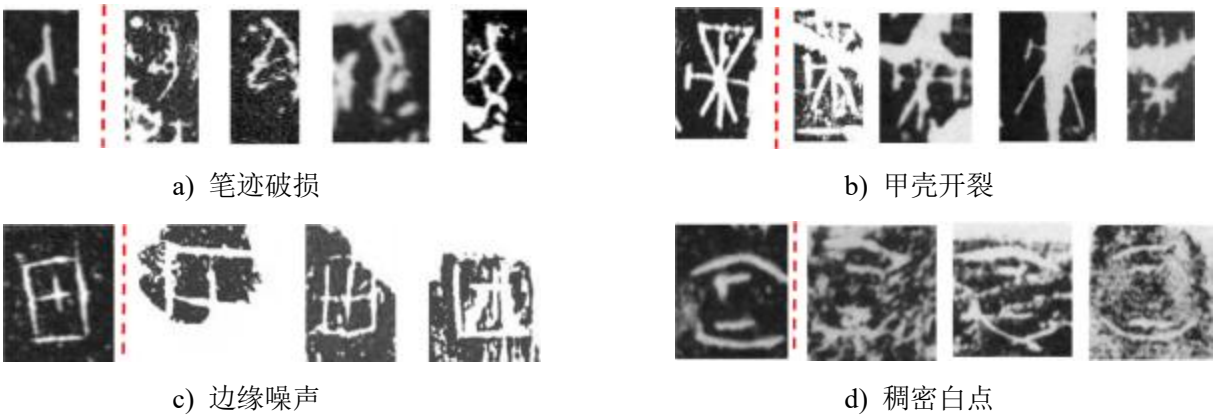
第二，由于殷商时期王室使用龟甲或兽骨进行占卜时，会将其炙烤至开裂，因此裂痕会经过部分字符，导致拓片图像上的字形被白色条状区域遮盖，图片样例如图 4-6 b)；

第三，由于挖掘时未进行保护等原因，许多甲骨在挖掘出来时已经是碎片，导致许多处于碎片边缘的字符不完整，对应拓片图像则体现为由图像边缘延伸出的大片白色区域将字形遮盖，导致字形残缺，影响字形识别，单字符图像样例如图 4-6 c)；

第四，由于掩埋过程中与沙石碰撞等原因，许多甲骨的表面出现严重磨损，对应拓片图像则体现为细密的白色噪声，导致字形变得模糊，图像样例如图 4-6 d)。



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net



a) 笔迹破损　　b) 甲壳开裂

c) 边缘噪声　　d) 稠密白点

图 4-6 数据集 OBC306 中的样本噪声样例展示，这些样例同时也是 4.3 节实验的错误样例

## 4.2 基于循环式生成对抗网络的甲骨文字符增广算法

由 4.1 节对 OBC306 统计属性分析可以知道，当前甲骨文单字符识别任务存在样本数量分布不均问题。甲骨文字符数据的获取非常困难，需要出土新的甲骨文物，并且通过甲骨学专家耗费大量时间和精力进行辨认标注之后，才能获得新的数据。因此，通过获取新的真实数据提升识别效果的方式是难以实现的，因此需要通过算法合成数据提供给识别模型强化训练。

另一方面，由 3.2.1 节的阐述知，本文拥有的字模数据集中每一类字的数据偏少，自身的多样性偏弱，手动模拟真实拓片上的噪声风格不仅令合成图像不够真实，而且图像多样性比较差，因此这里需要一个可以自行学习甲骨拓片风格的、可以生成具有丰富多样性的字符图片的算法。生成对抗网络可以学习风格之间的映射关系，训练好的生成器能够将某种风格的数据转化为新风格的数据，因此本文引入生成对抗网络[59]作为甲骨文字符数据合成的算法。由于本文使用的甲骨文字模数据集与 OBC306 的甲骨文字符在空间上没有完全对应，无法构成配对图像，因此我们引入的是不需要配对图像也能合成指定属性图像的循环式生成对抗神经网络（CycleGAN）算法。下面详细介绍算法的基本原理与增广流程。

### 4.2.1 循环式生成对抗网络基本原理

经典的生成对抗网络[59]的基本思想在于生成器 $G$ 和判别器 $D$ 的零和博弈，其中生成器 $G$ 的任务是接收从输入分布 $P_I$ 采样的输入向量，经过自身处理后输出高维向量，而输出的高维向量服从的分布 $P_G$ 尽可能接近目标分布 $P_D$。判别器的任务是接收来自某个分布的高维向量，并判断该向量是否真的来自目标分布 $P_D$。因此在这个过程中两者形成竞争关系，生成器的目标是令输出分布接近目标分布，使输出向量能欺骗判别器，



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

而判别器的目标是分辨出输入向量来自目标分布还是来自生成器，两者交替提升，最终完成模型训练。由于生成对抗网络可以令生成器 $G$ 的输出分布接近目标分布，因此它被应用于图像合成任务[60-64]。然而，最初生成对抗网络[59]无法控制合成图像包含的具体属性，尽管文献[61]通过增加图像作为模型的指导信息，从而实现指定图像的生成，但缺点是需要目标位置完全对应的配对图像作训练。

CycleGAN[62]的基本思想仍然与经典网络相同，但是与以往的工作[59,61]相比，既可以做到控制特定图像风格之间的转换，同时这项工作又通过网络层面的设计，可以摆脱配对训练图像的要求，因此，本文选择使用 CycleGAN 进行伪甲骨字符生成。CycleGAN 的基本框架如图 4-7 所示，为了令训练难度下降，本文将真实字模图像进行黑白反转，修改成黑底白字图像。以往工作只训练一个生成器 $G$ 和一个判别器 $D$，最后实现的是其中一种风格向另一种风格的单向转换。而循环式生成对抗神经网络则训练了两个生成器 $G_{AB}$ 和 $G_{BA}$、两个判别器 $D_A$ 和 $D_B$。如图 4-7 红色箭头所示，在训练过程中形成两个对称的循环，在每一次迭代中都会执行这两个循环，通过计算设计的损失函数并更新网络参数，同时学习从 $A$ 到 $B$、从 $B$ 到 $A$ 两种映射，从而摆脱了以往工作中需要的配对图像数据，不需要对数据作出特殊要求就可以控制指定图像风格的转换。在本文中，$A$ 指代字模风格领域，$B$ 指代真实甲骨文风格领域，因此本文最后可以实现字模数据风格以及真实甲骨文风格之间的双向转换。

