

## 摘要

甲骨文是商周晚期王室用于占卜和记载事件的文字，因被刻写在龟甲和兽骨上而得名，是中国乃至整个东亚已知最早的成熟且系统化的文字形式。目前全球所发现的殷商时期甲骨文字符不足五千个，但其中真正被解读和理解的字符数量仅有两千余字。借助计算机对甲骨片上甲骨文字符进行识别和检测，对于甲骨文识别、考古、文化探源等领域的深入研究具有极其重要的意义。

甲骨文字符识别的目标是对甲骨文中的单个字符进行精确的类别分析，依托已有的研究成果，为揭示更多的字符变种提供坚实的基础。而甲骨文字符检测的任务则是从甲骨片图像中精确定位字符的位置，这不仅简化了从大量图像中提取甲骨文字的过程，而且为后续深入研究提供了极大的便利。

线性注意力（Linear Attention）是一种降低计算复杂度、降低模型参数量的简单有效的方式。本研究使用 Transformer 架构，引入线性注意力，提出一种新的自注意力计算方式，能够显著降低模型参数量，并使用该注意力计算方法构建一种新的主干网络，基于上述内容完成甲骨文字符识别和甲骨文字符检测任务。本文围绕甲骨文字符识别和甲骨文字符检测两项基本任务开展研究，主要工作包有：

1. 本文提出一种新的线性注意力计算方法 MS Attention，将该方法与金字塔视觉 Transformer 相结合，构建一个多尺度特征提取网络 Linear-PVT，减少了模型参数量，提高了甲骨文字符识别的准确性和效率，且其提取多尺度特征的能力为之后甲骨文字符检测任务提供支持。在公开甲骨文识别数据集 OBC306 上验证，本方法取得92.31%的分类准确率。

2. 提出一种新的甲骨文字符检测算法，该算法基于 Linear-PVT 主干网络，结合 GiraffeDet 的 Neck 网络设计，并使用 Mask-RCNN 的检测头来共同构建检测模型。在已公开的甲骨文字检测数据集上，该检测算法的 mAP 为0.459。通过引入 CBAM 注意力模块的方式，将 Neck 网络中的特征融合方式由简单拼接，改为利用 CBAM 注意力模块进行动态特征融合。CBAM 模块包含通道和空间注意力机制，其中通道注意力机制是关注通道方向上各通道间的重要程度，空间注意力机制关注的是图像中的重要区域。通过通道和空间注意力机制有效地加强了模型对于关键特征的识别能力，从而使得特征融合过程更加精细和高效。这种基于注意力的特征融合方式不仅提升了甲骨文字符的检测精

度,而且增加了模型对复杂背景下甲骨文字符检测的鲁棒性。实验结果表明,引入 CBAM 后,甲骨文字符检测的 mAP 有了显著提高,达到0.483,证明了该方法在甲骨文字符检测任务上的有效性和优越性。

关键词: 甲骨文, 字符识别, 字符检测, 线性注意力, 深度学习

# 目 录

摘 要.....	III
Abstract.....	V
第 1 章 引 言.....	1
1.1 研究背景与意义.....	1
1.2 研究现状.....	3
1.2.1 甲骨文识别方法.....	3
1.2.2 甲骨文检测方法.....	4
1.2.3 数据增强方法.....	4
1.3 主要研究工作.....	4
1.4 文章组织架构.....	6
第 2 章 相关理论和技术.....	7
2.1 视觉 Transformer 模型.....	7
2.2 线性注意力.....	10
2.3 特征融合网络.....	11
2.4 扩散模型.....	12
2.5 本章小结.....	13
第 3 章 基于线性注意力的甲骨文识别模型.....	14
3.1 甲骨文识别模型.....	14
3.1.1 线性注意力算法.....	14
3.1.2 甲骨文识别模型结构.....	16
3.2 数据集处理.....	17
3.2.1 数据集介绍.....	18
3.2.2 有条件扩散模型原理.....	19
3.2.3 基于有条件扩散模型的数据扩充算法.....	20
3.3 实验结果与结论.....	21
3.3.1 实验环境与评价指标.....	21
3.3.2 自然图像数据集实验.....	22