CycleGAN 的目标函数如式（4-1）所示。目标函数包括两个部分，生成对抗损失 $L_{GAN}$ 与循环一致性损失 $L_{cyc}$，$\lambda_{cyc}$ 是循环一致性损失的权重。接下来一一介绍这几项损失。

$$L_{total}=L_{GAN}+\lambda_{cyc}L_{cycle} \tag{4-1}$$

首先是生成对抗损失 $L_{GAN}$，这一项生成对抗损失与经典网络的形式相同。对于从 $A$ 到 B 的转换路径，其生成对抗损失 $L_{GAN}(G_{AB}, D_B, A,B)$如公式（4-2）所示。

$$L_{GAN}\left(G_{AB}, D_B, A, B\right)=E_{y\sim P_B(y)}[\log D_B(y)]+E_{x\sim P_A(x)}[1-\log D_B(G_{AB}(x))] \tag{4-2}$$

其中 $P_A$ 为字模数据的数据分布，$P_B$ 为真实甲骨文数据分布，这部分损失对应图 4-7 右半部分两个“真实字符图像?”环节得到的损失。这里的生成对抗损失与经典网络的损失形式相同，目的在于令甲骨文合成图像与甲骨文真实图像尽量相似。同理，对于从 $B$ 到 $A$ 的转换路径，生成对抗损失 $L_{GAN}(G_{BA}, D_A, B,A)$ 如公式（4-3）所示。

$$L_{GAN}\left(G_{BA}, D_A, B, A\right)=E_{x\sim P_A(x)}[\log D_A(x)]+E_{y\sim P_B(y)}[1-\log D_A(G_{BA}(y))] \tag{4-3}$$



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

这部分损失对应图 4-7 中左半部分两个“真实字模图像?”环节得到的损失。

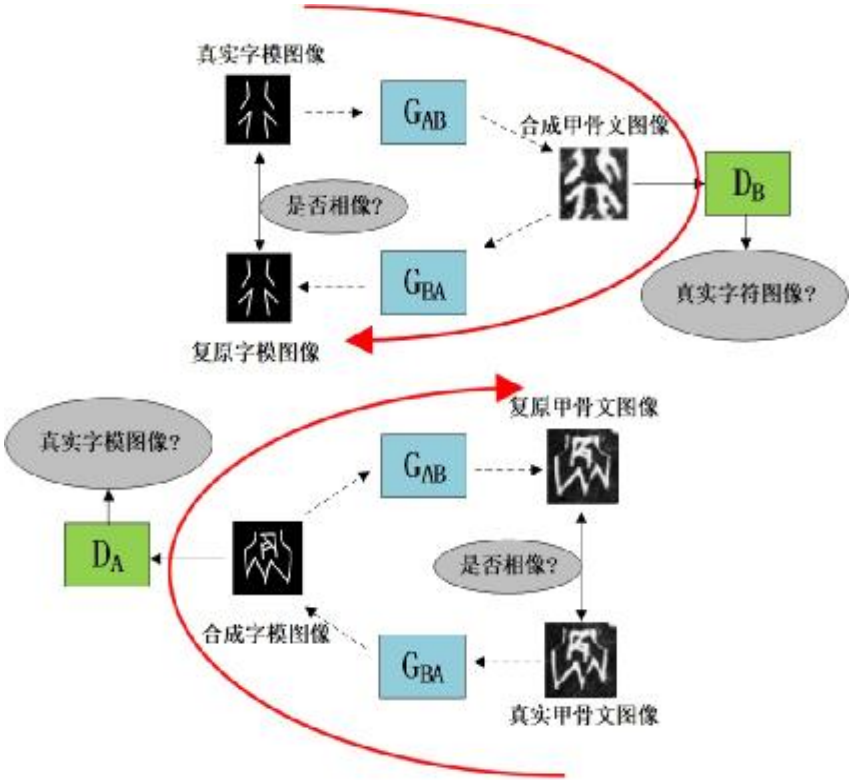



图 4-7 循环式生成对抗神经网络的基本框架

然后是循环一致性损失 $L_{cyc}$，其计算方法如公式（4-4）所示。

$$L_{cyc}\left(G_{AB}, G_{BA}\right)=E_{x\sim p_A(x)}[\left\|G_{BA}(G_{AB}(x))-x\right\|_1]+E_{y\sim p_B(y)}[\left\|G_{AB}(G_{BA}(y))-y\right\|_1] \quad （4\text{-}4）$$

这部分损失对应图 4-7 中两个“是否相像?”环节得到的损失，损失的形式是原图像与复原图像的 $L_1$ 范数。在仅仅使用生成对抗损失 $L_{GAN}$ 的情况下，如果网络容量足够大，那生成器 $G_{AB}$ 完全能做到将任意的字模数据转换成一模一样的甲骨文合成图像，虽然还符合输出分布接近目标分布的给定条件，但这种结果显然不符合本文的目的。为了约束映射函数的解空间，算法引入了循环一致性损失 $L_{cyc}$，通过令网络多学习一个风格转换方向，并约束复原图像和原图尽量接近，使生成器学习到正确的风格转换方式。

由图 4-6 可知，数据集 OBC306 存在许多噪声严重的真实样本，为了令生成对抗网络在学习两种风格的映射关系的同时，能够生成字形相对清晰的合成图像，本文在计算循环一致性损失 $L_{cyc}$ 时引入权值进行引导。在计算式(4-4)的第一项循环一致性损失时，即真实字模图像与复原字模图像的循环一致性损失时，对笔画区域的 $L_1$ 范数赋予较高的权重 $\lambda_f$，对图像背景区域的 $L_1$ 范数采用较低的权重 $\lambda_b$，具体公式如式(4-5)所示。



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$$\left\| G_{BA}(G_{AB}(x)) - x \right\|_1 = \lambda_f \sum_{i \in S} \| p_i - p'_i \|_1 + \lambda_b \sum_{i \notin S} \| p_i - p'_i \|_1 \quad (4\text{-}5)$$

其中 $S$ 为图像的笔画区域，$p$ 和 $p'$ 分别为真实字模图像与复原字模图像的像素值。对笔画区域设置更高的权重可以令生成器更注重字形区域，从而生成字形更加清晰的合成图像。这里我们将 $\lambda_f$ 和 $\lambda_b$ 分别设为 10 和 1。

模型训练过程中，每次迭代都会更新生成器 $G_{AB}$、$G_{BA}$ 和判别器 $D_A$、$D_B$ 的参数。更新生成器 $G_{AB}$ 和 $G_{BA}$ 的参数时，将使用 4.2.1 节中的 3 个损失。由于算法结构中 $G_{AB}$ 和 $G_{BA}$ 属于成对存在的关系，因此两个生成器共用同一个目标函数 $L_G$ 及其梯度，并一起更新参数。生成器的目标函数 $L_G$ 公式如（4-6）所示，我们将 $\lambda_{cyc}$ 分别设为 10。

$$L_G = \frac{L_{GAN}(G_{BA}, D_A, B, A) + L_{GAN}(G_{AB}, D_B, A, B)}{2} + \lambda_{cyc} L_{cycle} \quad (4\text{-}6)$$