3.3.3 甲骨文识别数据集实验.....	25
3.4 本章小结.....	26
第4章 基于注意力特征融合的甲骨文检测模型.....	28
4.1 特征融合网络.....	28
4.1.1 基于 GFPN 的多层特征融合网络.....	28
4.1.2 引入注意力机制的 GFPN 特征融合网络.....	31
4.2 甲骨文检测模型框架.....	33
4.2.1 目标检测头.....	33
4.2.2 整体框架.....	34
4.3 实验结果与结论.....	35
4.3.1 实验数据集.....	35
4.3.2 评价指标.....	36
4.3.3 实验结果及分析.....	36
4.4 本章小结.....	38
第5章 总结与展望.....	40
5.1 研究总结.....	40
5.2 未来工作展望.....	41
参考文献.....	42
致  谢.....	49
攻读学位期间取得的科研成果清单.....	50

# 第1章 引言

## 1.1 研究背景与意义

甲骨文，也被称作契文或甲骨卜辞，是商代末期王族用于占卜及纪事目的，在龟甲与兽骨上刻录的文字体系。作为中国及东亚地区已知最古老的系统化文字，甲骨文的发现对于汉字的研究具有里程碑意义。迄今为止，考古学者们已挖掘并发现超过十五万片卜骨，这些遗物不仅记录了当时的占卜实践，也涵盖了多个邦国与部族的符号及称谓。与《说文解字》中所著录定型古汉字符号系统相较，甲骨文展现了汉字构型的多样性，昭示了殷商晚期时，甲骨文字已在构型方式上达到成熟。商朝晚期的甲骨文，从字形结构、符号化水平、书写形态及其功能性等多个维度来看，均表现出了其文字系统经过了长期演变、结构完善且功能全面的特点，证明汉字已进入了一个成熟的发展阶段，并形成了一个完整的文字体系。甲骨文所记载的内容不仅丰富了商王朝的历史文献，也为《史记》中《殷本纪》的文字记载提供了实物证据，明确了殷墟的具体位置，并将中国的可考历史年代向前推了一千年。

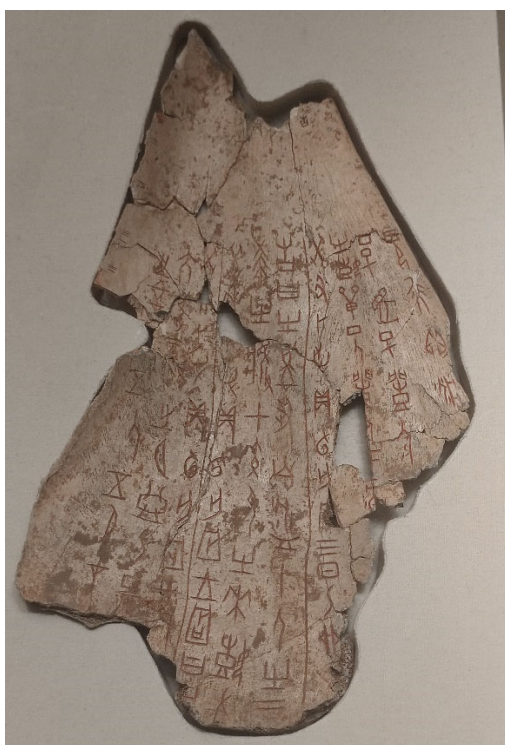


图 1.1 刻有甲骨文的甲骨骨片

习近平总书记高度重视数字技术对中华优秀传统文化的保护传承重要作用，指出要运用现代科技手段加强古籍典藏的保护修复和综合利用<sup>[1]</sup>。《关于实施中华优秀传统文化传承发展工程的意见》<sup>[2]</sup>明确强调了要推进数字化保存和传播。因此，随着当下以大数据、人工智能为代表的新技术的不断发展，必须高效地利用这些新技术，使其成为中华优秀传统文化传承发展的关键动力。