在更新判别器 $D_A$ 和 $D_B$ 的参数时，由于两者使用的训练数据不同，因此两者单独计算目标函数，并且只使用 4.2.1 节中的生成对抗损失 $L_{GAN}$，两者的计算公式如下。

$$L_{D_A} = L_{GAN}(G_{AB}, D_B, A, B) \quad (4\text{-}7)$$

$$L_{D_B} = L_{GAN}(G_{BA}, D_A, B, A) \quad (4\text{-}8)$$

### 4.2.2 增广数据来源

与 3.2.1 节相同，我们同样引入甲骨文字模数据集作为数据增广的来源。如 3.2.1 所述，甲骨文字模数据集包含 5490 类字形，而 OBC306 的字类基本被字模数据集涵盖，因此后续对 OBC306 的每一类字进行数据增广时，均使用对应字类的字模数据作为输入数据。

### 4.2.3 数据增广算法

接下来本文介绍数据增广算法的流程。先对 CycleGAN 模型进行训练，再利用训练好的模型进行图像合成。为了兼顾字模数据与甲骨文字符数据的尺寸，本文将判别器与生成器的输入图像均定为 64×64 的单通道图像，并根据当前任务确定生成器与判别器的网络结构。以 $G_{AB}$ 为例，生成器的模型结构如图 4-8，首先是 3 个卷积层模块作初步特征提取以及图像的下采样，接着是 4 个残差网络模块，以跳跃式连接结合低层特征和高层特征，然后是 2 个反卷积层模块，通过上采样逐步将特征表征尺度复原至原图大小，最后接上一个卷积层模块，调整输出通道数为原图的通道数，因此生成器最后的输出仍



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

为 64×64 的单通道图像。以 $D_B$ 为例，判别器的模型结构如图 4-9，模型使用全卷积网络对输入图像进行逐层下采样，最终得到当前输入图像为真实甲骨文图像的预测概率。本文参照文献[60]的结论，除了生成器的最后一个卷积模块以及判别器的第一个卷积模块，其他所有模块里都使用了实例归一化层，进行图像实例级别的归一化，从而使模型训练更加稳定。

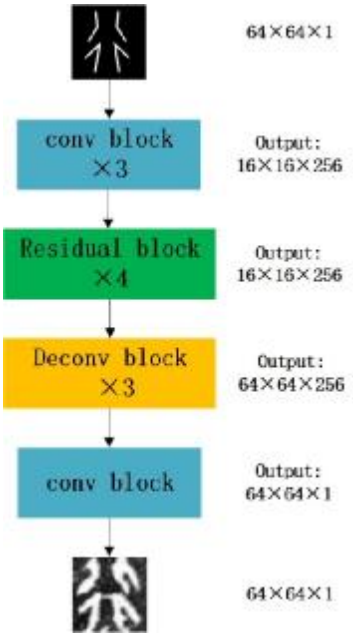



图 4-8 生成器的模型结构

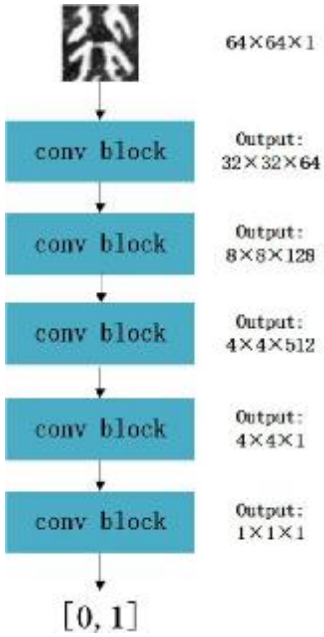



图 4-9 判别器的模型结构

确定网络结构之后，本文开始对模型进行训练，训练时需要甲骨字模风格和真实甲骨拓片风格的训练数据。在甲骨字模风格数据方面，本文选择对应 OBC306 字类的字模数据，并对其进行简单扩充，包括笔画粗细变化、随机仿射变换、随机翻转等，最后获得 127957 张字模数据；在真实甲骨文风格数据方面，本文对 OBC306 中每个字类随机选取 75%数据作为训练集，共 232236 张数据，最后模型一共训练了 12 个 epoch。然后本文开始将模型用于合成甲骨文字符图像。由于模型参数固定的情况下，相同的输入图像只能得到相同的输出图像，为了能从少数的字模数据获得大量的具有丰富多样性的合成图像，我们首先对字模数据进行随机的图像处理，再将处理后的字模图像作为生成器的输入进行图像合成，图像处理操作包括以下几个步骤：

1. 加入随机噪声。
2. 以 10%的概率进行随机水平翻转。
3. 进行仿射变换，其中平移幅度为 0.7～0.9 倍宽高的随机数，旋转角度为-15° 到+15° 之间。



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

4. 进行透视变换，具体操作是将图像分成 8×8 的网格，并让每个网格点向随机方向偏移一次，偏移距离服从均值为 0、方差为 0.01 的正态分布。
5. 进行局部均值模糊，滤波器核大小为 2×2。

## 4.3 实验结果与分析

### 4.3.1 数据集 OBC306 实验

#### 4.3.1.1 实验设置

算法方面，为了应用在当前计算机视觉领域效果较好的深度学习技术，本文选用 5 种目前在该领域应用较广的卷积神经网络模型，分别是 AlexNet[65]，Inception-v4[57]，VGG-16[66]，ResNet-50 以及 ResNet-101[67]，训练模型时使用 Adam 优化策略，训练至模型收敛为止。除了神经网络模型，本文也引入传统的 HOG 特征[68]与 SVM 模型[34]的组合，用于与神经网络模型的对比。

数据方面，本文从每一类数据中随机选取 75%数据作为该类的训练样本，剩下的 25%作为测试样本。对于 29 个仅有 1 张样本的类别，本文将样本作为该类的测试样本，因此该类没有训练样本。

评价方式方面，以往的评价方式为识别准确率 $p'$，计算公式如（4-9），其中 $T$ 为测试集中识别正确的样本数，$N$ 为测试集总样本数。但由于数据集中样本分布不均，识别准确率会受样本量大的类别主导。因此本文使用类平均准确率 $p$ 作为评价标准，公式如（4-10），其中 $T_i$ 为第 $i$ 个类的识别正确样本数，$N_i$ 为第 $i$ 个类的总测试样本数，$n$ 为类别数。（4-10）的优点在于每一类的识别结果被平等看待，后续的识别实验均以此作为评价标准。

$$p' = \frac{T}{N} \tag{4-9}$$

$$p = \frac{\sum_{i=1}^{n} \frac{T_i}{N_i}}{n} \tag{4-10}$$

实验在 Pytorch 代码框架[69]下进行，使用的设备为 Intel i7-6700K 处理器, Nvidia TITAN X 图像处理器以及 32GB 内存。

#### 4.3.1.2 结果分析

所有方法在 OBC306 上的类平均准确率如表 4-2 所示。由表中可得，首先 HOG 结合 SVM 模型的结果比所有神经网络算法的结果差很多；然后在所有 5 种神经网络算法，



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

Inception-v4 在 Top-1 和 Top-3 取得了最好的识别结果，在 Top-5 和 Top-10 取得次优的结果，总体上是表现最好的算法，但显然识别结果还不够高，仍有提升空间。总的来说，与传统的识别算法相比，深度神经网络算法在甲骨文识别领域具有更好的应用效果，但是目前的效果远远未能令人满意，接下来还需要进行提升。

表 4-2 不同方法在 OBC306 上的类平均准确率

| 方法 | Top-1 | Top-3 | Top-5 | Top-10 |
|---|---|---|---|---|
| HOG+SVM | 14.29% | 25.93% | 32.15% | 41.26% |
| AlexNet | 66.75% | 75.31% | 77.46% | 80.56% |
| VGG16 | 67.20% | 76.21% | 78.57% | 81.51% |
| ResNet-50 | 68.72% | 78.12% | 80.69% | 82.63% |
| ResNet-101 | 69.50% | 77.92% | 80.00% | 81.66% |
| Inception-v4 | 70.28% | 78.74% | 80.28% | 82.28% |

接下来本文进行了更深入的结果分析。由于数据集的样本分布存在长尾效应，本文按照单类样本量将字符分为 0～9、10～99、100～999、1000～9999 和 10000 以上共 5 组，并计算使用不同神经网络算法时的组内类平均准确率，结果如图 4-10 所示。由图 4-10 可知，当样本量很少时，所有模型的准确率效果都非常低，随着样本量的提升，组内类平均准确率有明显的提升，由此可以看出，样本量的分布不平衡是甲骨字符识别的一个难点。

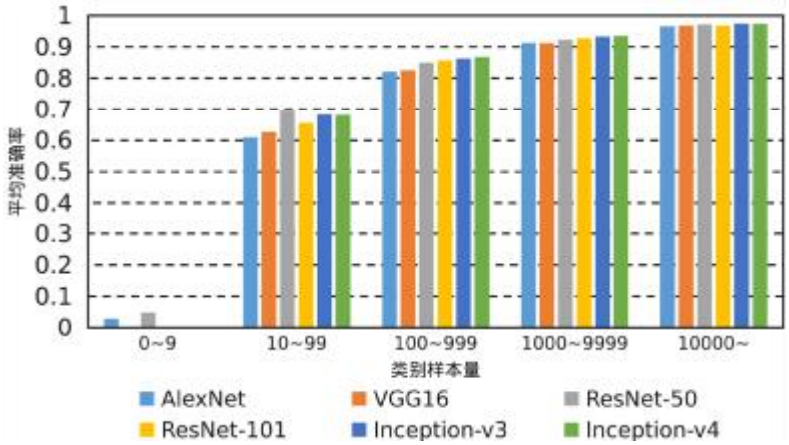



图 4-10 以样本量分组后每组不同算法的类平均准确率

另一方面，异体字以及噪声的问题同样影响算法准确率的提升。图 4-5 展示了测试样本中因异体字问题而导致识别错误的样例，图 4-6 则展示了因噪声问题导致样本识别错误的样例。同时，图 4-11 展示了样本量在 1000 张以上的类别的样本数及其类别准确率，并且类别按照样本量的降序排序，由图可以看到，类别准确率的变化趋势并没有和



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

样本量的变化趋势保持一致，而是总体呈现波动，表明准确率并不仅仅受长尾效应影响。同时图 4-12 展示了图 4-11 中准确率处于低谷的部分类别的错误样本，可以看到这些样本的噪声情况比较严重，这也表明噪声是影响甲骨文单字符识别结果的因素之一。

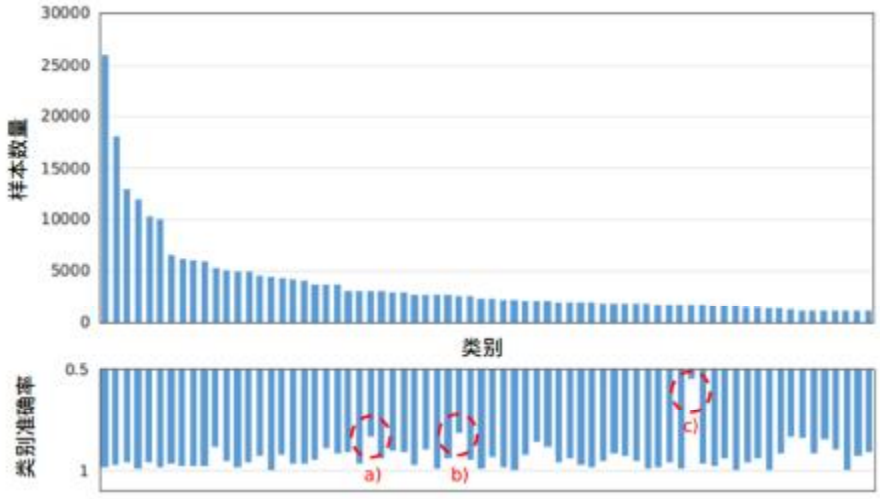
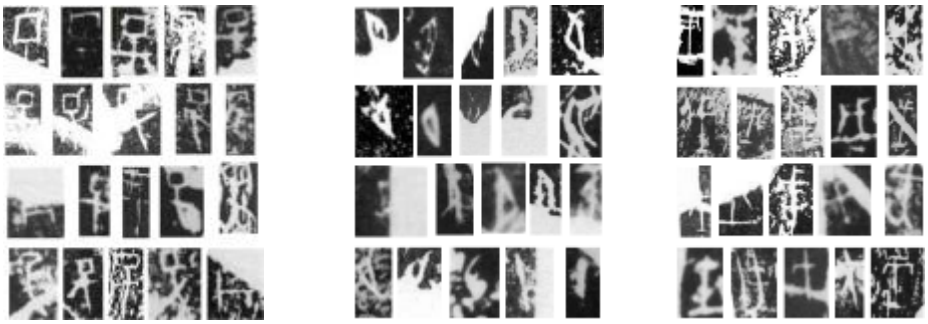



图 4-11 样本量在 1000 以上的类别的样本量与类别准确率



a) b) c)