习近平总书记在甲骨文发现和研究 120 周年的贺信中提到<sup>[3]</sup>：甲骨文是汉字的源头和中华优秀传统文化的根脉，值得倍加珍视、更好传承发展。要确保甲骨文等古文字研究“有人做、有传承”。故深入研究甲骨文的历史和文化价值，不仅有利于中华文明传承、发展，还能够为促进文明交流互鉴，推动人类社会进步做出巨大贡献。在殷墟，随着数字技术被运用到甲骨文研究中，不仅建设了甲骨文大数据平台，还利用深度学习技术对甲骨碎片进行自动缀合。随着甲骨文材料的数字化，对这些古文字符的识别与翻译研究亦步步深入，特别是通过整理甲骨文拓片印制品等资料，构建了甲骨文文字数据库。在此过程中，部分甲骨文的语义已获得诠释。然而，此类研究成果大多依赖于专家的手工分析，这不仅要求较高的专业知识水平，也提升了研究成本，进而限制了甲骨文识别与解读的效率。根据中国国家文物局举行的“考古中国”重大项目进展工作会所介绍的甲骨文发现与研究 120 周年以来的重要成果<sup>[4]</sup>，我国已发掘出逾十五万片甲骨片<sup>[5,6]</sup>，记录有超过四千个不同的甲骨文单字，涉及商朝时期政治生活等各方面内容，但其中仅约两千余字符被成功解读。由于未来可能有更多的甲骨材料被发现，未知字符或变体的出现亦在所难免。因此，开发一种高效的甲骨文自动检测与识别系统，对降低研究成本具有重大意义。

近年来研究表明，视觉 Transformer (Vision Transformer, ViT)<sup>[7]</sup>能够在计算机视觉的各种任务中达到先进水平。其全局自注意力的计算，使模型能够捕捉图像中的长距离依赖关系，增强了特征提取的效率和图像识别的准确性。并且，自注意力机制有助于理解复杂场景，对目标检测和分割任务有所帮助，为甲骨文字符图像的研究提供了新视角。已有的、标注完备的甲骨文识别数据集如 OBC306<sup>[8]</sup>、甲骨文检测数据集<sup>[9]</sup>，为利用深度学习技术精确识别和检测甲骨文字符提供了坚实的数据基础。借助 Transformer 模型的卓越性能，构建一套甲骨文字符识别与检测系统，通过创新自注意力机制计算方法进一步提升甲骨文字符识别与检测的精度，并通过优化，减少模型参数和计算复杂度，有效

降低甲骨文研究的成本，加速数据处理过程，提高整体的研究效率。

## 1.2 研究现状

### 1.2.1 甲骨文识别方法

传统的甲骨文识别方法通常基于甲骨文字形结构或图形的拓扑特征进行特征提取，从而对甲骨文字符进行分类。周新伦<sup>[10]</sup>和李峰<sup>[11]</sup>将甲骨文字符转化为点和线，构成无向图。对该无向图的顶点数、块数、邻接子图的割集等图论特征进行编码，通过比较编码序列识别甲骨文。吕肖庆等人<sup>[12]</sup>提出基于曲率直方图的傅里叶描述子，对每个甲骨文字符计算特征向量后，通过相似度计算进行字符分类。顾邵通<sup>[13]</sup>依据文字具有拓扑特征的稳定性，利用拓扑配准的方法，计算甲骨文字符在拓扑特征之间的欧氏距离，实现基于拓扑结构的甲骨文字符配准。上面基于特征编码的分类方法依旧有局限性。特征编码对甲骨文文字中的异体字识别能力较弱，且存在特征编码重码较多、效率较差的问题。

随着机器学习的不断发展，甲骨文字识别任务也引入了机器学习方法。史小松<sup>[14]</sup>构建一个简单的甲骨文数据库，在此基础上使用支持向量机（Support Vector Machine, SVM），针对甲骨文字的部首进行分类。刘永革等人<sup>[15]</sup>通过对甲骨文字符进行分块，采用分块直方图提取字符特征，使用 SVM 进行识别，最终分类准确率达到 88%。基于机器学习的方法虽然提高了甲骨文识别的精度，但是模型依然需要相关领域专家手动设计特征作为模型输入，同时识别算法的步骤也较为复杂。