图 4-12 部分类别的预测错误样本，这些类别在图 4-11 中已标出

从上述结果可以得出结论，在甲骨文单字符识别研究中，数据量分布的长尾效应、异体字现象以及噪声问题是这项工作的难点所在。

## 4.3.2 数据增广算法实验

### 4.3.2.1 与一般增广算法的比较实验

本节探究 4.2 节所述的基于 CycleGAN 网络的数据增广算法的性能。为了进行对比，本文引入一般增广方法对真实甲骨文字符图像进行增广。对于待增广的真实图像，一般增广方法的具体操作与 4.2.3 节中对字模数据所作的图像处理操作相同。本文使用 Inception-v4 作为字符识别实验的算法，并沿用 4.3.1 节中的实验数据配置，然后对于总样本少于 1000 张的类别，即训练图像少于 750 张的类别，本文使用一般增广方法和基于 CycleGAN 网络的增广方法将训练数据补充至 750 张，而 1000 张以上的类别不作任



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

何处理。对于 OBC306 数据集中 29 个只有一张样本的类别，由于这些类别没有训练样本，一般增广方法无法对其进行增广，因此为了公平对比，使用 CycleGAN 网络进行数据增广时将跑两个版本的实验，分别是不对这 29 类作数据增广以及对这 29 类作数据增广。

实验结果如表 4-3 所示。由表 4-3 可以看出，数据增广操作可以提升当前甲骨文字符识别任务的类平均准确率，而且相比于一般增广方法，基于 CycleGAN 网络的数据增广算法对识别任务所有 4 个级别的类平均准确率都有更高的提升效果。除此之外，利用 CycleGAN 网络和字模数据对 29 个没有训练样本的类别进行数据增广，能进一步提升识别任务的类平均准确率，帮助原本完全没有训练数据的类别的识别，这是一般增广算法无法实现的。

接下来我们进行更深入的结果分析。与图 4-10 的分组方式相同，图 4-13 为不同的数据增广方式下，不同样本量的类组的 Top-1 平均准确率，其中增广方法只对样本量在 1000 以下的类别进行增广。可以看出，一般增广方法与基于 CycleGAN 网络的增广方法都令样本量少的类别的平均准确率有所提升，而相对于一般增广方法，基于 CycleGAN 网络的数据增广方法在样本量为 0~9 的类别上优势明显，但样本量为 10~99 与 100~999 的类别的平均准确率略低于一般增广方法，原因是由于部分类别的字模数据数量很少，且与真实甲骨文数据的字形差别较大，使用字模数据合成的图像改变了训练集的数据分布，从而影响了准确率。图 4-14 展示了部分字模数据与真实数据字形不一致的类别。

表 4-3 不同数据增广方法在 OBC306 上的类平均准确率

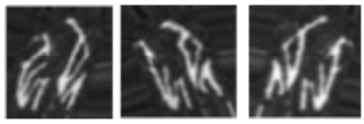
| 增广方式 | 增广 29 类? | Top-1 | Top-3 | Top-5 | Top-10 |
|---|---|---|---|---|---|
| 不作增广 | 否 | 70.28% | 78.74% | 80.28% | 82.28% |
| 一般增广 | 否 | 75.42% | 81.94% | 84.13% | 85.89% |
| 基于 CycleGAN | 否 | 77.76% | 84.66% | 86.19% | 87.44% |
| 基于 CycleGAN | 是 | 86.54% | 93.59% | 95.12% | 96.74% |

表 4-4 展示了不同来源的甲骨文字符样本，第 1、3 列分别为甲骨文字模数据以及对应字类的真实数据，第 2 列为第 1 列的字模数据使用 CycleGAN 网络合成的数据，第 4 列为第 3 列的真实数据使用一般增广方法得到的合成样本。由表 4-4 可以看出，基于 CycleGAN 网络获得的合成数据与作为其来源的字模数据相比，字形笔画上变化较大，同时这些合成数据之间也存在较大差异，具有较丰富的多样性。除此之外，所有合成图



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

表 4-4 两种数据增广算法的生成图像样例与对应的图像来源

| 字模数据 | 基于CycleGAN网络的增广算法 | | | 真实数据 | 一般增广算法 | | |
|---|---|---|---|---|---|---|---|
|  |  |  |  |  |  | | |
|  |  |  |  |  |  | | |

像都与真实甲骨文图像具有非常接近的图像风格。另一方面，一般增广方法得到的数据与作为来源的真实甲骨文数据相比，虽然也保留了原来的风格，但是图像整体变化不大，多样性与生成对抗网络合成的数据相比较低。而且由于一般的增广方法对整张图像进行了空间上的变换，容易使图像边缘区域失真扭曲，影响图片质量。综上所述，从视觉上看，基于生成对抗网络的增广方法获得的合成图像具有接近真实的图像风格，同时也有足够的多样性，与一般增广方法得到的图像相比，整体质量更高。

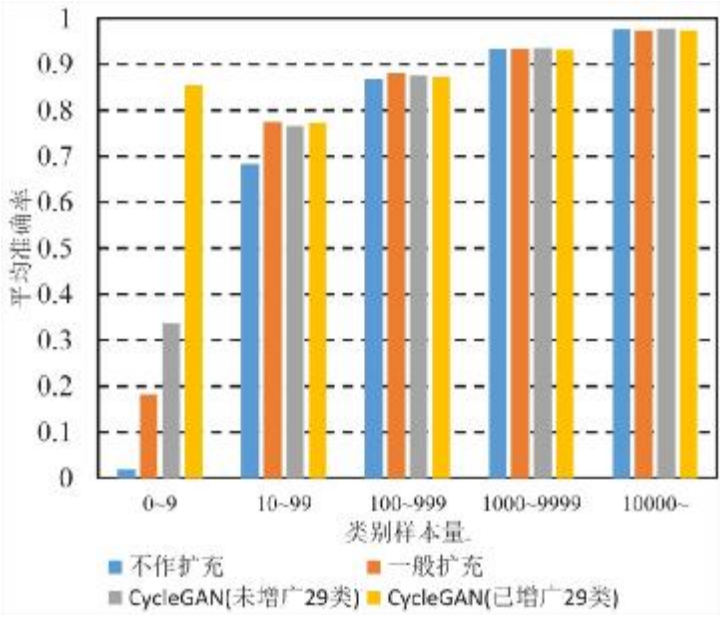



图 4-13 不同的数据增广方式下，不同样本量的类组的平均准确率

a） b） c）

图 4-14 部分字模数据与甲骨文真实数据字形差别较大的类别样例



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

#### 4.3.2.2 数据增广数量的探究实验

本节探究一般增广算法和基于 CycleGAN 网络的数据增广算法在不同增广数量下的性能。我们分别对样本量在 1000 以下、3000 以下和 5000 以下的类别进行训练数据的增广，探究在这些情况下两种增广算法对识别任务的类平均准确率的影响，本节实验均没有对 29 个没有训练样本的类别进行增广。实验结果如表 4-5 所示，由表可得基于 CycleGAN 网络的增广算法在 3 种实验条件下对类平均准确率的提升效果都比一般增广算法的更好，但由于合成数据的分布与真实数据分布存在一定差异，训练数据中加入过多合成数据后会改变数据分布，影响模型训练。因此随着增广量的上升，使用两种方法后的整体准确率都略微下降。

表 4-5 不同的训练数据增广量下，甲骨文字符识别任务的 Top-1 类平均准确率

| 增广方式 | 增广 1000 以下 | 增广 3000 以下 | 增广 5000 以下 |
|---|---|---|---|
| 一般增广 | 75.42% | 75.17% | 75.09% |
| 基于 CycleGAN | 77.76% | 77.52% | 76.70% |

#### 4.3.2.3 当前增广算法存在的问题

由上述结果可得，本章提出的基于 CycleGAN 网络的数据增广算法可以提升甲骨文字符的识别准确率，同时比一般增广算法的提升效果更好，然而当前的增广算法还存在一些问题。尽管当前算法成功学习到甲骨文字符数据的字符风格，但是如 4.3.2.1 节实验所述，当前基于 CycleGAN 网络的数据增广算法对甲骨文的字形变化较小，暂时还不能生成字形更加多变的甲骨文合成数据。如果需要进一步提升算法效果，需要进一步调整当前数据增广算法使其能生成字形丰富程度更高的合成图像。

## 4.4 本章小结

这一章首先介绍了本文构建的甲骨文字符数据集 OBC306，这是目前已知数据量最大的甲骨文单字符数据集，本文给出了数据集的具体构建过程，并提供详细的数据统计以及难点挑战总结，然后给出多种算法的识别结果以及具体的结果分析；然后针对甲骨文字符识别任务中类别样本的分布不平衡问题，提出了基于 CycleGAN 网络的甲骨文字符数据增广算法，文中详细叙述了算法基本原理以及数据增广流程，并从多个角度的实验与一般增广算法进行对比，证明基于 CycleGAN 网络的甲骨文字符数据增广算法比一般增广算法更能有效提升甲骨文字符识别任务的准确率，最后提出了当前该数据增广算法未来可以改进的方向。



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

# 结　论

甲骨文是当前中国乃至东亚最早的成熟文字体系之一，记录了 3600 多年前的殷商时期的历史，研究甲骨文既有利于探究汉字的发展脉络，同时也对了解中国乃至整个世界的历史有着非常重要的意义，因此甲骨文的研究价值非常高。作为甲骨文字符字形破译的基础，甲骨文字符的检测与识别一直是重要的研究内容，然而目前基于计算机技术的甲骨文字符自动检测工作还非常少，而甲骨文字符自动识别工作仍然比较依赖甲骨学专家的特征工程，对甲骨学知识水平要求较高，且设计的系统也比较复杂。除此之外，甲骨文字形研究领域也缺乏大数据量的、字符级别的甲骨文数据集，这同样也影响了基于计算机技术的甲骨文字形研究工作的发展。基于这些情况，本文做了以下几个方面的工作：

1. 在甲骨文字形检测方面，本文首先构建了甲骨文字符检测数据集 OBCD，这是甲骨学研究领域首个甲骨文单字符检测数据集，为后续深度学习应用于甲骨文字符检测工作提供基准数据库。接着本文构建了基于 R-FCN 和 FPN 深度卷积网络的甲骨字符识别算法框架，并针对甲骨拓片上复合字符难检测且没有足够多的标注良好的训练数据的问题，本文实现了基于字模数据的甲骨文字符动态增广算法，在增加整体数据规模的同时针对性地增加难检字的训练实例，令检测模型更加关注难检字。除此之外，本文实现甲骨文字符识别辅助检测算法，在当前存在较多易与甲骨文字符混淆的甲骨划痕的情况下帮助检测模型减少错误的检测结果。实验证明基于字模数据的甲骨文字符动态增广算法能帮助检测算法更好地预测复合难检字，而在单尺度测试和多尺度测试的情况下，识别模型都能对检测算法的预测结果作出辅助修正，最终 F-measure 指标达到 83.79%。
2. 在甲骨文字形识别方面，本文构建了一个大数据量的甲骨文单字符数据集 OBC306，数据量达到 309551 张，覆盖 306 类不同的甲骨文字符，这是至今数据规模最大的甲骨文单字符数据集，为利用深度学习解决甲骨字形识别提供了丰富的数据。接着接着论文建立了基于 Inception-v4 深度卷积网络的字符识别算法框架，并基于当前任务中类别之间数据不平衡的问题，实现了基于 CycleGAN 网络的甲骨文字符数据增广算法，用以合成更加多样化的甲骨字符图像，平衡类样本分布，解决长尾效应对甲骨字符识别性能的影响。实验证明利用 CycleGAN 合成的增广数据提升了甲骨文



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

字符任务的识别准确率，而且相比一般数据增广算法，本文实现的利用 CycleGAN 的数据增广算法的提升效果更加明显，最终识别准确率达到 86.54%。

本文未来的研究可以有以下几个方面：

1. 在甲骨文字形检测方面，进一步提升增广算法合成的字符实例的真实程度，并处理其他几类难检字符。本文实现的甲骨文动态增广算法能成功帮助检测算法更好地检测难检字，从而提升整体 F-measure 指标。不过目前合成的字符实例在模拟真实风格方面可以进一步改进，与甲骨拓片的融合程度也可以进一步提升。除此之外，当前检测任务中尚有几类较难检测的字符，包括类似汉字“一二三”的甲骨文字符以及处于甲骨拓片边缘或者裂痕边缘的甲骨文字符，前者与甲骨划痕容易混淆，同时相对大小比较小，因此较难正确检测，而后者则因为字形残缺而不容易检测。
2. 在甲骨文字形识别方面，提升生成对抗网络合成的样本的字形多样性。基于 CycleGAN 的数据增广算法已初步取得成效，而由合成样本的可视化可以看出，当前 CycleGAN 生成的样本虽然已具有接近真实样本的风格，但是 CycleGAN 算法本身暂时只能实现甲骨文真实数据风格的迁移，难以对字形本身作出改变，从而限制了合成数据的字形多样性。如果能解决这个问题，使合成样本在具有真实样本风格的同时，拥有更高的字形多样性，那么可以使识别任务的性能进一步提升。



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

# 参考文献

[1] 胡厚宣. 八十五年来甲骨文材料之再统计[J]. 史学月刊, 1984, 5: 15-22.

[2] 陈梦家. 解放后甲骨的新资料和整理研究[J]. 文物参考资料, 1954, 5: 3-8.