基于深度学习的甲骨文识别方法近年来也是研究的重点。刘芳等人<sup>[16]</sup>基于 Mask R-CNN 模型进行甲骨文拓片的识别，闫升等人<sup>[17]</sup>进一步改进 Mask R-CNN，结合类别屏蔽和自动识别矫正，首次对甲骨拓片图像同时进行甲骨文的识别与检测任务。林小渝<sup>[18,19]</sup>提出从甲骨文偏旁角度进行甲骨文识别，其方法提高了识别准确率，还有利于帮助发掘未知的甲骨文字符。Guo 等人<sup>[20]</sup>基于 Inception-v3 模型<sup>[21]</sup>进行甲骨文识别，相较于 AlexNet<sup>[22]</sup>、VGG-19<sup>[23]</sup>等模型性能更好，在字符图像存在模糊、遮挡等情况下依旧有良好的表现。Du 等人<sup>[24]</sup>提出一种双分支自监督模型，通过在甲骨文字符数据集上进行自监督训练，提高甲骨文字符识别准确率。张颐康等<sup>[25]</sup>提出一种跨模态的深度学习技术，在卷积神经网络（CNN）的基础上，配合临摹、基于文字拓扑的特征编码，实现跨模态的特征空间建模，与单模态识别方法相比，多模态方法的识别准确率有很大提升。

### 1.2.2 甲骨文检测方法

近年来,基于深度学习的甲骨文字符检测方法的研究成果颇丰。Liu<sup>[26]</sup>优化 Faster R-CNN 模型,解决其负样本较多的问题,提高了甲骨文检测的精度。Li 等人<sup>[27]</sup>改进 SSD 模型,提高该模型在检测小目标时的准确率,提高了检测有摩擦甲骨文字符的精度。Wang 等人<sup>[28]</sup>使用 YOLOv4 模型在甲骨拓片图像上进行甲骨文字符检测。陈双浩<sup>[29]</sup>基于沙漏网络 (Hourglass Network) 采用非锚点框设计,使用自适应形状高斯核表示甲骨文字符的空间区域,将甲骨文字符的检测转换为对应高斯图的预测。王浩彬构建基于 R-FCN 和 FPN 的检测算法框架,在甲骨文检测数据集 OBCD 上进行甲骨文检测。Zhang 等人<sup>[30]</sup>构建了大规模数据集 OracleBone-8000 并采用基于深度学习的场景文本检测算法进行甲骨文定位。

### 1.2.3 数据增强方法

已发现的甲骨文字符中,存在部分罕见字,使得甲骨文字符数据集存在数据不平衡的问题,整体呈现长尾效应。为了在小样本的情况下进行甲骨文字符的识别,Han 等人<sup>[31]</sup>提出通过自监督学习预训练的数据增强方法 Orc-Bert。该方法利用在大规模未标记中文字符上训练的 BERT 模型来生成少量甲骨文字样本的增强样本。该方法提高了甲骨文识别的性能,还解决了因数据有限导致的数据不平衡所带来的挑战。该方法在训练时,需要大量未标注数据来训练模型,无法直接利用已有的有标注小样本数据集进行训练。针对该问题,毕晓君等人<sup>[32]</sup>提出基于有监督对比学习的小样本甲骨文识别方法,可以直接使用有标注小样本数据集进行识别,有效提升识别效果。张颐康等<sup>[25]</sup>开发了一种新颖的度量学习方法,这种方法采用三元组损失函数,并借鉴了对抗训练的原理,实现了将拓片上的甲骨文字符与临摹的甲骨文字符映射至同一特征空间。这一方法巧妙地通过隐式方式解决了甲骨文识别任务中经常遇到的类别不均衡问题。Li 等<sup>[33]</sup>从数据增强的角度出发,提出了一种创新的混合式数据增强策略。该策略通过在训练阶段将同一类别内的不同样本特征进行随机组合,创造出新的样本特征,有效地促进了数据分布的均衡化。王浩彬<sup>[34]</sup>引入了循环式生成对抗网络算法,这一算法在缺乏配对图像的条件下,能够使网络学习拓片甲骨文字的独特风格,并生成具有高度多样性的字符图像。