[3] 郭沫若，胡厚宣. 甲骨文合集[M]. 北京:中华书局，1978

[4] 中国社会科学院考古研究所. 小屯南地甲骨[M]. 北京:中华书局，1980

[5] 李学勤，齐文心，Sarah Allan. 英国所藏甲骨集[M]. 北京:中华书局，1982

[6] 胡厚宣. 苏德美日所见甲骨集[M]. 成都:四川辞书出版社，1988

[7] 彭邦炯，谢济南，马季凡. 甲骨文合集补编[M]. 北京:语文出版社，1999

[8] 许进雄. 怀特氏等收藏甲骨文集[M]. 多伦多:加拿大皇家安大略博物馆，1979

[9] 松丸道雄. 东京大学东洋文化研究所藏甲骨文字[M]. 东京:东京大学出版会，1983

[10]伊藤道治. 天理大学附属天理参考馆甲骨文字[M]. 奈良:天理大学天理教道友社，1987

[11]董作宾. 甲骨文断代研究例[M]. 中央研究院历史语言研究所, 1965.

[12]陈梦家. 殷墟卜辞综述[M]. 北京:中华书局，1988.

[13]李学勤，彭裕商. 殷墟甲骨分期研究[M].上海:上海古籍出版社. 1996

[14]Cheung C. The Chinese History That Is Written in Bone [EB/OL]. https://www.sapiens.org/archaeology/chinese-oracle-bones-history/. 2018.1.23.

[15]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,

[16]顾绍通. 基于拓扑配准的甲骨文字形识别方法[J]. 计算机工程与应用, 2016, 44(10):2001-2006.

[17]李锋，周新伦. 甲骨文自动识别的图论方法[J]. 电子科学学刊, 1996, 18(S1):41-47.

[18]周新伦，李锋，华星城，韦剑. 甲骨文计算机识别方法研究[J]. 复旦学报(自然科学版), 1996, 35(5):481-486.

[19]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.

[20]CHANT. A computerized database of oracular inscriptions on tortoise shells and bones[DB/OL].http://www.chineseupress.com/chinesepress/promotion/chant/jiaguwen_folder/jiaguwen_info.html. 1996.

[21]刘鹗. 铁云藏龟[M]. 抱残守缺斋（石印本）, 1903.



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

[22]罗振玉.殷虚书契菁华[M]. 北京图书馆出版社, 2000.

[23]于省吾. 双剑誃古器物图录[M]. 中华书局, 1940.

[24]明义士. 殷虚卜辞[M]. 艺文印书馆, 1972.

[25]中央研究院历史语言研究所. 小学堂甲骨文[DB/OL]. http://xiaoxue.iis.sinica.edu.tw/jiaguwen.2013.

[26]Sara Chiesura. Digitising Chinese oracle bones[EB/OL]. https://www.bl.uk/projects/chinese-oracle-bones. 2018

[27]Kim K I, Jung K, Kim J H. Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(12): 1631-1639.

[28]Zhong Y, Zhang H, Jain A K. Automatic caption localization in compressed video[J]. IEEE transactions on pattern analysis and machine intelligence, 2000, 22(4): 385-392.

[29]Zhong Y, Karu K, Jain A K. Locating text in complex color images[J]. Pattern recognition, 1995, 28(10): 1523-1535.

[30]Gllavata J, Ewerth R, Freisleben B. Text detection in images based on unsupervised classification of high-frequency wavelet coefficients[C]. Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. IEEE, 2004, 1: 425-428.

[31]Ali S A, Hashim A T. Wavelet transform based technique for text image localization[J]. Karbala International Journal of Modern Science, 2016, 2(2): 138-144.

[32]Lee S H, Seok J H, Min K M, et al. Scene text extraction using image intensity and color information[C]. 2009 Chinese Conference on Pattern Recognition. IEEE, 2009: 1-5.

[33]Liu Z, Sarkar S. Robust outdoor text detection using text intensity and shape features[C].2008 19th International Conference on Pattern Recognition. IEEE, 2008: 1-4.

[34]Evgeniou T, Pontil M. Support vector machines: Theory and applications[C]. Advanced Course on Artificial Intelligence. Springer, Berlin, Heidelberg, 1999: 249-257.

[35]Bradski G R. Real time face and object tracking as a component of a perceptual user interface[C]. Proceedings Fourth IEEE Workshop on Applications of Computer Vision. WACV'98 (Cat. No. 98EX201). IEEE, 1998: 214-219.

[36]B. Epshtein, E. Ofek, and Y. Wexler. Detecting text in natural scenes with stroke width transform[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010: 963-2970.

[37]Neumann L, Matas J. A method for text localization and recognition in real-world



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

images[C].Asian Conference on Computer Vision. Springer, Berlin, Heidelberg, 2010: 770-783.

[38]Liu Y, Goto S, Ikenaga T. A contour-based robust algorithm for text detection in color images[J]. IEICE transactions on information and systems, 2006, 89(3): 1221-1230.

[39]Pan Y F, Hou X, Liu C L. A hybrid approach to detect and localize texts in natural scene images[J]. IEEE transactions on image processing, 2011, 20(3): 800-813.

[40]Liao M, Shi B, Bai X, et al. Textboxes: A fast text detector with a single deep neural network[C].Thirty-First AAAI Conference on Artificial Intelligence. 2017.

[41]Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C].European conference on computer vision. Springer, Cham, 2016: 21-37.

[42]Liu Y, Jin L. Deep matching prior network: Toward tighter multi-oriented text detection[C].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1962-1969.

[43]Zhou X, Yao C, Wen H, et al. EAST: an efficient and accurate scene text detector[C].Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2017: 5551-5560.

[44]Kim K H, Hong S, Roh B, et al. Pvanet: Deep but lightweight neural networks for real-time object detection[J]. arXiv preprint arXiv:1608.08021, 2016.

[45]吕肖庆，李沫楠，蔡凯伟，王晓唐，英敏. 一种基于图形识别的甲骨文分类方法[J]. 北京信息科技大学学报, 2010, 25(Z2):92-96.

[46]Guo J, Wang C, Roman-Rangel E, et al. Building hierarchical representations for oracle character and sketch recognition[J]. IEEE Transactions on Image Processing, 2016, 25(1): 104-118.

[47]高峰,吴琴霞,刘永革,熊晶. 基于语义构件的甲骨文模糊字形的识别方法[J].科学技术与工程,2014,14(30):67-70.

[48]刘永革, 刘国英. 基于 SVM 的甲骨文字识别[J]. 安阳师范学院学报, 2017 (2): 54-56.

[49]Meng L. Recognition of Oracle Bone Inscriptions by Extracting Line Features on Image Processing[C]. ICPRAM. 2017: 606-611.

[50]刘永革, 李雪山. 甲骨文数字化平台建设[EB/OL]. http://www.guoxue.com/?p=5338. 2009.8.18.

[51]Tzutalin. LabelImg[CP/OL]. https://github.com/tzutalin/labelImg. 2015.

[52]Everingham M, Van Gool L, Williams C K I, et al. The PASCAL visual object classes



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

challenge 2007 (VOC2007) results[J]. 2007.