## 1.3 主要研究工作



本研究主要针对甲骨文字符识别任务和甲骨文检测任务设计新的深度学习方法，提高识别和检测精度。

甲骨文字符的识别和检测任务中有以下几个难点：首先，甲骨文字符的边缘等特征明显，但因甲骨埋藏时间久远，甲骨拓片存在裂纹、划痕等与字符本身无关的标记，为样本图像添加更多噪声影响。除此之外，还存在部分字符出现破损、磨损等情况，导致甲骨文字符不完整或者分散在多个碎片中。这些情况对后面的识别和检测工作带来一定困难。其次，被发现并识别出来的单字有限，且甲骨文在发展的过程中，可能会出现一字多形的异体字情况，因此在数据集制作时，需要大量甲骨文领域专业知识辅助，数据集标注成本高且很难形成较大规模的数据集。不仅如此甲骨文数据集中样本数量不平均的问题严重，这些问题为模型训练带来较大的障碍。本文为克服以上难点，做出如下工作。

本文的甲骨文识别研究基于 OBC-306 数据集进行训练。数据集中数据分布不平衡，整体呈现长尾分布。为减轻长尾分布所带来的模型训练困难问题，考虑使用数据扩充的方式，对数据集的数据分布进行平衡。近年来，扩散模型（Diffusion Model）<sup>[35]</sup>在图像生成领域达到了强于生成对抗网络的效果，因此考虑使用条件可控的扩散模型 Iterative Latent Variable Refinement（ILVR）<sup>[36]</sup>，通过添加参考图像信息作为条件约束的方式对数据集指定类别进行扩充。

对于甲骨文字识别任务，本文引入视觉 Transformer 作为骨干网络。但视觉 Transformer 参数量多、训练时对数据量要求高。因此本文使用线性注意力机制，并提出新的线性注意力计算方法 MS Attention，使用核函数和 Sigmoid 函数，对自注意力中的 Softmax 函数进行拟合，改变自注意力中矩阵乘法计算的顺序来减少模型参数量，提高计算效率。将这一方法与金字塔视觉 Transformer（Pyramid Vision Transformer）<sup>[37,38]</sup>相结合，构建出一个新的多尺度特征提取网络 Linear-PVT，从而提高甲骨文字符识别的准确性和效率。并且 Linear-PVT 拥有多尺度特征提取能力，将其作为主干网络时，为其添加多尺度特征融合网络和检测头可以直接作为甲骨文检测任务的模型。

针对甲骨文字符检测任务，本文依托 Linear-PVT 构建检测模型。为减小拓片图像数据中的划痕等噪声问题，使用大特征融合网络（Neck 网络），提高模型对特征的提取和融合能力。结合 GiraffeDet<sup>[39]</sup>的 Neck 网络设计，使用 Mask-RCNN<sup>[40]</sup>的检测头，提出一

种新的甲骨文字符检测算法。为了提升模型的鲁棒性，引入了 CBAM (Convolutional Block Attention Module) 注意力模块<sup>[41]</sup>，该模块通过空间和通道注意力机制，优化了 Neck 网络中的特征融合方式。将原本简单的特征拼接方法替换为基于 CBAM 的动态特征融合策略，这一改进显著增强了模型对关键特征的识别能力。通过空间注意力机制，模型能够聚焦于图像中的关键区域，而通道注意力机制则评估并强调了各个通道在特征表达中的重要性。这种双重注意力机制的应用不仅提升了对甲骨文字符的检测精度，还增强了模型在面对复杂背景时的鲁棒性。

#### 1.4 文章组织架构

第 1 章绪论，介绍了本研究的研究背景和研究方式，阐述了甲骨文识别和检测的研究现状，说明了本研究的主要工作和创新点。

第 2 章相关理论和技术，介绍了与本研究提出的方法所相关的理论知识，从扩散模型、线性注意力和特征融合网络三方面进行介绍。

第 3 章对使用条件可控的扩散模型扩充数据集进行叙述，介绍主要方法和算法流程，最后对生成数据进行展示。然后对 Linear-PVT 模型进行详细阐述，介绍改进的线性注意力——MS Attention 的计算方式，并通过实验和理论分析表明方法的有效性。

第 4 章对改进特征融合网络进行说明，说明对 GiraffeDet 的 Neck 网络进行改进的方法，并在甲骨文检测数据集上实验，证明其检测精度有所提升。

第 5 章结论和未来展望，该部分对本研究进行了相关结论和成果的总结，针对不足之处，为之后甲骨文识别和检测的相关工作进行展望。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/548035071007007005>