[53]Jones E, Oliphant E, Peterson P, et al. SciPy: Open Source Scientific Tools for Python, 2001. http://www.scipy.org/

[54]Dai J, Li Y, He K, et al. R-fcn: Object detection via region-based fully convolutional networks[C].Advances in neural information processing systems. 2016: 379-387.

[55]Lin T Y, Dollár 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.

[56]Neubeck A, Van Gool L. Efficient non-maximum suppression[C].18th International Conference on Pattern Recognition (ICPR'06). IEEE, 2006, 3: 850-855.

[57]Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C].Thirty-First AAAI Conference on Artificial Intelligence. 2017.

[58]沈建华，曹锦炎. 甲骨文字形表[M]. 上海:上海辞书出版社，2008

[59]Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C].Advances in neural information processing systems. 2014: 2672-2680.

[60]Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.

[61]Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C].Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1125-1134.

[62]Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C].Proceedings of the IEEE International Conference on Computer Vision. 2017: 2223-2232.

[63]Mejjati Y A, Richardt C, Tompkin J, et al. Unsupervised Attention-guided Image-to-Image Translation[C].Advances in Neural Information Processing Systems. 2018: 3697-3707.

[64]Arjovsky M, Chintala S, Bottou L. Wasserstein gan[J]. arXiv preprint arXiv:1701.07875, 2017.

[65]Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C].Advances in neural information processing systems. 2012: 1097-1105.

[66]Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

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

[68]Dalal N, Triggs B. Histograms of oriented gradients for human detection[C].international Conference on computer vision & Pattern Recognition (CVPR'05). IEEE Computer Society, 2005, 1: 886--893.

[69]Paszke A, Gross S, Chintala S, et al. Automatic differentiation in pytorch[J]. 2017

[70]Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex[J]. The Journal of physiology, 1962, 160(1): 106-154.

[71]Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J]. Biological cybernetics, 1980, 36(4): 193-202.

[72]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.



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

# 攻读硕士学位期间取得的研究成果

一、已发表（包括已接受待发表）的论文，以及已投稿、或已成文打算投稿、或拟成文投稿的论文情况（只填写与学位论文内容相关的部分）：

| 序号 | 作者（全体作者，按顺序排列） | 题　　目 | 发表或投稿刊物名称、级别 | 发表的卷期、年月、页码 | 相当于学位论文的哪一部分（章、节） | 被索引收录情况 |
|---|---|---|---|---|---|---|
| 1 | Haobin Wang, Shuangping Huang, Lianwen Jin | Focus on Scene Text Using Deep Reinforcement Learning | 2018 24th International Conference on Pattern Recognition | 2018-August, 2018.11. pp. 3769-3765 | 第三章 | EI |
| 2 | Shuangping Huang, Zhuoyao Zhong, Lianwen Jin, Shuye Zhang, Haobin Wang | DropRegion training of inception font network for high-performance Chinese font recognition | Pattern Recognition | 77, 2018.5. pp. 395-411 | 第四章 | SCI |
| 3 | Shuangping Huang, Haobin Wang, Yongge Liu, Xiaosong Shi, Lianwen Jin | OBC306: A Large Oracle Bone Character Recognition Dataset | 2019 15th International Conference on Document Analysis and Recognition | 已接受 | 第四章 | EI |

二、与学位内容相关的其它成果（包括专利、著作、获奖项目等）

[1] 发明专利：基于深度强化学习的场景文本检测方法及系统

发明人：黄双萍，王浩彬，金连文

专利受理号：201711352220.0

进入实审日期：2018 年 5 月 29 日



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

# 致　　谢

三年时光转瞬即逝，不知不觉之间时间已经来到硕士生涯的最后一年，华园依旧木棉飘红，但此时此刻的我却即将告别母校。

感谢我的导师黄双萍老师，在这三年时间里不管是学术知识还是为人处事方面，都教会了我许多东西。黄老师治学严谨，对每一个学生都认真负责，对我和其他师兄师弟都非常关心。在我撰写毕业论文的过程中，黄老师在百忙之中抽出时间悉心指导我，让我得以完成论文的撰写。在这里我衷心感谢黄老师在这段时间的辛劳，学生铭记于心。

感谢河南省安阳师范学院、甲骨文信息处理教育部重点实验室与河南省甲骨文信息处理重点实验室，为甲骨文字符检测与识别课题提供开放课题支持。

感谢我在实验室的小伙伴们,这三年时光里有你们在身边一起研究讨论,一起聚餐,一起欢笑，为我的硕士生活增添了许多色彩，谢谢你们，能遇见你们真好。

感谢我的父母和兄弟，无论我身处何方，你们都会关心我；无论我遇到什么事情、做什么决定，你们都会尊重支持我，你们是我避风的港湾，也是我前进的动力，谢谢你们，我不会辜负你们的期望，未来一定会好好报答你们。

论文的撰写即将结束，学生生涯也即将完结，我希望在我离开学校之后，能继续奋斗，成为一名更优秀的人，将来回报社会，回报国家！



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

**3.答辩委员会对论文的评语**

（主要内容包括：1.对论文的综合评价；2.对论文主要工作和创造性成果的简要介绍；3.对作者掌握基础理论、专业知识程度、独立从事科研工作能力以及在答辩中表现的评价；4.存在的不足之处和建议；5.答辩委员会结论意见等）

论文《基于深度学习的甲骨文检测与识别研究》选题前沿，具有较好的学术价值与应用价值。

该论文的主要贡献有：

1. 构建了首个甲骨拓片检测数据集并进行分析总结，搭建了基于深度学习的检测算法框架，针对甲骨检测难点提出了基于字模数据的动态增广算法与识别辅助检测算法，并使甲骨文检测性能有所提升。

2. 构建了当前最大的甲骨文单字符识别数据集，分析总结了甲骨字符字形识别的难点。针对甲骨文识别任务中类别样本分布不均的问题，提出了基于 CycleGAN 的甲骨文字符扩充算法，解决分布长尾效应对识别性能的影响，提升了甲骨字符识别精准度。

论文撰写规范，结构合理，论证严谨。在答辩过程中，作者表达清楚，条理清晰，回答问题准确。

经答辩委员会讨论表决，一致认为该生的论文达到了硕士学位论文水平，同意通过论文答辩，建议授予工学硕士学位。

论文答辩日期：2019 年 6 月 4 日　　答辩委员会委员 4 人

表决票数：同意毕业及授予学位（4）票；

同意毕业，但不同意授予学位（0）票；

不同意毕业（0）票

表决结果（打"√"）：通过（√）；不通过（　）

决议：同意授予硕士学位（√）　不同意授予硕士学位（　）

| | |
|---|---|
| 答辩成员签名 | 刘仲民（主席） 薛涛 黄双萍 |
| 答辩秘书签名 | |



(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net